ACEMBL Expression System Series

*Multi*Coli

Multi-Protein Expression for *E.coli*

User Manual

Vers. 2.0 June 2011 This manual is based on the original ACEMBL manual (November 2009) written by Yan Nie, Christoph Bieniossek and Imre Berger but has been revised, updated and, wherever necessary, modified and expanded to meet customer demands.

ACEMBL was developed at the European Molecular Biology Laboratory, EMBL Grenoble Outstation, Grenoble, France.

Table of Contents	
A. The ACEMBL System Kit: Contents and Storage	3
B. Introduction	4
C. Synopsis	5
C. The ACEMBL System	6
C.1. ACEMBL vectors C.2. The multiple integration element (MIE) C.3. Tags, promoters, terminators C.4. Generating Plasmid Constructs for Complex Expression C.5. Complex Expression	6 7 8 9 10
D. Procedures	12
 D.1. Cloning into ACEMBL vectors D.1.1. Single gene insertion into the MIE by SLIC D.1.2. Polycistron assembly in MIE by SLIC D.1.3. Gene insertion by restriction/ligation D.1.4. Multiplication by using the HE and BstXI sites D.2. Cre-LoxP reaction of Acceptors and Donors D.2.1. Cre-LoxP fusion of Acceptors and Donors D.2.2. Deconstruction of fusion vectors by Cre D.3. Co-expression by Co-transformation E. ACEMBL multi-gene combination: Examples E.1. SLIC cloning into ACEMBL vectors: human TELE 	12 17 22 25 28 30 33 35 37
E.2. The Homing endonuclease/BstXI module: yeast RES complex	38
F. Appendix	39
 F.1. DNA sequence of MIE F.2. DNA sequences of ACEMBL vectors F.2.1. pACE1 F.2.2. pACE2 F.2.3. pDC F.2.4. pDK F.2.5. pDS F.3. ACEMBL plasmid maps F.4. Analytical restriction of ACEMBL vectors 	39 40 41 42 43 44 45 46 48
G. References	49

A. The ACEMBL System Kit: Contents and Storage

Reagents supplied in ACEMBL system kit:

Acceptor vectors: pACE1, pACE2 – approx. 2 µg plasmid DNA per vial

Donor vectors: pDC, pDK, pDS – approx. 2 µg plasmid DNA per vial

keep at 4°C for short-term storage and in a freezer at -20°C or lower for medium- and long-term storage (take care to avoid repeated freeze-thaw cycles, e.g. by aliquotting DNA prior to freezing)

E. coli transformed with Acceptor and Donor vectors, provided as agar stabs; for plating bacteria as a starting point for plasmid preparations

keep agar stabs at 4°C or room temperature

pirLC, pirHC cells[†]

[†] *E. coli* strains expressing the *pir* gene product for propagation of donor vectors (any other strain with *pir*⁺ background can be used as well). LC: low copy number propagation, HC: high copy number propagation of plasmids with R6K γ origin.

keep agar stabs at 4°C or room temperature

Additionally required reagents:

Antibiotics: ampicillin, chloramphenicol, kanamycin, spectinomycin, tetracycline

Enzymes: Cre recombinase

T4 DNA polymerase (for recombination insertion of genes) Phusion polymerase or any other proof-reading DNA polymerase (for

PCR amplification of DNA)

Restriction enzymes and T4 DNA ligase (for restriction-ligation cloning)

Standard laboratory *E.coli* strain for cloning (TOP10, HB101, DH5 α)

Expression strain(s) of choice, e.g. BL21(DE3), Rosetta, AD494, Origami (DE3), etc.

Choice of strain will depend on your target protein and its underlying DNA sequence.

B. Introduction

Protein complexes are the heart and soul of many cellular processes¹. Some researchers go as far as describing the cell as "a collection of protein machines"². Whether you think of replication, transcription³, translation⁴, DNA repair, the processing, import, trafficking as well as export of proteins or other biomolecules, or the maintenance of the structural stability and integrity of any cell, multi-subunit protein assemblies play an important role in all these biological phenomena.

In addition, other processes, e.g. entry of viruses into human cells, also critically hinge on multiple proteins or protein complexes⁵. Moreover, various prokaryotic microorganisms, with *E.coli* being the prototypical workhorse, are harnessed to express heterologous proteins and protein complexes but also to cost-efficiently produce known or novel compounds by means of metabolic engineering⁶.

Scientists wishing to study these processes in functional and structural detail, often require significant amounts of the protein complexes under investigation. While obtaining bulk protein usually is not a problem for protein complexes that are abundant in a steady-state cell, this becomes more difficult for complexes that are transient in nature, appear only periodically in cells or simply occur only in low abundance. In such cases, systems come in handy that allow homo- or heterologous expression of these complexes in large amounts.

While various methods and systems have been developed to address this problem, most of them are of little use for intense research efforts directed at generating and investigating scores of protein complexes in parallel, i.e. in an automated fashion. Such a system should be robust and easy-to-install in terms of manipulation steps / protocols and/or components used in the process⁷. The ACEMBL system exactly addresses these needs.

¹ Robinson et al., Nature **450**, 973 (2007); Charbonnier S et al., Biotechnol Annu Rev **14**, 1 (2008)

² Alberts, Cell **92**, 291 (1998).

³ Van Hijum et al., Microbiol Mol Biol Rev **73**, 481 (2009).

⁴ Estrozi et al., Nat Struct Mol Biol 18, 88

⁵ Bhattacharya, Nature **459**, 24 (2009).

⁶ Chemler and Koffas, Curr Opin Biotech **19**, 597 (2008); Chou, Appl Microbiol Biotech **76**, 521 (2007); Lee et al., Curr Opin Biotech **19**, 553 (2008).

⁷ Nie et al., Curr Genomics **10**: 558-72 (2009).

C. Synopsis

ACEMBL *Multi*Coli is a 3rd generation multi-gene expression system for complex production in *E. coli*, created at the European Molecular Biology Laboratory EMBL, at Grenoble. ACEMBL can be applied both manually and also in an automated set-up by using a liquid handling workstation. ACEMBL applies tandem recombination steps for rapidly assembling many genes into multi-gene expression cassettes. These can be single or polycistronic expression modules, or a combination of these elements. ACEMBL also offers the option to employ conventional approaches involving restriction enzymes and ligases if desired, which may be the methods of choice in laboratories not familiar with recombination approaches.

The following strategies for multi-gene assembly and expression are provided for in the ACEMBL system and detailed in Sections C and D:

- (1) Single gene insertions into vectors (recombination or restriction/ligation)
- (2) Multi-gene assembly into a polycistron (recombination or restriction/ligation)
- (3) Multi-gene assembly using homing endonucleases
- (4) Multi-gene plasmid fusion by Cre-LoxP reaction
- (5) Multi-gene expression by cotransformation

These strategies can be used individually or in conjunction, depending on the project and user.

In Section D, step-by-step protocols are provided for each of the methods for multi-gene cassette assembly that can be applied in the ACEMBL *Multi*Coli system. Each procedure is illustrated by corresponding complex expression experiments in Section D of this Supplement.

DNA sequences of ACEMBL vectors are provided in the Appendix and can be copied from there for further use.

C. The ACEMBL System

C.1. ACEMBL vectors

At the core of the technology are five small *de novo* designed vectors which are called "Acceptor" and "Donor" vectors (see Illustration 1). Acceptor vectors (pACE1, pACE2) contain origins of replication derived from ColE1 (low to medium copy) and resistance markers (ampicillin or tetracycline). Donor vectors contain conditional origins of replication (derived from phage R6Kγ), which make their propagation dependent on hosts expressing the *pir* gene. Donor vectors contain resistance markers kanamycin, chloramphenicol, or spectinomycin. Up to three Donor vectors can be used in conjunction with one Acceptor vector.



Illustration 1: ACEMBL system for multi-protein complex production

All Donor and Acceptor vectors contain a loxP imperfect inverted repeat and in addition, a multiple integration element (MIE). This MIE consists of an expression cassette with a promoter of choice (prokaryotic, mammalian, insect cell specific or a combination thereof, depending on the ACEMBL system) and a corresponding terminator (if required, e.g. the lac promoter does not require a matching terminator). These flank a DNA segment that contains a number of restriction sites which can be used for conventional cloning approaches or also for generating double-strand breaks for the integration of expression elements of choice (further promoters, ribosomal binding sites, terminators and genes). The MIE is completed by a homing endonuclease site and a specifically designed restriction enzyme site (BstXI) flanking the promoter and the terminator (see C.2.) Vector DNA sequences are provided in the Appendix. Maps of all vectors are shown at the end of this manual.



C.2. The multiple integration element (MIE)

The MIE was derived from a polylinker⁸ and allows several approaches for multigene assembly (Section D). Multiple genes can be inserted into the MIE of any one of the vectors by a variety of methods, for example BD-In-Fusion recombination⁹ or SLIC (sequence and ligation independent cloning)¹⁰. For this, the vector needs to be linearized, which can also be carried out efficiently by PCR reaction with appropriate primers, since the vectors are all small (2 to 3 kb). Use of ultrahigh-fidelity polymerases such as Phusion¹¹ is recommended. Alternatively, if more conventional approaches are preferred, i.e. in a regular wet lab setting without robotics, the vectors can also be linearized by restriction digestion, and a gene of interest can be integrated by restriction / ligation (Section D). The DNA sequence of the MIE is shown in the Appendix.

⁸ Tan et al. *Protein Expr. Purif.* **40**, 385 (2005)

⁹ ClonTech TaKaRa Bio Europe, www.clontech.com

¹⁰ Li and Elledge, *Nat. Methods* **4**, 251 (2007)

¹¹ Finnzymes/New England BioLabs, www.neb.com

C.3. Tags, promoters, terminators

Current vectors of the ACEMBL system for *Escherichia coli* contain the default promoters T7 and Lac, as well as the T7 terminator element (Illustr.1, 10). The T7 system is currently most commonly used; it requires bacterial strains which contain a T7 polymerase gene in the *E. coli* genome. The Lac promoter is a strong endogenous promoter which can be utilized in most strains. All ACEMBL vectors contain the lac operator element for repression of heterologous expression.

Evidently, all promoters and terminators present in ACEMBL Donor and Acceptor vectors, and in fact the entire multiple integration element (MIE) can be exchanged with a favored expression cassette by using restriction/ligation cloning with appropriate enzymes (for example ClaI/PmeI, Illustration 2) or insertion into linearized ACEMBL vectors where the MIE was removed by sequence and ligation independent approaches such as SLIC (sequence and ligation independent cloning). In an experimental variation the T7 promoter in pDC was substituted with a trc promoter (pDC^{trc}) and the T7 promoter in pACE with an arabinose promoter (pACE^{ara}). The resulting vectors were used successfully in co-expression experiments by inducing with arabinose and IPTG.

Currently, the ACEMBL system vectors do not contain DNA sequences encoding for affinity tags that enable purification or solubilization of the protein(s) of interest. We typically use C- or N-terminal oligohistidine tags, with or without protease sites for tag removal. We introduce these by means of the respective PCR primers used for amplification of the genes of interest prior to SLIC mediated insertion. We recommend outfitting Donors or Acceptors of choice by the array of custom tags that are favored in individual user laboratories prior to inserting recombinant genes of interest. This is best done by a design which will, after tag insertion, still be compatible with the recombination based principles of ACEMBL system usage.

8

C.4. Generating Plasmid Constructs for Complex Expression

To create your expression constructs (see illustration 3), introduce your gene or genes of interest - carrying any additional modifications such as purification or reporter tags - using your method of choice (conventional restriction-ligation cloning or SLIC) into any of the acceptor or donor vectors. You can then create acceptor-donor fusions with the help of Cre recombinase. Note that you need at least one acceptor vector if you wish to amplify the multi-gene constructs in standard laboratory strains. Select your multi-vector multi-gene fusions by subjecting transformed bacteria to multiple antibiotic selection on agar and/or multi-well plates. You will then have to extract the plasmid construct from your host strains since the expression strain will most likely be different (see chapter C.5).



Illustration 3: Schematic representation of process for generating multi-gene expression constructs.

If, for example, your requirements for an antibiotic resistance marker change, you can transfer entire expression cassettes (including promoters and terminators) from acceptor to acceptor or donor to donor by employing the homing endonuclease-BstXI module. Note that you cannot move cassettes from acceptors to donors or

vice versa since their homing endonuclease and BstXI recognition sites are incompatible.

C.5. Complex Expression

For expression in *E.coli*, the ACEMBL multi-gene expression vector fusions with appropriate promoters or terminators are transformed into the appropriate expression host of choice. In the current version (T7 and lac promoter elements), most of the wide array of currently available expression strains can be utilized. If particular expression strains already contain helper plasmids with DNA encoding for chaperones, lysozyme or other factors of interest, the design of the multi-gene fusion should ideally be such that the ACEMBL vector containing the resistance marker that is also present on the helper plasmid is not included in multi-gene vector construction to avoid issues with plasmid incompatibility (although this is probably not essential).

Alternatively, the issue can be resolved by creating new versions of the ACEMBL vectors containing resistance markers that circumvent the conflict. This can be easily performed by PCR amplifying the vectors minus the resistance marker, and combine the resulting fragments with a PCR amplified resistance marker by recombination (SLIC) or blunt-end ligation (using 5'phosphorylated primers). Note that resistance markers can also be exchanged in between ACEMBL vectors by restriction digestion with AlwNI and ClaI (for Donors) and AlwNI and PmeI (for Acceptors).

Donor vectors depend on the *pir* gene product expressed by the host, due to the R6Ky conditional origin of replication. In regular expression strains, they rely on fusion with an Acceptor for productive replication. Donors or Donor-Donor fusions can nonetheless be used even for expression when not fused with an Acceptor, by using expression strains carrying a genomic insertion of the *pir* gene. Such strains have recently become available (e.g. from Novagen Inc., Madison WI, USA).

Co-transformation of two plasmids can also lead to successful protein complex expression. The ACEMBL system contains two Acceptor vectors, pACE and

pACE2, which are identical except for the resistance marker (Illustration 1). Therefore, genes present on pACE1 or pACE2, respectively, can be expressed by co-transformation of the two plasmids and subsequent simultaneous exposure to tetracyclin and ampicillin. In fact, entire Acceptor-Donor fusions containing several genes, based on pACE1 or pACE2 as Acceptors, can in principle be co-transformed for multi-expression, if needed.

D. Procedures

D.1. Cloning into ACEMBL vectors

All Donors and Acceptors contain an identical MIE with exception of the homing endonuclease site / BstXI tandem than flanks the MIE (Illustrations 1 and 3, plasmid maps in the appendix). The MIE is tailored for sequence and ligation independent gene insertion methods. In addition, the MIE also contains a series of unique restriction sites, and therefore can be used as a classical polylinker for conventional gene insertion by restriction/ligation. We suggest to choose the methods a user lab is most familiar with. For automated applications, restriction/ligation is essentially ruled out. In this case, recombination approaches can be used efficiently for gene insertion (SLIC).



D.1.1. Single gene insertion into the MIE by SLIC

Several procedures for restriction/ligation independent insertion of genes into vectors have been published or commercialized (Novagen LIC, Becton-Dickinson BD In-Fusion and others), each with its own merit. All of these systems rely on the exonuclease activity of DNA polymerases. In the absence of dNTPs, 5' extensions are created from blunt ends or overhangs by digestion from the 3' end. If two DNA fragments contain the same approx. 20 bp sequence at their termini at opposite

ends, this results in overhangs that share complementary sequences capable of annealing. This can be exploited for ligation independent combination of two or several DNA fragments containing homologous sequences.

If T4 DNA polymerase is used, this can be carried out in a manner that is independent of the sequences of the homology regions (Sequence and Ligation Independent Cloning, SLIC) and detailed protocols have become available. In the context of multi-protein expression, this is particularly useful, as the presence of unique restriction sites, or their creation by mutagenesis, in the ensemble of encoding DNAs ceases to be an issue.

We adapted SLIC for the insertion of encoding DNAs amplified by Phusion polymerase into the ACEMBL Acceptor and Donor vectors according to the published protocols. This not only allows seamless integration of genes into the expression cassettes, but also concatamerization of expression cassettes into multigene constructs via a simple and repetitive routine that can be readily automated.

Illustration 4: **Single gene insertion by SLIC**. A gene of interest (GOI 1) is PCR amplified with specific primers and integrated into a vector (Acceptor, Donor) linearized by PCR with complementary primers (complementary regions are shaded in light gray or dark grey, respectively). Resulting PCR fragments contain homology regions at their ends. T4 DNA polymerase acts as an exonuclease in the absence of dNTP and produces long sticky overhangs. Mixing (optionally annealing) of T4DNA polymerase exonuclease treated insert and vector is followed by transformation, yielding a single gene expression cassette.



We use an improved protocol for SLIC which was modified from the original publication¹². This protocol, as applied manually, is detailed below (Protocol 1). If other systems are used (BD-InFusion etc.), follow the manufacturer's recommendations. For robotics applications, modifications of the protocol may be necessary and will be detailed elsewhere¹³.

14

Protocol 1: Single gene insertion by SLIC.

Reagents required:

Phusion Polymerase 5x HF Buffer for Phusion Polymerase dNTP mix (10 mM) T4 DNA polymerase (and10x Buffer) DpnI enzyme *E. coli* competent cells 100mM DTT, 2M Urea, 500 mM EDTA Antibiotics

Step 1: Primer design

Primers for the SLIC procedure are designed to provide the regions of homology that result in long sticky ends after treatment with T4 DNA polymerase in the absence of dNTP:

Primers for the <u>insert</u> contain a DNA sequence corresponding to this region of homology ("Adaptor sequence" in Illustration 4, inset), followed by a sequence stretch that specifically anneals to the insert to be amplified (Illustration 4, inset). Useful adaptor sequences for SLIC are listed below (Table I).

If the gene of interest (GOI) is amplified from a vector already containing expression elements (e.g. the pET vector series), this "insert specific sequence" can be located upstream of a ribosome binding site (rbs). Otherwise, the forward primer needs to be designed such that a ribosome binding site is also provided in the final construct (Illustration 4, inset).

¹² Li and Elledge, *Nat. Methods* **4**, 251 (2007)

¹³ Bieniossek, Nie et al., Nat. Methods 6, 447 (2009).

Primers for PCR linearization of the <u>vector</u> backbone are simply complementary

to the two adaptor sequences present in the primer pair chosen for insert amplification (Illustration 4).

Step 2: PCR amplification of insert and vector

Identical reactions are prepared in 100- μ l volume for DNA insert to be cloned and vector to be linearized by PCR:

ddH ₂ O	75 µl
5× Phusion HF Reaction buffer	20 µl
dNTPs (10 mM stock)	2 μl
Template DNA (100 ng/μl)	1 µl
5' SLIC primer (100 μM stock)	1 µl
3' SLIC primer (100 μM stock)	1 µl
Phusion polymerase (2 U/μl)	0.5 μl

PCR reactions are then carried out with a standard PCR program (unless very long DNAs are amplified, in which case you should the double extension time):

1 x 98° C for 2 min 30 x [98° C for 20 sec. -> 50°C for 30 sec. -> 72°C for 3 min] Hold at 10°C

Analysis of the PCR reactions by agarose gel electrophoresis and ethidium bromide staining is recommended.

Step 3: DpnI treatment of PCR products (optional)

PCR reactions are then supplied with 1 μ l DpnI enzyme which cleaves parental plasmids (that are methylated). For insert PCR reactions, DpnI treatment is not required if the resistance marker of the template plasmid differs from the destination vector.

Reactions are carried out as follows: Incubation: 37°C for 1-4h Inactivation: 80°C for 20 min

Step 4: Purification of PCR products

! PCR products must be purified to remove residual dNTPs!

Otherwise, the T4 DNA polymerase reaction (Step 5) may be compromised.

Product purification is best performed by using commercial PCR Purification Kits or NucleoSpin Kits (Qiagen, MacheryNagel or others). It is recommended to perform elution in the minimal possible volume indicated by the manufacturer.

Step 5: T4 DNA polymerase exonuclease treatment

Identical reactions are prepared in $20-\mu$ l volume for insert and for vector (eluted in Step 4):

10x T4 DNA polymerase buffer	2 μl
100mM DTT	1 µl
2M Urea	2 μl
DNA eluate from Step 3 (vector or insert)	14 µl
T4 DNA polymerase	1 µl

Reactions are then carried out as follows:

Incubation:	23°C for 20 min
Arrest:	Addition of 1 μl 500 mM EDTA
Inactivation:	75°C for 20 min

Step 6: Mixing and Annealing

Exonuclease treated insert and vector are then mixed, followed by an (optional) annealing step which was found to enhance efficiency¹⁴:

T4 DNA pol ti	reated insert:	10 µl
T4 DNA pol ti	reated vector:	10 µl
Annealing:	65°C for 10 min	
Cooling:	Slowly (in heat block) to RT	

Step 7: Transformation

Mixtures are next transformed into competent cells following standard transformation procedures.

Reactions for pACE1 and pACE2 derivatives are transformed into standard *E. coli* cells for cloning (such as TOP10, DH5 α , HB101) and after recovery (2-4h) plated on agar containing ampicillin (100 µg/ml) or tetracycline (25 µg/ml), respectively.

Reactions for Donor derivatives are transformed into *E. coli* cells expressing the *pir* gene (such as BW23473, BW23474, or PIR1 and PIR2, Invitrogen) and plated on agar containing chloramphenicol (25 μ g/ml, pDC), kanamycin (50 μ g/ml, pDK), and spectinomycin (50 μ g/ml, pDS).

Step 8: Plasmid analysis

Plasmids are cultured in small-scale in media containing the corresponding antibiotic, and analyzed by sequencing and (optionally) restriction mapping with an appropriate restriction enzyme.

¹⁴ Dr. Rolf Jaussi, PSI Villigen, personal communication

D.1.2. Polycistron assembly in MIE by SLIC

The multiple integration element can also be used to integrate genes of interest by using multi-fragment SLIC recombination as shown in Illustration 4. Genes preceded by ribosome binding sites (rbs) can be assembled into polycistrons in this way.

Illustration 5: **Generating a polycistron by SLIC.** Genes of interest (GOI 1,2,3) are PCR amplified with specific primers and integrated into a vector (Acceptor, Donor) linearized by PCR with primers complementary to the ends of the forward primer of the first (GOI 1) and the reverse primer of the last (GOI 3) gene to be assembled in the polycistron (complementary regions are shaded in light gray or dark grey, respectively). Resulting PCR fragments contain homology regions at the ends. T4 DNA polymerase acts as an exonuclease in the absence of dNTP and produces long sticky overhangs. Mixing (optionally annealing) of T4DNA polymerase exonuclease treated insert and vector is followed by transformation, yielding a polycistronic expression cassette.



Protocol 2. Polycistron assembly by SLIC.

Reagents required:

Phusion Polymerase

5x HF Buffer for Phusion Polymerase

dNTP mix (10 mM)

T4 DNA polymerase (and 10x Buffer)

E. coli competent cells

100mM DTT, 2M Urea, 500 mM EDTA

Antibiotics

Step 1: Primer design

The MIE element is composed of tried-and-tested primer sequences. These constitute the "Adaptor" sequences that can be used for inserting single genes or multi-gene constructs. Recommended adaptor sequences are listed below (Table I).

Adaptor sequences form the 5' segments of the primers used to amplify DNA fragments to be inserted into the MIE. Insert specific sequences are added at 3', DNA encoding a ribosome binding site can be inserted optionally if not already present on the PCR template

Step 2: PCR amplification of insert and primer

Identical reactions are prepared in $100-\mu$ l volume for all DNA insert (GOI 1,2,3) to be cloned and the vector to be linearized by PCR:

ddH ₂ O	75 µl
$5 \times$ Phusion HF Reaction buffer	20 µl
dNTPs (10 mM stock)	2 µl
Template DNA (100 ng/µl)	1 µl
5' SLIC primer (100 μM stock)	1 µl
3' SLIC primer (100 μM stock)	1 µl
Phusion polymerase (2 U/µl)	0.5 μl

PCR reactions are then carried out with a standard PCR program (unless very long DNAs are amplified, in which case you should double the extension time):

1 x 98° C for 2 min 30 x [98° C for 20 sec. -> 50°C for 30 sec. -> 72°C for 3 min] Hold at 10°C

Analysis of the PCR reactions by agarose gel electrophoresis and ethidium bromide staining is recommended.

Step 3: DpnI treatment of PCR products (optional)

PCR reactions are then supplied with 1 μ l DpnI enzyme which cleaves parental plasmids (that are methylated). For insert PCR reactions, DpnI treatment is not required if the resistance marker of the template plasmids differs from the destination vector.

Reactions are then carried out as follows: Incubation: 37°C for 1-4h Inactivation: 80°C for 20 min

Step 4: Purification of PCR products

! PCR products must be purified to remove residual dNTPs **!**

Otherwise, the T4 DNA polymerase reaction (Step 5) may be compromised.

Product purification is best performed by using commercial PCR Purification Kits or NucleoSpin Kits (Qiagen, MacheryNagel or others). It is recommended to perform elution in the minimal possible volume indicated by the manufacturer.

Step 5: T4 DNA polymerase exonuclease treatment

Identical reactions are prepared in $20-\mu$ l volume for each insert (GOI 1,2,3) and for the vector (eluted in Step 4):

10x T4 DNA polymerase buffer	2 µl
100mM DTT	1 µl
2M Urea	2 µl
DNA eluate from Step 3 (vector or insert)	14 µl
T4 DNA polymerase	1 µl

Reactions are then carried out as follows:

Incubation:	23°C for 20 min
Arrest:	Addition of 1 μl 500 mM EDTA
Inactivation:	75°C for 20 min

Step 6: Mixing and Annealing

T4 DNA polymerase exonuclease treated insert and vector are then mixed, followed by an (optional) annealing step which was found to enhance efficiency¹⁵:

T4 DNA pol t	reated insert 1 (GOI 1):	5 µl
T4 DNA pol t	reated insert 2 (GOI 2):	5 µl
T4 DNA pol ti	reated insert 3 (GOI 3):	5 µl
T4 DNA pol ti	reated vector:	5 µl
Annealing:	65°C for 10 min	

Cooling: Slowly (in heat block) to RT

Step 7: Transformation

Mixtures are next transformed into competent cells following standard transformation procedures.

¹⁵ Dr. Rolf Jaussi, PSI Villigen, personal communication.

Reactions for pACE1 and pACE2 derivatives are transformed into standard *E. coli* cells for cloning (such as TOP10, DH5 α , HB101) and after recovery plated on agar containing ampicillin (100 µg/ml) or tetracycline (25 µg/ml), respectively.

Reactions for Donor derivatives are transformed into *E. coli* cells expressing the *pir* gene (such as BW23473, BW23474, or PIR1 and PIR2, Invitrogen) and plated on agar containing chloramphenicol (25 μ g/ml, pDC), kanamycin (50 μ g/ml, pDK), and spectinomycin (50 μ g/ml, pDS).

Step 8: Plasmid analysis

Plasmids are cultured and correct clones are selected based on specific restriction digestion and DNA sequencing of the inserts.

Table I. Adaptor sequences

For single gene or multi-gene insertions into ACEMBL vectors by SLIC.

Adaptor ¹	Sequence	Description
T7InsFor	TAGGTATCGATAATAC GACTCACTATAGGG	Forward primer for <u>insert</u> amplification, if gene of interest (GOI) is present in a T7 system vector (i.e. pET series).
		No further extension (rbs, insert specific overlap) required.
T7InsRev	CCTCAAGACCCGTTTA GAGGCCCCAAGGGGT	Reverse primer for <u>insert</u> amplification, if GOI is present in a T7 system vector (i.e. pET series).
	TATGCTAG	No further extension (stop codon, insert specific overlap) required.
T7VecFor	CTAGCATAACCCCTTG GGGCCTCTAAACGGGT	Forward primer for <u>vector</u> amplification, reverse complement of T7InsRev.
	CTTGAGG	No further extension required.
T7VecRev	CCCTATAGTGAGTCGT ATTATCGATACCTA	Reverse primer for <u>vector</u> amplification, reverse complement of T7InsFor.
		No further extension required.
NdeInsFor	GTTTAACTTTAAGAAG GAGATATACATATG	Forward primer for <u>insert</u> amplification for insertion into MIE site I1 (Illustration 2).
		Further extension at 3' (insert specific overlap) required.
		Can be used with adaptor XhoInsRev in case of single fragment SLIC (Illustr. 3).
XhoInsRev	pACE1,pACE2,pDC (T7):	Reverse primer for <u>insert</u> amplification for insertion into MIE site I4 (Illustr. 2).
	CTTTGTTAGCAGCCGG ATCTCTCGAG	Further extension at 3' (stop codon, insert specific overlap) required.
	pDK,pDS (lac): GGGTTTAAACGGAACT AGTCTCGAG	Can be used with adaptor NdeInsFor in case of single fragment SLIC (Illustr. 3).
XhoVecFor	pACE1,pACE2,pDC (T7):	Forward primer for <u>vector</u> amplification, reverse complement of XhoInsRev

	CTCGAGAGATCCGGCT GCTAACAAAG	No further extension required.
	pDK,pDS (lac): CTCGAGACTAGTTCCGT TTAAACCC	
NdeVecRev	CATATGTATATCTCCTT CTTAAAGTTAAAC	Reverse primer for <u>vector</u> amplification, reverse complement of NdeInsFor.
		No further extension required.
SmaBam	GAATTCACTGGCCGTC GTTTTACAGGATCC	Reverse primer for <u>insert</u> amplification (GOI1) for insertion into MIE site I1 (Illustr. 2).
		Further extension at 3' (stop codon, insert specific overlap) required.
		Use with adaptor NdeInsFor.
BamSma	GGATCCTGTAAAACGA CGGCCAGTGAATTC	Forward primer for <u>insert</u> amplification (GOI2) for insertion into site I2 (Illustr. 2,4).
		Further extension at 3' (rbs, insert specific overlap) required.
		Use with adaptor SacHind (multifragment SLIC, Illustr. 4)
SacHind	GCTCGACTGGGAAA ACCCTGGCGAAGCT	Reverse primer for <u>insert</u> amplification (GOI2) insertion into MIE site I2 (Illustr. 2, 4).
	Т	Further extension at 3' (stop codon, insert specific overlap) required.
		Use with adaptor BamSma (multifragment SLIC, Illustr. 4)
HindSac	AAGCTTCGCCAGGG TTTTCCCAGTCGAGC	Forward primer for <u>insert</u> amplification (GOI3) for insertion into site I3 (Illustr. 2,4).
		Further extension at 3' (rbs, insert specific overlap) required.
		Use with adaptor BspEco (multifragment SLIC, Illustr. 4)
BspEco5	GATCCGGATGTGAA ATTGTTATCCGCTGG	Reverse primer for <u>insert</u> amplification (GOI3) insertion into MIE site I3 (Illustr. 2, 4).
	TACC	Further extension at 3' (stop codon, insert specific overlap) required.
		Use with adaptor HindSac.(multifragment SLIC, Illustr. 4)
Eco5Bsp	GGTACCAGCGGATA ACAATTTCACATCCG	Forward primer for <u>insert</u> amplification (GOI4) for insertion into site I4 (Illustr. 2,4).
	GATC	Further extension at 3' (rbs, insert specific overlap) required.
		Use with adaptor XhoInsRev (multifragment SLIC, Illustr. 4)

¹ All Adaptor primers (without extension) can be used as sequencing primers for genes of interest that were inserted into the MIE.

D.1.3. Gene insertion by restriction/ligation

The MIE can also be interpreted as a simple multiple cloning site with a series of unique restriction sites. The MIE is preceded by a promoter and a ribosome binding site, and followed by a terminator. Therefore, cloning into the MIE by classical restriction/ligation also yields functional expression cassettes.

Genes of interest can be subcloned by using standard cloning procedures into the multiple integration element (MIE) (see Appendix) of ACEMBL vectors (the MIE is identical in all vectors).

Protocol 3. Restriction/ligation cloning into the MIE.

Reagents required:

Phusion Polymerase 5x HF Buffer for Phusion Polymerase dNTP mix (10 mM) 10 mM BSA Restriction endonucleases (and 10x Buffer) T4 DNA ligase (and 10x Buffer) Calf or Shrimp intestinal alkaline phosphatase *E. coli* competent cells Antibiotics

Step 1: Primer design

For conventional cloning, PCR primers are designed containing chosen restriction sites, preceded by appropriate overhangs for efficient cutting (c.f. New England Biolabs catalog), and followed by \geq 20 nucleotides overlapping with the gene of interest that is to be inserted.

All MIEs are identical in the ACEMBL vectors. They contain a ribosome binding site preceding the Ndel site. For single gene insertions, therefore, a rbs does not need be included in the primer.

If multi-gene insertions are planned (for example into insertion sites I1-I4 of the MIE), primers need to be designed such that a rbs preceding the gene and a stop codon at its end are provided.

In particular for polycistron cloning by restriction/ligation, it is recommended to construct templates by custom gene synthesis. In the process, the restriction sites present in the MIE can be eliminated from the encoding DNAs.

Step 2: Insert preparation

PCR of insert(s):

Identical PCR reactions are prepared in 100 μI volume for genes of interest to be inserted into the MIE:

ddH ₂ O	75 µl
5× Phusion HF Reaction buffer	20 µl
dNTPs (10 mM stock)	2 µl
Template DNA (100 ng/μl)	1 µl
5' primer (100 μM stock)	1 µl
3' primer (100 μM stock)	1 µl
Phusion polymerase (2 U/µl)	0.5 μl

PCR reactions are then carried out with a standard PCR program (unless very long DNAs are amplified, then double extension time):

1 x 98° C for 2 min 30 x [98° C for 20 sec. -> 50°C for 30 sec. -> 72°C for 3 min] Hold at 10°C

Analysis of the PCR reactions by agarose gel electrophoresis and ethidium bromide staining is recommended.

Product purification is best performed by using commercial PCR Purification Kits or NucleoSpin Kits (Qiagen, MacheryNagel or others). It is recommended to perform elution in the minimal possible volume indicated by the manufacturer.

Restriction digestion of insert(s):

Restriction reactions are carried out in 40 μ l reaction volumes, using the specific restriction enzymes as specified by the manufacturer's recommendations (c.f. New England Biolabs catalog and others).

PCR Kit eluate (≥1 μg)	30 µl	
10x Restriction enzyme buffer	4 μΙ	
10 mM BSA	2 μΙ	
Restriction enzyme for 5'	2 μΙ	
Restriction enzyme for 3'	2 μl (in case of double digestion, e ddH ₂ O)	lse

Restriction digestions are performed in a single reaction with both enzymes (double digestion) or sequentially (two single digestions) if the buffer conditions required are incompatible.

Gel extraction of insert(s):

Processed insert is then purified by agarose gel extraction using commercial kits (Qiagen, MachereyNagel etc). It is recommended to elute the extracted DNA in the minimal volume defined by the manufacturer.

Step 3: Vector preparation

Restriction digestion of ACEMBL plasmid(s):

Restriction reactions are carried out in 40 μ l reaction volumes, using specific restriction enzymes as specified by manufacturer's recommendations (c.f. New England Biolabs catalog and others).

ACEMBL plasmid (≥0.5 μg) in ddH ₂ O	30 µl	
10x Restriction enzyme buffer	4 µl	
10 mM BSA	2 µl	
Restriction enzyme for 5'	2 µl	
Restriction enzyme for 3'	2 µl	(for double digestions, else ddH_2O)

Restriction digestions are performed in a single reaction with both enzymes (double digestion) or sequentially (two single digestions) if the buffer conditions required are incompatible.

Gel extraction of vector(s):

Processed vector is then purified by agarose gel extraction using commercial kits (Qiagen, MachereyNagel etc). It is recommended to elute the extracted DNA in the minimal volume defined by the manufacturer.

Step 4: Ligation

Ligation reactions are carried out in 20 μ l reaction volumes according to the recommendations of the supplier of T4 DNA ligase:

ACEMBL plasmid (gel extracted)	8 µl
Insert (gel extracted)	10 µl
10x T4 DNA Ligase buffer	2 µl
T4 DNA Ligase	0.5 μl

Ligation reactions are performed at 25°C (sticky end) for 1h or at 16°C (blunt end) overnight.

Step 5: Transformation

Mixtures are next transformed into competent cells following standard transformation procedures.

Reactions for pACE1 and pACE2 derivatives are transformed into standard *E. coli* cells for cloning (such as TOP10, DH5 α , HB101) and after recovery plated on agar containing ampicillin (100 µg/ml) or tetracycline (25 µg/ml), respectively.

Reactions for Donor derivatives are transformed into *E. coli* cells expressing the *pir* gene (such as BW23473, BW23474, or PIR1 and PIR2, Invitrogen) and plated

on agar containing chloramphenicol (25 μ g/ml, pDC), kanamycin (50 μ g/ml, pDK), and spectinomycin (50 μ g/ml, pDS).

Step 6: Plasmid analysis

Plasmids are cultured and correct clones are selected based on specific restriction digestion and DNA sequencing of the inserts.

D.1.4. Multiplication by using the HE and BstXI sites

All ACEMBL system vectors contain a homing endonuclease (HE) site and a designed BstXI site that envelop the multiple integration element (MIE). The homing endonuclease site can be used to insert entire expression cassettes, containing single genes or polycistrons, into a vector already containing one gene or several genes of interest. Homing endonucleases have long recognition sites (20-30 base pairs or more). Although not all equally stringent, homing endonuclease sites are most probably unique in the context of even large plasmids, or, in fact, entire genomes.

In the ACEMBL system, Donor vectors contain a recognition site for homing endonuclease PI-Scel (Illustr. 2). This HE site yields upon cleavage a 3' overhang with the sequence -GTGC. Acceptor vectors contain the homing endonuclease site I-Ceul, which upon cleavage will result in a 3' overhang of -CTAA. On Acceptors and Donors, the respective HE site is preceding the MIE. The 3' end of the MIE contains a specifically designed BstXI site, which upon cleavage will generate a matching overhang. The basis of this is the specificity of cleavage by BstXI. The recognition sequence of BstXI is defined as CCANNNNN'NTGG (apostrophe marks position of phosphodiester link cleavage). The residues denoted as N can be chosen freely. Donor vectors thus contain a BstXI recognition site of the sequence CCATGTGC'CTGG, and Acceptor vectors contain CCATCTAA'TTGG. The overhangs generated by BstXI cleavage in each case will match the overhangs generated by HE cleavage. Note that Acceptors and Donors have different HE sites.

The recognition sites are not symmetric. Therefore, ligation of a HE/BstXI digested fragment into a HE site of an ACEMBL vector will be (1) directional and (2) result in a hybrid DNA sequence where a HE halfsite is combined with a BstXI halfsite. This site will be cut by neither HE nor BstXI. Therefore, in a construct that

had been digested with a HE, insertion by ligation of HE/BstXI digested DNA fragment containing an expression cassette with one or several genes will result in a construct which contains all heterologous genes of interest, enveloped by an intact HE site in front, and a BstXI site at the end. Therefore, the process of integrating entire expression cassettes by means of HE/BstXI digestion and ligation into a HE site can be repeated iteratively.

Protocol 4. Multiplication by using homing endonuclease/BstXI.

Reagents required:

Homing endonucleases PI-Scel, I-Ceul

10x Buffers for homing endonucleases

Restriction enzyme BstXI (and 10x Buffer)

T4 DNA ligase (and 10x Buffer)

E. coli competent cells

Antibiotics

Step 2: Insert preparation

Restriction reactions are carried out in 40 μ l reaction volumes, using homing endonucleases PI-SceI (Donors) or I-CeuI (Acceptors) as recommended by the supplier (c.f. New England Biolabs catalog and others).

ACEMBL plasmid ($\geq 0.5 \ \mu g$) in ddH ₂ O	32 µl
10x Restriction enzyme buffer	4 μl
10 mM BSA	2 μl
PI-Scel (Donors) or I-Ceul (acceptors)	2 μl

Reactions are then purified by PCR extraction kit or acidic ethanol precipitation, and next digested by BstXI according to the recommendations of the supplier.

HE digested DNA in ddH ₂ O	32 µl
10x Restriction enzyme buffer	4 µl
10 mM BSA	2 µl
BstXI	2 μl

Gel extraction of insert(s):

Processed insert is then purified by agarose gel extraction using commercial kits (Qiagen, MachereyNagel etc). It is recommended to elute the extracted DNA in the minimal volume defined by the manufacturer.

Step 3: Vector preparation

Restriction reactions are carried out in 40 μ l reaction volumes, using homing endonucleases PI-SceI (Donors) or I-CeuI (Acceptors) as recommended by the supplier (c.f. New England Biolabs catalog and others).

ACEMBL plasmid ($\geq 0.5 \ \mu g$) in ddH ₂ O	33 µl
10x Restriction enzyme buffer	4 µl
10 mM BSA	2 µl
PI-Scel (Donors) or I-Ceul (acceptors)	1 µl

Reactions are then purified by PCR extraction kit or acidic ethanol precipitation, and next treated with intestinal alkaline phosphatase according to the recommendations of the supplier.

HE digested DNA in ddH ₂ O	17 µl
10x Alkaline phosphatase buffer	2 µl
Alkaline phosphatase	1 µl

Gel extraction of vector:

Processed vector is then purified by agarose gel extraction using commercial kits (Qiagen, MachereyNagel etc). It is recommended to elute the extracted DNA in the minimal volume defined by the manufacturer.

Step 4: Ligation

Ligation reactions are carried out in 20 μl reaction volumes:

4 µl
14 µl
2 µl
0.5 μl

Ligation reactions are performed at 25°C for 1h or at 16°C overnight.

Step 5: Transformation

Mixtures are next transformed into competent cells following standard transformation procedures.

Reactions for pACE and pACE2 derivatives are transformed into standard *E. coli* cells for cloning (such as TOP10, DH5 α , HB101) and after recovery plated on agar containing ampicillin (100 µg/ml) or tetracycline (25 µg/ml), respectively.

Reactions for Donor derivatives are transformed into *E. coli* cells expressing the *pir* gene (such as BW23473, BW23474, or PIR1 and PIR2, Invitrogen) and plated on agar containing chloramphenicol (25 μ g/ml, pDC), kanamycin (50 μ g/ml, pDK), and spectinomycin (50 μ g/ml, pDS).

Step 6: Plasmid analysis

Plasmids are cultured and correct clones selected based on specific restriction digestion and DNA sequencing of the inserts.

Note: Integration can likewise be performed by sequence and ligation independent cloning. It is recommended to carry out linearization of the vector by digestion with HE, if heterologous genes are already present, to avoid PCR amplifications over encoding regions. The fragment to be inserted is generated by PCR amplification resulting in a PCR fragment containing a 20-25 base pair stretch at its 5' end that is identical to the corresponding DNA sequence present at the HE site counted from the site of cleavage towards 5' (site of cleavage is position -4). At the 3' end of the PCR fragment, the homology region is 20-25 base pairs counted from the site of cleavage towards 3'.

D.2. Cre-LoxP reaction of Acceptors and Donors

Cre recombinase is a member of the integrase family (Type I topoisomerase from bacteriophage P1). It recombines a 34 bp loxP site in the absence of accessory protein or auxiliary DNA sequence. The loxP site is comprised of two 13 bp recombinase-binding elements arranged as inverted repeats which flank an 8 bp central region where cleavage and ligation reaction occur.

The site-specific recombination mediated by Cre recombinase involves the formation of a Holliday junction (HJ). The recombination events catalyzed by Cre recombinase are dependent on the location and relative orientation of the loxP sites. Two DNA molecules, for example an Acceptor and a Donor plasmid, containing single loxP sites will be fused. Furthermore, the Cre recombination is an equilibrium reaction with 20-30% efficiency in recombination. This provides useful options for multi-gene combinations for multi-protein complex expression.

Illustration 6: LoxP imperfect inverted repeat				
13bp	8bp	13bp		
5'ATAACTTCGTATA	GCATACAT	TATACGAAGTTAT3'		
3'TATTGAAGCATAT	CGTATGTA	ATATGCTTCAATA5'		
inverted repeat	Spacer	inverted repeat		

In a reaction where several DNA molecules such as Donors and Acceptors are incubated with Cre recombinase, the fusion/excision activity of the enzyme will result in an equilibrium state where single vectors (educt vectors) and all possible fusions coexist. Donor vectors can be used with Acceptors and/or Donors, likewise for Acceptor vectors. Higher order fusions are also generated where more than two vectors are fused. This is shown schematically in Illustration 6.

The fact that Donors contain a conditional origin of replication that depends on a pir^+ (*pir* positive) background now allows for selecting out from this reaction mix all desired Acceptor-Donor(s) combinations. For this, the reaction mix is used to transform to *pir* negative strains (TOP10, DH5 α , HB101 or other common laboratory cloning strains). Then, Donor vectors will act as suicide vectors when plated out on agar containing the antibiotic corresponding to the Donor encoded resistance marker, unless fused with an Acceptor. By using agar with the appropriate combinations of antibiotics, all desired Acceptor-Donor fusions can be selected for.

Using this approach, fusion vectors of 25 kb and larger have generated. In stability tests (serial passaging for more than 60 generations), even such large plasmids proved to be stable as checked by restriction mapping, even if only one of the antibiotics corresponding to the encoded resistance markers was provided in the growth medium.



D.2.1. Cre-LoxP fusion of Acceptors and Donors

This protocol is designed for generating multi-gene fusions from Donors and Acceptors by Cre-LoxP reaction.

Reagents:

Cre recombinase (from NEB or self made) Standard *E. coli* competent cells (*pir*⁻ strain) Antibiotics 96well microtiter plates 12 well tissue-culture plates (or petri dishes) w. agar/antibiotics 31

LB medium

- 1. For a 20 μ l Cre reaction, mix 1-2 μ g of each educt in approximately equal amounts. Add ddH₂O to adjust the total volume to 16-17 μ l, then add 2 μ l 10x Cre buffer and 1-2 μ l Cre recombinase.
- 2. Incubate Cre reaction at 37°C (or 30°C) for 1 hour.
- 3. Optional: load 2-5 μl of Cre reaction on an analytical agarose gel for examination.

Heat inactivation at 70°C for 10 minutes before the gel loading is strongly recommended.

4. For chemical transformation, mix 10-15μl Cre reaction with 200 μl chemical competent cells. Incubate the mixture on ice for 15-30 minutes. Then perform heat shock at 42°C for 45-60 s.

Up to 20 μ l Cre reaction (0.1 volumes of the chemical competent cell suspension) can be directly transformed into 200 μ l chemical competent cells.

For electrotransformation, up to 2 μ l Cre reaction could be directly mixed with 100 μ l electrocompetent cells, and transformed by using an electroporator (e.g. BIORAD E. coli Pulser) at 1.8-2.0 kV.

Larger volume of Cre reaction must be desalted by ethanol precipitation or PCR purification column before electrotransformation. The desalted Cre reaction mix should not exceed 0.1 volumes of the electrocompetent cell suspension.

The cell/DNA mixture could be immediately used for electrotransformation without prolonged incubation on ice.

- 5. Add up to 400 μ l of LB media (or SOC media) per 100 μ l of cell/DNA suspension immediately after the transformation (heat shock or electroporation).
- 6. Incubate the suspension in a 37°C shaking incubator overnight or for at least 4 hours (recovery period).

For recovering multifusion plasmid containing more than 2 resistance markers, it is strongly recommended to incubate the suspension at 37°C overnight.

- 7. Plate out the recovered cell suspension on agar containing the desired combination of antibiotics. Incubate at 37°C overnight.
- 8. Clones from colonies present after overnight incubation can be verified by restriction digestion at this stage (refer to steps 12-16).

Especially in the case that only one multifusion plasmid is desired.

For further selection by single antibiotic challenges on a 96 well microtiter plate, continue to step 9.

Several to many different multifusion plasmid combinations can be processed and selected on one 96 well microtiter plate in parallel.

- 9. For 96 well antibiotic tests, inoculate four colonies from each agar plate with different antibiotic combination into ~500 μ l LB media without antibiotics. Incubate the cell cultures in a 37°C shaking incubator for 1-2 hours.
- 10. During the incubation of colonies, fill a 96 well microtiter plate with 150 μ l antibiotic-containing LB media (following Illustration 7). It is recommended to add coloured dye (positional marker) in the wells indicated.

A typical arrangement of the solutions, which is used for parallel selections of multifusion plasmids, is shown in Illustration 7. The concept behind the 96 well plate experiment is that every cell suspension from single colonies needs to be challenged by all four single antibiotics for unambiguous interpretation.



Illustration 8: 96 well analysis of Cre assembly

11. Add 1 μ l aliquots of pre-incubated cell culture (Step 9) to the corresponding wells. Then incubate the inoculated 96 well microtiter plate in a 37°C shaking incubator overnight at 180-200 rpm.

Recommended: use parafilm to wrap the plate to avoid drying out. The remainder of the pre-incubated cell cultures could be kept at 4°C for further inoculations if necessary.

12. Select transformants containing desired multifusion plasmids based on antibiotic resistance, according to the combination of dense (positive) and clear (no growth) cell microcultures from each colony. Inoculate 10-20 μl cell

culture into 10 ml LB media with corresponding antibiotics. Incubate in a 37°C shaking incubator overnight.

- 13. Centrifuge the overnight cell cultures at 4000g for 5-10 minutes. Purify plasmid from the resulting cell pellets with common plasmid miniprep kits, according to manufacturers' recommendation.
- 14. Determine the concentrations of purified plasmid solutions by using UV absorption spectroscopy (e.g. by using a NanoDrop[™] 1000 machine).
- 15. Digest approx. 0.5-1 μ g of the purified plasmid solution in a 20 μ l restriction digestion with appropriate endonuclease(s). Incubate under recommended reaction condition for ~2 hours.
- 16. Use 5-10 μ l of the digestion for analytical agarose (0.8-1.2%) gel electrophoresis. Verify plasmid integrity by comparing the experimental restriction pattern to a restriction pattern predicted *in silico* (e.g. by using program VectorNTI from Invitrogen or similar programs).
- D.2.2. Deconstruction of fusion vectors by Cre

The protocol is suitable for releasing any single educt from multifusion constructs (deconstruction). This is achieved by Cre-LoxP reaction, transformation and plating on agar with appropriately reduced antibiotic resistance level (c.f. Illustration 6). In the liberated educt entity, encoding genes can be modified and diversified. Then, the diversified construct is resupplied by Cre-LoxP reaction (C.2.1.).

Reagents:

Cre recombinase (and 10x Buffer)

E. coli competent cells

(*pir*⁺ strains, *pir*⁻ strains could be used only when partially deconstructed Acceptor-Donor fusions are desired).

Antibiotics

- 1. Incubate approx. 1 μ g multifusion plasmid with 2 μ l 10x Cre buffer, 1-2 μ l Cre recombinase, add ddH₂O to adjust the total reaction volume to 20 μ l.
- 2. Incubate this Cre deconstruction reaction mixture at 30°C for 1-4 hour.
- 3. Optional: load 2-5 μl of the reaction on an analytical agarose gel for examination.

Heat inactivation at 70°C for 10 minutes before the gel loading is strongly recommended.

4. For chemical transformation, mix 10-15 μ l De-Cre reaction with 200 μ l chemical competent cells. Incubate the mixture on ice for 15-30 minutes. Then perform heat shock at 42°C for 45-60 s.

Up to 20 μ l De-Cre reaction (0.1 volumes of the chemical competent cell suspension) can be directly transformed into 200 μ l chemical competent cells.

For electrotransformation, up to 2 μ l De-Cre reaction can be directly mixed with 100 μ l electrocompetent cells, and transformed by using an electroporator (e.g. BIORAD *E. coli* Pulser) at 1.8-2.0 kV.

Larger volume of De-Cre reaction must be desalted by ethanol precipitation or PCR purification column before electrotransformation. The desalted De-Cre reaction mix should not exceed 0.1 volumes of the electrocompetent cell suspension.

The cell/DNA mixture could be immediately used for electrotransformation without prior incubation on ice.

- 5. Add up to 400 μ l of LB media (or SOC media) per 100 μ l of cell/DNA suspension immediately after the transformation (heat shock or electroporation).
- 6. Incubate the suspension in a 37°C shaking incubator (recovery).

For recovery of partially deconstructed double/triple fusions, incubate the suspension in a 37°C shaking incubator for 1 to 2 hours.

For recovery of individual educts such as single ACEMBL vectors from pACKS plasmid, incubate the suspension in a 37°C shaking incubator overnight or for at least 4 hours.

- 7. Plate out the recovered cell suspension on agar containing the desired (combination of) antibiotic(s). Incubate at 37°C overnight.
- 8. Colonies after overnight incubation might be verified directly by restriction digestion at this stage (refer to steps 12-16).

Especially recommended in the case that only one single educt or partially deconstructed multifusion plasmid is desired.

For further selection by single antibiotic challenge on a 96 well microtiter plate, continue with step 9.

Several different single educts/partially deconstructed multifusion plasmids can be processed and selected on one 96 well microtiter plate in parallel.

- 9. For 96 well analysis, inoculate four colonies each from agar plates containing a defined set of antibiotics into ~500 μ l LB media without antibiotics. Incubate the cell cultures in a 37°C shaking incubator for 1-2 hours.
- 10. During the incubation of colonies, fill a 96 well microtiter plate with 150 μ l antibiotic-containing LB media or coloured dye (positional marker) in the corresponding wells.

Refer to Illustrations 7 and 12 for the arrangement of the solutions in the wells, which are used for parallel selection of single educts or partially deconstructed multifusion plasmids. The concept is that every cell suspension from a single colony needs to be challenged by all four antibiotics separately for unambiguous interpretation.

11. Add 1 μ l aliquots from the pre-incubated cell cultures (Step 9) into the corresponding wells. Then incubate the 96 well microtiter plate in a 37°C shaking incubator overnight at 180-200 rpm.

Recommended: use parafilm to wrap the plate to prevent desiccation. The remainder of the pre-incubated cell cultures can be kept in 4°C fridge for further inoculations if necessary.

- 12. Select transformants containing desired single educts or partially deconstructed multifusion plasmids according to the combination of dense (growth) and clear (no growth) cell cultures from each colony. Inoculate 10-20 μl cell cultures into 10 ml LB media with corresponding antibiotic(s). Incubate in a 37°C shaking incubator overnight.
- 13. Centrifuge the overnight cell cultures at 4000g for 5-10 minutes. Purify plasmid from cell pellets with common plasmid miniprep kits, according to manufacturers' information.
- 14. Determine the concentrations of purified plasmid solutions by using UV absorption spectroscopy (e.g. NanoDrop[™] 1000).
- 15. Digest approx. 0.5-1 μ g of the purified plasmid solution in a 20 μ l restriction digestion (with 5-10 unit endonuclease). Incubate under recommended reaction condition for ~2 hours.
- 16. Use 5-10 μ l of the digestion for analytical agarose gel (0.8-1.2%) electrophoresis. Verify the plasmid integrity by comparing the actual restriction pattern to predicted restriction pattern *in silico* (e.g. by using VectorNTI, Invitrogen, or any other similar program).
- 17. Optional: Possibly, a deconstruction reaction is not complete but yields partially deconstructed fusions which still retain entities to be eliminated. In this case, we recommend to pick these partially deconstructed fusions containing and perform a second round of Cre deconstruction reaction (repeat steps 1-8) by using this construct as starting material.

In our hands, two sequential deconstruction reactions were always sufficient to recover all individual modules, for instance all four single ACEMBL vectors from a pACKS plasmid. Liberation of single educts from double/triple fusions were found to be often more efficient than from quadruples such as the pACKS plasmid of the system kit (Section E).

D.3. Co-expression by Co-transformation

Protein complexes can be expressed also from two separate vectors co-transformed in expression strains. The co-transformed vectors can have the same or different origins of replication, however, they must encode different resistance markers. Plasmids pACE1 (ampicillin resistance marker) and pACE2 (tetracycline resistance marker) have both a ColE1 derived replicon and can therefore be used with all common expression strains. pACE1 and pACE2 derivatives (also including fused Donors if needed) can be co-transformed into expression strains, and double transformants selected for by plating on agar plates containing both ampicillin and tetracyclin antibiotics.

Transformations are carried out by using standard transformation protocols.

E. ACEMBL multi-gene combination: Examples

Examples of multi-protein expressions by ACEMBL illustrating the gene combination procedures detailed in Section C are shown in the following. Reactions presented were carried out manually following the protocols provided, and also on a Tecan Freedom Evoll 200 robot with adapted protocols. These and other examples have been published in Bieniossek et al., Nat. Methods 6, 447 (2009) plus supplementary material.

E.1. SLIC cloning into ACEMBL vectors: human TFIIF

Genes encoding full-length human RAP74 with a C-terminal oligo-histidine tag and full-length human RAP30 were amplified from pET-based plasmid template¹⁶ following the protocols described above. Linearized vector backbones were generated by PCR amplification from pACE1 and pDC. SLIC following Protocol 1 (Section D), yielded pACE1-RAP30 and pDC-RAP74his (Fig 8). These plasmids were fused by Cre-LoxP reaction (Section C). Results from restriction mapping by BstZ17I/BamHI double digestion of 11 double resistant (Cm, Ap) colonies are shown by a gel section from 1% E-gel electrophoresis (M: NEB 1kb DNA marker). All clones tested showed the expected pattern (5.0 + 2.8 kb). One clone was transformed in BL21(DE3) cells. Expression and purification by Ni²⁺-capture and S200 chromatography resulted in human TFIIF complex.



Illustration 9: ACEMBLing TFIIF (Image courtesy of Dr. I. Berger, EMBL Grenoble).

¹⁶ Gaiser, Tan and Richmond, *J. Mol. Biol.* **302**, 1119 (2000).

E.2. The Homing endonuclease/BstXI module: yeast RES complex

Plasmids pCDFDuet-Pml1p, pRSFDuet-Snu17p-NHis and pETDuet-Bud13p, encoding for yeast proteins (all full-length) Pml1p, Snu17p and Bud13p, respectively, were a kind gift from Dr. Simon Trowitzsch and Dr. Markus Wahl (MPI Göttingen). Snu17p contains a six-histidine tag fused to its N-terminus. The gene encoding for His6tagged Snu17p was excised from pRSFDuet-Snu17p-NHis by Ncol/XhoI restriction, and ligated into a Ncol/XhoI digested pACE1 construct (containing an unrelated gene between the NcoI and XhoI sites) resulting in pACE1-Snu17. The gene encoding for Bud13p was liberated from pETDuet-Bud13p by restriction digestion with XbaI and EcoRV, and placed into XbaI/PmeI digested pDC resulting in pDC-Bud13. The gene encoding Pm1Ip was liberated from pCDFDuet-Pml1p by restriction digestion with NdeI and XhoI, and placed into NdeI/XhoI digested pDC resulting in pDC-Pml1. Next, the expression cassette for Bud13p was liberated from pDC-Bud13 by digestion with PI-SceI and BstXI. The liberated fragment was inserted into PI-SceI digested and alkaline phosphatase treated pDC-Pml1p resulting in pDC-Bud13p-Pml1p.

pACE-Snu17 and pDC-BudPmI were then fused by Cre-LoxP reaction and selected for by plating on agar plates containing ampicillin and chloramphenicol. Fusion plasmids were transformed into BL21(DE3) cells. Expression and purification by Ni²⁺-capture and S200 size exclusion chromatography resulted in the trimeric RES complex.



Illustration 10: The HE/BstXI multiplication module.

F. Appendix

F.1. DNA sequence of MIE

Below are the sequence and map of the MIE fragment between T7/lac promoter and T7 terminator in ACEMBL vectors. Forward and reverse primers for sequencing can be standard vector primers for T7 and lac. Adaptor primer sequences (c.f. Table I) are indicated. DNA sequences in these homology regions contain tried-and-tested sequencing primers¹⁷. Sites of insertion (I1-I4) are shown.. The adaptor sequences, and probably any sequence in the homology regions, can be used as adaptors for multifragment insertions. The ribosome binding site present in the MIE (rbs) is boxed in red.



¹⁷ Tan S et al. *Protein Expr. Purif.* **40**, 385 (2005).

F.2. DNA sequences of ACEMBL vectors

Below are the complete sequences of all acceptor and donor vectors.

Please note though that copy-pasting from a PDF may result in faulty data transfer into other applications, i.e. DNA sequences may lack base pairs.

To avoid any trouble, simply request the vector sequence data at <u>order@atg-biosynthetics.de</u>. We will provide them either in text format or as ApE files that will give you ample annotation of the features contained on the vectors. The ApE plasmid editor can be downloaded for free from its author's web site (M. Wayne Davis, University of Utah).

F.2.1. pACE1

1	GGTACCGCGG	CCGCGTAGAG	GATCTGTTGA	TCAGCAGTTC	AACCTGTTGA
51	TAGTACTTCG	TTAATACAGA	TGTAGGTGTT	GGCACCATGC	АТААСТАТАА
101	CGGTCCTAAG	GTAGCGACCT	AGGTATCGAT	AATACGACTC	ACTATAGGGG
151	AATTGTGAGC	GGATAACAAT	TCCCCTCTAG	AAATAATTTT	GTTTAACTTT
201	AAGAAGGAGA	TATACATATG	AGGCCTCGGA	TCCTGTAAAA	CGACGGCCAG
251	TGAATTCCCC	GGGAAGCTTC	GCCAGGGTTT	TCCCAGTCGA	GCTCGATATC
301	GGTACCAGCG	GATAACAATT	TCACATCCGG	ATCGCGAACG	CGTCTCGAGA
351	GATCCGGCTG	CTAACAAAGC	CCGAAAGGAA	GCTGAGTTGG	CTGCTGCCAC
401	CGCTGAGCAA	TAACTAGCAT	AACCCCTTGG	GGCCTCTAAA	CGGGTCTTGA
451	GGGGTTTTTT	GGTTTAAACC	CATCTAATTG	GACTAGTAGC	CCGCCTAATG
501	AGCGGGCTTT	TTTTTAATTC	CCCTATTTGT	TTATTTTTCT	AAATACATTC
551	AAATATGTAT	CCGCTCATGA	GACAATAACC	CTGATAAATG	CTTCAATAAT
601	ATTGAAAAAG	GAAGAGTATG	AGTATTCAAC	ATTTCCGTGT	CGCCCTTATT
651	CCCTTTTTTG	CGGCATTTTG	CCTTCCTGTT	TTTGCTCACC	CAGAAACGCT
701	CGTGAAAGTA	AAAGACGCAG	AGGACCAATT	GGGGGCACGA	GTGGGATACA
751	TAGAACTGGA	CTTGAATAGC	GGTAAAATCC	TTGAGAGTTT	TCGCCCTGAA
801	GAGCGTTTTC	CAATGATGAG	CACTTTCAAA	GTTCTGCTAT	GTGGAGCAGT
851	ATTATCCCGT	GTAGATGCGG	GGCAAGAGCA	ACTCGGACGA	CGAATACACT
901	ATTCGCAGAA	TGACTTGGTT	GAATACTCCC	CAGTGACAGA	AAAGCACCTT
951	ACGGACGGAA	TGACGGTAAG	AGAATTATGT	AGTGCCGCCA	TAACGATGAG
1001	TGATAACACT	GCGGCGAACT	TACTTCTGAC	AACCATCGGT	GGACCGAAGG
1051	AATTAACCGC	TTTTTTGCAC	AATATGGGAG	ACCATGTAAC	TCGCCTTGAC
1101	CGTTGGGAAC	CAGAACTGAA	TGAAGCCATA	CCAAACGACG	AGCGAGACAC
1151	CACAATGCCT	GCGGCAATGG	CAACAACATT	ACGCAAACTA	TTAACTGGCG
1201	AACTACTTAC	TCTGGCTTCA	CGGCAACAAT	TAATAGACTG	GCTTGAAGCG
1251	GATAAAGTTG	CAGGACCACT	ACTGCGTTCG	GCACTTCCTG	CTGGCTGGTT
1301	TATTGCTGAT	AAATCTGGGG	CAGGAGAGCG	TGGTTCACGG	GGTATCATTG
1351	CCGCACTTGG	ACCAGATGGT	AAGCCTTCCC	GTATCGTAGT	TATCTACACG
1401	ACGGGTAGTC	AGGCAACTAT	GGACGAACGA	AATAGACAGA	TTGCTGAAAT
1451	AGGGGCTTCA	CTGATTAAGC	ATTGGTAAAC	CGATACAATT	AAAGGCTCCT
1501	TTTGGAGCCT	TTTTTTTGG	ACGGACCGGT	AGAAAAGATC	AAAGGATCTT
1551	CTTGAGATCC	TTTTTTTCTG	CGCGTAATCT	GCTGCTTGCA	ААСАААААА
1601	CCACCGCTAC	CAGCGGTGGT	TTGTTTGCCG	GATCAAGAGC	TACCAACTCT
1651	TTTTCCGAAG	GTAACTGGCT	TCAGCAGAGC	GCAGATACCA	AATACTGTCC
1701	TTCTAGTGTA	GCCGTAGTTA	GGCCACCACT	TCAAGAACTC	TGTAGCACCG
1751	CCTACATACC	TCGCTCTGCT	AATCCTGTTA	CCAGTGGCTG	CTGCCAGTGG
1801	CGATAAGTCG	TGTCTTACCG	GGTTGGACTC	AAGACGATAG	TTACCGGATA
1851	AGGCGCAGCG	GTCGGGCTGA	ACGGGGGGTT	CGTGCACACA	GCCCAGCTTG
1901	GAGCGAACGA	CCTACACCGA	ACTGAGATAC	CTACAGCGTG	AGCTATGAGA
1951	AAGCGCCACG	CTTCCCGAAG	GGAGAAAGGC	GGACAGGTAT	CCGGTAAGCG
2001	GCAGGGTCGG	AACAGGAGAG	CGCACGAGGG	AGCTTCCAGG	GGGAAACGCC
2051	TGGTATCTTT	ATAGTCCTGT	CGGGTTTCGC	CACCTCTGAC	TTGAGCGTCG
2101	ATTTTTGTGA	TGCTCGTCAG	GGGGGCGGAG	CCTATGGAAA	AACGCCAGCA
2151	ACGCGGCCTT	TTTACGGTTC	CTGGCCTTTT	GCTGGCCTTT	TGCTCACATG
2201	TTCTTTCCTG	CGTTATCCCC	TGATTCTGTG	GATAACCGTA	TTACCGCCTT
2251	TGAGTGAGCT	GATACCGCTC	GCCGCAGCCG	AACGACCGAG	CGCAGCGAGT
2301	CAGTGAGCGA	GGAAGCGGAA	GAGCGCCTGA	TGCGGTATTT	TCTCCTTACG
2351	CATCTGTGCG	GTATTTCACA	CCGCAATGGT	GCACTCTCAG	TACAATCTGC
2401	TCTGATGCCG	CATAGTTAAG	CCAGTATACA	CTCCGCTATC	GCTACGTGAC
2451	TGGGTCATGG	CTGCGCCCCG	ACACCCGCCA	ACACCCGCTG	ACGCGCCCTG
2501	ACGGGCTTGT	CTGCTCCCGG	CATCCGCTTA	CAGACAAGCT	GTGACCGTCT
2551	CCGGGAGCTG	CATGTGTCAG	AGGTTTTCAC	CGTCATCACC	GAAACGCGCG
2601	AGGCAGGGGG	AATTCCAGAT	AACTTCGTAT	AATGTATGCT	ATACGAAGTT
2651	AT				

F.2.2. pACE2

ATGAAATCTA ACAATGCGCT CATCGTCATC CTCGGCA51TGCTGTAGGC ATAGGCTTGG TTATGCCGGT ACTGCGGG51CTATATGCTCA TTCCGACAGC ATCGCCGCCAC CATCATGGG201CGACGCTTT GGCCGCCCC CACTCTCT CGCTGGG201GCCGGACGAA TCGTGGCCGG CATCACCGCG CGCACAGG301GCCGGACGAA TCGTGGCCGG CATCACCGCG CGCACAGG351CGCCTATATC GCCGACATCA CCGATGGGGA AGATCGGG401GGCTCATAGG CGCTTGTTC GCGGTGGAA TGGTGGCG451GGGGGACTGT TGGGCGCAC CTCCTACAT GCACCAT501GGTCCTACAG GGCCTCAACC TACTACTGGG CTGCTCC551CGCATAAGG AGAGCGCCG CCCATGCCT TGAGAGCG601ACCTCCTCC GGTGGCCGG GGCATGACT ATCGTCGG611TCATTTTGGG CAGGACACCT TCGCGAGAC GGCGCACC751TCGCTTGCGG TATTCGGAAT CTTGCACGC CTCGCTCZ801TGGTCCCCC ACCAAACCTT TCGGCGAGAA GCAGCGT901TGGATGGCCT TCCCCATTAT GATCTTCC GCTCGGG951GCCCGCGTG CAGGCCATGC TGTCCAGGCA GGTAGAT1001GACAGCTTCA AGGATCGCT GCGGCCTCTA CCAGCCT951GCCCGCGTG CGGTGCATGG AGCCGCCCC ATACCTTC1051GGACGCTGC CGGTGCATGG AGCCGGCCA CCTCGAC1251ATCAAAGGAT CTTCTTGAGA TCTTTTTT TGGACGGT1251ACCAAACAAA AAACCACCCC TACCAGCGT GCTCAGC1451CCTGCTGCCAG TGGCGATAGA CCGCTAGCA GGCTGGCC TAACCCT1551AGCTACCAAC TCTTTTGGA CCTTTTTTT TGGACGGT1551AACAACAAC TCTTTTTGG AAGGTCATA CCGGCTGGC TAACCCT1551ACCAAACACA TCTTTTTC TGGAGCAACGG GCTTAACCG1551AACAACCAAC TCTTTTTTC TGAGCGCAC CGACCAGC1551AACACCAAC TCTTTTTC TGAGCGCA CGCTCTGC GCGAACGG1651GTAGTACCAC TGCGGAACC ACCCTCCCC AAGGTA <td< th=""><th>CCG TCACCCTGGA GGC CTCTTGCGGG CGT GCTGCTAGCG TCG GAGCACTGTC CTA CTTGGAGCCA GTG GATTCTCTAC GTG CGGTTGCTGG GCT CGCCACTTCG</th><th>CTCGGCACCG ACTGCCGGGC ACTATGGCGT</th><th>CATCGTCATC TTATGCCGGT</th><th>ACAATGCGCT ATAGGCTTGG</th><th>ATGAAATCTA TGCTGTAGGC</th><th>1 51 101</th></td<>	CCG TCACCCTGGA GGC CTCTTGCGGG CGT GCTGCTAGCG TCG GAGCACTGTC CTA CTTGGAGCCA GTG GATTCTCTAC GTG CGGTTGCTGG GCT CGCCACTTCG	CTCGGCACCG ACTGCCGGGC ACTATGGCGT	CATCGTCATC TTATGCCGGT	ACAATGCGCT ATAGGCTTGG	ATGAAATCTA TGCTGTAGGC	1 51 101
51TGCTGTAGGCATAGGCTTGGTATAGCCGTACTGCCGT101ATATCGTCCATTCGAGCAGCATCGCCGCACTATGGC151CTATATGGCTTGATGCAATGGCCACCACACCCGCTTCG201CGACCGCTTGCCGGACCACACCGCGCCGCCAGTCCTGCT301GCCGGACGCATCGTGGCCGCCACTACCGGCGCCCGCCTATAC351CGCCTATACGGCCGCACATCCCGATGGGAAATGGTGGC401GCTCATGAGGCCGTGTTTCGGGGGGACGTTGGTGGCCA451GGGGGACTGTTGGGCGCCACTCCTTACATGCACCATT501GGTGCTCAACGGTGGCCGACCCATGCCTTGGAGGCG511CGCATAAGGAAGCGCCGACCCATGCCTTGGAGGCG601ACCTCCTTCCGGTGGGCGACGCGCGCGGGTGCCGC751TGGCTGCGGTATCGAGAACCTTGCACGCCCTCGCGCGCG751TGGCGGCCGACGCGCGGGGCTACGTTGCAGGGCGGCGA751TGGCGGCCGACGCGCGGGGCCACCACCTCGCGCGCTG751TGGCGGCCGACGCGCGGGCCCTCCGGCTGCGGGCGGCCA751TGGCGGCCGACGCGCGGCCGCTCCGCGCTGGGGCGGCCA751TGGCGGCGACGGCCGGCCGCTCCGCGCTGGGCTCCGGC751TGGCGGCGACGCCGCGCGCCTCCGCGCGGCCGCGCCC751GGACGGCGACGCCGCGCCCGGGCGAGCGGCTCCGGCA751GGACGCGCGATGGTCAGGCGCTCCGGCAGCTCCGGCA751GGACGCGCGATGGTCAGGCGCTCGGCAGCTCCGGCA751GGACGCGCGATGGTCAGCG<	GGC CTCTTGCGGG CGT GCTGCTAGCG TCG GAGCACTGTC CTA CTTGGAGCCA GTG GATTCTCTAC GTG CGGTTGCTGG GCT CGCCACTTCG	ACTGCCGGGC ACTATGGCGT	TTATGCCGGT ATCGCCAGTC	ATAGGCTTGG	TGCTGTAGGC	51
101ATATCGTCCATTCCGACAGCATCGCCAGTCACTATGG151CTATATGCGTTGATGCAATTTCTATGCGCACCGCTTCG201CGACCGCTTGGCGCCCCCAGTCCTGCCGCTCGTC301GCCGGACGCATCGTGGCCGCCATCACCGGCGCCCACAGC301GCCGGACGATCGTGGCCGGCATCACCGGCGCCCACAGC301GGCGCACGACGCGTGTTCGGCGGCGAACGCCACAGC401GGCTCATGAGGCCTGTTCTGGCGGCGAACGCCATATC501GGTGCTCAACGGCCTCAACCTACTACTGGGTGCTCTCT501GGTGCTCAACGGCCGCGACGCCCATGCCCTTGGAGCGCG501GGTGCTCACCGGTGGCCGCGCCGAGACGCTTCGCTGGG501GGTGCTCCCCGGTGGCGCGGCCGCGCGCTTGGCCGCC601ACCTCTTCTATCAGGACACCCTTCGCTGGGGGCCGCGC701TCATTTCGCGAGGCACCCTTCCCATTACGGCGCGCCC71TCGCTGCGCGACCGCCACGCTTCCCATTACGCCGCGCTC801TGGCCGCCGACCGCCATGCTGCCCGCGTGGACAGCTTA901TGGATGGCACGGCGCATGCGCCCCGCCCATACAGCG901GCAACCATCACCGTTCTGGAGCCCCGCCCCATACACCGC1001GAACGCTGCCCTTTTGGAGCCTTTTTTTGGGCGCG1011GGAGGTACAGCGCCGCCCACCACCGCGCTCAACCA1011CGGGTGGCACCTTTTTGGAGCCTTTTTTTGGCGGCG1151CGCACACACACCTTTTTTTTTGGCGGCGG1201ATTAAAGGACCTTTTTTTTTTTTGGGGGCGG <td>CGT GCTGCTAGCG TCG GAGCACTGTC CTA CTTGGAGCCA GTG GATTCTCTAC GTG CGGTTGCTGG GCT CGCCACTTCG</td> <td>ACTATGGCGT</td> <td>ATCGCCAGTC</td> <td>TTCCCACACC</td> <td>ΔΤΑΤΟΓΤΟΑ</td> <td>101</td>	CGT GCTGCTAGCG TCG GAGCACTGTC CTA CTTGGAGCCA GTG GATTCTCTAC GTG CGGTTGCTGG GCT CGCCACTTCG	ACTATGGCGT	ATCGCCAGTC	TTCCCACACC	ΔΤΑΤΟΓΤΟΑ	101
151CTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCG201CGACCGCTTGGCCGCCCCCATCCTCTCGCTTCGT301GCCGGACCATCCGACCACCCCGCTCCTGC351CGCCTATATCGCCGACATCACCGATGGGGAAGATCGGC401GGCTCATAGGGGCTCATACTGCCGACGCATGGTGGCCAT501GGGGACTGATGGCGCTCAACCGGCCTCATACTGCCGATGACG501GGTGCTCACGGCGCCCACTACTACTGGGTACGTCGCG501GGTCCTTCCGGTGGGCCGACCCATGCCCTGGTGCCGG601AGCTCCTTCCGGTGGGGCGGGGCCATGACTATCGTCGG611TCATTTCGGCGAGGACGCTTTCGCTGGAGCGCGCCC701TCATTTCGGGGAGCGCCATCTCGCTGGAGCGGCCCC801TGGTCCCGCCACCAAACGTTTCGCGGCGAGAGCAGGCCC901TGGATGGCCTTCCCCATTATGATTCTTCCGCCCGCTGC901TGGATGGCCATCCTCATATGGCCCGCCCCATTACCGG901GGACGCTGATCGTCACGGCGATTATGCCGCCCGCCCC901GGACGCTGATCGTCAGGAGCCCCGCCCATACAGGA1001GACAGCTTCAAGGATCGACGGCCCGGCCCCATACCGCC1011CGGGTGGCATCTTTTGGAGCCTTTTTTTTGGGCGCGGC1151CGTGCGCAACTCTTTTCGGAGGCTTCCGGC1201ATTAAAGGACCCTCTTTAGGAGTCTCTGGGC1301GCAAACAAAAAACCACCCTACCGCCTCCACGGCGGCA1451CTCTGTAGCACGCGCCTACATACCGGGTGGACGACGCTACG	TCG GAGCACTGTC CTA CTTGGAGCCA GTG GATTCTCTAC GTG CGGTTGCTGG GCT CGCCACTTCG	~~~~~~~~~~	1110000011010	11000101000	111110010011	TOT
201CGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGT301GCCGGACGATCGTGGCCGGCCACCACAGCCCCCCCTCTGT301GCCGGACGCATCGTGGCCGGCCCACCAGCGCCCCCACAGC351CGCCTATATCGCCGACATCACCCATGGGCAATCGTCGCC451GGGGGACTGTTGGGCGCCACTCCCTTACATGCACCATT501GCTCTATCAGGGCGCACACCTACTACTGGGCTGCTCCG551CGCATAAGGAGGCCCCACCCCATGCCTTGGAGGCGCG601AGCTCCTTCCGGTGGCGACTTCGCTGGAGGGCGCGC651TGTCTTCTTTATCATGCAACTCGTAGGACGGGCGCGC701TCATTTCGGCGAGGACCGCTTTCGCTGGAGCGCGCGC801TGGTCCGCCAACCAAACGTTTCGCGGCGAGGCCGCTTC901TGGATGGCCTTCCCCATTATGATTCTTCCGCTTCCGG951GCCCGCGGTCGCAGGCCATGCTTTTCGCAGGAGGTAGAT1001GACAGCTTCAAGGATCGCTGGCCCGGCCCCATACTTCT151CGTTGCGTGCCGGTTGAGGAGCTCCAGGGGCTTCAGG151AGCTACCAACTCTTTTGAGACCCGCGCTACCTCGGGC151AGCTACCAACTCTTTTTCCGAAGGCCGCAGCTCCAGGG151GCAACCAACATCTTTTTCCGAAGGCCGAGCTCCAGGG151AGCTACCAACTCTTTTTCCGAGCTCAGGGGCTTCAGGG151AGCTACCAACTCTTTTTCCGAGCTCAGGGGCTTCAGGG151AGCTACCAACTGCTTCTAGGGCTCTAGGGGGTTCAGGG151TAGTACCAGCTGCGTG	CTA CTTGGAGCCA GTG GATTCTCTAC GTG CGGTTGCTGG GCT CGCCACTTCG	CCCGTTCTCG	TCTATGCGCA	TGATGCAATT	CTATATGCGT	151
251CTATCGACTACGCGATCATGGCGACCACACCGTCTTG301GCCGGACGCATCGTGGCGGCATCACCGGCGCCACAGG351CGCCTATATCGCCGACATCACCCATCACGGCGGCGTGGTA401GGCTCATGAGCGCTTGTTCGCGCGGGATTGGTGGCG501GGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCT501GGTGCTCACACGGCGCGACCTCCTTACATGCACACT501GGTGCTCACGGTGGCCGACCCATGCCTTGGAGGCG601AGCTCCTTCTATCATGCAACTCGTAGGACAGGTGCCGCG701TCATTTCGGGCAGGACCGTTCGCTGGCGCGCGCGCG751TCGCTTGCGGTATTCGGAATCTTGCAGGCGCCGCGTG901TGGATGGCCTTCCCCATATGATCCTTCTGCTCCGCT901GGACGCTGATCGTCAGGCTATTCAGACAGCTCGGCT901GGACGCTGATCGTCAGGCGATCCTTCGCTCGGC901GGACGCTGATCGTCAGGCGCCCCGCTTACCAGCTTA901GGACGCTGATCGTCAGGCGCCCCGCCTATACCAGCTTA1001GACAGCTTCAAGGATGCATGGCGCGCCCCCATACCTT101CGGTGGCGACGCGTCTTTTTTTGGCGCGCG1151CGTTGGCGGCGTTTTTTTTGGCGCGGC1201ATTAAAGGTCCTCTTCAGAAGCTGCGCGA1301GCAAACAAAAAAACCACCGCTACCAGCGGTAGTTAGCGCCGCGC1401CCAAATACTGTCCTTCTAGAGTCGTGTCTTACCGGGGGGG151AGGTACCAGTGGCGCGGAAAGGGCGCTCGGCAGCGGGGGGGGGGGGGGGGGGGGGGGGG	GTG GATTCTCTAC GTG CGGTTGCTGG GCT CGCCACTTCG	CGCTTCGCTA	CAGTCCTGCT	GGCCGCCGCC	CGACCGCTTT	201
301GCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGG351CGCCTATATCGCCGACATCACCGATGGGGAAGATCGG401GGCGCTATATCGCGCTGGTTCGGCGTGGGCATGGTGGCA501GGGGACTGACGCCTCAACCTACTACTGGGCTCGTTACAT551CGCATAAGGGAGAGCGCCGACCCATGCCCTTGGAGGCGG601AGCTCCTTCCGGTGGGCGACCCATGCCCTTGGAGGCCGA701TCATTTCGGCTAGGACGCCTTCGCTGGGGGCGGCGCG751TCGCTGCGCACCAAACGTTTCGGCGAGAAGCAGGCCG801TGGTCCCGCCACCAAACGTTTCGCGGCGAGCGCGCTGC951GCCCGCGTGCGCGCCTGGCTGTCAGGCAGGTGGCT901GGACGCTGATCGTCAGGCGCCCGGCCGAGCCCGCGCGC1001GACAGCTCAAGGATCGCCCGCCCGGCCCAAGACCGCCA1011CGGGTTGGCATGGATGGCATGTTTTTTTGGCGGC1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGCGGCCA1301GCAAACAAAAAACCACCGCTACCAGCGTGCTCAGG1401CCAAACAAAAAAACCACCGCACCTGCCTCTGCGGTCGGC1511AGCTACCAACTTCTTTTTCCGAAGGTACCGGGGTGTCAGG1521AGCTACCAACTCTTTTTTCCGAGGGGAACGGGGTCGGGC1531AGCTACCAACTCTTTTTCCCGGGGGCCGGCGGTTCAGG1541CTCTGGACGACGCCCTTCTACTGACCGCGCAGGGGCCGGG1551TAGTACCGGTGGCGCGCACCGGGGCGGCGGAACGGA1651GTGACCTAG <td< td=""><td>GTG CGGTTGCTGG GCT CGCCACTTCG</td><td>CCGTCCTGTG</td><td>GCGACCACAC</td><td>CGCGATCATG</td><td>CTATCGACTA</td><td>251</td></td<>	GTG CGGTTGCTGG GCT CGCCACTTCG	CCGTCCTGTG	GCGACCACAC	CGCGATCATG	CTATCGACTA	251
351CGCCTATATCGCCGACATCACCGATGGGGAAGATCGGA401GGCTCATGAGCGCTTGTTCGGCGTGGGAATGGTGGCAA501GGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCT501GGTGCTCAACGGCGCGCGACCCATGCCCTTGAGAGCA601AGCTCCTTCCGGTGGCGCGACCCATGCCCTTGAGAGCA611AGCTCCTTCTTATCATGCAACTCGTAGGACAGGTGCCGCA701TCATTTCGGGAGGACCGCTTTCGCTGGAGCCGCGCCC801TGGCCGCCCAACCAACGTTTCGGCGAGAAGCAGGCCCA901TGGATGGCCTTCCCCATTATGATTCTTCCGCTTCCGG951GCCCGCGTGCAGGCCATGCTGTCCAGGCAGGTGGAA901GGACGCTGATGGTCAGGCGGGGCTCATCCAGGCCCC951GCCCGCGTGATGGTAGCGCGGGGCTCATCCAGCCCGGTG1051GGACGCTGATGGTGCAGAGCCCGCCCTATACTTC1051GGACGCTGATGTTTTGGAGCCTTTTTTTTGGACGGG1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGCGGG1351AGCTACAAAAAACCACCGTACCAGCGGTGGCTAACGG1401CCAAACAAAAAACCACCGACCTGGCTCAACCGGGTGGGC151CGTGCGCAGTGGCGCAACGACCGCTACACGCATAACGG151AGCTACCAACTCTTTTCAGAGCTGGGCGGGGGGGGG151AGCTACCAACTCTTTTCTGGGAACGGGGGGGCGGGGG151CTCTGTAGCATGCTGCCCCAACCGGGGGGGGCGGGGGG151AGCTACCAGTGGCGGGAA<	GCT CGCCACTTCG	GCCACAGGTG	CATCACCGGC	TCGTGGCCGG	GCCGGACGCA	301
401GGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCG451GGGGGACTGTTGGGCGCCATCTCCTTACATGCACCAT501GGTGCTCAACGGCCTCAACCTACTACTGGGTGCTTC551CGCATAAGGAGACGCCCACCCATGCCCTTGAGAGG601AGCTCCTTCGTGGGCGCGGGCATGACTATCGTCGG651TGTCTTTTATCATGCAACTCGTAGGACAGCGCGCCG701TCATTTCGGGAGGACCGCTTTCGCTGGAGCGCGCCC801TGGCCCCGCACCAAACGTTTCGGCGGAGAGCAGGCCG901TGGATGCCTTCGCCAGGCAGCGCGCCCCGCTCCGG951GCCCGCGTGCAGGCCACCTGTCCAGGCAGCTCCGGT951GCCCGCGTGATCGTCAGGCAGCGCGCCCTATACTTT1001GACAGCTTCAAGGATGCCTGCGCGGCCCATACTTT1051GGACGCTGATCGTTGAGAGCCCGGCCTATACTTG1101CGGTTGGCATGGTTCTAGAACCCTTTTTTGGAGGG1201ATTAAAGGCTCCTTTTTGAGATCCTTTTTTTGGCGGC1301GCAAACAAAAAACCACCGTACCGCGGTGGCTTCAG1401CCAAATACAATCCTTCTAGATCGTGCTCAAGCTCAGGA151AGCTACCAGTCTTTTTCTGAAGGTACCACCGAACGAA151AGCTACCAGTCTTTTCTAGAAGGTACACACGACTGGGC151AGCTACCAGTCTTTTCTAGAACCTGCGGGCTGACTGGGC151AGCTACCAGTGGCCGCCCCAAGGTACCGCGGGGCAGG151AGCTACCAGTGGCGGCAGCACGGGTGGGC<		AGATCGGGCT	CCGATGGGGA	GCCGACATCA	CGCCTATATC	351
451GGGGGACTGTTGGGCGCATCTCCTTACATGCACCAT501GGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTC551CGCATAAGGGAGAGCGCCGACCCATGCCCTTGAGAGCC601AGCTCCTTCCGGTGGGGCCGACCCATGCCCTTGAGAGCCG651TGCTTCTTTATCATGCAACTCGCTGGGAGGTGCCGC701TCATTTCGGGAGGACCGCTTTGCTGCGTGAGCGCGCGCCC801TGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCT901TGGATGGCCTCCCCCATTATGATTCTTCCGGTGAGATC901GGACGCTGACGGCCAGCCTGCCAGCCAGGTAGATC901GGACGCTGATGGTCCAGCAGGTGAGATCGCACGCTGA901GGACGCTGATCGTCAGGCAGGTGAGATCGCACGCCGC901GGACGCTGATGGTCCAGCCGCGCCGCCCTATACCTCC901GGACGCTGATGGTCCAGCCGCGCCGCCCTATACCTCC901GGACGCTGCAGCGTCCAGCCGCCCGCGCCTATACCTCC101GCACGCTGCATGGTCCAGCCGCCCCGCCCTATACCTCC1101CGGGTTGGCATGGTTGCTTCTGGCGCGCGCTTCAGC121ATTAAAGGCTCCTTTTTGGAGCCTTTTTTTTGGCGCGC1301GCAAACAAAAAACACCGCTACCTGCTCTGCTTAGC1401CCAAATACAGTCTTTTTTCTCGAGGGTCGGCTTAGCGGC151TAGTACCAGTGGCGCCAGGACGACTACACGAACGGCAA151TAGTACCGCTGGGGCGGCACGGGCGGCCGCGGGCGGCA151TAGTACCAGTGGCGGCAGG <td></td> <td>TGGTGGCAGG</td> <td>GGCGTGGGTA</td> <td>СССТТСТТС</td> <td>GGCTCATGAG</td> <td>401</td>		TGGTGGCAGG	GGCGTGGGTA	СССТТСТТС	GGCTCATGAG	401
501GGTGCTCAACGGCCCCAACCTACTACTGGGCTGCTTC551CGCATAAGGGAGAGCGCCGACCCATGCCCTTGAGAGC601AGCTCCTTCCGGTGGGCCGAGGGCATGACTATCGTCGCG651TGTCTTCTTTATCATGCAACTCGTAGGACGGTGCCGCG701TCATTTCGGGAACGACACCTTGCAGGACGGTGCCGCG801TGGTCCCGCCACCAAACGTTTCGGCGGAGAGCAGGCCCC901TGGATGGCCTTCCCCATTATGATTCTTCCGCTCCGG901TGGATGGCCTTCCCCATTATGATTCTTCCGCTCCGGC901GGACGCGCTGATCGTCACGGCGGTGCCGCGGAGCCGCCCC901GGACGCGCTGATCGTCACGGCGGTGCCGCGGGCTCCGGC901GGACGCGCGATGGATGCGCCGGGCCGCCCTATACCTTC1001GACAGCTTCAAGGATCGCCGCGCGCCCCATACCTTC1011CGGTGGGCATGGATGGCAGCGCGCCCCATACCTTC1021ATTAAAGGCTCCTTTTTGGAGCCTTTTTTTTGGCGCG1031GCAAACAAAAAAACCACCGCTACCAGCGGTGGATGGCA1251AGCTACCAACTCTTTTTCCGAAGGCCAACGCTTCAGG1351AGCTACCAACTCTTTTTTTCCGGAGGTCGGCGGAACGGC1451CTCTGTAGCATTGGGCGAACGCGCTCCCGGGAACGGC1551TAGTTACCGCTTGGAGCAACGGAACGGCAGGACCGCCCC1551TAGTTACCGGTGGGCGAACAGGAGGACCGCCGCGGACGGCGAGC1551TAGTTACCGGTTGGGGCGAACGGGCGCGGCGGCGGACGGCGAAC1601ACAGCCA	TCC TTGCGGCGGC	GCACCATTCC	СТССТТАСАТ	TGGGCGCCAT	GGGGGACTGT	4.51
551CGCATAAGGGAGAGCGCCGACCCATGCCCTTGAGAGCG601AGCTCCTTCCGGTGGGCCGACGCATGACTATCGTCGCG651TGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGG701TCATTTCGGCGAGGACCGCTTTCGCTGGAGCCGCGCGCG801TGGTCCCGCCACCAAACGTTTCGGCGGCGAGCCGCGCGC801TGGCGCGCGACGCGCTGGGCTACGTCTGCGCCGCGCTG901TGGATGGCCTTCCCCATTATGATCTTCCCGCTCCGG951GCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATG1001GACAGCTTCAAGGATCGCTCGCGGCCCCTATCAACCTCG1051GGACCGCTGATCGTCAGGCGATTATGCCGCCTCGGC1051GGACCGCTGATCGTTGTGGAGCCTTTTTTTTGGACGCG1051GGACCGCTGATCTTTTGGAGCCTTTTTTTTGGACGCG1051GGACCGCTGATCTTTTGGAGCCTTTTTTTTGGACGCG1201ATTAAAGGCTCCTTTTTGGAGCCTTTTTTTTGGCGCG1201ATTAAAGGATCTTCTTGAGATCCTTTTTTTTGGCGCG1301GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTCTGT1351AGCTACCAACTCTTTTTTTTTTTTTGGGGGCAACGCTTCTGGC1451CTCTGTAGCATGGGCGAAGAGCGCTCGCGCGGACGGCAGC1551TAGTTACCGGTTGGACGCGCTGGACCGCAGCGGTCGGGC1551TAGTTACCGGTTGGACGCGCCCGGCACGCGGGCGCGCC1551TAGTACCGGTTGGACGCGCCCGGCGGCGCGGGGGCGCAC1551GTGAG	CTA ATGCAGGAGT	CTGCTTCCTA	TACTACTGGG	GGCCTCAACC	GGTGCTCAAC	501
SoliAGENERATING A GENERACCO CONTRACT601AGETCETTECGETGGGCGCGGGCATGACT701TCATTTTEGGCGAGGACCGCTTTEGETGGA751TCGCTTGCGGTATTEGGAATCTTGCACGCC751TCGCTTGCGGACCAAACGTTTCGGCGGCGGA801TGGTCCCGCCACCAAACGTTTCGGCGGCGGG901TGGATGGCCTTCCCCATTATGATTCTTCC901GGCCGCGTGCAGGCCATGCTGTCCAGGCA901GGACGCTGAAGGATCGCTCGCGGCTCTA901GGACGCTGAAGGATCGCCGGCTCCAGGCA901GGACGCTGAAGGATCGCTCGCGCGCCCC901GGACGCTGAAGGACCGCCAGCTCCAGGCA901GGACGCTGAAGGACGCTCAGCGCCGCCCT1001GACAGCTTCAAGGACAGCTGCTCCAGGC1011CGGGTTGGCATGGTTGTGAGGCCCGCGCCC1101CGGGTTGGCATGGTTGTGAGAGCCGCGCC1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTT1301GCAAACAAAAAACCACGCTAACCAGGCT1401CCAAACAAAAAACCACGCTAACCGGCTAG1501CTCTGTGCAGTGGCGATAAGTCGTGTCTAC1501CTGCTGCCAGTGGCGAGGAAGACCTACAC151AGGTACCAGCTTTGGCAGAGGACAGCA151TAGTTACCGGTTTAGCGCAGCGGGCAGG151TAGTACCGGTGGCGCAGGAGCCTGGGCC151CTGTGTGCCAGTGGCGCAGGAGCCTGGGCCAGGC151CTGTGTACGAGTGGCGGCAGGAGCCTGGGCAGGA1651GGAGCAAGG </td <td></td> <td></td> <td></td> <td>ACACCCCCA</td> <td>CCCATAACCC</td> <td>551</td>				ACACCCCCA	CCCATAACCC	551
OOTAGETCUTTECGUIGGEGEGACGUGGAGACTGUGGGCGA651TGTCTTTTATCATGCAACTCGTAGGACAGUGGCGGAC701TCATTTCGGGAAGGACCGCTTTGCCTGGAGCGCGCCC801TGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCT901TGGATGGCCTTCCCCATTATGATTCTTCCGGTCCGGC951GCCCGCGTTGCAGGCCATGCGGGGCCGAAGCCGCCGCG951GGACGCTTCAAGGATCGCCGCGGCCCCCATACCTTC1001GGACGCTGATCGTCACGGCGATTATGCCGCCCGCGTG1051GGACGCTGATCGTCACGGCGACCGCCCCATACCTTC1151CGTTGCGTCGCGGTGCATGGAGCCGGGGCCCCTCGACG1201ATTAAAGGCTCCTTTTGAGATCCTTTTTTTGGCGGG1201ATTAAAGGATCTTCTTGAGATCCTTTTTTTGGCGGGG1201ATTAAAGGATCTTCTTGAGATCCTTTTTTTGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	CCC CACTTATCAC	A TCCTCCCCC	CCCCATCACT	CCTCCCCCCC	ACCTCCTTCC	601
TOTTRATTATEGGCGAGGACCGCTUTGGUGGG701TCATTTTCGGCGAGGACCGTUTGCGUGG801TGGTCCCGCCACCAAACGTTTUTGCCGGCGACA801TGGCGGCCGACGCGCTGGGCTACGTUTGC901TGGATGGCCTTUCCCATTATGATTCTTCTC901GGACGGCTGACGCGCATGCTGTCCAGGCA901GGACGCTGAAGGATCGCTCGCGCGCTTA1001GACAGUTTCAAGGATCGCTCGCGCGCCCCT1051GGACCGCTGATUTTTGTAGGACGCUUCGCGTTG1051GGACCGCTGATUTTTGTGAGGCUUCGCCCCCT1051GGACCGCTGATUTTTGTGAGGCUUCGCCCCCC1051GGACCGCTGATUTTTGGAGGUUCGCCCCCCC1051GGACCGCTGATUTTTGTGAGGUUCGCCCCCCC1051GGACCGCTGCGGGTGGCAGUUCGCCCCCCCCC1051GGACCACAAAAAAACCACCGCTACCAGCGGGG1251ATUAAAGGCTCUTTTTGGAGGUUCGCCAGC1301GUAAACAAAAAAACCACCGCTACCGGCGTGG1401CUAAATACTGTUTTTTTTTTTTCTGGGCGTTG151CTCTGTAGCACUGUCTAATAACCGGACAGG151TAGTTACCGGTUTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	CCA CCCCTCTCCC	CCTCCCCCCA	TCCTACCACA			651
751TCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTGC801TGGTCCCGCCACCAAACGTTTCGCGAGAAGCAGGCCT901TGGATGCCCTCCCCATTATGATTCTTCTCGCTCCCGC901TGGATGCCTTCCCCATTATGATTCTTCTCGCTCCGG901GACAGCTTCAAGGATCGCTCGCGGCGCCCCGCTCCGGC1001GACAGCTTCAAGGATCGCTCGCGGCGCCCCCATACCTTC1011CGGGTTGCATCGTCACGGCGACCGCCGCCCATACCTTC1151CGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACG1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGACGGZ1201ATTAAAGGATCTTCTTGAGATCCTTTTTTTTGGCGGZ1351AGCTACCAACTCTTTTTCGAAGGTAACTGGCTTCAGG1401CCAAATACTGTCCTTCTAGTGTAGCGTAGTTAGGCZ1501CTGCTGCAGTGGCGATAAGTCCTTCTAGTGCTAATCG151AGCTACCAACTCTTTTTCCGAAGGTACTGGCTCAGGC1401CCAAATACTGTCCTTCTAGTGTGAGCTAGGCGCCGCACAT1501CTGCTGCAGTGGCGCAACATAGTCGTCAACCGCTGAGCGTAG151TAGTTACCGGATAAGGCGAACGGCTCACATACCGAACGGC151TAGTACCGCTTGGAGCGAACGGCTCACATCGGCGCGCACG151TAGTACCGCTTGGAGCGAACGGCTCACACGAACGGCAAG151TAGTACCGCTTGGAGCGAACGGAACAGGAGACCGGCCACG151AGGGGGAAACGCCGGAGGGACGGAACAGGAGACCGGACGG151AGGGGGAA	CA GUGUIUIGGG	CCCCCACCAT	TUGIAGGACA	CCACCACCC		701
751TEGETTEGEGETATTEGERATE CTTEGEREGE801TEGETTEGEGEACCAAACGTTTEGECAGCCC851TEGECAGCCACGCCACAACGTTTCGCCAGCAC901TEGATEGECTTCCCCATTATGATTCTTCTCGCTTCCGG951GCCCGCGTGACCCCCATTATGATTCTTCTCGCTCCGGC1001GACAGCTTCAAGGATCGCTCGCGCGCCCCCATACCTTC1051GGACCGCTGATCGTCACGGCGATTTATGCCGCCCCGGGC1101CGGGTTGGCATGGATGTAGGAGCCGCGCCCCATACCTTC1151CGTTGCGTCGCGGTGCATGGAGCCGCGCCCCATACCTTC1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGCGCG1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGCGCGG1351AGCTACCAACTCTTTTTCCGAAGGTAACTGGCTAATCCG1351AGCTACCAACTCTTTTTCCGAAGGTACTGGCTAATCCG1401CCAAATACTGTCCTTCTAGTGTGGGCTATGGCGCATACC1551TAGTTACCGATGGCGCAAGGCGGTCGGCCTAGGCGT1561CTCGTGCCAGTGGCGATAAGTCGGTCTCACGCAACGGG1551TAGTTACCGATTGGAGCGCAACGCGCATACCCGAACGGG1551TAGGTACCAGCTTGGAGCTATGACGGCAGGAGCGCCAGGA1551TAGGTACCAGCTTGGAGCGAGGCGGGAACGGAGAGCGCAGG1551TAGGTACCGTTGGAGCGAACGCGGAACGGAGGGCTAGGGCA1551TAGGCTAGGGTCGGGTATCTTTAAGGGGCGGGGGGGGGA1561GTGAGCGAGCGCGGAACGGGCCTTTTTACG </td <td>GAI GAICGGCCIG</td> <td>GUGUGAUGAI</td> <td>CTTCGCIGGA</td> <td></td> <td>TCATTICGG</td> <td>701 751</td>	GAI GAICGGCCIG	GUGUGAUGAI	CTTCGCIGGA		TCATTICGG	701 751
801TGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCTA901TGGATGGCCTTCCCCATTATGATTCTTCCGCTTCCGG901GGCCGCGTGCAGGCATGCTGTCCAGGCAGGTAGATC1001GACAGCTTCAAGGATCGCCCGCGCGCGCCTATACCTTC1051GGACCGCTGATCGTCACGGCGATTTATGCCGCCCGCGC1151CGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACG1201ATTAAAGGCTCCTTTTGGAGACCTTTTTTTTGGACGGC1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTTGGCGCGC1201ATTAAAGGCTCCTTTTTGGAGCCTTTTTTTTTGGCGCGCG1201ATTAAAGGCTCCTTTTTGGAGCCTTTTTTTTTGGCGCGCG1201ATTAAAGGCTCTTTTTTGGAGCCTTTTTTTTTGGCGCGGC1201ATTAAAGGCTCTTTTTTGGAGGCTTCAGGGGTTGGT1301GCAAACAAAAAAACCACCGCTACCAGCGAGGCTTCAGG1401CCAAATACTGTCCTTCTAGTGTAGCCTAGGCTTCAGG1501CTGCTGCCAGTGGCGATAAGTCGTGTCTCAGCAGGGT151TAGTTACCGGTTGGGCGAAGCGCTACACGAACCGGAGGA151TAGTACCGCTTGGGCGAGGAGGCGAGGAGCGCGCGCGAGGA151TAGTACCGCTTGGGCGAGGAGGCGCAGGGGGCCGGGC151TAGGCGCAGGTGGGCGAGGGCGGGTGGCCGGGGCGAGG151AGGGGGAAACGCCTGGGGCCTTTTACGGAGGCGAGG151AGGGGGAAACGCCTGGGGCCTTTTACGGAGGGGGGG151AGGGGAAAC <td>AAG UUTTUGTUAU</td> <td>CTCGCTCAAG</td> <td>TTGCACGUU</td> <td>TATTCGGAAT</td> <td>TUGUTTGUGG</td> <td>/51</td>	AAG UUTTUGTUAU	CTCGCTCAAG	TTGCACGUU	TATTCGGAAT	TUGUTTGUGG	/51
851TGGCGGCCGACGCGGCTGGGCTAGCTCTTGCTGGCTGCCA901TGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCGGG951GCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATG1001GACAGCTTCAAGGATCGCTGGCGGCCCCTATACCTTC1051GGACCGCTGATCGTACGGCGATTTATGCCGCCCGGGC1101CGGGTTGGCATGGATTGTAGGCCGGGCCCTATACCTTG1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGACGGC1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGCGCGC1201ATCAAAGGATCTTCTTGAGATCCTTTTTTTTGGCGCGCG1201ACAAACAAAAAAACCACCGCTACCAGCGGTGGTTCAGC1301GCAAACAAAAAAACCACCGCTACCAGCGGTGTTAGCCG1401CCAAATACTGTCCTTTTTTCCGAAGGTACCAGCTTAGGCCG1501CTGCTGCCAGTGGCGGATAAGTCGTGCTGCTGGCAACGGC1501CTGCTGCCAGTGGCGGCAAGCGGTCGGGCTGAACGGC1501CTGCTGCCAGTGGAGCTATGAGGGGCAACGCGCTTCCG1551TAGTTACCGGATAAGGCCACGAACCAGCACGAACAGGA1551TAGTACCGGTGGAGCTATGAGGGGCAACGCGCTGGGC1551AGGGGGAAACGCCTGGGGCCTTTTACGGCGGGGGGG1561GTGAGCTATGAGGGCGAGGCCTTTTTTTCCGGGGGGGG1561GAGGGGAAACGCCTGGGGCCTGCTGGGCCTGGGGGGGG1561AAAAACGCCAGCAACGGGGCCTTTTTACGGCCGGGGGGGGG1501G	ATT ATCGCCGGCA	GCAGGCCATT	TCGGCGAGAA	ACCAAACGTT	TGGTCCCGCC	801
901TGGATGGCCT TCCCCATTATGATTCTTCTC GCTTCCGG951GCCCGCGTTG CAGGCCATGC TGTCCAGGCA GGTAGATG1001GACAGCTTCA AGGATCGCTC GCGGCTCTTA CCAGCCT1051GGACCGCTGA TCGTCACGGC GATTATGCC GCCTCGGG1101CGGGTGGCA TGGATTGTAG GCGCCGCCCT ATACCTTG1151CGTTGCGTCG CGGTGCATGG AGCCGGGCCA CCTCGACG1201ATTAAAGGCT CCTTTTGGAG CCTTTTTTT TGGACGGT1351AGCAACAAAA AAACCACCGC TACCAGCGGT GGTTGGT1351AGCTACCAAC TCTTTTCCG AAGGTAACTG GCTTCAGG1401CCAAATACTG TCCTTCTAGT GTAGCCGTAG TTAGGCC1451CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCC1551TAGTTACCGG ATAAGGCGCA GCGGTCGGGC TGAACGG1651GTGAGCTATG AGAAAGCGCC ACGACTACAC CGAACTG1551TAGTTACCGG ATAAGGCGCA GCGGTCGGGC TGAACGGC1661ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTG1651GTGAGCTATG AGAAAGCGCC ACGCTTCCCG AAGGAGG1701TATCCGGTAA GCGGCAGGGT CGGAACAGGA GAGCGCAC1751AGGGGAAAC GCCTGGTATC TTTATAGTCC TGTCGGGC1751AGGGGAAAC GCCTGGTATC TTTATAGTCC TGCCGGC1801GACTTGAGCG TCGATTTTT TGGAGCGGC CTTTTACGG1951GTATTACCGC CTTTGAGTGA CCGAGAAGCG GAAGAGC1951GTATTACCGC CTTTGAGTGA CCGCATAGTT AAGCCAGC2001GAGCGCAGCG AGTCAGTGAG CGAGAAGCG GAAGAGC201CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGC2101CAGTACAATC TGCTCTGGTCC CGGCATCC2201CTGACGCGC CTGACGGCC TGCCTGCCC CCGACACC2201CTGACGCGCC CTGACGGCC TGCCATGTT CAGAGGTC2301ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATACTC	CGC GACGCGAGGC	TGGCGTTCGC	TACGTCTTGC	CGCGCTGGGC	TGGCGGCCGA	851
951GCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATC1001GACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCT1051GGACCGCTGATCGTCACGGCGATTATGCCGCCTCGGC1151CGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGCGC1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGACGGC1251ATCAAAGGATCTTCTTGAGATCCTTTTTTTGGACGGC1301GCAAACAAAAAACCACCGCTACCAGCGTGCTTCAGC1401CCAAATACTGTCCTTCTAGTGTAGCCGTAGGTTAGGCC1401CCAAATACTGTCCTTCTAGTGTAGCCGTGGCTACGC1501CTGCTGCCAGTGGCGCAAAGTCGTGTCTTACCGGGTCGGC1501CTGCTGCCAGTGGCGCAAGGCGGTCGGCTGAGCGCA1501CTGCTGCCAGTGGCGGAAAGTCGCTGCGCGGAACGGC1501CTGCTGCCAGTGGCGCAAGGGGGTCGGCTGAACGGC1551TAGTTACCGGATAGGGGAAAGGACCTACACCGAACGGC1651GTGAGCTATGAGGGGAAACGCGCAGGACGACCTACAC1701TATCCGGTAAGCGCAGGGACGGAACAGGAGAGCGCAGC1751AGGGGAAACGCCTGGTGTTTTTGTCTGGCGTTCCTGGG1801GACTTGAGCGTCGTGTTTTCCTGGCGGCACCGCAAGGA1901TTTGCTCACATGTCTCTGTCTGGCGCACGGAGCGCAGC1951GTATTACCGCCTTTGAGTGACGGGAAACGCCCGCAAGCG2001GAGCGCAGCGACTGACGGCCCTGACGGCCCCGAAACGC2101CAGTACAATC <td>GCG GCA'I'CGGGA'I'</td> <td>GCTTCCGGCG</td> <td>GATTCTTCTC</td> <td>TCCCCATTAT</td> <td>TGGATGGCCT</td> <td>901</td>	GCG GCA'I'CGGGA'I'	GCTTCCGGCG	GATTCTTCTC	TCCCCATTAT	TGGATGGCCT	901
1001GACAGCTTCA AGGATCGCTC GCGGCTCTTA CCAGCCT1051GGACCGCTGA TCGTCACGGC GATTTATGCC GCCTCGGG1101CGGGTTGCA TGGATTGTAG GCGCCGCCCT ATACCTT1151CGTTGCGTCG CGGTGCATGG AGCCGGGCCA CCTCGACG1201ATTAAAGGCT CCTTTTGGAG CCTTTTTTT TGGACGG1251ATCAAAGGAT CTTCTTGAGA TCCTTTTTT CTGCGCG1301GCAAACAAAA AAACCACCGC TACCAGCGGT GGTTTGT1351AGCTACCAAC TCTTTTCCG AAGGTAACTG GCTTCAGG1401CCAAATACTG TCCTTCTAGT GTAGCCGTAG TTAGGCC1451CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCC1501CTGCTGCCAG TGGCGATAAG TCGTGTCTTA CCGGGTT1551TAGTTACCGG ATAAGGCGCA GCGGTCGGGC TGAACGGG1601ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTG1651GTGAGCTATG AGAAAGCGCC ACGCTTCCCG AAGGAGG1701TATCCGGTAA GCGGCAGGGT CGGAACAGGA GAGCGCAG1751AGGGGGAAAC GCCTGGTATC TTTATAGTCC TGTCGGG1801GACTTGAGCG TCGATTTTG TGATGCTGCT CAGGGGG1901TTTTGCTCAC ATGTTCTTC CTGCGTATC CCCTGAT1951GTATTACCGC ATGTGTGAG CGAGGAAGCG GAAGAGCG2001GAGCGCAGCG AGTCAGTGA GCGCATAGTT ACCCGC2101CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAG211CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAG212CTGACGCGCC CTGACGGCT TGTCTGCCC CGGCATCC2201CTGACGCGCC CTGACGGCT TGTCTGCTCC CGCCATC2201CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATC2301ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT	GAC GACCATCAGG	GGTAGATGAC	TGTCCAGGCA	CAGGCCATGC	GCCCGCGTTG	951
1051GGACCGCTGA TCGTCACGGC GATTTATGCC GCCTCGGG1101CGGGTTGGCA TGGATTGTAG GCGCCGCCCT ATACCTTG1151CGTTGCGTCG CGGTGCATGG AGCCGGGCCA CCTCGACG1201ATTAAAGGCT CCTTTTGGAG CCTTTTTTT TGGACGGG1251ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGG1301GCAAACAAAA AAACCACCGC TACCAGCGGT GGTTTGT1351AGCTACCAAC TCTTTTCCG AAGGTAACTG GCTTCAGG1401CCAAATACTG TCCTTCTAGT GTAGCCGTAG TTAGGCCG1451CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCG1551TAGTTACCGG ATAAGGCGCA GCGGTCGGGC TGAACGGG1601ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTGG1651GTGAGCTATG AGAAAGCGCC ACGCTTCCCG AAGGGAGG1701TATCCGGTAA GCGGCAGGGT CGGAACAGGA GAGCGCAG1751AGGGGGAAAC GCCTGGTATC TTTATAGTCC TGTCGGGG1801GACTTGAGCG TCGATTTTG TGATGCTCGT CAGGGGGG1901TTTTGCTCAC ATGTCTTTC CTGCGTATC CCCTGATG1951GTATTACCGC CTTTGAGTGA GCTGATACCG CTCGCCG2001GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCG2011CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGC2151ATCGCTACGT GACTGGGCT TGTCTGCCC CGGCATCC2201CTGACGCGC CTGACGGGCT TGTCTGCCC CGGCATCC2201CTGACGCCC CTGACGGGCT TGTCTGCTCC CGCCACC2201ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT2301ACCGAAACGC GCGAGGCAGG GGGAATCCA GATAACT	AAC TTCGATCATT	CCAGCCTAAC	GCGGCTCTTA	AGGATCGCTC	GACAGCTTCA	1001
1101CGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTC1151CGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACG1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGACGGG1251ATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGG1301GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTT1351AGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGG1401CCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCC1451CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCG1501CTGCTGCCAGTGGCGATAGTCGTGTCTTACCGGGTTG1551TAGTTACCGGATAAGGCGACGGCTCGGGCTGAACGGG1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAGG1801GACTTGAGCGTCGATTTTTTGATGCTCGTCAGGGGGG1951GTATTACCGCCTTTGAGTGACGGGTAATCCCTCGCGGC1951GTATTACCGCAGCGAACGGGGCCTTTTTTCCCTGATGCCGCATAGTT201CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCACC2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCACC2251GCTGTGACCGCTGCCGGGGCTGTCTGCTCCCGGCATCC2301ACCGAACGCGCGAGGCAGGGGGAATTCCAGATACCT	CGA GCACATGGAA	GCCTCGGCGA	GATTTATGCC	TCGTCACGGC	GGACCGCTGA	1051
1151CGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACG1201ATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGACGG1251ATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGG1301GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTGTT1351AGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGG1401CCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCC1451CTCTGTAGCACCGCCTACATACCTCGCTGTGCTAATCG1501CTGCTGCCAGTGGCGATAGTCGTGTCTTACCGGGGTTG1551TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGG1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACGGG1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGGGAAC1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAGG1801GACTTGAGCGTCGATTTTTTGAGGGGGGGCTGGCGGGG1901TTTTGCTCACATGTTCTTT<	GTC TGCCTCCCCG	ATACCTTGTC	GCGCCGCCCT	TGGATTGTAG	CGGGTTGGCA	1101
1201ATTAAAGGCTCCTTTTTGGAGCCTTTTTTTTGGACGGA1251ATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGG1301GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTT1351AGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGG1401CCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCA1451CTCTGTAGCACCGCCTACATACCTCGGTCGCTAATCG1551TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGG1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACGGG1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGGGA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAGG1751AGGGGGAAACGCCTGGTATCTTTTAAGTCCTGTCGGGG1801GACTTGAGCGTCGATTTTTTTGGAGCTGGCCTGGCGGG1901TTTTGCTCACATGTTCTTTCCTGCGGTAACCGCTCCGGGG1951GTATTACCGCCTTTGAGTGAGCGGAAAGCGGAAGAGCG2001GAGCGCAGCGAGCCAACTCTGTGCGGAAACGCGCGCAACGC2101CAGTACAATCTGCTTGAGTGACGGCAAGCGGAGCCGCC2251GCTGTGACCGCTGCGGGGACTGCAGGGGCACTGCAGGGTCA2301ACCGAAACGCGCGAGGCAGGGGGAAATCCAGATAACT	CTG AACCGATACA	CCTCGACCTG	AGCCGGGCCA	CGGTGCATGG	CGTTGCGTCG	1151
1251ATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGCG1301GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGT1351AGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGG1401CCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCA1451CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCG1501CTGCTGCCAGTGGCGATAAGTCGTGTGCTACCGGGGTGGGC1551TAGTTACCGGATAAGGCGCAGCGGCTCGGGCTGAACGGG1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGG1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGGGA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAGG1751AGGGGGAAACGCCTGGTATCTTTTATAGTCCTGTCGGGG1801GACTTGAGCGTCGATTTTTTTGATGCTCGTCAGGGGGG1901TTTTGCTCACATGTTCTTTCCTGCGGTATCCTCCGGGG1951GTATTACCGCCTTTGAGTGACGGGAAAGCGGAAGAGCG2001GAGCGCAGCGAGCCAACTCTGTGCGGAAGCGGAAGAGCG2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGC2151ATCGCTACGTGACTGGGTCATGCCTGCCCCCGACACG2201CTGACGCGCCTGACGGGGCTGCCTGCTCCCGGCATCG2251GCTGTGACCGCCCGAGGGCAGGGGAATTCCAGATAACT	ACC GGTAGAAAAG	TGGACGGACC	CCTTTTTTTT	CCTTTTGGAG	ATTAAAGGCT	1201
1301GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTT1351AGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGG1401CCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCA1451CTCTGTAGCACCGCCTACATACCTCGCTGTGCTAATCG1501CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTG1551TAGTTACCGGATAAGGCGCAGCGCTCGGGCTGAACGGG1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGGGA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAGG1751AGGGGGAAACGCCTGGTATCTTTTATAGTCCTGTCGGGG1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG1901TTTTGCTCACATGTTCTTTCCTGCGGTATCCTCGCGGG1951GTATTACCGCCTTTGAGTGAGCGGAAGCGGAAGAGCG2001GAGCGCAGCGAGCCAACTCTGTGCGGTATTCACACCGGG2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGC2201CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCG2251GCTGTGACCGTCTCCGGGAGGGAATTCCAGATACCT2301ACCGAAACGCGCGAGGCAGGGGGAATTCCAGATAACT	TAA TCTGCTGCTT	CTGCGCGTAA	TCCTTTTTTT	CTTCTTGAGA	ATCAAAGGAT	1251
1351AGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGG1401CCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCA1451CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCG1501CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTC1551TAGTTACCGGATAAGGCGAAGCGCTACACCGAACTGA1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGGAA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAC1751AGGGGGAAACGCCTGGTATCTTTTATAGTCCTGTCGGGT1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGC1901TTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATT1951GTATTACCGCCTTTGAGTGAGCGGTATCTAACACGCG2001GAGCGCAGCGAGCCAACTCTGTGCGGTAAGCGGAACAGC2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGC2201CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCC2251GCTGTGACCGTCTCCGGGAGGGAAATCCAGATAACT2301ACCGAAACGCGCGAGGCAGGGGAATTCCAGATAACT	TTG CCGGATCAAG	GGTTTGTTTG	TACCAGCGGT	AAACCACCGC	GCAAACAAAA	1301
1401CCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCA1451CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCG1501CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTG1551TAGTTACCGGATAAGGCGAAGCGCTACACCGAACTGA1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAG1751AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGG1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG1901TTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATT1951GTATTACCGCCTTTGAGTGAGCGGTATCTAAGCCAGG2001GAGCGCAGCGAGTCAGTGAGCGGGTATTCACACCGGC2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGG2201CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCG2201CTGACGCGCCCTGACGGGGGCTGCATGTGTCAGAGGT2301ACCGAAACGCGCGAGGCAGGGGAATTCCAGATAACT	CAG AGCGCAGATA	GCTTCAGCAG	AAGGTAACTG	TCTTTTTCCG	AGCTACCAAC	1351
1451CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCO1501CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTT1551TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGC1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAC1751AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGT1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGA1951AAAAACGCCAGCAACGCGGCCTTTTAACGGTTCCTGGC1951GTATTACCGCCTTTGAGTGAGCGGTATTCACACCGCA2001GAGCGCAGCGAGTCAGTGAGCGGGTATTCACACCGCA2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGT2151ATCGCTACGTGACTGGGTCATGCCTGCCCCCGACACC2201CTGACGCGCCCTGACGGGGCTGTCTGCTCCCGGCATCC2201ACCGAAACGCGCGAGGCAGGGGAATTCCAGATAACT2301ACCGAAACGCGCGAGGCAGGGGAATTCCAGATAACT	ACC ACTTCAAGAA	TTAGGCCACC	GTAGCCGTAG	TCCTTCTAGT	CCAAATACTG	1401
1501CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTG1551TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGC1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAG1751AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGT1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGA1951AAAAACGCCAGCAACGCGGCCTTTTACGGTTCCTGGGT1951GTATTACCGCCTTTGAGTGAGCGGTATTCACACCGCZ2001GAGCGCAGCGAGTCAGTGAGCGGGTATTCACACCGCZ2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGC2201CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCC2201ACCGAAACGCGCGAGGCAGGGGGAAATCCAGATAACT2301ACCGAAACGCGCGAGGCAGGGGAAATCCAGATAACT	CTG TTACCAGTGG	GCTAATCCTG	ACCTCGCTCT	CCGCCTACAT	CTCTGTAGCA	1451
1551TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGG1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAG1751AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGT1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG1851AAAAACGCCAGCAACGCGGCCTTTTACGGTTCCTGGG1901TTTTGCTCACATGTTCTTCCTGCGTTATCCCCTGATT1951GTATTACCGCCTTTGAGTGAGCGGAAACGGAAGAGCG2001GAGCGCAGCGAGTCAGTGAGCGGGTATTCACACCGCZ2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGT2151ATCGCTACGTGACTGGGTCATGCCTGCCCCCGACACG2201CTGACGCGCCCTGACGGGGCTGTCTGCTCCCGGCATCC2301ACCGAAACGCGCGAGGCAGGGGAATTCCAGATAACT	GGA CTCAAGACGA	CCGGGTTGGA	TCGTGTCTTA	TGGCGATAAG	CTGCTGCCAG	1501
1601ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAG1751AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGG1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGG1851AAAAACGCCAGCAACGCGGCCTTTTACGGTTCCTGGG1901TTTTGCTCACATGTTCTTCCTGCGTTATCCCCTGATT1951GTATTACCGCCTTTGAGTGAGCTGATACCGGCGCGCG2001GAGCGCAGCGAGTCAGTGAGCGAGAGAGCGGAAGAGCG2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGC2151ATCGCTACGTGACTGGGTCATGCCTGCCCCCGCATCG2201CTGACGCGCCCTGACGGGGCTTCCTGCTCCCGGCATCG2301ACCGAAACGCGCGAGGCAGGGGAATTCCAGATAACT	GGG GTTCGTGCAC	TGAACGGGGG	GCGGTCGGGC	ATAAGGCGCA	TAGTTACCGG	1551
1651GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAG1751AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGT1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGG1851AAAAACGCCAGCAACGCGGCCTTTTACGGTTCCTGGG1901TTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATT1951GTATTACCGCCTTTGAGTGAGCTGATACCGCTCCCGGG2001GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCG2051TTTTCTCCTTACGCATCTGTGCGGTATTCACACCGCZ2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGC2201CTGACGCGCCCTGACGGGCTTGCTGTGCTCCCGGCATCG2251GCTGTGACCGTCTCCGGGAGGGGAATTCCAGATAACT2301ACCGAAACGCGCGAGGCAGGGGGAATTCCAGATAACT	AGA TACCTACAGC	CGAACTGAGA	CGACCTACAC	TTGGAGCGAA	ACAGCCCAGC	1601
1701TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAG1751AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGT1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGG1851AAAAACGCCAGCAACGCGGCCTTTTACGGTTCCTGGG1901TTTTGCTCACATGTTCTTCCTGCGTTATCCCCTGATT1951GTATTACCGCCTTTGAGTGAGCTGATACCGCTCCCGGG2001GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCG2051TTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGGZ2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGC2151ATCGCTACGTGACTGGGGCATGGCTGCGCCCCGACACG2201CTGACGCGCCCTGACGGGGCTTGTCTGCTCCCGGCATCG2251GCTGTGACCGTCTCCGGGAGGGGAATTCCAGATAACT2301ACCGAAACGCGCGAGGCAGGGGGAATTCCAGATAACT	AAA GGCGGACAGG	AAGGGAGAAA	ACGCTTCCCG	AGAAAGCGCC	GTGAGCTATG	1651
1751AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGG1801GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG1851AAAAACGCCAGCAACGCGGCCTTTTACGGTTCCTGGG1901TTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGAT1951GTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCGGG2001GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCG2051TTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGGZ2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGG2151ATCGCTACGTGACTGGGGTCATGGCTGCGCCCCGACACC2201CTGACGCGCCCTGACGGGGCTTGTCTGCTCCCGGCATCG2251GCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGT2301ACCGAAACGCGCGAGGCAGGGGGAATTCCAGATAACT	CGA GGGAGCTTCC	GAGCGCACGA	CGGAACAGGA	GCGGCAGGGT	TATCCGGTAA	1701
1801GACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGG1851AAAAACGCCAGCAACGCGGCCTTTTACGGTTCCTGGG1901TTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGAT1951GTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGG2001GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCG2051TTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCZ2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGG2151ATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACG2201CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCG2251GCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGT2301ACCGAAACGCGCGAGGCAGGGGGAATTCCAGATAACT	TTT CGCCACCTCT	TGTCGGGTTT	TTTATAGTCC	GCCTGGTATC	AGGGGGAAAC	1751
1851AAAAACGCCAGCAACGCGGCCTTTTACGGTTCCTGGG1901TTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGAT1951GTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGG2001GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCG2051TTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCG2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGG2151ATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACG2201CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCG2251GCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGT2301ACCGAAACGCGCGAGGCAGGGGGAATTCCAGATAACT	GCG GAGCCTATGG	CAGGGGGGGGG	TGATGCTCGT	TCGATTTTTG	GACTTGAGCG	1801
1901TTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGAT1951GTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGG2001GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCG2051TTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCG2101CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGT2151ATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACG2201CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCG2251GCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTT2301ACCGAAACGCGCGAGGCAGGGGGAATTCCAGATAACT	ССФ 0/10001/1100 ССФ ФФФССФСССС	TTCCTGCCCT		CCARCECCEC	AAAAACCCCA	1851
1951ITTIGETERE ALGITETITE CIGEGITATE CECTIGAT1951GTATTACCGC CTTTGAGTGA GCTGATACCG CTCGCCGC2001GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCC2051TTTTCTCCTT ACGCATCTGT GCGGTATTTC ACACCGCZ2101CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAG'2151ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACC2201CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCC2251GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGT'2301ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT'	ССІ ІІІССІООСС ТСТ СТССАТААСС					1001
2001GAGCGCAGCG AGTCAGTGAG GCIGATACCG CICGCCG2001GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCG2051TTTTCTCCTT ACGCATCTGT GCGGTATTTC ACACCGC22101CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGG2151ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACG2201CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCG2251GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGT2301ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT	CAC CCCAACCACC	CUCIGATICI	CIGCGIIAIC	CTTTCACTCA	CTATTACCCC	1051
2001GAGEGEAGEG AGTEAGTEAGTEAG EGAGGAAGEG GAAGAGEG2051TTTTCTCCTT ACGCATCTGT GCGGTATTTC ACACCGC22101CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAG2151ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACG2201CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCG2251GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGT'2301ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT'	CCC TCATCCCCTA	CIUGUUGUAG	GCIGAIACCG	ACTICACTOR	GIATIACCGC	2001
2001ITTECTECTT ACGEATETGT GEGGTATITE ACACEGEA2101CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAG'2151ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACC2201CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCC2251GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGT'2301ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT'	JUC IGAIGUGGIA	GAAGAGCGCC		AGICAGIGAG		2001
2101CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAG2151ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACC2201CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCC2251GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGT2301ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT	AAI GGIGCACICI	ACACCGCAAI	GCGGIAIIIC	ACGCAICIGI		2001
2151ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACC2201CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCC2251GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGT2301ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT	TAT ACACTUGUT	AAGUCAGTAT	CCGCATAGTT	TGCTCTGATG		2101
2201CTGACGCGCCCTGACGCGCCCGGCATCC2251GCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGT2301ACCGAAACGCGCGAGGCAGGGGGAATTCCAGATAACT	CCG CCAACACCCG	CCGACACCCG	TGGCTGCGCC	GACTGGGTCA	ATCGCTACGT	2151
2251 GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGT 2301 ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT	CGC TTACAGACAA	CGGCATCCGC	TGTCTGCTCC	CTGACGGGCT	CTGACGCGCC	2201
2301 ACCGAAACGC GCGAGGCAGG GGGAATTCCA GATAACT	TTT CACCGTCATC	CAGAGGTTTT	CTGCATGTGT	TCTCCGGGAG	GCTGTGACCG	2251
	ТСС ТАТААТСТАТ	GATAACTTCG	GGGAATTCCA	GCGAGGCAGG	ACCGAAACGC	2301
2351 GCTATACGAA GTTATGGTAC CGCGGCCGCG TAGAGGA	TCT GTTGATCAGC	TAGAGGATCT	CGCGGCCGCG	GTTATGGTAC	GCTATACGAA	2351
2401 AGTTCAACCT GTTGATAGTA CTTCGTTAAT ACAGATG	TAG GTGTTGGCAC	ACAGATGTAG	CTTCGTTAAT	GTTGATAGTA	AGTTCAACCT	2401
2451 CATGCATAAC TATAACGGTC CTAAGGTAGC GACCTAG	GTA TCGATAATAC	GACCTAGGTA	CTAAGGTAGC	TATAACGGTC	CATGCATAAC	2451
2501 GACTCACTAT AGGGGAATTG TGAGCGGATA ACAATTCO	CCC TCTAGAAATA	ACAATTCCCC	TGAGCGGATA	AGGGGAATTG	GACTCACTAT	2501
2551 ATTTTGTTTA ACTTTAAGAA GGAGATATAC ATATGAG	GCC TCGGATCCTG	ATATGAGGCC	GGAGATATAC	ACTTTAAGAA	ATTTTGTTTA	2551
2601 TAAAACGACG GCCAGTGAAT TCCCCGGGAA GCTTCGCC	CAG GGTTTTCCCA	GCTTCGCCAG	TCCCCGGGAA	GCCAGTGAAT	TAAAACGACG	2601
2651 GTCGAGCTCG ATATCGGTAC CAGCGGATAA CAATTTC	ACA TCCGGATCGC	CAATTTCACA	CAGCGGATAA	ATATCGGTAC	GTCGAGCTCG	2651
2701 GAACGCGTCT CGAGAGATCC GGCTGCTAAC AAAGCCCC	GAA AGGAAGCTGA	AAAGCCCGAA	GGCTGCTAAC	CGAGAGATCC	GAACGCGTCT	2701
	CCC CTTGGGGGCCT	AGCATAACCC	AGCAATAACT	GCCACCGCTG	GTTGGCTGCT	2751
2751 GTTGGCTGCT GCCACCGCTG AGCAATAACT AGCATAA	TCT AATTGGACTA	AAACCCATCT	TTTTTGGTTT	CTTGAGGGGT	CTAAACGGGT	2801
2751 GTTGGCTGCT GCCACCGCTG AGCAATAACT AGCATAAC 2801 CTAAACGGGT CTTGAGGGGT TTTTTGGTTT AAACCCA		AATTCCCCTA	GCTTTTTTTT	TAATGAGCGG	GTAGCCCGCC	2851
 2751 GTTGGCTGCT GCCACCGCTG AGCAATAACT AGCATAAC 2801 CTAAACGGGT CTTGAGGGGGT TTTTTGGTTT AAACCCA 2851 GTAGCCCGCC TAATGAGCGG GCTTTTTTTT AATTCCCC 	CTA TTTGTTTATT	CATGAGACAA	TGTATCCGCT	CATTCAAATA	TTTCTAAATA	2901
 2751 GTTGGCTGCT GCCACCGCTG AGCAATAACT AGCATAAG 2801 CTAAACGGGT CTTGAGGGGT TTTTTGGTTT AAACCCA 2851 GTAGCCCGCC TAATGAGCGG GCTTTTTTTT AATTCCCG 2901 TTTCTAAATA CATTCAAATA TGTATCCGCT CATGAGAG 	CTA THIGHHIATT CAA TAACCCTGAT					

1	ATCAACGTCT	CATTTTCGCC	AAAAGTTGGC	CCAGATCTAT	GTCGGGTGCG
51	GAGAAAGAGG	TAATGAAATG	GCACCTAGGT	ATCGATAATA	CGACTCACTA
101	TAGGGGAATT	GTGAGCGGAT	AACAATTCCC	CTCTAGAAAT	AATTTTGTTT
151	AACTTTAAGA	AGGAGATATA	CATATGAGGC	CTCGGATCCT	GTAAAACGAC
201	GGCCAGTGAA	TTCCCCGGGA	AGCTTCGCCA	GGGTTTTTCCC	AGTCGAGCTC
251	GATATCGGTA	CCAGCGGATA	ACAATTTCAC	ATCCGGATCG	CGAACGCGTC
301	TCGAGAGATC	CGGCTGCTAA	CAAAGCCCGA	AAGGAAGCTG	AGTTGGCTGC
351	TGCCACCGCT	GAGCAATAAC	TAGCATAACC	CCTTGGGGCC	TCTAAACGGG
401	TCTTGAGGGG	TTTTTTGGTT	TAAACCCATG	TGCCTGGCAG	ATAACTTCGT
451	ATAATGTATG	CTATACGAAG	TTATGGTACC	GCGGCCGCGT	AGAGGATCTG
501	TTGATCAGCA	GTTCAACCTG	TTGATAGTAC	GTACTAAGCT	CTCATGTTTC
551	ACGTACTAAG	CTCTCATGTT	TAACGTACTA	AGCTCTCATG	TTTAACGAAC
601	TAAACCCTCA	TGGCTAACGT	ACTAAGCTCT	CATGGCTAAC	GTACTAAGCT
651	CTCATGTTTC	ACGTACTAAG	CTCTCATGTT	TGAACAATAA	AATTAATATA
701	AATCAGCAAC	TTAAATAGCC	TCTAAGGTTT	TAAGTTTTAT	AAGAAAAAAA
751	AGAATATATA	AGGCTTTTAA	AGCTTTTAAG	GTTTAACGGT	TGTGGACAAC
801	AAGCCAGGGA	TGTAACGCAC	TGAGAAGCCC	TTAGAGCCTC	TCAAAGCAAT
851	TTTGAGTGAC	ACAGGAACAC	TTAACGGCTG	ACAGAATTAG	CTTCACGCTG
901	CCGCAAGCAC	TCAGGGCGCA	AGGGCTGCTA	AAGGAAGCGG	AACACGTAGA
951	AAGCCAGTCC	GCAGAAACGG	TGCTGACCCC	GGATGAATGT	CAGCTGGGAG
1001	GCAGAATAAA	TGATCATATC	GTCAATTATT	ACCTCCACGG	GGAGAGCCTG
1051	AGCAAACTGG	CCTCAGGCAT	TTGAGAAGCA	CACGGTCACA	CTGCTTCCGG
1101	TAGTCAATAA	ACCGGTAAAC	CAGCAATAGA	CATAAGCGGC	TATTTAACGA
1151	CCCTGCCCTG	AACCGACGAC	CGGGTCGAAT	TTGCTTTCGA	ATTTCTGCCA
1201	TTCATCCGCT	TATTATCACT	TATTCAGGCG	TAGCAACCAG	GCGTTTAAGG
1251	GCACCAATAA	CTGCCTTAAA	AAAATTACGC	CCCGCCCTGC	CACTCATCGC
1301	AGTACTGTTG	TAATTCATTA	AGCATTCTGC	CGACATGGAA	GCCATCACAA
1351	ACGGCATGAT	GAACCTGAAT	CGCCAGCGGC	ATCAGCACCT	TGTCGCCTTG
1401	CGTATAATAT	TTGCCCATGG	TGAAAACGGG	GGCGAAGAAG	TTGTCCATAT
1451	TGGCCACGTT	TAAATCAAAA	CTGGTGAAAC	TCACCCAGGG	ATTGGCTGAG
1501	ACGAAAAACA	TATTCTCAAT	AAACCCTTTA	GGGAAATAGG	CCAGGTTTTC
1551	ACCGTAACAC	GCCACATCTT	GCGAATATAT	GTGTAGAAAC	TGCCGGAAAT
1601	CGTCGTGGTA	TTCACTCCAG	AGCGATGAAA	ACGTTTCAGT	TTGCTCATGG
1651	AAAACGGTGT	AACAAGGGTG	AACACTATCC	CATATCACCA	GCTCACCGTC
1701	TTTCATTGCC	ATACGGAATT	CCGGATGAGC	ATTCATCAGG	CGGGCAAGAA
1751	TGTGAATAAA	GGCCGGATAA	AACTTGTGCT	TATTTTTCTT	TACGGTCTTT
1801	AAAAAGGCCG	TAATATCCAG	CTGAACGGTC	TGGTTATAGG	TACATTGAGC
1851	AACTGACTGA	AATGCCTCAA	AATGTTCTTT	ACGATGCCAT	TGGGATATAT
1901	CAACGGTGGT	ATATCCAGTG	ATTTTTTTCT	CCATTTTAGC	TTCCTTAGCT
1951	CCTGAAAATC	TCGATAACTC	AAAAATACG	CCCGGTAGTG	ATCTTATTTC
2001	ATTATGGTGA	AAGTTGGACC	CTCTTACGTG	CCGATCAACG	TCTCATTTTC
2051	GCCAAAAGTT	GGCCCAG			

F.2.3. pDC

1951

2001

2051

CTATGTCGGG	TGCGGAGAAA	GAGGTAATGA	AATGGCACCT	AGGTATCGAT
GGCTTTACAC	TTTATGCTTC	CGGCTCGTAT	GTTGTGTGGA	ATTGTGAGCG
GATAACAATT	TCACACAGGA	AACAGCTATG	ACCATGATTA	CGAATTTCTA
GAAATAATTT	TGTTTAACTT	TAAGAAGGAG	ATATACATAT	GAGGCCTCGG
ATCCTGTAAA	ACGACGGCCA	GTGAATTCCC	CGGGAAGCTT	CGCCAGGGTT
TTCCCAGTCG	AGCTCGATAT	CGGTACCAGC	GGATAACAAT	TTCACATCCG
GATCGCGAAC	GCGTCTCGAG	ACTAGTTCCG	TTTAAACCCA	TGTGCCTGGC
AGATAACTTC	GTATAATGTA	TGCTATACGA	AGTTATGGTA	CGTACTAAGC
TCTCATGTTT	CACGTACTAA	GCTCTCATGT	TTAACGTACT	AAGCTCTCAT
СТТТААССАА	CTAAACCCTC	ATCCCTAACC	TACTAACCTC	TCATCCTAA

Κ

1

51

F.2.4. pDł

101	GATAACAATT	TCACACAGGA	AACAGCTATG	ACCATGATTA	CGAATTTCTA
151	GAAATAATTT	TGTTTAACTT	TAAGAAGGAG	ATATACATAT	GAGGCCTCGG
201	ATCCTGTAAA	ACGACGGCCA	GTGAATTCCC	CGGGAAGCTT	CGCCAGGGTT
251	TTCCCAGTCG	AGCTCGATAT	CGGTACCAGC	GGATAACAAT	TTCACATCCG
301	GATCGCGAAC	GCGTCTCGAG	ACTAGTTCCG	TTTAAACCCA	TGTGCCTGGC
351	AGATAACTTC	GTATAATGTA	TGCTATACGA	AGTTATGGTA	CGTACTAAGC
401	TCTCATGTTT	CACGTACTAA	GCTCTCATGT	TTAACGTACT	AAGCTCTCAT
451	GTTTAACGAA	CTAAACCCTC	ATGGCTAACG	TACTAAGCTC	TCATGGCTAA
501	CGTACTAAGC	TCTCATGTTT	CACGTACTAA	GCTCTCATGT	TTGAACAATA
551	AAATTAATAT	AAATCAGCAA	CTTAAATAGC	CTCTAAGGTT	TTAAGTTTTA
601	TAAGAAAAAA	AAGAATATAT	AAGGCTTTTA	AAGCTTTTAA	GGTTTAACGG
651	TTGTGGACAA	CAAGCCAGGG	ATGTAACGCA	CTGAGAAGCC	CTTAGAGCCT
701	CTCAAAGCAA	TTTTCAGTGA	CACAGGAACA	CTTAACGGCT	GACAGAATTA
751	GCTTCACGCT	GCCGCAAGCA	CTCAGGGCGC	AAGGGCTGCT	AAAGGAAGCG
801	GAACACGTAG	AAAGCCAGTC	CGCAGAAACG	GTGCTGACCC	CGGATGAATG
851	TCAGCTACTG	GGCTATCTGG	ACAAGGGAAA	ACGCAAGCGC	AAAGAGAAAG
901	CAGGTAGCTT	GCAGTGGGCT	TACATGGCGA	TAGCTAGACT	GGGCGGTTTT
951	ATGGACAGCA	AGCGAACCGG	AATTGCCAGC	TGGGGCGCCC	TCTGGTAAGG
1001	TTGGGAAGCC	CTGCAAAGTA	AACTGGATGG	CTTTCTTGCC	GCCAAGGATC
1051	TGATGGCGCA	GGGGATCAAG	ATCTGATCAA	GAGACAGGAT	GAGGATCGTT
1101	TCGCATGATT	GAACAAGATG	GATTGCACGC	AGGTTCTCCG	GCCGCTTGGG
1151	TGGAGAGGCT	ATTCGGCTAT	GACTGGGCAC	AACAGACAAT	CGGCTGCTCT
1201	GATGCCGCCG	TGTTCCGGCT	GTCAGCGCAG	GGGCGCCCGG	TTCTTTTTGT
1251	CAAGACCGAC	CTGTCCGGTG	CCCTGAATGA	ACTGCAGGAC	GAGGCAGCGC
1301	GGCTATCGTG	GCTGGCCACG	ACGGGCGTTC	CTTGCGCAGC	TGTGCTCGAC
1351	GTTGTCACTG	AAGCGGGAAG	GGACTGGCTG	CTATTGGGCG	AAGTGCCGGG
1401	GCAGGATCTC	CTGTCATCTC	ACCTTGCTCC	TGCCGAGAAA	GTATCCATCA
1451	TGGCTGATGC	AATGCGGCGG	CTGCATACGC	TTGATCCGGC	TACCTGCCCA
1501	TTCGACCACC	AAGCGAAACA	TCGCATCGAG	CGAGCACGTA	CTCGGATGGA
1551	AGCCGGTCTT	GTCGATCAGG	ATGATCTGGA	CGAAGAGCAT	CAGGGGCTCG
1601	CGCCAGCCGA	ACTGTTCGCC	AGGCTCAAGG	CGCGCATGCC	CGACGGCGAG
1651	GATCTCGTCG	TGACACATGG	CGATGCCTGC	TTGCCGAATA	TCATGGTGGA
1701	AAATGGCCGC	TTTTCTGGAT	TCATCGACTG	TGGCCGGCTG	GGTGTGGCGG
1751	ACCGCTATCA	GGACATAGCG	TTGGCTACCC	GTGATATTGC	TGAAGAGCTT
1801	GGCGGCGAAT	GGGCTGACCG	CTTCCTCGTG	CTTTACGGTA	TCGCCGCTCC
1851	CGATTCGCAG	CGCATCGCCT	TCTATCGCCT	TCTTGACGAG	TTCTTCTGAG
1901	CGGGACTCTG	GGGTTCGAAA	TGACCGACCA	AGCGACGCCC	AACCTGCCAT

CACGAGATTT CGATTCCACC GCCGCCTTCT ATGAAAGGTT GGGCTTCGGA

ATCGTTTTCC GGGACGCCGG CTGGATGATC CTCCAGCGCG GGGATCTCAT

GCTGGAGTTC TTCGCCCACC CCGGGAT

F.2.5. pDS

1	CTATGTCGGG	TGCGGAGAAA	GAGGTAATGA	AATGGCACCT	AGGTATCGAT
51	GGCTTTACAC	TTTATGCTTC	CGGCTCGTAT	GTTGTGTGGA	ATTGTGAGCG
101	GATAACAATT	TCACACAGGA	AACAGCTATG	ACCATGATTA	CGAATTTCTA
151	GAAATAATTT	TGTTTAACTT	TAAGAAGGAG	ATATACATAT	GAGGCCTCGG
201	ATCCTGTAAA	ACGACGGCCA	GTGAATTCCC	CGGGAAGCTT	CGCCAGGGTT
251	TTCCCAGTCG	AGCTCGATAT	CGGTACCAGC	GGATAACAAT	TTCACATCCG
301	GATCGCGAAC	GCGTCTCGAG	ACTAGTTCCG	TTTAAACCCA	TGTGCCTGGC
351	AGATAACTTC	GTATAATGTA	TGCTATACGA	AGTTATGGTA	CGTACTAAGC
401	TCTCATGTTT	CACGTACTAA	GCTCTCATGT	TTAACGTACT	AAGCTCTCAT
451	GTTTAACGAA	CTAAACCCTC	ATGGCTAACG	TACTAAGCTC	TCATGGCTAA
501	CGTACTAAGC	TCTCATGTTT	CACGTACTAA	GCTCTCATGT	TTGAACAATA
551	AAATTAATAT	AAATCAGCAA	CTTAAATAGC	CTCTAAGGTT	TTAAGTTTTA
601	TAAGAAAAAA	AAGAATATAT	AAGGCTTTTA	AAGCTTTTAA	GGTTTAACGG
651	TTGTGGACAA	CAAGCCAGGG	ATGTAACGCA	CTGAGAAGCC	CTTAGAGCCT
701	CTCAAAGCAA	TTTTGAGTGA	CACAGGAACA	CTTAACGGCT	GACATAATTC
751	AGCTTCACGC	TGCCGCAAGC	ACTCAGGGCG	CAAGGGCTGC	TAAAGGAAGC
801	GGAACACGTA	GAAAGCCAGT	CCGCAGAAAC	GGTGCTGACC	CCGGATGAAT
851	GTCAGCTGGG	AGGCAGAATA	AATGATCATA	TCGTCAATTA	TTACCTCCAC
901	GGGGAGAGCC	TGAGCAAACT	GGCCTCAGGC	ATTTGAGAAG	CACACGGTCA
951	CACTGCTTCC	GGTAGTCAAT	AAACCGGTAA	GTAGCGTATG	CGCTCACGCA
1001	ACTGGTCCAG	AACCTTGACC	GAACGCAGCG	GTGGTAACGG	CGCAGTGGCG
1051	GTTTTCATGG	CTTGTTATGA	CTGTTTTTTT	GGGGTACAGT	CTATGCCTCG
1101	GGCATCCAAG	CAGCAAGCGC	GTTACGCCGT	GGGTCGATGT	TTGATGTTAT
1151	GGAGCAGCAA	CGATGTTACG	CAGCAGGGCA	GTCGCCCTAA	AACAAAGTTA
1201	AACATCATGA	GGGAAGCGGT	GATCGCCGAA	GTATCGACTC	AACTATCAGA
1251	GGTAGTTGGC	GTCATCGAGC	GCCATCTCGA	ACCGACGTTG	CTGGCCGTAC
1301	ATTTGTACGG	CTCCGCAGTG	GATGGCGGCC	TGAAGCCACA	CAGTGATATT
1351	GATTTGCTGG	TTACGGTGAC	CGTAAGGCTT	GATGAAACAA	CGCGGCGAGC
1401	TTTGATCAAC	GACCTTTTGG	AAACTTCGGC	TTCCCCTGGA	GAGAGCGAGA
1451	TTCTCCGCGC	TGTAGAAGTC	ACCATTGTTG	TGCACGACGA	CATCATTCCG
1501	TGGCGTTATC	CAGCTAAGCG	CGAACTGCAA	TTTGGAGAAT	GGCAGCGCAA
1551	TGACATTCTT	GCAGGTATCT	TCGAGCCAGC	CACGATCGAC	ATTGATCTGG
1601	CTATCTTGCT	GACAAAAGCA	AGAGAACATA	GCGTTGCCTT	GGTAGGTCCA
1651	GCGGCGGAGG	AACTCTTTGA	TCCGGTTCCT	GAACAGGATC	TATTTGAGGC
1701	GCTAAATGAA	ACCTTAACGC	TATGGAACTC	GCCGCCCGAC	TGGGCTGGCG
1751	ATGAGCGAAA	TGTAGTGCTT	ACGTTGTCCC	GCATTTGGTA	CAGCGCAGTA
1801	ACCGGCAAAA	TCGCGCCGAA	GGATGTCGCT	GCCGACTGGG	CAATGGAGCG
1851	CCTGCCGGCC	CAGTATCAGC	CCGTCATACT	TGAAGCTAGA	CAGGCTTATC
1901	TTGGACAAGA	AGAAGATCGC	TTGGCCTCGC	GCGCAGATCA	GTTGGAAGAA
1951	TTTGTCCACT	ACGTGAAAGG	CGAGATCACC	AAGGTAGTCG	GCAAATAATG
2001	TCTAACAATT	CGTTCAAGCC	GACGGAT		

F.3. ACEMBL plasmid maps



Acceptor vectors pACE1 (pACE) and pACE2, containing a T7 promoter and terminator, are shown. **Donor vectors** pDK, pDS and pDC contain conditional origins of replication. pDS and pDK have a lac promoter. pDC has a T7 promoter. Resistance markers are shown in gray, origins of replication in yellow. LoxP imperfect inverted repeat sequences are shown as spheres. Homing endonuclease sites and corresponding BstXI sites are boxed. The restriction enzyme sites in the multiple integration element (MIE) are indicated. All MIEs

have the same DNA sequence between Clal and Pmel. Differences in unique restriction site composition stem from differences in the plasmid backbone sequences.

F.4. Analytical restriction of ACEMBL vectors

All ACEMBL vectors were analyzed by BamHI restriction digestion. The undigested and digested ACEMBL vectors are shown below:



Restriction mapping of ACEMBL vectors. Both undigested as well as BamHI-digested Acceptor and Donor vectors are. All restriction reactions yield the expected sizes. Lane 1-5 show uncut pACE1, pACE2, pDC, pDK, and pDS vectors; lane M shows λ Styl marker; lane A-E show BamHI-digested pACE1, pACE2, pDC, pDK, and pDS vectors.

Additionally, the MultiColi vectors yield characteristic restriction digest patterns when plasmid DNA is restricted with HindIII and HinfI.

Vector	Length [bp]	Restriction	Fragment(s) [bp]
pACE1	2652	HindIII	2652
		Hinfl	1690, 491, 396, 75
pACE2	2982	HindIII	2982
		Hinfl	772, 618, 491, 396, 256, 220, 154, 75
pDC	2067	HindIII	1517, 550
		Hinfl	1275, 792
pDK	2077	HindIII	1681, 396
		Hinfl	1796, 134, 58, 52, 37
pDS	2027	HindIII	1631, 396
		Hinfl	1814, 213

G. References

- 1. Alberts B (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92: 291-294.
- Ausubel F, Brent R, et al. (eds., 1994). Current Protocols in Molecular Biology. John Wiley & Sons, New York, electronic version DOI 10.1002/0471142727
- 3. Bhattacharya A (2009). Protein structure: Structures of desire. Nature 459: 24-27.
- Bieniossek C, Nie Y, Frey D, Olieric N, Schaffitzel C, Collinson I, Romier C, Berger P, Richmond TJ, Steinmetz MO, Berger I (2009). Automated unrestricted multigene recombineering for multiprotein complex production. *Nature Methods* 6: 447-450.
- 5. **Charbonnier S, Gallego O, Gavin AC** (2008). The social network of a cell: Recent advances in interactome mapping. *Biotechnology Annual Review* 14: 1-28.
- 6. **Chemler JA and Koffas MAG** (2008). Metabolic engineering for plant natural product biosynthesis in microbes. *Current Opinion in Biotechnology* 19: 597-605.
- 7. **Chou CP** (2007). Engineering cell physiology to enhance recombinant protein production in *Escherichia coli. Applied Microbiology and Biotechnology* 76: 521-532.
- 8. Estrozi LF, Boehringer D, Shan SO, Ban N, Schaffitzel C (2011). Cryo-EM structure of the E. coli translating ribosome in complex with SRP and its receptor. *Nature Structural and Molecular Biology* 18: 88-90.
- 9. Gaiser F, Tan S, and Richmond TJ (2000). Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. *Journal of Molecular Biology* 302: 1119-1127.
- 10. Inoue H, Nojima H and Okayama H (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96: 23-28.
- 11. Lee SK, Chou H, Ham TS, Lee TS, and Keasling JD (2008). Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology* 19: 556-563.
- 12. Li M and Elledge SJ (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature Methods* 4: 251-256.
- 13. Nie Y, Viola C, Bieniossek C, Trowitzsch S, Vijay-Achandran LS, Chaillet M, Garzoni F, Berger I (2009). Getting a grip on complexes. *Current Genomics* 10: 558-72.
- 14. Robinson CV, Sali A, and Baumeister W (2007). The molecular sociology of the cell. *Nature* 450: 973-982.
- 15. Sambrook J and Russell DW (2000). Molecular Cloning, A Laboratory Manual, 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, available online at http://cshprotocols.cshlp.org
- 16. Tan S, Kern RC, and Selleck W (2005). The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli*. *Protein Expression and Purification* 40: 385-395.
- 17. Van Hijum SA, Medema MH, and Kuipers OP (2009). Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiology and Molecular Biology Reviews* 73: 481-509.

Please consider quoting the boxed article as original reference when publishing results obtained with the *Multi*Coli^o kit.