GATEWAY Binary Vector (pGWB)

1. All pGWB vectors confer resistance to both kanamycin and hygromycin in plant.
2. Some of the vectors contain the 35S promoter upstream of the cloning site while others have none.
3. The N- or C-terminal tags are automatically fused subsequent to the LR reaction.
4. Please read the GATEWAY instruction manual (Invitrogen) for procedures,   etc.

　Please note that our pGWB seriese are available only for basic research.   If you plan to use these pGWBs for commercial research, please contact Invitrogen about licencing of GATEWAY^{TM}.
　Please do not distribute to other researchers without permission.

Tsuyoshi Nakagawa
Research Institute of Molecular Genetics
Shimane University
Matsue 690, Japan
tnakagaw@life.shimane-u.ac.jp
(July 26, 2002)

The pGWB Series

(1) pGWB1 : [(no promoter, no tag)   (--R1-CmR-ccdB-R2 -)]

(2) pGWB2 : [(35S promoter, no tag)   (--35S promoter–R1-CmR-ccdB-R2--)]

(3) pGWB3 : [(no promoter, C-GUS)   (--R1- CmR-ccdB-R2–GUS--)]

(4) pGWB4 : [(no promoter, C-sGFP)   (--R1- CmR-ccdB-R2–sGFP--)]

(5) pGWB5: [(35S promoter, C-sGFP)   (--35S promoter–R1- CmR-ccdB-R2–sGFP--)]

(6) pGWB6 : [(35S promoter, N-sGFP)   (--35S promoter-sGFP-R1- CmR-ccdB-R2--)]

(7) pGWB7 : [(no promoter, C-6xHis)   (--R1- CmR-ccdB-R2-6xHis--)]

(8) pGWB8 : [(35S promoter, C-6xHis)   (--35S promoter-R1- CmR-ccdB-R2-6xHis--)]

(9) pGWB9 : [(35S promoter, N-6xHis)   (--35S promoter-6xHis-R1- CmR-ccdB-R2--)]

(10) pGWB10 : [(no promoter, C-FLAG)   (--R1- CmR-ccdB-R2-FLAG--)]

(11) pGBW11 : [(35S promoter, C-FLAG)   (--35S promoter-R1- CmR-ccdB-R2-FLAG--)]

(12) pGWB12 : [(35S promoter, N-FLAG)   (--35S promoter-FLAG-R1- CmR-ccdB-R2--)]

(13) pGWB13 : [(no promoter, C-3xHA)   (--R1- CmR-ccdB-R2-3xHA--)]

(14) pGWB14 : [(35S promoter, C-3xHA)   (--35S promoter-R1- CmR-ccdB-R2-3xHA--)]

(15) pGWB15 : [(35S promoter, N-3xHA)   (--35S promoter-3xHA-R1- CmR-ccdB-R2--)]

(16) pGWB16 : [(no promoter, C-4xMyc)   (--R1- CmR-ccdB-R2-4xMyc--)]

(17) pGWB17 : [(35S promoter, C-4xMyc)   (--35S promoter-R1- CmR-ccdB-R2-4xMyc--)]

(18) pGWB18 : [(35S promoter, N-4xMyc)   (--35S promoter-4xMyc-R1- CmR-ccdB-R2--)]

(19) pGWB19 : [(no promoter, C-10xMyc)   (- -R1- CmR-ccdB-R2-10xMyc)]

(20) pGWB20 : [(35S promoter, C-10xMyc) (--35S promoter-R1- CmR-ccdB-R2-10xMyc)]

(21) pGWB21 : [(35S promoter, N-10xMyc) (--35S promoter-10xMyc-R1- CmR-ccdB-R2)]

(22) pGWB22 : [(no promoter, C-GST)   (--R1- CmR-ccdB-R2-GST--)]

(23) pGWB23 : [(35S promoter, C-GST)   (--35S promoter-R1- CmR-ccdB-R2-GST)]

(24) pGWB24 : [(35S promoter, N-GST)   (--35S promoter-GST-R1- CmR-ccdB-R2)]

(25) pGWB25 : [(no promoter, C-T7)   (--R1- CmR-ccdB-R2-T7--)]

(26) pGWB26 : [(35S promoter, C-T7)   (--35S promoter-R1- CmR-ccdB-R2-T7)]

(27) pGWB27 : [(35S promoter, N-T7)   (--35S promoter-T7-R1- CmR-ccdB-R2)]

(28) pGWB28 : [(no promoter, C-TAP)   (--R1- CmR-ccdB-R2-TAP--)]

(29)pGWB29 :[(35S promoter, C-TAP)   (--35S promoter-R1- CmR-ccdB-R2-T7)]

Notes:

1. pGWB1 is used for simple cloning, pGWB2 is for expression driven by the 35S promoter, pGWB3 for promoter-GUS analysis, pGWB4-6 for fusion of GFP, and pGWB7-24 are for fusion of tags.

2. Tags could be added to either the N- or C- terminus of *Your Clone*. For example, in pGWB10 "C-FLAG" means that after the LR reaction with the vector, FLAG will be fused at the C-terminal of *Your Clone*.

3. DNA sequences enclosed in [ ] are introduced into the *Hin*dIII-*Sac*I site of the binary vector pABH-Hm1 [Mita et al. (1995) Plant Physiol. **107**: 895-904; see next page]. R = the *att*R site.

Construction of pGWBs

The binary vector used as the starting plasmid for the construction of pGWBs was pABH-Hm1 [Mita et al. (1995) Plant Physiol. **107**: 895-904]. pABH-Hm1 itself was made by introducing the beta amylase promoter and HPT unit into pBI101. The *Hin*dIII-*Sac*I region in pABH-Hm1 harboring the *beta amylase promoter::GUS* was replaced by the GATEWAY cassette and tag in the pGWB as shown below. The average size of pGWB is 17kb.

| Nos-P | NPTII | Nos-T- | (HindIII)- | GATEWAY | -(SacI)- | Nos-T | Nos-T | HPT | 35S |

The DNA fragment diagrammed above is present between LB and RB of binary vector. Note that the HPT unit is in reverse orientation.

All pGWBs contain NPTII gene for kanamycin selection and HPT gene for hygromycin selection in plant.

Antibiotics can be used for selection in bacteria are as below.

pGWBs: kanamycin, hygromycin, chloramphenicol

recombinant pGWB (after LR reaction) : kanamycin, hygromycin.

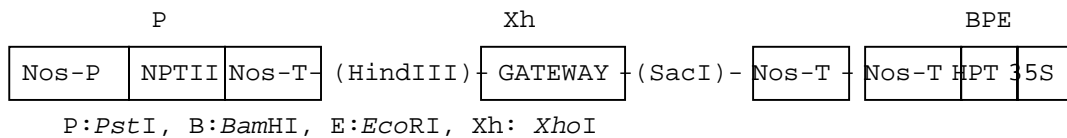The princple behind the GATEWAY technology and the pGWB vector series could be generalized as follow.

(1) Subsequent to the LR reaction with an Entry Clone and a pGWB vector, the *att*R1-(CmR, *ccd*B)-*att*R2 cassette in the pGWB will be replaced by the *att*B1-(*Your Clone*)-*att*B2.

(2) The tag will be fused automatically in-frame to *Your Clone*, if the primer was designed according to the GATEWAY system (see instruction manual by Invitrogen). Note that the peptide encoded by the *att*B and linker region will be included in your construct, linking the tag to *Your Clone*.

(3) N indicates the nucleotide (and X the amino acid) that shold be included in *Your Clone*.

Important Note

(1) Linearize pGWBs by digestion with *Xho*I before performing the LR reaction.   The efficiency of the LR reaction depends on the topology of the plasmids in the following order (see instruction manual of Invitrogen).

Either or both plasmids linear > both plasmids relaxed >> both plasmids supercoiled

The *Xho*I linker was introduced into *Not*I site (of GATEWAY cassette just downstream of *att*R1) in all pGWBs.   This *Xho*I site is unique.

```
          P                        Xh                      BPE
  ┌───────┬───────┬───────┐       ┌─────────┐       ┌───────┬───────────────┐
  │ Nos-P │ NPTII │ Nos-T ┤(HindIII)┤ GATEWAY ├(SacI)-│ Nos-T ┤ Nos-T│HPT│35S │
  └───────┴───────┴───────┘       └─────────┘       └───────┴───────────────┘
       P:PstI, B:BamHI, E:EcoRI, Xh: XhoI
```

The *Hin*dIII、*Xho*I and *Sac*I sites are unique in pGWBs.

(2) Select with kanamycin (50 ug/ml) AND hygromycin (50 ug/ml) after the LR reaction. Because the Entry Clone vector pDONR201 is also encode kanamycin resistance, hygromycin is necessary to screen recombinant pGWBs.   Kanamycin, should be used in addition as many petite colonies will appear on plate without kanamycin.

(3) Use the *Escherichia coli* strain DH5alpha for cloning the product of the LR (BP) reactions, since negative selection with *ccd*B does not work in the *E.coli* which has F' episome (see instruction manual of Invitrogen).

(4) Sometimes, *E.coli* with both the Entry Clone and recombinant pGWB are obtained subsequent to the cloning.   I recommend that only the *E. coli* colonies with only the recombinant pGWB be used.

pGWB1 : [(no promoter, no tag)   (--R1-ccdB-R2 --)]


```
(AAGCTT) --//
(TCTAGA)TC ACA AGT TTG TAC AAA AAA --(CmR, ccdB) -- TTC TTG TAC AAA GTG
GTG ATC TAG AGG (GAG CTC)
```


(1) The *Hin*dIII (AAGCTT),   *Xba*I (TCTAGA) and *Sac*I (GAGCTC) sites are shown
(2) The *att*R1-CmR-*ccd*B-*att*R2 cassette is underlined.

After the LR reaction, the following sequence will be obtained.


```
(AAGCTT) --//
(TCTAGA)TC ACA AGT TTG TAC AAA AAA GCA GGC TNN (Your Clone) NAC CCA GCT
TTC TTG TAC AAA GTG GTG ATC TAG AGG (GAG CTC)
```


(1) The *att*B1 and *att*B2 sequences are underlined.
(2) The in-frame termination codon TAG (double-underlined) is present downstream of
    *att*B2.

pGWB2 : [(35S promoter, no tag)   (--35S promoter-R1-ccdB-R2 --)]


```
(AAGCTT) -- 35S promoter --//
(TCTAGA)GTTA TCA ACA AGT TTG TAC AAA AAA --(CmR,ccdB)--TTC TTG TAC AAA GTG
GTT GAT AAC AGC GCT TA(G AGCTC)
```


(1) The *Hin*dIII (AAGCTT),　*Xba*I (TCTAGA) and *Sac*I (GAGCTC) sites are shown
(2) The *att*R1-CmR-*ccd*B-*att*R2 cassette is underlined.
(3) The sequence upstream of the *Xba*I site is the same as in pBI121

After LR reaction, the following sequence will be obtained


```
(AAGCTT) -- (35S promoter) --//
(TCTAGA)GTTA TCA ACA AGT TTG TAC AAA AAA GCA GGC TNN (Your Clone) NAC CCA
                                          M -------   (X) (P)
GCT TTC TTG TAC AAA GTG GTT GAT AAC AGC GCT TA(G AGCTC)
(A) (F) (L) (Y) (K) (V) (V) (D) (N) (S) (A)  *
```


(1) Include the initiation codon ATG in *Your Clone* as pGWB2 does not have one.
(2) The inclusion of a termination codon in *Your Clone* is not necessary since one is
    present downstream of attB2.   The amino acid residues in ( ) are encoded by the
    *att*B2 and the linker region, and will be included if the termination codon TAG
    (double-underlined) of the vector is used.
(3) N indicates the nucleotide (and X the amino acid) that should be included in *Your
    Clone*.

pGWB3 :　[(no promoter, C-GUS)　(--R1-ccdB-R2–GUS--)]


```
(AAGCTT) --//
(TCTAA)T CAA ACA AGT TTG TAC AAA AAA --(CmR, ccdB)-- TTC TTG TAC AAA GTG
GTT CGA TCT AGA GGA TCC CCG GGT GGT CAG TCC CTT ATG  --(GUS sequence)
--(GAGCTC)
```


(1) As a result of a mistake during the fill-in reaction while constructing pGWB3, the *Xba*I site TCTAGA was changed to TCTAA.　However, I think it would not affect the result of your experiments.

(2) The initiation codon ATG (double-underlined) is from the GUS coding sequence.

(3) The *Sac*I site (GAGCTC) is present just upstream of the Nos terminator.

After the LR reaction, the following sequence will be obtained.


```
(AAGCTT) --//
(TCTAA)T CAA ACA AGT TTG TAC AAA AAA GCA GGC TNN (Your Clone) NAC CCA GCT
                                                             (X) (P) (A)
TTC TTG TAC AAA GTG GTT CGA TCT AGA GGA TCC CCG GGT GGT CAG TCC CTT ATG
(F) (L) (Y) (K) (V) (V) (R) (S) (R) (G) (S) (P) (G) (G) (Q) (S) (V)  M
--(GUS sequence) --(GAGCTC)
```


(1) If *Your Clone* has only a 5' region/promoter without an initiation codon, translation will start at the ATG (double-underlined) of GUS.

(2) If *Your Clone* contains at least a part of an ORF, C-terminal GUS fusion will be made (if in frame).　In that case, the peptide sequence made up of the amino acids in ( ) will link *Your Clone* and GUS.

(3) N indicates the nucleotide (and X the amiono acid) that should be included in *Your Clone*, if C-terminal GUS fusion is to be made.

(4) The DNA sequence from *GUS* to the *Sac*I site is the same as in pBI121.

.

pGWB4 : [(no promoter, C-sGFP)   (--R1-ccdB-R2–sGFP--)]


```
(AAGCTT) --//
(TCTA A)TC ACA AGT TTG TAC AAA AAA --(Cm-R, ccdB)-- TTC TTG TAC AAA GTG
GTG ATC ATG --(GFP sequence) --(GAGCTC)
```


(1) As a result of a mistake during the fill-in reaction while constructing pGWB4, the *Xba*I site   TCTAGA was changed to TCTAA.   However, I think it would not affect the results of your experiments.

(2) The initiation codon ATG (double-underlined) is from the GFP coding sequence.

After the LR reaction, the following sequence will be obtained.


```
(AAGCTT) --//
(TCTA A)TC ACA AGT TTG TAC AAA AAA GCA GGC TNN (Your Clone) NAC CCA GCT
                                                            (X)(P)(A)
TTC TTG TAC AAA GTG GTG ATC ATG --(GFP sequence) --(GAGCTC)
(F) (L) (Y) (K) (V) (V) (I)  M
```


(1) If *Your Clone* has only a 5' region/promoter without an initiation codon, tranalation will start at the ATG (double-underlined) of *GFP*.

(2) If *Your Clone* contains at leaset a part of an ORF, C-terminal GFP fusion will be made.   In that case, the peptide sequence made up of the amino acid in ( ) will link *Your Clone* and GFP.

(3) N indicates the nucleotide (and X the amino acid) that should be included in Your Clone, if C-terminal GUS fusion is to be made.

(4) The sequences from the termination codon of *GFP* to *Sac*I site (see below) was generated by the follwing steps: (i) *Not*I digest, (ii) fill-in, and (iii) *Sac*I linker (8-mer) ligation.

```
TAAAGCGGCCC(GAGCTC)
```

pGWB5 : [(35S promoter, C-sGFP) (--35Spromoter–R1-ccdB-R2–sGFP--)]

(AAGCTT) -- 35S promoter --//

(TCT AA)T CAA <u>ACA AGT TTG TAC AAA AAA</u> --(CmR, ccdB)—TTC TTG TAC AAA GTG

<u>GTT</u> CGA TCT AGA GGA TCC <u><u>ATG</u></u> --(*GFP* sequence) --(GAGCTC)

(1) As a result of a mistake during the fill-in reaction while constructing pGWB5, the XbaI site TCTAGA was changed to TCTAA.   However, I think it would not affect the result of your experiments.

(2) The initiation codon ATG (double-underlined) is from the *GFP* coding sequence.

After LR reaction, the following sequence will be obtained.

35S promoter –(TCT AA)T CAA <u>ACA AGT TTG TAC AAA AAA GCA GGC T</u>NN (*Your Clone*)

N<u>AC CCA GCT TTC TTG TAC AAA GTG GTT</u> CGA TCT AGA GGA TCC <u><u>ATG</u></u> --

-X- -P- -A- -F- -L- -Y- -K- -V- -V- -R- -S- -R- -G- -S- -M-

(*GFP sequence*) -- (GAGCTC)

(1) Include an initiation codon ATG in *Your Clone*.   Protein fusion with the C-terminal GFP will be linked by the peptide sequence derived from the *att*B2 and linker region as shown in the above sequence.

(2) N indicates the nucleotide (and X the amino acid) that is needed to be included in *Your Clone*.

(3) The DNA sequence downstream of *GFP* is the same ase in pGWB4.

(4) Hyphenations flanking the amino acid residues are used for proper spacing.

pGWB6 : [(35S promoter, N-sGFP)   (--35S promoter-sGFP-R1-ccdB-R2--)]


```
(AAGCTT) -- 35S promoter --(GFP sequence)--//
--- C(TG TAC A)TC ACA AGT TTG TAC AAA AAA --(CmR, ccdB)—TTC TTG TAC AAA
GTG GTG ATG TAC AAG TAA AGCGGCCC(GAGCTC)
```

(1) The GATEWAY conversion cassette was introduced into the *Bsr*GI (TGTACA) site (in the 3' end of *GFP*).
(2) The termination codon TAA (double-underlined) is derived from *GFP*.

After LR reaction, the following sequence will be obtained.


```
(AAGCTT) -- 35S promoter --(GFP sequence)--//
--- C(TG TAC A)TC ACA AGT TTG TAC AAA AAA GCA GGC TNN (Your Clone) NAC
--- -L- -Y- -I- -T- -S- -L- -Y- -K- -K- -A- -G- -X- ---- ----- (X)
CCA GCT TTC TTG TAC AAA GTG GTG ATG TAC AAG TAA AGCGGCCC(GAGCTC)
(P) (A) (F) (L) (Y) (K) (V) (V) (M) (Y) (K)  *
```


(1) The inclusion of a termination codon in *Your Clone* is not necessary since one is present downstream of *att*B2.   The amino acid residues in ( ) is encoded by the *att*B2 and the linker region, and will be included if the termination codon TAG of the vector is used.
(2) The C-terminus of the original GFP is LYK*.
(3) The 5' end sequence of *GFP* is as shown below.


```
    TCTAGAGGATCC ATG --(GFP sequence)--
     XbaI  BamHI  M
```


(4) The sequence upstream of *Xba*I site is the same as pBI121.

pGWB7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23 , 25, 26, 28, 29

(no promoter or 35S promoter, C-tag)

No promoter --- 7, 10, 13, 16, 19, 22, 25, 28

35S promoter --- 8, 11, 14, 17, 20, 23, 26, 29

```
(AAGCTT) -- (no promoter or 35S promoter) --//

(TCTAGA)GTTA TCA ACA AGT TTG TAC AAA AAA --(CmR,ccdB)--TTC TTG TAC AAA GTG

GTT GAT AAC AGC tag GCT TA(G AGCTC)
```

(1) The *Hin*dIII (AAGCTT), *Xba*I (TCTAGA) and *Sac*I (GAGCTC) sites are shown.

(2) The *att*R1-CmR-*ccd*B-*att*R2 cassette is underlined.

(3) The sequence upstream of the *Xba*I site is the same as in pBI101 or pBI121.

(4) The sequence of each tag is indicated in the "Tags" section.

After LR reaction, the following sequence will be obtained.

```
(AAGCTT) -- (no promoter or 35S promoter) --//

(TCTAGA)GTTA TCA ACA AGT TTG TAC AAA AAA GCA GGC TNN (Your Clone) NAC CCA
                                                    M -------   -X- -P-
GCT TTC TTG TAC AAA GTG GTT GAT AAC AGC tag GCT TA(G AGCTC)
-A- -F- -L- -Y- -K- -V- -V- -D- -N- -S- tag -A-   *
```

(1) Include an initiation codon ATG in *Your Clone*.   Protein fusion with the C-terminal
   tag will be linked by the peptide sequence derived from the *att*B2 and linker region
   as shown in the above sequence.

(2) Translation will stop at the termination codon TAG (double-underlined) after the
   tag.

pGWB9, 12, 15, 18,2 1, 24, 27   (35S promoter, N-tag)


```
(AAGCTT) -- 35S promoter --//
(TCTAGA) ATG AGC tag GCT GTT ATC ACA AGT TTG TAC AAA AAA --(CmR,ccdB)-
-TTC TTG TAC AAA GTG GTG ATA ACC TA(G AGCTC)
```


(1) The *Hin*dIII (AAGCTT), *Xba*I (TCTAGA) and *Sac*I (GAGCTC) sites are shown.

(2) The *att*R1-CmR-*ccd*B-*att*R2 cassette is underlined.

(3) Sequence upstream of the *Xba*I site is the same as in pBI121.

(4) The sequence of each tags is indicated in the "Tags" section.

After the LR reaction, the following sequence will be obtained.


```
(AAGCTT) -- 35S promoter --//
(TCTAGA) ATG AGC tag GCT GTT ATC ACA AGT TTG TAC AAA AAA GCA GGC TNN
          M-  -S-  tag –A-  -V-  -I-  -T-  -S-  -L-  -Y-  -K-  -K-  -A-  -G-  -X-
(Your Clone) NAC CCA GCT TTC TTG TAC AAA GTG GTG ATA ACC TA(G AGCTC)
  --- ---     (X) (P) (A) (F) (L) (Y) (K) (V) (V) (I) (T)  *
```


(1) The inclusion of a termination codon in *Your Clone* is not necessary since one is
    present downstream of *att*B2. The amino acid residues in ( ) is encoded by the *att*B2
    and the linker region, and will be included if the termination codon TAG (double-
    underlined) of the vector is used.

Hyphenations in the amino acid sequences are used for proper spacing.

pGWB4, 5, 6 (sGFP) 720bp

The sequence of sGFP (ATG to TAA) is as described by Chiu et al., (1996) Current Biol., **6**: 325-330 .

Note that the C-terminal K is converted to I in pGWB6 (35S promoter, N-GFP). (see the section on pGWB6)

```
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTG
 M— V— S— K— G— E— E— L— F— T— G— V— V— P— I— L— V— E— L— D— G— D— V—
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
 N- G- H- K- F- S- V- S- G- E- G- E- G- D- A- T- Y- G- K- L- T- L- K-
TTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCACCTACGGCGTG
 F- I- C- T- T- G- K- L- P- V- P- W- P- T- L- V- T- T- F- T- Y- G- V-
CAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC
 Q- C- F- S- R- Y- P- D- H- M- K- Q- H- D- F- F- K- S- A- M- P- E- G-
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTC
 Y- V- Q- E- R- T- I- F- F- K- D- D- G- N- Y- K- T- R- A- E- V- K- F-
GAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
 E- G- D- T- L- V- N- R- I- E- L- K- G- I- D- F- K- E- D- G- N- I- L-
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC
 G- H- K- L- E- Y- N- Y- N- S- H- N- V- Y- I- M- A- D- K- Q- K- N- G-
ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
 I- K- V- N- F- K- I- R- H- N- I- E- D- G- S- V- Q- L- A- D- H- Y- Q-
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCC
 Q- N- T- P- I- G- D- G- P- V- L- L- P- D- N- H- Y- L- S- T- Q- S- A-
CTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
 L- S- K- D- P- N- E- K- R- D- H- M- V- L- L- E- F- V- T- A- A- G- I
ACTCACGGCATGGACGAGCTGTACAAGTAA
 T- H- G- M- D- E- L- Y- K- *
```

pGWB7, 8, 9   (6xHis)   18bp

CATCATCATCATCATCAT

 H- H- H- H- H- H-

pGWB10, 11, 12   (FLAG)   24bp

GACTACAAGGATGACGATGACAAG

 D- Y- K- D- D- D- D- K-

pGWB13, 14, 15   (3xHA)   120bp

Please consult   http://pingu.salk.edu/users/forsburg/ for further information

GGGTTAATTAACATCTTTTACCCATACGATGTTCCTGACTATGCGGGCTATCCCTATGACGTCCCGGAC

 G- L- I- N- I- F- <u>Y- P- Y- D- V- P- D- Y- A-</u> G- <u>Y- P- Y- D- V- P- D-</u>

TATGCAGGATCCTATCCATATGACGTTCCAGATTACGCTGCTCAGTGCAGC

 <u>Y- A-</u> G- S- <u>Y- P- Y- D- V- P- D- Y- A-</u> A- Q- C- S-

(1) The three HA units are underlined.

(2) The 54th nucleotide is a G in the reported sequence, but it is a C in our vectors. This does not affect the

   identity of the amino acid, however [CCG (P) to CCC (P)].

pGWB16, 17, 18   (4xMyc)   171bp

Please consult   http://pingu.salk.edu/users/forsburg/ for further information

GGGTTAATTAACGGTGAACAAAAGCTAATCTCCGAGGAAGACTTGAACGGTGAACAAAAATTAATCTCA

 G- L- I- N- G- <u>E- Q- K- L- I- S- E- E- D- L-</u> N- G- <u>E- Q- K- L- I- S-</u>

GAAGAAGACTTGAACGGACTCGACGGTGAACAAAAGTTGATTTCTGAAGAAGATTTGAACGGTGAACAA

 <u>E- E- D- L-</u> N- G- L- D- G- <u>E- Q- K- L- I- S- E- E- D- L-</u> N- G- <u>E- Q-</u>

AAGCTAATCTCCGAGGAAGACTTGAACGGTAGC

 <u>K- L- I- S- E- E- D- L-</u> N- G –S

The four Myc units aare underlined.

pGWB19, 20, 21 (10xMyc) 402bp

GGGTTAATTAACGGTGAACAAAAGCTAATCTCCGAGGAAGACTTGAACGGTGAACAAAAATTAATCTCA

 G- L- I- N- G- <u>E- Q- K- L- I- S- E- E- D- L-</u> N- G- <u>E- Q- K- L- I- S-</u>

GAAGAAGACTTGAACGGACTCGACGGTGAACAAAAGTTGATTTCTGAAGAAGATTTGAACGGTGAACAA

 <u>E- E- D- L-</u> N- G- L- D- G- <u>E- Q- K- L- I- S- E- E- D- L-</u> N- G- <u>E- Q-</u>

AAGCTAATCTCCGAGGAAGACTTGAACGGTGAACAAAAATTAATCTCAGAAGAAGACTTGAACGGACTC

 <u>K- L- I- S- E- E- D- L-</u> N- G- <u>E- Q- K- L- I- S- E- E- D- L-</u> N- G- L-

GACGGTGAACAAAAGTTGATTTCTGAAGAAGATTTGAACGGTGAACAAAAGCTAATCTCCGAGGAAGAC

 D- G<u>- E- Q- K- L- I- S- E- E- D- L-</u> N- G- <u>E- Q- K- L- I- S- E- E- D-</u>

TTGAACGGTGAACAAAAATTAATCTCAGAAGAAGACTTGAACGGACTCGACGGTGAACAAAAGTTGATT

 <u>L-</u> N- G- <u>E- Q- K- L- I- S- E- E- D- L-</u> N- G- L- D- G- <u>E- Q- K- L- I-</u>

TCTGAAGAAGATTTGAACGGTGAACAAAAGCTAATCTCCGAGGAAGACTTGAACGGTAGC

 <u>S- E- E- D- L-</u> N- G- <u>E- Q- K- L- I- S- E- E- D- L-</u> N- G- S-

 (1) The ten Myc units are underlined.

 (2) The last codon AGC (Ser) is not included in pGWB19 and 20.

pGWB22,23,24 (GST)　678bp

The full-length *GST* sequence was derived from pGEX2T

```
ATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATAT
 M- S- P- I- L- G- Y- W- K- I- K- G- L- V- Q- P- T- R- L- L- L- E- Y-
CTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTT
 L- E- E- K- Y- E- E- H- L- Y- E- R- D- E- G- D- K- W- R- N- K- K- F-
GAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATG
 E- L- G- L- E- F- P- N- L- P- Y- Y- I- D- G- D- V- K- L- T- Q- S- M-
GCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATT
 A- I- I- R- Y- I- A- D- K- H- N- M- L- G- G- C- P- K- E- R- A- E- I-
TCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTT
 S- M- L- E- G- A- V- L- D- I- R- Y- G- V- S- R- I- A- Y- S- K- D- F-
GAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGT
 E- T- L- K- V- D- F- L- S- K- L- P- E- M- L- K- M- F- E- D- R- L- C-
CATAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTT
 H- K- T- Y- L- N- G- D- H- V- T- H- P- D- F- M- L- Y- D- A- L- D- V-
GTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAA
 V- L- Y- M- D- P- M- C- L- D- A- F- P- K- L- V- C- F- K- K- R- I- E-
GCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAA
 A- I- P- Q- I- D- K- Y- L- K- S- S- K- Y- I- A- W- P- L- Q- G- W- Q-
GCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGTTCCGCGTGGATCC
 A- T- F- G- G- G- D- H- P- P- K- S- D- L- V- P- R- G- S-
```

pGWB25, 26, 27　(T7)　33bp

```
ATGGCTAGCATGACTGGTGGACAGCAAATGGGT
 M- A- S- M- T- G- G- Q- Q- M- G
```

PGWB28, 29　(TAP)　546bp

AAGAGAAGATGGAAAAAGAATTTCATAGCCGTCTCAGCAGCCAACCGCTTTAAGAAAATCTCATCCTCC

K- R- R- W- K- K- N- F- I- A- V- S- A- A- N- R- F- K- K- I- S- S- S-

GGGGCACTTGATTATGATATTCCAACTACTGCTAGCGAGAATTTGTATTTTCAGGGTGAGCTCAAAACC

G- A- L- D- Y- D- I- P- T- T- A- S- E- N- L- Y- F- Q- G- E- L- K- T-

GCGGCTCTTGCGCAACACGATGAAGCCGTGGACAACAAATTCAACAAAGAACAACAAAACGCGTTCTAT

A- A- L- A- Q- H- D- E- A- V- D- N- K- F- N- K- E- Q- Q- N- A- F- Y

GAGATCTTACATTTACCTAACTTAAACGAAGAACAACGAAACGCCTTCATCCAAAGTTTAAAAGATGAC

E- I- L- H- L- P- N- L- N- E- E- Q- R- N- A- F- I- Q- S- L- K- D- D-

CCAAGCCAAAGCGCTAACCTTTTAGCAGAAGCTAAAAAGCTAAATGATGCTCAGGCGCCGAAAGTAGAC

P- S- Q- S- A- N- L- L- A- E- A- K- K- L- N- D- A- Q- A- P- K- V- D-

AACAAATTCAACAAAGAACAACAAAACGCGTTCTATGAGATCTTACATTTACCTAACTTAAACGAAGAA

N- K- F- N- K- E- Q- Q- N- A- F- Y- E- I- L- H- L- P- N- L- N- E- E-

CAACGAAACGCCTTCATCCAAAGTTTAAAAGATGACCCCAAGCCAAAGCGCTAACCTTTTAGCAGAAGCT

Q- R- N- A- F- I- Q- S- L- K- D- D- P- S- Q- S- A- N- L- L- A- E- A-

AAAAAGCTAAATGGTGCTCAGGCGCCGAAAGTAGACGCGAATTCCGCGGGGAAGTCAACCTGA

K- K- L- N- G- A- Q- A- P- K- V- D- A- N- S- A- G- K- S- T- *

(1) The CBP sequence is underlined.

(2) TEV site is underlined in bold.

(3) The protein A sequence is double-underlined.

Acknowledgements