

US007736851B2

(12) United States Patent

Stewart et al.

(54) DNA CLONING METHOD

- (75) Inventors: Francis Stewart, Leimen (DE);
 Youming Zhang, Heidelberg (DE);
 Frank Buchholz, Bremen (DE)
- (73) Assignee: Europaisches Laboratorium fur Molekularbiologie (EMBL), Heidelberg (DE)
- (*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 611 days.

This patent is subject to a terminal disclaimer.

- (21) Appl. No.: 10/842,534
- (22) Filed: May 11, 2004

(65) **Prior Publication Data**

US 2004/0203057 A1 Oct. 14, 2004

Related U.S. Application Data

(62) Division of application No. 10/231,013, filed on Aug. 30, 2002, now Pat. No. 6,787,316, which is a division of application No. 09/555,510, filed as application No. PCT/EP98/07945 on Dec. 7, 1998, now Pat. No. 6,509, 156.

(30) Foreign Application Priority Data

Dec. 5, 1997	(EP)	 97121462
Oct. 5, 1998	(EP)	 98118756

- (51) Int. Cl. *C12Q 1/68* (2006.01) *C07H 21/02* (2006.01) *C07H 21/04* (2006.01)
- (52) U.S. Cl. 435/6; 536/23.1; 536/24.3
- (58) **Field of Classification Search** None See application file for complete search history.

(10) Patent No.: US 7,736,851 B2

(45) **Date of Patent:** *Jun. 15, 2010

(56) **References Cited**

U.S. PATENT DOCUMENTS

FOREIGN PATENT DOCUMENTS

WO WO 95/22625 8/1995

OTHER PUBLICATIONS

Murphy, K.C., "Use of Bacteriophage Lambda Recombination Functions to Promote Gene Replacement in *Escherichia coli*", J. Bacter., vol. 180, pp. 2063-2071 (1998).*

Takahashi et al., PNAS USA, vol. 87, pp. 2790-2794 (1990).*

Nussbaum, A. and Cohen A., Use of Bioluminescence Gene Reporter for the Investigation of Red-dependent and Gam-dependent Plasmid Recombination in *Escherichia coli* K12, J. Mol. Biol., vol. 203, pp. 391-402 (1988).*

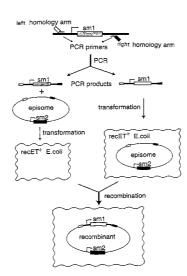
(Continued)

Primary Examiner—Teresa E Strzelecka (74) Attorney, Agent, or Firm—Rothwell, Figg, Ernst & Manbeck, P.C.

(57) ABSTRACT

The invention relates to methods for cloning DNA molecules using recE/recT-mediated homologous recombination mechanism between at least two DNA molecules where one DNA molecule is a circular or linear DNA molecule and the second DNA molecule is a circular DNA molecule, and the second DNA molecule contains two regions with sequence homology to the first DNA molecule. Competent cells and vectors are also described.

24 Claims, 65 Drawing Sheets



OTHER PUBLICATIONS

Nussbaum A. et al., Genetics, vol. 130, pp. 37-49 (1992).*

Degryse et al., "Homologous recombination based modification in *Escherichia coli* and germ line transmission in transgenic mice of a bacterial artificial chromosome", Nature Biotechnology, vol. 15, Sep. 1997, pp. 859-865.

Degryse et al., "Evaluation of *Escherichia coli* recBC sbcBC mutants for cloning by recombination in vivo", Journal of Biotechnology, vol. 2, No. 39, Apr. 15, 1995, pp. 181-187.

GeneBank Accession No. V00638, "Lambda genomes from map unit 74 backward to map unit 67".

Hall, S.D., et al., "Identification and Characterization of the *Escherichia coli* RecT Protein, a Protein Encoded by the recE Region That Promotes Denaturation of Homologous Single-Stranded DNA", Journal of Bacteriology, Washington, D.C., vol. 175, No. 1, Jan. 1993, p. 277-287.

Hall, D.S. et al., "Homologous Pairing and strand exchange promoted by the *Escherichia coli* RecT protein", PNAS USA, vol. 91, pp. 3205-3209 (1994).

Kolodner et al., "homologous pairing proteins encloded by the *Escherichia coli* recE and recT genes" Molecular Microbiology, vol. 11, No. 1, 1994, pp. 23-30.

Luisi-Deluca et al., "Genetic and physical analysis of plasmid recombination in recB recC sbcB and recB recC sbcA *Escherichia coli* K-12 mutants", Genetics, vol. 122, 1989, pp. 269-278.

Murphy, "Lambda Gam protein inhibits the helicase and Chi-stimulated recombination activities of *Escherichia coli* RecBCD enzyme," Journal of Bacteriology, vol. 173, No. 18, Sep. 1991, pp. 5808-5821.

Nussbaum et al., "Restrictuion-stimulated homologous recombination of plasmids by the RecE pathway of *Escherichia coli*", Genetics, vol. 130, No. 1, Jan. 1992, pp. 37-49. Oliner et al., "In vivo cloning of PCR Products in *E. coli*", Nucleic Acids Research, vol. 21, No. 22, pp. 5192-5197 (1993). Stratagene Catalog, p. 39 (1988).

A. Clark, et al.: "Genetic and Molecular Analyses of the C-Terminal Region of the recE Gene from the Rac Prophage of *Escherichia coli* K-12 Reveal the recT Gene," J. Bacteriol., vol. 175, No. 23, pp. 7673-7682, Dec. 1993.

NCBI Accession No. M24905.1, Search Pages, 2 pages.

NCBI Accession No. L23927.1, Search Pages, 3 pages.

NCBI Accession No. J02459 M17233 M24325 V00636 00906, Search Pages, 37 pages.

TaKaRa Biotechnology Guide: "Guide Genetic Engineering Research," pp. 163-169, 1994-1995.

B. Sauer, et al.: "Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1," Proc. Natl. Acad. Sci. USA, vol. 85, pp. 5166-5170, Jul. 1988.

Y. Sakaki, et al.: "Purification and Properties of the γ -Protein Specified by Bacteriophage λ : An Inhibitor of the Host RecBC Recombination Enzyme," Proc. Nat. Acad. Sci. USA, vol. 70, No. 8, pp. 2215-2219, Aug. 1973.

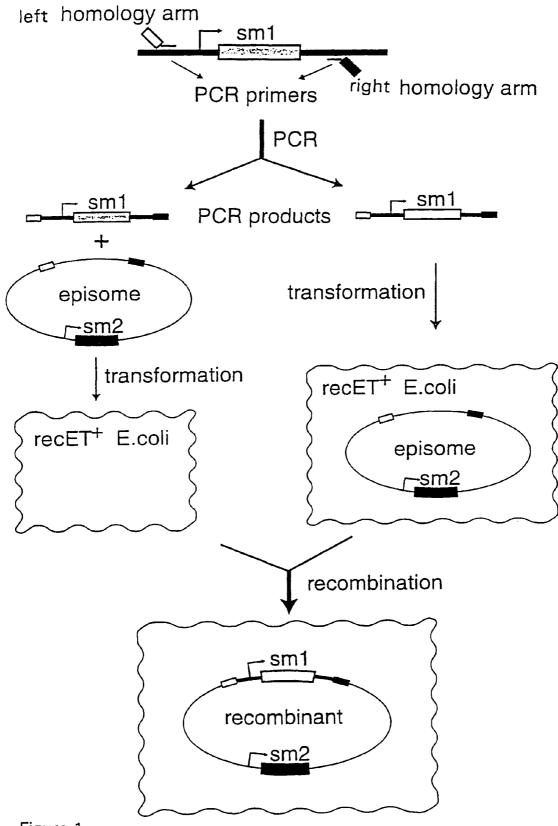
A. Poteete, et al.: "Activation of RecF-Dependent Recombination in *Escherichia coli* by Bacteriophage λ - and P22-Encoded Functions," J. Bacteriol., vol. 170, No. 9, pp. 4379-4381, Sep. 1988.

A. Poteete, et al.: "Modulation of *Escherichia coli* RecBCD Activity by the Bacteriophage λ Gam and P22 Abc Functions," J. Bacteriol., vol. 171, No. 5, pp. 2012-2021, May 1988.

M. Behme, et al.: "Postinfection Control by Bacteriophage T4 of *Escherichia coli* recBC Nuclease Activity," J. Virol., vol. 18, No. 1, pp. 20-25, Apr. 1976.

K. Murphy, et al.: "Properties of *Escherichia coli* Expressing Bacteriophage P22 Abc (Anti-RecBCD) Proteins, Including Inhibition of Chi Activity," J. Bacteriol., vol. 175, No. 6, pp. 1756-1766, Mar. 1993.

* cited by examiner





Three ways to select recombinants

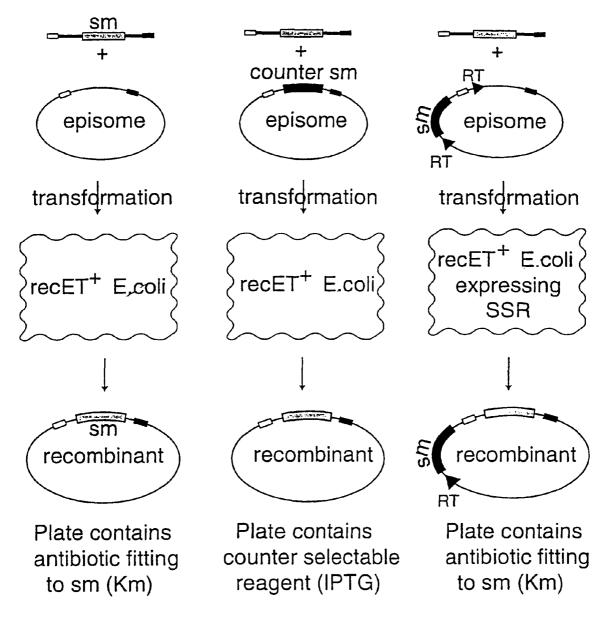
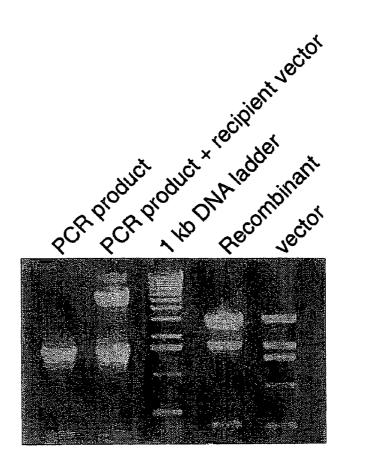
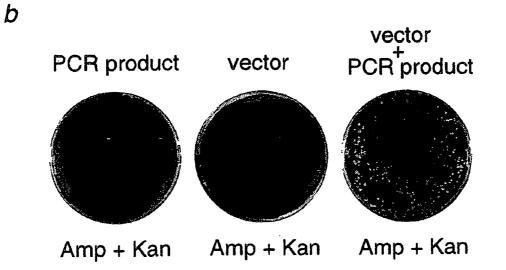


Figure 2



а





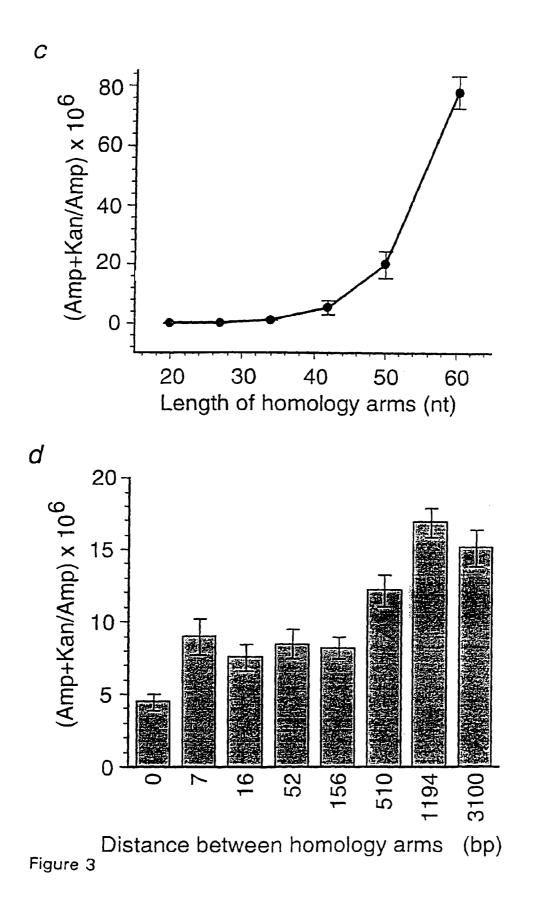


Figure 4a

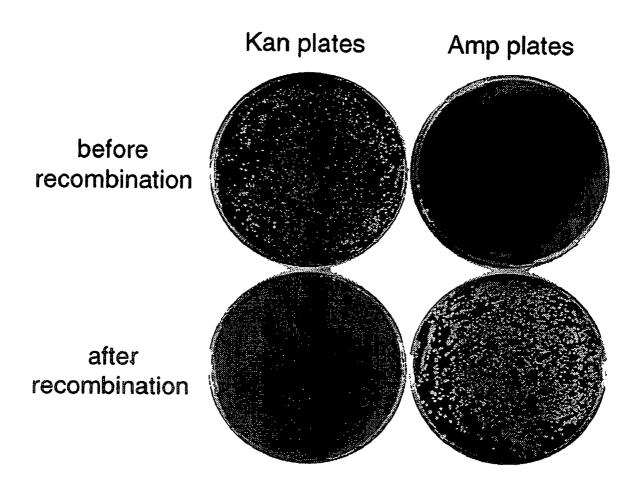
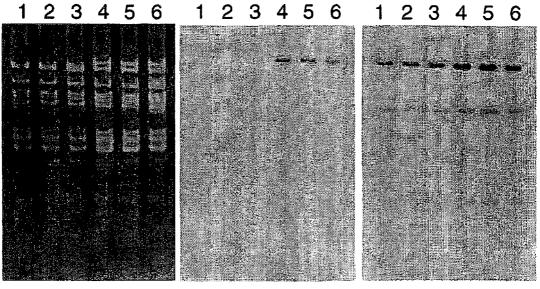


Figure 4b



P1 DNA digested with EcoR I

hybridized with a bla probe (Amp)

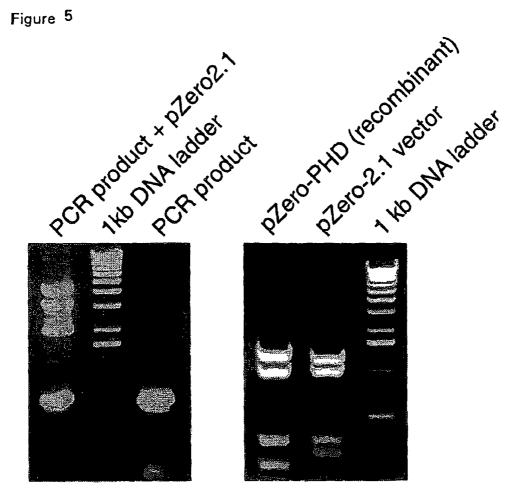
hybridized with a Hoxa-3 probe

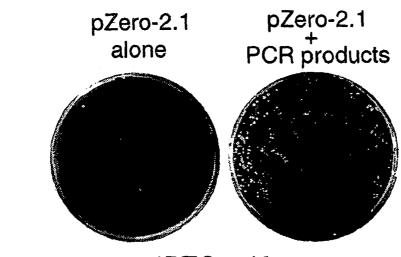
- Lane 1: 1 of P1-Hox clone in NS3145 original bacterial strain (Kan resistance)
- Lane 2-3: 2 of P1-Hox clones in JC9604 before homologous recombination (Kan resistance)
- Lane 4-6: 3 of P1-Hox clones in JC9604 after homologous recombination (Amp resistance)

a

b

Figure 5





IPTG + Kan

Figure 6

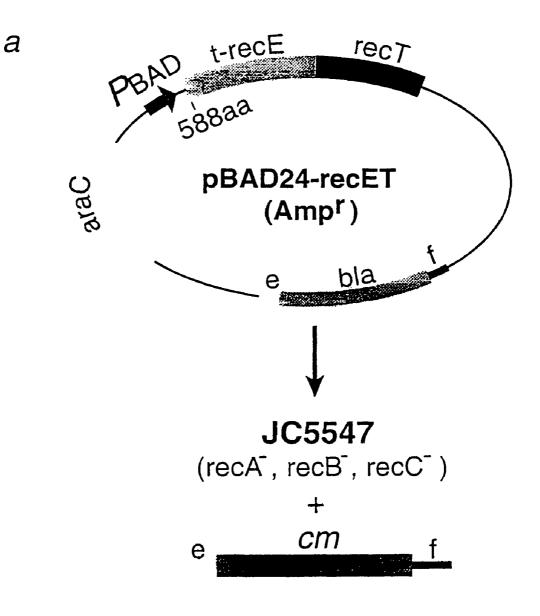
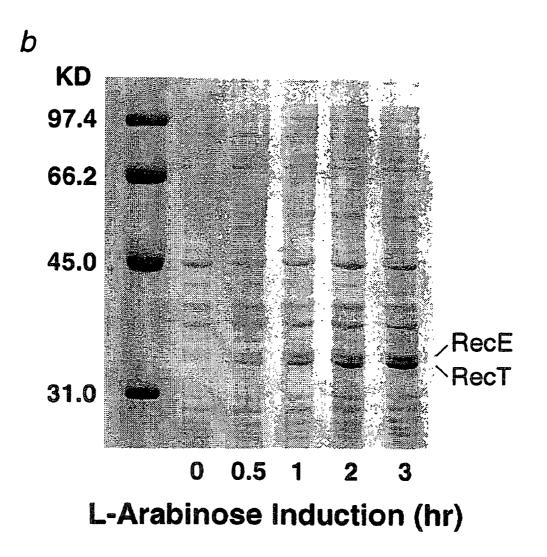


Figure 6



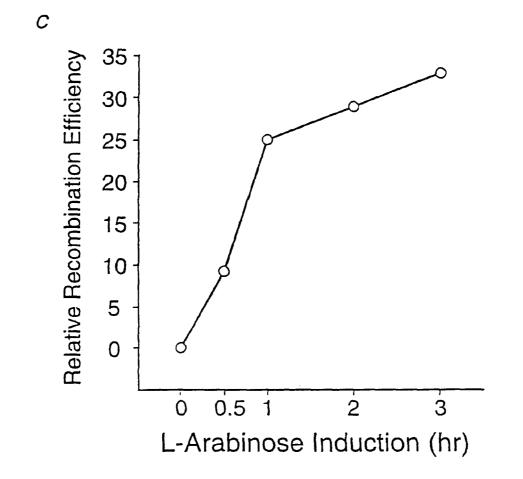
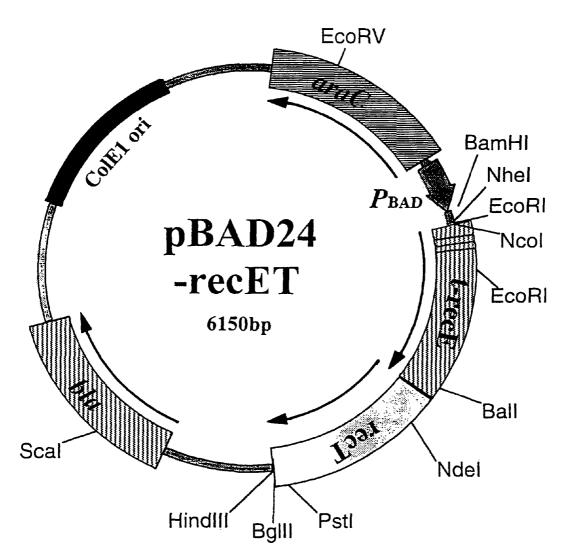


Figure 6

Figure 7a



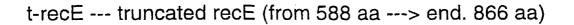


Figure 7b

1 ATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGGGATTC 44 TGCAAACCCTATGCTACTCCGTCAAGCCGTCAATTGTCTGATT 87 CGTTACCAA TTA TGA CAA CTT GAC GGC TAC ATC 293 ◀••• Ser Leu Lys Val Ala Val Asp 120 ATT CAC TIT TIC TIC ACA ACC GGC ACG GAA CIC 285 Asn Val Lys Glu Glu Cys Gly Ala Arg Phe Glu 153 GCT CGG GCT GGC CCC GGT GCA TIT TIT AAA TAC 274 Ser Pro Ser Ala Gly Thr Cys Lys Lys Phe Val 186 CCG CGA GAA ATA GAG TTG ATC GTC AAA ACC AAC 263 Arg Ser Phe Tyr Leu Gin Asp Asp Phe Gly Val 219 ATT GCG ACC GAC GGT GGC GAT AGG CAT CCG GGT 252 Asn Arg Gly Val Thr Ala lie Pro Met Arg Thr 252 GGT GCT CAA AAG CAG CTT CGC CTG GCT GAT ACG 241 Thr Ser Leu Leu Leu Lys Ala Gin Ser lie Arg 285 TTG GTC CTC GCG CCA GCT TAA GAC GCT AAT CCC 230 € GIn Asp Glu Arg Trp Ser Leu Val Ser Ile Gly 318 TAA CTG CTG GCG GAA AAG ATG TGA CAG ACG CGA 2194 Leu Gin Gin Arg Phe Leu His Ser Leu Arg Ser 351 CGG CGA CAA GCA AAC ATG CTG TGC GAC GCT GGC 208 Pro Ser Leu Cys Val His Gln Ala Val Ser Ala EcoRV 384 GAT ATC AAA ATT GCT GTC TGC CAG GTG ATC GCT 1974 He Asp Phe Asn Ser Asp Ala Leu His Asp Ser 417 GAT GTA CTG ACA AGC CTC GCG TAC CCG ATT ATC 1864Ile Tvr Gln Cys Ala Glu Arg Val Arg Asn Asp

Figure 7b (cont'd)

450 CAT CGG TGG ATG GAG CGA CTC GTT AAT CGC TTC 175 € Met Pro Pro His Leu Ser Glu Asn Ile Ala Glu 483 CAT GCG CCG CAG TAA CAA TTG CTC AAG CAG ATT 164 Met Arg Arg Leu Leu Leu Gin Giu Leu Leu Asn 516 TAT CGC CAG CAG CTC CGA ATA GCG CCC TTC CCC 153 Ile Ala Leu Leu Glu Ser Tyr Arg Gly Glu Gly 549 TIG CCC GGC GTT AAT GAT TIG CCC AAA CAG GTC 142 € Gin Giy Ala Asn lle lle Gin Giy Phe Leu Asp 582 GCT GAA ATG CGG CTG GTG CGC TTC ATC CGG GCG 131 Ser Phe His Pro Gin His Ala Glu Asp Pro Arg 615 AAA GAA CCC CGT ATT GGC AAA TAT TGA CGG CCA 648 GTT AAG CCA TTC ATG CCA GTA GGC GCG CGG ACG 109 Asn Leu Trp Glu His Trp Tyr Ala Arg Pro Arg 681 AAA GTA AAC CCA CTG GTG ATA CCA TTC GCG AGC 98 Phe Tyr Val Trp Gln His Tyr Trp Glu Arg Ala 714 CTC CGG ATG ACG ACC GTA GTG ATG AAT CTC TCC 87 € Glu Pro His Arg Gly Tyr His His Ile Glu Gly 747 TGG CGG GAA CAG CAA AAT ATC ACC CGG TCG GCA 76 € Pro Pro Phe Leu Leu IIe Asp Gly Pro Arg Cys 780 AAC AAA TIC TCG TCC CTG ATT TIT CAC CAC CCC 65 Val Phe Glu Arg Gly Gln Asn Lys Val Val Gly 813 CTG ACC GCG AAT GGT GAG ATT GAG AAT ATA ACC 54 €GIn Gly Arg IIe Thr Leu Asn Leu IIe Tyr Gly 846 TTT CAT TCC CAG CGG TCG GTC GAT AAA AAA ATC 43 € Lys Met Gly Leu Pro Arg Asp lie Phe Phe Asp

Figure 7b (cont'd) 879 GAG ATA ACC GTT GGC CTC AAT CGG CGT TAA ACC 32 € Leu Tyr Gly Asn Ala Glu lle Pro Thr Leu Gly 912 CGC CAC CAG ATG GGC ATT AAA CGA GTA TCC CGG 21 Ala Val Leu His Ala Asn Phe Ser Tyr Gly Pro 945 CAG CAG GGG ATC ATT TTG CGC TTC AGC CAT 10 € Leu Leu Pro Asp Asn Gin Ala Giu Ala Met 975 ACTTTTCATA CTCCCGCCAT TCAGAGAAGA AACCAATTGT 1015 CCATATIGCA TCAGACATIG CCGTCACIGC GTCTITIACT 1055 GGCTCTTCTC GCTAACCAAA CCGGTAACCC CGCTTATTAA 1095 AAGCATTCTG TAACAAAGCG GGACCAAAGC CATGACAAAA 1135 ACGCGTAACA AAAGTGTCTA TAATCACGGC AGAAAAGTCC 1175 ACATTGATTA TITGCACGGC GTCACACTTT GCTATGCCAT BamHI 1215 AGCATTTTTA TCCATAAGAT TAGCGGATCC TACCTGACGC 1255 TITTTATCGC AACTCTCTAC TGTTTCTCCA TACCCGTTTT EcoRI Ncol Nhel 1295 TTTGGGCTAG CAGGAGGAAT TCACC ATG GAT CCC GTA 1 Met Asp Pro Val 1332 ATC GTA GAA GAC ATA GAG CCA GGT ATT TAT TAC 5▶ lle Val Glu Asp lle Glu Pro Gly lle Tyr Tyr 1365 GGA ATT TCG AAT GAG AAT TAC CAC GCG GGT CCC 16 Gly Ile Ser Asn Glu Asn Tyr His Ala Gly Pro

1398 GGT ATC AGT AAG TCT CAG CTC GAT GAC ATT GCT

Figure 7b (cont'd) 27 Gly lle Ser Lys Ser Gln Leu Asp Asp lle Ala 1431 GAT ACT CCG GCA CTA TAT TTG TGG CGT AAA AAT 38▶Asp Thr Pro Ala Leu Tyr Leu Trp Arg Lys Asn 1464 GCC CCC GTG GAC ACC ACA AAG ACA AAA ACG CTC 49▶Ala Pro Val Asp Thr Thr Lys Thr Lys Thr Leu 1497 GAT TTA GGA ACT GCT TTC CAC TGC CGG GTA CIT 60 Asp Leu Gly Thr Ala Phe His Cys Arg Val Leu EcoRI 1530 GAA CCG GAA GAA TIC AGT AAC CGC TIT ATC GTA 71 Glu Pro Glu Glu Phe Ser Asn Arg Phe Ile Val 1563 GCA CCT GAA TTT AAC CGC CGT ACA AAC GCC GGA 82 Ala Pro Glu Phe Asn Arg Arg Thr Asn Ala Gly 1596 AAA GAA GAA GAG AAA GCG TTT CTG ATG GAA TGC 93 ▶ Lvs Glu Glu Glu Lys Ala Phe Leu Met Glu Cys 1629 GCA AGC ACA GGA AAA ACG GTT ATC ACT GCG GAA 104 Ala Ser Thr Gly Lys Thr Val IIe Thr Ala Glu 1662 GAA GGC CGG AAA ATT GAA CTC ATG TAT CAA AGC 115▶Glu Gly Arg Lys Ile I Leu Met Tyr Gln Ser

1695 GTT ATG GCT TIG CCG CTG GGG CAA TGG CTT GTT 126 Val Met Ala Leu Pro Leu Gly Gin Trp Leu Val 1728 GAA AGC GCC GGA CAC GCT GAA TCA TCA ATT TAC 137 Blu Ser Ala Gly His Ala Glu Ser Ser Ile Tyr 1761 TGG GAA GAT CCT GAA ACA GGA ATT TTG TGT CGG 148 Trp Glu Asp Pro Glu Thr Gly Ile Leu Cys Arg 1794 TGC CGT CCG GAC AAA ATT ATC CCT GAA TTT CAC 159 Cys Arg Pro Asp Lys IIe IIe Pro Glu Phe His 1827 TGG ATC ATG GAC GTG AAA ACT ACG GCG GAT ATT 170 ▶ Trp lle Met Asp Val Lys Thr Thr Ala Asp lle 1860 CAA CGA TTC AAA ACC GCT TAT TAC GAC TAC CGC 181 Gln Arg Phe Lys Thr Ala Tyr Tyr Asp Tyr Arg 1893 TAT CAC GTT CAG GAT GCA TTC TAC AGT GAC GGT 192 Fyr His Val Gln Asp Ala Phe Tyr Ser Asp Gly 1926 TAT GAA GCA CAG TIT GGA GTG CAG CCA ACT TIC 203 ▶ Tyr Glu Ala Gln Phe Gly Val Gln Pro Thr Phe 1959 GTT TTT CTG GTT GCC AGC ACA ACT ATT GAA TGC 214 Val Phe Leu Val Ala Ser Thr Thr Ile Glu Cys 1992 GGA CGT TAT CCG GTT GAA ATT TTC ATG ATG GGC

Figure 7b (cont'd)

Figure 7b (cont'd)

225 Bly Arg Tyr Pro Val Glu IIe Phe Met Met Gly 2025 GAA GAA GCA AAA CTG GCA GGT CAA CAG GAA TAT 236 Glu Glu Ala Lys Leu Ala Gly Gln Gln Glu Tyr 2058 CAC CGC AAT CTG CGA ACC CTG TCT GAC TGC CTG 247 His Arg Asn Leu Arg Thr Leu Ser Asp Cys Leu Ball 2091 AAT ACC GAT GAA TGG CCA GCT ATT AAG ACA TTA 258 Asn Thr Asp Glu Trp Pro Ala Ile Lys Thr Leu 2124 TCA CTG CCC CGC TGG GCT AAG GAA TAT GCAA 269 Ser Leu Pro Arg Trp Ala Lys Glu Tyr AlaA 2155 ATG ACT AAG CAA CCA CCA ATC GCA AAA GCC GAT 1 Met Thr Lys Gin Pro Pro Ile Ala Lys Ala Asp

279 s nAs p• • • 2188 CTG CAA AAA ACT CAG GGA AAC CGT GCA CCA GCA 12 Leu Gin Lys Thr Gin Gly Asn Arg Ala Pro Ala 2221 GCA GTT AAA AAT AGC GAC GTG ATT AGT TTT ATT 23 Ala Val Lys Asn Ser Asp Val IIe Ser Phe IIe 2254 AAC CAG CCA TCA ATG AAA GAG CAA CTG GCA GCA 34 Asn GIn Pro Ser Met Lys GIu GIn Leu Ala Ala Ndel 2287 GCT CTT CCA CGC CAT ATG ACG GCT GAA CGT ATG 45 Ala Leu Pro Arg His Met Thr Ala Glu Arg Met

Figure 7t	o (cont'd)					
2320 A 56▶1	TC CG					
2353 C 67▶P	CG GCC ro Ala					
2386 G 78▶V	TC AG al Ser					GGA GI y
2419 C 89▶L	TT GAG eu Glu					
2452 T 100⊧L						
2485 G 111▶G						
2518 C 122►A						
2551 C 133▶G						
2584 G 144 ⊧ G						
2617 G 155►A						
2650 G 166⋫G						
2683 G 177▶A						
2716 G 188▶ Va						

Figure 7b (cont'd) 2749 AGC CTG AGT AAA GCT GGT AAT AAC GGG CCG TGG 199 Ser Leu Ser Lys Ala Gly Asn Asn Gly Pro Trp 2782 GTA ACT CAC TGG GAA GAA ATG GCA AAG AAA ACG 210 Val Thr His Trp Glu Glu Met Ala Lys Lys Thr 2815 GCT ATT CGT CGC CTG TTC AAA TAT TTG CCC GTA 221 Ala Ile Arg Arg Leu Phe Lys Tyr Leu Pro Val 2848 TCA ATT GAG ATC CAG CGT GCA GTA TCA ATG GAT 232 Ser lle Glu lle Gln Arg Ala Val Ser Met Asp Pstl 2881 GAA AAG GAA CCA CTG ACA ATC GAT CCT GCA GAT 243 Glu Lys Glu Pro Leu Thr lle Asp Pro Ala Asp 2914 TCC TCT GTA TTA ACC GGG GAA TAC AGT GTA ATC 254 Ser Ser Val Leu Thr Gly Glu Tyr Ser Val IIe HindIII Ball 2947 GAT AAT TCA GAG GAA TAG ATCTAAGCTT 265▶Asp Asn Ser Glu Glu ••• 2975 GGCTGTTTTG GCGGATGAGA GAAGATTTTC AGCCTGATAC 3015 AGATTAAATC AGAACGCAGA AGCGGTCTGA TAAAACAGAA 3055 TTTGCCTGGC GGCAGTAGCG CGGTGGTCCC ACCTGACCCC 3095 ATGCCGAACT CÁGAAGTGAA ACGCCGTAGC GCCGATGGTA 3135 GTGTGGGGTC TCCCCATGCG AGAGTAGGGA ACTGCCAGGC 3175 ATCAAATAAA ACGAAAGGCT CAGTCGAAAG ACTGGGCCTT 3215 TCGTTTTATC TGTTGTTTGT CGGTGAACGC TCTCCTGAGT 3255 AGGACAAATC CGCCGGGAGC GGATTTGAAC GTTGCGAAGC 3295 AACGGCCCGG AGGGTGGCGG GCAGGACGCC CGCCATAAAC 3335 TGCCAGGCAT CAAATTAAGC AGAAGGCCAT CCTGACGGAT Figure 7b (cont'd)

3375 GGCCTTTTTG CGTTTCTACA AACTCTTTTG TTTATTTTTC

3415 TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC

3455 CCTGATAAAT GCTTCAATAA TATTGAAAAA GGAAGAGT AT 1▶Me

3495 G AGT ATT CAA CAT TTC CGT GTC GCC CTT ATT 1▶t Ser lle GIn His Phe Arg Val Ala Leu lle 3526 CCC TTT TIT GCG GCA TTT TGC CTT CCT GTT TTT 12 Pro Phe Phe Ala Ala Phe Cys Leu Pro Val Phe 3559 GCT CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT 23 Ala His Pro Glu Thr Leu Val Lys Val Lys Asp 3592 GCT GAA GAT CAG TIG GGT GCA CGA GTG GGT TAC 34 Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr 3625 ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT 45 lle Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu 3658 GAG AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG 56 Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met 3691 ATG AGC ACT TTT AAA GTT CTG CTA TGT GGC GCG 67 Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala 3724 GTA TTA TCC CGT GTT GAC GCC GGG CAA GAG CAA 78 Val Leu Ser Arg Val Asp Ala Gly Gin Glu Gin 3757 CTC GGT CGC CGC ATA CAC TAT TCT CAG AAT GAC 89▶Leu Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Scal 3790 TTG GTT GAG TAC TCA CCA GTC ACA GAA AAG CAT 100 ▶ Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His 3823 CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC 111 Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys

Figure 7b (c	ont'd)									
3856 AGT 122 ▶ Ser										
3889 GCC 133▶Ala										
3922 AAG 144⊧Lys										
3955 GAT 155▶Asp										
3988 GAG 166∳Glu										
4021 GAC 177▶Asp										
4054 TTG 188▶Leu										
4087 CTA 199▶Leu										
4120 GAG 210 È Gl u										
4153 TCG 221 ▶ Ser										
4186 AAA 232⊧Lys										
4219 ATC 243 ► IIe	lle	Ala	Ala	Leu	Gl y	Pro	Asp	GI y	Lys	Pro
4252 TCC 254 ► Ser	CGT A rg	ATC e	GTA Va I	GTT Val	ATC 11e	TAC Tyr	ACG Thr	ACG Thr	GGG Gl y	AGT Ser

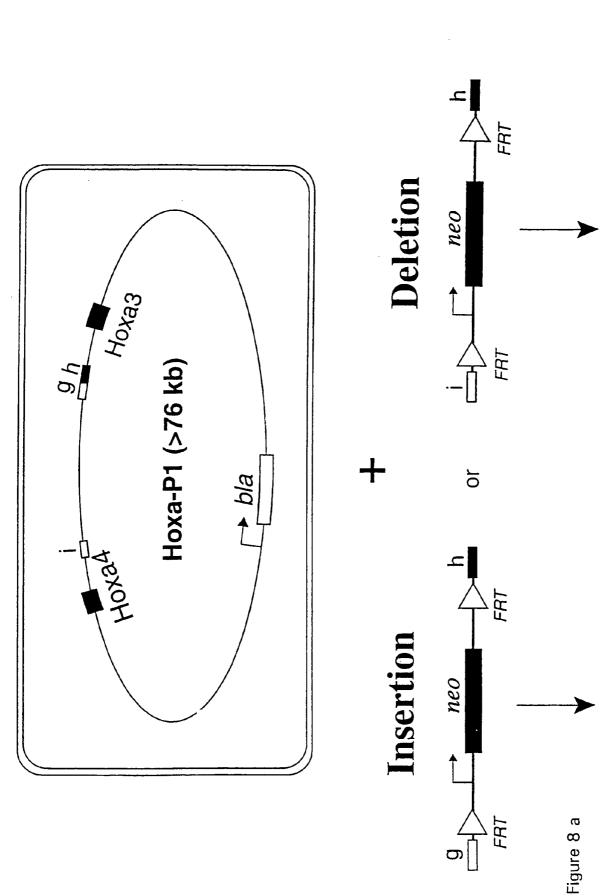
Figure /b (cont'd)

4285 CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC 265 ▶ Gin Ala Thr Met Asp Giu Arg Asn Arg Gin ile 4318 GCT GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG 276 Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp 4351 TAA CTGTCAGACC AAGTTTACTC ATATATACTT 287 . . . 4384 TAGATTGATT TACGCGCCCT GTAGCGGCGC ATTAAGCGCG 4424 GCGGGTGTGG TGGTTACGCG CAGCGTGACC GCTACACTTG 4464 CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT TCTTCCCTTC 4504 CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TCAAGCTCTA 4544 AATCGGGGGC TCCCTTTAGG GTTCCGATTT AGTGCTTTAC 4584 GGCACCTCGA CCCCAAAAAA CTTGATTIGG GIGATGGTTC 4624 ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT 4664 TTGACGTTGG AGTCCACGTT CTTTAATAGT GGACTCTTGT 4704 TCCAAACTIG AACAACACTC AACCCTATCT CGGGCTATTC 4744 TTTIGATTA TAAGGGATTT TGCCGATTTC GGCCTATTGG 4784 TTAAAAAATG AGCTGATTTA ACAAAAATTT AACGCGAATT 4824 TTAACAAAAT ATTAACGTTT ACAATTTAAA AGGATCTAGG 4864 TGAAGATCCT TTTTGATAAT CTCATGACCA AAATCCCTTA 4904 ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCGTAGAA 4944 AAGATCAAAG GATCTTCTTG AGATCCTTTT TTTCTGCGCG 4984 TAATCTGCTG CTTGCAAACA AAAAAACCAC CGCTACCAGC 5024 GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTT

Figure	e 7b (cont'd)			
5064	CCGAAGGTAA	CTGGCTTCAG	CAGAGCGCAG	ATACCAAATA
5104	CTGTCCTTCT	AGTGTAGCCG	TAGTTAGGCC	ACCACTTCAA
5144	GAACTCTGTA	GCACCGCCTA	CATACCTCGC	TCTGCTAATC
5184	CTGTTACCAG	TGGCTGCTGC	CAGTGGCGAT	AAGTCGTGTC
5224	TTACCGGGTT	GGACTCAAGA	CGATAGTTAC	CGGATAAGGC
5264	GCAGCGGTCG	GGCTGAACGG	GGGGTTCGTG	CACACAGCCC
5304	AGCTTGGAGC	GAACGACCTA	CACCGAACTG	AGATACCTAC
5344	AGCGTGAGCT	ATGAGAAAGC	GCCACGCTTC	CCGAAGGGAG
5384	AAAGGCGGAC	AGGTATCCGG	TAAGCGGCAG	GGTCGGAACA
5424	GGAGAGCGCA	CGAGGGAGCT	TCCAGGGGGA	AACGCCTGGT
5464	ATCTTTATAG	TCCTGTCGGG	TTTCGCCACC	TCTGACTTGA
5504	GCGTCGATTT	TIGIGATGCT	CGTCAGGGGG	GCGGAGCCTA
5544	TGGAAAAACG	CCAGCAACGC	GGCCTTTTTA	CGGTTCCTGG
5584	CCTTTTGCTG	GCCTTTTGCT	CACATGTTCT	TICCIGCGIT
5624	ATCCCCTGAT	TCTGTGGATA	ACCGTATTAC	CGCCTTTGAG
5664	TGAGCTGATA	CCGCTCGCCG	CAGCCGAACG	ACCGAGCGCA
5704	GCGAGTCAGT	GAGCGAGGAA	GCGGAAGAGC	GCCTGATGCG
5744	GTATITICIC	CTTACGCATC	TGTGCGGTAT	TTCACACCGC
5784	ATAGGGTCAT	GGCTGCGCCC	CGACACCCGC	CAACACCCGC
5824	TGACGCGCCC	TGACGGGCTT	GTCTGCTCCC	GGCATCCGCT
5864	TACAGACAAG	CTGTGACCGT	CTCCGGGAGC	TGCATGTGTC
5904	AGAGGTTTTC	ACCGTCATCA	CCGAAACGCG	CGAGGCAGCA
5944	AGGAGATGGC	GCCCAACAGT	CCCCCGGCCA	CGGGGCCTGC

Figure 7b (cont'd)

5984	CACCATACCC	ACGCCGAAAC	AAGCGCTCAT	GAGCCCGAAG
6024	TGGCGAGCCC	GATCTTCCCC	ATCGGTGATG	TCGGCGATAT
6064	AGGCGCCAGC	AACCGCACCT	GIGGCGCCGG	TGATGCCGGC
6104	CACGATGCGT	CCGGCGTAGA	GGATCTGCTC	ATGTTTGACA
6144	GCTTATC			



σ

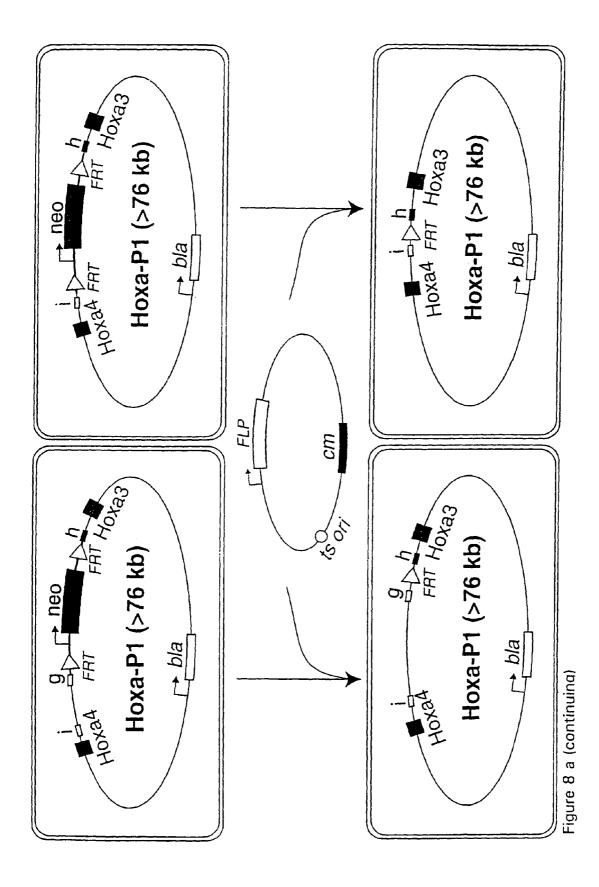


Figure 8 b

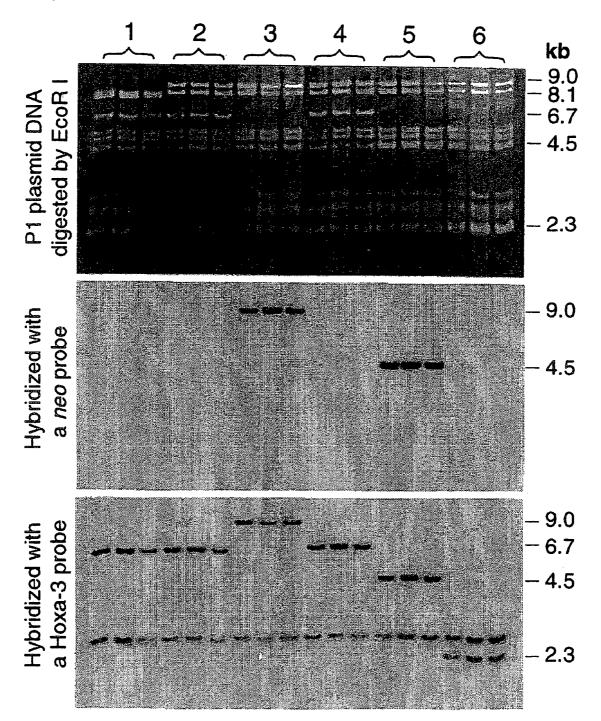


Figure 9a

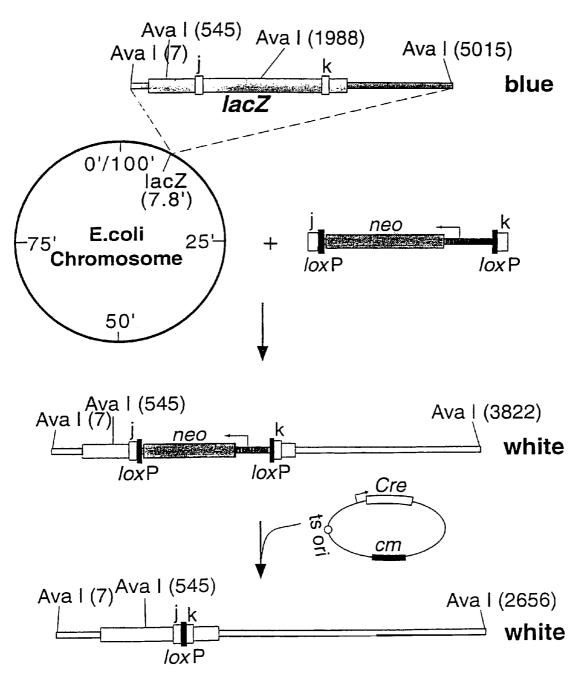
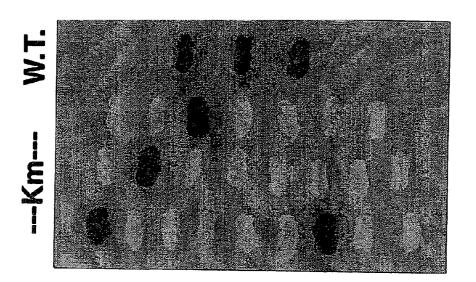


Figure 9

b



C

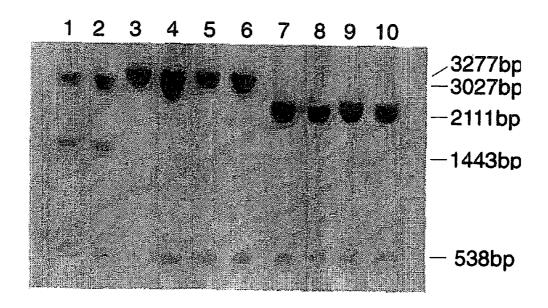


Figure 10a

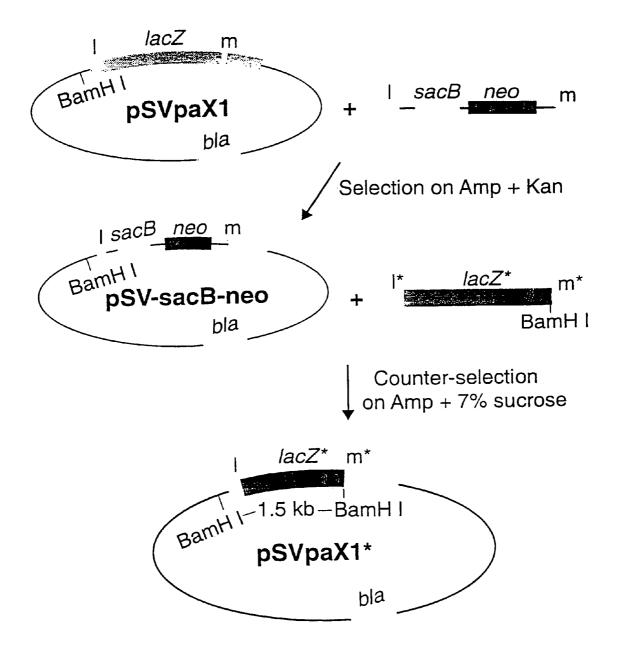
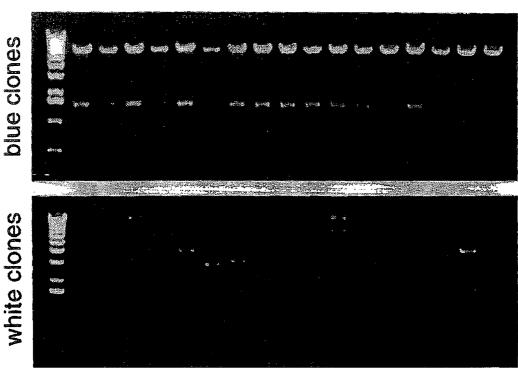


Figure 10







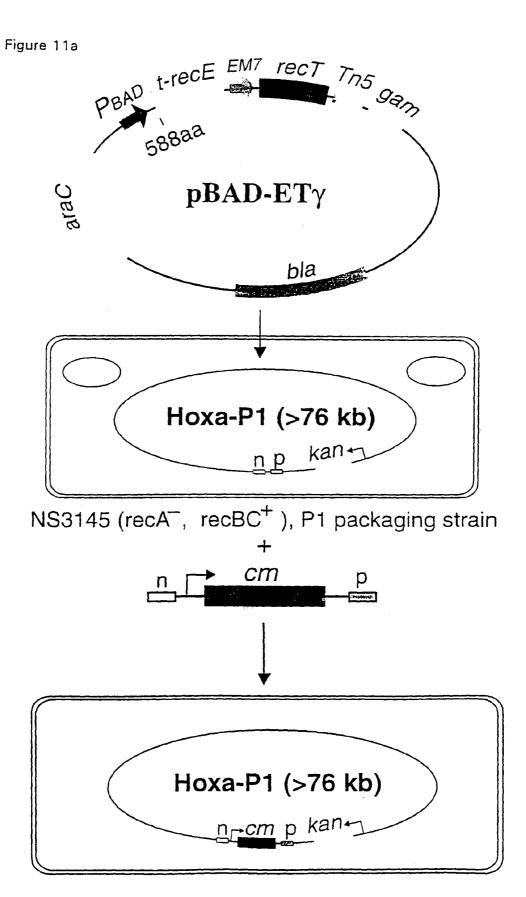
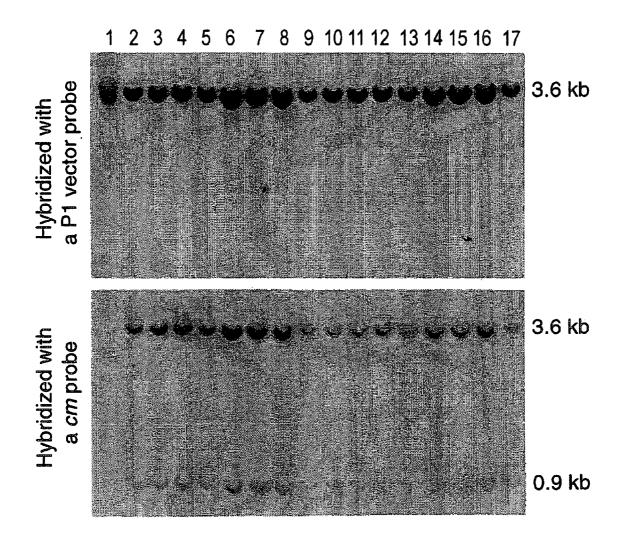


Figure 11 b



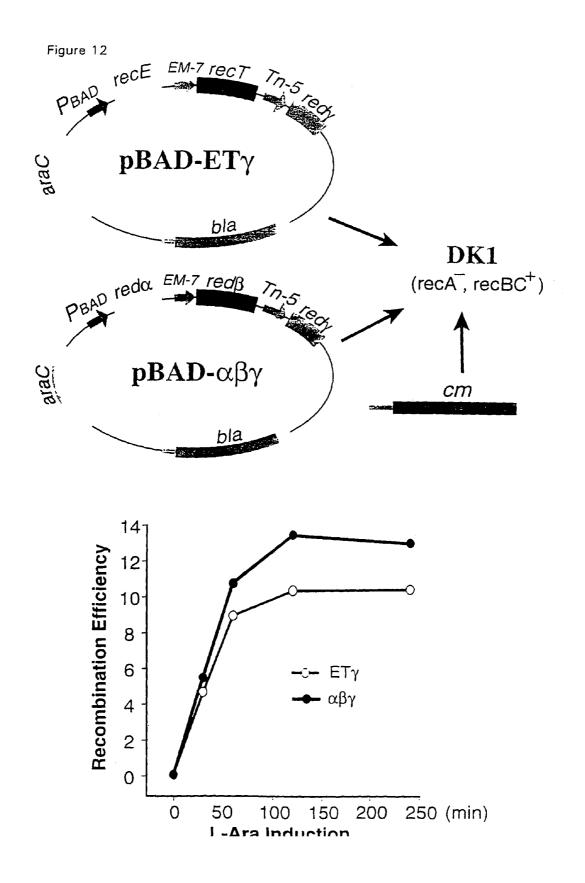


Figure 13 a

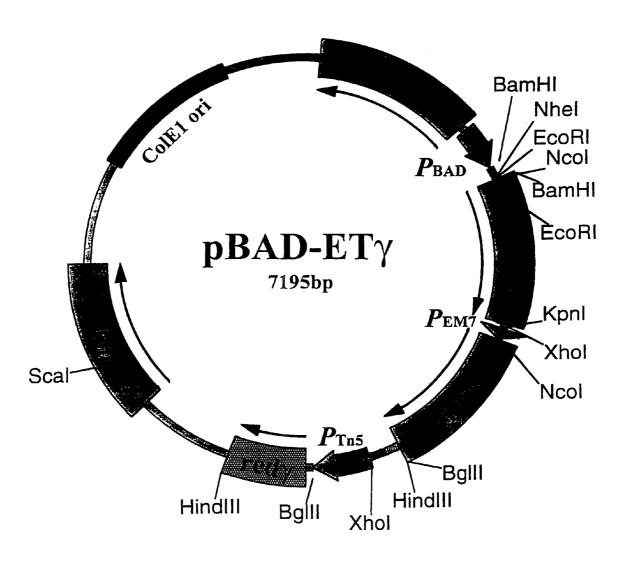


Figure 13b

1 ATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGGG 40 ATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAATT 79 GTCTGATTCGTTACCAA TTA TGA CAA CTT GAC 293 . • • Ser Leu Lys Val 111 GGC TAC ATC ATT CAC TIT TIC TIC ACA ACC 288 Ala Val Asp Asn Val Lys Glu Glu Cys Gly 141 GGC ACG GAA CTC GCT CGG GCT GGC CCC GGT 278 Ala Arg Phe Glu Ser Pro Ser Ala Gly Thr 171 GCA TIT TIT AAA TAC CCG CGA GAA ATA GAG 268 Cys Lys Lys Phe Val Arg Ser Phe Tyr Leu 201 TIG ATC GTC AAA ACC AAC ATT GCG ACC GAC 258 Gin Asp Asp Phe Gly Val Asn Arg Gly Val 231 GGT GGC GAT AGG CAT CCG GGT GGT GCT CAA 248 Thr Ala Ile Pro Met Arg Thr Thr Ser Leu 261 AAG CAG CTT CGC CTG GCT GAT ACG TTG GTC 238 Leu Leu Lys Ala Gin Ser Ile Arg Gin Asp 291 CTC GCG CCA GCT TAA GAC GCT AAT CCC TAA 228 Glu Arg Trp Ser Leu Val Ser Ile Gly Leu 321 CTG CTG GCG GAA AAG ATG TGA CAG ACG CGA 218 GIn GIn Arg Phe Leu His Ser Leu Arg Ser 351 CGG CGA CAA GCA AAC ATG CTG TGC GAC GCT 208 Pro Ser Leu Cys Val His Gln Ala Val Ser 381 GGC GAT ATC AAA ATT GCT GTC TGC CAG GTG 198 Ala lle Asp Phe Asn Ser Asp Ala Leu His 411 ATC GCT GAT GTA CTG ACA AGC CTC GCG TAC

Figure 13b (cor	nt'd)								
188¶Asp	Ser	lle	Тyr	GIn	Cys	Ala	Gl u	Arg	Val
441 CCG 178∮Arg									
471 GTT 168∮Asn									
501 TTG 158∮GIn									
531 CGA 148 ∢Ser									
561 AAT 138¶∣∣e									
591 CGG 128∮Pro									
621 CCC 118 ∢G Iy									
651 AAG 108∢Leu									
681 AAA 98 ∢P he									
711 AGC 88∮Ala									
741 CTC 78∮Glu									
771 CGG 68∮Pro									

					ACC Gl y					
					AAA Phe					
					CGT Thr					
					CGA Ser					
					CGC Ala				ACTI	TTC
982	ATAC	CTCCC	CGCCF	TTCF	AGAGA		ACCZ	ATTO	STCCA	TAT
1021	TGCZ	ATCA(GACAI	TGCC	CGTCA	ACTGO	GTCI	TTTA	LCTGC	GCTC
1060	TTCI	rcgci	TAACC		CCGGI	AACC	CCGC	TTAT	TAAA	AGC
1099	ATTO	CTGTZ	ACAA	AGCO	GGAC		GCCA	TGAC		ACG
1138	CGTZ	ACAZ	AAGT	GTCI		TCAC	GGCA	GAAA	AGTC	CAC
1177	ATTO	GATTA	TTTC	CACO	GCGI	CACA	CTTI	GCTA	TGCC	ATA
1216	GCAI	L.L.L.I.	TATCC	ATAA	GATT		BamHI GATC	CTAC	CTGA	.CGC

801 TTT CAC CAC CCC CTG ACC GCG AAT GGT GAG 58◀Lys Val Val Gly Gln Gly Arg lle Thr Leu

Figure 1	3b (cont	:'d)								
1255	TTT	TATC	GCAA	ACTCI	ICTAC	TGTI	TCTC	CATA	CCCC	TTT
		Nł	nel		E∞	RI	Nco	l Ba	amHl	
1294	TTT	rgggC	TAGC	AGGA	GGAA	T TC	ACC	AIG	GAT	CCC
							1)	Met	Asp	Pro
1329	GTA	ATC	GTA	GAA	GAC	ATA	GAG	CCA	GGT	ATT
4	Val	lle	Val	Glu	Asp	lle	Gl u	Pro	Gl y	lle
1359	TAT	TAC	GGA	ATT	TCG	AAT	GAG	AAT	TAC	CAC
	Tyr									
1389	GCG	GGT	CCC	GGT	ATC	AGT	AAG	TCT	CAG	CTC
	Ala									
1419	GAT	GAC	ATT	GCT	GAT	ACT	CCG	GCA	CTA	TAT
	Asp									Туr
1449	TIG	TGG	CGT	AAA	AAT	GCC	CCC	GTG	GAC	ACC
	Leu									
1479	ACA	AAG	ACA	AAA	ACG	CTC	GAT	TTA	GGA	ACT
54	Thr	Lys	Thr	Lys	Thr	Leu	Asp	Leu	Gl y	Thr
1509	GCT	TTC	CAC	TGC	CGG	GTA	CTT	GAA	CCG	GAA
641	Ala	Phe	Hi s	Cys	Arg	Val	Leu	Glu	Pro	Glu
	EcoF	RI								
1539	GAA	TIC	AGT	AAC	CGC	TTT	ATC	GIA	GCA	CCT Pro
	Gl u									
1569										
	Gl u									
15 9 9	GAA	GAA	GAG	AAA	GCG	TTT	CIG	ATG	GAA	TGC
94	Gi u	Gl u	Glu	Lys	Ala	Phe	Leu	Met	Glu	Cys
1629	GCA	AGC	ACA	GGA	AAA	ACG	GTT	ATC	ACT	GCG
104	Ala	Ser	Thr	GI y	Lys	Thr	Val	lle	Thr	Ala

Figure 13b (cont'd) 1659 GAA GAA GGC CGG AAA ATT GAA CTC ATG TAT 114 Glu Glu Gly Arg Lys IIe Glu Leu Met Tyr 1689 CAA AGC GTT ATG GCT TIG CCG CTG GGG CAA 124 Gin Ser Val Met Ala Leu Pro Leu Giy Gin 1719 TGG CTT GTT GAA AGC GCC GGA CAC GCT GAA 134 Trp Leu Val Glu Ser Ala Gly His Ala Glu 1749 TCA TCA ATT TAC TGG GAA GAT CCT GAA ACA 144 Ser Ser Ile Tyr Trp Glu Asp Pro Glu Thr 1779 GGA ATT TIG TGT CGG TGC CGT CCG GAC AAA 154 Glv IIe Leu Cys Arg Cys Arg Pro Asp Lys 1809 ATT ATC CCT GAA TTT CAC TGG ATC ATG GAC 164 lle lle Pro Glu Phe His Trp lle Met Asp 1839 GTG AAA ACT ACG GCG GAT ATT CAA CGA TTC 174 Val Lys Thr Thr Ala Asp Ile GIn Arg Phe 1869 AAA ACC GCT TAT TAC GAC TAC CGC TAT CAC 184 Lys Thr Ala Tyr Tyr Asp Tyr Arg Tyr His 1899 GTT CAG GAT GCA TTC TAC AGT GAC GGT TAT 194 Val Gin Asp Ala Phe Tyr Ser Asp Gly Tyr 1929 GAA GCA CAG TTT GGA GTG CAG CCA ACT TTC 204 Glu Ala Gin Phe Giy Val Gin Pro Thr Phe 1959 GTT TTT CTG GTT GCC AGC ACA ACT ATT GAA 214 Val Phe Leu Val Ala Ser Thr Thr Ile Glu 1989 TGC GGA CGT TAT CCG GTT GAA ATT TTC ATG 224 Cys Gly Arg Tyr Pro Val Glu IIe Phe Met 2019 ATG GGC GAA GAA GCA AAA CTG GCA GGT CAA 234 Met Gly Glu Glu Ala Lys Leu Ala Gly Gln

Figure 13b (cont'd) 2049 CAG GAA TAT CAC CGC AAT CTG CGA ACC CTG 244 GIn Glu Tyr His Arg Asn Leu Arg Thr Leu 2079 TCT GAC TGC CTG AAT ACC GAT GAA TGG CCA 254 Ser Asp Cys Leu Asn Thr Asp Glu Trp Pro 2109 GCT ATT AAG ACA TTA TCA CTG CCC CGC TGG 264 Ala Ile Lys Thr Leu Ser Leu Pro Arg Trp Xhol Kpnl 2139 GCT AAG GAA TAT GCA AAT GAC TAGATCTCGAG 274 Ala Lys Glu Tyr Ala Asn Asp 2171 GTACCCGAGCACGTGTTGACAATTAATCATCGGCATAGT 2210 ATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAA Ncol 2249 CC ATG GCT AAG CAA CCA CCA ATC GCA AAA 1 Met Ala Lys Gln Pro Pro Ile Ala Lys 2278 GCC GAT CTG CAA AAA ACT CAG GGA AAC CGT 10 Ala Asp Leu Gin Lys Thr Gin Gly Asn Arg 2308 GCA CCA GCA GCA GTT AAA AAT AGC GAC GTG 20 Ala Pro Ala Ala Val Lys Asn Ser Asp Val 2338 ATT AGT TIT ATT AAC CAG CCA TCA ATG AAA 30 ▶ Ile Ser Phe Ile Asn GIn Pro Ser Met Lys 2368 GAG CAA CTG GCA GCA GCT CTT CCA CGC CAT 40 Glu Gln Leu Ala Ala Ala Leu Pro Arg His 2398 ATG ACG GCT GAA CGT ATG ATC CGT ATC GCC 50 Met Thr Ala Glu Arg Met Ile Arg Ile Ala 2428 ACC ACA GAA ATT CGT AAA GTT CCG GCG TTA 60 Thr Thr Glu IIe Arg Lys Val Pro Ala Leu

Figure 13b (co	ont'd)					
2458 GGA 70 È GI y						-
2488 GCG 80▶Ala	•					
2518 GAG 90 È GI u						
2548 TTA 100▶Leu						
2578 AGC 110▶ Ser	-	-	-			 ATT e
2608 GGC 120 ► GI y						
2638 CGT 130►Arg						
2668 CGT 140▶Arg						
2698 TTC 150 ▶ Phe						
2728 CAC 160 His						
2758 GTT 170 ► Val		*				
2788 AAA 180⊧Lys					-	
2818 ACG 190▶Thr						

Figure 1			<u>ה</u> גר ג	രന്ന	CCIT	אסג
2848 200				Ala		
2878	GTA	ACT	CAC	TGG	GAA	GAA

Asn Gly Pro Trp 2878 GTA ACT CAC TGG GAA GAA ATG GCA AAG AAA 210 Val Thr His Trp Glu Glu Met Ala Lys Lys 2908 ACG GCT ATT CGT CGC CTG TTC AAA TAT TTG 220 ▶ Thr Ala IIe Arg Arg Leu Phe Lys Tyr Leu 2938 CCC GTA TCA ATT GAG ATC CAG CGT GCA GTA 230 Pro Val Ser Ile Glu Ile Gln Arg Ala Val 2968 TCA ATG GAT GAA AAG GAA CCA CTG ACA ATC 240 Ser Met Asp Glu Lys Glu Pro Leu Thr Ile 2998 GAT CCT GCA GAT TCC TCT GTA TTA ACC GGG 250 Asp Pro Ala Asp Ser Ser Val Leu Thr Gly 3028 GAA TAC AGT GTA ATC GAT AAT TCA GAG GAA 260 Glu Tvr Ser Val IIe Asp Asn Ser Glu Glu Balll HindIII 3058 TAG ATCTAAGCTTCCTGCTGAACATCAAAGGCAAGAAA 270 • • • 3096 ACATCTGTTGTCAAAGACAGCATCCTTGAACAAGGACAA 3135 TTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGA 3174 TATCCTATTGGCATTTTCTTTATTTCTTATCAACATAA Xhol

- 3213 AGGTGAATCCCATACCTCGAGCTTCACGCTGCCGCAAGC
- 3252 ACTCAGGGCGCAAGGGCTGCTAAAAGGAAGCGGAACACG
- 3291 TAGAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGGATG
- 3330 AATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCA
- 3369 AGCGCAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACA

AAC GGG CCG TGG

Figure 13b (cont'd) 3408 TGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGC 3447 GAACCGGAATTGCCAGCTGGGGGCGCCCTCTGGTAAGGTT 3486 GGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCG Balll 3525 CCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAA 3564 GAGACAGGATGAGGATCGTTTCGC ATG GAT ATT 1 Met Asp Ile 3597 AAT ACT GAA ACT GAG ATC AAG CAA AAG CAT 4▶Asn Thr Glu Thr Glu lle Lys Gln Lys His 3627 TCA CTA ACC CCC TTT CCT GTT TTC CTA ATC 14 Ser Leu Thr Pro Phe Pro Val Phe Leu Ile 3657 AGC CCG GCA TIT CGC GGG CGA TAT TIT CAC 24 Ser Pro Ala Phe Arg Gly Arg Tyr Phe His 3687 AGC TAT TTC AGG AGT TCA GCC ATG AAC GCT 34 ▶ Ser Tyr Phe Arg Ser Ser Ala Met Asn Ala 3717 TAT TAC ATT CAG GAT CGT CTT GAG GCT CAG 44▶Tyr Tyr Ile Gin Asp Arg Leu Giu Ala Gin 3747 AGC TGG GCG CGT CAC TAC CAG CAG CTC GCC 54 ▶ Ser Trp Ala Arg His Tyr Gln Gln Leu Ala 3777 CGT GAA GAG AAA GAG GCA GAA CTG GCA GAC 64▶Arg Glu Glu Lys Glu Ala Glu Leu Ala Asp 3807 GAC ATG GAA AAA GGC CTG CCC CAG CAC CTG 74 ▶ Asp Met Glu Lys Gly Leu Pro Gln His Leu 3837 TTT GAA TCG CTA TGC ATC GAT CAT TTG CAA 84 Phe Glu Ser Leu Cys Ile Asp His Leu Gln 3867 CGC CAC GGG GCC AGC AAA AAA TCC ATT ACC 94 Arg His Gly Ala Ser Lys Lys Ser Ile Thr

Figure 13b (cont'd) 3897 CGT GCG TTT GAT GAC GAT GTT GAG TTT CAG 104 ▶ Arg Ala Phe Asp Asp Asp Val Glu Phe Gin 3927 GAG CGC ATG GCA GAA CAC ATC CGG TAC ATG 114 ▶ Glu Arg Met Ala Glu His Ile Arg Tyr Met 3957 GTT GAA ACC ATT GCT CAC CAC CAG GTT GAT 124 Val Glu Thr lle Ala His His Gln Val Asp HindIII 3987 ATT GAT TCA GAG GTA TAA AACGAGTAGA AGCT 134 Ile Asp Ser Glu Val ••• 4019 TGGCTGTTTTGGCGGATGAGAGAGATTTTCAGCCTGAT 4058 ACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACA 4097 GAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGA 4136 CCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGA 4175 TGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTG 4214 CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACT 4253 GGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTC 4292 TCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACG 4331 TTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCC 4370 CGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCA 4409 TCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTT 4448 TGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTC 4487 ATGAGACAATAACCCTGATAAATGCTTCAATAATATTGA 4526 AAAAGGAAGAGT ATG AGT ATT CAA CAT TTC 1 Met Ser lle Gln His Phe

4556 CGT GTC GCC CTT ATT CCC TTT TTT GCG GCA 7▶Arg Val Ala Leu Ile Pro Phe Phe Ala Ala 4586 TIT TGC CTT CCT GTT TIT GCT CAC CCA GAA 17 Phe Cys Leu Pro Val Phe Ala His Pro Glu 4616 ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT 27 Fhr Leu Val Lys Val Lys Asp Ala Glu Asp 4646 CAG TTG GGT GCA CGA GTG GGT TAC ATC GAA 37 Gin Leu Gly Ala Arg Val Gly Tyr lle Glu 4676 CTG GAT CTC AAC AGC GGT AAG ATC CTT GAG 47 Leu Asp Leu Asn Ser Gly Lys IIe Leu Glu 4706 AGT TIT CGC CCC GAA GAA CGT TIT CCA ATG 57▶ Ser Phe Arg Pro Glu Glu Arg Phe Pro Met 4736 ATG AGC ACT TTT AAA GTT CTG CTA TGT GGC 67 Met Ser Thr Phe Lys Val Leu Leu Cys Gly 4766 GCG GTA TTA TCC CGT GTT GAC GCC GGG CAA 77▶Ala Val Leu Ser Arg Val Asp Ala Gly Gln 4796 GAG CAA CTC GGT CGC CGC ATA CAC TAT TCT 87 ▶ Glu Gln Leu Gly Arg Arg Ile His Tyr Ser Scal 4826 CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC 97 Gln Asn Asp Leu Val Glu Tyr Ser Pro Val 4856 ACA GAA AAG CAT CTT ACG GAT GGC ATG ACA 107 Thr Glu Lys His Leu Thr Asp Gly Met Thr 4886 GTA AGA GAA TTA TGC AGT GCT GCC ATA ACC 117 Val Arg Glu Leu Cys Ser Ala Ala Ile Thr 4916 ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT 127 Met Ser Asp Asn Thr Ala Ala Asn Leu Leu

4946 CTG ACA ACG ATC GGA GGA CCG AAG GAG CTA 137▶Leu Thr Thr IIe Gly Gly Pro Lys Glu Leu 4976 ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT 147 Thr Ala Phe Leu His Asn Met Gly Asp His 5006 GTA ACT CGC CTT GAT CGT TGG GAA CCG GAG 157 Val Thr Arg Leu Asp Arg Trp Glu Pro Glu 5036 CTG AAT GAA GCC ATA CCA AAC GAC GAG CGT 167 Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg 5066 GAC ACC ACG ATG CCT GTA GCA ATG GCA ACA 177▶Asp Thr Thr Met Pro Val Ala Met Ala Thr 5096 ACG TTG CGC AAA CTA TTA ACT GGC GAA CTA 187 Thr Leu Arg Lys Leu Leu Thr Gly Glu Leu 5126 CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA 197 Leu Thr Leu Ala Ser Arg Gln Gln Leu IIe 5156 GAC TGG ATG GAG GCG GAT AAA GTT GCA GGA 207 Asp Trp Met Glu Ala Asp Lys Val Ala Gly 5186 CCA CTT CTG CGC TCG GCC CTT CCG GCT GGC 217 Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly 5216 TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT 227 Trp Phe IIe Ala Asp Lys Ser Gly Ala Gly 5246 GAG CGT GGG TCT CGC GGT ATC ATT GCA GCA 237 Glu Arg Gly Ser Arg Gly Ile Ile Ala Ala 5276 CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC 247 Leu Giy Pro Asp Giy Lys Pro Ser Arg IIe 5306 GTA GTT ATC TAC ACG ACG GGG AGT CAG GCA 257 Val Val IIe Tyr Thr Thr Gly Ser Gln Ala

Figure 13b (cont'd) 5336 ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT 267 Fhr Met Asp Glu Arg Asn Arg Gln Ile Ala 5366 GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG 277 ▶ Glu Ile Gly Ala Ser Leu Ile Lys His Trp 5396 TAA CTGTCAGACCAAGTTTACTCATATATACTTTAGAT 287 • • • 5434 TGATTTACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGG 5473 GTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCA 5512 GCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCT 5551 TTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAA 5590 ATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC 5629 GGCACCTCGACCCCAAAAAACTTGATTTGGGTGATGGTT 5668 CACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC 5707 CTTIGACGTIGGAGTCCACGTICTTTAATAGTGGACTCT 5746 TGTTCCAAACTTGAACAACACTCAACCCTATCTCGGGCT 5785 ATTCTTTGATTTATAAGGGATTTTGCCGATTTCGGCCT 5824 ATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACG 5863 CGAATTTTAACAAAATATTAACGTTTACAATTTAAAAGG 5902 ATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAA 5941 ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGAC 5980 CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTT 6019 TTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCA 6058 CCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTA 6097 CCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCG

Figure 1	3b (cont'd)
6136	CAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTA
6175	GGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATAC
6214	CTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGT
6253	GGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA
6292	TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG
6331	GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTAC
6370	ACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGC
6409	GCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCG
6448	GTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAG
6487	CTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTC
6526	GGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGA
6565	TGCTCGTCAGGGGGGGGGGGGGGGGCGAGCCTATGGAAAAACGCCAGC
6604	AACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCT
6643	TTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCT
6682	GTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC
6721	GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTG
6760	AGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTC
6799	CTTACGCATCTGTGCGGTATTTCACACCGCATAGGGTCA
6838	TGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGC
6877	CCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGAC
6916	AAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGT
6955	TTTCACCGTCATCACCGAAACGCGCGAGGCAGCAAGGAG

- 6994 ATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACC
- 7033 ATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGG
- 7072 CGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAG
- 7111 GCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCC
- 7150 ACGATGCGTCCGGCGTAGAGGATCTGCTCATGTTTGACA
- 7189 GCTTATC

Figure 14 a

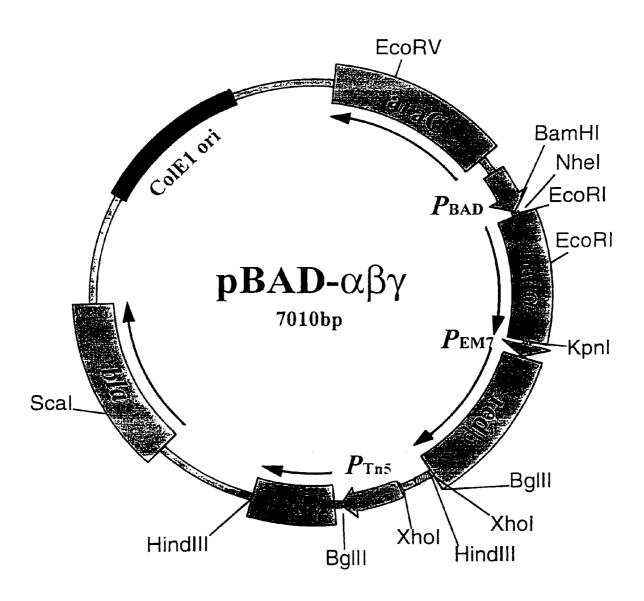


Figure 14b

Nsil

1 ATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGGG

40 ATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAATT

79 GTCTGATTCGTTACCAA TTA TGA CAA CTT GAC 293 **↓ • • • Ser** Leu Lys Val

111 GGC TAC ATC ATT CAC TIT TIC TIC ACA ACC 288 Ala Val Asp Asn Val Lys Glu Glu Cys Gly 141 GGC ACG GAA CTC GCT CGG GCT GGC CCC GGT 278 Ala Arg Phe Glu Ser Pro Ser Ala Gly Thr 171 GCA TIT TTT AAA TAC CCG CGA GAA ATA GAG 268 € Cys Lys Lys Phe Val Arg Ser Phe Tyr Leu 201 TTG ATC GTC AAA ACC AAC ATT GCG ACC GAC 258∢Gin Asp Asp Phe Gly Val Asn Arg Gly Val 231 GGT GGC GAT AGG CAT CCG GGT GGT GCT CAA 248◀Thr Ala Ile Pro Met Arg Thr Thr Ser Leu 261 AAG CAG CTT CGC CTG GCT GAT ACG TTG GTC 238∢Leu Leu Lys Ala GIn Ser lle Arg GIn Asp 291 CTC GCG CCA GCT TAA GAC GCT AAT CCC TAA 228 € Glu Arg Trp Ser Leu Val Ser lle Gly Leu 321 CTG CTG GCG GAA AAG ATG TGA CAG ACG CGA 218 Gin Gin Arg Phe Leu His Ser Leu Arg Ser 351 CGG CGA CAA GCA AAC ATG CTG TGC GAC GCT 208 ♥ Pro Ser Leu Cys Val His Gln Ala Val Ser **E**coRV 381 GGC GAT ATC AAA ATT GCT GTC TGC CAG GTG 198∢Ala Ile Asp Phe Asn Ser Asp Ala Leu His

411 ATC GCT GAT GTA CTG ACA AGC CTC GCG TAC 188 Asp Ser Ile Tyr Gin Cys Ala Glu Arg Val 441 CCG ATT ATC CAT CGG TGG ATG GAG CGA CTC 178 ¶Arg Asn Asp Met Pro Pro His Leu Ser Glu 471 GTT AAT CGC TTC CAT GCG CCG CAG TAA CAA 168 Asn lie Ala Glu Met Arg Arg Leu Leu Leu 501 TTG CTC AAG CAG ATT TAT CGC CAG CAG CTC 1584GIn Glu Leu Leu Asn Ile Ala Leu Leu Glu 531 CGA ATA GCG CCC TTC CCC TTG CCC GGC GTT 148 € Ser Tyr Arg Gly Glu Gly Gin Gly Ala Asn 561 AAT GAT TTG CCC AAA CAG GTC GCT GAA ATG 138 Ile Ile GIn Gly Phe Leu Asp Ser Phe His 591 CGG CTG GTG CGC TTC ATC CGG GCG AAA GAA 128 ¶ Pro GIn His Ala Glu Asp Pro Arg Phe Phe 621 CCC CGT ATT GGC AAA TAT TGA CGG CCA GTT 118 € Gly Thr Asn Ala Phe Ile Ser Pro Trp Asn 651 AAG CCA TTC ATG CCA GTA GGC GCG CGG ACG 1084 Leu Trp Glu His Trp Tyr Ala Arg Pro Arg 681 AAA GTA AAC CCA CTG GTG ATA CCA TTC GCG 98 € Phe Tyr Val Trp Gln His Tyr Trp Glu Arg 711 AGC CTC CGG ATG ACG ACC GTA GTG ATG AAT 88 Ala Glu Pro His Arg Gly Tyr His His Ile 741 CTC TCC TGG CGG GAA CAG CAA AAT ATC ACC 78 € Glu Gly Pro Pro Phe Leu Leu IIe Asp Gly 771 CGG TCG GCA AAC AAA TTC TCG TCC CTG ATT 68 Pro Arg Cys Val Phe Glu Arg Gly Gln Asn

Figure 14b (cont'd) 801 TTT CAC CAC CCC CTG ACC GCG AAT GGT GAG 58 Lys Val Val Gly Gln Gly Arg lle Thr Leu 831 ATT GAG AAT ATA ACC TIT CAT TCC CAG CGG 48 Asn Leu IIe Tyr Gly Lys Met Gly Leu Pro 861 TCG GTC GAT AAA AAA ATC GAG ATA ACC GTT 38 Arg Asp lie Phe Phe Asp Leu Tyr Giv Asn 891 GGC CTC AAT CGG CGT TAA ACC CGC CAC CAG 28 Ala Glu Ile Pro Thr Leu Gly Ala Val Leu 921 ATG GGC ATT AAA CGA GTA TCC CGG CAG CAG 18 His Ala Asn Phe Ser Tyr Gly Pro Leu Leu 951 GGG ATC ATT TIG CGC TIC AGC CAT ACTITIC 8 Pro Asp Asn Gin Ala Giu Ala Met 982 ATACTCCCGCCATTCAGAGAAGAAACCAATTGTCCATAT 1021 TGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTC 1060 TTCTCGCTAACCAAACCGGTAACCCCGCTTATTAAAAGC 1099 ATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACG 1138 CGTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCAC 1177 ATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATA BamHI 1216 GCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGC 1255 TTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTT Nhei EcoRI 1294 TTTTGGGCTAGCAGGAGGAATTCACC ATG ACA CCG 1 Met Thr Pro Pstl 1329 GAC ATT ATC CTG CAG CGT ACC GGG ATC GAT

4 Asp lie lie Leu Gin Arg Thr Gly lie Asp 1359 GTG AGA GCT GTC GAA CAG GGG GAT GAT GCG 14 Val Arg Ala Val Glu Gln Gly Asp Asp Ala 1389 TGG CAC AAA TTA CGG CTC GGC GTC ATC ACC 24 Frp His Lys Leu Arg Leu Gly Val lle Thr 1419 GCT TCA GAA GTT CAC AAC GTG ATA GCA AAA 34 Ala Ser Glu Val His Asn Val IIe Ala Lys 1449 CCC CGC TCC GGA AAG AAG TGG CCT GAC ATG 44 Pro Arg Ser Gly Lys Lys Trp Pro Asp Met 1479 AAA ATG TCC TAC TTC CAC ACC CTG CTT GCT 54 Lys Met Ser Tyr Phe His Thr Leu Leu Ala 1509 GAG GTT TGC ACC GGT GTG GCT CCG GAA GTT 64 Giu Val Cys Thr Gly Val Ala Pro Glu Val 1539 AAC GCT AAA GCA CTG GCC TGG GGA AAA CAG 74 Asn Ala Lys Ala Leu Ala Trp Gly Lys Gln EcoRI 1569 TAC GAG AAC GAC GCC AGA ACC CTG TTT GAA 84 Fyr Glu Asn Asp Ala Arg Thr Leu Phe Glu 1599 TTC ACT TCC GGC GTG AAT GTT ACT GAA TCC 94 Phe Thr Ser Gly Val Asn Val Thr Glu Ser 1629 CCG ATC ATC TAT CGC GAC GAA AGT ATG CGT 104 Pro IIe IIe Tyr Arg Asp Glu Ser Met Arg 1659 ACC GCC TGC TCT CCC GAT GGT TTA TGC AGT 114 Thr Ala Cys Ser Pro Asp Gly Leu Cys Ser 1689 GAC GGC AAC GGC CTT GAA CTG AAA TGC CCG 124 Asp Gly Asn Gly Leu Glu Leu Lys Cys Pro

1719 134		ACC Thr								
1749 144)		GGT Gl y								
1779 154		ATG Met								TGG T rp
1809 164		ACG Thr								
1839 174		TAT Tyr								GGC Gl y
1869 184		CAT Hi s								
1899 194)		TAC Tyr					-			
1929 204		GAG GI u								
1959 214		GCT Ala			Gl y					
1989 224		ТGG T rp				•	CCGF	AGCAC	GTGI	TGA
2025	CAAT	rtaat	CATO	CGGCZ	TAGT	ראדאי	CGGC		TATA	ATA
2064	CGAC	CAAGO	STGAC	GAAC	TAAA		ATG A Net S			
2098 5►		GCA Ala								

Sall 2128 CGT GTC GGC ATG GAT TCT GTC GAC CCA CAG 15 Arg Val Gly Met Asp Ser Val Asp Pro Gln 2158 GAA CTG ATC ACC ACT CTT CGC CAG ACG GCA 25 Glu Leu IIe Thr Thr Leu Arg Gln Thr Ala 2188 TTT AAA GGT GAT GCC AGC GAT GCG CAG TTC 35 Phe Lys Gly Asp Ala Ser Asp Ala Gln Phe 2218 ATC GCA TTA CTG ATC GTT GCC AAC CAG TAC 45 Ile Ala Leu Leu Ile Val Ala Asn Gln Tyr 2248 GGC CTT AAT CCG TGG ACG AAA GAA ATT TAC 55 Gly Leu Asn Pro Trp Thr Lys Glu Ile Tyr 2278 GCC TIT CCT GAT AAG CAG AAT GGC ATC GTT 65 Ala Phe Pro Asp Lys Gln Asn Gly IIe Val 2308 CCG GTG GTG GGC GTT GAT GGC TGG TCC CGC 75 Pro Val Val Gly Val Asp Gly Trp Ser Arg 2338 ATC ATC AAT GAA AAC CAG CAG TTT GAT GGC 85 ▶ Ile Ile Asn Glu Asn Gln Gln Phe Asp Gly 2368 ATG GAC TIT GAG CAG GAC AAT GAA TCC TGT 95 Met Asp Phe Glu Gln Asp Asn Glu Ser Cys 2398 ACA TGC CGG ATT TAC CGC AAG GAC CGT AAT 105 Thr Cys Arg Ile Tyr Arg Lys Asp Arg Asn 2428 CAT CCG ATC TGC GTT ACC GAA TGG ATG GAT 115 His Pro Ile Cys Val Thr Glu Trp Met Asp 2458 GAA TGC CGC CGC GAA CCA TTC AAA ACT CGC 125 Glu Cys Arg Arg Glu Pro Phe Lys Thr Arg 2488 GAA GGC AGA GAA ATC ACG GGG CCG TGG CAG 135 Glu Gly Arg Glu Ile Thr Gly Pro Trp Gln

Figure 14b (cont'd) 2518 TCG CAT CCC AAA CGG ATG TTA CGT CAT AAA 145 Ser His Pro Lys Arg Met Leu Arg His Lys 2548 GCC ATG ATT CAG TGT GCC CGT CTG GCC TTC 155 Ala Met lie Gin Cys Ala Arg Leu Ala Phe 2578 GGA TTT GCT GGT ATC TAT GAC AAG GAT GAA 165 Gly Phe Ala Gly Ile Tyr Asp Lys Asp Glu 2608 GCC GAG CGC ATT GTC GAA AAT ACT GCA TAC 175 Ala Glu Arg lle Val Glu Asn Thr Ala Tyr Pstl 2638 ACT GCA GAA CGT CAG CCG GAA CGC GAC ATC 185 Thr Ala Glu Arg Gln Pro Glu Arg Asp lle 2668 ACT CCG GTT AAC GAT GAA ACC ATG CAG GAG 195▶Thr Pro Val Asn Asp Glu Thr Met Gln Glu 2698 ATT AAC ACT CTG CTG ATC GCC CTG GAT AAA 205 Ile Asn Thr Leu Leu Ile Ala Leu Asp Lys. 2728 ACA TGG GAT GAC GAC TTA TIG CCG CIC TGT 215 Thr Trp Asp Asp Asp Leu Leu Pro Leu Cys 2758 TCC CAG ATA TTT CGC CGC GAC ATT CGT GCA 225 ▶ Ser Gin Ile Phe Arg Arg Asp Ile Arg Ala 2788 TCG TCA GAA CTG ACA CAG GCC GAA GCA GTA 235 Ser Ser Glu Leu Thr Gln Ala Glu Ala Val 2818 AAA GCT CTT GGA TTC CTG AAA CAG AAA GCC 245 Lys Ala Leu Gly Phe Leu Lys Gln Lys Ala Balll Xhol 2848 GCA GAG CAG AAG GTG GCA GCA TAGATCTCGAG 255 Ala Glu Gin Lys Val Ala Ala •••

Figure 14	4b (cont'd) Hind[]]
2880	AAGCTTCCTGCTGAACATCAAAGGCAAGAAAACATCTGT
2919	TGTCAAAGACAGCATCCTTGAACAAGGACAATTAACAGT
2958	TAACAAATAAAAACGCAAAAGAAAATGCCGATATCCTAT
2997	TGGCATTTTCTTTTTATTTCTTATCAACATAAAGGTGAAT Xhol
3036	CCCATACCTCGAGCTTCACGCTGCCGCAAGCACTCAGGG
3075	CGCAAGGGCTGCTAAAAGGAAGCGGAACACGTAGAAAGC
3114	CAGTCCGCAGAAACGGTGCTGACCCCGGATGAATGTCAG
3153	CTACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAA
3192	GAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATA
3231	GCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGA
3270	Pvull ATTGCCAGCTGGGGGCGCCCTCTGGTAAGGTTGGGAAGCC
3309	CTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGAT
3348	BgIII CTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGG
3387	ATGAGGATCGTTTCGC ATG GAT ATT AAT ACT 1 Met Asp lie Asn Thr
	GAA ACT GAG ATC AAG CAA AAG CAT TCA CTA Glu Thr Glu lle Lys Gin Lys His Ser Leu
	ACC CCC TIT CCT GIT TIC CIA AIC AGC CCG Thr Pro Phe Pro Val Phe Leu lle Ser Pro
3478 26	GCA TTT CGC GGG CGA TAT TTT CAC AGC TAT Ala Phe Arg Gly Arg Tyr Phe His Ser Tyr
	TTC AGG AGT TCA GCC ATG AAC GCT TAT TAC Phe Arg Ser Ser Ala Met Asn Ala Tyr Tyr

3538 ATT CAG GAT CGT CTT GAG GCT CAG AGC TGG 46▶ lle Gin Asp Arg Leu Giu Ala Gin Ser Trp 3568 GCG CGT CAC TAC CAG CAG CTC GCC CGT GAA 56 Ala Arg His Tyr Gln Gln Leu Ala Arg Glu 3598 GAG AAA GAG GCA GAA CTG GCA GAC GAC ATG 66▶Glu Lys Glu Ala Glu Leu Ala Asp Asp Met 3628 GAA AAA GGC CTG CCC CAG CAC CTG TTT GAA 76 ▶ Glu Lys Gly Leu Pro Gln His Leu Phe Glu 3658 TCG CTA TGC ATC GAT CAT TTG CAA CGC CAC 86▶ Ser Leu Cys IIe Asp His Leu GIn Arg His 3688 GGG GCC AGC AAA AAA TCC ATT ACC CGT GCG 96▶Gly Ala Ser Lys Lys Ser lle Thr Arg Ala 3718 TTT GAT GAC GAT GTT GAG TTT CAG GAG CGC 106 Phe Asp Asp Asp Val Glu Phe Gln Glu Arg 3748 ATG GCA GAA CAC ATC CGG TAC ATG GTT GAA 116 Met Ala Glu His Ile Arg Tyr Met Val Glu 3778 ACC ATT GCT CAC CAC CAG GTT GAT ATT GAT 126 Thr Ile Ala His His Gln Val Asp Ile Asp HindIII 3808 TCA GAG GTA TAA AACGAGTAGA AGC TIG GCT 136▶ Ser Giu Val ••• 3839 GTT TIG GCG GAT GAG AGA AGA TTT TCA GCC 3869 TGA TACAGATTAAATCAGAACGCAGAAGCGGTCTGATA 3907 AAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCA 3946 CCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGC 3985 GCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGG

4024 AACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAA

- 4063 AGACTGGGCCTTTCGTTTTATCTGTTGTTGTCGGTGAA
- 4102 CGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTT
- 4141 GAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGG
- 4180 ACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAA
- 4219 GGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAAC
- 4258 TCTTTTGTTTATTTTTCTAAATACATTCAAATATGTATC
- 4297 CGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT
- 4336 ATTGAAAAAGGAAGAGT ATG AGT ATT CAA CAT 1 Met Ser Ile GIn His

4368 TTC CGT GTC GCC CTT ATT CCC TTT TTT GCG 6 Phe Arg Val Ala Leu IIe Pro Phe Phe Ala 4398 GCA TIT TGC CTT CCT GTT TIT GCT CAC CCA 16 Ala Phe Cys Leu Pro Val Phe Ala His Pro 4428 GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA 26 Glu Thr Leu Val Lys Val Lys Asp Ala Glu 4458 GAT CAG TTG GGT GCA CGA GTG GGT TAC ATC 36 Asp GIn Leu Gly Ala Arg Val Gly Tyr IIe 4488 GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT 46 Glu Leu Asp Leu Asn Ser Gly Lys IIe Leu 4518 GAG AGT TTT CGC CCC GAA GAA CGT TTT CCA 56 Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro 4548 ATG ATG AGC ACT TTT AAA GTT CTG CTA TGT 66 Met Met Ser Thr Phe Lys Val Leu Leu Cys

4578 GGC GCG GTA TTA TCC CGT GTT GAC GCC GGG 76 Gly Ala Val Leu Ser Arg Val Asp Ala Gly 4608 CAA GAG CAA CTC GGT CGC CGC ATA CAC TAT 86▶GIn Glu GIn Leu Gly Arg Arg Ile His Tyr Scal 4638 TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA 96▶ Ser Gin Asn Asp Leu Val Giu Tyr Ser Pro 4668 GTC ACA GAA AAG CAT CTT ACG GAT GGC ATG 106 Val Thr Glu Lys His Leu Thr Asp Gly Met 4698 ACA GTA AGA GAA TTA TGC AGT GCT GCC ATA 116 Thr Val Arg Glu Leu Cys Ser Ala Ala lle 4728 ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA 126 Thr Met Ser Asp Asn Thr Ala Ala Asn Leu 4758 CTT CTG ACA ACG ATC GGA GGA CCG AAG GAG 136 Leu Leu Thr Thr Ile Gly Gly Pro Lys Glu 4788 CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT 146 Leu Thr Ala Phe Leu His Asn Met Gly Asp 4818 CAT GTA ACT CGC CTT GAT CGT TGG GAA CCG 156 His Val Thr Arg Leu Asp Arg Trp Glu Pro 4848 GAG CTG AAT GAA GCC ATA CCA AAC GAC GAG 166 Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu 4878 CGT GAC ACC ACG ATG CCT GTA GCA ATG GCA 176 Arg Asp Thr Thr Met Pro Val Ala Met Ala 4908 ACA ACG TTG CGC AAA CTA TTA ACT GGC GAA 186 Thr Thr Leu Arg Lys Leu Leu Thr Gly Glu 4938 CTA CTT ACT CTA GCT TCC CGG CAA CAA TTA 196▶Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu

4968 ATA GAC TGG ATG GAG GCG GAT AAA GTT GCA 206 lie Asp Trp Met Glu Ala Asp Lys Val Ala 4998 GGA CCA CTT CTG CGC TCG GCC CTT CCG GCT 216 Giv Pro Leu Leu Arg Ser Ala Leu Pro Ala 5028 GGC TGG TTT ATT GCT GAT AAA TCT GGA GCC 226 Gly Trp Phe IIe Ala Asp Lys Ser Gly Ala 5058 GGT GAG CGT GGG TCT CGC GGT ATC ATT GCA 236 Gly Glu Arg Gly Ser Arg Gly Ile Ile Ala 5088 GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT 246 Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg 5118 ATC GTA GTT ATC TAC ACG ACG GGG AGT CAG 256 lle Val Val IIe Tyr Thr Thr Gly Ser GIn 5148 GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC 266 Ala Thr Met Asp Glu Arg Asn Arg Gln Ile 5178 GCT GAG ATA GGT GCC TCA CTG ATT AAG CAT 276 Ala Glu Ile Gly Ala Ser Leu Ile Lys His 5208 TGG TAA CTGTCAGACCAAGTTTACTCATATATACTTT 286 Trp ••• 5245 AGATTGATTTACGCGCCCTGTAGCGGCGCATTAAGCGCG 5323 GCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCT 5362 TCCTTTCTCGCCACGTTCGCCGGCTTTCCCCCGTCAAGCT 5401 CTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCT 5440 TTACGGCACCTCGACCCCAAAAAACTTGATTTGGGTGAT 5479 GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTT

Figure 1	4b (cont'd)
5518	CGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGA
5557	CTCTTGTTCCAAACTTGAACAACACTCAACCCTATCTCG
5596	GGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCG
5635	GCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTT
5674	AACGCGAATTTTTAACAAAATATTAACGTTTACAATTTAA
5713	AAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGAC
5752	CAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTC
5791	AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
5830	TTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
5869	ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGA
5908	GCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAG
5947	AGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTA
5986	GTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTAC
6025	ATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGC
6064	CAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAG
6103	ACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
6142	GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGAC
6181	CTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGA
6220	AAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
6259	TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAG
6298	GGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCC
6337	TGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTT

6376	GTGATGCTCGTCAGGGGGGGGGGGGGGGGGCGAGCCTATGGAAAAACGC
6415	CAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
6454	GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGA
6493	TTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGA
6532	TACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTC
6571	AGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTT
6610	TCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAGG
6649	GTCATGGCTGCGCCCGACACCCGCCAACACCCGCTGAC
6688	GCGCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTAC
6727	AGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAG
6766	AGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCAA
6805	GGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGC
6844	CACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAA
6883	GTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGAT
6922	ATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCC
6961	GGCCACGATGCGTCCGGCGTAGAGGATCTGCTCATGTTT
7000	GACAGCTTATC

DNA CLONING METHOD

This application is a divisional of U.S. Ser. No. 10/231,013, filed Aug. 30, 2002, and issued as U.S. Pat. No. 6,787,316 on Sep. 7, 2004, which is a divisional application of U.S. Ser. No. 5 09/555,510, filed Jun. 5, 2000 and issued as U.S. Pat. No. 6,509,156 on Jan. 21, 2003, which is a 371 of PCT/EP98/ 07945 filed Dec. 7, 1998, which claims priority of EP 98 118 756.0 filed Oct. 5, 1998 and EP 97 121 462.2 filed Dec. 5, 1997. The disclosure of the prior application(s) is hereby 10 incorporated by reference herein in their entirety.

DESCRIPTION

The invention refers to a novel method for cloning DNA 15 molecules using a homologous recombination mechanism between at least two DNA molecules. Further, novel reagent kits suitable for DNA cloning are provided.

Current methods for cloning foreign DNA in bacterial cells usually comprise the steps of providing a suitable bacterial ²⁰ vector, cleaving said vector with a restriction enzyme and in vitro-inserting a foreign DNA fragment in said vector. The resulting recombinant vectors are then used to transform bacteria. Although such cloning methods have been used successfully for about 20 years they suffer from several drawbacks. These drawbacks are, in particular, that the in vitro steps required for inserting foreign DNA in a vector are often very complicated and time-consuming, if no suitable restriction sites are available on the foreign DNA or the vector.

Furthermore, current methods usually rely on the presence 30 of suitable restriction enzyme cleavage sites in the vector into which the foreign DNA fragment is placed. This imposes two limitations on the final cloning product. First, the foreign DNA fragment can usually only be inserted into the vector at the position of such a restriction site or sites. Thus, the cloning 35 product is limited by the disposition of suitable restriction sites and cloning into regions of the vector where there is no suitable restriction site, is difficult and often imprecise. Second, since restriction sites are typically 4 to 8 base pairs in length, they occur a multiple number of times as the size of the $_{40}$ DNA molecules being used increases. This represents a practical limitation to the size of the DNA molecules that can be manipulated by most current cloning techniques. In particular, the larger sizes of DNA cloned into vectors such as cosmids, BACs, PACs and P1 s are such that it is usually 45 impractical to manipulate them directly by restriction enzyme based techniques. Therefore, there is a need for providing a new cloning method, from which the drawbacks of the prior art have at least partly been eliminated.

According to the present invention it was found that an 50 efficient homologous recombination mechanism between two DNA molecules occurs at usable frequencies in a bacterial host cell which is capable of expressing the products of the recE and recT genes or functionally related genes such as the red α and red β genes, or the phage P22 recombination 55 system (Kolodner et al., Mol. Microbiol. 11(1994) 23-30; Fenton, A. C. and Poteete, A. R., Virology 134 (1984)148-160; Poteete, A. R. and Fenton, A. C., Virology 134 (1984) 161-167). This novel method of cloning DNA fragments is termed "ET cloning". 60

The identification and characterization of the *E. coli* RecE and RecT proteins is described Gillen et al. (J. Bacteriol. 145 (1981), 521-532) and Hall et al. (J. Bacteriol. 175 (1993), 277-287). Hall and Kolodner (Proc. Natl. Acad. Sci. USA 91 (1994), 3205-3209) disclose in vitro homologous pairing and 65 strand exchange of linear double-stranded DNA and homologous circular single-stranded DNA promoted by the RecT

protein. Any references to the use of this method for the cloning of DNA molecules in cells cannot be found therein.

The recET pathway of genetic recombination in *E. coli* is known (Hall and Kolodner (1994), supra; Gillen et al. (1981), supra). This pathway requires the expression of two genes, recE and recT. The DNA sequence of these genes has been published (Hall et al., supra). The RecE protein is similar to bacteriophage proteins, such as $\lambda \exp or \lambda \operatorname{Red}\alpha$ (Gillen et al., J. Mol. Biol. 113 (1977), 27-41; Little, J. Biol. Chem. 242 (1967), 679-686; Radding and Carter, J. Biol. Chem. 246 (1971), 2513-2518; Joseph and Kolodner, J. Biol. Chem. 258 (1983), 10418-10424). The RecT protein is similar to bacteriophage proteins, such as $\lambda \beta$ -protein or $\lambda \operatorname{Red}\beta$ (Hall et al. (1993), supra; Muniyappa and Radding, J. Biol. Chem. 256 (1981), 12636-12639). The content of the above-cited documents is incorporated herein by reference.

Oliner et al. (Nucl. Acids Res. 21 (1993), 5192-5197) describe in vivo cloning of PCR products in *E. coli* by intermolecular homologous recombination between a linear PCR product and a linearized plasmid vector. Other previous attempts to develop new cloning methods based on homologous recombination in prokaryotes, too, relied on the use of restriction enzymes to linearise the vector (Bubeck et al., Nucleic Acids Res. 21 (1993), 3601-3602; Oliner et al., Nucleic Acids Res. 21 (1993), 5192-5197; Degryse, Gene 170 (1996), 45-50) or on the host-specific recA-dependent recombination system (Hamilton et al., J. Bacteriol. 171 (1989), 4617-4622; Yang et al., Nature Biotech. 15 (1997), 859-865; Dabert and Smith, Genetics 145 (1997), 877-889). These methods are of very limited applicability and are hardly used in practice.

The novel method of cloning DNA according to the present invention does not require in vitro treatments with restriction enzymes or DNA ligases and is therefore fundamentally distinct from the standard methodologies of DNA cloning. The method relies on a pathway of homologous recombination in *E. coli* involving the recE and recT gene products, or the redα and red β gene products, or functionally equivalent gene products. The method covalently combines one preferably linear and preferably extrachromosomal DNA fragment, the DNA fragment to be cloned, with one second preferably circular DNA vector molecule, either an episome or the endogenous host chromosome or chromosomes. It is therefore distinct from previous descriptions of cloning in *E. coli* by homologous recombination which either rely on the use of two linear DNA fragments or different recombination pathways.

The present invention provides a flexible way to use homologous recombination to engineer large DNA molecules including an intact >76 kb plasmid and the E. coli chromosome. Thus, there is practically no limitation of target choice either according to size or site. Therefore, any recipient DNA in a host cell, from high copy plasmid to the genome, is amenable to precise alteration. In addition to engineering large DNA molecules, the invention outlines new, restriction enzyme-independent approaches to DNA design. For example, deletions between any two chosen base pairs in a target episome can be made by choice of oligonucleotide homology arms. Similarly, chosen DNA sequences can be inserted at a chosen base pair to create, for example, altered protein reading frames. Concerted combinations of insertions and deletions, as well as point mutations, are also possible. The application of these strategies is particularly relevant to complex or difficult DNA constructions, for example, those intended for homologous recombinations in eukaryotic cells, e.g. mouse embryonic stem cells. Further, the present invention provides a simple way to position site specific recombi-

65

nation target sites exactly where desired. This will simplify applications of site specific recombination in other living systems, such as plants and mice.

A subject matter of the present invention is a method for cloning DNA molecules in cells comprising the steps:

- a) providing a host cell capable of performing homologous recombination,
- b) contacting in said host cell a first DNA molecule which is capable of being replicated in said host cell with a second DNA molecule comprising at least two regions 10 of sequence homology to regions on the first DNA molecule, under conditions which favour homologous recombination between said first and second DNA molecules and
- c) selecting a host cell in which homologous recombina- 15 tion between said first and second DNA molecules has occurred

In the method of the present invention the homologous recombination preferably occurs via the recET mechanism, i.e. the homologous recombination is mediated by the gene 20 products of the recE and the recT genes which are preferably selected from the E. coli genes recE and recT or functionally related genes such as the phage λ red α and red β genes.

The host cell suitable for the method of the present invention preferably is a bacterial cell, e.g. a gram-negative bacte- 25 rial cell. More preferably, the host cell is an enterobacterial cell, such as Salmonella, Klebsiella or Escherichia. Most preferably the host cell is an Escherichia coli cell. It should be noted, however, that the cloning method of the present invention is also suitable for eukaryotic cells, such as fungi, plant or 30 animal cells.

Preferably, the host cell used for homologous recombination and propagation of the cloned. DNA can be any cell, e.g. a bacterial strain in which the products of the recE and recT, or red α and red β , genes are expressed. The host cell may 35 comprise the recE and recT genes located on the host cell chromosome or on non-chromosomal DNA, preferably on a vector, e.g. a plasmid. In a preferred case, the RecE and RecT, or Red α and Red β , gene products are expressed from two different regulatable promoters, such as the arabinose-induc- 40 ible BAD promoter or the lac promoter or from non-regulatable promoters. Alternatively, the recE and recT, or red α and red β , genes are expressed on a polycistronic mRNA from a single regulatable or non-regulatable promoter. Preferably the expression is controlled by regulatable promoters.

Especially preferred is also an embodiment, wherein the recE or red α gene is expressed by a regulatable promoter. Thus, the recombinogenic potential of the system is only elicited when required and, at other times, possible undesired recombination reactions are limited. The recT or red β gene, 50 on the other hand, is preferably overexpressed with respect to recE or red α . This may be accomplished by using a strong constitutive promoter, e.g. the

EM7 promoter and/or by using a higher copy number of recT, or red β , versus recE, or red α , genes.

For the purpose of the present invention any recE and recT genes are suitable insofar as they allow a homologous recombination of first and second DNA molecules with sufficient efficiency to give rise to recombination products in more than 1 in 10⁹ cells transfected with DNA. The recE and recT genes 60 may be derived from any bacterial strain or from bacteriophages or may be mutants and variants thereof. Preferred are recE and recT genes which are derived from E. coli or from E. *coli* bacteriophages, such as the red α and red β genes from lambdoid phages, e.g. bacteriophage λ .

More preferably, the recE or red α gene is selected from a nucleic acid molecule comprising(a) the nucleic acid sequence from position 1320 (ATG) to 2159 (GAC) as depicted in FIG. 7B or SEQ ID No. 2,

(b) the nucleic acid sequence from position 1320 (ATG) to 1998 (CGA) as depicted in FIG. 14B or SEQ ID No. 11,

- 5 (c) a nucleic acid encoding the same polypeptide within the degeneracy of the genetic code and/or
 - (d) a nucleic acid sequence which hybridizes under stringent conditions with the nucleic acid sequence from (a), (b) and/or (c).

More preferably, the recT or red β gene is selected from a nucleic acid molecule comprising

- (a) the nucleic acid sequence from position 2155 (ATG) to 2961 (GAA) as depicted in FIG. 7B or SEQ ID No. 4,
- (b) the nucleic acid sequence from position 2086 (ATG) to 2868 (GCA) as depicted in FIG. 14B or SEQ ID No. 11,
- (c) a nucleic acid encoding the same polypeptide within the degeneracy of the genetic code and/or
- (d) a nucleic acid sequence which hybridizes under stringent conditions with the nucleic acid sequences from (a), (b) and/or (c).

It should be noted that the present invention also encompasses mutants and variants of the given sequences, e.g. naturally occurring mutants and variants or mutants and variants obtained by genetic engineering. Further it should be noted that the recE gene depicted in FIG. 7B is an already truncated gene encoding amino acids 588-866 of the native protein. Mutants and variants preferably have a nucleotide sequence identity of at least 60%, preferably of at least 70% and more preferably of at least 80% of the recE and recT sequences depicted in FIGS. 7B and 13B, and of the red α and red β sequences depicted in FIG. 14B.

According to the present invention hybridization under stringent conditions preferably is defined according to Sambrook et al. (1989), infra, and comprises a detectable hybridization signal after washing for 30 min in 0.1×SSC, 0.5% SDS at 55° C., preferably at 62° C. and more preferably at 68° C.

In a preferred case the recE and recT genes are derived from the corresponding endogenous genes present in the E. coli K12 strain and its derivatives or from bacteriophages. in particular, strains that carry the sbcA mutation are suitable. Examples of such strains are JC8679 and JC 9604 (Gillen et al. (1981), supra). Alternatively, the corresponding genes may also be obtained from other coliphages such as lambdoid 45 phages or phage P22

The genotype of JC 8679 and JC 9604 is Sex (Hfr, F+, F-, or F'): F-.JC 8679 comprises the mutations: recBC 21, recC 22, sbcA 23, thr-1, ara-14, leu B 6, DE (gpt-proA) 62, lacY1, tsx-33, gluV44 (AS), galK2 (Oc), LAM-, his-60, relA 1, rps L31 (strR), xyl A5, mtl-1, argE3 (Oc) and thi-1. JC 9604 comprises the same mutations and further the mutation recA 56.

Further, it should be noted that the recE and recT, or red α and red β , genes can be isolated from a first donor source, e.g. a donor bacterial cell and transformed into a second receptor source, e.g. a receptor bacterial or eukaryotic cell in which they are expressed by recombinant DNA means.

In one embodiment of the invention, the host cell used is a bacterial strain having an sbcA mutation, e.g. one of E. coli strains JC 8679 and JC 9604 mentioned above. However, the method of the invention is not limited to host cells having an sbcA mutation or analogous cells. Surprisingly, it has been found that the cloning method of the invention also works in cells without sbcA mutation, whether recBC+ or recBC-, e.g. also in prokaryotic recBC+ host cells, e.g. in E. coli recBC+ cells. In that case preferably those host cells are used in which the product of a recBC type exonuclease inhibitor gene is

expressed. Preferably, the exonuclease inhibitor is capable of inhibiting the host recBC system or an equivalent thereof. A suitable example of such exonuclease inhibitor gene is the λ red γ gene (Murphy, J. Bacteriol. 173 (1991), 5808-5821) and functional equivalents thereof, respectively, which, for 5 example, can be obtained from other coliphages such as from phage P22 (Murphy, J. Biol. Chem. 269 (1994), 22507-22516).

More preferably, the exonuclease inhibitor gene is selected from a nucleic acid molecule comprising

- (a) the nucleic acid sequence from position 3588 (ATG) to 4002 (GTA) as depicted in FIG. **13**B or SEQ ID No. 10 or 11,
- (b) a nucleic acid encoding the same polypeptide within the degeneracy of the genetic code and/or

(c) a nucleic acid sequence which hybridizes under stringent conditions as defined above with the nucleic acid sequence from (a) and/or (b).

Surprisingly, it has been found that the expression of an exonuclease inhibitor gene in both recBC+ and recBC- strains leads to significant improvement of cloning efficiency.

The cloning method according to the present invention employs a homologous recombination between a first DNA molecule and a second DNA molecule. The first DNA mol-25 ecule can be any DNA molecule that carries an origin of replication which is operative in the host cell, e.g. an E. coli replication origin. Further, the first DNA molecule is present in a form which is capable of being replicated in the host cell. The first DNA molecule, i.e. the vector, can be any extrachromosomal DNA molecule containing an origin of replication which is operative in said host cell, e.g. a plasmid including single, low, medium or high copy plasmids or other extrachromosomal circular DNA molecules based on cosmid, P1, BAC or PAC vector technology. Examples of such vectors 35 are described, for example, by Sambrook et al. (Molecular Cloning, Laboratory Manual, 2nd Edition (1989), Cold Spring Harbor Laboratory Press) and loannou et al. (Nature Genet. 6 (1994), 84-89) or references cited therein. The first DNA molecule can also be a host cell chromosome, particu- $_{40}$ larly the E. coli chromosome. Preferably, the first DNA molecule is a double-stranded DNA molecule.

The second DNA molecule is preferably a linear DNA molecule and comprises at least two regions of sequence homology, preferably of sequence identity to regions on the 45 first DNA molecule. These homology or identity regions are preferably at least 15 nucleotides each, more preferably at least 20 nucleotides and, most preferably, at least 30 nucleotides each. Especially good results were obtained when using sequence homology regions having a length of about 40 50 or more nucleotides, e.g. 60 or more nucleotides. The two sequence homology regions can be located on the linear DNA fragment so that one is at one end and the other is at the other end, however they may also be located internally. Preferably, also the second DNA molecule is a double-stranded DNA 55 molecule.

The two sequence homology regions are chosen according to the experimental design. There are no limitations on which regions of the first DNA molecule can be chosen for the two sequence homology regions located on the second DNA mol-60 ecule, except that the homologous recombination event cannot delete the origin of replication of the first DNA molecule. The sequence homology regions can be interrupted by nonidentical sequence regions as long as sufficient sequence homology is retained for the homologous recombination 65 reaction. By using sequence homology arms having nonidentical sequence regions compared to the target site muta-

tions such as substitutions, e.g. point mutations, insertions and/or deletions may be introduced into the target site by ET cloning.

The second foreign DNA molecule which is to be cloned in the bacterial cell may be derived from any source. For example, the second DNA molecule may be synthesized by a nucleic acid amplification reaction such as a PCR where both of the DNA oligonucleotides used to prime the amplification contain in addition to sequences at the 3'-ends that serve as a primer for the amplification, one or the other of the two homology regions. Using oligonucleotides of this design, the DNA product of the amplification can be any DNA sequence suitable for amplification and will additionally have a sequence homology region at each end.

A specific example of the generation of the second DNA molecule is the amplification of a gene that serves to convey a phenotypic difference to the bacterial host cells, in particular, antibiotic resistance. A simple variation of this procedure involves the use of oligonucleotides that include other sequences in addition to the PCR primer sequence and the sequence homology region. A further simple variation is the use of more than two amplification primers to generate the amplification product. A further simple variation is the use of more than one amplification reaction to generate the amplification product. A further variation is the use of DNA fragments obtained by methods other than PCR, for example, by endonuclease or restriction enzyme cleavage to linearize fragments from any source of DNA.

It should be noted that the second DNA molecule is not necessarily a single species of DNA molecule. It is of course possible to use a heterogenous population of second DNA molecules, e.g. to generate a DNA library, such as a genomic or cDNA library.

The method of the present invention may comprise the contacting of the first and second DNA molecules in vivo. In one embodiment of the present invention the second DNA fragment is transformed into a bacterial strain that already harbors the first vector DNA molecule. In a different embodiment, the second DNA molecule and the first DNA molecule are mixed together in vitro before co-transformation in the bacterial host cell. These two embodiments of the present invention are schematically depicted in FIG. 1. The method of transformation can be any method known in the art (e.g. Sambrook et al. supra). The preferred method of transformation.

After contacting the first and second DNA molecules under conditions which favour homologous recombination between first and second DNA molecules via the ET cloning mechanism a host cell is selected, in which homologous recombination between said first and second DNA molecules has occurred. This selection procedure can be carried out by several different methods. In the following three preferred selection methods are depicted in FIG. **2** and described in detail below.

In a first selection method a second DNA fragment is employed which carries a gene for a marker placed between the two regions of sequence homology wherein homologous recombination is detectable by expression of the marker gene. The marker gene may be a gene for a phenotypic marker which is not expressed in the host or from the first DNA molecule. Upon recombination by ET cloning, the change in phenotype of the host strain conveyed by the stable acquisition of the second DNA fragment identifies the ET cloning product.

In a preferred case, the phenotypic marker is a gene that conveys resistance to an antibiotic, in particular, genes that convey resistance to kanamycin, ampillicin, chlorampheni-

col, tetracyclin or any other substance that shows bacteriocidal or bacteriostatic effects on the bacterial strain employed.

A simple variation is the use of a gene that complements a deficiency present within the bacterial host strain employed. 5 For example, the host strain may be mutated so that it is incapable of growth without a metabolic supplement. In the absence of this supplement, a gene on the second DNA fragment can complement the mutational defect thus permitting growth. Only those cells which contain the episome carrying the intended DNA rearrangement caused by the ET cloning step will grow.

In another example, the host strain carries a phenotypic marker gene which is mutated so that one of its codons is a stop codon that truncates the open reading frame. Expression 15 of the full length protein from this phenotypic marker gene requires the introduction of a suppressor tRNA gene which, once expressed, recognizes the stop codon and permits translation of the full open reading frame. The suppressor tRNA gene is introduced by the ET cloning step and successful 20 recombinants identified by selection for, or identification of, the expression of the phenotypic marker gene. In these cases, only those cells which contain the intended DNA rearrangement caused by the ET cloning step will grow.

A further simple variation is the use of a reporter gene that 25 conveys a readily detectable change in colony colour or morphology. In a preferred case, the green fluorescence protein (GFP) can be used and colonies carrying the ET cloning product identified by the fluorescence emissions of GFP. In another preferred case, the lacZ gene can be used and colonies 30 carrying the ET cloning product identified by a blue colony colour when X-gal is added to the culture medium.

In a second selection method the insertion of the second DNA fragment into the first DNA molecule by ET cloning alters the expression of a marker present on the first DNA 35 molecule. In this embodiment the first DNA molecule contains at least one marker gene between the two regions of sequence homology and homologous recombination may be detected by an altered expression, e.g. lack of expression of the marker gene. 40

In a preferred application, the marker present on the first DNA molecule is a counter-selectable gene product, such as the sacB, ccdB or tetracycline-resistance genes. In these cases, bacterial cells that carry the first DNA molecule unmodified by the ET cloning step after transformation with 45 the second DNA fragment, or co-transformation with the second DNA fragment and the first DNA molecule, are plated onto a medium so the expression of the counter-selectable marker conveys a toxic or bacteriostatic effect on the host. Only those bacterial cells which contain the first DNA mol- 50 ecule carrying the intended DNA rearrangement caused by the ET cloning step will grow.

In another preferred application, the first DNA molecule carries a reporter gene that conveys a readily detectable change in colony colour or morphology. In a preferred case, 55 the green fluorescence protein (GFP) can be present on the first DNA molecule and colonies carrying the first DNA molecule with or without the ET cloning product can be distinguished by differences in the fluorescence emissions of GFP. In another preferred case, the lacZ gene can be present on the 60 first DNA molecule and colonies carrying the first DNA molecule with or without the ET cloning product identified by a blue or white colony colour when X-gal is added to the culture medium

In a third selection method the integration of the second 65 DNA fragment into the first DNA molecule by ET cloning removes a target site for a site specific recombinase, termed

here an RT (for recombinase target) present on the first DNA molecule between the two regions of sequence homology. A homologous recombination event may be detected by removal of the target site.

In the absence of the ET cloning product, the RT is available for use by the corresponding site specific recombinase. The difference between the presence or not of this RT is the basis for selection of the ET cloning product. In the presence of this RT and the corresponding. site specific recombinase, the site specific recombinase mediates recombination at this RT and changes the phenotype of the host so that it is either not able to grow or presents a readily observable phenotype. In the absence of this RT, the corresponding site specific recombinase is not able to mediate recombination.

In a preferred case, the first DNA molecule to which the second DNA fragment is directed, contains two RTs, one of which is adjacent to, but not part of, an antibiotic resistance gene. The second DNA fragment is directed, by design, to remove this RT. Upon exposure to the corresponding site specific recombinase, those first DNA molecules that do not carry the ET cloning product will be subject to a site specific recombination reaction between the RTs that remove the antibiotic resistance gene and therefore the first DNA molecule fails to convey resistance to the corresponding antibiotic. Only those first DNA molecules that contain the ET cloning product, or have failed to be site specifically recombined for some other reason, will convey resistance to the antibiotic.

In another preferred case, the RT to be removed by ET cloning of the second DNA fragment is adjacent to a gene that complements a deficiency present within the host strain employed. In another preferred case, the RT to be removed by ET cloning of the second DNA fragment is adjacent to a reporter gene that conveys a readily detectable change in colony colour or morphology.

In another preferred case, the RT to be removed by ET cloning of the second DNA fragment is anywhere on a first episomal DNA molecule and the episome carries an origin of replication incompatible with survival of the bacterial host cell if it is integrated into the host genome. In this case the host genome carries a second RT, which may or may not be a mutated RT so that the corresponding site specific recombinase can integrate the episome, via its RT, into the RT sited in the host genome. Other preferred. RTs include RTs for site specific recombinases of the resolvase/transposase class. RTs include those described from existing examples of site specific recombination as well as natural or mutated variations thereof.

The preferred site specific recombinases include Cre, FLP, Kw or any site specific recombinase of the integrase class. Other preferred site specific recombinases include site specific recombinases of the resolvase/transposase class.

There are no limitations on the method of expression of the site specific recombinase in the host cell. In a preferred method, the expression of the site specific recombinase is regulated so that expression can be induced and quenched according to the optimisation of the ET cloning efficiency. In this case, the site specific recombinase gene can be either integrated into the host genome or carried on an episome. In another preferred case, the site specific recombinase is expressed from an episome that carries a conditional origin of replication so that it can be eliminated from the host cell.

In another preferred case, at least two of the above three selection methods are combined. A particularly preferred case involves a two-step use of the first selection method above, followed by use of the second selection method. This combined use requires, most simply, that the DNA fragment to be cloned includes a gene, or genes that permits the iden-

tification, in the first step, of correct ET cloning products by the acquisition of a phenotypic change. In a second step, expression of the gene or genes introduced in the first step is altered so that a second round of ET cloning products can be identified. In a preferred example, the gene employed is the 5 tetracycline resistance gene and the first step ET cloning products are identified by the acquisition of tetracycline resistance. In the second step, loss of expression of the tetracycline gene is identified by loss of sensitivity to nickel chloride, fusaric acid or any other agent that is toxic to the host cell 10 when the tetracycline gene is expressed. This two-step procedure permits the identification of ET cloning products by first the integration of a gene that conveys a phenotypic change on the host, and second by the loss of a related phenotypic change, most simply by removal of some of the DNA 15 sequences integrated in the first step. Thereby the genes used to identify ET cloning products can be inserted and then removed to leave ET cloning products that are free of these genes.

In a further embodiment of the present invention the ET ²⁰ cloning may also be used for a recombination method comprising the steps of

a) providing a source of RecE and RecT, or Red α and Red β , proteins,

b) contacting a first DNA molecule which is capable of ²⁵ in a host cell and being replicated in a suitable host cell with a second DNA molecule comprising at least two regions of sequence homology to regions on the first DNA molecule, under conditions which favour homologous recombination between said first and second DNA molecules and ³⁰

c) selecting DNA molecules in which a homologous recombination between said first and second DNA molecules has occurred.

The source of RecE and RecT, or Red α and Red β , proteins may be either purified or partially purified RecE and RecT, or Red α and Red β , proteins or cell extracts comprising RecE and RecT, or Red α and Red β , proteins.

The homologous recombination event in this embodiment may occur in vitro, e.g. when providing a cell extract containing further components required for homologous recombination. The homologous recombination event, however, may also occur in vivo, e.g. by introducing RecE and RecT, or Red α and Red β , proteins or the extract in a host cell (which may be recET positive or not, or red $\alpha\beta$ positive or not) and contacting the DNA molecules in the host cell. When the recombination occurs in vitro the selection of DNA molecules may be accomplished by transforming the recombination mixture in a suitable host cell and selecting for positive clones as described above. When the recombination occurs in vivo the selection methods as described above may directly be applied.

A further subject matter of the invention is the use of cells, preferably bacterial cells, most preferably, *E. coli* cells capable of expressing the recE and recT, or red α and red β , 55 genes as a host cell for a cloning method involving homologous recombination.

Still a further subject matter of the invention is a vector system capable of expressing recE and recT, or red α and red β , genes in a host cell and its use for a cloning method involving 60 homologous recombination. Preferably, the vector system is also capable of expressing an exonuclease inhibitor gene as defined above, e.g. the λ red γ gene. The vector system may comprise at least one vector. The recE and recT, or red α and red β , genes are preferably located on a single vector and more 65 preferably under control of a regulatable promoter which may be the same for both genes or a single promoter for each gene.

Especially preferred is a vector system which is capable of overexpressing the recT, or red β , gene versus the recE, or red β , gene.

Still a further subject matter of the invention is the use of a source of RecE and RecT, or Red α and Red β , proteins for a cloning method involving homologous recombination.

A still further subject matter of the invention is a reagent kit for cloning comprising

(a) a host cell, preferably a bacterial host cell,

- (b) means of expressing recE and recT, or red α and red β , genes in said host cell, e.g. comprising a vector system, and
- (c) a recipient cloning vehicle, e.g. a vector, capable of being replicated in said cell.

On the one hand, the recipient cloning vehicle which corresponds to the first DNA molecule of the process of the invention can already be present in the bacterial cell. On the other hand, it can be present separated from the bacterial cell.

In a further embodiment the reagent kit comprises

(a) a source for RecE and RecT, or Red α and Red $\beta,$ proteins and

(b) a recipient cloning vehicle capable of being propagated in a host cell and

(c) optionally a host cell suitable for propagating said recipient cloning vehicle.

The reagent kit furthermore contains, preferably, means for expressing a site specific recombinase in said host cell, in particular, when the recipient ET cloning product contains at least one site specific recombinase target site. Moreover, the reagent kit can also contain DNA molecules suitable for use as a source of linear DNA fragments used for ET cloning, preferably by serving as templates for PCR generation of the linear fragment, also as specifically designed DNA vectors from which the linear DNA fragment is released by restriction enzyme cleavage, or as prepared linear fragments included in the kit for use as positive controls. or other tasks. Moreover, the reagent kit can also contain nucleic acid amplification primers comprising a region of homology to said vector. Preferably, this region of homology is located at the 5'-end of the nucleic acid amplification primer.

The invention is further illustrated by the following Sequence listings, Figures and Examples.

- SEQ ID NO. 1: shows the nucleic acid sequence of the plasmid pBAD24-rec ET (FIG. 7).
- SEQ ID NOs 2/3: show the nucleic acid and amino acid sequences of the truncated recE gene (t-recE) present on pBAD24-recET at positions 1320-2162.
- SEQ ID NOs 4/5: show the nucleic acid and amino acid sequences of the recT gene present on pBAD24-recET at position 2155-2972.
- SEQ ID NOs 6/7: show the nucleic acid and amino acid sequences of the araC gene present on the complementary stand to the one shown of pBAD24-recET at positions 974-996.
- SEQ ID NOs 8/9: show the nucleic acid an amino acid sequences of the bla gene present on pBAD24-recET at positions 3493-4353.
- SEQ ID NO 10: shows the nucleic acid sequence of the plasmid pBAD-ET_Y (FIG. **13**).
- SEQ ID No 11: shows the nucleic acid sequence of the plasmid pBAD- $\alpha\beta\gamma$ (FIG. **14**) as well as the coding regions for the genes red α (1320-200), red β (2086-2871) and red γ (3403-3819).

SEQ ID NOs 12-14: show the amino acid sequences of the Reda, Red β and Red γ proteins, respectively. The red γ sequence is present on each of pBAD-ET γ (FIG. 13) and pBAD- $\alpha\beta\gamma$ (FIG. 14).

FIG. **1**

A preferred method for ET cloning is shown by diagram. The linear DNA fragment to be cloned is synthesized by PCR using oligonucleotide primers that contain a left homology arm chosen to match sequences in the recipient episome and a sequence. for priming in the PCR reaction, and a right 10 homology arm chosen to match another sequence in the recipient episome and a sequence for priming in the PCR reaction. The product of the PCR reaction, here a selectable marker gene (sm1), is consequently flanked by the left and right homology arms and can be mixed together in vitro with 15 the episome before co-transformation, or transformed into a host cell harboring the target episome. The host cell contains the products of the recE and recT genes. ET cloning products are identified by the combination of two selectable markers, sm1 and sm2 on the recipient episome. 20

FIG. 2

Three ways to identify ET cloning products are depicted. The first, (on the left of the figure), shows the acquisition, by ET cloning, of a gene that conveys a phenotypic difference to the host, here a selectable marker gene (sm). The second (in 25 the centre of the figure) shows the loss, by ET cloning, of a gene that conveys a phenotypic difference to the host, here a counter selectable marker gene (counter-sm). The third shows the loss of a target site (RT, shown as triangles on the circular episome) for a site specific recombinase (SSR), by ET cloning. In this case, the correct ET cloning product deletes one of the target sites required by the SSR to delete a selectable marker gene (sm). The failure of the SSR to delete the sm gene identifies the correct ET cloning product.

FIG. 3

A simple example of ET cloning is presented. (a) Top panel-PCR products (left lane) synthesized from oligonucleotides designed as described in FIG. 1 to amplify by PCR a kanamycin resistance gene and to be flanked by homology arms present in the recipient vector, were mixed in vitro with 40 the recipient vector (2nd lane) and cotransformed into a recET+E. coli host. The recipient vector carried an ampillicin resistance gene. (b) Transformation of the sbcA E. coli strain JC9604 with either the PCR product alone (0.2 µg) or the vector alone (0.3 µg) did not convey resistance to double 45 selection with ampicillin and kanamycin (amp+kan), however cotransformation of both the PCR product and the vector produced double resistant colonies. More than 95% of these colonies contained the correct ET cloning product where the kanamycin gene had precisely integrated into the recipient 50 vector according to the choice of homology arms. The two lanes on the right of (a) show Pvu II restriction enzyme digestion of the recipient vector before and after ET cloning. (c) As for b, except that six PCR products $(0.2 \,\mu\text{g each})$ were cotransformed with pSVpaZ11 (0.3 µg each) into JC9604 and 55 plated onto Amp+Kan plates or Amp plates. Results are plotted as Amp+Kan-resistant colonies, representing recombination products, divided by Amp-resistant colonies, representing the plasmid transformation efficiency of the competent cell preparation, $\times 10^6$. The PCR products were equivalent to 60 the a-b PCR product except that homology arm lengths were varied. Results are from five experiments that used the same batches of competent cells and DNAs. Error bars represent standard deviation. (d) Eight products flanked by 50 bp homology arms were cotransformed with pSVpaZ 11 into 65 JC9604. All eight PCR products contained the same left homology arm and amplified neo gene. The right homology

arms were chosen from the pSVpaZ11 sequence to be adjacent to (0), or at increasing distances (7-3100 bp), from the left. Results are from four experiments.

FIGS. **4**(*a*) and (*b*)

ET cloning in an approximately 100 kb P1 vector to exchange the selectable marker.

A P1 clone which uses a kanamycin resistance gene as selectable marker and which contains at least 70 kb of the mouse Hox a gene cluster was used. Before ET cloning, this episome conveys kanamycin resistance (top panel, upper left) to its host E. coli which are ampillicin sensitive (top panel, upper right). A linear DNA fragment designed to replace the kanamycin resistance gene with an ampillicin resistance gene was made by PCR as outlined in FIG. 1 and transformed into E. coli host cells in which the recipient Hox a/P1 vector was resident. ET cloning resulted in the deletion of the kanamycin resistance gene, and restoration of kanamycin sensitivity (top panel, lower left) and the acquisition of ampillicin resistance (top panel, lower right). Precise DNA recombination was verified by restriction digestion and Southern blotting analyses of isolated DNA before and after ET cloning (lower panel).

FIGS. **5**(*a*) and (*b*)

ET cloning to remove a counter selectable marker A PCR fragment (upper panel, left, third lane) made as outlined in FIGS. **1** and **2** to contain the kanamycin resistance gene was directed by its chosen homology arms to delete the counter selectable ccdB gene present in the vector, pZero-2.1. The PCR product and the pZero vector were mixed in vitro (upper panel, left, 1 st lane) before cotransformation into a recE/ 35 recT+*E*. *coli* host. Transformation of pZero-2.1 alone and plating onto kanamycin selection medium resulted in little colony growth (lower panel, left). Cotransformation of pZero-2.1 and the PCR product presented ET cloning products (lower panel, right) which showed the intended molecular event as visualized by Pvu II digestion (upper panel, right).

FIG. 6

ET cloning mediated by inducible expression of recE and recT from an episome.

RecE/RecT mediate homologous recombination between linear and circular DNA molecules. (a) The plasmid pBAD24-recET was transformed into E. coli JC5547, and then batches of competent cells were prepared after induction of RecE/RecT expression by addition of L-arabinose for the times indicated before harvesting. A PCR product, made using oligonucleotides e and f to contain the chloramphenicol resistance gene (cm) of pMAK705 and 50 bp homology arms chosen to flank the ampicililin resistance gene (bla) of pBAD24-recET, was then transformed and recombinants identified on chloramphenicol plates. (b) Arabinose was added to cultures of pBAD24-recET transformed JC5547 fordifferenttimes immediately before harvesting for competent cell preparation. Total protein expression was analyzed by SDS-PAGE and Coomassie blue staining. (c) The number of chloramphenicol resistant colonies per µg of PCR product was normalized against a control for transformation efficiency, determined by including 5 pg pZero2.1, conveying kanamycin resistance, in the transformation and plating an aliquot onto Kan plates.

The plasmid pBAD24-recET is shown by diagram. The plasmid contains the genes recE (in a truncated form) and recT under control of the inducible BAD promoter (P_{BAD}) The plasmid further contains an ampillicin resistance gene 5 (Amp') and an araC gene.

13

FIG. 7B

FIG. 7A

The nucleic acid sequence and the protein coding portions of pBAD24-recET are depicted.

FIG. 8

Manipulation of a large E. coli episome by multiple recombination steps. FIG. 8a depicts the scheme of the recombination reactions. A P1 clone of the Mouse Hoxa complex, resident in JC9604, was modified by recombination with PCR products that contained the neo gene and two Flp recombina- 15 tion targets (FRTs). The two PCR products were identical except that one was flanked by g and h homology arms (insertion), and the other was flanked by i and h homology arms (deletion). In a second step, the neo gene was removed by Flp recombination between the FRTs by transient transformation 20 of a Flp expression plasmid based on the pSC101 temperature-sensitive origin (ts ori). FIG. 8b (upper panel): ethidium bromide stained agarose gel showing EcoR1 digestions of P1 DNA preparations from three independent colonies for each step. FIG. 8b (middle panel): a Southern blot of the upper 25 panel hybridized with a neo gene probe. FIG. 8b (lower panel): a Southern blot of the upper panel hybridized with a Hoxa3 probe to visualize the site of recombination. Lane 1 in each of the panels shows the original Hoxa3 P1 clone grown in E. coli strain NS3145. Lane 2 in each of the panels shows 30 that replacement of the Tn903 kanamycin resistance gene in the P1 vector with an ampicillin resistance gene, increased the 8.1 kb band (lane 1) to 9.0 kb. Lane 3 in each of the panels shows that insertion of the Tn5-neo gene with g-h homology arms upstream of Hoxa3, increased the 6.7 kb band (lanes 35 1,2) to 9.0 kb. Lane 4 in each of the panels shows that Flp recombinase deleted the g-h neo gene reducing the 9.0 kb band (lane 3) back to 6.7 kb. Lane 5 in each of the panels shows that deletion of 6 kb of Hoxa3-4 intergenic DNA by replacement with the i-h neo gene, decreased the 6.7 kb band 40 (lane 2) to 4.5 kb. Lane 6 in each of the panels shows that Flp recombinase deleted the i-h neo gene reducing the 4.5 kb band to 2.3 kb.

FIG. **9**

Manipulation of the E. coli chromosome. FIG. 9a depicts 45 the scheme of the recombination reactions. The endogenous lacZ gene of JC9604 at 7.8' of the E. coli chromosome, shown in expanded form with relevant Ava I sites and coordinates, was targeted by a PCR fragment that contained the neo gene flanked by homology arms j and k, and loxP sites, as depicted. 50 Integration of the neo gene removed most of the lacZ gene including an Ava I site to alter the 1443 and 3027 bp bands into a 3277 bp band. In a second step, the neo gene was removed by Cre recombination between the loxPs by transient transformation of a Cre expression plasmid based on the 55 pSC101 temperature-sensitive origin (ts ori). Removal of the neo gene by Cre recombinase reduces the 3277 band to 2111 bp. FIG. 9b shows β -galactosidase expression evaluated by streaking colonies on X-Gal plates. The top row of three streaks show β -galactosidase expression in the host JC9604 60 strain (w.t.), the lower three rows (Km) show 24 independent primary colonies, 20 of which display a loss of β-galactosidase expression indicative of the intended recombination event. FIG. 9c shows the results from Southern analysis of E. coli chromosomal DNA digested with Ava I using a random 65 primed probe made from the entire lacZ coding region; lanes 1,2, w.t.; lanes 3-6, four independent white colonies after

integration of the j-k neo gene; lanes 7-10; the same four colonies after transformation with the Cre expression plasmid.

FIG. 10

Two rounds of ET cloning to introduce a point mutation. FIG. 10a depicts the scheme of the recombination reactions. The lacZ gene of pSVpaX1 was disrupted in JC9604lacZ, a strain made by the experiment of FIG. 9 to ablate endogenous lacZ expression and remove competitive sequences, by a sacB-neo gene cassette, synthesized by PCR to pIB279 and flanked by l and m homology arms. The recombinants, termed pSV-sacB-neo, were selected on Amp+Kan plates. The lacZ gene of pSV-sacB-neo was then repaired by a PCR fragment made from the intact lacZ gene using 1* and m* homology arms. The m* homology arm included a silent C to G change that created a BamH1 site. The recombinants, termed pSVpaX1*, were identified by counter selection against the sacB gene using 7% sucrose. FIG. 10b shows that β -galactosidase expression from pSVpaX1 was disrupted in pSV-sacB-neo and restored in pSVpaX1*. Expression was analyzed on X-gal plates. Three independent colonies of each pSV-sacBneo and pSVpaX1* are shown. FIG. 10c shows Ethidium bromide stained agarose gels of BamH1 digested DNA prepared from independent colonies taken after counter selection with sucrose. All β -galactosidase expressing colonies (blue) contained the introduced BamH1 restriction site (upper panel). All white colonies displayed large rearrangements and no product carried the diagnostic 1.5 kb BamH1 restriction fragment (lower panel).

FIG. **11**

Transferance of ET cloning into a recBC+ host to modify a large episome. FIG. 11a depicts the plasmid, pBAD-ETy, which carries the mobile ET system, and the strategy employed to target the Hoxa P1 episome. pBAD-ETy is based on pBAD24 and includes (i) the truncated recE gene (t-recE) under the arabinose-inducible P_{BAD} promoter; (ii) the recT gene under the EM7 promoter; and (iii) the redy gene under the Tn5 promoter. It was transformed into NS3145, a recA E. coli strain which contained the Hoxa P1 episome. After arabinose induction, competent cells were prepared and transformed with a PCR product carrying the chloramphenicol resistance gene (cm) flanked by n and p homology arms. n and p were chosen to recombine with a segment of the P1 vector. FIG. 11b shows the results from Southern blots of Pvu II digested DNAs hybridized with a probe made from the P1 vector to visualize the recombination target site (upper panel) and a probe made from the chloramphenicol resistance gene (lower panel). Lane 1, DNA prepared from cells harboring the Hoxa P1 episome before ET cloning. Lanes 2-17, DNA prepared from 16 independent chloramphenicol resistant colonies.

FIG. 12

Comparison of ET cloning using the recE/recT genes in pBAD-ET γ with red α /red β genes in pBAD- $\alpha\beta\gamma$.

The plasmids pBAD-ET γ or pBAD- $\alpha\beta\gamma$, depicted, were transformed into the *E. coli* recA–, recBC+ strain, DK1 and targeted by a chloramphenicol gene as described in FIG. **6** to evaluate ET cloning efficiencies. Arabinose induction of protein expression was for 1 hour.

FIG. **13**A

The plasmid pBAD-ETγ is shown by diagram.

FIG. **13**B

The nucleic acid sequence and the protein coding portions of pBAD-ET γ are depicted.

FIG. **14**A

The plasmid pBAD- $\alpha\beta\gamma$ is shown by diagram. This plasmid substantially corresponds to the plasmid shown in FIG.

10

25

55

60

13 except that the recE and recT genes are substituted by the red α and red β genes.

FIG. **14**B

The nucleic acid sequence and the protein coding portions of pBAD- $\alpha\beta\gamma$ are depicted.

1. Methods

1.1 Preparation of Linear Fragments

Standard PCR reaction condition were used to amplify linear DNA fragments.

Table 1

The Tn5-neo gene from pJP5603 (Penfold and Pemberton, Gene 118 (1992), 145-146) was amplified by using oligo pairs a/b and c/d. The chloramphenicol (cm) resistant gene from pMAK705 (Hashimoto-Gotoh and Sekiguchi, J. Bacteriol. 131 (1977), 405-412) was amplified by using primer pairs e/f and n/p. The Tn5-neo gene flanked by FRT or loxP sites was amplified from pKaZ or pKaX (http://www.emblheidelberg.de/ExternalInfo/stewart) using oligo pairs i/h, g/h and j/k. The sacB-neo cassette from plB279 (Blomfield et al., Mol. Microbiol. 5 (1991), 1447-1457) was amplified by using oligo pair l/m. The lacZ gene fragment from pSVpaZ11 (Buchholz et al., Nucleic Acids Res. 24 (1996), 4256-4262) was amplified using oligo pair 1*/m*. PCR products were purified using the QIAGEN PCR Purification Kit and eluted with H₂O₂, followed by digestion of any residual template DNA with Dpn I. After digestion, PCR products were extracted once with Phenol:CHCl₃, ethanol precipitated and resuspended in H₂O at approximately 0.5 µg/µl.

1.2 Preparation of Competent Cells and Electroporation

Saturated overnight cultures were diluted 50 fold into LB medium, grown to an OD600 of 0.5, following by chilling on ice for 15 min. Bacterial cells were centrifuged at 7,000 rpm for 10 min at 0° C. The pellet was resuspended in ice-cold 35 10% glycerol and centrifuged again (7,000 rpm, -5° C., 10 min). This was repeated twice more and the cell pellet was suspended in an equal volume of ice-cold 10% glycerol. Aliquots of 50 µl were frozen in liquid nitrogen and stored at -80° C. Cells were thaved on ice and 1 µl DNA solution 40 (containing, for co-transformation, 0.3 µg plasmid and 0.2 µg PCR products; or, for transformation, 0.2 µg PCR products) was added. Electroporation was performed using ice-cold cuvettes and a Bio-Rad Gene Pulser set to 25 µFD, 2.3 kV with Pulse Controller set at 200 ohms. LB medium (1 ml) was 45 added after electroporation. The cells were incubated at 37° C. for 1 hour with shaking and then spread on antibiotic plates.

1.3 Induction of RecE and RecT Expression E. coli JC5547 carrying pBAD24-recET was cultured overnight in LB medium plus 0.2% glucose, 100 µg/ml ampicillin. Five parallel LB cultures, one of which (0) included 0.2% glucose, were started by a 1/100 inoculation. The cultures were incubated at 37° C. with shaking for 4 hours and 0.1% L-arabinose was added 3, 2, 1 or 1/2 hour before harvesting and processing as above. Immediately before harvesting, 100 µl was removed for analysis on a 10% SDS-polyacrylamide gel. E. coli NS3145 carrying Hoxa-P1 and pBAD-ETy was induced by 0.1% L-arabinose for 90 min before harvesting.

1.4 Transient Transformation of FLP and Cre Expression Plasmids

The FLP and Cre expression plasmids, 705-Cre and 705-FLP (Buchholz et al, Nucleic Acids Res. 24 (1996), 3118-3119), based on the pSC101 temperature sensitive origin, 65 were transformed into rubidium chloride competent bacterial cells. Cells were spread on 25 µg/ml chloramphenicol plates,

and grown for 2 days at 30° C., whereupon colonies were picked, replated on L-agar plates without any antibiotics and incubated at 40° C. overnight. Single colonies were analyzed on various antibiotic plates and all showed the expected loss of chloramphenicol and kanamycin resistance.

1.5 Sucrose Counter Selection of SacB Expression

The E. coli JC9604lacZ strain, generated as described in FIG. 11, was cotransformed with a sacB-neo PCR fragment and pSVpaX1 (Buchholz et al, Nucleic Acids Res. 24 (1996), 4256-4262). After selection on 100 µg/ml ampicillin, 50 µg/ml kanamycin plates, pSVpaX-sacB-neo plasmids were isolated and cotransformed into fresh JC9604lacZ cells with a PCR fragment amplified from pSVpaX1 using primers 1*/m*. Oligo m* carried a silent point mutation which generated a BamHl site. Cells were plated on 7% sucrose, 100 µg/ml ampicillin, 40 µg/ml X-gal plates and incubated at 28° C. for 2 days. The blue and white colonies grown on sucrose 20 plates were counted and further checked by restriction analysis.

1.6 Other Methods

DNA preparation and Southern analysis were performed according to standard procedures. Hybridization probes were generated by random priming of fragments isolated from the Tn5 neo gene (PvuII), Hoxa3 gene (both HindIII fragments), lacZ genes (EcoR1 and BamH1 fragments from pSVpaX1), cm gene (BstB1 fragments from pMAK705) and P1 vector fragments (2.2 kb EcoR1 fragments from P1 vector).

2. Results

2.1 Identification of Recombination Events in E. coli

To identify a flexible homologous recombination reaction in E. coli, an assay based on recombination between linear and circular DNAs was designed (FIG. 1, FIG. 3). Linear DNA carrying the Tn5 kanamycin resistance gene (neo) was made by PCR (FIG. 3a). Initially, the oligonucleotides used for PCR amplification of neo were 60 mers consisting of 42 nucleotides at their 5' ends identical to chosen regions in the plasmid and, at the 3' ends, 18 nucleotides to serve as PCR primers. Linear and circular DNAs were mixed in equimolar proportions and co-transformed into a variety of E. coli hosts. Homologous recombination was only detected in sbcA E. coli hosts. More than 95% of double ampicillin/kanamycin resistant colonies (FIG. 3b) contained the expected homologously recombined plasmid as determined by restriction digestion and sequencing. Only a low background of kanamycin resistance, due to genomic integration of the neo gene, was apparent (not shown).

The linear plus circular recombination reaction was characterized in two ways. The relationship betweeen homology arm length and recombination efficiency was simple, with longer arms recombining more efficiently (FIG. 3c). Efficiency increased within the range tested, up to 60 bp. The effect of distance between the two chosen homology sites in the recipient plasmid was examined (FIG. 3d). A set of eight PCR fragments was generated by use of a constant left homology arm with differing right homology arms. The right homology arms were chosen from the plasmid sequence to be 0-3100 bp from the left. Correct products were readily obtained from all, with less than 4 fold difference between them, although the insertional product (0) was least efficient. Correct products also depended on the presence of both homology arms, since PCR fragments containing only one arm failed to work.

2.2 Involvement of RecE and RecT

The relationship between host genotype and this homologous recombination reaction was more systemically examined using a panel of E. *coli* strains deficient in various recombination components (Table 1)

Only the two sbcA strains, JC8679 and JC9604 presented the intended recombination products and RecA was not required. In sbcA strains, expression of RecE and RecT is activated. Dependence on recE can be inferred from comparison of JC8679 with JC8691. Notably no recombination products were observed in JC9387 suggesting that the sbcBC background is not capable of supporting homologous recombination based on 50 nucleotide homology arms.

To demonstrate that RecE and RecT are involved, part of the recET operon was cloned into an inducible expression 15 vector to create pBAD24-recET (FIG. 6a). the recE gene was truncated at its N-terminal end, as the first 588 a.a.s of RecE are dispensable. The recBC strain, JC5547, was transformed with pBAD24-recET and a time course of RecE/RecT induction performed by adding arabinose to the culture media at 20 various times before harvesting for competent cells. The batches of harvested competent cells were evaluated for protein expression by gel electrophoresis (FIG. 6b) and for recombination between a linear DNA fragment and the endogenous pBAD24-recET plasmid (FIG. 6c). Without 25 induction of RecE/RecT, no recombinant products were found, whereas recombination increased in approximate concordance with increased RecE/RecT expression. This experiment also shows that co-transformation of linear and circular DNAs is not essential and the circular recipient can be endog- 30 enous in the host. From the results shown in FIGS. 3, 6 and Table 2, we conclude that RecE and RecT mediate a very useful homologous recombination reaction in recBC E. coli at workable frequencies. Since RecE and RecT are involved, we refer to this way of recombining linear and circular DNA 35 fragments as "ET cloning".

2.3 Application of ET Cloning to Large Target DNAs

To show that large DNA episomes could be manipulated in E. coli, a >76 kb P1 clone that contains at least 59 kb of the intact mouse Hoxa complex, (confirmed by DNA sequencing and Southern blotting), was transferred to an E. coli strain having an sbcA background (JC9604) and subjected to two rounds of ET cloning. In the first round, the Tn903 kanamycin resistance gene resident in the P1 vector was replaced by an ampicillin resistance gene (FIG. 4). In the second round, the interval between the Hoxa3 and a4 genes was targeted either by inserting the neo gene between two base pairs upstream of the Hoxa3 proximal promoter, or by deleting 6203 bp between the Hoxa3 and a4 genes (FIG. 8a). Both insertional and deletional ET cloning products were readily obtained (FIG. 8b, lanes 2, 3 and 5) showing that the two rounds of ET cloning took place in this large E. coli episome with precision and no apparent unintended recombination.

The general applicability of ET cloning was further examined by targeting a gene in the *E. coli* chromosome (FIG. 9*a*). The β -galactosidase (lacZ) gene of JC9604 was chosen so that the ratio between correct and incorrect recombinants could be determined by evaluating β -galactosidase expression. Standard conditions (0.2 µg PCR fragment; 50 µl competent cells), produced 24 primary colonies, 20 of which were correct as determined by β -galactosidase expression (FIG. 9*b*), and DNA analysis (FIG. 9*c*, lanes 3-6).

2.4 Secondary Recombination Reactions to Remove Operational Sequences

The products of ET cloning as described above are limited by the necessary inclusion of selectable marker genes. Two

18

different ways to use a further recombination step to remove this limitation were developed. In the first way, site specific recombination mediated by either FIp or Cre recombinase was employed. In the experiments of FIGS. 8 and 9, either FIp recombination target sites (FRTs) or Cre recombination target sites (loxPs) were included to flank the neo gene in the linear substrates. Recombination between the FRTs or loxPs was accomplished by FIp or Cre, respectively, expressed from plasmids with the pSC101 temperature sensitive replication origin (Hashimoto-Gotoh and Sekiguchi, J. Bacteriol. 131 (1977), 405-412) to permit simple elimination of these plasmids after site specific recombination by temperature shift. The precisely recombined Hoxa P1 vector was recovered after both ET and FIp recombination with no other recombination products apparent (FIG. 8, lanes 4 and 6). Similarly, Cre recombinase precisely recombined the targeted lacZ allele (FIG. 9, lanes 7-10). Thus site specific recombination can be readily coupled with ET cloning to remove operational sequences and leave a 34 bp site specific recombination target site at the point of DNA manipulation.

In the second way to remove the selectable marker gene, two rounds of ET cloning, combining positive and counter selection steps, were used to leave the DNA product free of any operational sequences (FIG. 10a).

Additionally this experiment was designed to evaluate, by a functional test based on β -galactosidase activity, whether ET cloning promoted small mutations such as frame shift or point mutations within the region being manipulated. In the first round, the lacZ gene of pSVpaX1 was disrupted with a 3.3 kb PCR fragment carrying the neo and B. subtilis sacB (Blomfield et al., Mol. Microbiol. 5 (1991), 1447-1457) genes, by selection for kanamycin resistance (FIG. 10a). As shown above for other positively selected recombination products, virtually all selected colonies were white (FIG. 10b), indicative of successful lacZ disruption, and 17 of 17 were confirmed as correct recombinants by DNA analysis. In the second round, a 1.5 kb PCR fragment designed to repair lacZ was introduced by counter selection against the sacB gene. Repair of lacZ included a silent point mutation to create a BamHl restriction site. Approximately one quarter of sucrose resistant colonies expressed β -galactosidase, and all analyzed (17 of 17; FIG. 10c) carried the repaired lacZ gene with the BamHl point mutation. The remaining three quarters of sucrose resistant colonies did not express β -galactosidase, and all analyzed (17 of 17; FIG. 10c) had undergone a variety of large mutational events, none of which resembled the ET cloning product. Thus, in two rounds of ET cloning directed at the lacZ gene, no disturbances of β -galactosidase activity by small mutations were observed, indicating the RecE/RecT recombination works with high fidelity. The significant presence of incorrect products observed in the counter selection step is an inherent limitation of the use of counter selection, since any mutation that ablates expression of the counter selection gene will be selected. Notably, all incorrect products were large mutations and therefore easily distinguished from the correct ET product by DNA analysis. In a different experiment (FIG. 5), we observed that ET cloning into pZero2.1 (InVitroGen) by counter selection against the ccdB gene gave a lower background of incorrect products (8%), indicating that the counter selection background is variable according to parameters that differ from those that influence ET cloning efficiencies.

2.5 Transference of ET Cloning between E. coli Hosts

60

65

The experiments shown above were performed in recBC-*E. coli* hosts since the sbcA mutation had been identified as a suppressor of recBC (Barbour et al., Proc. Natl. Acad. Sci. USA 67 (1970), 128-135; Clark, Genetics 78 (1974), 259-271). However, many useful E. coli strains are recBC+, including strains commonly used for propagation of P1, BAC or PAC episomes. To transfer ET cloning into recBC+ strains, we developed pBAD-ETy and pBAD- $\alpha\beta\gamma$ (FIGS. 13 and 14). 5 These plasmids incorporate three features important to the mobility of ET cloning. First, RecBC is the major E. coli exonuclease and degrades introduced linear fragments. Therefore the RecBC inhibitor, Redy (Murphy, J. Bacteriol. 173 (1991), 5808-5821), was included. Second, the recombi- 10 nogenic potential of RecE/RecT, or Reda/Redb, was regulated by placing recE or red α under an inducible promoter. Consequently ET cloning can be induced when required and undesired recombination events which are restricted at other times. Third, we observed that ET cloning efficiencies are 15 enhanced when RecT, or Red β , but not RecE, or Red α , is overexpressed. Therefore we placed recT, or red β , under the strong, constitutive, EM7 promoter.

pBAD-ET γ was transformed into NS3145 *E. coli* harboring the original Hoxa P1 episome (FIG. 11*a*). A region in the 20 P1 vector backbone was targeted by PCR amplification of the chloramphenicol resistance gene (cm) flanked by n and p homology arms. As described above for positively selected ET cloning reactions, most (>90%) chloramphenicol resistant colonies were correct. Notably, the overall efficiency of 25 ET cloning, in terms of linear DNA transformed, was nearly three times better using pBAD-ET γ than with similar experiments based on targeting the same episome in the sbcA host,

<160> NUMBER OF SEQ ID NOS: 14

JC9604. This is consistent with our observation that overexpression of RecT improves ET cloning efficiencies.

A comparison between ET cloning efficiencies mediated by RecE/RecT, expressed from pBAD-ET γ , and Red α /Red β , expressed from pBAD- $\alpha\beta\gamma$ was made in the recA-, recBC+*E. coli* strain, DK1 (FIG. **12**). After transformation of *E. coli* DK1 with either pBAD-ET γ or pBAD- $\alpha\beta\gamma$, the same experiment as described in FIG. **6***a,c*, to replace the bla gene of the pBAD vector with a chloramphenicol gene was performed. Both pBAD-ET γ or pBAD- $\alpha\beta\gamma$ presented similar ET cloning efficiencies in terms of responsiveness to arabinose induction of RecE and Red α , and number of targeted events.

TABLE 1

<i>E. coli</i> Strains	Genotypes	Amp + Kan	Amp × 10 ⁸ /μg
JC8679	recBC sbcA	318	2.30
JC9604	recA recBC sbcA	114	0.30
JC8691	recBC sbcA recE	0	0.37
JC5547	recA recBC	0	0.37
JC5519	recBC	0	1.80
JC15329	recA recBC sbcBC	0	0.03
JC9387	recBC sbcBC	0	2.20
JC8111	recBC sbcBC recF	0	2.40
JC9366	recA	0	0.37
JC13031	recJ	0	0.45

SEQUENCE LISTING

<210>	SEQ 3	ID NO	1					
<211>	LENG	FH: 61	50					
<212>								
			Artifi	icial Sequer	nce			
<220>								
<221>	NAME,	/KEY:	misc_f	eature				
<222>	LOCA	FION:	(1)	(6150)				
<223>	OTHE	R INFO	RMATIC	N: plasmid	pBAD24-rec	ET		
<220>	FEATU	JRE :						
<221>	NAME	/KEY:	misc_f	eature				
<222>	LOCA	FION:	(96).	(974)				
<223>	OTHER	R INFO	ORMATIC	N: product	= "araC"			
<220>	FEATU	JRE :						
<221>	NAME	/KEY:	misc_f	eature				
<222>	LOCA	FION:	(1320)	(2162)				
<223>	OTHER	R INFO	RMATIC	N: product	= "t-recE"			
<220>	FEATU	JRE :						
<221>	NAME	/KEY:	misc_f	eature				
<222>	LOCA	FION:	(2155)	(2972)				
<223>	OTHER	R INFO	RMATIC	N: product	= "recT"			
<220>	FEAT	JRE :		-				
<221>	NAME	/KEY:	misc_f	eature				
<222>	LOCA	FION:	(3493)	(4353)				
<223>	OTHER	R INFO	RMATIC	N: product	= "bla"			
				-				
<400>	SEQUI	ENCE :	1					
atcgat	tgcat	aatgt	gcctg	tcaaatggac	gaagcaggga	ttctgcaaac	cctatgctac	
tccgto	caage	cgtca	attgt	ctgattcgtt	accaattatg	acaacttgac	ggctacatca	
_	-	-	-		-	-		
ttcact	tttt	cttca	acaacc	ggcacggaac	tcgctcgggc	tggccccggt	gcattttta	
aataco	caca	aqaaa	atagag	ttgatcgtca	aaaccaacat	tgcgaccgac	qqtqqcqata	
	5 5	5	5 5	5 5		5 5 5	55 55 5	
gggato	ecaaa	taata	rctcaa	aagcagette	gcctggctga	tacattaatc	ctcacaccaa	
		- 59-5	,		559009a		<u>-</u>	
attaar	racco	taato	ractas	atactaacaa	aaaagatgtg	adadaddaa	caacaacaaa	
cccaaų	Jacyc	caalt	liciad	0090099099	uuaayatyty	ucayacycya	cyycyacady	

caaacatgct gtgcgacgct	ggcgatatca	aaattgctgt	ctgccaggtg	atcgctgatg	420
tactgacaag cctcgcgtac	ccgattatcc	atcggtggat	ggagcgactc	gttaatcgct	480
tccatgcgcc gcagtaacaa	ttgctcaagc	agatttatcg	ccagcagctc	cgaatagcgc	540
cetteccett geeeggegtt	aatgatttgc	ccaaacaggt	cgctgaaatg	cggctggtgc	600
gcttcatccg ggcgaaagaa	ccccgtattg	gcaaatattg	acggccagtt	aagccattca	660
tgccagtagg cgcgcggacg	aaagtaaacc	cactggtgat	accattcgcg	agcctccgga	720
tgacgaccgt agtgatgaat	ctctcctggc	gggaacagca	aaatatcacc	cggtcggcaa	780
acaaattete gteeetgatt	tttcaccacc	ccctgaccgc	gaatggtgag	attgagaata	840
taacctttca ttcccagcgg	tcggtcgata	aaaaaatcga	gataaccgtt	ggcctcaatc	900
ggcgttaaac ccgccaccac	ı atgggcatta	aacgagtatc	ccggcagcag	gggatcattt	960
tgcgcttcag ccatactttt	catactcccg	ccattcagag	aagaaaccaa	ttgtccatat	1020
tgcatcagac attgccgtca	ctgcgtcttt	tactggctct	tctcgctaac	caaaccggta	1080
accccgctta ttaaaagcat	tctgtaacaa	agcgggacca	aagccatgac	aaaaacgcgt	1140
aacaaaagtg tctataatca	cggcagaaaa	gtccacattg	attatttgca	cggcgtcaca	1200
ctttgctatg ccatagcatt	tttatccata	agattagcgg	atcctacctg	acgcttttta	1260
tegeaactet etaetgttte	tccatacccg	ttttttggg	ctagcaggag	gaattcacca	1320
tggatcccgt aatcgtagaa	gacatagagc	caggtattta	ttacggaatt	tcgaatgaga	1380
attaccacgc gggtcccggt	atcagtaagt	ctcagctcga	tgacattgct	gatactccgg	1440
cactatattt gtggcgtaaa	aatgcccccg	tggacaccac	aaagacaaaa	acgctcgatt	1500
taggaactgc tttccactgo	cgggtacttg	aaccggaaga	attcagtaac	cgctttatcg	1560
tagcacctga atttaaccgc	cgtacaaacg	ccggaaaaga	agaagagaaa	gcgtttctga	1620
tggaatgcgc aagcacagga	aaaacggtta	tcactgcgga	agaaggccgg	aaaattgaac	1680
tcatgtatca aagcgttatg	getttgeege	tggggcaatg	gcttgttgaa	agcgccggac	1740
acgctgaatc atcaatttac	tgggaagatc	ctgaaacagg	aattttgtgt	cggtgccgtc	1800
cggacaaaat tatccctgaa	tttcactgga	tcatggacgt	gaaaactacg	gcggatattc	1860
aacgattcaa aaccgcttat	tacgactacc	gctatcacgt	tcaggatgca	ttctacagtg	1920
acggttatga agcacagttt	ggagtgcagc	caactttcgt	ttttctggtt	gccagcacaa	1980
ctattgaatg cggacgttat	ccggttgaaa	ttttcatgat	gggcgaagaa	gcaaaactgg	2040
caggtcaaca ggaatatcac	cgcaatctgc	gaaccctgtc	tgactgcctg	aataccgatg	2100
aatggccagc tattaagaca	ttatcactgc	cccgctgggc	taaggaatat	gcaaatgact	2160
aagcaaccac caatcgcaaa	agccgatctg	caaaaaactc	agggaaaccg	tgcaccagca	2220
gcagttaaaa atagcgacgt	gattagtttt	attaaccagc	catcaatgaa	agagcaactg	2280
gcagcagete ttecaegeea	tatgacggct	gaacgtatga	tccgtatcgc	caccacagaa	2340
attogtaaag ttooggogtt	aggaaactgt	gacactatga	gttttgtcag	tgcgatcgta	2400
cagtgttcac agctcggact	tgagccaggt	agcgccctcg	gtcatgcata	tttactgcct	2460
tttggtaata aaaacgaaaa	gagcggtaaa	aagaacgttc	agctaatcat	tggctatcgc	2520
ggcatgattg atctggctcg	ccgttctggt	caaatcgcca	gcctgtcagc	ccgtgttgtc	2580
cgtgaaggtg acgagtttag	cttcgaattt	ggccttgatg	aaaagttaat	acaccgcccg	2640
ggagaaaacg aagatgcccc	ggttacccac	gtctatgctg	tcgcaagact	gaaagacgga	2700

				COILCTI	lucu	
ggtactcagt	ttgaagttat	gacgcgcaaa	cagattgagc	tggtgcgcag	cctgagtaaa	2760
gctggtaata	acgggccgtg	ggtaactcac	tgggaagaaa	tggcaaagaa	aacggctatt	2820
cgtcgcctgt	tcaaatattt	gcccgtatca	attgagatcc	agcgtgcagt	atcaatggat	2880
gaaaaggaac	cactgacaat	cgatcctgca	gattcctctg	tattaaccgg	ggaatacagt	2940
gtaatcgata	attcagagga	atagatctaa	gcttggctgt	tttggcggat	gagagaagat	3000
tttcagcctg	atacagatta	aatcagaacg	cagaagcggt	ctgataaaac	agaatttgcc	3060
tggcggcagt	agcgcggtgg	tcccacctga	ccccatgccg	aactcagaag	tgaaacgccg	3120
tagcgccgat	ggtagtgtgg	ggtctcccca	tgcgagagta	gggaactgcc	aggcatcaaa	3180
taaaacgaaa	ggctcagtcg	aaagactggg	cctttcgttt	tatctgttgt	ttgtcggtga	3240
acgctctcct	gagtaggaca	aatccgccgg	gagcggattt	gaacgttgcg	aagcaacggc	3300
ccggagggtg	gcgggcagga	cgcccgccat	aaactgccag	gcatcaaatt	aagcagaagg	3360
ccatcctgac	ggatggcctt	tttgcgtttc	tacaaactct	tttgtttatt	tttctaaata	3420
cattcaaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	ataatattga	3480
aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	ttttgcggca	3540
ttttgccttc	ctgtttttgc	tcacccagaa	acgctggtga	aagtaaaaga	tgctgaagat	3600
cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	gatccttgag	3660
agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	ttaaagttct	gctatgtggc	3720
gcggtattat	cccgtgttga	cgccgggcaa	gagcaactcg	gtcgccgcat	acactattct	3780
cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	tggcatgaca	3840
gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	acactgcggc	caacttactt	3900
ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcacaacat	gggggatcat	3960
gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	cgacgagcgt	4020
gacaccacga	tgcctgtagc	aatggcaaca	acgttgcgca	aactattaac	tggcgaacta	4080
cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	agttgcagga	4140
ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaatc	tggagccggt	4200
gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	ctcccgtatc	4260
gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	acagatcgct	4320
gagataggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	ctcatatata	4380
ctttagattg	atttacgcgc	cctgtagcgg	cgcattaagc	gcggcgggtg	tggtggttac	4440
gcgcagcgtg	accgctacac	ttgccagcgc	cctagcgccc	gctcctttcg	ctttcttccc	4500
tteetttete	gccacgttcg	ccggctttcc	ccgtcaagct	ctaaatcggg	ggctcccttt	4560
agggttccga	tttagtgctt	tacggcacct	cgaccccaaa	aaacttgatt	tgggtgatgg	4620
ttcacgtagt	gggccatcgc	cctgatagac	ggtttttcgc	cctttgacgt	tggagtccac	4680
gttctttaat	agtggactct	tgttccaaac	ttgaacaaca	ctcaacccta	tctcgggcta	4740
ttcttttgat	ttataaggga	ttttgccgat	ttcggcctat	tggttaaaaa	atgagctgat	4800
ttaacaaaaa	tttaacgcga	attttaacaa	aatattaacg	tttacaattt	aaaaggatct	4860
aggtgaagat	cctttttgat	aatctcatga	ccaaaatccc	ttaacgtgag	ttttcgttcc	4920
actgagcgtc	agaccccgta	gaaaagatca	aaggatcttc	ttgagatcct	tttttctgc	4980
gcgtaatctg	ctgcttgcaa	acaaaaaaac	caccgctacc	agcggtggtt	tgtttgccgg	5040
atcaagagct	accaactctt	tttccgaagg	taactggctt	cagcagagcg	cagataccaa	5100

-						
atactgtco	ct tctagtgtag	ccgtagttag	gccaccactt	caagaactct	gtagcaccgc	5160
ctacataco	ct cgctctgcta	atcctgttac	cagtggctgc	tgccagtggc	gataagtcgt	5220
gtcttacco	gg gttggactca	agacgatagt	taccggataa	ggcgcagcgg	tcgggctgaa	5280
cggggggtt	tc gtgcacacag	cccagcttgg	agcgaacgac	ctacaccgaa	ctgagatacc	5340
tacagegte	ga gctatgagaa	agcgccacgc	ttcccgaagg	gagaaaggcg	gacaggtatc	5400
cggtaagco	gg cagggtcgga	acaggagagc	gcacgaggga	gcttccaggg	ggaaacgcct	5460
ggtatett	ta tagtcctgtc	gggtttcgcc	acctctgact	tgagcgtcga	tttttgtgat	5520
gctcgtcaç	gg ggggcggagc	ctatggaaaa	acgccagcaa	cgcggccttt	ttacggttcc	5580
tggcctttt	tg ctggcctttt	gctcacatgt	tctttcctgc	gttatcccct	gattetgtgg	5640
ataaccgta	at taccgccttt	gagtgagctg	ataccgctcg	ccgcagccga	acgaccgagc	5700
gcagcgagt	tc agtgagcgag	gaagcggaag	agcgcctgat	gcggtattt	ctccttacgc	5760
atctgtgco	gg tatttcacac	cgcatagggt	catggctgcg	ccccgacacc	cgccaacacc	5820
cgctgacgo	cg ccctgacggg	cttgtctgct	cccggcatcc	gcttacagac	aagctgtgac	5880
cgtctccg	gg agctgcatgt	gtcagaggtt	ttcaccgtca	tcaccgaaac	gcgcgaggca	5940
gcaaggaga	at ggcgcccaac	agtcccccgg	ccacgggggcc	tgccaccata	cccacgccga	6000
aacaagcgo	ct catgageeeg	aagtggcgag	cccgatcttc	cccatcggtg	atgtcggcga	6060
tataggcgo	cc agcaaccgca	cctgtggcgc	cggtgatgcc	ggccacgatg	cgtccggcgt	6120
agaggatct	tg ctcatgtttg	acagettate				6150
<pre><212> TYH <213> ORC <220> FEA <221> NAM <222> LOO <223> OTH <220> FEA <220> FEA <220> FEA <221> NAM <222> LOO </pre>	NGTH: 843 PE: DNA GANISM: Artif ATURE: ME/KEY: misc_ CATION: (1) HER INFORMATI	feature (843) ON: t-recE ((843)		pBAD24-recE	I at 1320-216	2
<400> SEQ	QUENCE: 2					
	ccc gta atc g Pro Val Ile V 5	al Glu Asp				48
-	aat gag aat t Asn Glu Asn T 20				s Ser Gln	96
	gac att gct g Asp Ile Ala A 35			0 00 0		144
	gtg gac acc a Val Asp Thr T					192
	tgc cgg gta c Cys Arg Val L					240
	cct gaa ttt a Pro Glu Phe A 85		-			288
	ttt ctg atg g Phe Leu Met G					336

100 105 110	
gcg gaa gaa ggc cgg aaa att gaa ctc atg tat caa agc gtt atg gct Ala Glu Glu Gly Arg Lys Ile Glu Leu Met Tyr Gln Ser Val Met Ala 115 120 125	384
ttg ccg ctg ggg caa tgg ctt gtt gaa agc gcc gga cac gct gaa tca Leu Pro Leu Gly Gln Trp Leu Val Glu Ser Ala Gly His Ala Glu Ser 130 135 140	432
tca att tac tgg gaa gat cct gaa aca gga att ttg tgt cgg tgc cgt Ser Ile Tyr Trp Glu Asp Pro Glu Thr Gly Ile Leu Cys Arg Cys Arg 145 150 155 160	480
ccg gac aaa att atc cct gaa ttt cac tgg atc atg gac gtg aaa act Pro Asp Lys Ile Ile Pro Glu Phe His Trp Ile Met Asp Val Lys Thr 165 170 175	528
acg gcg gat att caa cga ttc aaa acc gct tat tac gac tac cgc tat Thr Ala Asp Ile Gln Arg Phe Lys Thr Ala Tyr Tyr Asp Tyr Arg Tyr 180 185 190	576
cac gtt cag gat gca ttc tac agt gac ggt tat gaa gca cag ttt gga His Val Gln Asp Ala Phe Tyr Ser Asp Gly Tyr Glu Ala Gln Phe Gly 195 200 205	624
gtg cag cca act ttc gtt ttt ctg gtt gcc agc aca act att gaa tgc Val Gln Pro Thr Phe Val Phe Leu Val Ala Ser Thr Thr Ile Glu Cys 210 215 220	672
gga cgt tat ccg gtt gaa att ttc atg atg ggc gaa gaa gca aaa ctg Gly Arg Tyr Pro Val Glu Ile Phe Met Met Gly Glu Glu Ala Lys Leu 225 230 235 240	720
gca ggt caa cag gaa tat cac cgc aat ctg cga acc ctg tct gac tgc Ala Gly Gln Gln Glu Tyr His Arg Asn Leu Arg Thr Leu Ser Asp Cys 245 250 255	768
ctg aat acc gat gaa tgg cca gct att aag aca tta tca ctg ccc cgc Leu Asn Thr Asp Glu Trp Pro Ala Ile Lys Thr Leu Ser Leu Pro Arg 260 265 270	816
tgg gct aag gaa tat gca aat gac taa Trp Ala Lys Glu Tyr Ala Asn Asp * 275 280	843
<pre><210> SEQ ID NO 3 <211> LENGTH: 280 <211> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <221> NAME/KEY: misc_feature <222> LOCATION: (1)(280) <223> OTHER INFORMATION: t-recE on plasmid pBAD24-recET at 1320-2162 <400> SEQUENCE: 3</pre>	
Met Asp Pro Val Ile Val Glu Asp Ile Glu Pro Gly Ile Tyr Tyr Gly	
1 5 10 15	
Ile Ser Asn Glu Asn Tyr His Ala Gly Pro Gly Ile Ser Lys Ser Gln 20 25 30	
Leu Asp Asp Ile Ala Asp Thr Pro Ala Leu Tyr Leu Trp Arg Lys Asn 35 40 45	
Ala Pro Val Asp Thr Thr Lys Thr Lys Thr Leu Asp Leu Gly Thr Ala 50 55 60	
Phe His Cys Arg Val Leu Glu Pro Glu Glu Phe Ser Asn Arg Phe Ile 65 70 75 80	
Val Ala Pro Glu Phe Asn Arg Arg Thr Asn Ala Gly Lys Glu Glu Glu 85 90 95	
Lys Ala Phe Leu Met Glu Cys Ala Ser Thr Gly Lys Thr Val Ile Thr 100 105 110	

Ala Glu Gl 11														
	-	Arg	Lys	Ile	Glu 120	Leu	Met	Tyr	Gln	Ser 125	Val	Met	Ala	
Leu Pro Le 130	u Gly	Gln	Trp	Leu 135	Val	Glu	Ser	Ala	Gly 140	His	Ala	Glu	Ser	
Ser Ile Ty 145	r Trp		Asp 150	Pro	Glu	Thr	Gly	Ile 155	Leu	Сүз	Arg	Cys	Arg 160	
Pro Asp Ly	s Ile	Ile 165	Pro	Glu	Phe	His	Trp 170	Ile	Met	Asp	Val	Lys 175	Thr	
Thr Ala As	p Ile 180	Gln	Arg	Phe	Lys	Thr 185	Ala	Tyr	Tyr	Asp	Tyr 190	Arg	Tyr	
His Val Gl 19	-	Ala	Phe	Tyr	Ser 200	Asp	Gly	Tyr	Glu	Ala 205	Gln	Phe	Gly	
Val Gln Pr 210	o Thr	Phe	Val	Phe 215	Leu	Val	Ala	Ser	Thr 220	Thr	Ile	Glu	Суз	
Gly Arg Ty 225	r Pro		Glu 230	Ile	Phe	Met	Met	Gly 235	Glu	Glu	Ala	Lys	Leu 240	
Ala Gly Gl	n Gln	Glu 245	Tyr	His	Arg	Asn	Leu 250	Arg	Thr	Leu	Ser	Asp 255	Сүз	
Leu Asn Th	r Asp 260	Glu	Trp	Pro	Ala	Ile 265	Lys	Thr	Leu	Ser	Leu 270	Pro	Arg	
Trp Ala Ly 27		Tyr	Ala	Asn	Asp 280									
<213> ORGA <220> FEAT <221> NAME	URE :				-									
<pre><222> LOCA <223> OTHE <220> FEAT <221> NAME <222> LOCA <223> OTHE <400> SEQU</pre>	R INFO URE: /KEY: TION: R INFO	(1). DRMAT CDS (1). DRMAT	. (81 'ION: . (81	.0) rec .0)	T or	n pla	asmic	ł pBł	4D24 -	-recl	ST at	219	55-2972	
<pre><223> OTHE <220> FEAT <221> NAME <222> LOCA <223> OTHE <400> SEQU atg act aa</pre>	R INF(URE: /KEY: TION: R INF(ENCE: g caa	(1). DRMAT CDS (1). DRMAT 4 cca	. (81 'ION : . (81 'ION : 'ION : cca	.0) rec .0) rec atc	T or T gca	aaa	gcc	gat	ctg	саа	aaa	act	cag	48
<223> OTHE <220> FEAT <221> NAME <222> LOCA <223> OTHE <400> SEQU atg act aa Met Thr Ly	R INFO URE: /KEY: TION: R INFO ENCE: g caa s Gln 285	(1). DRMAT CDS (1). DRMAT 4 cca Pro	. (81 'ION : . (81 'ION : 'ION : cca Pro	.0) rec .0) rec atc Ile	T or T gca Ala	aaa Lys 290	gcc Ala	gat Asp	ctg Leu	caa Gln	ааа Lys 295	act Thr	cag Gln	
<pre><223> OTHE <220> FEAT <221> NAME <222> LOCA <223> OTHE <400> SEQU atg act aa</pre>	R INF(URE: /KEY: TION: R INF(ENCE: g caa s Gln 285 t gca g Ala	(1). DRMAT CDS (1). DRMAT 4 cca Pro cca	. (81 'ION: . (81 'ION: cca Pro gca	.0) rec .0) rec atc Ile gca	T or T gca Ala gtt	aaa Lys 290 aaa	gcc Ala aat	gat Asp agc	ctg Leu gac	caa Gln gtg	aaa Lys 295 att	act Thr agt	cag Gln ttt	48 96
<pre><223> OTHE <220> FEAT <221> NAME <222> LOCP <223> OTHE <400> SEQU atg act aa Met Thr Ly gga aac cg Gly Asn Ar</pre>	R INF(URE: /KEY: TION: R INF(ENCE: g caa s Gln 285 t gcaa g Ala 0 g cca	(1). DRMAT CDS (1). DRMAT 4 cca Pro cca Pro tca	. (81 ION: . (81 ION: CCA Pro gca Ala atg	.0) rec .0) rec Ile gca Ala aaa	T or T gca Ala gtt Val 305 gag	aaa Lys 290 aaa Lys caa	gcc Ala Asn ctg	gat Asp agc Ser gca	ctg Leu gac Asp gca	caa Gln gtg Val 310 gct	aaa Lys 295 att Ile ctt	act Thr agt Ser cca	cag Gln ttt Phe cgc	
<pre><223> OTHE <220> FEAT <221> NAME <222> LOCP <223> OTHE <400> SEQU atg act aa Met Thr Ly gga aac cg Gly Asn Ar 3C att aac ca Ile Asn Gl</pre>	R INFO URE: /KEY: TION: R INFO ENCE: g caa g caa	<pre>(1). DRMAT CDS (1). DRMAT 4 cca Pro cca Pro tca Ser gaa Glu</pre>	. (81 ION: . (81 ION: CCCA Pro gca Ala atg Met cgt	.0) rec .0) rec Ile gca Ala aaa Lys 320 atg	T or T gca Ala gtt Val 305 gag Glu atc	aaa Lys 290 aaa Lys Caa Gln cgt	gcc Ala aat Asn ctg Leu atc	gat Asp agc Ser gca Ala gcc	ctg Leu gac Asp gca Ala 325 acc	caa Gln Val 310 gct Ala aca	aaa Lys 295 att Ile ctt Leu gaa	act Thr agt Ser Cca Pro att	cag Gln ttt Phe cgc Arg cgt	96
<pre><223> OTHE <220> FEAT <221> NAME <222> LOCA <223> OTHE <222> COCH atg act aa Met Thr Ly gga aac cg Gly Asn Ar 3C att aac ca Ile Asn Gl 315 cat atg act His Met Th</pre>	R INFO URE: /KEY: TION: R INFO ENCE: g caa s Gln 285 t gcaa g Ala 0 g cca n Pro g gct r Ala g gcg	<pre>(1). RMAT CDS (1). RMAT 4 cca Pro cca Pro tca Ser gaa Glu tta</pre>	. (81 ION: . (81 ION: Pro gca Ala atg Met cgt Arg 335 gga	.0) rec atc Ile gca Ala aaa Lys 320 atg Met aac	T or T gca Ala gtt Val 305 gag glu atc Ile tgt	aaa Lys 290 aaaa Lys caa Gln cgt Arg gac	gcc Ala aat Asn ctg Leu atc Ile act	gat Asp agc Ser gca Ala 340 atg	ctg Leu gac Asp gca Ala 325 acc Thr agt	caa Gln gtg Val 310 gct Ala aca Thr ttt	aaa Lys 295 att Leu gaa Glu gtc	act Thr agt Ser Cca Pro att Ile agt	cag Gln ttt Phe cgc Arg cgt Arg 345 gcg	96 144
<pre><223> OTHE <220> FEAT <221> NAME <222> LOCA <223> OTHE <222> LOCA <223> OTHE <400> SEQU atg act aa Met Thr Ly gga aac cg Gly Asn Ar gga aac cg Gly Asn Ar 300 att aac ca lle Asn Gl 315 cat atg ac His Met Th 330 aaa gtt cc</pre>	R INFO URE: /KEY: TION: R INFO ENCE: g caa s Gln 285 t gcaa g Ala 0 g cca n Pro g gct r Ala g gcg o Ala g tgt	<pre>(1).(1). RMAT CDS (1). RMAT 4 cca Pro cca Pro tca Ser gaa Glu tta Leu 350 tca</pre>	. (81 ION: . (81 ION: ccca Pro gca Ala atg Met cgt Arg 335 gga Gly cag	0) rec atc Ile gca Ala aaa Lys 320 atg Met aac Asn ctc	CT or CT gca Ala gtt Val 305 gag Glu atc Ile tgt Cys gga	aaa Lys 290 aaaa Lys Caa Gln cgt Arg gac Asp ctt	gcc Ala aat Asn ctg Leu atc Ile act Thr 355 gag	gat Asp agc Ser gca Ala gcc Ala 340 atg Met cca	ctg Leu gac Asp gca Ala 325 acc Thr agt Ser ggt	caa Gln gtg Val 310 gct Ala aca Thr ttt Phe agc	aaa Lys 295 att Ile ctt Leu gaa Glu gtc Val gcc	act Thr agt Ser cca Pro att Ile agt Ser 360 ctc	cag Gln ttt Phe cgc Arg cgt Arg 345 gcg Ala ggt	96 144 192
<pre><223> OTHE <220> FEAT <221> NAME <222> LOCA <223> OTHE <222> OTHE <400> SEQU atg act aa Met Thr Ly gga aac cg Gly Asn Ar gGly Asn Ar gCly Asn Gl att aac ca lle Asn Gl 315 cat atg act His Met Th 330 aaa gtt cc Lys Val Pr atc gta ca</pre>	R INFO URE: /KEY: TION: R INFO ENCE: g caa s Gln 285 t gca g Ala 0 g cca n Pro g gct r Ala g gcg o Ala g tgt n Cys 365 t tta r Leu	<pre>(1). CDS (1). CDS (1). CDS (1). CCA Pro tca Ser gaa Glu tta Leu 350 tca Ser ctg</pre>	. (81 ION: . (81 ION: Pro gca Ala atg Met cgt Alg 335 gga Gly cag Gln cct	.0) rec atc Ile gca Ala aaa Lys 320 atg Met aaac Asn ctc Leu ttt	T or gca Ala gtt Val 305 gag Glu atc Cys gga Gly ggt	aaa Lys 290 aaa Lys Caa Gln Cgt Arg gac Asp ctt Leu 370 aat	gcc Ala aat Asn ctg Leu atc Ile act Thr 355 gag Glu aaa	gat Asp agc Ser gca Ala 340 atg Met cca Pro aac	ctg Leu gac Asp gca Ala 325 acc Thr agt Ser ggt Gly gaa	caa Gln Val 310 gct Ala aca Thr ttt Phe agc Ser aag	aaa Lys 295 att Leu gaa Glu gtc Val gcc Ala 375 agc	act Thr agt Ser cca Pro att Ile agt Ser 360 ctc Leu ggt	cag Gln ttt Phe cgc Arg 345 gcg Ala ggt Gly aaa	96 144 192 240

		400				405					
cgc cgt tct ggt Arg Arg Ser Gly 410				Ser .							432
ggt gac gag ttt Gly Asp Glu Phe	-	-			-	-	-				480
cgc ccg gga gaa Arg Pro Gly Glu 445	-		-	-			-		-	-	528
gca aga ctg aaa Ala Arg Leu Lys 460											576
cag att gag ctg Gln Ile Glu Leu 475	Val Arg										624
tgg gta act cac Trp Val Thr His 490				Lys							672
ctg ttc aaa tat Leu Phe Lys Tyr		•					<u> </u>	-	<u> </u>		720
atg gat gaa aag Met Asp Glu Lys 525											768
tta acc ggg gaa Leu Thr Gly Glu 540								tag *			810
<210> SEQ ID NO	5										
<211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <223> OTHER INF	59 Artifici misc_feat (1)(26 DRMATION:	ture 9)		smid	pBF	\D24 -	recI	ST at	: 215	5-2972	
<pre><211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: 220> FEATURE: <221> NAME/KEY: <222> LOCATION: <223> OTHER INF <400> SEQUENCE:</pre>	59 Artifici misc_feat (1)(26 DRMATION: 5	ture 9) recT or	n pla								
<pre><211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <223> OTHER INF <400> SEQUENCE: Met Thr Lys Gln 1</pre>	Artifici misc_feat (1)(26 DRMATION: 5 Pro Pro 5	ture 9) recT or Ile Ala	n pla Lys	Ala . 10	Asp	Leu	Gln	Гла	Thr 15	Gln	
<pre><211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <223> OTHER INF <400> SEQUENCE: Met Thr Lys Gln</pre>	Artifici misc_feat (1)(26 DRMATION: 5 Pro Pro 5	ture 9) recT or Ile Ala	n pla Lys	Ala . 10	Asp	Leu	Gln	Гла	Thr 15	Gln	
<pre><211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <222> LOCATION: <223> OTHER INF <400> SEQUENCE: Met Thr Lys Gln 1 Gly Asn Arg Ala</pre>	59 Artifici misc_feat (1)(26 DRMATION: 5 Pro Pro 5 Pro Ala .	ture 9) recT or Ile Ala Ala Val	Lys Lys 25	Ala . 10 Asn	Asp Ser	Leu Asp	Gln Val	Lys Ile 30	Thr 15 Ser	Gln Phe	
<pre><211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <223> OTHER INF <400> SEQUENCE: Met Thr Lys Gln 1 Gly Asn Arg Ala 20 Ile Asn Gln Pro</pre>	59 Artifici misc_feau (1)(26 DRMATION: 5 Pro Pro 5 Pro Pro 5 Pro Ala . Ser Met	ture 9) recT or Ile Ala Ala Val Lys Glu 40	Lys Lys 25 Gln	Ala . 10 Asn Leu .	Asp Ser Ala	Leu Asp Ala	Gln Val Ala 45	Lys Ile 30 Leu	Thr 15 Ser Pro	Gln Phe Arg	
<pre><211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <223> OTHER INF <400> SEQUENCE: Met Thr Lys Gln 1 Gly Asn Arg Ala 20 Ile Asn Gln Pro 35 His Met Thr Ala</pre>	59 Artifici misc_feat (1)(26 DRMATION: 5 Pro Pro 5 Pro Pro 5 Pro Ala Ser Met Glu Arg 1	ture 9) recT or Ile Ala Ala Val Lys Glu 40 Met Ile 55	Lys Lys 25 Gln Arg	Ala . 10 Asn Leu . Ile .	Asp Ser Ala Ala	Leu Asp Ala Thr 60	Gln Val Ala 45 Thr	Lys Ile 30 Leu Glu	Thr 15 Ser Pro Ile	Gln Phe Arg Arg	
<pre><211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <223> OTHER INF <400> SEQUENCE: Met Thr Lys Gln 1 Gly Asn Arg Ala 20 Ile Asn Gln Pro 35 His Met Thr Ala 50 Lys Val Pro Ala</pre>	Artifici misc_feat (1)(26 ORMATION: 5 Pro Pro 5 Pro Ala. Ser Met Glu Arg : Leu Gly . 70	ture 9) recT or Ile Ala Ala Val Lys Glu 40 Met Ile 55 Asn Cys	Lys Lys Gln Arg Asp	Ala . 10 Asn Leu . Ile . Thr :	Asp Ser Ala Ala Met 75	Leu Asp Ala Thr 60 Ser	Gln Val Ala 45 Thr Phe	Lys Ile 30 Leu Glu Val	Thr 15 Ser Pro Ile Ser	Gln Phe Arg Arg Ala 80	
<pre><211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <223> OTHER INF <400> SEQUENCE: Met Thr Lys Gln 1 Gly Asn Arg Ala 20 Ile Asn Gln Pro 35 His Met Thr Ala 50 Lys Val Pro Ala 65</pre>	Artifici misc_feat (1)(26 DRMATION: 5 Pro Pro 5 Pro Ala Ser Met Glu Arg 1 Leu Gly 70 Ser Gln 85	ture 9) recT or Ile Ala Ala Val Lys Glu 40 Met Ile 55 Asn Cys Leu Gly	Lys Lys 25 Gln Arg Asp Leu	Ala . 10 Asn Leu . Ile . Thr : Glu 90	Asp Ser Ala Ala Met 75 Pro	Leu Asp Ala Thr 60 Ser Gly	Gln Val Ala 45 Thr Phe Ser	Lys Jle 30 Leu Glu Val Ala	Thr 15 Ser Pro Ile Ser Leu 95	Gln Phe Arg Arg Ala 80 Gly	
<pre><211> LENGTH: 2 <212> TYPE: PRT <213> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <223> OTHER INF <400> SEQUENCE: Met Thr Lys Gln 1 Gly Asn Arg Ala 20 Ile Asn Gln Pro 35 His Met Thr Ala 50 Lys Val Pro Ala 65 Ile Val Gln Cys His Ala Tyr Leu</pre>	Artifici misc_feat (1)(26 DRMATION: 5 Pro Pro 5 Pro Ala Ser Met Glu Arg 1 Leu Gly 70 Ser Gln 85 Leu Pro	ture 9) recT or Ile Ala Ala Val Lys Glu 40 Met Ile 55 Leu Gly Phe Gly	n pla Lys 25 Gln Arg Leu Leu Asp	Ala . 10 Asn Leu . Thr : Glu 90 Lys .	Asp Ser Ala Ala Met 75 Pro Asn	Leu Asp Ala Thr 60 Ser Gly Glu	Gln Val Ala 45 Thr Phe Ser Lys	Lys Jle 30 Leu Glu Val Ala Ser 110	Thr 15 Ser Pro Ile Ser Leu 95 Gly	Gln Phe Arg Arg Ala 80 Gly Lys	
<pre><211> LENGTH: 2 <212> TYPE: PRT <212> ORGANISM: <220> FEATURE: <221> NAME/KEY: <222> LOCATION: <222> LOCATION: <223> OTHER INF <400> SEQUENCE: Met Thr Lys Gln 1 Gly Asn Arg Ala 20 Ile Asn Gln Pro 35 His Met Thr Ala 50 Lys Val Pro Ala 65 Ile Val Gln Cys His Ala Tyr Leu 100 Lys Asn Val Gln</pre>	Artifici misc_feat (1)(26 DRMATION: 5 Pro Pro 5 Pro Ala Ser Met Glu Arg : Leu Gly Ser Gln 85 Leu Pro Leu Ile Gln Ile	ture 9) recT or Ile Ala Ala Val Lys Glu 40 Met Ile 55 Leu Gly Phe Gly Ile Gly 120	lys Lys Gln Arg Leu Leu Asp Leu Tyr	Ala . 10 Asn Leu . Ile . Thr : Glu 90 Lys . Arg .	Asp Ser Ala Ala Met 75 Pro Asn Gly	Leu Asp Ala Thr 60 Ser Gly Glu Met	Gln Val Ala 45 Thr Phe Ser Lys Ile 125	Lys Ile 30 Leu Glu Val Ala Ser 110 Asp	Thr 15 Ser Pro Ile Ser Leu 95 Gly Leu	Gln Phe Arg Arg Ala 80 Gly Lys Ala	

-continued

Arg Pro Gly Glu Asn Glu Asp Ala Pro Val Thr His Val Tyr Ala Val 170 175 165 Ala Arg Leu Lys Asp Gly Gly Thr Gln Phe Glu Val Met Thr Arg Lys 185 180 190 Gln Ile Glu Leu Val Arg Ser Leu Ser Lys Ala Gly Asn Asn Gly Pro 195 200 205 Trp Val Thr His Trp Glu Glu Met Ala Lys Lys Thr Ala Ile Arg Arg 210 215 220 Leu Phe Lys Tyr Leu Pro Val Ser Ile Glu Ile Gln Arg Ala Val Ser 225 230 235 240 Met Asp Glu Lys Glu Pro Leu Thr Ile Asp Pro Ala Asp Ser Ser Val 245 250 255 Leu Thr Gly Glu Tyr Ser Val Ile Asp Asn Ser Glu Glu 260 265 <210> SEQ ID NO 6 <211> LENGTH: 876 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <221> NAME/KEY: misc_feature <222> LOCATION: (1)..(876) <223> OTHER INFORMATION: araC on plasmid pBAD24-recET at 974-996 <220> FEATURE: <221> NAME/KEY: CDS <222> LOCATION: (1)..(876) <223> OTHER INFORMATION: araC <400> SEQUENCE: 6 tgacaacttg acggctacat cattcacttt ttcttcacaa ccggcacgga actcgctcgg gctggccccg gtgcattttt taaatacccg cgagaaatag agttgatcgt caaaaccaac attqcqaccq acqqtqqcqa taqqcatccq qqtqqtqctc aaaaqcaqct tcqcctqqct gatacgttqg tcctcqcqcc aqcttaaqac qctaatccct aactqctqqc qqaaaaqatq tgacagacgc gacggcgaca agcaaacatg ctgtgcgacg ctggcgatat caaaattgct qtctqccaqq tqatcqctqa tqtactqaca aqcctcqcqt acccqattat ccatcqqtqq atggagegae tegttaateg ettecatgeg eegcagtaae aattgeteaa geagatttat cgccagcagc tccgaatagc gcccttcccc ttgcccggcg ttaatgattt gcccaaacag gtcgctgaaa tgcggctggt gcgcttcatc cgggcgaaag aaccccgtat tggcaaatat tgacggccag ttaagccatt catgccagta ggcgcgcgga cgaaagtaaa cccactggtg ataccattcg cgagcctccg gatgacgacc gtagtgatga atctctcctg gcgggaacag caaaatatca cccggtcggc aaacaaattc tcgtccctga tttttcacca ccccctgacc gcgaatggtg agattgagaa tataaccttt cattcccagc ggtcggtcga taaaaaaatc gagataaccg ttggcctcaa tcggcgttaa acccgccacc agatgggcat taaacgagta teceggeage aggggateat tttgegette ageeat <210> SEQ ID NO 7 <211> LENGTH: 292

LENGTH: 292
<211> LENGTH: 292
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(292)
<223> OTHER INFORMATION: araC on plasmid pBAD24-recET at 974-996

60

120

180

240

300

360

420

480

540

600

660

720

780

840

-continued

Met Ala Glu Ala Gln Asn Asp Pro Leu Leu Pro Gly Tyr Ser Phe Asn Ala His Leu Val Ala Gly Leu Thr Pro Ile Glu Ala Asn Gly Tyr Leu Asp Phe Phe Ile Asp Arg Pro Leu Gly Met Lys Gly Tyr Ile Leu Asn Leu Thr Ile Arg Gly Gln Gly Val Val Lys Asn Gln Gly Arg Glu Phe Val Cys Arg Pro Gly Asp Ile Leu Leu Phe Pro Pro Gly Glu Ile His His Tyr Gly Arg His Pro Glu Ala Arg Glu Trp Tyr His Gln Trp Val Tyr Phe Arg Pro Arg Ala Tyr Trp His Glu Trp Leu Asn Trp Pro Ser Ile Phe Ala Asn Thr Gly Phe Phe Arg Pro Asp Glu Ala His Gln Pro His Phe Ser Asp Leu Phe Gly Gln Ile Ile Asn Ala Gly Gln Gly Glu Gly Arg Tyr Ser Glu Leu Leu Ala Ile Asn Leu Leu Glu Gln Leu Leu Leu Arg Arg Met Glu Ala Ile Asn Glu Ser Leu His Pro Pro Met Asp Asn Arg Val Arg Glu Ala Cys Gln Tyr Ile Ser Asp His Leu Ala Asp Ser Asn Phe Asp Ile Ala Ser Val Ala Gln His Val Cys Leu Ser Pro 200 205 Ser Arg Leu Ser His Leu Phe Arg Gln Gln Leu Gly Ile Ser Val Leu Ser Trp Arg Glu Asp Gln Arg Ile Ser Gln Ala Lys Leu Leu Ser Thr Thr Arg Met Pro Ile Ala Thr Val Gly Arg Asn Val Gly Phe Asp Asp Gln Leu Tyr Phe Ser Arg Val Phe Lys Lys Cys Thr Gly Ala Ser Pro Ser Glu Phe Arg Ala Gly Cys Glu Glu Lys Val Asn Asp Val Ala Val Lys Leu Ser <210> SEQ ID NO 8 <211> LENGTH: 861 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <221> NAME/KEY: misc_feature <222> LOCATION: (1)..(861) <223> OTHER INFORMATION: bla gene on plasmid pBAD24-recET at 3493-4353 <220> FEATURE: <221> NAME/KEY: CDS <222> LOCATION: (1)..(861) <223> OTHER INFORMATION: bla <400> SEQUENCE: 8 atg agt att caa cat ttc cgt gtc gcc ctt att ccc ttt ttt gcg gca Met Ser Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala ttt tqc ctt cct qtt ttt qct cac cca qaa acq ctq qtq aaa qta aaa

-continued

Phe C	Суз 310	Leu	Pro	Val	Phe	Ala 315	His	Pro	Glu	Thr	Leu 320	Val	ГЛа	Val	Lys	
gat g Asp # 325																144
ctc a Leu A																192
cca a Pro N																240
cgt g Arg \																288
cag a Gln A 3																336
gat g Asp (405																384
gat a Asp A																432
gag d Glu I																480
gat d Asp A																528
gac a Asp 1 4																576
act o Thr (485																624
atg g Met (672
gct q Ala (720
cgc g Arg (768
gta g Val N																816
aga (Arg (565																861
<2103 <2113 <2123 <2203 <2203 <2213 <2223 <2223 <2223 <2235	> LE > TY > OF > FE > NA > LC > OT	ENGTH PE: CGANI CATUF ME/F CATI CATI	H: 28 PRT [SM: RE: RE: (EY: [ON: INF(Art: DOMA (1) DRMA	AIN (2:	36)			n pla	asmio	d pBł	4D24	-recl	ET at	: 349	3-4353

<400> SEQUENCE: 9

-continued

												0011	CIII	ucu	
Met 1	Ser	Ile	Gln	His 5	Phe	Arg	Val	Ala	Leu 10	Ile	Pro	Phe	Phe	Ala 15	Ala
Phe	Cys	Leu	Pro 20	Val	Phe	Ala	His	Pro 25	Glu	Thr	Leu	Val	Lys 30	Val	Lys
Asp	Ala	Glu 35	Asp	Gln	Leu	Gly	Ala 40	Arg	Val	Gly	Tyr	Ile 45	Glu	Leu	Asp
Leu	Asn 50	Ser	Gly	Гла	Ile	Leu 55	Glu	Ser	Phe	Arg	Pro 60	Glu	Glu	Arg	Phe
Pro 65	Met	Met	Ser	Thr	Phe 70	Lys	Val	Leu	Leu	Cys 75	Gly	Ala	Val	Leu	Ser 80
Arg	Val	Asp	Ala	Gly 85	Gln	Glu	Gln	Leu	Gly 90	Arg	Arg	Ile	His	Tyr 95	Ser
Gln	Asn	Asp	Leu 100	Val	Glu	Tyr	Ser	Pro 105	Val	Thr	Glu	Гла	His 110	Leu	Thr
Asp	Gly	Met 115	Thr	Val	Arg	Glu	Leu 120	Cys	Ser	Ala	Ala	Ile 125	Thr	Met	Ser
Asp	Asn 130	Thr	Ala	Ala	Asn	Leu 135	Leu	Leu	Thr	Thr	Ile 140	Gly	Gly	Pro	Lys
Glu 145	Leu	Thr	Ala	Phe	Leu 150	His	Asn	Met	Gly	Asp 155	His	Val	Thr	Arg	Leu 160
Asp	Arg	Trp	Glu	Pro 165	Glu	Leu	Asn	Glu	Ala 170	Ile	Pro	Asn	Asp	Glu 175	Arg
Asp	Thr	Thr	Met 180	Pro	Val	Ala	Met	Ala 185	Thr	Thr	Leu	Arg	Lys 190	Leu	Leu
Thr	Gly	Glu 195	Leu	Leu	Thr	Leu	Ala 200	Ser	Arg	Gln	Gln	Leu 205	Ile	Asp	Trp
Met	Glu 210	Ala	Asp	ГЛЗ	Val	Ala 215	Gly	Pro	Leu	Leu	Arg 220	Ser	Ala	Leu	Pro
Ala 225	Gly	Trp	Phe	Ile	Ala 230	Asp	Lys	Ser	Gly	Ala 235	Gly	Glu	Arg	Gly	Ser 240
Arg	Gly	Ile	Ile	Ala 245	Ala	Leu	Gly	Pro	Asp 250	Gly	Lys	Pro	Ser	Arg 255	Ile
Val	Val	Ile	Tyr 260	Thr	Thr	Gly	Ser	Gln 265	Ala	Thr	Met	Asp	Glu 270	Arg	Asn
Arg	Gln	Ile 275	Ala	Glu	Ile	Gly	Ala 280	Ser	Leu	Ile	ГÀа	His 285	Trp		
<21: <21: <22: <22: <22: <22: <22: <22:	0> FH L> NA 2> LC 3> O 3> O 5> FH L> NA 2> LC	ENGTH YPE: RGANI EATUH AME/I DCATI THER EATUH AME/I DCATI	H: 7: DNA ISM: RE: KEY: ION: INF RE: KEY: ION:	195 Art: miso (1) DRMA	c_fea (7: TION (7:	ature 195) : pla 195)	e asmio	l pB2	AD-ET	[-gai	nma				
<400)> SH	EQUEI	ICE :	10											
															igctac acatca
															tttta
															gcgata
															cgccag
								-		-				-	-

cttaagacgc taat	ceetaa etgetggege	aaaagatgtg	acagacgcga	cggcgacaag	360
caaacatgct gtgc	cgacget ggegatatea	aaattgctgt	ctgccaggtg	atcgctgatg	420
tactgacaag ccto	egegtae eegattatee	atcggtggat	ggagcgactc	gttaatcgct	480
tccatgcgcc gcag	gtaacaa ttgctcaago	agatttatcg	ccagcagctc	cgaatagcgc	540
cetteccett geee	cggcgtt aatgatttgo	ccaaacaggt	cgctgaaatg	cggctggtgc	600
getteateeg ggeg	gaaagaa ccccgtatto	gcaaatattg	acggccagtt	aagccattca	660
tgccagtagg cgcg	geggaeg aaagtaaaee	cactggtgat	accattcgcg	agcctccgga	720
tgacgaccgt agtg	gatgaat ctctcctggo	gggaacagca	aaatatcacc	cggtcggcaa	780
acaaattete gtee	cctgatt tttcaccacc	ccctgaccgc	gaatggtgag	attgagaata	840
taacctttca ttcc	ccagcgg tcggtcgata	aaaaaatcga	gataaccgtt	ggcctcaatc	900
ggcgttaaac ccgc	ccaccag atgggcatta	aacgagtatc	ccggcagcag	gggatcattt	960
tgcgcttcag ccat	actttt catactccc	ccattcagag	aagaaaccaa	ttgtccatat	1020
tgcatcagac attg	geegtea etgegtettt	tactggctct	tctcgctaac	caaaccggta	1080
accccgctta ttaa	aaagcat tctgtaacaa	agcgggacca	aagccatgac	aaaaacgcgt	1140
aacaaaagtg tcta	ataatca cggcagaaaa	gtccacattg	attatttgca	cggcgtcaca	1200
ctttgctatg ccat	agcatt tttatccata	agattagcgg	atcctacctg	acgcttttta	1260
tegeaactet etae	stgttte tecatacee	l ttttttggg	ctagcaggag	gaattcacca	1320
tggatcccgt aato	ogtagaa gacatagago	caggtattta	ttacggaatt	tcgaatgaga	1380
attaccacgc gggt	cccggt atcagtaagt	ctcagctcga	tgacattgct	gatactccgg	1440
cactatattt gtgg	gegtaaa aatgeeeee	tggacaccac	aaagacaaaa	acgctcgatt	1500
taggaactgc tttc	ceactge egggtaette	aaccggaaga	attcagtaac	cgctttatcg	1560
tagcacctga attt	caacege egtacaaace	ccggaaaaga	agaagagaaa	gcgtttctga	1620
tggaatgcgc aago	cacagga aaaacggtta	tcactgcgga	agaaggccgg	aaaattgaac	1680
tcatgtatca aago	egttatg getttgeege	tggggcaatg	gcttgttgaa	agcgccggac	1740
acgctgaatc atca	aatttac tgggaagato	ctgaaacagg	aattttgtgt	cggtgccgtc	1800
cggacaaaat tato	eeetgaa ttteaetgga	tcatggacgt	gaaaactacg	gcggatattc	1860
aacgattcaa aacc	egettat taegaetaee	gctatcacgt	tcaggatgca	ttctacagtg	1920
acggttatga agca	acagttt ggagtgcago	caactttcgt	ttttctggtt	gccagcacaa	1980
ctattgaatg cgga	acgttat ccggttgaaa	ttttcatgat	gggcgaagaa	gcaaaactgg	2040
caggtcaaca ggaa	atatcac cgcaatctgo	gaaccctgtc	tgactgcctg	aataccgatg	2100
aatggccagc tatt	laagaca ttatcactgo	cccgctgggc	taaggaatat	gcaaatgact	2160
agatetegag gtae	cccgagc acgtgttgac	aattaatcat	cggcatagta	tatcggcata	2220
gtataatacg acaa	aggtgag gaactaaaco	atggctaagc	aaccaccaat	cgcaaaagcc	2280
gatctgcaaa aaac	ctcaggg aaaccgtgca	ccagcagcag	ttaaaaatag	cgacgtgatt	2340
agttttatta acca	agccatc aatgaaagag	caactggcag	cagctcttcc	acgccatatg	2400
acggctgaac gtat	gateeg tategeeace	acagaaattc	gtaaagttcc	ggcgttagga	2460
aactgtgaca ctat	gagttt tgtcagtgco	atcgtacagt	gttcacagct	cggacttgag	2520
ccaggtagcg ccct	coggtca tgcatattta	ctgccttttg	gtaataaaaa	cgaaaagagc	2580
ggtaaaaaga acgt	tcaget aatcattgge	tatcgcggca	tgattgatct	ggctcgccgt	2640

				-0011011	iueu	
tctggtcaaa	tcgccagcct	gtcagcccgt	gttgtccgtg	aaggtgacga	gtttagcttc	2700
gaatttggcc	ttgatgaaaa	gttaatacac	cgcccgggag	aaaacgaaga	tgccccggtt	2760
acccacgtct	atgctgtcgc	aagactgaaa	gacggaggta	ctcagtttga	agttatgacg	2820
cgcaaacaga	ttgagctggt	gcgcagcctg	agtaaagctg	gtaataacgg	gccgtgggta	2880
actcactggg	aagaaatggc	aaagaaaacg	gctattcgtc	gcctgttcaa	atatttgccc	2940
gtatcaattg	agatccagcg	tgcagtatca	atggatgaaa	aggaaccact	gacaatcgat	3000
cctgcagatt	cctctgtatt	aaccgggggaa	tacagtgtaa	tcgataattc	agaggaatag	3060
atctaagctt	cctgctgaac	atcaaaggca	agaaaacatc	tgttgtcaaa	gacagcatcc	3120
ttgaacaagg	acaattaaca	gttaacaaat	aaaaacgcaa	aagaaaatgc	cgatatccta	3180
ttggcatttt	cttttatttc	ttatcaacat	aaaggtgaat	cccatacctc	gagetteacg	3240
ctgccgcaag	cactcagggc	gcaagggctg	ctaaaaggaa	gcggaacacg	tagaaagcca	3300
gtccgcagaa	acggtgctga	ccccggatga	atgtcagcta	ctgggctatc	tggacaaggg	3360
aaaacgcaag	cgcaaagaga	aagcaggtag	cttgcagtgg	gcttacatgg	cgatagctag	3420
actgggcggt	tttatggaca	gcaagcgaac	cggaattgcc	agctgggggg	ccctctggta	3480
aggttgggaa	gccctgcaaa	gtaaactgga	tggctttctt	gccgccaagg	atctgatggc	3540
gcaggggatc	aagatctgat	caagagacag	gatgaggatc	gtttcgcatg	gatattaata	3600
ctgaaactga	gatcaagcaa	aagcattcac	taaccccctt	tcctgttttc	ctaatcagcc	3660
cggcatttcg	cgggcgatat	tttcacagct	atttcaggag	ttcagccatg	aacgcttatt	3720
acattcagga	tcgtcttgag	gctcagagct	gggcgcgtca	ctaccagcag	ctcgcccgtg	3780
aagagaaaga	ggcagaactg	gcagacgaca	tggaaaaagg	cctgccccag	cacctgtttg	3840
aatcgctatg	catcgatcat	ttgcaacgcc	acgggggccag	caaaaaatcc	attacccgtg	3900
cgtttgatga	cgatgttgag	tttcaggagc	gcatggcaga	acacatccgg	tacatggttg	3960
aaaccattgc	tcaccaccag	gttgatattg	attcagaggt	ataaaacgag	tagaagcttg	4020
gctgttttgg	cggatgagag	aagattttca	gcctgataca	gattaaatca	gaacgcagaa	4080
gcggtctgat	aaaacagaat	ttgcctggcg	gcagtagcgc	ggtggtccca	cctgacccca	4140
tgccgaactc	agaagtgaaa	cgccgtagcg	ccgatggtag	tgtggggtct	ccccatgcga	4200
gagtagggaa	ctgccaggca	tcaaataaaa	cgaaaggctc	agtcgaaaga	ctgggccttt	4260
cgttttatct	gttgtttgtc	ggtgaacgct	ctcctgagta	ggacaaatcc	gccgggagcg	4320
gatttgaacg	ttgcgaagca	acggcccgga	gggtggcggg	caggacgccc	gccataaact	4380
gccaggcatc	aaattaagca	gaaggccatc	ctgacggatg	gcctttttgc	gtttctacaa	4440
actcttttgt	ttatttttct	aaatacattc	aaatatgtat	ccgctcatga	gacaataacc	4500
ctgataaatg	cttcaataat	attgaaaaag	gaagagtatg	agtattcaac	atttccgtgt	4560
cgcccttatt	cccttttttg	cggcattttg	ccttcctgtt	tttgctcacc	cagaaacgct	4620
ggtgaaagta	aaagatgctg	aagatcagtt	gggtgcacga	gtgggttaca	tcgaactgga	4680
tctcaacagc	ggtaagatcc	ttgagagttt	tcgccccgaa	gaacgttttc	caatgatgag	4740
cacttttaaa	gttctgctat	gtggcgcggt	attatcccgt	gttgacgccg	ggcaagagca	4800
actcggtcgc	cgcatacact	attctcagaa	tgacttggtt	gagtactcac	cagtcacaga	4860
aaagcatctt	acggatggca	tgacagtaag	agaattatgc	agtgctgcca	taaccatgag	4920
tgataacact	gcggccaact	tacttctgac	aacgatcgga	ggaccgaagg	agctaaccgc	4980
tttttgcac	aacatggggg	atcatgtaac	tcgccttgat	cgttgggaac	cggagctgaa	5040

-continued

tgaagccata	ccaaacgacg	agcgtgacac	cacgatgcct	gtagcaatgg	caacaacgtt	5100
gcgcaaacta	ttaactggcg	aactacttac	tctagcttcc	cggcaacaat	taatagactg	5160
gatggaggcg	gataaagttg	caggaccact	tctgcgctcg	gcccttccgg	ctggctggtt	5220
tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	ggtatcattg	cagcactggg	5280
gccagatggt	aagccctccc	gtatcgtagt	tatctacacg	acgggggagtc	aggcaactat	5340
ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	ctgattaagc	attggtaact	5400
gtcagaccaa	gtttactcat	atatacttta	gattgattta	cgcgccctgt	agcggcgcat	5460
taagcgcggc	gggtgtggtg	gttacgcgca	gcgtgaccgc	tacacttgcc	agcgccctag	5520
cgcccgctcc	tttcgctttc	ttcccttcct	ttctcgccac	gttcgccggc	tttccccgtc	5580
aagctctaaa	tcggggggtc	cctttagggt	tccgatttag	tgctttacgg	cacctcgacc	5640
ccaaaaaact	tgatttgggt	gatggttcac	gtagtgggcc	atcgccctga	tagacggttt	5700
ttcgcccttt	gacgttggag	tccacgttct	ttaatagtgg	actcttgttc	caaacttgaa	5760
caacactcaa	ccctatctcg	ggctattctt	ttgatttata	agggattttg	ccgatttcgg	5820
cctattggtt	aaaaatgag	ctgatttaac	aaaaatttaa	cgcgaatttt	aacaaaatat	5880
taacgtttac	aatttaaaag	gatctaggtg	aagatccttt	ttgataatct	catgaccaaa	5940
atcccttaac	gtgagttttc	gttccactga	gcgtcagacc	ccgtagaaaa	gatcaaagga	6000
tcttcttgag	atcctttttt	tctgcgcgta	atctgctgct	tgcaaacaaa	aaaaccaccg	6060
ctaccagcgg	tggtttgttt	gccggatcaa	gagctaccaa	ctctttttcc	gaaggtaact	6120
ggcttcagca	gagcgcagat	accaaatact	gtccttctag	tgtagccgta	gttaggccac	6180
cacttcaaga	actctgtagc	accgcctaca	tacctcgctc	tgctaatcct	gttaccagtg	6240
gctgctgcca	gtggcgataa	gtcgtgtctt	accgggttgg	actcaagacg	atagttaccg	6300
gataaggcgc	agcggtcggg	ctgaacgggg	ggttcgtgca	cacagcccag	cttggagcga	6360
acgacctaca	ccgaactgag	atacctacag	cgtgagctat	gagaaagcgc	cacgcttccc	6420
gaagggagaa	aggcggacag	gtatccggta	agcggcaggg	tcggaacagg	agagcgcacg	6480
agggagcttc	caggggggaaa	cgcctggtat	ctttatagtc	ctgtcgggtt	tcgccacctc	6540
tgacttgagc	gtcgattttt	gtgatgctcg	tcagggggggc	ggagcctatg	gaaaaacgcc	6600
agcaacgcgg	cctttttacg	gttcctggcc	ttttgctggc	cttttgctca	catgttcttt	6660
cctgcgttat	cccctgattc	tgtggataac	cgtattaccg	cctttgagtg	agctgatacc	6720
gctcgccgca	gccgaacgac	cgagcgcagc	gagtcagtga	gcgaggaagc	ggaagagcgc	6780
ctgatgcggt	attttctcct	tacgcatctg	tgcggtattt	cacaccgcat	agggtcatgg	6840
ctgcgccccg	acacccgcca	acacccgctg	acgcgccctg	acgggcttgt	ctgctcccgg	6900
catccgctta	cagacaagct	gtgaccgtct	ccgggagctg	catgtgtcag	aggttttcac	6960
cgtcatcacc	gaaacgcgcg	aggcagcaag	gagatggcgc	ccaacagtcc	cccggccacg	7020
gggcctgcca	ccatacccac	gccgaaacaa	gcgctcatga	gcccgaagtg	gcgagcccga	7080
tcttccccat	cggtgatgtc	ggcgatatag	gcgccagcaa	ccgcacctgt	ggcgccggtg	7140
atgccggcca	cgatgcgtcc	ggcgtagagg	atctgctcat	gtttgacagc	ttatc	7195

<210> SEQ ID NO 11 <211> LENGTH: 7010 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE:

<pre><221> NAME/KEY: misc_feature <222> LOCATION: (1)(7010) <223> OTHER INFORMATION: plasmid pBAD-alpha-beta-gamma <220> FEATURE: <221> NAME/KEY: CDS <222> LOCATION: (1320)(2000) <223> OTHER INFORMATION: red alpha <220> FEATURE: <221> NAME/KEY: CDS <222> LOCATION: (2086)(2871) <223> OTHER INFORMATION: red beta <220> FEATURE: <221> NAME/KEY: CDS <222> LOCATION: (3403)(3819) <223> OTHER INFORMATION: red gamma</pre>									
<400> SEQUENCE: 11									
atcgatgcat aatgtgcctg tcaaatggac gaagcaggga ttctgcaaac cctatgctac	60								
tccgtcaage cgtcaattgt ctgattcgtt accaattatg acaacttgac ggctacatca	120								
ttcacttttt cttcacaacc ggcacggaac tcgctcgggc tggccccggt gcattttta	180								
aatacccgcg agaaatagag ttgatcgtca aaaccaacat tgcgaccgac ggtggcgata	240								
ggcatccggg tggtgctcaa aagcagcttc gcctggctga tacgttggtc ctcgcgccag	300								
cttaagacgc taatccctaa ctgctggcgg aaaagatgtg acagacgcga cggcgacaag	360								
caaacatgct gtgcgacgct ggcgatatca aaattgctgt ctgccaggtg atcgctgatg	420								
tactgacaag cctcgcgtac ccgattatcc atcggtggat ggagcgactc gttaatcgct	480								
tccatgcgcc gcagtaacaa ttgctcaagc agatttatcg ccagcagctc cgaatagcgc	540								
ccttcccctt gcccggcgtt aatgatttgc ccaaacaggt cgctgaaatg cggctggtgc	600								
gcttcatccg ggcgaaagaa ccccgtattg gcaaatattg acggccagtt aagccattca	660								
tgccagtagg cgcgcggacg aaagtaaacc cactggtgat accattcgcg agcctccgga	720								
tgacgaccgt agtgatgaat ctctcctggc gggaacagca aaatatcacc cggtcggcaa	780								
acaaattoto gtoootgatt tttoaccaco oootgacogo gaatggtgag attgagaata	840								
taacctttca ttcccagcgg tcggtcgata aaaaaatcga gataaccgtt ggcctcaatc	900								
ggcgttaaac ccgccaccag atgggcatta aacgagtatc ccggcagcag gggatcattt	960								
tgcgcttcag ccatactttt catactcccg ccattcagag aagaaaccaa ttgtccatat	1020								
tgcatcagac attgccgtca ctgcgtcttt tactggctct tctcgctaac caaaccggta	1080								
accccgctta ttaaaagcat tctgtaacaa agcgggacca aagccatgac aaaaacgcgt	1140								
aacaaaagtg tctataatca cggcagaaaa gtccacattg attatttgca cggcgtcaca	1200								
ctttgctatg ccatagcatt tttatccata agattagcgg atcctacctg acgcttttta	1260								
tegcaactet etactgttte tecataceeg tttttttggg etagcaggag gaatteace	1319								
atg aca ccg gac att atc ctg cag cgt acc ggg atc gat gtg aga gct	1367								
Met Thr Pro Asp Ile Ile Leu Gln Arg Thr Gly Ile Asp Val Arg Ala 290 295 300									
gtc gaa cag ggg gat gat gcg tgg cac aaa tta cgg ctc ggc gtc atc	1415								
Val Glu Gln Gly Asp Asp Ala Trp His Lys Leu Arg Leu Gly Val Ile 305 310 315									
acc gct tca gaa gtt cac aac gtg ata gca aaa ccc cgc tcc gga aag	1463								
Thr Ala Ser Glu Val His Asn Val Ile Ala Lys Pro Arg Ser Gly Lys 320 325 330 335									
aag tgg cct gac atg aaa atg tcc tac ttc cac acc ctg ctt gct gag	1511								
Lys Trp Pro Asp Met Lys Met Ser Tyr Phe His Thr Leu Leu Ala Glu 340 345 350									
gtt tgc acc ggt gtg gct ccg gaa gtt aac gct aaa gca ctg gcc tgg	1559								

				-
_	cont	- 1	nu	ed

Val																
	Суз	Thr	Gly 355	Val	Ala	Pro	Glu	Val 360	Asn	Ala	Lys	Ala	Leu 365	Ala	Trp	
	aaa Lys															1607
	gtg Val 385															1655
	acc Thr															1703
	ctg Leu															1751
	ggt Gly															1799
	atg Met															1847
	cgt Arg 465															1895
	aag Lys															1943
	atg Met															1991
	cga Arg	tag *	atco	cggt	acc (cgago	cacgt	g ti	gaca	aatta	a ato	catco	ggca			2040
	5															
tagi	tatat	teg g	gcata	agtai	ta ai	tacga	acaaç	g gtç	gagga	aact	aaa		tg ag et Se 1			2094
gca	-	gca	acg	ctg	gct	<u>a</u> aa	aag	ctg	gct	gaa	cgt	Me gtc	et Se 1 ggc	er Th atg	nr gat	2094 2142
gca Ala tct	ctc Leu	gca Ala gac	acg Thr cca	ctg Leu cag	gct Ala gaa	ggg Gly 10 ctg	aag Lys atc	ctg Leu acc	gct Ala act	gaa Glu ctt	cgt Arg 15 cgc	Me gtc Val cag	et Se 1 ggc Gly acg	er Th atg Met gca	gat Asp ttt	
gca Ala tct Ser 20 aaa	ctc Leu 5 gtc	gca Ala gac Asp gat	acg Thr cca Pro gcc	ctg Leu cag Gln agc	gct Ala gaa Glu 25 gat	999 Gly 10 ctg Leu gcg	aag Lys atc Ile cag	ctg Leu acc Thr ttc	gct Ala act Thr atc	gaa Glu ctt Leu 30 gca	cgt Arg 15 cgc Arg tta	Me gtc Val cag Gln ctg	et Se 1 ggc Gly acg Thr atc	er Th atg Met gca Ala gtt	gat Asp ttt Phe 35 gcc	2142
gca Ala tct Ser 20 aaa Lys aac	ctc Leu 5 gtc Val ggt	gca Ala gac Asp gat Asp tac	acg Thr Cca Pro gcc Ala ggc	ctg Leu cag Gln agc Ser 40 ctt	gct Ala gaa Glu 25 gat Asp aat	ggg Gly 10 ctg Leu gcg Ala ccg	aag Lys atc Ile cag Gln tgg	ctg Leu acc Thr ttc Phe acg	gct Ala act Thr atc Ile 45 aaa	gaa Glu ctt Leu 30 gca Ala gaa	cgt Arg 15 cgc Arg tta Leu att	Me gtc Val cag Gln ctg Leu tac	et se 1 ggc Gly acg Thr atc Ile gcc	er Th atg Met gca Ala gtt Val 50 ttt	gat Asp ttt Phe 35 gcc Ala cct	2142 2190
gca Ala tct Ser 20 aaa Lys aac Asn gat	ctc Leu 5 gtc Val ggt Gly cag	gca Ala gac Asp gat Asp tac Tyr cag	acg Thr cca Pro gcc Ala ggc Gly 55 aat	ctg Leu cag Gln agc Ser 40 ctt Leu ggc	gct Ala gaa Glu 25 gat Asp aat Asn atc	ggg Gly 10 ctg Leu gcg Ala ccg Pro gtt	aag Lys atc Ile cag Gln tgg Trp ccg	ctg Leu acc Thr ttc Phe acg Thr 60 gtg	gct Ala act Thr atc Ile 45 aaa Lys gtg	gaa Glu ctt Leu 30 gca Ala Glu ggc	cgt Arg 15 cgc Arg tta Leu att Ile gtt	Me gtc Val cag Gln ctg Leu tac Tyr gat	et Se 1 ggc Gly acg Thr atc Thr gcc Ala 65 ggc	atg Met gca Ala gtt Val 50 ttt Phe	gat Asp ttt Phe 35 gcc Ala cct Pro tcc	2142 2190 2238
gca Ala tct Ser 20 aaa Lys aac Asn gat Asp cgc	ctc Leu 5 gtc Val ggt Gly cag Gln aag	gca Ala gac Asp gat Asp tac Tyr cag Gln 70 atc	acg Thr cca Pro gcc Ala ggc Gly 55 aat Asn aat	ctg Leu cag Gln agc Ser 40 ctt Leu ggc Gly gaa	gct Ala Glu 25 gat Asp aat Asn atc Ile aac	ggg Gly 10 ctg Leu gcg Ala ccg Pro gtt Val cag	aag Lys atc Ile cag Gln tgg Trp ccg Pro 75 cag	ctg Leu acc Thr ttc Phe acg Thr 60 gtg Val ttt	gct Ala act Thr atc 1le 45 aaa Lys gtg Val gat	gaa Glu ctt Leu 30 gca Ala Glu ggc Gly ggc	cgt Arg 15 cgc Arg tta Leu att Ile gtt Val atg	Me gtc Val cag Gln ctg Leu tac Tyr gat Asp 80 gac	et Se 1 ggc Gly acg Thr atc Thr atc Ala 65 ggc Gly ttt	atg Met gca Ala gtt Val 50 ttt Phe tgg Trp gag	gat Asp ttt Phe 35 gcc Ala cct Pro tcc Ser cag	2142 2190 2238 2286
gca Ala tct Ser 20 aaa Lys aac Asn gat Asp cgc Arg gac	ctcc Leu 5 gtc Val ggt Gly cag Gln aag Lys atc Ile	gca Ala gac Asp gat Asp tac Tyr cag Gln 70 atc Ile gaa	acg Thr cca Pro gcc Ala ggc Gly 55 aat Asn aat Asn tcc	ctg Leu cag Gln agc Ser 40 ctt Leu ggc Gly gaa Glu tgt	gct Ala Glu 25 gat Asp aat Asn atc Ile aac Asn	ggg gGly Gly 10 ctg Leu gcg Ala ccg Pro gtt Val cag Gln 90 tgc	aag Lys atc Ile cag Gln tgg Trp ccg Pro 75 cag Gln cgg	ctg Leu acc Thr ttc Phe acg Thr 60 gtg Val ttt Phe att	gct Ala act Thr atc Ile 45 aaa Lys gtg Val gat Asp tac	gaa Glu ctt Leu 30 gca Ala Glu ggc Gly ggc Gly cgc	cgt Arg 15 cgc Arg tta Leu att Ile gtt Val atg Met 95 aag	Me gtc Val cag Gln ctg Leu tac Tyr gat Asp 80 gac Asp	et Se 1 ggc Gly acg Thr atc Thr atc Ile gcc Ala 65 ggc Gly ttt Phe cgt	atg Met gca Ala gtt Val 50 ttt Phe tgg Glu aat	gat Asp ttt Phe 35 gcc Ala cct Pro tcc Ser cag Gln cat	2142 2190 2238 2286 2334
gca Ala tct Ser 20 aaa Lys aac Asn gat Asp cgc Arg gac Asp 100 ccg	ctc Leu 5 gtc Val ggt Gly cag gly cag glu Lys atc Ile 85 aat	gca Ala gac Asp gat Asp tac Tyr cag Gln ile gaa Glu tgc	acg Thr cca Pro gcc Ala ggc Gly 55 aat Asn aat Asn tcc Ser gtt	cttg Leu cag Gln agc Ser 40 ctt Leu ggc Gly gaa Glu tgt Cys acc	gct Ala Glu 25 gat Asp atc Ile aacc Asn aca Thr 105 gaa	gggg Gly 10 ctg Leu gcg Ala ccg Pro gtt Val cag Gln 90 tgc Cys	aag Lys atc Ile cag Gln tgg Trp Ccg Pro 75 cag Gln cgg Arg atg	ctg Leu acc Thr ttc Phe acg Thr 60 gtg Val ttt Phe att Ile gat	gct Ala act Thr atc Ile 45 aaa Lys gtg Val gat Asp tac Tyr gaa	gaa Glu ctt Leu 30 gca Ala ggc Glu ggc Gly ggc Gly cgc Arg 110 tgc	cgt Arg 15 cgc Arg tta Leu att Ile gtt Val atg Met 95 aag Lys cgc	Me gtc Val cag Gln ctg Leu tac Tyr gat Asp 80 gac Asp cgc	et Se 1 ggc Gly acg Thr atc Thr atc Ile ggc Gly ttt Phe cgt Arg gaa	atg Met gca Ala gttl Yal 50 ttt Phe tgg Glu aat Asn cca	nr gat Asp ttt Phe 35 gcc Ala cct Pro tcc Ser cag Gln cat His 115 ttc	2142 2190 2238 2286 2334 2382

Lys Thr Arg Glu Gly Arg Glu Ile Thr Gly Pro Trp Gln Ser His Pro 135 140 145	
aaa cgg atg tta cgt cat aaa gcc atg att cag tgt gcc cgt ctg gcc Lys Arg Met Leu Arg His Lys Ala Met Ile Gln Cys Ala Arg Leu Ala 150 155 160	2574
ttc gga ttt gct ggt atc tat gac aag gat gaa gcc gag cgc att gtc Phe Gly Phe Ala Gly Ile Tyr Asp Lys Asp Glu Ala Glu Arg Ile Val 165 170 175	2622
gaa aat act gca tac act gca gaa cgt cag ccg gaa cgc gac atc act Glu Asn Thr Ala Tyr Thr Ala Glu Arg Gln Pro Glu Arg Asp Ile Thr 180 185 190 195	2670
ccg gtt aac gat gaa acc atg cag gag att aac act ctg ctg atc gcc Pro Val Asn Asp Glu Thr Met Gln Glu Ile Asn Thr Leu Leu Ile Ala 200 205 210	2718
ctg gat aaa aca tgg gat gac gac tta ttg ccg ctc tgt tcc cag ata Leu Asp Lys Thr Trp Asp Asp Asp Leu Leu Pro Leu Cys Ser Gln Ile 215 220 225	2766
ttt cgc cgc gac att cgt gca tcg tca gaa ctg aca cag gcc gaa gca Phe Arg Arg Asp Ile Arg Ala Ser Ser Glu Leu Thr Gln Ala Glu Ala 230 235 240	2814
gta aaa gct ctt gga ttc ctg aaa cag aaa gcc gca gag cag aag gtg Val Lys Ala Leu Gly Phe Leu Lys Gln Lys Ala Ala Glu Gln Lys Val 245 250 255	2862
gca gca tag atctcgagaa gcttcctgct gaacatcaaa ggcaagaaaa Ala Ala * 260	2911
catctgttgt caaagacagc atccttgaac aaggacaatt aacagttaac aaataaaaac	2971
gcaaaagaaa atgccgatat cctattggca ttttctttta tttcttatca acataaaggt	3031
gaateecata eetegagett eaegetgeeg eaageaetea gggegeaagg getgetaaaa	3091
ggaageggaa caegtagaaa gecagteege agaaaeggtg etgaeeeegg atgaatgtea	3151
gctactgggc tatctggaca agggaaaacg caagcgcaaa gagaaagcag gtagcttgca	3211
gtgggcttac atggcgatag ctagactggg cggttttatg gacagcaagc gaaccggaat	3271
tgccagctgg ggcgccctct ggtaaggttg ggaagccctg caaagtaaac tggatggctt	3331
tettgeegee aaggatetga tggegeaggg gateaagate tgateaagag acaggatgag	3391
gatcgtttcg c atg gat att aat act gaa act gag atc aag caa aag cat Met Asp Ile Asn Thr Glu Thr Glu Ile Lys Gln Lys His 1 5 10	3441
tca cta acc ccc ttt cct gtt ttc cta atc agc ccg gca ttt cgc ggg Ser Leu Thr Pro Phe Pro Val Phe Leu Ile Ser Pro Ala Phe Arg Gly 15 20 25	3489
cga tat ttt cac agc tat ttc agg agt tca gcc atg aac gct tat tac Arg Tyr Phe His Ser Tyr Phe Arg Ser Ser Ala Met Asn Ala Tyr Tyr 30 35 40 45	3537
att cag gat cgt ctt gag gct cag agc tgg gcg cgt cac tac cag cag Ile Gln Asp Arg Leu Glu Ala Gln Ser Trp Ala Arg His Tyr Gln Gln 50 55 60	3585
ctc gcc cgt gaa gag aaa gag gca gaa ctg gca gac gac atg gaa aaa Leu Ala Arg Glu Glu Lys Glu Ala Glu Leu Ala Asp Asp Met Glu Lys 65 70 75	3633
ggc ctg ccc cag cac ctg ttt gaa tcg cta tgc atc gat cat ttg caa Gly Leu Pro Gln His Leu Phe Glu Ser Leu Cys Ile Asp His Leu Gln 80 85 90	3681
cgc cac ggg gcc agc aaa aaa tcc att acc cgt gcg ttt gat gac gat Arg His Gly Ala Ser Lys Lys Ser Ile Thr Arg Ala Phe Asp Asp Asp 95 100 105	3729

Trig3777String	Val Glu Phe Glu Glu Arg Met Ala Glu His Ile Arg Tyr Met Val Glu 125 acc att gct cac cac cag gtt gat att gat tag ag gta taa Thr Ile Ala His His Glu Val Asp Ile Arp Ser Glu Val * 130 3819 aacagataga agcttggtg ttttggggga tgagagaaga ttttcagoct gatacagat 3879 aaacagataga agcttggtg tttggagaa tgagagaaga ttttcagoct gatacagat 3879 gggcccacctg accacage gaacacaga ggaaacage caggacacag agggcgat tgagagagt 3939 gggcccacct gaccacage gaacacage caggacacag agggcgat tgagagagt 4059 gaaagactgg gccttcgtt ttactcgtg tgagagaacage caggaggg ggacggagg ggaggggg 4119 aaatcageeg ggageggat tgaacgte caggeaacag cccgagagg ggagggggg 4239 tttagegtt cacaaacte tttgttat ttttctaaa acattaa agtatagaa 4239 tttagegtt cacaaacte tttgttat ttttctaaa cattagagaag agtagaga 4119 catagaaa ataccetga aaagtaga ggctgada tgggaattg gaggtgg attg gaggacaga 4119 catagaaaa ataccetga taaatgeta ataatgega atagtaga tagtagaa 4119 cttaacaag a caggaggt taaagtata ag tggagaa taagtaga atagaga 4119 ctaacaaga acggagact tggacaca ag tggagaga tagtagaaga 4119 ctaacacag aagaact ttggacgaa tagtagtag atagtagtag 4119 ctaacacag aagaact ttgaacgag agacacag gacacacag ataggagaa 4119 ctaacacag aagagacatt tggacgaca agacacag agacacag agacacag agacacag agagacag	-continued	
Thr 11e Ål His Glň Val Åpp 11e Åpp Ser Glű Val * 130 aacagataga agcttggetg tittiggegga tagagaaga titteageet gatacagat gaacagataga agcttggetg tittiggegga tagagaaga titteageet gatacagat giggtotecee algeegaagt agggaate caggeeteaa ataaacegaa aggeteagte gaaageetig geetteegt titatetgtig titgteegig aacgeetee to tagataggae gaaageetig geetteege gaaceaga eggacagag geetagee gaaageetig geettee gaaageetig geettee gaaageetig gegaete gaaageetig geetaeaa gaaageetig geetaeaa gaaageetig geetaeaa geetaeaet geegeegee tittgeegtt etaeaete eegeeggee tittgeegtt etaeaete eegeeggee gittacateg eegeeggee geegeggee geegeeggee geegeeggee geegeeggee geeeaeeeee	Thr 11e Àia His Gin Val Aøp 11e Áøp ser Ölü Val * 135 aacagataga agcttagetg tittgagoga tagagaaga tittcageet gatacagat 3879 aacagataga agcttagetg tittgagoga tagagaaga tittcageet gatacagat 3989 ggetteccer atgegaagt agggaarde caggoratea ataaacgaa aggeteagt 4069 gaaagatgg octittett tattettit tittgeogt aacgetete tagataggee 4119 aaateegeeg ggaeeggat tagaegtee caggoratea ataaacgaa aggeteagte 4239 tittgeogtt etacaate tattettit tittetaat acatteaat atgateege 4299 teacagagae ataaceetg etacaate tattageagaag geeateetg etacagate 4419 cetacagaa ataaceetga taatgete ataatatt gaaaaggaag agtatgad 4419 cetacaaate ettittyttat tittetaat acatteaat atgatee 4539 gittacatega aeggeaget taacagega aggeeggata tageteggg eeggatat teegetggg 4779 gittacatega aeggaage act titgeege cagatage tagateggage 4779 gittacaatg agageact titgeegge caacttet teagaatga tagageggae 4779 citgeadaa cagagtgat accettage geggatat atgeeggege 4799 citgeadaag agetagag ecatacea aggeggae gaecacteg geggaegae 4799 gitacatega aeggaage cattetaage gegegaegae gaecacteg geggaegae 4779 citgeacaat agatgagaa cattage gegegaegae gaecacteg geggaegae 4799 citge	Val Glu Phe Gln Glu Arg Met Ala Glu His Ile Arg Tyr Met Val Glu	3777
aaataagaac gacgagacg tetgataaaa cagaattige etggegega tagegeggeg3939gicccaactig acccaatgee gaactaaga gigaaacgee giagegeega tagegaegig4059giagaattig goetteeet titatetgtig titigteggig aaegeeteet etgatagigge4119aaateegeeg gigaeeggat tigaaegtige gaageaegig eeeggaggi geeggaegig4129acgeeegeea taaaetgeea geetaaat taageaggag geetaeetig eigaaggaegig4239tittigegitt etaeaaetigeea geetaeaat taageaggaa geetaegig eigaggaegi4359tittigegitt etaeaaetig eitaeaetig eigaagaa geataggag agtatgaga4359tittigegitt eidaeaaete tittigtitat titteega attitegee eergaaga geataggag4379gittaeaetig eigagee eitatteeet tittigege attitege eergaaga4359gittaeaetig aaegeggig aaagtaaaga tagetggag geeggaata4359gittaeaetig ataggaeaet titaaagte tigetaggig geeggaata teeegggge4599acgeegggea aaggaeaet gigeegega taaeataet eergaaga tagaaggaa tatageag4599acgeegggea aaggaeaet gigeeggea taaeataet eergaagaa tatageagge4799cigeeaaaa eergettit tigeaaaaa tigegagaa tagaagaa tatageagge4799cigeeaaaa eaegettit tigeaaaaa tigegaaga getaadagaa tagaagaa4799giggaacega getgaatgaa geeateeaa aeggegge gaageaga diadagaa4799giggaacega getgaatgaa geeacaeaa tiggggaaea gaagaagaa tageaggee4799giggaacgaag aedetaegg geeagaaga getaeggage4799ciggeaaagae aaegtig gaageegga aagtaegag geeagaaga4799giggaacgga geeggaaag aeeteegg geeagaaga geeagaaga4799giggaaagae aaegtig geeagegeg geeagaaga geeagaaga geeagaaga4799giggaaagae aaegtig geeagaaga geeagaaga geeagaaga4799giggaaagae aaegeeggag aeegeegg geeag	aaatcagaac gcagaaqeeg tetgataaa cagaattege etgegeeg tagegeege a gigeteecee ageegaagt agegaatege cageaceea ataaacega ageteegee 4000 giggeteece ageegaagt tgaacege cageaceaa ataaacega ageteegee 4100 gaaagaeteg geetteegt taategteg gaaceaeg eccegaggget geeggeege 4179 acceecee ageegaagt tgaacegteg gaaceaeg eccegaggget geeggeege 4179 acgeecee taaaetegee geeteette tettetteg attegeege geeggeege 4179 cageecee tettee tettegee attegee gaaceaeg eccegaggget geeggeege 4179 tetaaetee tetteette tettettet tetteette acaeteeaa atgateege 4299 tetaaetee tetteette tettette tetteette acaeteeaa atgateege 4299 tetaaetee tetteette acaegeget agateette gaatgaga getadgage 44394 tetaaetee tetteette acaegeget agateette gaaged gedgegeege 4499 gitteeaat gatgagaet tetaaagte tgetagge geeggeate 4599 gitteeaat gatgagaet tetaaagte tgetagge geeggeate 4599 acceeggea agaegaete tetaagte tgetagge geeggeate 14599 gitteeaat gatgagaet tetaaagte tgetagge geeggeate 4599 acceedge aacegette tegeace acaegege caacetee tetagaagaa tatageage 4799 gigaacegga getgaatgaa geetaceaa agaegagge tgaacaeag atgetagge 4799 eaagagget aacegette tegeacaa agaegage tgaacaeag atgetagge 4799 eaagagget aacegette tegeacaa agaegage tgaacaeag atgetagge 4799 eaagaagga getgaatgaa geetaceaa acgeegage tgaacaeag atgetegge 4999 eaagaegag getgaatgaa geetaceaa acgeegage tgaacaeag atgetegge 4999 eaagaagga daacegette tegeaaaa gaaggage tgaacaeag atgetegge 4999 eaagaacega getgataga geetaceaa acgeegage tgaacaeag atgetegge 4999 eaagaegeg actatega gaaceaceg tegaagae tegaagag tetageege 4999 eaagaegeg actatega gaaceacge tegaagae tegaagae tegaagae 4999 eaagaete acaetgee egeetee egaacege tgaacaege tgaacaege 3999 eaagaegeg actatega gaacgaaga cacteeagaeg acteeaga 3999 eaagaegeg actatag tegaagaa agetegageg tgaatgaag 20000000000000000000000000000000000	Thr Ile Ala His His Gln Val Asp Ile Asp Ser Glu Val 🔺	3819
gitcccacctg acccatge gaacteagaa gigaacege giagegeega tiggtagtegi9999giggteteece atgegagat agggaactge caggacatea ataaacgaa aggeteagte4059gaaagaetig geettegt ttatetigtig tittgeeggig aacgeeteet gagaaggae4119aaateegeeg gigaeggatt tigaegtig gaageaaegg eeegeggig geeggeegge4239tittgegtt etaeaaete tittigtitat titteeaaa acatteaaa atgetaeceg4299teaacatte eegegee etaaaeteege gaaceaaat aageagaag geeateetig eeegagae4319cteaacatte eegigtee etatteeet tittigege attetgeet eeegagae4319cteaacatte eegigtee etatteeet tittigege attetgeet eeegagae4319cteaacatga aageetigt aaatgate gaedeaaga geeateetig geeegagae4319gittacateg atgagaet titaaagte tigetagge eeegagaa gatagagat4359acceecaga aageagaet eigeteege ataeetatte teagaatgae taggetigg4319ctgeacaaga gageaacte gitegeega taeetatte teagaatgae taggetigg4319ctgeacaag eaegagag eateetigg eeegagaa4319ctgeacaaga eaegtigg aacaetaegeg eeaaetae tetgaaagaa tategeagg4319ctgeacaaga eaegtigg aacaetaege geegagae attegeege4319ctgeacaaga eaegtigg aacaetaege geegagae attaeetae4319ctgeacaaag eaectaegeg eeaaetae tetgaaagaa tategeaga4319ctgeacaaaa eaegtigge eaaetataa etgegagae attaeetae tetgaaagaa tategeagae4319ctagaagaa daadagaga gatagaat agaegaae4319ctagaagaae aaegtigge aaetataa etgegagae gatgaagae tageagae4319ctageacaa aaegtigge aaetataa etgegagae gatgaagae tageagae4319ctaacaacaa aagtgaga gacaataa aggacagag gaagaaetagaa tageeggea4319ctaacaacaa aaettgag gaegeggata aggeagage gaagaaet accecagae4319 </td <td>gtcccacctg acccatgc gaactagaa gtgaaacgc gtagcgcqa tggtagtgt999gggtctcccc atgcgaqat agggaactgc caggcatcaa ataaacgaa aggctcagt4059gaaagactg gccttcgtt ttatctgttg tttgtcggt aacgctcc tgagtagac4119aatccgccq gaacggatt tgaacgttg gaagcacg cccggagggt gcggggcgg4129cttgggtt ctacaactc tttgttat tttctaaat acctcaaat atgtaccg4299tttaggtt ctacaactc tttgttat tttctaaat accttcaat atgtacg419ctacacatt ccgtgtege cttattccct ttttggg atttgg cccgaagat439gttacaccg accggggt aggtggg aggtggg gcggggg419gttacatcg atggact aaatggtt agatgtaga tcagtggg gcacgaagat439gttacatcg actgggtg aagtaaag atgctggg agttttgg cccgaagaa459gttacatga actgggtg aagtacagg gagtactg gagtgtgg gcacgagat459acgceggca aggcaact ggtcgcgca tacactat ccagaatgac ttggggg479ctgaagagg acccaggag agtacggg gcagaaga attaggag479ctgcaataa catggtgg aactacaa acggcg gcaactac gtagagga ttaggaga479ctgcaataa gatggaat acctggg gcaactac gtagggg gcgggata479ctgcaataa catggtgg aactacg gtaggagg gcgaaga attaggaga479ctgcaataa agatggag actttgg gagcggata acgtcgg gcgaatga gtaggagg479ctgcaataa agatggag actttgg gagcggata agatgaga ttaggaga479ctacgaaga accgttg gacggacga agatgaga gtaggaga479ctgaaggagt aacggtgg gagggagg ggggggg479ctgcaataa agatggag actatagg gacgacga gtaggaga479ctgcaacaa acggtgg aacggaga gtaggaga gtaggaga479ctgcaacaa acgtggg acctacg gggcgga gacgacga gtaggaga479ctgcaacag accggtgg accacaac gtgggggg gtagaaga gtaggaga479ctgcacaacg accgggg gtaggaga actat</td> <td>aacgagtaga agcttggctg ttttggcgga tgagagaaga ttttcagcct gatacagatt</td> <td>3879</td>	gtcccacctg acccatgc gaactagaa gtgaaacgc gtagcgcqa tggtagtgt999gggtctcccc atgcgaqat agggaactgc caggcatcaa ataaacgaa aggctcagt4059gaaagactg gccttcgtt ttatctgttg tttgtcggt aacgctcc tgagtagac4119aatccgccq gaacggatt tgaacgttg gaagcacg cccggagggt gcggggcgg4129cttgggtt ctacaactc tttgttat tttctaaat acctcaaat atgtaccg4299tttaggtt ctacaactc tttgttat tttctaaat accttcaat atgtacg419ctacacatt ccgtgtege cttattccct ttttggg atttgg cccgaagat439gttacaccg accggggt aggtggg aggtggg gcggggg419gttacatcg atggact aaatggtt agatgtaga tcagtggg gcacgaagat439gttacatcg actgggtg aagtaaag atgctggg agttttgg cccgaagaa459gttacatga actgggtg aagtacagg gagtactg gagtgtgg gcacgagat459acgceggca aggcaact ggtcgcgca tacactat ccagaatgac ttggggg479ctgaagagg acccaggag agtacggg gcagaaga attaggag479ctgcaataa catggtgg aactacaa acggcg gcaactac gtagagga ttaggaga479ctgcaataa gatggaat acctggg gcaactac gtagggg gcgggata479ctgcaataa catggtgg aactacg gtaggagg gcgaaga attaggaga479ctgcaataa agatggag actttgg gagcggata acgtcgg gcgaatga gtaggagg479ctgcaataa agatggag actttgg gagcggata agatgaga ttaggaga479ctacgaaga accgttg gacggacga agatgaga gtaggaga479ctgaaggagt aacggtgg gagggagg ggggggg479ctgcaataa agatggag actatagg gacgacga gtaggaga479ctgcaacaa acggtgg aacggaga gtaggaga gtaggaga479ctgcaacaa acgtggg acctacg gggcgga gacgacga gtaggaga479ctgcaacag accggtgg accacaac gtgggggg gtagaaga gtaggaga479ctgcacaacg accgggg gtaggaga actat	aacgagtaga agcttggctg ttttggcgga tgagagaaga ttttcagcct gatacagatt	3879
gggtctcccc atgoggagt agggaactgc caggcatcaa ataaacgaa aggctcagtc 4059 gaagactgg gcctttcgtt ttatcgttg tttgtcggtg aacgctcctc tgagtaggac 4119 aaatccgccg ggagcggatt tgaacgtgc gaagcaacgg cccggaggg ggoggagg 4179 acgcccgcca taaactgca ggcatcaat taagcagaag gccatcctga cggatggct 4239 ttttgcgtt ctacaaact tttgtttat ttttctaaat acattcaaat atgtatccgc 4299 tcatgagaca ataacctga taaatgctc aataattg aaaaaggaag agtatgagta 4359 ttcaacatt ccgtgtcgc cttatccct ttttgcggc atttgcct cctgttttg 4419 ctcaaccag acggggt aaagtaaaag atgctgaaga tcagtgggt gcacgagtgg 4479 gttacatcga atggatct aaaggcga agatcatg agatcttg gagtttcgc ccgaagac 4539 gttttccaat gatgagcact ttaaagtt tgctagg gggtata tccgtggtgg 4479 gttacacag aaggcagat agagcacg agaccaga tggggtag 4479 gttacacag acggggg aggcgaata agagcgg agatgaga taaggtgg 4479 gttacacag agggaact ggtgcgcga tacactat tcagaatga ttagcagg 4479 gttacacag acgggga aggcaact ggtgcgcg taacactat tcagaatga ttagcagtg 479 ccgaaggag aggcaact ggtgcgcg taacactat tcagaatgac ttggtggg 4479 gtggaaccgg agggaag acctggg gcagcgg tggacacg atgaggag 4479 cgaaggaga accggtgg aacatataa tgggggaca gaagaaga ttagcagtg 479 cgaaggagc accggtgg aacatataa ctgggggaca gtaagaga ttagcagtg 479 cgaaggaga accggtgg aacatataa ctgggggaca gtaacacg atgcgaggg 4479 gggaaccgg gcgaatgaa gccatacaa acgacgagg tgaacacag atgctgtag 4899 caatggcaa aaggtgga gagcggata aggtaggg acacttct gcgtcggcc 5019 ttccggctg ctggtttat gcgataaa ctggggga acacttct gcgccggcg 5019 ttacggag actaggga gagcgaata gacgaacg tgagacgg taggtagg tctcgcgg 5139 ggagtcagg actagggc gaagaaat gacagacg tgagaggg gacacttct gcgcaggg 5139 ggagtcagg actaggg gacgaggg tgggggg tacggagg gacgacgaa 5199 ttaagcattg gtaactgca gaccaagtt accacaga tggggtag gaccacga 5199 ttaggcac tcggcca aaacttga tgggggg ggt gggtgg gtcacgga ggggggg 5199 ccctgtagg gcgattaa gcgggggg tgggggg ggt gggtggg ggt ggggggg gggggg	gggtctcccc atgcgqagt agggaactgc caggcatca ataaaacgaa aggccagtc 4059 gaaagactgg gccttcgtt ttatctgttg tttgtcggtg aacgctcc tgagtaggac 4119 aaatccgccg ggagcggatt tgaacgttge gaagcaacgg cccggagggt ggcggcagg 4117 acgcccgcca taaactgcca ggcatcaaat taagcagaag gccatoctg cggatggcct 4239 ttttgcgttt ctacaaactc tttgttat tttctaaat acattcaaat atgtatccgc 4299 tcatgagaca ataaccctga taaatgctc ataatattg aaaaaggaag agtatgagta 4359 tttacacatt cogtgtcgc cttattccct ttttgggg atttgggt cccgaagac 4539 gtttccacag acgcgggt aagtaaaag atgctgag a gatctggg gccggaaca 4539 gtttccacag acggggta aagtacaag agtcttgg cgcggtata tcccgtgttg 4599 acgccgggca aggcaact ggtcgccgca tacactatc tcagaatgac ttgggtgggt 4779 gttacaccag cacagaaag catcttacgg atggcatgac agtaggag attaggtg 4779 cgaaggagct accggttg aacactggg ccaactat ctcgaaaga atggagga 4779 cgaaggagct accggttt ttggcacaa tggggggata agtacatgg tggcagag 4779 cgaaggagct accggttg gaaggagata agtggagg cgaactag ttggctggg 4779 cgaaggagc acacgttt ttggcacaac tggggggata agtacacg atggagga 4779 cgaaggagc acacgttt ttggcacaac tggggggata agtgcagg agtgcggg 4899 caatggcaag acggtgg gagggggata agtggagg gccgccg 5019 ttccggctgg ctggttat gcggaata acgtggag accattctg cgctcggcc 5019 ttccggctgg ctggttat gdgagaag ccccccgat cgtagttg 139 gggatcagge actatggat gaggaggg tgaagg gcgaacag tggacggg 132 gaggacag actggggc agtggaag ccccccgat cgtagtag 5199 ttaagcatg daggggc agtagaag ccccccgat cgtagtag 5199 ttaagcatg gaactgag cgacaagt accattat acttagat gattacgg 5259 ccctgtage gcgattaag cgcggggg tggggggg tagggggg gacggag gggggcg 539 ttaggcacg ccataggc cgctcctt gcttcc ctccttt cgccacgt 5379 gccggettt cccgtcaag tcaaacgg ggggcgg tgggggg gtcacggag gggggccg 539 ttaggcac tggaccaa accatgg tgggggg tgggggg gtcacggag gggggggg 539 ttaggcaca caaaactg ttggtgag tgggggg attcaagg gggcggag 539 ttaggcad cggccca aaaactgg tggggggg ttagggggg gadgag tcatggg gggcgccg 539 ttggcagg cctaag ccaaagt ctaaacgg ggggggg tgaatgag tggggggg 539 ttaggcacg tggcdaag ccatcacca acacggggg tggggggg atagggg gggggggg 539 ttggcaggg ggggggg ggaggggggggggggggggggg	aaatcagaac gcagaagcgg tctgataaaa cagaatttgc ctggcggcag tagcgcggtg	3939
gaaagactgg gocttegtt ttatetgttg tttgteggtg aacgetetee tggtaggac 4119 aaateegeeg ggageggatt tgaaegtte gaageaaegg eeeggaggg ggegggeagg 4179 acgeeegee taaaetgee ggeateaat taageagag geeateega eggatggeet 4239 ttttgegtt etaeaaete ttttgttat tttetaat aeatteaat atgtateege 4299 teatgagae ataaeeegg eetateee ttttgegge atttgeet eetgggg gaeggagg 4479 gttacatet eegtgee etateeet ttttgegge atttgeet eetgggg geeaggagg 4479 gttacateg aeeggggg aagtaaag atgetgaga teagttggg geeagagg 4479 gttacateg aeegggge aagtaaag atgetgag teagttggg geeagagg 4479 gttacateg aeegggge aagtaeaeg ggeeggat agateettg ggggtatta teeegtgg 4479 gttacateg aeegggee agageaee ggeegee teattee teagaatgae ttggetggg 4479 gtttecaat gatgageaet ttaaagte tgetatggg egeggtatta teeegtgg 4479 gtttecaat gatgageaet ggeegee teaeettee teagaatgae ttggetggg 4479 gegaagaegg aaggeaaee ggeegee taeeettee teagaatgae ttggetggg 4479 edgaagagee aaeegtggg acaeet ggeegee taeeettee teagaatgae ttggetggg 4479 gegaagggg a gaggagaa eaeettaegg atggeatgae agtaagaga ttatgeagt 4779 egaaggage aaeegtgg aaeettaegg atggeatgae agtaagaga ttatgeagt 4779 egaaggage aaeegtgg acaeet ggeegee caeettae teagaaegg atggaggee 4779 egaaggage aaeegtgg aaeetatae etggeggaea agtaeaegg atgeeggae 4799 eaaegeeggg eegaatgaa geeateae teggeggaea agtaeaegg tgaeeegg atgeeggee 4899 eaatggeae aaeegttge aaeetatee tegggage agtaeeeg atgegggge 4899 eaatggeae aaeegtgg gaegggaa aegteegg tgaeeegg tegeeggee 5019 tteageaeg eegaatga geeggaaaa eggeageegg tgaeeggg teeeeggg 5019 teateggaeg eaetggg gaeggagaa agaegaege tgaeegg geegetae 5079 teateggeeg eegaatag eggegggg tggegggeg gaegggeg gaeeggeegg 5139 ggageeagg eesataag egeeggegg tggeggga aeeettee geeegge 5199 teategeag eesatag eggeggeg ggegggeg ggeggegg gaeegge gaeegeegg 5199 teategeag eesatag eegegggeg ggeggegg ggegegegg aeeeggeg gaeegeegg 5199 teategeag eesatag eegegggeg ggegggeg atgagaegg gaeeggegg gaee	gaaagactgg octtecgtt ttatctgttg tttgtcggtg accgetetee tggaaggae 4119 aaateegeegeg ggaggatt tgaacgttg gaageaacgg eeggaagg ggeagtegg ggggggg 4179 aegeeegee taaaetgee ggeateaat taageagaag geeateetga eggatggeet 4239 tettgegtt etaeaaete tttgttat tttetaaat acateaaat atgtateege 4299 teatgagae ataaeeegg eegataaag atgetgag accategg ggeagggg ggatgggt 4479 gttaeaeatt eegtgeege etatteeet ttttgegge atttgeet eeggaggg geageggg 4479 gttaeatega actggeg etattee aaeagegga agatetggg geaggggg gagetggg 4479 gttaeaega actggege etattee aaeagegga agatetgg gagetggg geaggggg 4479 gttaeaega actggegg aagtaaag atgetgaga teagtgggg geaggggg 4479 gttaeaega actggagee etattee aaeagegga agatetgg geeggtatt eeeggage 4539 gttteeat gatgageaet ggeegee teaeatet teagaatgae ttggtgagg 4479 gegaagaage accggggg aggeegga agaeede ggeegge atgageaga taggegga 4779 egaaggage aggegaate ggeegee teegaag eegggaat ttageagg 4779 egaaggage accggtgg aaeettaag tggegggae agtagaega taggagga 4779 egaaggage accggtgg aaeetaae aggeggae agtaacege etgaagag attatgeagg 4799 acategeaga eaeggtgg aaeetaaa etggeggae agtaaceg atgegagae 4779 egaaggage accggtgg aaeetataa etggeggae agtaacege etgaagag attatgeagg 4799 eaaeattaa agaetggag gageggaa aagtgegga aaetatea teggeggae tegeegga 4799 eaaeaattaa agaetggag gageggaa aagtgegga accattet geeteegg 5139 ggageaegga etggttat geagaaat gaeggaegg tageeggag 509 teatgeage actggggee gatggtaage eeeeegg 5139 geagerage accategae gaegaaat gaegaaege tgagatgg geeeetaa 5319 ettaeegg eeeataag eegegegg tgggtggt geeeggag gaeegetaa 5319 ettaeegg eeeataag eegegegg tgggtggt geeeggag gaeegetaa 5319 ettaeegge geeataag eeeee geeeggegg gegeegegg attageege 5439 ettaeegge eeestaag eegegegg tgggtggg gteeggaegg gaeegetaa 5319 ettaeegge eeestaag eegegegg tgggtggg gateggaegg gaeegetaa 5319 ettaeegge eeestaag eegegegg tgggtggg gteeggaegg gaeegeegg 5439 ettaeegge eegataag eegegegg	gtcccacctg accccatgcc gaactcagaa gtgaaacgcc gtagcgccga tggtagtgtg	3999
aaatccgeeg ggaggggatt tgaacgttge gaageaacge ceeggaggt ggegggagg g4179acgeecegee taaactgee ggeateaat taageagaa geeateetaa atgtaecee4239tittgegtt etaeaacte tittgttat tittetaaat acatteaaa atgtaecee4399teadgagaa ataaceetga taaatgette aataattig aaaaagaga gatagagta4359teaacatti eegtgeege ettateeet tittgegge attegeet eegtgeggaga4479gttaeatega acgeggga aagtaaaga atgetgaga teagtiggi geeegagaga4599acgeegggea agagaacte ggtegeega taaattet eegaagaa ttaegetgg4599acteaceag acaggaga ettateeet tittgegg eegggtatt teeeggggaga479gttaeatag etagggaet titaagte tgetaggg geeggtata teeegtgggag479ctgeedaa eegggga agageace ggeegge acaettaet etagaaga attageagg479ctgeedaae etaeggg eegaatta eegtggg eegaatta teeggggaga479ctgeedaae acagtgeg eegaataa agaggeg tgaeedaga teggaggae479ctgeedaae acgetge aaeettae aeggegge taeaeeeg ateggagae479cgaaggaet accegttt tigeecaae aeggegge tgaeeeeg ategeggae479cgaaggaet accegtge aaeetataa etgeggae accecteg etgateet4899caatgedaa acgetgeg aaeetataa etgegaeg accaette geeteegg5019teeggeeg eegattaa geeggaata agatgeag accaette geeteegg5019teeggaeg aaeetaga gaegaaat agaagae eeceegaa gaegaaga teaggaeg519teagaeg geeetaa geeggegg tgeggegt aegeaged gaeeetae519teageegg eecetaag eegeegg gegeetee taegeageg gaeeetee539geggeetag geeetaa geeggegg tgeggegat aeeetaa eegeagge539gegaeagg eectaag eegeegg gegeetee taegeagg gaeeetee539teageegg geeetaag eegeegg gegeegg gegeetaa eegeeggg539ceetagaeg eegeegg geee	aaatccgccgggagggggtttgaacgttgcgaagcaacggcccggagggtggggggggg4179acgcccgccataaatgtcaggatcaaattaagcagaaggccatcctgacggtgggc4239ttttgcgtttctacaaactttttgttttttttcaaatacttcaaatatgttccc4299tcaatgagaaataacctgataaatgttcataatattgaaaaggaagagttaggta4359ttcaacattcccgtgtcgccttattcctttttgcggatttgcgtcccgagagg4479gttacatcgaacgcgggtaagagcactggacggggat45394599acgccgggcaagagcaactggtcgcgatccagtgggag4779ctgccataaccatgagtaaacactgggccaatgaagattrgggggacgacgcagggaagagcaactggtcgcggaatgagagattrggggagggggaccgacatgagtaaacactgggccaatgaca4779cgaaggagctaccagttttttgccaacatgggggaca4899aactaccagtcaaggtgaacactgggcaatgcaacg4779cgaaggagcaccgctttttgccaacatgggggaca4899caatgcaacacggtgggatacgctgggggtgacaccacg4899caatggaacacagttgggccatcacatggtgggat4899caatggaacacagtggggccatcacagtggtgggg5019ttccggcagctggtggggagctggggtgacaccagg5019ttccggacgcatggacggacgagggggccatggacg5019ttccgacaggcaggggacgacgagggggacgacggg5029ccatggacgccatggacg </td <td>gggtctcccc atgcgagagt agggaactgc caggcatcaa ataaaacgaa aggctcagtc</td> <td>4059</td>	gggtctcccc atgcgagagt agggaactgc caggcatcaa ataaaacgaa aggctcagtc	4059
acgccggca taaactgcCa ggcatcaaat taagcagaag gccatcctga cggatggcct 4239 tittgcgtti ctacaaacte tittgtttat tittetaaat acatteaaat atgtatecge 4299 teatgagaea ataaceetga taaatgette aataatattg aaaaggaag agtatgagta 4359 titeaacatti eegtgtegee ettatteet tittgegge attitgeett eetgttitg 4419 eteaacatga acgetggtg aaagtaaaag atgetgaaga teagttgggt geaegagtgg 4479 gttacatega acgggatete aacageggta agateettga gagtttege eeegagata 4539 gttteeaat gatgageaet titaaagte tgetatgtgg egeggtatta teeegtgg 4599 acgeegggea agageaaete ggtegeega taeaetatte teagaatgae titggtaggt 4719 etgeeataae eatgagtga aaceetgeg eegagtata teeegtggg 4719 etgeeataae eatgaggaa eaceetggg eegagtata teeegtggg 4719 etgeeataae eatgaggaa eaceetgg eegagtata teeegtggg 4719 eggaggaget aacegtggt aaeetategg eegagtata teeegtggg 4719 etgeeataae eatgagtga aaeetgegg eeaaetaet teegaeaeg ateggaggae 4779 egaaggaget aacegttit tigeeaaea igggggatea igaaeaeg ateggaggae 4899 eeaatggeaae aaegttgee aaeetataa etggeggaet attateet geeteggee 5019 tieeggetgg eegattat geegataaa eegagaeg igaaeetge geeteggee 5019 tieeggeegg eeggttatt geegataaa eegagaege igaagegggg teeegeggta 5079 teattgeage aeetatggat gaacgaaata eegagaeeg igagegtggg teeegegga 5139 ggagteagge eaetatggat gaacgaaata geeggeegg igagegtgg geeeaegag 5139 geggetegg eegattaag eegageggg iggeggata eegetaeg 5139 eeetgageg geeeataag eegeeggg iggeggta eegeegg igaeegeeae 5139 eeetgaege geeeataag eegeeggg iggeggata eegeeaeg igaeegeeae 5139 eeetgaege geeeataag eegeeggg iggeggata eegeeaeg igaeegeeae 5139 eeetgaeeg eeetaegee eegeeette gettettee etteettee eeeeegee 5139 eeetgaaeg eeetaegee eegeeegg iggeggeteet taeggeegg igaeegetae 5139 eeetgaaeg eeetaegee eegeeette gettettee etteette eeeeaegt 5139 eeetgaaeg eegetttee eeetteegge itegeggeegg iggegeeete 5379 eeetgaaeg eegetttee eeetteege eetteegeeet isess itgeteeaaa ettgaaeaa acteaaeee itegegge aceetgae eegeeege 5559 tigtteeaaa ettgaaeaa acteaaeee iteeggeegg atteeggeegg iteegegg ieegeegg 5579 aatttiaaea aaatataae gttaeeaat taaaggeeg iteegeegg eegeeeegeeegeeeeeegeeeeeegeeeeee	acgcccgcca taaactgcca ggcatcaat taagcagaag gccatcctga cggatggcct 4239 ttttgcgtt ctacaaact tttgttta ttttctaaat acattcaaat atgtatccgc 4299 tcatgagaca ataaccctga taaatgctt aataatattg aaaaaggaag agtatgagta 4359 ttcaacatt ccgtgtcgc cttatccct ttttgcgge atttgcct cctgttttg 4419 ctcacccaga aacgctggtg aaagtaaaag atgctgaaga tcagtgggg gccagagtgg 4479 gttacatcga actggatct aacagcggta agatcctga gagtttcgc cccgaagac 4539 gttttccaat gatgagcact ttaaagtt tgctatgtgg ggcggtatt tcccgtgttg 4599 acgccgggca agagcaact ggtcgccga tacactatt ctagaatgac ttggtgagt 4559 actcaccagt cacagaaag catctacg atggggatca gtagagaga ttatgcagtg 4779 cgaaggagct aacgcttt ttgcacaaca tgggggatca tgtaactgc cttgatcgtt 4839 gggaaccgga gctgaatga gccataccaa acgacgagcg tgaacaccag atccggagca 4959 acatggcaac aacgttgge aacatataa ctgggggatca tgtaactcg cttgatcgtt 4839 gggaaccgga gctgaatga gccataccaa acgacgagcg tgaacaccag atccggag 4999 caatggcaac aacgttgge aaactataa ctggggaaca actatct a gcttcccgg 4959 aacaattaat agactggatg aggcggata aagtgcagg accacttcg gctcggcc 5019 ttccggctgg ctggtttat gctgataaa ctggagcgg gacacactg gctcgggcc 5019 ttccggctgg ctggttat gcgaaaat gacagaacg tgagcggg gctcaccgg 5139 ggagtcagg actaggatg gagcgggg tgggggg taggtggg tccccggg 5139 gtagctagg cacatagg gagcgggg tgggggg taggaggg gaccactaca 519 ttaagcatg gtaactgca gaccaagtt actatat acttagat gattacgg 5259 ccctgtagcg gccattaag cgcgggggg tgggtggt acgcgcagg gaccactag 5139 gccggcttt cccgtcaag tctaaatcgg gggctccct tagggtcog attaatgcg 5439 ttaagcatg cacgaccaa aaaactgg tgggtggt acgcgacgat gtggcatag 5439 ttaagcaac actggacc gctcctt gctttctc cttccttct cgccacgtc 5439 ccctgataga cggttttcg cccttgacg ttgggtggt atccggag tgggcatag 5439 ttatcgcga tttcggccta ttggtaaaa aatgggtg ttccggag 5439 ccctgataga cggttttcg cccttgacg ttgggtgat atcatgag tgggcatag 5439 ttatccaaa atattaac gtttacaat taaaggat gttggtgat atcattaga 3439 ttatccaaa atattaac gtttacaat taaaggat attatagag 5439 aattttaaca aaatattaac gtttacaat taaaggat 2559 ttgttccaaa attaaac actaaccc atcaccgg attcggt aagataga 5539 acaaaaaaa ccacgcta ctaaaccg agttttgtc cactgaggt cggaaaaa	gaaagactgg gcctttcgtt ttatctgttg tttgtcggtg aacgctctcc tgagtaggac	4119
LittiggittiCitacaaacticLittigtitatLittictaaatacattcaaatatgtatcogc4299tittiggigaaataaaccigataaatgitticaataatattgaaaaaggaagagtatgagta4359titcaacatticogtgitegoccitattcocttittiggigaacagtiggiggacgaggigg4479gitacategaactggatcicaacageggtaagattcigggaggitticgcocggaggig4479gitacategaactggatcicaacageggtaagatccifggaggitticgcocggagacc4539gitticcaatgatgagcacttitaaagtictigetatgigcocggagata4559actcaccagtcacagaaaagcatcitaciggatgacaagaatitatgoagti4719cigeataaccatgagtaaaacactgigcigaagaagattitagcagig4719cigeataaccatgagtaaacactgiggcacadaagaa4779cigaaggagetaaccgctttititgcaacaatitgacaacg4799aacaactaacatgaggagacatgacagatitgcagaga4899caatggcaacaacgetiggcacaactacgtitgcagaga4779cigaaggagetaacgetiggcacaactacgtitgcagaga4899caatagcaacaacgetiggcacaactacgtitgcagagagad4899caatggcaacaacgetiggcacaactacgtitgcagagagad4899caatagcaacaacgetigggagcagagggtitgcacacg5019ticcaggcagciggittatgetigaaaciggitgagacattaccaga5079ticaggcagciggittatgetiggitgaagcigaitagagagagagagagagagagagagagagagaga	LittiggittiCitaaaactiLittigittatLitticaaatacattcaaatatgatcogo4299LicatgagacaataacoctgaLaaatgittiataatattigaaaaaggaagagtatgagta4359LicaacatticogtgtogoccitattooctLittigoggaattigoottcoggaggg4479gitacatogaacgotggitgaaagtaaaagagtotgaggatooggitgitgoggagg4479gitacatogaactggatotiaacagoggitagatottigggogggitticgitcoggagga4539acgoogggaagagacactiggtgooggitaagatottigggogggittittooggaggit4599acgoogggaagagacactiggtgooggitaagatagagatitgooggittit4599acgoogggaagagaaacatagagaataaactiggitgit4659actacocagtcacagaaaagcattatooggittigagagagaa4779cigaaggagtaacactgooggi caactactatodgagagaatitgooggittig4799cigaaggagetaacooggitaacactgoog caactatatictodgacaagagaa4779cigaaggagetaacooggittitittitgoocaacatitgooggittig4899caatagcaacaacgotggitgitgagacagagagtitgooggittig4899caatagcaacaacgotggittittitgoocaacatitgooggittig4999caatagcaacaacgittiggitgacacacagagtooggittig4999caatagcaacaacgittiggittitcodgacaagtitgooggittiggittititig4999caatagcaacaacgittiggittititiggagacacacaaftooggittititig5079ticatigcaacaacgitti	aaatccgccg ggagcggatt tgaacgttgc gaagcaacgg cccggagggt ggcgggcagg	4179
tcatgagaca ataaccctga taaatgctte aataattg aaaaaggaag agtatgagta4359ttoaacattt eegtgtegee ettatteeet ttttgegge atttgeett eegtgt4419ctcacccaga aacgetggtg aaagtaaaag atgetgaga teagtgggt geeeggaga4479gttacatega actggatete aacageggta agateettga gagtttege eegagaae4539gttteeaat gatgageaet tttaaagte tgetatgtg egeggtatta teeegtgg4719acgeegggea agageaaete ggtegeega tacactatte teagaatgaa ttatgeagtg4719ctgecataac catgagtgat aacaetgeg ceaacttaet tetgaaaaga attatgeagtg4719ctgecataac eatgagtgat aacaetgeg ceaacttaet tetgaaaag atgetgggg4879cgaaggaget aaccgettt ttgecaaca tgggggatea tgtaactege ettgategg4899caatggeaac aacgttgege aaactattaa etggegaaet acttaetta getteeegg4899caatggeaac aacgttgege aaactatta etggagegg tgaaccacag atgeetggg5019tteeggetgg etggttatt getgataaat etggageegg tagaegtggg tetegeggat5019tteeggetgg etggttat getgataga geetgatege etgagatege tegetggg5019ttaageatg gageegga gatagaaag etecteegg tagaeggegg tagaegtggg5019ttaageatg gaegeegge ggtggataa accactteg egedeggg5019ttaageatg gaegeegga gatggaaat agaegatege tgagatagg geeceegag5019ttaageatg gaegeegge ggtgggggg tagagaggg geeceet tegeeggegg5019ttaageatg geeetaag egegegggg tggtggtgta egegaagg gaeceetae5119ttaageatg geecetaage eegegeggg tggtggtgta egegaagg gaeceetae5139geggette eeggeegege tegaegge eggtggtgg ggtggggta gegegagg gaeceetae5139ttaageatg geecetaage eegegeggg gggggggg teaegtaegg tegeegaegg5439geeggette ceegteaage eegegeggg gggggggg tedegtae geegaegg5439ttaegeae ee	tcatgagaca ataaccctga taatgttc ataatattg aaaaggag agtatgagt4159ttcaacatt cogtgtogo ottattoot ttttgogge atttgoott octgttttg4419ctacaccaga aacgetggtg aaagtaaag atgetgaga toagtggg goacgagtgg4479gttacatoga actggatot aacagoggta agatoottga gagtttooc occgaagaac4539gtttocaat gatgacat tttaaagtte toctatgg ogcggtatta tocogtgtg4599acgeogggea agageaact ggtogoegea tacactatte toagaatgae ttagtgagg4779ctgacacagt cacagaaaag catctacgg atggagata agtaaggaa ttatgoagtg4789cgaaggaget aaccgtgtt ttgcacaaca tgggggatea tgtaactoge cttgatogt4839gggaaccgg gotgaatga gocatacaa acgacgagg tgaacacag atgoggage4779cgaaggaget aacggttg gacgatga gocatacaa acgacgage tgaaccacg atgotggag4899catggcaa aacgttgge aactatta toggggatea tgtaactoge cttgatogt4899catatgcaa acgttgge aactataa ctggogaca actatted gottocoge5019ttcoggetgg ctggtttatt gotgataaa ctggagegg taagtggg toctogegta5019ttcoggetgg ctggtttat gotgataag octocogtat cgtagttat tacacacag5139ggagtcagg acctaggge agtggeggg tggggggg toggetggg toctogegta5199ttaagcattg gtaactgte gaccaagtt acteatat acttagat gattage5199ctatgcage cctagege cgetoctte gettettee ctteette cgecacegt5199ctagcade togacceag totaactgg ggggegg tataggg gtteagt gggccade5319gtgcagae ttggcace togaccea aaactgg gggtggg ttaggtgg ttaggtgg tgggegg die5319gtgacacgg gcgegttaag oggeggggg tggggggg die5319gtggcacgg ccctagege cgtcette gettettee cttectte cgccacgte5319cttgccage ccgacge cgecetteg gggggggg die5319cttgccage ccgtaace tacaggg gggggggg die53	acgcccgcca taaactgcca ggcatcaaat taagcagaag gccatcctga cggatggcct	4239
ttcaacattt ccgtgtgcgc cttattcct tttttgggg atttggctt cctgttttt4419ctcacccaga acggtggt aagtaaag atgctgaga tcagttggg gccggtag4479gttacatcga acggatct acacagggt agatcttg ggggtttg cccggagaa4539gtttccaat gatgagcact tttaaagtt tgctatgg ggggtatta tccgtgtg4599acgccgggca agagcaacte ggtcgcoga tacactatte tcagaatgae ttggtaga4719ctgccataac catgagtga aacactggg ccaacttact tctgacaacg atcggagga4719ctgccataac catgagtga aacactgcg ccaacttact tctgacaacg atcggagga4719cgaaggagct aaccgettt ttgcacaac tgggggatca tgtaactge cttgatcgt4839gggaaccgg gctgaatga gccataccaa acgacgage tgacaccacg atcctgg4899caatggcaac aacgttgcg aaactatta ctggggaact acttatct agettcccg4959acaattaat agactggatg aggcggata aagtacgag tgacaccacg atgcctgtag4899caatggcaac aacgttgcg aaactataa ctggggaact acttactcta gctcccgg5019ttccggctgg ctggttatt gctgataat ctggagccg tgacgtgg tctcgcggt5019ttccggctg ctggttatt gctgataat ctggagccg tgagctggg tctcgcgg5139ggaaccagg cactatgga gaccaatta gacagtacg tgagatgg gcccactac5139gtggtcacg ccctageg c cgcccttt gcttgtat taccacgacg5139ttaagcattg gaccgaat gaccgagg gggggg tggtggtgta cgcgcagg gaccgcac5139ctgccaag cccage cgcccttt gctttctc ctcctttc cgccacgt5139gccggctttc ccgtcaage tcaaatcg gggctccctt tagggtccg attagtgca5139ctgccaaga cggttttcg cccttgacg tgggtggt tacggcag tgggcgata5139ctgccaage gccacaaca aaacttga tggggggg tccctt tagggtccg attagtgc5139ctgccaage ccgacccaa aaactgg gggcggg tccctt tagggtccg attagtgg5139ctgccaaga cggtccccac tggcgcc tgggggggg tccctt	ttcaacattt ccgtgtcgcc cttattcct tttttgggg attttgcctt ccgtttttg4419ctcacccaga aacgctggtg aagtaaaag atgctgaaga tcagttgggt gcacgagtgg4479gttacatcga actggatctc aacagcggta agatccttga gagtttcgc cccgaagaac4539gtttccaat gatgagcact tttaaagtc tgctatggg cgcggtatta tcccgtgtg4659acgccgggca agagcaactc ggtcgccgca tacactattc tcagaatgac ttggtagag4719ctgccataac catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac4779cgaaggagct aaccgcttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgt4839gggaaccgga gctgaatga gccatcaca acgacgagcg tgacaccacg atcgctgg4899caatgcaac aacgttgcg aactataa ctggcgaact acttactta gcttcccgc5019ttccggctgg ctggttatt gctgataac ctggagcgg tgagcggg tcccgcgg5019ttccggctgg ctggttatt gctgataac ctggagccg tgagcgtgg tcccgcgg5139gggatcagge aactatggat gaacgaaat gcagatgcg tgagtggg tcccgcgg5139ggagtcagge cactatggat gaacgaaat gcagatcg tgagtggg tcccgcg5139ttcagcatg gtgacggaca attagg ggcgggg tgggggta agcgcagg gacgcacaca5139ggagtcagg cactatggat gaccgaata gacagatcg tgagtggg tccccggacg5139gtggtcagg gccataag cgcggcggg tgggtggta cgccgcag gacgctaca5139gccggtttc cccgtcaagc tctaaatcgg gggctccct tagggtcg dtcaggagg5439ttaagcattg gtaactgca gaccattga ttgggtgg tggtggtg tcccgtag tgggcatcg5439ttacgcace tcgacccaa aaacttgat ttgggtgat gttcacgtag tgggcatcg5439cccgaaga cggttttcg cccttagc gttgggg tttaggg tggggag tggggag tggggga5439ttacgcace tcgacccaa aaacttgat tgggggag gtggagg tcccttta gtgggag5439ttacgcace tcgacccaa aaacttgat tggggag gtggagg gtgacg tacagtag tgggcgac5439 </td <td>ttttgcgttt ctacaaactc ttttgtttat ttttctaaat acattcaaat atgtatccgc</td> <td>4299</td>	ttttgcgttt ctacaaactc ttttgtttat ttttctaaat acattcaaat atgtatccgc	4299
ctcacccaga aacgetggt aaagtaaaag atgetgaaga teagttgggt geacgagtgg4479gttacatega actggatete aacageggt agateettg gagttttege eeegaagae4539gttttecaat gatgageaet ttaaagtte tgetatgtg egeggtatta teeegtggg4599acgeegggea agageaaete ggeegeega tacactatte teagaatgae ttaggetggg4719etgeeaaaag eaettaegg egegega tacactate teegaaagaa ttatgeagg4719ctgeeataae eatgagtga aaceetgeeg eeaactae teegaaegaa ttatgeagga4739gggaacegg egegatga aceettt ttgeeaaaa eggeggee tgeeaeeeg atgeetgeeg4839gggaacegg egegatga aaceettt ttgeeaaaa eggeggee tgeeeeeg atgeetgeeg4839gggaacegg egegatga aaceetataa etggegaaet aettaetet agetteege etgeege4959aacaattaat agaetggge aaactataa etggegaaet aettaetet geetgeegg5019tteeggeege eaettaggat gageeggata agetgeegg tgagegtggg teeeggegg5139ggagteagge aactatggat gaacgaaat gaeagaaee teggaageg geeteette5139ttaageattg gtaaetgee gaeeggegg ggegegget ageegaagg geeteette5139ttaageatg geegeggeg geeteette getteetee etteeteeteegeegge5139ctgeeaegg eeetaage egeegggg ggegegget gegegaegg gaeeggeat aettageeg5139ttaageatg geegeegeg egeettee getteete etteetee etteetee etteetee5379geeggeette eeegeege egeeette getteete etteetee etteeteetege5339ttaeggeae tegaaeaa etteege etteetegeege tegaaaga etteetege5339geeggeette eeegee egeeete taaateeg gggeeteett tagggteege attagtege5439ttaeggaee tegaaeae etteege gggeeteett tagggteege attagteege5439ttaeggeaee tegaaeae attege gggeeteet tagggteege attagtege5439ttaeggeaee tegaaeae aettege teggegge attagegeege etteegeege5439 </td <td>ctcacccaga aacgctggtg aaagtaaaag atgctgaaga tcagttgggt gcacgagtgg4479gttacatcga actggatctc aacagcggta agatccttga gagttttcgc cccgaagaac4539gttttccaat gatgagcact tttaaagttc tgctatgtgg cgcggtatta tcccgtgtgg4659acgccgggca agagcaactc ggtcgccgca tacactattc tcagaatgac ttggttgagt4719ctgacacagt cacagaaag catcttacgg atggcatgac agtaagagaa ttatgcagtg4739ctgacaggac accggtgtat aacactgcgg ccaacttact tctgacaacg atcggaggac4799cgaaggagct aaccgettt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt4839gggaaccgga gctgaatgaa gccatacca acgacgacg tgacaccacg atgcctgag4899caatggcaac aacgttgcgc aactatta ctggggatca tgtaactcgc cttgatcgt4959aacaattaat agactggatg gaggcggata aagttgcagg accattctg cgctcggcc5019ttccggctgg ctggttatt gctgataac ctggagccg tgagcgtgg tctccgggt5019ttccggctgg ctggttatt gctgataat ctggagccg tgagtgtgg tccccgtag5139ggagtCaggc aactatggat gaacgaatg gccgcaagt cgtgadtagg gccgcactacg5139ttaagcattg gtaactgca gaccagtt actcatat acttragatt gattacgc5139ctgcaggg ccctaage ccgccgcgg tggtggtta cgcgcagcg gaccgtaca5139ttaagcattg gtaactgca gaccagtt gctgtggt tccccgtag5139ctgcagg ccctaage ccgccgcgg tggtggtgta cgccacag tgccctacg5139ctgccagg ccctaage ccctaage tggtggtggt tggtggtgg tccccttt gcccacgt5139ctgccagg ccctaage cccccca aaacttgg gggcccct tagggtgg tggcgcag5439cttgccaag ccggtggt tggtggtgggt ggtggtggg tggcggag tgggcggac5439ccctgaaga cggttttcg cccttgacg ttggggg tggtggag tcccttta tagtggac5439ccctgaaga cggttttcg cccttgacg ttggggagg ttttatgggg tccc5439c</td> <td>tcatgagaca ataaccctga taaatgcttc aataatattg aaaaaggaag agtatgagta</td> <td>4359</td>	ctcacccaga aacgctggtg aaagtaaaag atgctgaaga tcagttgggt gcacgagtgg4479gttacatcga actggatctc aacagcggta agatccttga gagttttcgc cccgaagaac4539gttttccaat gatgagcact tttaaagttc tgctatgtgg cgcggtatta tcccgtgtgg4659acgccgggca agagcaactc ggtcgccgca tacactattc tcagaatgac ttggttgagt4719ctgacacagt cacagaaag catcttacgg atggcatgac agtaagagaa ttatgcagtg4739ctgacaggac accggtgtat aacactgcgg ccaacttact tctgacaacg atcggaggac4799cgaaggagct aaccgettt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt4839gggaaccgga gctgaatgaa gccatacca acgacgacg tgacaccacg atgcctgag4899caatggcaac aacgttgcgc aactatta ctggggatca tgtaactcgc cttgatcgt4959aacaattaat agactggatg gaggcggata aagttgcagg accattctg cgctcggcc5019ttccggctgg ctggttatt gctgataac ctggagccg tgagcgtgg tctccgggt5019ttccggctgg ctggttatt gctgataat ctggagccg tgagtgtgg tccccgtag5139ggagtCaggc aactatggat gaacgaatg gccgcaagt cgtgadtagg gccgcactacg5139ttaagcattg gtaactgca gaccagtt actcatat acttragatt gattacgc5139ctgcaggg ccctaage ccgccgcgg tggtggtta cgcgcagcg gaccgtaca5139ttaagcattg gtaactgca gaccagtt gctgtggt tccccgtag5139ctgcagg ccctaage ccgccgcgg tggtggtgta cgccacag tgccctacg5139ctgccagg ccctaage ccctaage tggtggtggt tggtggtgg tccccttt gcccacgt5139ctgccagg ccctaage cccccca aaacttgg gggcccct tagggtgg tggcgcag5439cttgccaag ccggtggt tggtggtgggt ggtggtggg tggcggag tgggcggac5439ccctgaaga cggttttcg cccttgacg ttggggg tggtggag tcccttta tagtggac5439ccctgaaga cggttttcg cccttgacg ttggggagg ttttatgggg tccc5439c	tcatgagaca ataaccctga taaatgcttc aataatattg aaaaaggaag agtatgagta	4359
gttacatcga actggatctc aacageggta agatccttga gagttttege eccgaagaac4539gttttecaat gatgageact ttaaagtte tgetatgtg egeggtatta teeetgttg4599acgeegggea agageaacte ggtegeegea tacactatte teagaatgae ttagetggg4659acteaceagt eacagaaag eacttaegg atggeatgae agtaagagaa ttatgeagtg4719etgeeataac eatgagtgat aacactgegg eccaacttaet tetgaeaaeg ateggaagae4739eggaacegga getgaatgaa geeataeea tgggeggatea tgtaaeteeg ettgateegt4839gggaacegga getgaatgaa geeataeea etgggegate atetaeet ageetegge4839ecaatggeaae aacgttgee aaactattaa etggegaaet actaeeta getteegeggta4899ecaatggeage ateggggg gageggata aagttgeagg aceaetteg egetegge5019tteeggetgg etggttatt getgataat etggageegg tgaegeggg tedegegge5019tteeggetgg eactatggat gaagegaata agetgeeg tgagatgg geteedeegeg5139ggagteagge aactatggat gaacgaaata etegaageeg tgagatgg geteedeegeg5139ggagteagge aactatggat gaeegaata gaeagatege tgagatagg geteeceta5139ttaageattg gtaactgtea gaeegagtg getgeggt gagegggg gaeegegg gaeegegegg geeedeegeg5139ttaggeateg geeetaag egegegggg gegegeggg teggegageg gaeegegeg gaeegegeg5139ttaegeattg gtaactgea gaeegaata gaeegatege taggetegg gaeegegeg gaeegegeg5319ttaegeattg geeetaag egegegggg gegeeteet tagggteeg gateegeg attaegeg5339cectgaagg geetaaag egegegggg gegeeteet tagggteeg gattag tegeegegg attaagege5319ttaeggeaee tegaeceaa aaacttga ttgggtgatg gteceetag tgggeeteege5319ttaeggeaee tegaeceaa aaacttga ttgggtgatg gteceetag tgggeeteege5319ttaegeaee tegaeceaa aaaacttga ttgggtgatg gteceetag tgggeeteege5439ttaeggeaee	gttacatcga actggatctc aacageggta agatcttga gagttttege eccgaagaac4539gtttteeaat gatgageact ttaaagtte tgetatgtg egeggtatta teeegtgg4599acegeegggea agageaacte ggtegeegea tacactatte teagaatgae ttagetggg4659acteaceagt eacagaaag eacttacgg atggeatgae agtaagagaa ttatgeagtg4719ctgeeataae eatgagtgat aacactgeeg eccaectaet tetgaeaaeg ateggaggae4739eggaacegga getgaatgaa gecataeea aeggeggatea tgtaacteeg ettgategt4839gggaacegga getgaatgaa gecataeea aeggeggaeg tgaeaeeag atgeetggee4959acaattaat agaetggatg gaggeggata aegttgeag accaettet getteeggee5019tteeggetgg etggttatt getgataat etggageegg tgagetggg teegeggta5079teattgeage aetggggee gatggtaage ecteceegtat egtagatagg geeteeteeggee5139ggagteagge daeetatggat gaeegaata gaeagateg tgagatagg geeteeteeggee5139teatgeage geetatagg egegeggg tgggeggta egegagg gaeegetae5179ttaageattg gtaaetgee geeteette getteetee etteegaatgge geeteeteeg5139geeggette eccegtae egeegggg tgggeggta egegaegg gaeegetae5139gttaceage geetataag egegegggg tgggtggta egegaegg ageegetae5139ttaageattg gtaaetgee geeteette getteetee etteegaatggee geeseteetee5379geeggette eccegaeg eectaegge tgggtggta egegaegg ageegeatea5139ttaggeaee tegaeee aaaettga ttggggegee etteegaag tgggeeteete5379geeggette eccegae ectaegee ttggggegg gggeteeett tagggteegg attaggee5499ccctgaaga eggtttteeg ecttageege ttgggeeteege ttggegaaga teggaage5499ccctgaaga eggtttteeg eccttageege ttgggeeteege ttggegaaga eccettag5599ttgteceaa ettggacea acteaeeet ateegggee ttggegaaga teggaege tege	ttcaacattt ccgtgtcgcc cttattccct tttttgcggc attttgcctt cctgtttttg	4419
gttttccaatgatgagcacttttaaagttctgctatgtggcgccggtattatcccgtgtgg4659accgccgggaaggcaactcggtcgccgatacactattctcagaatgacttggtggg4719ctgccataaccatgagtgaacactgcggccaactattctcagaagaatatgcagtg4739cgaaggagctaaccgctttttgcacaacatgggggatcatgtaactcgcttggagggg4839gggaaccggagctgaatgaagccataccaaacggcgggatcatgtaactcgcttgaggggaaccggagctgaatgaagccataccaaacggcgggatcaactactcac4839caatggcaacaacgttgcgcaacattatactggcgaacactactcacg4959aacaattaatagactgggtggagcgggataacgttgcggccatcgccgc5019ttccggctggctggtttattgctgataaactggagcggtacacattag5139ggagacaggaaactatggatgaacggagtcctccgtatggagatggg5139ggggtcaggacattggatgaacggggggggggggggggtggagtaggggccgacgg5139gccgggtggccgattaggccgggggggggggggggggggggggggggggggggg	gttttccaat gatgagcact tttaaagttc tgctatgtgg cgcggtatta tcccgtgtgg4599acgccgggca agacaactc ggtcgccgca tacactattc tcagaatgac ttggttggat4659actcaccagt cacagaaag catcttacgg atggcatgac agtaagagaa ttatgcagtg4719ctgccataac catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac4779cgaaggagct aaccgcttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt4839gggaaccgga gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgac4959aacaattaat agactggatg gagcggata aagttgcagg accacttctg cgctcggcc5019ttccggctgg ctggtttatt gctgataat ctggagccgg tgagcgtggg tcccgggta5079tcattgcagc actggggcca gatggtagc cctcccgtat cgtagttatc tacacgacgg5139ggagcacgg gcgattag gacgagata agttgcagg tgagcgggg gcccactta5139ggagtcagg acctatggat gaccgaggt gtggtggta cgcgacg gcccactga5379gcctgtagcg gcgcattaag cgcggcggg gtggtggta cgcgcacg gaccgcacg5319cttgccagc cccagcgcc cgctccttc gcttctcc cttccttct cgccacgtt5319cttgccagc tcgacccaa acaacttgg gggcgcac tagggtggg gtggggg gtgcgagg gtggggccactg5499ccctgtagcg ccctagcgc cccttagcg ttgggggg ggggggg gtgcccact taggggccactg5499ccctgtagc ccctagcgcc ccctttg gcttcttcc cttcctttct cgccacgttc559ttacggcact tcgacccaa acacttgat ttgggtggg attcacgag tggggaga tcataggg5619attttgccga tttggccaa attaga gtttacgta attaggg tccctaga gccgacac5619ccctgataga cggttttcg ccctttgacg ttggggt tttaacaaa attaacgcg5619atttgccaa ttggaccaa actcaccct atccgggc attaacaa attaacgc5619attttgccga tttggccaacaac actacgacct accgggg gttttcgtc cactgaggt cagcactac579accatgaaca actgggccaacaacc accaccct atccgggg	ctcacccaga aacgctggtg aaagtaaaag atgctgaaga tcagttgggt gcacgagtgg	4479
acgccgggcaaggcacactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatttacggatggcatgacagtaggagaattggttgagtctgccataaccatggdgataacactgcggccaacttacttctgacaacgatggaggagagggaaccggagctgaatgaaggcatgacaggccggcttgaactcgcttgaacggggggaaccggagctgaatgaagccataccaaacggcgggctgaaccacgatgccggggggaaccggagctgaatgaagccataccaaacggcgggctgaaccacgttgsgggaaccggagctgaatgaagccataccaaacggcgggctgacaccacgttgsgggaaccggagctggattaagttgcggtggggaacattgccggctasgggaaccggagctggatggcgaagttgcaggtgacaccacgttgs4959aacaattaatagactgggacaactattaactggggaacattactcatgctccggc5019ttccggctggctggtttattgctgataaactggagccggtgaggtcggg5079tcattgcagactatggatgaacgaatagacgagcggtgaggcggg5139ggagtcaggcactatggatgaacggagggaaggatgg5139ttaagcattggtaactgcagaacgggggggggtggtggttagcgcagcg5259cctgacaggcccgtcaaggtgggtgggggtggtggatg5439cttgccaagccgaccccaaaaactggggggtggg5439cttgccaagccgaccccaaaaactggggggaac5439cctgaccacccgaccccaaaaactggggggaac5439cccgatagacggccatac<	acqccqggca agagcaactc ggtcqccqca tacactattc tcagaatgac ttggttgagt4659actcaccagt cacagaaaag catcttacgg atggcatgac agtaagagaa ttatgcagtg4719ctgccataac catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac4779cgaaggagct aaccgcttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt4839gggaaccgga gctgaatgaa gccataccaa acgacgageg tgacaccacg atgcctgtag4899caatggcaac aacgttgcg aaactattaa ctggcgaact acttactcta gcttcccggc4959aacaattaat agactggatg gaggcggata aggtcgag accacttcg cgctcggcc5019ttccggctgg ctggtttat gctgataat ctggagcgg tgagcgtgg tccccggta5079tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttat tacacgacgg5139gggatcagg aactatggat gacgaaata gacgaata gcagatagt cgtagtgg gctcactga5139ggagtcagg gcgcattaag cgcggcggt gtggtggta cgcgacgg gacgcatca5139ggagtcagg gccattaag cgcgcgggg gtggtggta cgcgcacg5379ccctgtagcg gcgcattaag cgcgcgggg ggggctccctt taggttcg gaccgcatca5319cttgccagcg ccctagcgcc cgctccttc gcttctcc cttccttct cgccacgtc5499ccctgataga cggttttcg cccttgacg ttggggtag gttcacgtag tgggccatcg5499ccctgataga cggttttcg cccttgacg ttggggtca cgttcttaa tagtggact5419ttacggcacc tcgaccccaa aaacttgg ttggggtca cgttcttaa tagtggacc559ttgtccaaa cttgaccaa actaacga attagag ttggggt attacagtag tggggcatag5479gccggttttc cccgtcaagc tctaaatcg gggcggta ttggggtggg gtcccctt agggtgggg fcgcdatag5479gccgtagg gccattaag cgcgcgggg tggggggt cccttaaggt gaccgcatcg5379gccgggtttc cccggccc cgtcctttc gcttggatggg gtcgcdag gtgggggg5439cttgccaaa ctgaacaaca actcaacct atccggggt attcacgtag tggggacatcg	gttacatcga actggatctc aacagcggta agatccttga gagttttcgc cccgaagaac	4539
actcaccagtcacagaaaagcatctacggatggcatgacagtagagaattatgcagtgctgccataaccatgggtgataacactgcggccaacttacttctgacaacgatggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgtggggaaccggagctgaatgaagccataccaaacggcggggttgtaactcgctdtagcaatggcaacaacgttgcgcaactattaactggcggactacttactctagctcgcggcaacaattaatagactgggtggaggcggataaagttgcaggtgtactcgc5019ttccggctggctggtttattgctgataaactggagcgggtaccactacg5139ggagtcaggcaactatggatgaaggagcgtgtagttatctaccagacgg5139ggagtcaggcaactatggatgaacgaaatagacagaatgggactgatgg5139ggagtcagggtaactgtcagaacgaatagacagaatgggacagacgc5139ggagtcagggccataaggaccgaggggggtgtggttacgcgcaggg5259cctgtagcggccataggccgccacggcggaccgacgt5319gccggctttccccggtcaagttagggtgggtacggcacgg5439ttacggcacctcgacccaaaaaacttgatttgggtggttgggtgcacga5499cctgatagacgtttttcgccttggcgttgggtgca5439ttacggcacctcgaccacaacaacttggggggcccact5439ttacggcacctcgaccacaaaaacttgggtgggtcacgtatgg5619atttgccgattggttaaaactggggatgatatttttga5619attttgccgattggttaaaat	actcaccagtcacagaaaagcatttacggatggatgacatggaggattatggagggctgccataaccatggatgataacactgcggccaacttactttggacacgaatggaggacgggaaccggagctgaatgaagccataccaatgggggatcatgtaactcgccttgatcgtt4839gaggaaccggagctgaatgaagccataccaaacggcggggtaaacacttactttggacacga4899caatggcaacaacgttgcgcaaactattaactggcgaactactactcatgctcccggc4959aacaattaatagactgggatggaggcgggataaagttgcaggaccacttctggctcgggcc5019ttccggctggctggtttattgctgataaatctggagcgggtacagacgg5139ggagtcaggcaactatggatgaacgaaatagacgatggctacacgacgg5139ggagtcaggcaactatggatgaacgaaatagacgatggctacacgacgg5139ggagtcaggcaactatggatgaacgaaatagacgatggcgagataggt5139ggagtcaggcaactatggatgaacgaaatagacgatggcgagataggt5139ggagtcagggccatagggaccaaggttactactatatactacacgacg5139ggagtcagggccatagggccataggggggggggggagataggtgattacgg5139ggagtcagggccatagggaccaaggggggggggggagatggg5139ggagtcagggccatagggggggggggggagataggtgccatagg5139ggagtcagggccatagggggggggggggggggggggggagataggt5139gccggacttggaccaaggggggggggggggggggggggggggggggg <td< td=""><td>gttttccaat gatgagcact tttaaagttc tgctatgtgg cgcggtatta tcccgtgttg</td><td>4599</td></td<>	gttttccaat gatgagcact tttaaagttc tgctatgtgg cgcggtatta tcccgtgttg	4599
ctgccataac catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac 4779 cgaaggagct aaccgcttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt 4839 gggaaccgga gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgtag 4899 caatggcaac aacgttgcg aaactattaa ctggcgaact acttactcta gctcccggc 4959 aacaattaat agactggatg gaggcggata aagttgcagg accacttctg cgctcgcgcc 5019 ttccggctgg ctggtttatt gctgataat ctggagccgg tgagcgtggg tctcgcggta 5079 tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 5139 ggagtcagge aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 5199 ttaagcattg gtaactgtca gaccaagtt acttatta acttagatt gattacgcg 5259 ccctgtagcg gcgcattaag cgcggcggg gtggtggta cgcgcaggt gacgctaca 5319 ctgtccageg ccctagcgcc cgctccttc gcttctcc cttccttct cgccacgtt 5379 gccggctttc cccgtcaagc tctaaatcgg gggctcact tagggttccg attagtgct 5439 ttacggcacc tcgacccaa aaacttgat ttgggtgatg gtcacgtag tgggccatcg 5499 ccctgataga cggtttttcg cccttgacg ttggggtca cgtctttaa tagtggact 5559 ttgttccaaa cttggacca actaaccct atccgggct attattataggg 5619 attttgccga tttcggccta tggttaaaa aatgagctga tttaacaaaa attaacgg 5679 aattttaaca aaatattac gtttacaat taaaaggatc taggtgaag tccttttga 5739 taatctcatg accaaaatc cttaacgtg gtttcgtc cactgaggt cagacccgt 5799 agaaaagatc aaaggatct cttaagtg gtttcgtc cactgagcg cagacccgt 5799	ctgccataac catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac 4779 cgaaggagct aaccgcttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt 4839 gggaaccgga gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgtag 4899 caatggcaac aacgttgcgc aaactattaa ctggcgaact acttactcta gcttcccggc 4959 aacaattaat agactggatg gaggcggata aagttgcagg accacttcg cgctcggccc 5019 ttccgggctgg ctggtttatt gctgataat ctggagccgg tgagcgtggg tctcgcggta 5079 tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 5139 ggagtcaggc aactatggat gaccaagtt acttaatt acttagatt gattaccgcg 5139 ttaagcattg gtaactgca gaccaagtt actcattat acttagat gattaccg 5259 ccctgtagcg gcgcattaag cgcggcggg gtggtggtta cgcgcagcg gaccgctaca 5319 cttgccagc cctagcgcc cgctcctte gcttcttce cttccttet cgccacgtt 5379 gccggcttte cccgtcaage tctaaatcgg gggcgtag gtggtgatg gtcacgtag tgggccatcg 5499 ttaagcact tggaccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg 5499 ccctgataga cggttttteg cccttgacg ttggggtag ttcacgtag tgggccatcg 5499 ccctgataga cggttttteg cccttgacg ttggagtca tgttattata tagtggact 5559 ttgttccaaa cttggaccaa actcaaccct atctcgggct attctttga tttataggg 5619 attttgccga tttcggccta ttggttaaa aatgagtcg ttagtgaga tccttttga 5739 gagaaaagatc aaagattac gttacaatt taaaaggatc taggtgaag tccttttga 5739 aaatttaaca aaatataac gtttacaatt taaaaggatc taggtgaag tccttttga 5739 aaaaaaaaa ccaccgctac cgcggtgg tttttteg cactgaggt cagacccgt 5799 agaaaagatc aaaggatct cttgaggt ttttttg cggtaatc gctgctgca 5799 aacaaaaaaa ccaccgctac cagcggtgg tttgttgcg gatcaagag taccaactt 5799	acgccgggca agagcaactc ggtcgccgca tacactattc tcagaatgac ttggttgagt	4659
cgaaggaget aacegettt ttgeacaaca tgggggatea tgtaactege ettgategtt 4839 gggaacegga getgaatgaa geeataecaa aegaegageg tgaeaeeaeg atgeetgtag 4899 caatggeaae aacgttgege aaaetattaa etggegaaet aettaeteta getteeegge 4959 aaeaattaat agaetggatg gaggeggata aagttgeagg aeeaettetg egeteggeee 5019 tteeggetgg etggttatt getgataaat etggageegg tgagegtggg tetegeggta 5079 teattgeage aetggggeea gatggtaage eeteeegt egagataggt geeteaetgg 5139 ggagteagge aaetatggat gaacgaaata gaeagatege tgagataggt geeteaetga 5199 ttaageattg gtaaetgtea gaeegaata gaeagatege tgagataggt geeteaetga 5199 ttaageattg gtaaetgtea gaeegaggt gtggtggtta egeegageg gaeegetaea 5319 eetegeage eeteggegee egeteette getteettee etteette	cgaaggaget aacegettt ttgeacaaca tgggggatea tgtaactege ettgategtt 4839 gggaacegga getgaatgaa geeataeeaa aegaegageg tgaaceaeg atgeetgtag 4899 caatggeaae aaegttgege aaaetattaa etggegaaet aettaeteta getteeegge 4959 aaeaattaat agaetggatg gaggeggata aagttgeagg aeeaettetg egeteggeee 5019 tteeggetgg etggttatt getgataaat etggageegg tgagegtggg tetegeggta 5079 teattgeage aetggggeea gatggtaage eeteecgtat egtagttate taeaegaegg 5139 ggagteagge aaetatggat gaacgaaata gaeagatege tgagataggt geeteaegg 5139 ttaageattg gtaaetgtea gaeegaaata gaeagatege tgagataggt geeteaegg 5139 eeteggegg egeeattaag egeggeggt gtggtggtta egegeagegt gaeegetae 5319 ettgeeageg eeetagegee egeteette getteette etteettet egeeaegte 5379 geeggettte eeegteaage tetaaategg gggeteeet tagggtteeg attagtget 5439 ttaeggeaee tegaeeeeaa aaaaettgat ttgggtgatg gtteeegtag tggggeeateg 5499 eeetgataga eggtttteeg eeetteggeet atteettaa tagtggaete 5559 ttgtteeaaa ettggaeeae aeteaaeeet attegggtea eggtegat attagtgg 5619 atttggeega tteeggeeta ttggttaaaa aatgagetga tttaaeaaaa attaaegg 5679 aattttaeea aaatattae gtttaeeatt taaaaggate taggtgaaga teettttga 5739 taateetag aecaaaatee ettaaegg ggttteegte eaetgaegt eagaeeegt 5799 agaaaagate aaaggatett ettggatee tttttteg egetaatet getgetge 5799 agaaaagate aaaggatett ettagggteg tttetteg eaetgaegt eagaeeegt 5799 aacaaaaaaa eeaeegtae tetgagatee ttttttteg ggtaaegt eagaeeegt 5859	actcaccagt cacagaaaag catcttacgg atggcatgac agtaagagaa ttatgcagtg	4719
gggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgctgtagaacaattaatagactggatggaggcggataaagttgcaggaccattctagcttcccggc5019ttccggctggctggtttattgctgataaatctggggacggtgaggtgggtcccggcgg5079tcattgcagcactggggccagatggtaggcctcccgtatcgtagttatctacacgacgg5139gggatcaggcaactattggatgaacgaaatagacagatcgctgagataggtgctccactga5199ttaagcattggtaactgtcagaacgaaatagacagatcgctgagataggtgctccactga5299ccctgtagcggcgcattaaggcgggcgggtgtgtggtgttacgcgcagcgtgaccgctaca5319gctgcagcgccctagcgcccgctccttcgcggaccgg53795379gccggctttccccgtcaagctctaaatcgggggtcaggsttagggtccg5439ttacggcacctcgacccaaaaaacttgatttgggtggtsttatagggacc5559ttgttccaaacttgaccaaactacaccctactctgggc5619atttgccgatttggttcaaaatttaacgttacaaaattaacgcg5739aatttaacaaatattaacgttacaatttaaaggatcaatttaacg5739aatttaacaaatattaacgttacaatttaagggtcg3739taactcaagcctgaccaaaatttaacg5619atttgccgatttggttaaaaaatgggtcg3739taatctaatactacaatttaagggtcg5799aatttaacaaatattaacgttacaattcaatgggtcg5799agaaagatc<	gggaaccgga gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgtag 4899 caatggcaac aacgttgcgc aaactattaa ctggcgaact acttactcta gcttcccggc 4959 aacaattaat agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc 5019 ttccggctgg ctggttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta 5079 tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 5139 ggagtcaggc aactatggat gaacgaata gacagatcg tgagataggt gcctcactga 5199 ttaagcattg gtaactgtca gaccaagtt acteatat acttagatt gattacgcg 5259 ccctgtagcg gcgcattaag cgcggcgggt gtggtggtta cgcgcagcg gaccgctaca 5319 gttacgcagc ccctagcgc cgctccttc gcttctc cttccttct cgccacgtc 5379 gccggcttc cccgtcaage tctaaatcgg gggctccctt tagggttcg attagtgct 5439 ttacggcace tcgacccaa aaacttgat ttgggtgatg gttcacgtag tgggccatcg 5499 ccctgataga cggtttttcg ccctttgacg ttggggtc attatgtgg 5619 attttgccaa cttgaacaac actcaaccc atctcgggct attattag tttataggg 5619 attttgccaa cttgaacaac actcaacct atctcgggct attatatatagtggac 5679 aattttaaca aaatattac gtttacaatt taaaaggatc taggtgaaga tccttttga 5739 taatccatg accaaactc cttaacgtg gtttcgtc cactgacgt cagacccgt 579 agaaaagatc aaaggatct cttaaggt gtttcgtc cactgacgt cagacccgt 5799 agaaaagatc aaaggatct cttgagt ttgtttgcc gatcaagag taccaactt 5799 agaaaagatc aaaggatct ctgaggtc ttttttcg cgcgtaatc gctgcttgca 5859	ctgccataac catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac	4779
caatggcaac aacgttgcgc aaactattaa ctggcgaact acttactcta gcttcccggc 4959 aacaattaat agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc 5019 ttccggctgg ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta 5079 tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 5139 ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 5199 ttaagcattg gtaactgtca gaccaagttt actcatatat actttagatt gattacgcg 5259 ccctgtagcg gcgcattaag cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca 5319 gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg attagtgct 5439 ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg 5499 ccctgataga cggttttcg ccctttgacg ttggagtca cgttctttaa tagtggactc 5559 ttgttccaaa cttggacca cattaggt tagggtca tatctttga tttataaggg 5619 attttgccga tttcggccta ttggttaaaa aatgagctga ttaacaaaa attaacgcg 5679 aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tccttttga 5739 taatcccatg accaaaatcc cttaacgtg gtttcgtc cactgaggct cagacccgt 5799 agaaaagatc aaaggatctt cttgagatcc ttttttcg cgcgtaatct gctgcttgca 579	caatggcaac aacgttgcgc aaactattaa ctggcgaact acttactcta gcttcccggc 4959 aacaattaat agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc 5019 tteeggetgg etggttaat getgataaat etggageegg tgagegtggg tetegeggta 5079 teattgcage actggggeea gatggtaage eeteccegtat egtagttate tacaegaegg 5139 ggagteagge aactatggat gaacgaaata gacagatege tgagataggt geeteaetga 5199 ttaageattg gtaactgtea gaccaagtt acteatata aetttagatt gattaeege 5259 eeetgaage gegeattaag egeggeggt gtggtggtta egegeagegt gaeegetae 5319 geeggettte eeetgaee egeteette getteetee etteettee egeeaegte 5379 geeggettte eeetgaege eeetaaegg gggeteeett tagggtaeg attaagteg 5439 ttaacggeaee tegaeeeaa aaaaettgat ttgggtgatg gtteaegtag tgggeeateg 5439 eeetgataga eggtttteg eeetttage ttggggte attettaa tagtggaeee 5559 ttgtteeaaa ettgaaeaae aeteaaeee ateeegget attetttga tttataaggg 5619 attttgeega ttteggeeta ttggttaaaa aatgagetga ttaaeaaaa attaaegeg 5679 aattttaaea aaatattaae gtttaeaatt taaaaggaet taggtgaaga teeettttga 5739 taateteatg accaaaatee ettaaetg ggtttegte eaetgaeggt eagaeeegg 5739 agaaaagate aaaggatett ettgagteg gtttegte eaetgaeggt eagaeeegg 5739 agaaaagate aaaggatet ettaaetgtg ttgttgee gateaagage taceaetet 5859 aaccaaaaaaa eeeeegtae ettaaetgta ttgttgeeg gateaagage taceaetet 5739	cgaaggagct aaccgctttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt	4839
aacaattaat agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc 5019 ttocggotgg otggtttatt gotgataaat otggagoogg tgagogtggg totogoggta 5079 toattgcago actggggoca gatggtaago octocogtat ogtagttato tacaogaogg 5139 ggagtcaggo aactatggat gaacgaaata gacagatogo tgagataggt gootcaotga 5199 ttaagoattg gtaactgtca gaccaagttt actoatat actttagatt gattacogog 5259 cootgtagog gogoattaag ogoggoggt gtggtggtta ogoggagog gacogotaca 5319 gcoggottto occgtoago ogotootto gottotto ottootto ogoagge gacogotaca 5319 ttaagoatg gogoattaag ogoggoggg gtggtggtgat ogogoaggt gacogotaca 5319 gcoggottto occgtoago ogotootto gottotto ottootto ogoagge gacogotaca 5439 ttacggcaco togaccocaa aaaacttgat ttgggtgatg gttoaogtag tggggocatog 5499 cootgataga oggttttog ocotttgacg ttggagtoa ogttottaa tagtggacto 5559 ttgttocaaa ottgaacaac actoaaccot atotoggot attotttga ttataaggg 5619 attttgocga tttoggocta ttggttaaaa aatgagotga ttaacaaa attaacogo 5679 aattttaaca aaatattaac gtttacaatt taaaaggato taggtgaaga toottttga 5739 taactocatg accaaaatco ottaacgga gttttogtto cactgagogt cagacccogt 5799 agaaaagato aaaggatott ottgagato ttttttga tttatagog 5799	accaattaat agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc5019ttccggctgg ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta5079tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg5139ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga5199ttaagcattg gtaactgtca gaccaagttt actcatatat actttagatt gattaccgcg5259ccctgtagcg gcgcattaag cgcggcgggt gtggtggtta cgcgcagcg gaccgctaca5319gcaggctttc cccgtcaagc cgctccttc gcttctcc cttccttct cgccacgttc5379gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg atttagtgct5439ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg5499ccctgataga cggttttcg ccctttgacg ttggagtca cgttcttaa tagtggaccd5559ttgtccaaa cttgaacaac actcaaccct atccgggct attctttga tttataaggg5619aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tccttttga5739gaaaagatc aaaggatct cttaacgtga gtttcgtc cactgagcgt cagaccccgt5799agaaaagatc aaaggatctt cttggatcc ttttttcg cgcgtaatct gctgcttgca5859aatttaaca aaaact catcaacct tttttcg cactgagcgt cagaccccgt5799aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tccttttga5799agaaaagatc aaaggatctt cttgagtcc ttttttcg cgcgtaatct gctgcttgca5859acttaaaaaaaa ccaccgctac cagcggtgt ttgtttgccg gatcaagagt cacaactc5859acttaaaaaaaa ccaccgctac cagcggtgt ttgtttgccg gatcaagagc taccaactc5859	gggaaccgga gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgtag	4899
ttccggctgg ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta 5079 tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 5139 ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 5199 ttaagcattg gtaactgtca gaccaagtt actcatatat actttagatt gatttacgcg 5259 ccctgtagcg gcgcattaag cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca 5319 ettgccagcg ccctagcgcc cgctccttc gcttcttcc cttccttct cgccacgttc 5379 gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg atttagtgct 5439 ttacggcacc tcgaccccaa aaaacttgat ttgggtgagt gttgagtg gttcacgtag tgggcattcg 5499 ccctgataga cggttttcg ccctttgacg ttggagtca cgttcttaa tagtggactc 5559 ttgttccaaa cttgaacaac actcaaccct atctcgggct attctttga tttataaggg 5619 attttgccga tttcggccta ttggttaaaa aatgagctga ttaacaaaa atttaacgcg 5679 aaatttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tccttttga 5739 taatcccatg accaaaatcc cttaacgtga gtttcgtc cactgagcgt cagaccccgt 5799 agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaatct gctgcttgca 5859	ttccggctgg ctggttatt gctgataat ctggagccgg tgagcgtggg tctcgcggta 5079 tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 5139 ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 5199 ttaagcattg gtaactgtca gaccaagttt actcatatt actttagatt gatttacgcg 5259 ccctgtagcg gcgcattaag cgcggcgggt gtggtgggtta cgcgcagcgt gaccgctaca 5319 cttgccagcg ccctagcgcc cgctccttc gcttctcc cttccttct cgccacgtt 5379 gccggctttc cccgtcaage tctaaatcgg gggctccctt tagggttccg attragtgct 5439 ttacggcacc tcgaccccaa aaaacttgat ttgggtggt gttcacgtag tgggccatcg 5499 ccctgataga cggttttcg ccctttgacg ttggagtca cgttctttaa tagtggactc 5559 ttgttccaaa cttgaacaac actcaaccct atctcgggct attctttga tttataaggg 5619 attttgccga tttcggccta ttggttaaaa aatgagctga tttaacaaaa atttaacgcg 5739 taatctcatg accaaaatcc cttaacgtga gtttcgttc cactgagcgt cagaccccgt 5799 agaaaagatc aaaggatct cttgagatc ttttttctg cgcgtaatc gctgctgca 5859	caatggcaac aacgttgcgc aaactattaa ctggcgaact acttactcta gcttcccggc	4959
tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 5139 ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 5199 ttaagcattg gtaactgtca gaccaagtt actcatata actttagatt gatttacgcg 5259 ccctgtagcg gcgcattaag cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca 5319 cttgccagcg ccctagcgcc cgctccttc gcttcttcc cttccttct cgccacgttc 5379 gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg attagtgct 5439 ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tggggcatcg 5499 ccctgataga cggttttcg ccctttgacg ttggagtca cgttcttaa tagtggactc 5559 ttgttccaaa cttgaacaac actcaaccct atctcgggct attctttga tttataaggg 5619 attttgccga tttcggccta ttggttaaaa aatgagctga tttaacaaaa atttaacgcg 5679 aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tccttttga 5739 taatccatg accaaaatcc cttaacgtga gtttcgtc cactgagcgt cagaccccgt 5799 agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaatct gctgcttgca 5859	tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 5139 ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 5199 ttaagcattg gtaactgtca gaccaagtt actcatata actttagatt gatttacgcg 5259 ccctgtagcg gcgcattaag cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca 5319 cttgccagcg ccctagcgcc cgctccttc gcttcttcc cttccttct cgccacgttc 5379 gccggctttc cccgtcaage tctaaatcgg gggctccctt tagggttccg attagtgct 5439 ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg 5499 ccctgataga cggttttcg ccctttgacg ttggagtca cgttcttaa tagtggactc 5559 ttgttccaaa cttgaacaac actcaaccct atctcgggct attctttga tttataaggg 5619 attttgccga tttcggccta ttggttaaaa aatgagctga ttaacaaaa atttaacgcg 5679 aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tccttttga 5739 taaccaaaatc aaagatct cttgagtg gtttcgttc cactgagcgt cagaccccgt 5799 agaaaagatc aaaggatctt cttgagatce ttttttcg gcgctaatct gctgcttgca 5859	aacaattaat agactggatg gaggoggata aagttgcagg accacttotg cgotoggooo	5019
ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 5199 ttaagcattg gtaactgtca gaccaagtt actcatata actttagatt gatttacgcg 5259 ccctgtageg gcgcattaag cgcggegggt gtggtggtta cgcgcagegt gaccgctaca 5319 cttgccageg ccctagegec egeteette gettettee etteettet egecaegte 5439 gccggette eccegteaage tetaaategg gggeteet tagggtegg gtgggget gtgggget 5439 ttaeggeace tegaeceeaa aaaaettgat ttgggtgatg gtteaegtag tgggeeateg 5439 ccctgataga eggtttteg ecettegaeg ttggagtee egiteettaa tagtggaete 5559 ttgteceaaa ettgaacaac acteaaceet ateeegget atteettaa tagtggaete 5559 ttgtteceaa ettgaacaac acteaaceet ateeegget atteettega ttaaaggg 5619 attetgeega ttteggeeta ttggttaaaa aatgagetga ttaacaaaa attaacgeg 5679 aaatttaaca aaatataac gtttacaatt taaaaggate taggtgaaga teettettga 5739 taateeetag aceaaaatee ettaacgtga gtteegte eaetgaeget eagaeceetg 5739	ggagtcagge aactatggat gaacgaaata gacagatege tgagataggt geeteactga 5199 ttaageattg gtaactgtea gaceaagtt acteatata actttagatt gattaegeg 5259 eeetgageg gegeattaag egeggegggt gtggtggtta egegeagegt gaeegetaea 5319 ettgeeageg eeetagegee egeteette getteetee etteettee egeeaget 5439 ttaeggeace tegaceeaa aaaaettgat ttgggtgatg gtteaegtag tgggeeateg 5439 eeetgataga eggtttteg eeettgaeg ttggagteea egteettaa tagtggaete 5559 eeetgataga eggtttteg eeettgaeg ttggagteea egteettaa tagtggaete 5559 ettgeeaga ettegaeaaa acteaaeeet ateegget atteettega ttaaaggg 5619 attetgeega tteeggeeta ttggttaaaa aatgagetga ttaaeaaa attaaegeg 5679 aaatttaaea aaatataae gtttaeaatt taaaaggate taggtgaaga teettttga 5739 taateeetag aceaaaatee ettaaegtga gttteegte eaetgaeget eagaeeegt 5799 agaaaagate aaaggatett ettgagatee tttttteetg egegtaatet geegettgea 5859	ttccggctgg ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta	5079
ttaagcattg gtaactgtca gaccaagtt actcatata actttagatt gatttacgcg5259ccctgtagcg gcgcattaag cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca5319cttgccagcg ccctagcgcc cgctccttc gcttctcc cttccttct cgccacgtt5379gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg atttagtgct5439ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg5499ccctgataga cggttttcg ccctttgacg ttggagtca cgttcttaa tagtggacct5559ttgttccaaa cttgaacaac actcaaccct atctcgggct attctttga tttataaggg5619attttgccga tttcggccta ttggttaaa aatgagctga ttaacaaa atttaacgcg5679taatctcatg accaaaatce cttaacgtga gtttcgtc cactgagcgt cagaccccgt5739taatctcatg accaaaatce cttaacgtga gtttcgtc cactgagcgt cagaccccg5799agaaaagatc aaaggatctt cttgagatce ttttttcg cgcgtaatct gctgcttgca5859	ttaagcattg gtaactgtca gaccaagtt actcatata actttagatt gatttacgog5259ccctgtagcg gcgcattaag cgcggcggg gtggtggtgat cgcgcagcgt gaccgctaca5319cttgccagcg ccctagcgcc cgctccttte gcttcttee etteettee egecacgtte5379gccggcttte eccgtcaage tetaaategg gggeteeett tagggtteeg attagtget5439ttaeggcace tegaceccaa aaaacttgat ttgggtgatg gtteaegtag tggggecateg5499ecctgataga eggtttteeg ecettegaeg ttggagteea egttettaa tagtggaete5559ttgtteeaaa ettgaacaae acteaaceet ateteggget attetttga ttataaggg5619aattttaaca aaatattaae gtttacaatt taaaggaete taggtgaaga teettttga5739aaateteaag accaaaatee ettaaegtg gttteegte cactgagegt eagacecegt5799agaaaagate aaaggatett ettgagatee tttttteeg egegtaatet getgettgea5859aacaaaaaaa ecacegetae eageggtgt ttgtttgeeg gateaagage taccaactee5819	tcattgcage actgggggcca gatggtaage eetceegtat egtagttate tacaegaegg	5139
ccctgtagcg gcgcattaag cgcggcggt gtggtggtta cgcgcagcgt gaccgctaca5319cttgccagcg ccctagcgcc cgctccttc gcttcttcc cttccttct cgccacgttc5379gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg atttagtgct5439ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg5499ccctgataga cggttttcg ccctttgacg ttggagtcca cgttcttaa tagtggactc5559ttgttccaaa cttgaacaac actcaaccct atctcgggct attctttga tttaaagg5619attttgccga tttcggccta ttggttaaaa aatgagctga ttaacaaa atttaacgcg5739taatctcatg accaaaatce cttaacgtga gtttcgttc cactgagcgt cagaccccgt5799agaaaagatc aaaggatctt cttgagatce tttttttga cgcgtaatct gctgcttgca5799	ccctgtageg gegeattaag egeggeggt gtggtggtta egegeaget gaeegetae5319cttgeeageg eeetagegee egeteette getteetee etteetteet egeeaegte5439geeggettee eeegeeaegee tetaaategg gggeteeett tagggtteeg attagtget5439ttaeggeaee tegaeeeaa aaaaettgat ttgggtgatg gteaegtag tgggeeateg5499ccctgataga eggtttteeg eeettgaeg ttggagteea egteettaa tagtggaete5559ttgteeaaa ettgaeeaa aaeaettgat tagtgget attettega ttataaggg5619attttgeega ttteggeeta ttggttaaaa aatgagetga ttaaeaaa attaaeegeg5679aatttaaca aaatattaae gtttaeaatt taaaaggate taggtgaaga teettttega5739agaaaagate aaaggatet ettggagtee ttttttee eeetgaeget eagaeeege5859aaaaaaaa eeaeegetae eageggtggt ttgtttgeeg gateaagage taceaaetee5819	ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga	5199
cttgccagcg ccctagcgcc cgctccttc gcttcttcc cttccttct cgccacgttc5379gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg atttagtgct5439ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg5499ccctgataga cggttttcg ccctttgacg ttggagtcca cgttcttaa tagtggactc5559ttgttccaaa cttgaacaac actcaaccct atctcgggct attctttga tttataaggg5619attttgccga tttcggccta ttggttaaaa aatgagctga ttaacaaa atttaacgcg5679aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tccttttga5739taatctcatg accaaaatcc cttaacgtga gtttcgtc cactgagcgt cagaccccgt5799agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaatct gctgcttgca5859	cttgccagcgccctagcgcccgctcctttcgcttcttcccttcctttctcgccagcttc5379gccggctttccccgtcaagctctaaatcgggggctccctttagggttccgattagtgct5439ttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcg5499ccctgatagacggttttcgccctttgacgttggggtgatggttcacgtagtggggccatcg5559ttgttccaaacttgaacaacactcaaccctattctgggctattataaggg5619attttgccgatttcggcctattggttaaaaaatgagctgattaacaaaaattaacgg5679aatttaacaaaaaaaaagtttacaatttaaaaggatctaggtgaagatccttttga5739taatctcatgaccaaaatcccttaacgtgagttttcgtccacgagcgt cagaccccgt5799agaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctcgcgtcaacc5859aacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactct5919	ttaagcattg gtaactgtca gaccaagttt actcatatat actttagatt gatttacgcg	5259
gccggctttccccgtcaagctctaaatcgggggctccctttagggttccgattagtgct5439ttacggcacctcgaccccaaaaaacttgatttgggtgatggtcacgtagtgggccatcg5499ccctgatagacggttttcgccctttgacgttggggtgatggtcacgtagtggggccatcg5559ttgttccaaacttgaacaacactcaaccctattctgggctattctttgatttaaaggg5619attttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaattaacgcg5679aattttaacaaaatattaacgtttacaatttaaaaggatctaggtgaagatcctttttga5739taatctcatgaccaaaatcccttaacgtgagtttcgtccactgagcgt5799agaaaagatcaaaggatcttcttgagatccttttttctg5859	gccggctttccccgtcaagctctaaatcgggggctccctttagggttccgattagtgct5439ttacggcacctcgaccccaaaaaacttgatttgggtgatggtcacgtagtgggccatcg5499ccctgatagacggttttcgccctttgacgttgggtgatggtcacgtagtggggccatcg5559ttgttccaaacttgaacaacactcaaccctattctgggctattcttttgattataaggg5619attttgccgatttcggcctattggttaaaaaatgagctgattataacgcg5679aattttaacaaaatattaacgtttacaatttaaaaggatctaggtgaagatcctttttga5739taatctcatgaccaaaatcccttaacgtgagtttcgttccactgagcgt5859aacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagc5819	ccctgtagcg gcgcattaag cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca	5319
ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg5499ccctgataga cggttttcg cccttgacg ttggagtcca cgttcttaa tagtggactc5559ttgttccaaa cttgaacaac actcaaccct atctcgggct attctttga tttataaggg5619attttgccga tttcggccta ttggttaaaa aatgagctga tttaacaaaa atttaacgcg5679aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tccttttga5739taatctcatg accaaaatcc cttaacgtga gtttcgtc cactgagcgt cagaccccgt5799agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaatct gctgcttgca5859	ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg5499ccctgataga cggttttcg cccttgacg ttggagtcca cgttctttaa tagtggactc5559ttgttccaaa cttgaacaac actcaaccct atctcgggct attcttttga tttataaggg5619attttgccga tttcggccta ttggttaaaa aatgagctga tttaacaaaa atttaacgcg5679aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tcctttttga5739taatctcatg accaaaatcc cttaacgtga gtttcgttc cactgagcgt cagaccccgt5799agaaaagatc aaaggatctt cttggatcc ttttttcg cgcgtaatct gctgcttgca5859aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct5919	cttgccagcg ccctagcgcc cgctcctttc gctttcttcc cttcctttct cgccacgttc	5379
ccctgataga cggtttttcg ccctttgacg ttggagtcca cgttctttaa tagtggactc5559ttgttccaaa cttgaacaac actcaaccct atctcgggct attcttttga tttataaggg5619attttgccga tttcggccta ttggttaaaa aatgagctga tttaacaaaa atttaacgcg5679aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tcctttttga5739taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt5799agaaaagatc aaaggatctt cttgagatcc ttttttcg cgcgtaatct gctgcttgca5859	ccctgataga cggtttttcg ccctttgacg ttggagtcca cgttctttaa tagtggactc5559ttgttccaaa cttgaacaac actcaaccct atctcgggct attcttttga tttataaggg5619attttgccga tttcggccta ttggttaaaa aatgagctga tttaacaaaa atttaacgcg5679aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tcctttttga5739taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt5799agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaatct gctgcttgca5859aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct5919	gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg atttagtgct	5439
ttgttccaaa ettgaacaac acteaaceet ateteggget attetttga tttataaggg 5619 attetgeega ttteggeeta ttggttaaaa aatgagetga tttaacaaaa atttaaegeg 5679 aattetaaca aaatattaae gettacaatt taaaaggate taggtgaaga teetttttga 5739 taateteatg accaaaatee ettaaegtga getttegtte eaetgagegt eagaeeeeg 5799 agaaaagate aaaggatett ettgagatee ttettttetg egegtaatet getgettgea 5859	ttgttccaaa cttgaacaac actcaaccct atctcgggct attctttga tttataaggg 5619 attttgccga tttcggccta ttggttaaaa aatgagctga tttaacaaaa atttaacgcg 5679 aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tcctttttga 5739 taatctcatg accaaaatcc cttaacgtga gtttcgttc cactgagcgt cagaccccgt 5799 agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaatct gctgcttgca 5859 aaccaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct 5919	ttacggcacc tcgaccccaa aaaacttgat ttgggtgatg gttcacgtag tgggccatcg	5499
attttgocga tttcggocta ttggttaaaa aatgagotga tttaacaaaa atttaacgog 5679 aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tootttttga 5739 taatotoatg accaaaatoo ottaacgtga gtttogtto caotgagogt cagacocogt 5799 agaaaagato aaaggatott ottgagatoo tttttttotg ogogtaatot gotgottgoa 5859	attttgocga tttoggocta ttggttaaaa aatgagotga tttaacaaaa atttaacgog 5679 aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tootttttga 5739 taatotocatg accaaaatoo ottaacgtga gttttogtto cactgagogt cagacocogt 5799 agaaaagato aaaggatott ottgagatoo tttttttotg ogogtaatot gotgottgoa 5859 aacaaaaaaa ocacogotac cagoggtggt ttgtttgoog gatcaagago taccaactot 5919	ccctgataga cggtttttcg ccctttgacg ttggagtcca cgttctttaa tagtggactc	5559
aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tcctttttga 5739 taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt 5799 agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 5859	aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tcctttttga 5739 taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt 5799 agaaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 5859 aaccaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct 5919	ttgttccaaa cttgaacaac actcaaccct atctcgggct attcttttga tttataaggg	5619
taateteatg accaaaatee ettaacgtga gttttegtte eaetgagegt eagaeeeegt 5799 agaaaagate aaaggatett ettgagatee tttttttetg egegtaatet getgettgea 5859	taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt 5799 agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 5859 aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct 5919	attttgccga tttcggccta ttggttaaaa aatgagctga tttaacaaaa atttaacgcg	5679
agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 5859	agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 5859 aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct 5919	aattttaaca aaatattaac gtttacaatt taaaaggatc taggtgaaga tcctttttga	5739
	aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct 5919	taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt	5799
aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct 5919		agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca	5859
	ttttccqaaq qtaactqqct tcaqcaqaqc qcaqatacca aatactqtcc ttctaqtqta 5979	aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct	5919
ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgta 5979		ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgta	5979

geegtagtta ggeeaceaet teaagaaete tgtageaeeg eetaeataee tegetetget	6039
aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttggactc	6099
aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggtt cgtgcacaca	6159
gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga	6219
aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg	6279
aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt atagtcctgt	6339
cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag	6399
cctatggaaa aacgccagca acgcggcctt tttacggttc ctggcctttt gctggccttt	6459
tgeteacatg ttettteetg egttateeee tgattetgtg gataacegta ttacegeett	6519
tgagtgaget gatacegete geogeageeg aacgaeegag egeagegagt eagtgagega	6579
ggaagcggaa gagcgcctga tgcggtattt tctccttacg catctgtgcg gtatttcaca	6639
ccgcataggg tcatggctgc gccccgacac ccgccaacac ccgctgacgc gccctgacgg	6699
gettgtetge teceggeate egettaeaga eaagetgtga eegteteegg gagetgeatg	6759
tgtcagaggt tttcaccgtc atcaccgaaa cgcgcgaggc agcaaggaga tggcgcccaa	6819
cagtcccccg gccacggggc ctgccaccat acccacgccg aaacaagcgc tcatgagccc	6879
gaagtggcga gcccgatctt ccccatcggt gatgtcggcg atataggcgc cagcaaccgc	6939
acctgtggcg ccggtgatgc cggccacgat gcgtccggcg tagaggatct gctcatgttt	6999
gacagettat e	7010
<pre><210> SEQ ID NO 12 <211> LENGTH: 226 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <221> NAME/KEY: DOMAIN <222> LOCATION: (1)(226) <223> OTHER INFORMATION: Red-alpha from plasmid pBAD-alpha-beta-game </pre>	nma
<400> SEQUENCE: 12	
Met Thr Pro Asp Ile Ile Leu Gln Arg Thr Gly Ile Asp Val Arg Ala 1 5 10 15	
Val Glu Gln Gly Asp Asp Ala Trp His Lys Leu Arg Leu Gly Val Ile 20 25 30	
Thr Ala Ser Glu Val His Asn Val Ile Ala Lys Pro Arg Ser Gly Lys	
35 40 45	
Lys Trp Pro Asp Met Lys Met Ser Tyr Phe His Thr Leu Leu Ala Glu 50 55 60	
Val Cys Thr Gly Val Ala Pro Glu Val Asn Ala Lys Ala Leu Ala Trp 65 70 75 80	
Gly Lys Gln Tyr Glu Asn Asp Ala Arg Thr Leu Phe Glu Phe Thr Ser 85 90 95	
Gly Val Asn Val Thr Glu Ser Pro Ile Ile Tyr Arg Asp Glu Ser Met 100 105 110	
Arg Thr Ala Cys Ser Pro Asp Gly Leu Cys Ser Asp Gly Asn Gly Leu 115 120 125	
Glu Leu Lys Cys Pro Phe Thr Ser Arg Asp Phe Met Lys Phe Arg Leu 130 135 140	
Gly Gly Phe Glu Ala Ile Lys Ser Ala Tyr Met Ala Gln Val Gln Tyr	

-continued

Ser Met Trp Val Thr Arg Lys Asn Ala Trp Tyr Phe Ala Asn Tyr Asp Pro Arg Met Lys Arg Glu Gly Leu His Tyr Val Val Ile Glu Arg Asp Glu Lys Tyr Met Ala Ser Phe Asp Glu Ile Val Pro Glu Phe Ile Glu Lys Met Asp Glu Ala Leu Ala Glu Ile Gly Phe Val Phe Gly Glu Gln Trp Arg <210> SEQ ID NO 13 <211> LENGTH: 261 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <221> NAME/KEY: DOMAIN <222> LOCATION: (1)..(261) <223> OTHER INFORMATION: Red-beta from plasmid pBAD-alpha-beta-gamma <400> SEQUENCE: 13 Met Ser Thr Ala Leu Ala Thr Leu Ala Gly Lys Leu Ala Glu Arg Val Gly Met Asp Ser Val Asp Pro Gln Glu Leu Ile Thr Thr Leu Arg Gln Thr Ala Phe Lys Gly Asp Ala Ser Asp Ala Gln Phe Ile Ala Leu Leu Ile Val Ala Asn Gln Tyr Gly Leu Asn Pro Trp Thr Lys Glu Ile Tyr Ala Phe Pro Asp Lys Gln Asn Gly Ile Val Pro Val Val Gly Val Asp Gly Trp Ser Arg Ile Ile As
n Glu As
n Gl
n Gl
n Phe As
p Gly Met As
p $% \left({{\mathbb{F}}_{{\mathbb{F}}}} \right)$ Phe Glu Gln Asp Asn Glu Ser Cys Thr Cys Arg Ile Tyr Arg Lys Asp Arg Asn His Pro Ile Cys Val Thr Glu Trp Met Asp Glu Cys Arg Arg Glu Pro Phe Lys Thr Arg Glu Gly Arg Glu Ile Thr Gly Pro Trp Gln Ser His Pro Lys Arg Met Leu Arg His Lys Ala Met Ile Gln Cys Ala Arg Leu Ala Phe Gly Phe Ala Gly Ile Tyr Asp Lys Asp Glu Ala Glu Arg Ile Val Glu Asn Thr Ala Tyr Thr Ala Glu Arg Gln Pro Glu Arg Asp Ile Thr Pro Val Asn Asp Glu Thr Met Gln Glu Ile Asn Thr Leu Leu Ile Ala Leu Asp Lys Thr Trp Asp Asp Asp Leu Leu Pro Leu Cys Ser Gln Ile Phe Arg Arg Asp Ile Arg Ala Ser Ser Glu Leu Thr Gln Ala Glu Ala Val Lys Ala Leu Gly Phe Leu Lys Gln Lys Ala Ala Glu Gln Lys Val Ala Ala

			-	
- C	ont	1 11	ued	

50

<211 <212 <213 <220 <221 <222	L> LH 2> T 3> OH 0> FH L> NH 2> LO 3> OT	EATUR AME/R DCATI FHER	H: 13 PRT ISM: RE: KEY: ION: INF(38 Art: DOM2 (1)	AIN (13 FION	38) : Rea	- d-gar		Erom	plas	smid	pBAI	D-alı	pha-1	oeta-gamma
<400)> SI	EQUEI	ICE :	14											
Met 1	Asp	Ile	Asn	Thr 5	Glu	Thr	Glu	Ile	Lys 10	Gln	Lys	His	Ser	Leu 15	Thr
Pro	Phe	Pro	Val 20	Phe	Leu	Ile	Ser	Pro 25	Ala	Phe	Arg	Gly	Arg 30	Tyr	Phe
His	Ser	Tyr 35	Phe	Arg	Ser	Ser	Ala 40	Met	Asn	Ala	Tyr	Tyr 45	Ile	Gln	Asp
Arg	Leu 50	Glu	Ala	Gln	Ser	Trp 55	Ala	Arg	His	Tyr	Gln 60	Gln	Leu	Ala	Arg
Glu 65	Glu	Lys	Glu	Ala	Glu 70	Leu	Ala	Asp	Asp	Met 75	Glu	Lys	Gly	Leu	Pro 80
Gln	His	Leu	Phe	Glu 85	Ser	Leu	Суз	Ile	Asp 90	His	Leu	Gln	Arg	His 95	Gly
Ala	Ser	Lys	Lys 100	Ser	Ile	Thr	Arg	Ala 105	Phe	Asp	Asp	Asp	Val 110	Glu	Phe
Gln	Glu	Arg 115	Met	Ala	Glu	His	Ile 120	Arg	Tyr	Met	Val	Glu 125	Thr	Ile	Ala
His	His 130	Gln	Val	Asp	Ile	Asp 135	Ser	Glu	Val						

The invention claimed is:

1. A method for cloning DNA molecules in procaryotic cells comprising the steps of:

- a) providing a procaryotic host cell capable of performing $_{40}$ RecET homologous recombination, wherein the host cell expresses red α and red β genes,
- b) contacting in said host cell a circular first DNA molecule which is capable of being replicated in said host cell with a linear second DNA molecule comprising at least two 45 regions of sequence homology to regions on the first DNA molecule and further comprising a DNA fragment to be cloned under conditions which favour homologous recombination between said first and second DNA molecules and
- c) selecting a host cell in which homologous recombination between said first and second DNA molecules has occurred,
- wherein a second DNA molecule is introduced into the host cell in a form which allows recombination without fur- 55 ther modification; and
- wherein when said RecET homologous recombination occurs, it is mediated by gene products of said red α and redβ genes.

2. The method according to claim **1** wherein the host cell is $_{60}$ transformed with at least one vector capable of expressing red α and/or red β genes.

3. The method of claim 1 wherein the expression of the red α and/or red β genes is under control of a regulatable promoter.

4. The method of claim 2 wherein the red β gene is overexpressed versus the red α gene.

5. The method according to claim 1 wherein the host cell is a gram-negative bacterial cell.

6. The method according to claim 5 wherein the host cell is an Escherichia coli cell.

7. The method according to claim 6 wherein the host cell is an Escherichia coli K12 strain.

8. The method according to claim 7 wherein the E. coli strain is selected from JC 8679 and JC 9604.

9. The method according to claim 1 wherein the host cell further expresses a redy inhibitor gene.

10. The method according to claim 9 wherein the host cell is transformed with a vector expressing the redy inhibitor gene.

11. The method according to claim 1 wherein the first DNA molecule is an extra chromosomal DNA molecule containing an origin of replication which is operative in the host cell.

12. The method according to claim 11 wherein the first DNA molecule is selected from plasmids, cosmids, P1 vectors, BAC vectors and PAC vectors.

13. The method according to claim 1 wherein the first DNA molecule is a host cell chromosome.

14. The method according to claim 1 wherein the regions of sequence homology are at least 15 nucleotides each.

15. The method according to claim 1 wherein the second DNA molecule is obtained by an amplification reaction.

16. The method according claim 1 wherein the first and/or second DNA molecules are introduced into the host cells by 65 transformation.

17. The method according to claim 16 wherein the transformation method is electroporation.

18. The method according to claim **1** wherein the first and second DNA molecules are introduced into the host cell simultaneously by co-transformation.

19. The method according to claim **1** wherein the second DNA molecule is introduced into a host cell in which the first 5 DNA molecule is already present.

20. The method according to claim **1** wherein the second DNA molecule contains at least one marker gene placed between the two regions of sequence homology and wherein homologous recombination is detected by expression of said 10 marker gene.

21. The method according to claim 20 wherein the marker gene is selected from antibiotic resistance genes, deficiency complementation genes and reporter genes.

22. The method of claim 1 wherein the first DNA molecule contains at least one marker gene between the two regions of sequence homology and wherein homologous recombination is detected by lack of expression of said marker gene.

23. The method of any claim 1 wherein said marker gene is selected from genes which, under selected conditions, convey a toxic or bacteriostatic effect on the cell, and reporter genes.

24. A method according to claim 1 wherein the first DNA molecule contains at least one target site for a site specific recombinase between the two regions of sequence homology and wherein homologous recombination is detected by removal of said target site.

* * * * *