Fire Lab 1997 Vector Supplement- February 1997

***gfp* variants** **PCR-based *gfp* tagging**
restriction-site-based *gfp* tagging ***gfp* intron-tagging**
broad expression ***smg*-dependent expression vectors**

Si-Qun Xu, Bill Kelly, Brian Harfe[1], Mary Montgomery, Joohong Ahnn[2], Staci Getz[3], Andrew Fire

Carnegie Institution of Washington, Department of Embryology
115 West University Parkway
Baltimore, Md. 21210 USA
[1]Graduate Program in Biology, Johns Hopkins University, Baltimore, Md.
[2]Present Address: Kwangju Institute of Science and Technology, Kwangju, Korea
[3]Undergraduate Program in Biology, College of Notre Dame, Baltimore, Md.

The vector kit has four components

1. This handout, with preliminary description of the new vectors

2. The microtiter vector array, containing vectors described in this handout.

Contained in this package, store -20°C. Instructions for handling the samples are described on the microtiter wills and on p.20 of this handout

3. A C. elegans strain carrying the ts smg-1 mutation smg-1(cc546ts): One strain is being sent with this kit {PD8120 [smg-1(cc546ts)]}. Three other strains {PD8117 [smg-1(cc545ts)unc-54(r293)]}, {PD8118 [smg-1(cc546ts)unc-54(r293)]}, and {PD8119 [smg-1(cc545ts)]} are being sent to the *C. elegans* stock center [St. Paul, Mn.], and can be requested from there as needed.

4. The address of the ciw vector archive (web address: www.ciwemb.edu).

I. Summary

This kit contains several sets of vectors that should be useful in studies of gene expression and function in *C. elegans*. The vectors in this supplementary kit are in many cases derivatives of our earlier vectors. Documentation on the earlier vectors has been provided, and is still available from our web site (www.ciwemb.edu). In many cases, the vectors in this current kit have components or segments that can be simply excised and inserted in previous constructions to obtain a plasmid with desired properties. In working with this kit, we are assuming that the user has experience and knowledge of sequence-based design and construction of plasmids. We are providing both a general description and the full predicted DNA sequence for each of the constructs. Users should then be able to use their own software and design facilities to generate the restriction maps and other information needed for detailed construct design.

The present kit contains the following

A. Fusion constructs with the coding region for blue-shifted *gfp*
B. Fusion constructs with the coding regions for other *gfp* protein variants
C. Vectors for tagging of *C. elegans* exons with *gfp*
D. Vectors for tagging of *C. elegans* introns with *gfp*
E. Vectors for facilitated *gfp*-tagging using PCR
F. Vectors for general expression of specified coding region
G. Vectors for smg-dependent expression
H. Some notes on the use of excess Caenorhabditis genomic DNA as carrier to prevent transgene silencing.
I. A variety of miscellaneous constructs that may be of some general use.

Suggestions for the "vector novice" and the "vector pro"

This supplementary kit is not intended to replace earlier expression vector kits sent out by this laboratory [see reference 6]. If you are new to *C. elegans*, to studies with *gfp*, or to DNA-mediated transformation, you may wish to review literature from this lab and elsewhere concerning the nature of the system [see references]. If you have already produced *gfp* fusions and are hoping to obtain improved expression, we make the following suggestions (in order):

1. The use of excess Caenorhabditis genomic DNA as carrier has in many cases resulted in improved uniformity or accuracy in expression pattern. This procedure is described in detail in section H.

2. The most widely used vectors from the 1995 kit generally contain five introns (three within gfp and one on each side). A six-intron reporter segment (from pPD94.81 in the 1995 kit) and a more recent seven-intron reporter (from pPD104.91 in this supplement) may offer significant improvements in expression for certain fusions.

3. We tested a number of gfp variants for cytosolic activity. With our hardware (standard FITC filters), we find that gfp[S65C] shows the best photo-stability. The S65C (9) was included in most vectors in the 1995 kit. The double mutant F64LS65T has subsequently been reported as highly active [4]. We know of no well controlled experiment in which these two forms have been compared in a quantitative assay. Under our conditions we see no major difference in initial signal between gfp[S65C] and gfp[F64LS65T]. Thus for cytosolic (and likely mitochondrial components), we find that none of the currently available gfp variants is significantly preferable to gfp[S65C]. Nonetheless, we would certainly encourage comparisons by anybody who might have direct means of comparing the different gfp forms. The [F64L S65T] may be preferred for secreted components, although our experience with this form is limited.

II. Description of individual vectors

A. Blue GFP constructs

Heim and Tsien [7] reported an efficient blue fluorescing variant of GFP containing two amino acid substitutions: tyr66>his and tyr145>phe. We produced these substitutions in the context of our intron-containing *gfp* and *gfp*::*lacZ* cassettes. These give a blue fluorescent signal using standard "DAPI" filter sets (standard GFP version gives a green signal with DAPI filter sets). Under these illumination conditions, the Y66HY145F blue fluorescence fades much more rapidly than standard GFP. We are investigating the use of different filter sets to improve stability. In the meantime, for our applications, we are somewhat unsure of the usefulness of this variant for double labeling.

The blue *gfp* constructs are in test plasmids driven by the *unc-54* and *myo-3* promoters (which drive expression in body wall muscle). Restriction sites flanking and within *gfp* allow easy exchange of coding regions with equivalent regions in existing vectors or fusion constructs (Available rare sites 5' of the mutated *gfp* are *KpnI*, *AgeI*, *NcoI*; Rare sites 3' to the mutated *gfp* are *Acc*I, *MfeI*, *Eco*RI).

Table 1: constructs with blue-shifted *gfp* (Y66H;Y145F referred to as [gf22] below).
Tested by spectral analysis of fluorescence properties of transient transformants and heritably transformed lines)

plasmid	lig	promoter	localization	reporter	introns	3' end	decoy
pPD115.57	L3567	*myo-3*	cytoplasm	*gfp* (Y66H;Y145F)	Aåß∂L	*unc-54*	yes
pPD116.11	L3618	*unc-54*	SV40 NLS	*gfp* (Y66H;Y145F)::*lacZ*	Aåß∂B-L	*unc-54*	none

B. Other GFP variants

We have produced a number of other amino-acid sequence variants in the context of our intron-rich *gfp* coding region. The amino-acid variants include some based on published literature, and a few combinatorial mutations. At the outset we should say that for standard applications with cytosolic or mitochondrial *gfp* signals, we still recommend the S65C form which is the basis for many of the constructs in the original 1995 kit. From the literature, some of the newer variants might be expected to have different spectral properties and thus be advantageous in specific applications. We have no simple quantitative way of comparing GFP spectra; we would certainly encourage any enthusiastic laboratory to generate such data.

As with section **I** above, these GFP mutations have been produced in test constructs carrying the *gfp* variant driven in body wall muscle by the *unc-54* or *myo-3* promoters. Any of the variants can be exchanged into existing vectors or fusion constructs by a simple restriction fragment swap.

Table 2:
Variant GFP forms (including those in 1997 kit)

"Wild type"	[gf1]
S65T	[gf2]
S65C	[gf3]
F64L S65T	[gf4]
S65A V68L S72A	[gf5]
S65G S72A	[gf6]
Y66H	[gf7]
Y66W	[gf8]
F64L S65T Y145F	[gf9]
F64L S65T N146I	[gf10]
S65C M153A	[gf11]
F64L S65T N146I M153A	[gf12]
F64L S65T N146I M153T V163A	[gf13]
Y66H N146I M153T V163A	[gf14]
Y66W N146I M153T V163A	[gf15]
Y66H N146I	[gf16]
Y66W N146I	[gf17]
F64L S65T Y145F M153A	[gf18]
F64L S65T Y145F M153T V163A	[gf19]
Y66H Y145F M153T V163A	[gf20]
Y66W Y145F M153T V163A	[gf21]
Y66H Y145F "Blue *gfp* above"	[gf22]
Y66W Y145F	[gf23]
S65T Y145F	[gf24]
F64L S65T M153A	[gf25]
S65T M153A	[gf26]
F64L S65T M153T V163A	[gf27]
Y66H M153T V163A	[gf28]
Y66W M153T V163A	[gf29]
Y66H N146I M153A	[gf30]
Y66W N146I M153A	[gf31]
S65T N146I M153A	[gf32]
S65T N146I M153T V163A	[gf33]
S65T N146I	[gf34]
Y66H Y145F M153A	[gf35]
Y66W Y145F M153A	[gf36]
S65T Y145F M153A	[gf37]
S65T Y145F M153T V163A	[gf38]
Y66H M153A	[gf39]
Y66W M153A	[gf40]
S65T M153T V163A	[gf41]

The *gfp* modifications described above are from studies in which bacteria were used to select for improved *gfp* fluorescence: gf4-gf6 are from Cormack et al [3]; the other changes are inspired by the work of Heim et al. [7-9].

We now have some experience with variants 2, 3, 4, 5, 6, 13, 15, 19, 22, 23, 25, and 26. All of these show marked improvement in (as observed by standard FITC illumination) over wild-type *gfp*. All forms with S65 modifications cause an excitation "red shift", reducing fluorescence with near UV (e.g. DAPI, Hoechst filter sets), and yielding a surprisingly strong "red signal" using rhodamine filter sets (green illumination). We saw comparable initial activity levels from gf2, 3, 4, 5, 6 & 25. The gf3 [S65C] variant exhibited the best photo-stability, [gf4] and [gf5] showed intermediate photo-stability, while the gf2 [S65T], gf6 [S65G S72A] and gf25 [F64L S65T M153A] variants appeared less stable to photobleaching.

Table 3: constructs with variant *gfp* forms

The gfp fluorescence properties of these have not been fully characterized. Note that the constructs all express in body wall muscle but have slightly different expression signals. These differences shouldn't affect chromatic properties, but will effect expression level and uniformity. For any of these constructs, a restriction fragment (e.g. NcoI-MfeI) can be used to transfer the relevant mutated region of *gfp* to existing vectors or constructs without alteration in reading frame.

plasmid	lig	promoter[1]	localization[2]	reporter	designation	introns[3]	3' end	decoy
pPD104.53	L2944	*unc-54*	SV40 NLS	F64L S65T	[gf4]	AåßðΣL	*unc-54*	no
pPD104.33	L2946	*unc-54*	SV40 NLS	S65A V68L S72A	[gf5]	AåßðΣL	*unc-54*	no
pPD104.64	L2947	*unc-54*	SV40 NLS	S65G S72A	[gf6]	AåßðΣL	*unc-54*	no
pPD115.02	L3513	*myo-3*	cytoplasm	Y66H	[gf7]	AåßðL	*unc-54*	yes
pPD115.16	L3514	*myo-3*	cytoplasm	Y66W	[gf8]	AåßðL	*unc-54*	yes
pPD114.95	L3516	*myo-3*	cytoplasm	F64L S65T Y145F	[gf9]	AåßðL	*unc-54*	yes
pPD114.92	L3515	*myo-3*	cytoplasm	F64L S65T N146I	[gf10]	AåßðL	*unc-54*	yes
pPD114.98	L3517	*myo-3*	cytoplasm	S65C M153A	[gf11]	AåßðΣL	*unc-54*	yes
pPD115.42	L3557	*myo-3*	cytoplasm	F64L S65T N146I M153A	[gf12]	AåßðΣL	*unc-54*	yes
pPD115.44	L3558	*myo-3*	cytoplasm	F64L S65T N146I M153T V163A	[gf13]	AåßðΣL	*unc-54*	yes
pPD115.45	L3559	*myo-3*	cytoplasm	Y66H N146I M153T V163A	[gf14]	AåßðΣL	*unc-54*	yes
pPD115.46	L3560	*myo-3*	cytoplasm	Y66W N146I M153T V163A	[gf15]	AåßðΣL	*unc-54*	yes
pPD115.48	L3561	*myo-3*	cytoplasm	Y66H N146I	[gf16]	AåßðL	*unc-54*	yes
pPD115.50	L3562	*myo-3*	cytoplasm	Y66W N146I	[gf17]	AåßðL	*unc-54*	yes
pPD115.51	L3563	*myo-3*	cytoplasm	F64L S65T Y145F M153A	[gf18]	AåßðΣL	*unc-54*	yes
pPD115.53	L3564	*myo-3*	cytoplasm	F64L S65T Y145F M153T V163A	[gf19]	AåßðΣL	*unc-54*	yes
pPD115.54	L3565	*myo-3*	cytoplasm	Y66H Y145F M153T V163A	[gf20]	AåßðΣL	*unc-54*	yes
pPD115.55	L3566	*myo-3*	cytoplasm	Y66W Y145F M153T V163A	[gf21]	AåßðΣL	*unc-54*	yes
pPD115.57	L3567	*myo-3*	cytoplasm	Y66H Y145F	[gf22]	AåßðL	*unc-54*	yes
pPD115.60	L3568	*myo-3*	cytoplasm	Y66W Y145F	[gf23]	AåßðL	*unc-54*	yes
pPD115.61	L3569	*unc-54*	SV40 NLS	S65T Y145F	[gf24]	AåßðL	*unc-54*	yes
pPD115.62	L3570	*myo-3*	cytoplasm	F64L S65T M153A	[gf25]	AåßðΣL	*unc-54*	yes
pPD115.64	L3571	*unc-54*	SV40 NLS	S65T M153A	[gf26]	AåßðΣL	*unc-54*	yes
pPD115.66	L3572	*myo-3*	cytoplasm	F64L S65T M153T V163A	[gf27]	AåßðΣL	*unc-54*	yes
pPD115.69	L3573	*myo-3*	cytoplasm	Y66H M153T V163A	[gf28]	AåßðΣL	*unc-54*	yes
pPD115.70	L3574	*myo-3*	cytoplasm	Y66W M153T V163A	[gf29]	AåßðΣL	*unc-54*	yes
pPD115.95	L3602	*myo-3*	cytoplasm	Y66H N146I M153A	[gf30]	AåßðΣL	*unc-54*	yes
pPD115.97	L3603	*myo-3*	cytoplasm	Y66W N146I M153A	[gf31]	AåßðΣL	*unc-54*	yes
pPD115.99	L3604	*unc-54*	SV40 NLS	S65T N146I M153A	[gf32]	AåßðΣL	*unc-54*	yes
pPD115.101	L3605	*unc-54*	SV40 NLS	S65T N146I M153T V163A	[gf33]	AåßðΣL	*unc-54*	yes
pPD115.104	L3606	*unc-54*	SV40 NLS	S65T N146I	[gf34]	AåßðL	*unc-54*	yes
pPD115.105	L3607	*myo-3*	cytoplasm	Y66H Y145F M153A	[gf35]	AåßðΣL	*unc-54*	yes
pPD115.107	L3608	*myo-3*	cytoplasm	Y66W Y145F M153A	[gf36]	AåßðΣL	*unc-54*	yes
pPD115.109	L3609	*unc-54*	SV40 NLS	S65T Y145F M153A	[gf37]	AåßðΣL	*unc-54*	yes
pPD115.111	L3610	*unc-54*	SV40 NLS	S65T Y145F M153T V163A	[gf38]	AåßðΣL	*unc-54*	yes
pPD115.114	L3611	*myo-3*	cytoplasm	Y66H M153A	[gf39]	AåßðΣL	*unc-54*	yes
pPD115.116	L3612	*myo-3*	cytoplasm	Y66W M153A	[gf40]	AåßðΣL	*unc-54*	yes
pPD115.118	L3613	*unc-54*	SV40 NLS	S65T M153T V163A	[gf41]	AåßðΣL	*unc-54*	yes

[1]The two promoters (*unc-54* and *myo-3*) both function in bodywall muscle but have somewhat different levels; in addition, the *unc-54* promoter shows distinct mosaicism in repetitive array contexts.

[2]The NLS-GFP protein encoded by some of these constructs is relatively small and appears not to be efficiently retained in the nucleus. Thus all of the above constructs produce predominantly cytoplasmic localization.

[3]The Σ intron present in some of these constructs is just 3' to the *gfp* coding region. This intron significantly boosts expression of certain fusions. An S65C version of gfp carrying intron Σ is available in the 1995 kit (pPD94.81=L2406)

Table 4 *gfp::lacZ* fusion constructions with diverse *gfp*'s

These all have parallel structures, with the only differences being in the *gfp* coding sequence. Each construct is driven by the minimal *unc-54* promoter in body wall muscle, each produces a fusion protein with the SV40 nuclear localization signal attached to a *gfp*::lacZ fusion. The large *gfp*::lacZ fusion is well retained within the nucleus, so that the expression from these constructs is highly nuclear. pPD96.02 is the prototype for this series, and was included in the previous (1995) kit.

plasmid	lig	promoter	localization	reporter	gf#	introns	3' end	decoy
pPD96.02	L2472	*unc-54*	SV40 NLS	*gfp*(S65C)::lacZ	[gf2]	AåßðB-L	*unc-54*	no
pPD116.03	L3615	*unc-54*	SV40 NLS	*gfp*(F64L S65T N146I M153T V163A)::lacZ	[gf13]	AåßðB-L	*unc-54*	no
pPD116.06	L3616	*unc-54*	SV40 NLS	*gfp*(Y66W N146I M153T V163A)::lacZ	[gf15]	AåßðB-L	*unc-54*	no
pPD116.07	L3617	*unc-54*	SV40 NLS	*gfp*(F64L S65T Y145F M153T V163A)::lacZ	[gf19]	AåßðB-L	*unc-54*	no
pPD116.11	L3618	*unc-54*	SV40 NLS	*gfp*(Y66H Y145F)::lacZ	[gf22]	AåßðB-L	*unc-54*	no
pPD116.14	L3619	*unc-54*	SV40 NLS	*gfp*(Y66W Y145F)::lacZ	[gf23]	AåßðB-L	*unc-54*	no
pPD116.16	L3620	*unc-54*	SV40 NLS	*gfp*(F64L S65T M153A)::lacZ	[gf25]	AåßðB-L	*unc-54*	no
pPD116.21	L3621	*unc-54*	SV40 NLS	*gfp*(S65T M153A)::lacZ	[gf26]	AåßðB-L	*unc-54*	no

C. Vectors for tagging of *C. elegans* exons with *gfp*

In many cases, it is desirable to study a functional gene by the conservative insertion of *gfp* coding sequences (i.e. not losing any sequences from the original gene). In general, this is carried out by inserting *gfp* coding sequences "in-frame" while maintaining all the original sequences. The resulting GFP-tagged gene products often retain the function of the original gene while acquiring fluorescence from the GFP component. GFP appears remarkably well suited to such a tagging approach, since it can retain its fluorescence properties in the context of both C-terminal and N-terminal amino acid extensions. In addition, GFP can confer fluorescence properties in a wide variety of cellular compartments, including (but by no means limited to) cytoplasm, nucleus, mitochondria, and extracellular spaces.

Construction of in-frame fusions can be carried out by several means. The most general is to choose optimal points within protein sequence for GFP insertion, then use site directed mutagenesis and PCR to insert *gfp* precisely at those sites. In other cases, there are existing unique restriction sites in the coding region that can be used for direct insertion of *gfp*. In the latter case, PCR could be used to generate GFP coding sequences with appropriate linker sequences for in-frame insertion. This requires rather careful analysis of the final product to rule out mutations in the oligonucleotides used for PCR, or in the intervening material.

As an alternative, we have produced a set of six pre-made *gfp* cassettes flanked with numerous restriction sites placed in all different reading frames. The current set of vectors is sufficiently complete that virtually any common restriction site in any reading frame can be insertion-tagged with *gfp*. All of these constructs contain a *gfp* coding region with three internal intron sequences (åßð). Different amino acid sequence forms of GFP are available as noted.

Figure 1: General structure of exon tagging vectors

```
---R1-R2-R3-R4-R5---GfpCodingSequences---R1-R2-R3-R4-R5---
```

Where R1-Rn are restriction enzyme sites with different overhanging ends. Vectors are designed so that cutting with any single enzyme produces a single reading frame entering and leaving *gfp*. Six different vectors with distinct multiple-cloning-site regions have been produced and are sufficient to tag virtually any common restriction site.

D. Vectors for insertional tagging of *C. elegans* introns with *gfp*.

In some cases where a unique restriction site is found in an intron, it has been possible to "tag" the gene of interest by inserting *gfp* with flanking splice junctions. The geometry of this is shown below.

Figure 2. Construction of *gfp*-tagged genes using intron insertion

```
A.  Original Gene

  EXON-[5'splice site]-INTRON-[3'splice site]-EXON
                              |
                 unique restriction site [RX]

B. gfp segment for insertion into unique restriction site [RX].

  RX-[3' Splice site]-GFP CODING SEQUENCE-[5' Splice Site]-RX
```

We provide three such intron-insertion vectors, covering all reading frames. A variety of restriction sites (including *EcoRI, BamHI, XmaI, SalI, Hind3, MluI, NotI, SfcI, XbaI, ClaI, SacI, ApaI, PstI*, and several blunt sites) are duplicated both upstream and downstream of the splice junctions flanking *gfp*. These are otherwise unique, so that each is available for constructing insertion clones. To the extent that splice junctions can occur at hinges or domain boundaries in protein coding sequences, it might be expected that this scheme will often yield functional chimeric proteins. All of these constructs contain a *gfp* coding region with three internal intron sequences (åßð). Different sequence forms of GFP are available as noted.

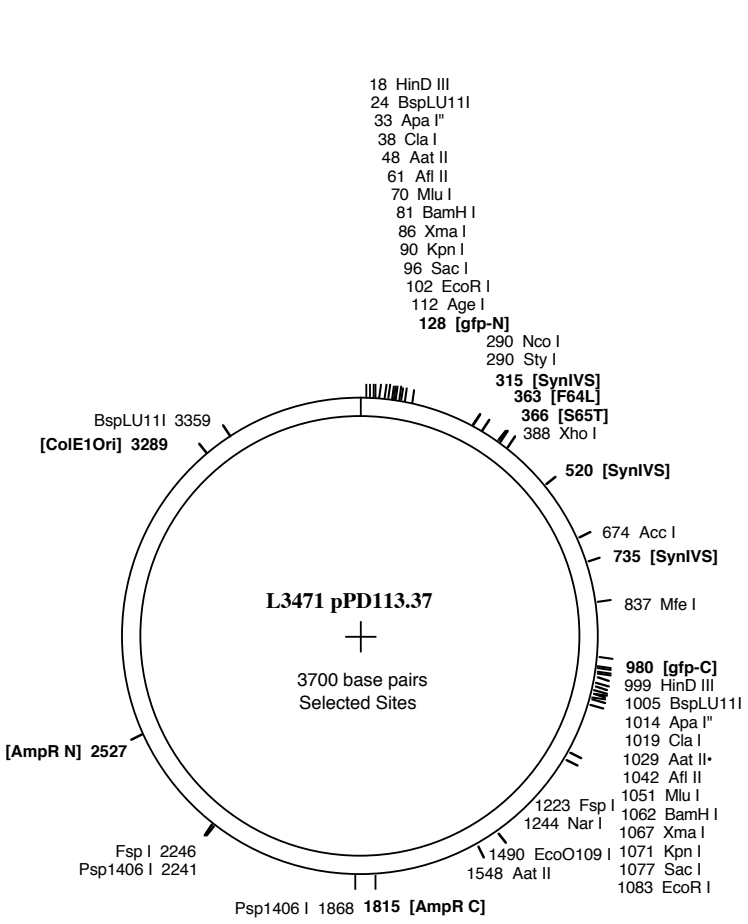
E. Vectors for assembly of *gfp*-tagged constructs using PCR.

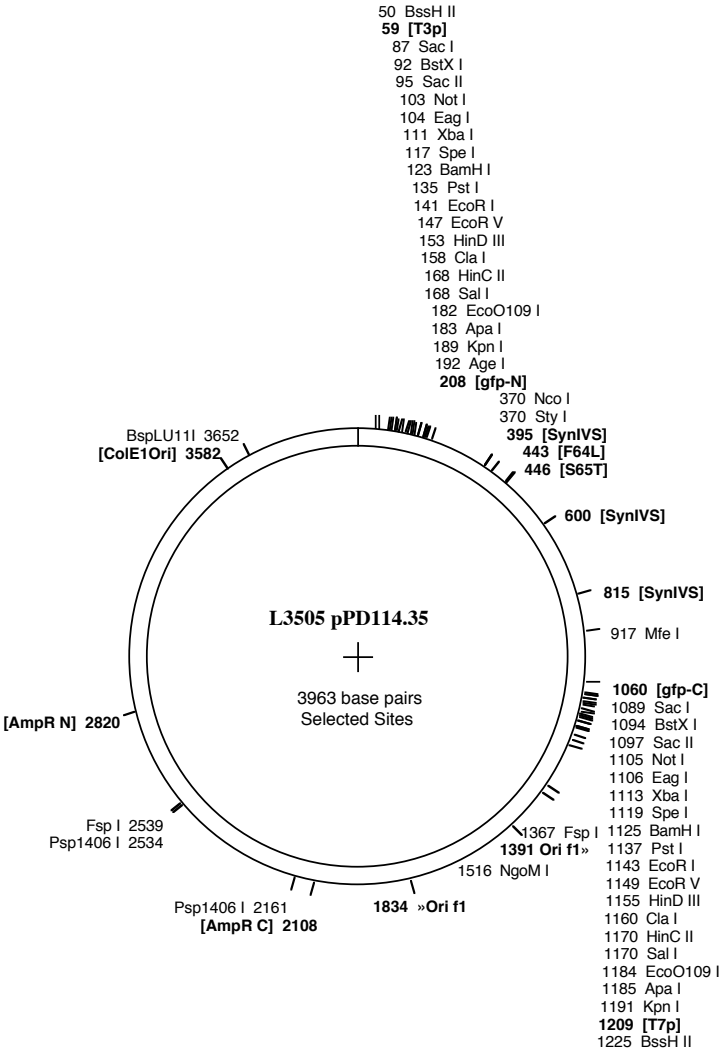
For PCR-based tagging schemes, we have constructed a set of *gfp* vectors which are slightly modified versions of currently available vectors. The modifications involve the addition of restriction sites in N-terminal and C-terminal coding regions flanking *gfp*. These vectors save several steps in PCR-based construction of chimeric genes in which the coding region of interest is to be placed either 5', 3', or surrounding *gfp*. In particular these vectors should be of significant utility in cases where the sequence of a gene is available (e.g., from the sequencing project) before the availability of a well-defined subclone. These vectors already have both a promoter (*mec-7*) and a 3' UTR region (from *let-858*). These have been included for two reasons

1) to allow the efficient cloning of asymmetric restriction fragment by replacement *and*
2) to allow characterization in (non-essential) touch cells of *geneX-gfp* fusion protein localization and toxicity.

Maps of exemplary exon insertion vectors

Note that *gfp* coding sequences (F64L S65T variant in these cases) are interrupted by three canonical intron sequences from *C. elegans* , and are flanked by a variety of rare-cutting restriction sites. Each of these sites are present in the same frame both 5' to and 3' to *gfp*. Thus excision with any of these enzymes produces a linear restriction fragment which can be inserted into an appropriate restriction site in a target molecule of choice. All vectors of this type are similar, with the major differences being the restriction sites flanking *gfp* and their reading frames. The "overhang" tables on the next page indicate the appropriate vector to use for the insertion, given the reading frame through the relevant restriction site in the target sequence. The vectors all have similar structures in the *gfp* region; note however that the KS0 and KS1 based vectors are constructed from a phagemid vector (Bluescript II KS+), while the MCS3-MCS6 vectors were constructed from pUC19-derived plasmid vectors.





Recommended scheme for using exon insertion vectors:

1) Generate a precise predicted sequence for the entire genomic clone that you will be starting with. This is a computer job, merging the genomic sequence that you are studying with the vector that you had used for subcloning. We use the program DNA Strider [14]; other software packages are also available for such manipulations [13]. We find that precise and complete "computer predicted restriction maps" are essential in design of DNA constructs.

2) Find restriction sites in the coding region of your gene which are unique in the entire genomic clone. Determine the overhang sites for these restriction sites and the reading frames through each site.

3) Find the overhang sequence in the left column of the following table. To the right of each overhang sequence, the three reading frames are diagrammed. For each reading frame and overhang, there is reference to one of the six "MCS" sequences (KS0, KS1, MCS3, MCS4, MCS5, or MCS6) and a restriction enzyme to cleave the corresponding vector.

4) Using the table below, choose a vector with the desired MCS sequence and a *gfp* form appropriate for the type of molecule to be tagged (we recommend S65C for cytosolic/mitochondrial components, F64LS65T for secretion).

5) Cut your genomic clone with the unique enzyme, and treat the restriction digest with alkaline phosphatase. Cut the chosen *gfp* vector with the appropriate enzyme. Gel purify both fragments, ligate, transform *E. coli*. Check miniprep DNA for *gfp* orientation and (if appropriate) for regeneration of the restriction sites at the ligation junctions.

6) We recommend assays from standard tandem arrays (e.g., plasmid co-injections with pRF4), as well as more physiological "complex" arrays made by diluting the transforming DNA with excess *C. elegans* genomic DNA [sect **H**].

Exon Insertion MCS table- overhang and reading frame determine relevant MCS

4-base 5'			
AATT	-aa tt- EcoRI KS1	aat t-- EcoRI KS0	--a att EcoRI MCS4
AGCT	-ag ct- HindIII KS1	agc t-- HindIII KS0	--a gct HindIII MCS4
CATG	-ca tg- FokI MCS4	cat g-- BspLU11I MCS4	--c atg BspLU11I MCS3
CCGG	-cc gg- XmaI MCS3	ccg g-- XmaI KS0	--c cgg XmaI MCS4
CGCG	-cg cg- MluI MCS6	cgc g-- MluI MCS3	--c gcg MluI MCS4
CTAG	-ct ag- XbaI KS1	cta g-- XbaI KS0	--c tag STOP
GATC	-ga tc- BamHI KS1	gat c-- BamHI KS0	--g atc BamHI MCS3
GGCC	-gg cc- Bsp120I MCS5	ggc c-- NotI KS1	--g gcc NotI KS0
GTAC	-gt ac- Asp718I MCS6	gta c-- Asp718I KS0	--g tac Asp718I MCS3
TCGA	-tc ga- SalI KS1	tcg a-- SalI KS0	--t cga SalI MCS5
TGCA	-tg ca- SfcI KS1	tgc a-- SfcI KS0	--t gca SfcI MCS6
TTAA	-tt aa- AflII MCS6	tta a-- AflII MCS3	--t taa STOP
3-base 5': Some EcoO109 sites can be tagged (not all), depending on overhang sequence			
GGC	-gg c-- EcoO109I MCS5	ggc --- EcoO109I KS0	
2-base 5'			
CG	--c g-- ClaI KS1	-cg --- ClaI KS0	--- cg- ClaI MCS3
TA	--t a-- AseI MCS5	-ta --- AseI MCS4	--- ta- [n.a.]
4-base 3'			
ACGT	-ac gt- AatII MCS5	acg t-- AatII MCS4	--a cgt AatII MCS3
AGCT	-ag ct- SacI MCS5	agc t-- SacI KS0	--a gct SacI MCS3
CATG	-ca tg-	cat g-- NspHI MCS4	--c atg NspHI MCS4
GGCC	-gg cc- ApaI MCS5	ggc c-- ApaI KS0	--g gcc ApaI MCS3
GTAC	-gt ac- KpnI MCS6	gta c-- KpnI KS0	--g tac KpnI MCS3
TGCA	-tg ca- PstI KS1	tgc a-- PstI KS0	--t gca PstI MCS6
2-base 3'			
GC	--g c-- SacII MCS5	-gc --- BsiE1 KS1	--- gc- SacII KS0
AT	--a t-- [n.a.]	-at --- [n.a.]	--- at- PacI MCS5
Blunt	--A^T--	-A^T	A^T-
	--C^G-- SmaI MCS3	-C^G Eco47III MCS6	C^G-
	--G^C-- Ecl136II MCS5	-G^C Ecl136II KS0	G^C- Ecl136II MCS3
	--T^A-- EcoRV KS1	-T^A EcoRV KS0	T^A-

"STOP", [n.a.] Vector not made; these frames would include stops with relevant enzymes

Intron insertion MCS's- each has all of the 5',3' & blunt sites shown below

Frame 0 nnn [intron] nnn nnn (MCS "IIV0")
Frame 1 nnn n [intron] nn nnn (MCS "IIV1")
Frame 2 nnn nn [intron] n nnn (MCS "IIV2")

5' Overhang: /AATT (EcoRI), /GATC (BamHI), /CCGG (XmaI), /TCGA (SalI), /AGCT (Hind3), /CGCG (MluI), /GGCC (NotI), /TGCA (SfcI), /CTAG (XbaI, SpeI), /GGC (EcoO109), /CG (ClaI)

3' Overhang: AGCT/ (SacI), CGCG/ (BstXI), GGCC/ (ApaI), TGCA/ (PstI), TGG/ (MwoI) GAG/ (MslI), T/ (XcmI)

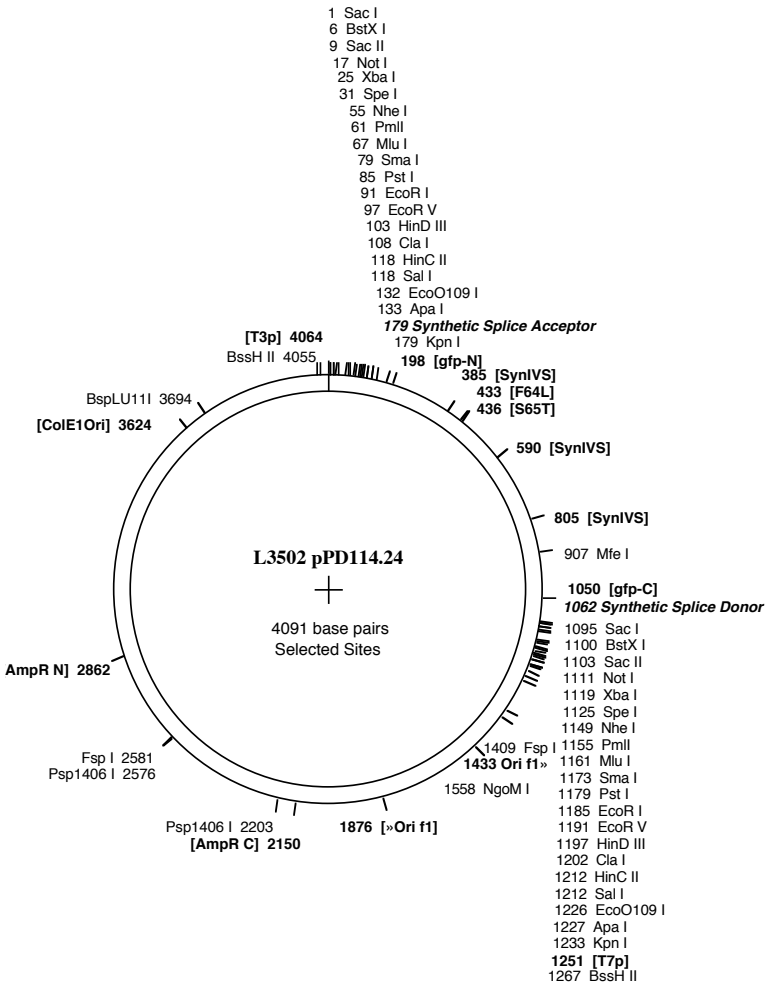
Blunt: GAG/CTC (Ecl136II), CACCG/CGGTG (MslI), CCG/CTC (BsrBI), CAC/GTG (PmlI) CCC/GGG (SmaI), GAT/ATC (EcoRV), GTC/GAC (HincII)

Vector names and ligation numbers

MCS	S65C	F64L & S65T
KS0	L2822 pPD102.33	L3505 pPD114.35
KS1	L2911 pPD103.87	L3506 pPD114.38
MCS3	L3829 pPD119.16	L3471 pPD113.37
MCS4	L3863 pPD119.45	L3465 pPD113.05
MCS5	L3827 pPD118.85	L3467 pPD113.54
MCS6	L3828 pPD118.90	L3469 pPD113.29
IIV0	L2908 pPD103.75	L3502 pPD114.24
IIV1	L3111 pPD107.45	L3503 pPD114.27
IIV2	L3110 pPD107.48	L3527 pPD114.42

Maps of exemplary intron insertion vectors

Note that *gfp* coding sequences (F64L S65T variant in these cases) are interrupted internally by three canonical intron sequences from *C. elegans* and are flanked by splice junctions (a 3' acceptor site upstream and a 5' donor site downstream). A variety of rare-cutting restriction sites are present outside the splice sites; each of these sites is present in the same frame both 5' to and 3' to the *gfp*. Thus excision with any of these enzymes produces a linear restriction fragment which can be inserted into an appropriate restriction site in an intron of choice. All vectors of this type are similar, with the major differences being the reading frame. The tables above indicate the appropriate vector to use for the insertion, given the nature by which the genomic intron interrupts the natural reading frame. This type of construct relies on the ability of the splicing machinery to treat a splice acceptor or donor site placed in an ectopic position as a true splicing site. This has been the case with several different genes that we have tested, but may not always be the case; hence it is expected that a fraction of constructs prepared using this type of scheme may be poorly functional or inactive.



not shown on the figure are BamHI sites present in duplicate in the 5' and 3' MCS regions... these allow insertion into all BamHI/BglII/BclI sites].

Recommended scheme for using intron insertion vectors:

1) Generate a precise predicted sequence for the entire genomic clone that you will be starting with. This is a computer job, merging the genomic sequence that you are studying with the vector that you had used for subcloning. We use the program DNA Strider [14]; other software packages are also available for such manipulations [13]. We find that precise and complete "computer predicted restriction maps" are essential in design of DNA constructs.

2) Find restriction sites in the introns within your gene which are unique in the entire genomic clone. Determine the type of interruption that each intron makes in the protein reading frame. Since the restriction sites used for the cloning fall within intron sequences in the final construct, there is no concern for the precise reading phase for these sites.

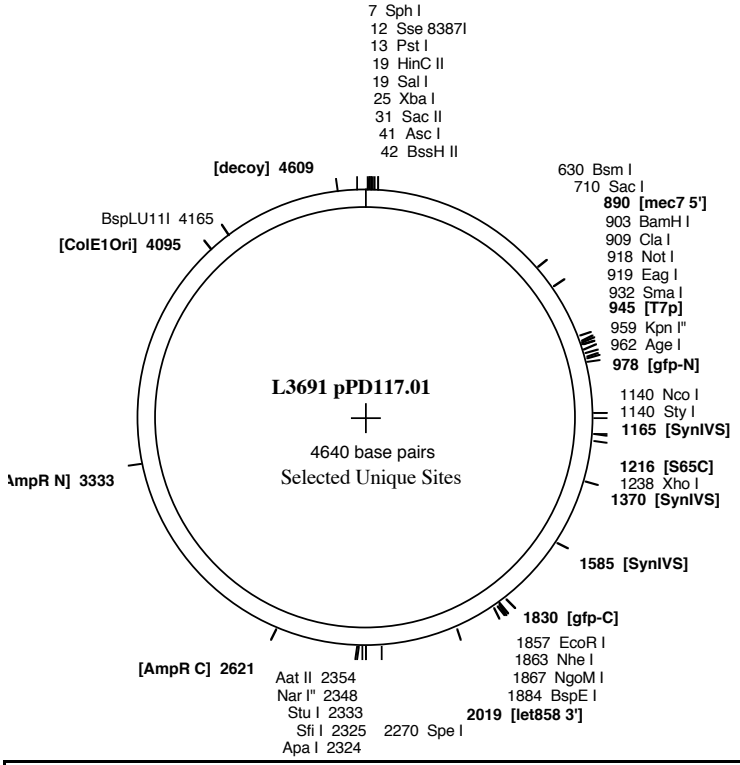
3) Choose the intron insertion MCS (IIV0, IIV1, or IIV2) corresponding to the interruption in the reading frame.

4) Using the table above, choose a vector with the desired MCS sequence and a *gfp* form appropriate for the type of molecule to be tagged ([S65C] is recommended for nuclear, cytoplasmic and mitochondiral proteins, while [F64L S65T] may have advantages for secreted components).

5) Cut your genomic clone with the unique enzyme, and treat the restriction digest with alkaline phosphatase. Cut the chosen *gfp* vector with the appropriate enzyme. Gel purify both fragments, ligate, transform *E. coli*. Check miniprep DNA for *gfp* orientation and (if appropriate) for regeneration of the restriction sites at the ligation junctions.

6) We recommend assays from standard tandem arrays (e.g., plasmid co-injections with pRF4), as well as more physiological "complex" arrays made by diluting the transforming DNA with excess *C. elegans* genomic DNA [sect **H**].

E. Vectors for construction of gfp-tagged constructs using PCR



1. Unique restriction sites are provided for directed insertion of PCR fragments upstream and downstream of *gfp*. Downstream fragments can either be inserted in frame (to make a fusion protein) or using stop codons in *gfp*. Insertions at *Eco*R*I*, *NheI*, or *Ngo*M*I* will translate into 3' sequences from the gene to be tagged. 3' insertions at *Bsp*E*I* will utilize the C-terminus of *gfp* that is provided.

2. Insertion of 5' sequences between *AscI* and *Bam*H*I* (or *XmaI*/*SmaI*) leaves numerous upstream restriction sites to allow insertion of additional 5' flanking DNA. Likewise, 3' insertions can be designed to allow additional material to be added.

3. T7 promoter can allow synthesis of antisense material from cloned PCR fragment. Note that for translational fusions with 5' material from the gene of interest, it is necessary to use 3' fusion sites (*AgeI*, *Asp*718*I*/*KpnI*) that avoid stop codons in the T7 promoter.

4. *NheI* and *Ngo*M*I* sites, along with *SfiI* downstream can be used in replacement of 3'-sequences with a smg-degradable 3' UTR.

5. Different variants in this vector series are available
pPD117.01 (L3691): S65C
pPD114.108 (L3522): F64L/S65T vector without T7, stops

General Scheme: Making a tagged version of *geneX*
1. Insert Long-range PCR "A" (5' end + N-terminus of *geneX*) between 1'st and 2'nd MCS
2. Insert Long-range PCR "B" (C-terminus + 3' end of *geneX*) between 3'rd and 4'th MCS
3. Sequence the coding junctions, also one could injection-test each construct independently for function.
4. Recombine the two constructs to make a *gfp*-tagged *geneX* (e.g., use the unique restriction sites in *gfp*)

First MCS:upstream of promoter (no relevant frame)

```
          SphI Sse8387I    Hinc2                          AscI
(HindIII)      PstI  SalI        XbaI    SacII          BssHII
 v          v     v    v v       v       v              v
 AAGCTT GCATGC CTGCAG   GTCGAC  TCTAGA CCGCGG TTTT GGCGCGCC ---mec7 sequences
```

Second MCS:joins promoter to reporter (pPD114.108 MCS2 sequence is slightly different...see notes)

```
BamHI     ClaI       NotI           XmaI  SmaI    ----T7pol--------------
 v          v         v                v  v        <-
gga tcc tat cga ttc gcg gcc gct gta cac ccg ggt gcc cta tag tga GTC GTA

 KpnI  AgeI
   v   v                  GFP coding region >>
TTG GTA CCG GTA GAA AAA ATG AGT AAA GGA GAA GAA CTT TTC
```

Third MCS: joins reporter to 3' UTR

```
     gfp C-terminus      (HindIII)   EcoRI   NheI   NgoMI
                >                v         v      v    v                    stop
GAc GAA CTA TAC AAA AGC CCA CAA GCT TTG GAA TTC GCT AGC CGG CCA TAC AAG TAA

BspEI
 v   >> let-858 3' coding     UTR>>>
TCC GGA TGATCGACGCCaACGTCGT TGA ATT TTC A
```

Fourth MCS: downstream of 3' end (no relevant frame)

```
                     Bsp120I  ApaI   StuI           NarI    AatII
let-858 3' flank       v   v       v              v       v >backbone
aacgatgagaattgtcctcgctt GGGCCC AAA AGGCCTTTTTTTTTT GGCGCC GACGTC AGG
```

Exemplary long-range PCR oligos (sites in gene may mandate use of different 5' adaptors on these oligos)
PCR "A":**Sense:** (Sse8387I+AscI+23bp of far upstream sense sequence): ttg ggc gcg cct gca gg $(N)_{23}$
Antisense: (XmaI+NotI+21bp of antisense just 5' to ATG): cc cac ccg ggc ggc cgc tAT $(\tilde{N})_{21}$
PCR "B": **Sense:** (NheI+NgoMI+23bp sense starting from ATG): c ggc cgc gct agc cgg cca ATG $(N)_{20}$
Antisense: (NotI+Bsp120I+23bp of downstream antisense sequence): ggc gga gcg gcc gca ggg ccc $(\tilde{N})_{23}$

F. Vectors (intended) for general expression of a specified coding region

Many people have asked for a vector that is capable of expressing any RNA sequence in all tissues at all times. This would indeed be useful. We don't necessarily have such a vector, but we are including in this kit a set of vectors which have the ability to express in a broad range of tissues (no guarantee of uniformity either). These vectors are all based on the *let-858* gene studies by Bill Kelly in this lab. The important observations [12] are

1. that *gfp*-tagged versions of *let-858* can express in all somatic tissues.
2. that some expression of *gfp*-tagged *let-858* constructs can be seen in the germline.

When introduced in standard transformation assays (e.g. by co-injection of circles with circular pRF4), germline expression can be seen at a low level in the first 2-5 generations. Following this, a generational silencing of germline expression is seen: germline expression is generally absent within a few generations, while somatic expression is generally maintained. We have found that this generational silencing can be at least partially overcome when the injection protocol is modified by linearizing the test and marker plasmids and diluting the mixture with an excess of mixed random genomic DNA fragments from *C. elegans*. This protocol is described in section **H** below.

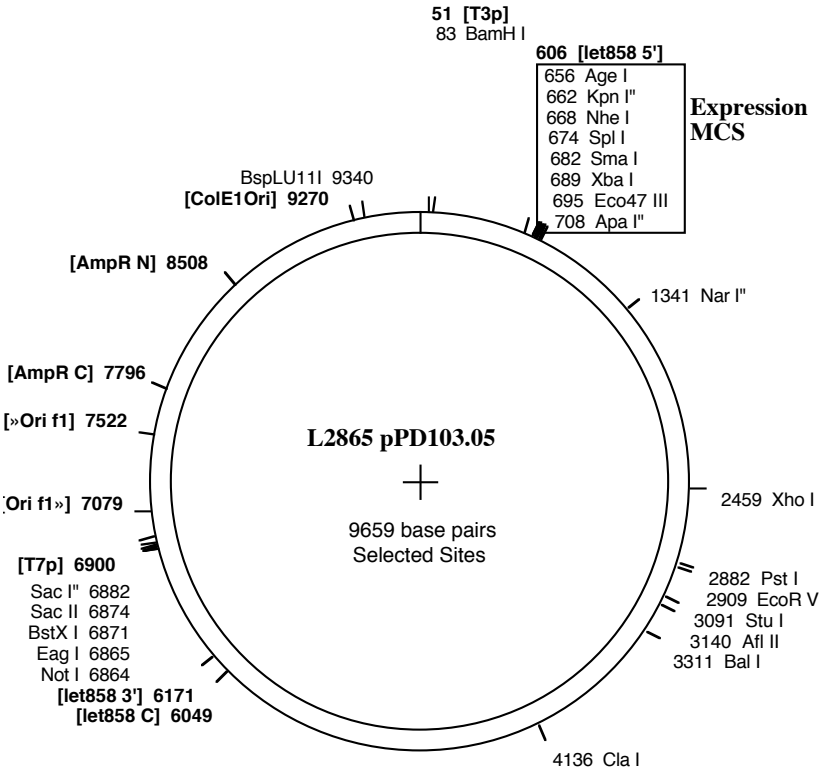
In producing *let-858* based expression vectors, we used a combination of *let-858* internal, upstream and downstream regions. We have no definitive data regarding germline expression from these vectors; as far as we can detect, insertion of a simple reporter (*gfp* or *lacZ*) doesn't give any detectable germline expression. It should be realized that these re-constructed reporter constructs differ significantly from the original *let-858* tagged constructs. Since we have a strong interest in the basis of germline expression, we are currently investigating the nature of both transcriptional and post-transcriptional limitations to germline expression. We would thus be particularly interested in updated information from anyone using these vectors who could detect germline expression. At present we are entertaining several hypotheses (not necessarily exclusive) regarding the apparent selectivity in germline expression

1. The germline may be very finicky about foreign DNA sequences and /or the relative placement of control signals.
2. The germline may be very finicky about what mRNAs are transported, stabilized, or translated.
3. The germline may be very finicky about what proteins are retained and for how long.
4. Certain types of intracellular localization signals may improve the ability to detect expression in the germline.

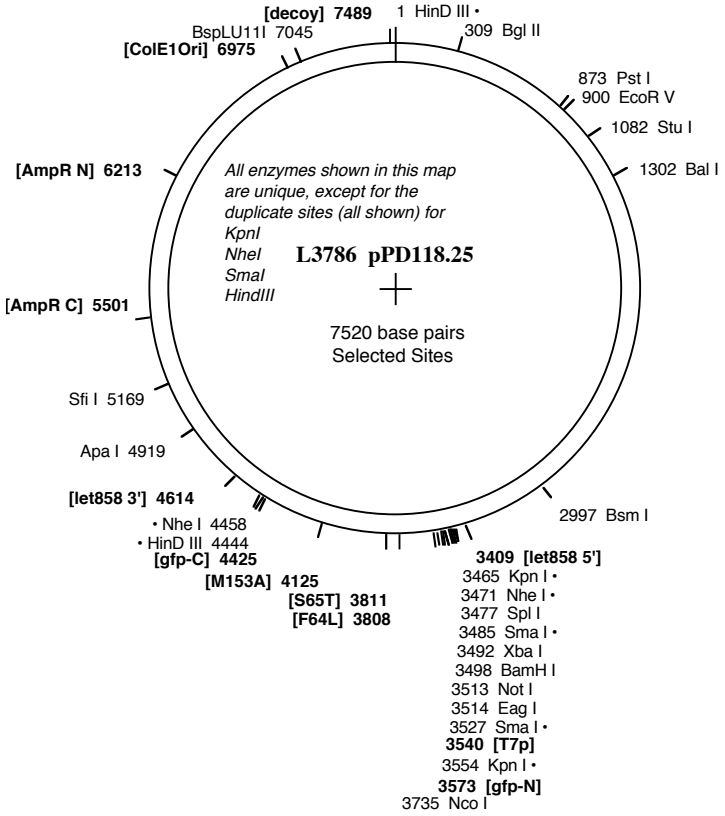
We are currently using the *let-858* system and the derived vectors to investigate the mysteries of germline/soma differences in gene expression. Meanwhile, we suspect that the included promoters will be useful for general somatic expression, for expression of some genes normally native to the germline, and perhaps for a subset of germline ectopic expression and gene fusion experiments.

Construct pPD103.05 (L2865)

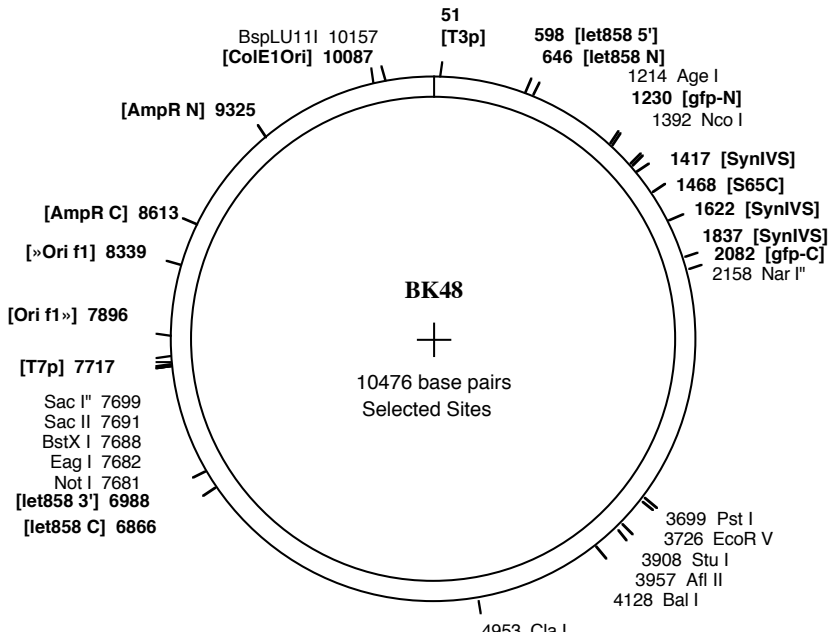
This is a simple *let-858* based expression vector with multiple cloning site inserted in place of the translational start (ATG) for the *let-858* protein. The expectation with this vector [cf ref. 23] is that the user will insert both coding region and 3'UTR into the MCS region, or alternatively insert just a coding region with the construct injected into a Smg$^-$ genetic background to overcome the long 3' sequence created by the presence of the *let-858* coding region.



Constructs pPD118.25 (L3786)
This is actually a *gfp* fusion driven by a hybrid *let-858* promoter. The vector contains (in order) an apparent internal enhancer from *let-858*, the *let-858* "promoter" region (sequences normally upstream of the ATG), a multiple cloning region, coding sequences for *gfp* , and the *let-858* 3' region. The *let-858* promoter can be excised with HindIII (upstream) and any of the restriction sites (*KpnI*, *NheI*, *SplI*, etc) that have been engineered just past the beginning of the *let-858* mRNA. To allow transcriptional fusions, there are no ATG's upstream of the *KpnI...KpnI* MCS. Tagged constructs can in some cases be constructed direcly in L3716 by substituting appropriate sequences from previous constructs. L3716 contains the S65C variant of *gfp*, which is recommended for cytosolic & mitochondrial components.



Construct **BK48**
This is the original *gfp*-tagged derivative of *let-858*, in a Bluescript (AmpR) backbone as described in Kelly et. al. [12]. It is not designed as an expression vector, but may be useful as a positive control for examining germline expression.



G. Vectors for Smg dependent expression.

At the 1995 international *C. elegans* meeting, Kevin Fitzgerald and Rock Pulak suggested to us a scheme for engineering conditional expression from virtually any promoter. This strategy takes advantage of the Smg RNA surveillance system in *C. elegans* [19]. Similar uses of conditional Smg activity had also been proposed by Papp [18] and Ransom et al. [20] The proposed scheme is as follows:

I. Isolate conditional mutations in the Smg system that are unable to degrade aberrant mRNAs at a non-permissive temperature, but are able to do so at a permissive temperature.

II. Construct a fusion DNA that has three components

a. A promoter with a defined activity pattern

b. A coding region to be expressed conditionally

c. A 3' UTR which is arbitrarily extended by irrelevant "junk" sequence.

III. Transform the fusion DNA into the Smg-conditional (e.g., temperature sensitive) host.

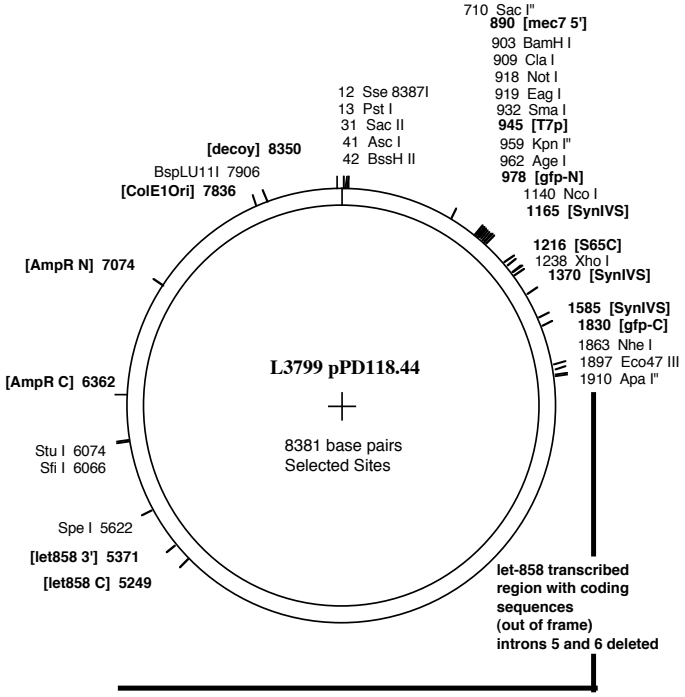
IV. Use temperature shifts to control activity of the transgene.

The expectation is that the extended 3' untranslated region will destabilize the mRNA in a wild type (Smg+) genetic background, while in a Smg- background, the mRNA surveillance will fail and the mRNA will be translated. Thus the transcript is under dual control of the promoter and the state of Smg activity. As one example, a tissue specific promoter could be used to limit expression to one cell type, while a conditional smg background and precisely timed temperature shifts could be used to limit expression to a single stage in the life cycle.

Previously, Papp had reported some smg mutants with partial cold sensitivity. Also *smg-7* alleles show some degree of temperature sensitivity [1]. Since both of these show incomplete temperature dependence, we carried out additional screens. *smg* mutations can readily be isolated following mutagenesis of the smg-suppressible *unc-54* allele, *r293* [19]. Performing this screen at 25°C, we isolated 75 putative *smg* mutations [24]. Of these, three (*cc543*, *cc545*, and *cc546*) showed temperature sensitivity. Two of these (*cc545* and *cc546*) have been shown to be allelic with *smg-1*, the third (*cc543*) has not been analyzed further.

We have used a variety of reporter constructs to test the feasibility of using *smg-1*(*cc546ts*) to generate conditional expression. Previous studies [17,23] had shown that (as with endogenous genes) unusual 3' UTR regions in transgenes could cause smg-dependent decay. We find that a *myo-3*::*gfp* fusion transgene with an artificially long 3' UTR (consisting of *lacZ* sequences) showed temperature-dependent activity in a *smg-1*(*cc546*) background. We would like to extend the technical usefulness of this by finding a "junk" 3' UTR which would lead to smg-dependent degradation and would not interfere with other aspects of experimental design (in particular the sequence should be lacking any reporter sequences or *cis*-acting transcriptional signals). The best candidate for this at the moment makes use of coding sequences from *let-858*. We have included with the kit a *mec-7*::*gfp* fusion that is "conditionally handicapped" by the extension of 3' UTR sequences by sequences from *let-858*. This clone can easily be used as a source for restriction fragments that will transfer the "conditional handicap" sequence.

Using vectors with this "extended" 3' UTR, we have found that the conditional ts-Smg system does not yield an absolute on-off difference between temperatures; nonetheless, the system offers a considerable degree of inducibility. This corresponds (by eye) to a difference between "strong and obvious" (at 25°C) and just detectable (at 16°C or in a Smg+ background). Rough quantitation suggests that the Smg+/Smg- differences are on the order of 10-20 fold.



Two variants of this plasmid also provided are L3774 (pPD118.10): identical to the plasmid shown at the right with the gfp variant being [F64L S65T M153A], and **L3720 (pPD117.58)**: similar to L3774 (pPD118.10), but retaining the fifth and sixth introns of let-858. The fifth intron may increase expression, but also may contain (uncharacterized) control elements.

Strains containing ts *smg-1* alleles

PD8117 *smg-1*(*cc545ts*)*unc-54*(*r293*)-
Movement: very Unc at 16°C, Slightly Unc at 25°C

PD8118 *smg-1*(*cc546ts*)*unc-54*(*r293*)
Movement: very Unc at 16°C, Almost wild type at 25°C

PD8119 *smg-1*(*cc545ts*)

PD8120 *smg-1*(*cc546ts*)

These strains are being furnished (as starved stocks) with the initial mailing of the 1997 worm kit. Subsequent requests should be addressed to the *C. elegans* genetic stock center.

H. Some notes on the use of excess N2 genomic DNA as carrier to prevent transgene silencing.

Summary

This section describes a relatively quick and straightforward means by which you may be able to achieve both increased expression and improved uniformity of transgene expression. This procedure doesn't help with all transgenes, and you'll need to make several lines to have any hope of interpreting an expression pattern. Also don't forget (even with this or other modifications to the injection protocol) that TRANSGENE ACTIVITY PATTERNS CANNOT UNDER ANY CIRCUMSTANCES BE USED AS THE SOLE MEANS TO DETERMINE THE PHYSIOLOGICAL EXPRESSION PATTERN OF AN ENDOGENOUS GENE.

We recently reported that inclusion of complex carrier DNA in *C. elegans* transformation experiments can improve the function of certain co-injected reporter constructs [12]. Of the carrier DNAs we've tried so far, *C. elegans* genomic DNA has been the most effective {Actually C. briggsae DNA is comparable, but we couldn't think off hand of any significant advantages}. Yeast, fly, mammalian, bacterial, and random DNAs were significantly less effective as carrier DNA; we don't know why this is, but we are actively investigating both the mechanism of silencing and the means by which it can be avoided. For the interim, here is our protocol for preparation of worm DNA and corresponding injection mixes. Note that the genomic DNA needs to be very clean to avoid sterility of the injected animals; this preparation protocol may be overly compulsive, but usually works.

Three types of expression constructs have been particularly helped by the co-injection of *C. elegans* DNA as carrier.

1. We have examined the ability of certain broadly expressed *C. elegans* genes (or their reporter-tagged derivatives) to express from diverse contexts (see Kelly et al., [12]). In standard "simple" arrays (e.g., those made by co-injection of a reporter plasmid with a selectable marker such as rol-6), we see initial (low level) germline expression followed by a generational "silencing" of the transgene so that expression is completely somatic within the first few generations. Somatic expression of these particular transgenes is partly or completely resistant to this generational silencing.

Inclusion of an excess population of carrier DNA consisting of random *C. elegans* DNA fragments greatly improves the level and maintenance of germline expression for these transgenes. Assuming copy number to be roughly proportional to relative composition of the injection mix, we also infer a modest increase in "per-copy" somatic expression of transgenes in an "*C. elegans*-DNA" environment (compared with lines produced by co-injection of the reporter construct at a similar concentration with a simple carrier DNA instead of *C. elegans* genomic DNA carrier).

2. Many constructs which use a defined enhancer (e.g., *unc-54*, *hlh-1*) to drive reporter expression by a heterologous promoter (e.g., *myo-2*, *pes-10*) show strong enhancement in the first generation following microinjection, but do not show this pattern (or show it only sporadically) in subsequent generations of heritable lines established by selection for a co-injected marker. Expression in these lines can frequently be dramatically improved by the carrier co-injection procedure described below.

3. Certain promoter constructs (e.g., *unc-54*) are active in a random subset of cells in each transgenic animal. This mosaicism seems dependent not on the inheritance of the extrachromosomal array (indeed the mosaicism is evident even with integrated arrays), but rather seems due to a stochastic activation of gene expression in a subset of cells. We have found that the injection of complex DNA as carrier (and particularly *C. elegans* DNA) can greatly increase the frequency of expression for some promoters (e.g. *unc-54*). Other promoters with similar expression patters (e.g. *myo-3*) appear to be less affected (or even resistant) to the mosaic silencing, so that not all constructs are improved by the co-injection of *Caenorhabditis* DNA.

We do not know the mechanism of the stimulatory effects of complex carrier DNA on the expression of certain constructs. Some points are worth keeping in mind, however. First, almost every line produced with N2 DNA as carrier shows some relief from silencing. Since each transgenic array is likely to contain only a small portion of the worm genome (1% by extrapolation of some previous estimates for array size [10,16]), it seems unlikely that a specific single-copy sequence is responsible for the improved expression. Rather we think that two processes are responsible. First is the increase in complexity afforded by dilution with complex carrier DNA. In model experiments with artificially produced multimers of a test sequence, we find evidence for a repeat-dependent silencing mechanism for sequences repeated several times within a single plasmid. Second, there may be some specific sequence element or sequence feature that is frequently found in *Caenorhabditis* DNA and which can serve as an anti-silencing element for nearby sequences. We are in the process of trying to identify such elements using the assays described above.

Procedures for transformation by co-injection with nematode carrier DNA

Overview: Your test construct (e.g. reporter fusion) and a selectable marker (e.g. *rol-6* plasmid) are linearized and diluted with a vast excess of random restriction fragments of *C. elegans* DNA. This mixture is injected using standard procedures. Although frequency of F1 animals expressing the selectable marker may be somewhat lower, the number of transgenic lines derived from these injections is comparable to that derived from standard injections.

Worm Genomic DNA Mini-prep for Microinjection

1. Add 450ul of worm lysis buffer to a frozen 50ul aliquot of worms.
2. Add 20ul of 20mg/ml proteinase K to worms and vortex.
3. Incubate at 62°C for 60 minutes. Vortex 4-5 times during the incubation. The solution should clear as the worms disintegrate.
4. Add 80ul of 5M NaCl. Mix thoroughly by inversion (important).
5. Add 80ul CTAB/NaCl solution. Incubate 10 minutes at 37°C.
6. Add 700ul chloroform, mix and briefly spin. Recover the aqueous phase.
7. Add 700ul phenol/chloroform(1:1), mix, briefly spin, recover aqueous phase.
8. Add 0.6 volume (about 400ul) of -20C isopropanol. Invert to mix. The stringy white DNA should be obvious. Spin in a microfuge for 5 minutes.
9. Decant and discard the supernatant. Wash the DNA **twice** with 70% ethanol at room temperature
10. Dry and resuspend DNA in 300ul TE pH 7.4.
11. Add 5ul RNase A (Standard RNase, pre-heat-treated to kill DNase); Incubate 42°C for 2 hr
12. Add PvuII (~200 units) and 36ul [10x PvuII buffer] to make volume 360ul; Incubate 37°C for 3 hr
13. Add 20ul 20%SDS + 10ul 0.5M EDTA pH7.5 + 20ul Protease K; Incubate 65°C for 2 hr. Add 40ul 10M Ammonium Acetate
14. Extract Twice with Phenol/Chloroform, Once With Chloroform, Add 1ml of Ethanol, Mix and spin 10min. Wash pellet with Ethanol and resuspend in 50ul.
15. Check on gel for indication of concentration and digestion completeness.

This material is generally not clean enough for injection into worms. We have obtained more reproducible results by further purifying over commercial miniprep columns-- one hypothesis is that these columns help by removing small RNA fragments left over from the RNase A digestion.

Column Cleanup of PvuII cut DNA

Manufacturer's instructions can essentially be followed here. We use the Promega Wizard prep system for this. The only unique part is to get the purified DNA in a mix equivalent to the initial cleared bacterial lysate: For this we pre-make and clear a "virtual lysate mixture".

"Virtual Lysate Mixture":
150ul GTE (50mM glucose, 25mM Tris-Cl pH8, 10mM EDTA)
150ul1M NaOH/0.1% SDS
150ul 5M KOAc (3MKAc, 2MHAc)
the SDS and KOAc will form a precipitate that can be removed by spinning 5min in a microcentrifuge at top speed.

1. Add 400ul of the cleared Virtual Lysate Mixture to approximately 20ug of cut DNA
2. Add 1 ml of suspended Promega Wizard prep resin (in manufacturer's resin suspension solution) to tube and mix by inverting several times.
3. Load mix into 3 ml syringe fitted with spin-column; follow manufacturer's instructions for washing and recovery of DNA from column.
4. Add Ammonium Acetate to 1M. Phenol/$CHCl_3$-extract once, $CHCl_3$ extract once, EtOH precipitate, EtOH wash and resuspend pellet in 20ul TE.

Injection Mixes: We have been using injection mixes with 50-100 ug/ml of PvuII cut N2 genomic DNA, with blunt-cut digested marker (*rol-6*) and test (e.g. reporter) plasmids at 0.5-2.0 ug/ml each. For some constructs, a somewhat increased test plasmid concentration can be used without significant silencing effects.

Appendix
1) Worm Lysis Buffer
0.1M Tris-Cl pH 8.5
0.1M NaCl
50mM EDTA
1% SDS
2) TE 7.4: 10mM Tris-Cl pH 7.4; 1 mM EDTA
3) Protease K:20mg/ml stock in TE 7.4
[Protease K from Boeringer Mannheim Biochemicals; store -20C in small aliquots]
4) CTAB/NaCl Solution : 10% CTAB in 0.7M NaCl
Dissolve 4.1g NaCl in 80mls distilled water and slowly add 10g CTAB (Mixes Alkyltrimethyl Ammonium Bromide; Sigma #M-7635) while heating and stirring. If necessary, heat to 65°C to dissolve. Adjust final volume to 100ml.

For current version of this protocol and updated information on this and other protocols, check the "protocols" folder on our **www** server ("www.ciwemb.edu")

Thanks to: Scott Emmons (Original Miniprep); S. L'Hernault (CTAB step).

I. Other potentially useful vectors (***** constructs are described in additional detail in the text)

modular *gfp* reporter constructs with S65C mutation and let-858 3' UTR and flank

pPD117.01 **L3691** *mec-7* promoter driving modular *gfp* [S65C]**[see section E above] ***
pPD118.17 **L3784** *mec-3* promoter driving modular *gfp* [S65C]
pPD118.20 **L3785** *myo-3* promoter driving modular *gfp* [S65C]
pPD118.25 **L3786** *let-858* enhancer/promoter driving modular *gfp* [S65C]**[see section F above] ***
pPD118.26 **L3787** *hsp16/2* promoter driving modular *gfp* [S65C]
pPD118.28 **L3788** *hsp16/41* promoter driving modular *gfp* [S65C]
pPD118.33 **L3790** *myo-2* promoter driving modular *gfp* [S65C]
pPD118.15 **L3781** no promoter (MCS) driving modular *gfp* [S65C]

modular *gfp* reporter constructs with Smg sensitive 3' UTR [see section G above] [all are S65C]

pPD118.44 **L3691** *mec-7* promoter *** [see section G above]**
pPD118.60 **L3808** *myo-3* promoter
pPD118.64 **L3809** *let-858* enhancer/promoter
pPD118.71 **L3811** *myo-2* promoter
pPD118.74 **L3812** no promoter (MCS)

enhancer assay constructs

pPD97.78 **L2603** Minimal *pes-10* promoter fused to 5-intron *gfp*[S65C] reporter with *unc-54* 3' end. This has little or no activity alone, but can be used as an enhancer assay vector.
pPD97.82 **L2604** Minimal *myo-2* promoter fused to 5-intron *gfp*[S65C] reporter with *unc-54* 3' end. This has little or no activity alone, but can be used as an enhancer assay vector.
pPD97.84 **L2605** Minimal *glp-1* promoter fused to 5-intron *gfp*[S65C] reporter with *unc-54* 3' end. This has some activity alone, but can be used as an enhancer assay vector.

pPD107.94 **L3135** Minimal *pes-10* promoter fused to 15-intron NLS-*gfp*[S65C]-lacZ reporter with *unc-54* 3' end. This has little or no activity alone, but can be used as an enhancer assay vector.
pPD107.97 **L3136** Minimal *myo-2* promoter fused to 15-intron NLS-*gfp*[S65C]-lacZ reporter with *unc-54* 3' end. This has little or no activity alone, but can be used as an enhancer assay vector.
pPD107.101 **L3137** Minimal *glp-1* promoter fused to 15-intron NLS-*gfp*[S65C]-lacZ reporter with *unc-54* 3' end. This has some activity alone, but can be used as an enhancer assay vector.

heat shock reporter fusions

pPD99.39 **L2678** *hsp16/2*::*lacZ*::*unc-54* 3' with 12 artificial introns
pPD99.47 **L2682** *hsp16/41*::*lacZ*::*unc-54* 3' with 12 artificial introns
pPD99.44 **L2680** *hsp16/2*::*gfp*[S65C]::*unc-54* 3' with 6 artificial introns
pPD99.52 **L2684** *hsp16/41*::*gfp*[S65C]::*unc-54* 3' with 6 artificial introns
pPD99.16 **L2681** *hsp16/2*::*gfp*[S65C]-*lacZ*::*unc-54* 3' with 15 artificial introns
pPD99.55 **L2685** *hsp16/41*::*gfp*[S65C]-*lacZ*::*unc-54* 3' with 15 artificial introns
pPD118.26 **L3787** *hsp16/2* promoter driving modular *gfp* [S65C] with 3 artificial introns
pPD118.28 **L3788** *hsp16/41* promoter driving modular *gfp* [S65C] with 3 artificial introns

other constructs

pPD98.41 **L2630** Useful as a positive control for silencing effects. *unc-54* enhancer driving myo-2::*gfp* [S65C] fusion. Expression in pharyngeal muscle occurs whether silenced or not. Body wall muscle expression is seen only if the construct has escaped silencing.

pPD104.91 **L2963** *unc-54*::*gfp* fusion with an additional synthetic intron interrupting the coding region of *gfp*. Denoted **Π**, this synthetic intron has an internal multiple cloning site for testing of sequences for their effects in an intervening sequence context.

pPD116.81 **L3670** Mito-gfp [S65C] driven by *let-858* promoter/enhancer (labels somatic mitochondria)

Appendix B. Explanation of a few of the more obscure terms in describing these vectors.

What follows below are overall descriptions of the vectors being distributed with this kit. The convenience of the microtiter well distribution format allows us to distribute a large number of vectors that might have applications in several different types of studies. It is anticipated that only a fraction of the vectors will be used by any given lab. These descriptions should allow the experimenter to see which vectors are most appropriate for the task at hand. In order to actually carry out the constructions, it is anticipated that the experimenter will retrieve the complete sequence of the vector from the **web** archive and use this to choose specific restriction sites and cloning strategies. All of the predicted structures are based on knowledge of the individual elements and a reconstruction of the cloning steps used to produce the vectors. At each step, we've done considerable checking with restriction enzymes to confirm structures. We can't rule out unexpected sequence changes, although most of the vectors below are at most one construct removed from a vector that has been tested in vivo for efficacy.

Plasmid #: This gives the unique number of the DNA preparation that is being distributed.

Ligation #: Ligation numbers uniquely indicate the structure of the construct described. These allow you to reference the proper sequence in the electronic archive.

Promoter: The new reporter segments have been tested with a variety of previously characterized promoter segments. The precise details of fragments used for promoter activity can be derived from the sequences.

IVS: Number and distribution of synthetic intervening sequences

𝔸 is the original synthetic IVS inserted upstream of the reporter coding region in the older *lacZ* vectors [4]

𝔹, ℂ , 𝔻 , 𝔼 , 𝔽 , 𝔾 , ℍ , 𝕀 , 𝕁 , 𝕂 are unique synthetic introns inserted into the coding region of *lacZ*

å , ß , ∂ are unique synthetic introns inserted into the *gfp* coding region

𝕀 is a synthetic intron with an internal multiple cloning region inserted into *gfp*

𝕊 is a synthetic intron inserted just upstream of the *unc-54* 3' UTR sequence

𝕃 is a synthetic intron inserted into the 3' UTR of *unc-54*

3' end: Vector 3' end sequences from *unc-54* and *let-858* are used.

decoy: decoy **+** vectors have a (synthetic intron--->short coding region) minigene upstream of the MCS, to decrease background from readthrough transcription. "double" decoy vectors have two different decoys upstream.

References

1. Cali, B., and P. Anderson (1995) Characterization of a new smg gene, smg-7. *C. elegans* meeting 1995: 142; Cali, B., Kuchma, S. and P. Anderson, personal communication.
2. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W., and Prasher, D. (1994) Green Fluorescent protein as a marker for gene expression. Science 263, 802-805.
3. Cormack, B., Valdivia., R.H., and Falkow, S. 1996. FACS optimized mutants of the green fluorescent protein (*gfp*). Gene 173:33-38.
4. Fire, A, Harrison, S., and Dixon, D. (1990). A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. Gene 93, 189-198.
5. Fire, A., Ahnn, J., Kelly, W., Harfe, B., Kostas, S., Hsieh, J., Hsu, M., and Xu, S. (1997) GFP applications in C. elegans. in GFP: Green Fluorescent Protein Strategies and Applications, M. Chalfie and s. Kain eds, John Wiley and Sons press, NY. (in press)
6. Fire, A., Ahnn, J., Seydoux, G., & S. Xu (1995) FireLab 1995 Vector kit documentation. www.ciwemb.edu
7. Heim, R., and Tsien, R.Y. 1996. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. Current Biology 6:178-182.
8. Heim, R., Cubitt, A., and Tsien, R. (1995) Improved Green Fluorescence. Nature 373, 663-664.
9. Heim, R., Prasher, D. and Tsien, R. (1994) Wavelength mutations and posttranslational autooxidation of green fluorescent protein. Proc. Natl. Acad. Sci. 91, 12501-12504.
10. Hope, I.A. (1991). 'Promoter trapping' in *Caenorhabditis elegans*. Development. 113, 399-408.
11. Kelly, B., Xu., S., and Fire (1995) Reporter transgene expression in the germline- a start. Worm Breeder's Gazette 14(1): 64
12. Kelly, W.G., Xu, S., Montgomery, M., and Fire, A. (1997) Distinct Requirements for Somatic and Germline Expression of a Generally Expressed C. elegans Gene. Genetics, in press
13. MacVector [IBI incorporated] and Vector NTI [Informax Inc] are examples of commercially developed packages for interactive sequence manipulation. We have not used either of these.
14. Marck, C.: 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. 16 (1988) 1829-1836.
15. Mello, C., and Fire, A. "DNA transformation" (Methods in Cell Biology: *C. elegans*, D. Shakes and H. Epstein eds., Academic Press [publication 1995].)
16. Mello, C.C., Kramer, J.M., Stinchcomb, D. and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J. 10:3959-3970.

17. Okkema, P., White-Harrison, S., Plunger, V., Aryana, A., and Fire, A. (1993). Sequence requirements for myosin gene expression and regulation in *C. elegans*. Genetics *135,* 385-404.
18. Papp, A. (1991) Graded suppression of lin-29 by cold-sensitive smg mutations reveals differential reouirements for lin-29 activity in vulval morphogenesis, cuticle gene expression, and the larval to adult (ua) switch in hypodermal cell fate. *C. elegans* meeting 1991: 229.
19. Pulak, R., and Anderson, P. (1993) Messenger RNA surveillance by the *C. elegans smg* genes. Genes & Development 7, 1885-1897.
20. Ransom, D., Sharp, S., Alberico, C., and B. De Stasio (1995) Disruptive Myosin Heads. *C. elegans* meeting 1995:431.
21. Seydoux, G. and Fire, A. (1994). Soma-germline asymmetry in the distributions of embryonic RNAs in *Caenorhabditis elegans*. Development *120*, 2823-2834.
22. Stringham, E.G., Dixon, D.K., Jones, D., and Candido, E.P.M. (1992) Temporal and spatial expression patterns of the small heat-shock (HSP16) genes in transgenic *Caenorhabditis elegans*. Mol. Biol. Cell 3, 221-233.
23. Wilkinson, H., Fitzgerald, K., and Greenwald, I. (1994) Reciprocal changes in expression of the receptor *lin-12* and its ligand lag-2 prior to commitment in a *C. elegans* cell fate decision.Cell 79:1187-1198.
24. These mutations were isolated in the F2 following mutagenesis and growth at 25°C. Since alleles of some smg mutations show a maternal rescue effect [19], it is possible that the screen was biased toward genes, such as *smg-1*, that do not show maternal rescue.

Acknowledgments:

We would like to thank S. Kostas, J. Hsieh, M. Hsu, K. Liu, C. Macarah, A. Pinder, M. Krause, S. Dymecki, N. Marsh-Armstrong, M. Chalfie, R. Heim, S. L'Hernault, S. Emmons, R. Tsien, R. Pulak, K. Fitzgerald, B. Cali, P. Anderson, S. Kuchma, I. Greenwald, T. Dill, and colleagues at this institution for their help and suggestions. Work in this lab is supported by the National Institutes of Health (Grant R01-GM37706) and the Carnegie Institution. Joohong Ahnn was supported in part by a Genetic Engineering Grant (G60200) from the Ministry of Education, Korea.

Fire Lab 1997 Vector Supplement- Physical Instructions

This is a microtiter rack with vectors from the Fire Lab. Each well has one individual vector. Stocks were dried from 10μl [10mM Tris-Cl (pH8) 1mM EDTA 5% Glycerol 1μg/ml tRNA 0.1mg/ml BSA]. DNA stocks used were very dilute (~~1μg/ml). The vectors should be rehydrated by adding 40μl H20, then transformed into a suitable *E. coli* host to obtain working stocks. The bacteria we use are DH5-alpha. All plasmids provide amp resistance.

Do not peel off the sealing film on top of the microtiter plate. To obtain a given plasmid, puncture the corresponding well with a standard pipette tip prefilled with 40μl of H2O(wipe the cover gently with a wet, then a dry kimwipe before this to avoid contamination). Pipet up and down to resuspend DNA, then take 1-2μl to transform *E. coli*. Cover opened wells with small pieces of lab tape or adhesive sealing film. As each plasmid is grown up, use restriction digestion (e.g. DdeI) compared with known sequence to help confirm that no cross- contamination has occurred. **Store -20°C this side up**

For locations of vectors, see the map with the kit documentation.

Remember to check the web archive for corrections. Always wear your seatbelt. "transgene activity patterns cannot under any circumstances be used as the sole means to determine the physiological expression pattern of an endogenous gene." We are sending this kit out (one per institution) as of February 1997. You are expected to share this kit with others at your institution. If you are leaving the institution, you can "duplicate" the kit by pipetting small aliquots from each well into duplicate wells. We make no guarantee to continue to distribute the kit.

History, corrections and other info

February 1997: Version 1.0. This is the first posting of documentation for 1997 Kit. This is a preliminary version of the documentation. Users are advised to continue to check the archive for updates over the next few weeks.

March 1997: Version 1.01. This is a (prerelease) update that deals with two aspects of gfp sequence. Neither of these should affect any gfp constructions or ongoing experiments.
1. The sequence of one of the synthetic introns (ivsə) within the gfp coding region as reported in the original vector kit had several errors. These have been corrected, and updated sequences for both the older (1995) vectors and the 1997 vectors are being posted on line. No changes in restriction sites of in activity result from the different sequence.

2. In tracing back through the construction of the wild type and mutant gfp forms, it was noted that a mutation was present in the original *gfp* coding sequence used for the Chalfie vectors (pTU60-65). This mutation (resulting in the amino acid substitution Q80R). This substitution has no reported effect on gfp function, but should be taken into account for any detailed studies of gfp spectra for the different gfp forms. In some later constructs, the Q80R mutation is corrected back to the (presumably wild-type) gfp. Of the vectors in the present and previous kit. the Q80R substitution is present as follows:

a. All of our constructs with unmutated "wild-type" *gfp* have the [Q80R] substitution.

b. All constructs reported with "single-mutant" [S65T] or [S65C] *gfp* also carry the [Q80R] substitution.

c. All constructs reported with gfp variants [gf4]-[gf6], [gf9]-[gf24], and [gf30-38] have the Q80R substitution corrected to Q80Q (i.e., so that the 80'th amino acid is really gluatmine).

d. In out limited experience, we have found no signficant difference between Q80Q and Q80R constructs.

April 1997: Supplementary vector kits were sent out a few weeks ago. If you expected one and haven't recieved it, give us a holler. For those that requested 1995 kits... these should be on there way. We're currently testing a new batch of these and hope to have them on the way next week.

Several minor fixes and suggestions:

a. The vector originally referred to as pPD117.83 (L3716) was updated before the release, but the documentation was not updated. The new version of this vector is called pPD118.25 (L3786). This vector should still correspond to the map shown for L3716, and the predicted sequence for L3786 has been posted on the web site.

b. Just a reminder to check all plasmid preps from the kit as soon as they are made with multiple restriciton enzymes. We recommend PvuII and DdeI as such enzymes. If these (or any subsequent digests) don't look as expected then you should i) try a few other colonies transformed from that well, and if this doesnt settle things ... ii) give us a holler

c. If you have trouble getting colonies from any of the wells, it may be a problem in resuspending the DNA properly from the glycerol stock. In some cases, this might be improved by performing the resuspension in 40ul of 100mM NaCl, instead of using water.