Fire Lab Vector Kit- June 1995 (updated to August 1996)

Credits:
Making & Testing New Vectors: Si-Qun Xu
Vector Design, Strategic Mistakes, Late Night Gel Photos: Andy Fire
Analysis of mRNA localization, lacZ in early embryos, pes-10 promoter: Geraldine Seydoux
Deleted myo-2 promoter analysis: Pete Okkema
Testing, Suggestions, Analysis: Joohong Ahnn, Brian Harfe, Bill Kelly, Verena Plunger
Also thanks to: Sue Dymecki, Mary Montgomery,

The vector kit has three components

1. The microtiter vector array, containing vectors described in this handout and some older vectors from this lab that are described in previous papers.

2. This handout, with preliminary information on the new vectors

3. The address of the ciw vector archive (web address www.ciwemb.edu) which will contain up-to-date information on these (and older) vectors.

Abstract

This kit contains three classes of vectors that should be useful in studies of gene expression and function in *C. elegans*. Reporter vectors with *lacZ* and/or *gfp* can be useful in following the expression pattern of a defined promoter region, and for defining signals responsible for the observed expression. Ectopic expression vectors allow a coding region of interest to be expressed in a pattern that allows the function of the encoded protein to be assayed. Enhancer assay vectors allow individual transcriptional control elements to be studied. The new vectors provide greater sensitivity than those previously described. This has advantages for many experiments, but it is important to realize that background as well as signal is boosted. In particular, it is critical to recall that TRANSGENE ACTIVITY PATTERNS CANNOT UNDER ANY CIRCUMSTANCES BE USED AS THE SOLE MEANS TO DETERMINE THE PHYSIOLOGICAL EXPRESSION PATTERN OF AN ENDOGENOUS GENE.

I. Background: Limitations of the older vectors

C. elegans provides a facile system to determine mechanisms responsible for developmentally regulated gene expression and pattern formation. Visibility of all cells throughout the life cycle has allowed complete determination of the cell lineage and a detailed description of morphogenic changes during development. Ongoing genetic analysis in a large number of laboratories, combined with a concerted genome sequencing effort, have been providing a great variety of molecular handles, some with defined functional roles and others with roles yet to be defined. Rapid methods for introduction of cloned DNA into the *C. elegans* germline provide means to analyze molecular function and expression in vivo. Gene fusion vectors can be a useful tool for these experiments, allowing the expression of reintroduced genes to be monitored, and giving an indication of the distribution of the resulting products.

The most frequently used reporter molecules for *C. elegans* have been the *E. coli lacZ* coding region (producing the enzyme ß-galactosidase, which can readily be located using a histochemical stain) and the *Aequora victoria gfp* coding region (producing the intrinsically fluorescent protein GFP). Previously described fusion vectors from this lab and elsewhere have proven useful for many studies, but the observed expression patterns have exhibited several unexpected properties. Expression has been examined for a large number of *lacZ* fusion constructs. In many cases the expression pattern of the endogenous gene was never determined, so it is not possible to correlate the transgene expression pattern with physiological expectation. Where both physiological and transgene expression patterns have been determined, there are some cases (examples) with a good correlation between the timing of distribution of *lacZ* expression and expression of the endogenous gene. In other cases, this correlation is incomplete, with the transgene:*lacZ* construct exhibiting ectopic expression or lacking aspects of the physiological expression pattern. Ectopic expression is frequently seen with short promoter segments, and occurs most prominently in the gut and pharynx. This may reflect weak promoter and/or enhancer signals in the original vectors which can specifically stimulate expression if placed in proximity to inserted regulatory elements (e.g., [9]). Deficits in the expression patterns of *lacZ* fusions can be classed in three groups. Often, a given transgene will express in the correct tissue, but exhibit mosaic expression with only a subset of cells staining. This has been observed even with integrated high-copy transgenes, and even in cases for which the endogenous gene is known to express uniformly in the tissue. In a few cases, expression in a single tissue or set of cells (which expresses the endogenous gene) is not seen with the transgene. Finally, an extremely frustrating problem has been the failure of any transgene fusion construct to show expression in embryonic, larval or adult germ line, or in the pre-12-cell embryo.

The *gfp* fusion vectors have been available for a shorter period of time. Although some successes have been reported in reproducing the expression patterns of endogenous genes, we have seen a number of cases in which the GFP fluorescence pattern is more restricted than that exhibited by equivalent *lacZ* fusions and by the endogenous gene. A possible explanation for this was suggested by studies of Heim et al. [6], and Inoye and Tsugi [19], who suggested that a complex and relatively slow chemical reaction is needed to convert the newly formed GFP peptide chain into a fluorescent protein. Their studies in *E. coli* have succeeded in identifying a number of mutations which result in more rapidly acquired fluorescent signal.

In this handout, we describe several approaches toward producing *lacZ* and *gfp* reporter vectors which express more efficiently and can more faithfully reproduce the expression pattern for the endogenous gene. This has yielded vectors with greatly increased sensitivity and somewhat decreased levels of mosaicism. These vectors do not solve the problem of germline expression, although somewhat more appropriate expression of genes has been seen in the early embryo They likewise have not solved (and may in some cases accentuate) the problem of ectopic expression. The vectors should be useful in many applications, but TRANSGENE ACTIVITY PATTERNS CANNOT UNDER ANY CIRCUMSTANCES BE USED AS THE SOLE MEANS TO DETERMINE THE PHYSIOLOGICAL EXPRESSION PATTERN OF AN ENDOGENOUS GENE.

In addition to the reporter vectors described above, two other sets of vectors have found widespread use for studies of *C. elegans*. The first set consists of a group of ectopic expression vectors. Each of these contains an active promoter region with a characterized expression pattern, followed by signals to allow processing and expression of the resultant transcripts. Using these vectors, an arbitrary coding region can in principle be expressed in any of a variety of defined patterns. These vectors have been used to express (or mis-express) a wide variety of homologous and heterologous proteins, and in antisense RNA and ribozyme experiments.

In studies of gene expression, a different type of vector (called an enhancer-assay vector) has been useful. These vectors contain a "basal" promoter fused to a reporter gene (e.g. *lacZ*). Generally the basal promoter is an element with only limited expression. These vectors are used as follows in studying enhancer function: a candidate enhancer sequence is placed upstream of the basal promoter. The resulting construct is then assayed for expression pattern to determine the effect of this element on expression. The interpretation of results with such vectors depends on several parameters, with one key concern being the ability of the "basal" promoter to respond in diverse locations and to diverse enhancer types.

II. New *lacZ* vectors

Insertion of multiple introns into the *lacZ* coding region produces a more sensitive reporter.

In our initial analysis of myosin gene expression in *C. elegans*, we found that an intervening sequence near the 5' end of a primary transcript can greatly improve the yield of protein product. More recent studies [Fire, Seydoux, and Xu ...] have revealed that insertion of additional intron sequences results in additional increases in gene expression. These studies are summarized below.

Quantitative comparisons using the *unc-54* promoter driving coding regions punctuated with 1-12 intron sequences have revealed a stimulation of expression by 2-3 orders of magnitude by the additional introns. Several types of data have suggested that the need for multiple introns throughout a transcript for optimal activity may be a general property of *C. elegans* gene expression:

1. A variety of promoter::*lacZ* constructs are stimulated by insertion of introns into the *lacZ* coding region. With weak and moderately strong promoters (*skn-1, glp-1, cey-2, hlh-1*, and minimal segments from *unc-54, ceh-24, myo-3, and myo-2*) the additional introns cause a dramatic increase in both the frequency and intensity of somatic expression. Intron-stimulation was somewhat less dramatic starting with already-strong promoters; these may already be saturating enzyme assays or ability of cells to make product.
2. Intron-based-stimulation has been seen in a wide variety of somatic tissue types. The stimulated promoters listed above include expression in gut, muscle (all classes), hypodermal and pharyngeal epithelial cells, and neurons.
3. Stimulation by intron sequences is not specific to the *lacZ* coding region. In particular, *gfp* constructs with multiple added introns show a stimulation of expression analogous to that seen with *lacZ*.
4. Stimulation by introns is seen with a variety of 3' UTRs. (Constructs with multiple intron sequences still require a 3' UTR for expression, but this 3' UTR need not be that from *unc-54*: stimulation has also been seen with *skn-1* and *hlh-1* 3' UTR regions.)
5. Stimulation by introns occurs both in Smg+ and Smg- genetic backgrounds (smg mutants increase the stability of many aberrant mRNAs; see Pulak et al. [15] for a review).
6. Stimulation by introns can be seen in transiently transformed animals (F1 animals from injected parents) and in inherited transgenics (array-containing transformed lines).
7. Stimulation by intron sequences does not require that the intron be placed in the coding region. We had previously observed stimulation by an intron in the 5' UTR of myosin gene::*lacZ* fusions. An intron in the 3' UTR of this construct produces substantial additional accumulation [fig].

In situ localization of transgene RNA products has shed some light on the mechanism of stimulated expression (these experiments have been carried out with *unc-54*::*lacZ* constructs with 0,1,8 and 12 intron sequences). Transcripts from single intron (or intron-lacking) constructs accumulate in the nucleus, with little or no RNA detected in the cytoplasm by this assay. In homozygous integrated transgenic lines, the nuclear RNA is frequently seen in two spots, which presumably correspond to the chromosomal sites of transcription. Equivalent intron-rich constructs show a very different pattern of intracellular RNA accumulation, with copious amounts of *lacZ* transcript seen in the cytoplasm. An intriguing possibility is that this difference in intracellular distribution might reflect an ability of introns to increase the efficiency of transport. Alternatively, this difference could reflect an increase in processing efficiency or transcript stability.

In designing vectors for high expression level, it was important to consider the question of how the spacing of introns affects the level of stimulation. In one model, the level of stimulation would depend on absolute number of introns, with little or no dependence on distribution within the transcript. Alternatively, the critical feature might be the need to break up long unspliced primary transcripts into smaller exonic region. In the latter case, the optimal distribution of introns would be spaced evenly within the transcript. Our limited experience to date is most consistent with the latter hypothesis. In particular, it appears that additional introns within *lacZ* can stimulate even fusion constructs that already contain introns from the fused gene.

Use of multiple-intron *lacZ* vectors:

We have constructed a set of promoterless vectors which are equivalent to our original *lacZ* vectors with the inclusion of varying numbers of intervening sequences. These have a convenient set of unique restriction sites and can readily be used

1. For construction of new *lacZ* fusions or
2. For modification of existing *lacZ* (or *gfp*) fusion constructs. (A intron-rich reporter coding region can be excised from the new vectors and used to replace the reporter in fusion constructs which contain older *lacZ* or *gfp* reporters.)

As noted above, we have seen substantial stimulation of expression in somatic tissues with a variety of weak and moderately strong promoters. The intron-rich vectors increase both the fraction of cells staining and the intensity of staining (these may both reflect a simple increase in expression). The mosaicism in expression which has been a problem with older (zero or one intron *lacZ* constructs) is partially relieved. The mosaicism is not

eliminated however: Even with the most active (12 intron) vector, we still see considerable mosaicism with certain promoter fusions, particularly in high-copy-number extrachromosomal arrays selected following co-injection with *rol-6*d (pRF4) [11].

With our original vectors, we had difficulty in obtaining robust expression in early embryonic lineages (prior to 28 cell stage). The intron-containing vectors improve expression in the early embryo: using the *pes-10* promoter we observe both earlier and more reproducible expression in the embryo: the onset of expression for the endogenous *pes-10* gene is approximately 12 cells, with transient expression in each somatic lineages following divergence from the germline. *pes-10*::*lacZ* fusions with a single intron show only expression in the later somatic lineages (C and to a lesser extent D lineages), while equivalent multi-intron *lacZ* fusions express in a pattern much closer to that of the endogenous protein.

A less tractable problem has been that of obtaining expression of *lacZ* fusion constructs (or indeed expression of any reporter construct) in the germline. Even with the most active reporter regions (*lacZ* with 12 introns) driven by upstream, internal and downstream sequences from a germ-line expressed gene (e.g. *skn-1*, *glp-1*), we have seen no expression in the germline. We are actively trying several approaches to remedy this problem.

Can a vector be too sensitive?

Our initial promoterless *lacZ* vectors (with zero or one intron) had the satisfying property that injection directly into worms (with no promoter inserted) did not result in any observed ß-galactosidase expression. Because we had increased the sensitivity of the vectors by insertion of introns, we revisited this question. The intron-rich vectors show varying levels of background expression, ranging from faint and rare expression (for the vector with two introns) to relatively strong and reproducible expression (vectors with 12 introns). The observed expression occurred predominantly in gut and pharyngeal tissue. Previous experiments [9] suggested a possible source for this expression: from these experiments we had postulated the presence of weak enhancer and/or promoter elements in the plasmid backbone sequence.

Under certain circumstances, readthrough transcription into the *lacZ* coding region could account for the background of staining with no promoter. The fate of putative readthrough transcripts will in this case determine the nature and level of background expression. If the primary transcripts efficiently yield a cytoplasmic mRNA with *lacZ* as the first open reading frame, then a background of ß-galactosidase activity will be seen. The sequences upstream of *lacZ* will thus influence the observed background, with some potentially unpredictable features to this dependence. First, the presence of out-of-frame ATG's or short ORFs upstream of *lacZ* could inhibit expression of ß-galactosidase protein by a readthrough message. This would be less significant if a trans-splice acceptor just upstream of the *lacZ* allows 5' ORFs to be removed from the mRNA molecule (generating instead a transcript with *lacZ* as the first ORF). The splice acceptor in the first synthetic intervening sequence might conceivably act in this way. Any trans-spliced gene fused to *lacZ* would certainly have the potential to process any readthrough transcript to generate trans-spliced material that would have a first open reading frame that could read into *lacZ*. In a promoter-containing construct, the background level will also depend on the presence of sequence elements in the promoter region that can block readthrough transcription or decrease the translational efficiency of the resulting mRNAs. The signals for transcriptional termination and efficient mRNA translation are still unexplored, precluding precise predictions of these effects based on sequence. One reasonable expectation might be that longer the upstream sequence, the more likely the attenuation of readthrough activity. Consistent with this, we have observed with a variety of promoters (e.g. [9], this work), that deletion of sequences upstream of a promoter frequently leads to ectopic expression in a pattern that can depend on flanking vector sequences.

A second contribution to background expression could come from enhancer sequences in flanking vector DNA. The level might depend partly on the presence of a bone fide promoter upstream of *lacZ*, although cryptic promoter sites in plasmid sequences could also yield expression. As with transcriptional readthrough discussed above, the activity of a vector-resident enhancer on a promoter just upstream of *lacZ* would be expected to depend on the nature of inserted sequences, with longer upstream regions more likely to contain (as yet uncharacterized) sequences that might block enhancer activity.

We considered several means to minimize background activities. Experiments with several different flanking plasmid backbones have shown differing background expression patterns (see ref [9]), but we were not able to identify any existing background-free vector. A second approach (commonly used in vertebrate systems) might be to cut the plasmid backbone sequences away from insert sequences before transformation. We have no experience with this, but the high efficiency with which incoming molecules are ligated to each other following transformation [17,11,5] might yield a variety of products with unexpected properties; in addition, free ends of unligated molecules might have distinct effects on nearby transcription units.

5' Decoy vectors

As an alternative approach to limit the effects of readthrough transcription, we have chosen to place a "decoy minigene" at the junction between the plasmid backbone and the restriction sites just upstream of the *lacZ* coding region. Several different decoy segments have been used successfully for this purpose. Each decoy consists of a short intron (splice donor + splice acceptor) followed by a short open reading frame with a consensus

C. elegans translational start. The open reading frame terminates just before the multiple cloning site for promoter insertion. The background with promoterless vectors is essentially eliminated by the decoy sequences tested. This is useful, but could in some senses be misleading, since the decoy would not be expected to block either readthrough transcription or activation by an enhancer. Although the decoy-containing vectors produce a clean background with no promoter, insertion of short or deleted promoter fragments might be expected to yield a degree of vector-specific background activity.

III. New *gfp* vectors

Making a more sensitive *gfp* reporter: I. intron insertion

In our initial experiments with *gfp* fusions in *C. elegans*, we and others had some success in obtaining expression, but in several cases, *gfp* fusions were much more difficult to detect and follow than their *lacZ* equivalents. By analogy to *lacZ*, we wished to test whether we could increase GFP expression by insertion of introns into the coding region. This is indeed the case, as indicated by qualitative assessments of fluorescence in animals transformed with *unc-54*::*gfp* fusions with the intron-containing reporter constructs. We have primarily used a *gfp* coding region with three internal introns. Vectors with a total of five or six intervening sequences were obtained by making use of intervening sequences that had been placed in the 5' and 3' UTR sequences in equivalent *lacZ* vectors.

As noted in Table 3, these vectors produce a strong stimulation in the level of *gfp* produced from an *unc-54*:*gfp* transcriptional fusion construct.

Making a more sensitive *gfp* reporter II. improved fluorescence

In our initial experiments with the wild type GFP coding region in the vectors provided by Chalfie et al. [1], several circumstantial observations suggested that GFP protein could accumulate without acquiring fluorescence [2]. This was a particularly a problem in the early and midstage embryo (pre-differentiation).

The variability in acquisition of fluorescence appears to be a common problem with *gfp* as a reporter in several different biological systems. Heim, Cubitt and Chen [6,7 tackled this problem elegantly by first characterizing a time dependent process required for formation of the fluorochrome, and second selecting mutant forms of the protein in E.. coli that could perform this process much more quickly. They obtained two mutations, S65T and S65C (nomenclature indicates mutation of serine-65 to threonine and cysteine respectively), that modify the site of fluorochrome formation. These provide several advantages for use as a reporter: a stronger fluorescent signal, faster folding of the fluorochrome and improved resistance to fading.

We have incorporated these mutations into intron-containing *gfp* vectors. As indicated by Heim et al., these provide an improved reporter for already-active *gfp* constructs (e.g. the *unc-54*::*gfp* fusions). In addition, these allow production of fluorescent protein in the early embryo: *gfp* fusions to the *pes-10* and *hlh-1* promoters (which show no fluorescence with the wild type *gfp* in early embryos), give strong expression with the S65T or S65C mutations.

Targeted *gfp* vectors.

For some applications of *gfp* reporters, labeling of a specific molecule in its normal context is required, while for other applications the critical aim is to label the cells expressing the fusion. We thought that it might be possible to improve the fluorescent signal from GFP in cell labeling experiments by targeting the produced protein to different intracellular compartments, potentially allowing a more amenable environment for fluorescence. In lines with native *gfp* expressed (no other protein sequences attached), the fluorescent signal is uniformly distributed through the cytoplasm, leaking also into the nucleus (and in some cases actually somewhat more concentrated there). We've examined the effects of several different targeting signals on GFP localization and embryonic fluorescence:

Nuclear localization signal:: Attachment of the SV40 nuclear localization signal (NLS) from our older *lacZ* vectors [3] results in fluorescence activity which is distinctly stronger in the nucleus than the cytoplasm. Curiously, the degree of nuclear localization is greatly improved in fusions which contain *gfp*, the nuclear localization signal, and an appended mass of protein (ß-galactosidase and pieces of myosin and HLH-1 are variably sufficient for this). The incomplete nuclear localization of the simple NLS-*gfp* protein may be due to the ability of this relatively small protein to diffuse out of the nucleus.

Secretion signal: We have previously used a synthetic secretion signal to drive secretion of several different expressed molecules [e.g., Perry et al., 14]. Secreted *gfp* constructs driven by the *myo-2*, *myo-3, pes-10* and *mec-7* promoters have been tested. The resulting fluorescence accumulation is both intracellular and extracellular. The intracellular staining with the *myo-2* promoter appeared reticular and might represent the internal secretion apparatus of the cell. Interestingly, coelomocytes appeared to scavenge secreted *gfp* expressed postembryonically

from *pes-10, myo-3,* or *mec-7*, storing the fluorescent material in internal vacuoles or droplets. We plan to use this as a screening tool to find mutants that affect the distribution and function of coelomocyte cells during development. Preliminary results with *gfp* mutant S65T appeared to give a stronger fluorescent signal. S65C was somewhat less active, suggesting the possibility that the novel cysteine might be subject to oxidation or other modification reactions in the extracellular environment.

Mitochondrial matrix localization signal: An N-terminal mitochondrial matrix localization signal from chicken mitochondrial aspartate aminotransferase [18] was synthesized and incorporated into the *gfp* expression cassette. This signal was sufficient to localize *gfp* to mitochondrial structures in body wall muscle, in hypodermis, and in embryos. At high expression levels, residual staining was observed in the cytoplasm. Although quantitative comparisons have not been performed, it appears that the mitochondrial *gfp* constructs may be somewhat brighter than their nuclear or cytoplasmic equivalents.

***gfp*-tagging of constructs**:

For many experiments, it will be useful to produce chimeric genes which would retain function of the original gene product while adding the fluorescent properties of GFP. Construction generally starts with a cloned copy of the gene of interest which has been shown to be functional by mutant rescue or other in vivo assay (e.g. a genomic clone). The *gfp* coding region is then inserted at either terminus of the coding region or into internal restriction sites. The hope is to find a position within the coding sequence at which insertion of *gfp* does not interfere with in vivo functions. The *gfp* insert for tagging these constructs will be a restriction fragment with compatible ends for insertion into the target site. In many cases, the *gfp* fragment is most easily generated by high-fidelity (e.g. *PfuI*) PCR starting with any of the above vectors: the freedom in designing PCR primers allows complete flexibility in choice of restriction sites and reading frame.

As an alternative, we have produced two modified *gfp* vectors which have a *gfp* coding region which is open at both ends, flanked by a variety of restriction sites (pPD79.44, has the wild type *gfp*, and pPD95.02 carries three internal introns and S65C). These allow in-frame insertion using any of the sites flanking the *gfp* segment. Ten restriction sites are available on the 5' side of *gfp* for in frame fusion; 31 sites are available on the 3' side.

We have used these as sources for *gfp* in producing a *gfp-lacZ* fusion constructs. The resulting constructs exhibit both *gfp* and *lacZ* activities. Since this could be useful in making fusion constructs for promoter analysis, we have incorporated the *gfp-lacZ* reporter into a set of promoterless vectors.

IV. Ectopic expression vectors

We have previously distributed a set of vectors which carried promoter elements, RNA processing signals, but no coding regions. These were designed for insertion of chosen coding regions in experiments to examine the effects of ectopic expression. Some re-evaluation of the design and use of these vectors is appropriate given the observations discussed above suggesting that long, unspliced regions in primary transcripts tend to decrease expression. In particular, it is likely that coding regions punctuated by native introns will be more efficiently expressed than long cDNA constructs. In some experiments, a certain degree of inefficiency in expression can be tolerated; in other cases, it will be important to optimize. For expression of *C. elegans* genes, a simple solution is to insert a genomic copy of the coding region to be expressed instead of a cDNA copy. As an alternative, a cDNA coding region from *C. elegans* or another species could be punctuated by introns in a manner analogous to that used for *lacZ* or *gfp* (see materials and methods).

In comparing activities of construct with different intron number, we found a substantial stimulation in expression in constructs with just the first and last intron (in 5' UTR and 3' UTR regions respectively). Since the 5' and 3' UTR regions are included in the ectopic expression vectors, we could redesign these to contain additional noncoding introns. We have made the corresponding vectors for the *mec-7*, *myo-3*, and *myo-2* promoters. These have not yet been tested, but are being made available on an as-is basis as part of this kit. If the extra-intron expression vectors indeed boost expression of intronless cDNAs, then it will be useful to produce additional equivalent vectors with the other promoters.

Two vectors that have been extensively utilized for ectopic expression utilize the promoters for the small heat shock genes *hsp16-2* (pPD49.78) and *hsp16-41* (pPD49.83). We do not know whether introns will stimulate gene expression under heat shock conditions: in some other organisms, heat shock has been reported to inhibit splicing. *C. elegans* heat shock genes are spliced, indicating that splicing cannot be completely inhibited. Nonetheless, the ability of multiple-intron constructs to be spliced under heat shock conditions has not been assessed.

V. Enhancer assay constructs

A deleted promoter segment fused to *lacZ* provides a vector for use in assaying for enhancer activities. We have extensively characterized three different promoters for use in such vectors. The key parameters that must be taken into account are the background with promoter segment alone (no enhancer) and the ability of the promoter used to respond to different enhancer types and in different tissues.

The deleted *myo-2* promoter (in pOK1134 [8,12]) gives only a faint background of ßgal expression. Experiments with a variety of enhancer elements has shown that this promoter segment can respond to enhancement in body wall muscles, pharyngeal muscles, and in non-muscle cells of the pharynx [8,12]. The response of the *myo-2* promoter may not be completely general (and may be biased toward expression in muscle), as seen in experiments with more generally active promoter elements [13].

A deleted *glp-1* promoter has provided us a vector with no evident bias toward expression in muscle. This promoter element may be less sensitive than the deleted *myo-2* promoter. In addition, the vector shows a sporadic background staining (with most frequent expression in posterior gut, pharynx, hyp-10, and a set of posterior hypodermal cells; expression is also seen in embryonic cells [probably hypodermal precursors] at approximately the 60 cell stage. The latter expression is similar to elements of the expression pattern for the endogenous *glp-1* gene [16] and may therefore represent physiological elements of the *glp-1* promoter).

The deleted *pes-10* promoter appears to be a much more sensitive responder to enhancer activities. With decoy-protected 1 or 2 intron *lacZ* reporters, very little background is seen with the vector alone. This promoter can respond to enhancement in a variety of tissues including gut, muscle, and hypodermis.

Some potentially more sensitive derivatives of the Δ*myo-2*, Δ*glp-1*, and Δ*pes-10* vectors have constructed by replacing the *lacZ* coding region with intron-rich segments (an upstream decoy was added to minimize background expression). These vector have not been extensively tested: for the *pes-10* and *glp-1* promoters, the level of background expression without an enhancer sequence is somewhat greater then with the original vectors (the pattern of *glp-1* expression is similar to that described above; expression from the deleted *pes-10* promoter occurs in gut and pharyngeal tissue).

One feature of the enhancer assays that we have performed is important to consider: many of the enhancer-assay constructs which behave well in F1 assays (inject DNA and stain F1) do not function well in high-copy extrachromosomal lines selected using pRF4 (*rol-6 su1006*d). We are investigating the reasons for the inefficiency of this assay. At present we strongly recommend that all enhancer assays be tested using both assays.

VI. Practical Considerations

Insertion of introns into a coding region.

Two different schemes have been used for intron insertion. The simpler scheme starts with a synthetic population of portable intron "transposon" elements which are blunt ended oligonucleotides (double stranded) that can be inserted into any existing blunt ended restriction site in the coding region

```
GTAAGTTTAAAC*******TACTAACTAAC*******ATTTAAATTTTCAG    where *'s are degenerate
CATTCAAATTTG*******ATGATTGATTG*******TAAATTTAAAAGTC    bases (34%GC+67%AT)
```

In principle, any blunt end restriction site in the coding region could be used. Surprisingly every insertion site so far used in *lacZ* (also in *gfp*) appears to allow splicing out of the transposon oligonucleotide (i.e., the resulting transcript produces active protein). Note that non-blunt sites in the coding region cannot be used, since filling in or chewing back a restriction site will generate insertions or deletions in the final coding region.

After blunt-ended sites in the coding region of *lacZ* had been exhausted, we used a second PCR-based scheme for insertion of introns. The PCR based scheme is more general in that any site can be used for insertion of intron sequences. Pfu polymerase was used to avoid errors, and the resulting products have been checked functionally by analyzing ß-galactosidase activity, but we have not re-sequenced the gene to rule out silent base changes.

In creating the intron-rich *lacZ* and *gfp* vectors, we were careful of several sequence features: first we wished all introns to have slightly different sequences (hence the use of the degenerate oligo in the blunt-end insertion strategy). This should minimize the possibility of intramolecular recombination after transformation into the worm. Second, where possible, we have mutated cryptic splice sites that flank the intron insertion sites. This was done to prevent the occurrence of abberrantly spliced side-products.

Insertion of multiple introns into an arbitrary coding region is relatively straightforward, although multiple cloning steps are required. Aliquots of the blunt ended intron-transposon material are available on request, as are detailed protocols for both intron insertion schemes.

Replacing intron-poor reporter genes in existing *lacZ* or *gfp* fusions with intron-rich regions from new vectors:

The new vectors are designed to be directly compatible with previously distributed *lacZ* and *gfp* vectors. This means that there are a variety of convenient restriction sites for simple exchange of the modified *lacZ* and *gfp* coding region for the original coding region. We have frequently used the enzymes *AgeI* and *ApaI* to carry out this exchange (this allows both *lacZ* and *gfp* fusion constructs to be inserted into any existing fusion). For shorter promoter regions (particularly those <1kb), a decoy sequence is desirable to decrease background due to readthrough transcription. To carry both intron-rich reporter and decoy sequences to an existing construct, another combination of enzymes will be required (we frequently use *SphI+AgeI* or *HindIII+ AgeI*).

Suggestions for making new *lacZ* and *gfp* fusions:

In our experience, TRANSGENE ACTIVITY PATTERNS CANNOT UNDER ANY CIRCUMSTANCES BE USED AS THE SOLE MEANS TO DETERMINE THE PHYSIOLOGICAL EXPRESSION PATTERN OF AN ENDOGENOUS GENE. Nonetheless, *lacZ* and *gfp* fusions can provide a valuable initial indication of expression pattern, and once an expression pattern has been independently confirmed by other means (antibody staining, in situ hybridization, and to some extent biological function or mosaic analysis), the reporter fusions can prove extremely valuable in further characterizing the signals responsible for the expression pattern.

Three questions are paramount in designing reporter fusion experiments:

1. How much sequence upstream of the coding region should be included in the construct?
2. How much of the internal (coding and intron) sequence should be included?
3. How should the fusion constructs be introduced and assayed?

Although there are no unique answers to these questions, a reasonable starting point is to have 4-5 kb of upstream sequence as well as the first few exons fused in frame to a *lacZ* or *gfp* vector with the *unc-54* 3' end. This should be assayed both by staining of F1 following injection and by making array-containing heritable lines. Often these two methods will give the same expression pattern, but in some cases we've seen differences. In our experience, the F1 expression pattern is slightly more likely to reflect the endogenous pattern, although (you knew we were going to say this) the only definitive experiment is to look directly at the endogenous protein using in situ hybridization or antibodies.

Frequently Asked Questions:

Q-Is there a generally expressed promoter or enhancer that is active in all tissues?
A-We don't have a characterized promoter segment that functions this way. Curiously, several genes presumed to be ubiquitously expressed (e.g. RNA polymerase II large subunit, histones) give restricted patterns in transgene fusions.

Q-How can I express a coding region in the germ line
A-We don't know. So far no reporter construct with *lacZ* or *gfp* has been active in the germ line. In some cases, we have made test constructs by inserting *lacZ* or *gfp* into plasmids which are initially sufficient for rescue of a mutant phenotype. The failure of these constructs suggests some interesting feature of gene expression in the germ line remains to be discovered.

Q-Expression of my transgene in extrachromosomal lines is mosaic (only a few cells of a given type, apparently at random). Should I integrate the transgene using gamma ray?
A-In some cases, the expression of a transgene is mosaic because of mosaic inheritance of the array. In other cases, the mosaicism appears to result from other mechanisms, i.e. there are cells which contain the transgene DNA but fail to express it (e.g., Krause et al., 1994). Thus the integration is by no means guaranteed to eliminate mosaicism.

Q-Why do the mitochondrial and secreted *gfp* expression vectors come in only one frame?
A-In order to function, these targeting signals need to be on the N terminus of the protein. This requires that a promoter fragment without any coding region be inserted upstream (hence there is no need to be concerned about reading frame).

Q-There seem to be too many vectors in this kit. Which vectors should I use for my experiment?
A- Our choice of constructs to make at the outset would be as follows:

To determine the activity pattern of a promoter fragment:
- 1 fusion in an old 1-intron *lacZ* vector (pPD21.28, 22.04, 22.11, or 81.34 depending on the reading frame)
- 1 fusion in a decoy protected *lacZ* vector with 12 introns (pPD95.03, 95.07, 95.10, or 95.57)
- 1 fusion in a maximal activity *gfp* vector (pPD95.75, 95.77, 95.79 or 95.81 or equivalent NLS-*gfp* vectors)

To assay an enhancer for function.. first choice: F1-stain assay of the segment in Δ*pes-10* construct pPD95.21
Use more sensitive vectors (more introns) if this doesn't show a signal

To examine the effects of misexpression of a defined open reading frame, try expressing the cDNA in heat shock vectors pPD49.79 or 49.83. *mec-7* and *myo-3* vectors are also worth trying with cDNA or genomic sequence.

VII. Descriptions of Distributed 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 **www** archive and use this to choose specific restriction sites and cloning strategies. Initially we plan to post sequences on the www archive, but if there is a demand, we can also post complete restriction maps produced using DNA strider. 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. Because the functional fusion constructs (e.g. with the *unc-54* promoter) are the best guarantee of an active coding region, we are making these available for applications in which just the reporter is needed.

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 *lacZ* and *gfp* reporters were tested with an *unc-54* promoter (*MluI* to *SfaNI*) segment previously shown to have weak, tissue-specific, activity in body wall muscle cells [13]
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 [3]
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
∑ is a synthetic intron inserted just upstream of the *unc-54* sequence
𝕃 is a synthetic intron inserted into the 3' UTR of *unc-54*

Frame: For making translational fusions with promoterless vectors, it is critical that the fusion between your inserted DNA and the reporter be in frame. The reading frame that continues into *lacZ* is shown. The entire upstream multiple-cloning sites reproduced below is present in each of the promoterless vectors (see also Fire et al. [3]; The frame need not be considered in enhancer assay vectors.) The TAG sequence in the *XbaI* site of frame I precludes the use of upstream sites in this set of vectors. This has been remedied by producing a special set of vectors (labeled frame "S"), which provide the missing frame for HindIII, SphI, PstI, and SalI fusion junctions.

```
HindIII      SphI    PstI  SalI   XbaI   BamHI   SmaI       BalI
  v           v       v  v       v       v       v          v       Frame 0
A AGC TTG CAT GCC TGC AGG TCG ACT CTA GAG GAT CCC CGG GAT TGG CCA ...

HindIII     SphI    PstI SalI    XbaI   BamHI    SmaI       BalI
 v           v       v v       v       v        v          v       Frame 1
aa gct tgc atg cct gca ggt cga ctc tag AGG ATC CCC GGG ATT GGC CA ...

HindIII    SphI   PstI   SalI   XbaI   BamHI     SmaI       BalI
 v          v       v  v       v       v        v          v       Frame 2
AAG CTT GCA TGC CTG CAG GTC GAC TCT AGA GGA TCC CCG GGA TTG GCC A ...

HindIII     SphI   PstI  SalI          BamHI  SmaI       BalI
 v          v       v v              v       v          v         Frame S
AA GCT TGC ATG CCT GCA GGT CGA CTA GAG GAT CCC CGG GAT TGG CCA ...
```

Note that the AgeI and KpnI sites are downstream of the point at which the frames have been shifted. These are thus in the same frame in all vectors (this reporter coding regions to be swapped between constructs as AgeI-ApaI or AgeI-EcoRI segments without regard to reading frame)

```
      KpnI
        AgeI
G GTA CCG GT
```

3' end: The 3' end from *unc-54* is the *EagI* to *DraI* segment described in Fire et al., 1990 (see [13] for precise placement of the polyadenylation/cleavage site in the message). Vectors with "No 3' end" have the 3' multiple cloning site from Fire et al [3] (unique sites include *StuI-SpeI-ApaI*).
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.

NLS-*lacZ* constructs with minimal *unc-54* promoter
Tested by transformation of *C. elegans* and analysis of *lacZ* expression pattern.

plasmid	lig	promoter	IVS	A	B	C	D	E	F	G	H	I	J	K	L	frame	3' end	decoy
pPD87.18	L2082	*unc-54*	1	+	-	-	-	-	-	-	-	-	-	-	-	0	*unc-54*	none
pPD88.22	L2128	*unc-54*	2	+	-	-	-	-	-	-	-	-	-	-	+	0	*unc-54*	none
pPD87.50	L2107	*unc-54*	4	+	-	-	-	+	-	-	+	-	-	+	-	0	*unc-54*	none
pPD88.27	L2129	*unc-54*	8	+	-	+	-	+	+	-	+	-	+	+	+	0	*unc-54*	none
pPD88.75	L2154	*unc-54*	12	+	+	+	+	+	+	+	+	+	+	+	+	0	*unc-54*	none

***gfp* constructs with minimal *unc-54* promoter**
Tested by transformation of *C. elegans* and analysis of *gfp* fluorescence pattern.

plasmid	lig	promoter	localization	mut?	IVS	A	å	ß	∂	∑	L	frame	3' end	decoy
pPD91.14	L2245	*unc-54*	SV40 NLS	w.t.	1	+	-	-	-	-	-	0	*unc-54*	none
pPD91.15	L2246	*unc-54*	SV40 NLS	w.t.	2	+	-	-	-	-	+	0	*unc-54*	none
pPD96.12	L2496	*unc-54*	SV40 NLS	w.t.	5	+	+	+	+	-	+	0	*unc-54*	none
pPD93.51	L2346	*unc-54*	SV40 NLS	S65T	5	+	+	+	+	-	+	0	*unc-54*	none
pPD93.48	L2345	*unc-54*	SV40 NLS	S65C	5	+	+	+	+	-	+	0	*unc-54*	none
pPD93.65	L2354	*unc-54*	SV40 NLS	S65T	6	+	+	+	+	+	+	0	*unc-54*	none
pPD94.81	L2406	*unc-54*	SV40 NLS	S65C	6	+	+	+	+	+	+	0	*unc-54*	none

promoterless NLS-*lacZ* vectors with introns
These vectors are one construction step removed tested, active *unc-54* constructs. Junction points and a variety of restriction patterns were tested to confirm the structures. Note: for making simple fusions with the 12-intron vectors, the decoy-containing vectors will minimize readthrough background and thus be preferred.

plasmid	lig	promoter	IVS	A	B	C	D	E	F	G	H	I	J	K	L	frame	3' end	decoy
pPD90.23	L2220	none	2	+	-	-	-	-	-	-	-	-	-	-	+	0	*unc-54*	none
pPD90.28	L2221	none	4	+	-	-	-	+	-	-	+	-	-	+	-	0	*unc-54*	none
pPD90.31	L2222	none	8	+	-	+	-	+	+	-	+	-	+	+	+	0	*unc-54*	none
pPD89.03	L2167	none	12	+	+	+	+	+	+	+	+	+	+	+	+	0	*unc-54*	none
pPD89.17	L2169	none	12	+	+	+	+	+	+	+	+	+	+	+	+	1	*unc-54*	none
pPD89.09	L2168	none	12	+	+	+	+	+	+	+	+	+	+	+	+	2	*unc-54*	none
pPD89.20	L2170	none	12	+	+	+	+	+	+	+	+	+	+	+	+	S	*unc-54*	none

promoterless NLS-*lacZ* vectors with introns and a synthetic decoy sequence
These vectors are one construction step removed tested, active *unc-54* constructs. Junction points and a variety of restriction patterns were tested to confirm the structures.

plasmid	lig	promoter	IVS	A	B	C	D	E	F	G	H	I	J	K	L	frame	3' end	decoy
pPD95.03	L2425	none	12	+	+	+	+	+	+	+	+	+	+	+	+	0	*unc-54*	+
pPD95.07	L2427	none	12	+	+	+	+	+	+	+	+	+	+	+	+	1	*unc-54*	+
pPD95.57	L2457	none	12	+	+	+	+	+	+	+	+	+	+	+	+	2	*unc-54*	+
pPD95.10	L2428	none	12	+	+	+	+	+	+	+	+	+	+	+	+	S	*unc-54*	+
pPD95.11	L2431	none	12	+	+	+	+	+	+	+	+	+	+	+	+	0	none (MCS)	+

promoterless *gfp* vectors
These vectors have been tested by insertion of the *myo-3* promoter and subsequent assay for fluorescence.

plasmid	lig	promoter	localization	mut?	IVS	A	å	ß	∂	∑	L	frame	3' end	decoy
pPD95.67	L2459	none	SV40 NLS	S65C	5	+	+	+	+	-	+	0	*unc-54*	+
pPD95.70	L2461	none	SV40 NLS	S65C	5	+	+	+	+	-	+	1	*unc-54*	+
pPD95.69	L2460	none	SV40 NLS	S65C	5	+	+	+	+	-	+	2	*unc-54*	+
pPD95.73	L2462	none	SV40 NLS	S65C	5	+	+	+	+	-	+	S	*unc-54*	+
pPD95.75	L2463	none	none	S65C	5	+	+	+	+	-	+	0	*unc-54*	+
pPD95.79	L2465	none	none	S65C	5	+	+	+	+	-	+	1	*unc-54*	+
pPD95.77	L2464	none	none	S65C	5	+	+	+	+	-	+	2	*unc-54*	+
pPD95.81	L2466	none	none	S65C	5	+	+	+	+	-	+	S	*unc-54*	+
pPD95.85	L2469	none	secretion	S65T	5	+	+	+	+	-	+	0	*unc-54*	none
pPD96.32	L2509	none	mitochondrial	S65C	5	+	+	+	+	-	+	0	*unc-54*	+

open-ended *gfp* cassettes
These vectors have been tested by use of the cassette in construction of an active *unc-54*:*gfp*:*lacZ* fusion.

plasmid	lig	promoter	localization	mut?	IVS	A	å	ß	∂	∑	L	frame	3' end	decoy
pPD79.44	L1729	*unc-54*	none	w.t.	0	-	-	-	-	-	-	n.a.	-	none
pPD95.02	L2419	*unc-54*	none	S65C	6	-	+	+	+	-	-	n.a.	-	none

NLS-*gfp-lacZ* fusion constructs
The promoterless vectors have been confirmed only by restriction digestion; the promoter containing vectors have been tested for *lacZ* and *gfp* in vivo. The multiple intron (15ivs) constructs give substantially better signal (although this may still be less than the individual reporters alone); localization of the chimeric protein is punctate, perhaps representing precipitated inclusion bodies.

plasmid	lig	promoter	localization	mut?	IVS	frame	3' end	decoy
pPD80.08	L1735	*myo-2*	NLS	w.t.	1 (A) 0	*unc-54*	none	
pPD95.93	L2471	*unc-54*	NLS	S65C	4 (Aåß∂)	0	*unc-54*	none
pPD96.02	L2472	*unc-54*	NLS	S65C	15 (Aåß∂BCDEFGHIJKL)	0	*unc-54*	none
pPD96.04	L2492	none	NLS	S65C	15 (Aåß∂BCDEFGHIJKL)	0	*unc-54*	+
pPD96.62	L2538	none	NLS	S65C	15 (Aåß∂BCDEFGHIJKL)	1	*unc-54*	+

enhancer assay vectors ("minimal" promoter::NLS::*lacZ*)
The 1134 and 69.39 vectors have been used extensively (see refs [8,9,12,13] for more extensive description of these vectors and enhancer assays in general). The Δ*pes-10* vectors have been tested for spontaneous background (which increases with increasing intron number), these vectors are single-step derivatives of constructs that and stimulation by muscle- and posterior-specific enhancers. Note that pPD95.16 still gives considerable signal in midstage embryos and in later stages, even with the synthetic "decoy" sequence.

plasmid	lig	promoter	IVS	A	B	C	D	E	F	G	H	I	J	K	L	frame	3' end	decoy
pOK1134	VP/PO	Δ*myo-2*	1	+	-	-	-	-	-	-	-	-	-	-	-	1	*unc-54*	none
pPD95.62	L2458	Δ*myo-2*	12	+	+	+	+	+	+	+	+	+	+	+	+	1	*unc-54*	+
pPD69.39	L1399	Δ*glp-1*	0	-	-	-	-	-	-	-	-	-	-	-	-	0	*unc-54*	none
pPD95.16	L2434	Δ*glp-1*	11	-	+	+	+	+	+	+	+	+	+	+	+	0	*unc-54*	+
pPD95.21	L2437	Δ*pes-10*	2	+	-	-	-	-	-	-	-	-	-	-	+	1	*unc-54*	double
pPD95.27	L2438	Δ*pes-10*	5	+	-	-	-	+	-	-	+	-	-	+	+	1	*unc-54*	double
pPD95.25	L2439	Δ*pes-10*	8	+	-	+	-	+	+	-	+	-	+	+	+	1	*unc-54*	double
pPD95.18	L2436	Δ*pes-10*	12	+	+	+	+	+	+	+	+	+	+	+	+	1	*unc-54*	double

new ectopic expression vectors
Two additional promoters have been used ectopic expression vectors have been produced since the initial set of such vectors was handed out. The *myo-3* vector gives a more uniform expression pattern in body wall muscles than was seen with either of the initial *unc-54* vectors. The *mec-3* vector may be useful for experiments designed for early touch-cell expression. The three-intron expression vectors are a (so-far untested) approach toward obtaining more uniform and robust expression.

plasmid	lig	promoter	IVS	Expression MCS (unique sites)	3' end	decoy
pPD95.86	L2470	*myo-3*	1 (A)	*NheI-AccI-SalI-KpnI-NcoI-SacI*	*unc-54*	none
pPD57.56	L1026	*mec-3*	1 (A)	*NheI-AccI-HincII-SalI-NcoI-EcoRV-SacI*	*unc-54*	none
pPD96.41	L2528	*mec-7*	3 (A∑L)	*StyI-NheI-KpnI-AgeI-EcoRV-XhoI-BglII-Eco47III-EspI*-ivs∑-*EcoRI*	*unc-54*	+
pPD96.48	L2531	*myo-2*	3 (A∑L)	*StyI-NheI-KpnI-AgeI-XhoI-SacI-BglII-Eco47III-EspI-BsmI*-ivs∑-*EcoRI*	*unc-54*	+
pPD96.52	L2534	*myo-3*	3 (A∑L)	*NheI-AccI-SalI-KpnI-AgeI-XhoI-SacI-Eco47III-EspI-BsmI*-ivs∑-*EcoRI*	*unc-54*	+

Other potentially useful vectors

pPD61.125 Bluescript-like vector with additional cloning sites that can be useful for cloning

pPD47.52 A lorist-like cosmid/plasmid vector (Kan-R) with a lambda based replicon. Can be used to as a cloning vector with no homology to standard pUC, bluescript constructs (also no homology to any of our other vectors). Produces low DNA yields.

pPD10.46 *unc-22* antisense RNA driven by *unc-54* promoter/enhancer. Useful as a dominant selectable marker (see ref [5])

pPD81.34 A single intron *lacZ* vector equivalent to those from the original set (pPD21.28, pPD22.04, pPD22.11) except that the polylinker has been modified to generate a sequence shown above as Frame S" This generates a polylinker which allows utilization of sites upstream of *XbaI* in the frame of pPD22.11.

pPD5.41 A complete copy of the *unc-54* gene (9kb *HpaI* fragment cloned into *SmaI* site of Bluescript KS+. Can be used as a selectable marker for *unc-54* mutant rescue [see ref 13].

pPD93.97 a body-wall muscle gfp fusion [L2370 *myo-3* N*gfp* (3 introns, S65C)]

Classic Vectors
(published and/or previously distributed)

***lacZ* vectors with 0 or 1 intron [ref 3]**
pPD8.02
pPD8.33
pPD16.43
pPD16.51
pPD18.32
pPD21.28
pPD22.04
pPD22.11
pDD16.01
pPD26.77
pPD34.110

***sup-7* vectors [ref 4]**
pAST18b
pAST19a
pAST19b
pICT19h
pICT19r
pPD26.14

ectopic expression (e.g. ref. [14]; Review in ref. [10])
pPD49.78 *hsp16-2*
pPD49.83 *hsp16-41* (Note: some early documentation mistakenly refers to this as a 16-48 vector)
pPD52.102 *mec-7*
pPD52.37 mtl-1
pPD54.01 mtl-2
pPD30.69 *myo-2*
pPD30.35 *unc-54* promoter
pPD30.38 *unc-54* enhancer/promoter
pPD49.26 no promoter (canonical vector)

some classic *lacZ* fusions
pPD50.14 *hsp16-2*
pPD50.21 *hsp16-41*
pPD20.97 *myo-2* [ref. 13]
pPD18.49 *myo-3* [ref. 13]

References:

1. 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.
2. Fire, A. , Ahnn, J., Seydoux, G., and Xu, S-Q (1994) Vectorology II: Green Fluorescent Protein (or "It isn't easy being green"). Worm Breeder's Gazette 13 (4), 30-31.
3. 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.
4. Fire, A., Kondo, K., and Waterston, R. (1990). Vectors for low copy transformation of *C. elegans*. Nucl. Acids Res. 18, 4269-4270.

5. Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. Development 113, 503-514.
6. Heim, R., Prasher, D. and Tsien, R. (1994) Wavelength mutations and posttranslational autooxidation of green fluorescent protein. Proc. Natl. Acad. Sci. 91, 12501-12504.
7. Heim, R., Cubitt, A., and Tsien, R. (1995) Improved Green Fluorescence. Nature 373, 663-664.
8. Jantsch-Plunger, V., and Fire, A. (1994) Modular structure of a nematode body muscle enhancer. J. Biol. Chem. 269, 27021-27028.
9. Krause, M., White Harrison, S., Xu, S., Chen, L., and Fire, A. (1994) Elements regulating cell and stage-specific expression of the *C. elegans* MyoD homolog *hlh-1*. Developmental Biology, *166,* 133-148.
10. Mello, C., and Fire, A. "DNA transformation" (Methods in Cell Biology: *C. elegans*, D. Shakes and H. Epstein eds., Academic Press [publication 1995].)
11. 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.
12. Okkema, P. and Fire, A. (1994) The *C. elegans* NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in *C. elegans* muscle. Development, *120,* 2175-2186.
13. 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.
14. Perry, M., Li, W., Trent, C., Robertson, B., Fire, A., Hageman, J., and Wood, W. (1993). Molecular characterization of the her-1 gene suggests a direct role in cell signaling during Caenorhabditis elegans sex determination. Genes and Devel. *7*, 216-228.
15. Pulak, R., and Anderson, P. (1993) Messenger RNA surveillance by the *C. elegans smg* genes. Genes & Development 7, 1885-1897.
16. Seydoux, G. and Fire, A. (1994). Soma-germline asymmetry in the distributions of embryonic RNAs in *Caenorhabditis elegans*. Development *120*, 2823-2834.
17. Stinchcomb, D.T., Shaw, J.E., Carr, S.H. and Hirsh, D. (1985). Extrachromosomal DNA transformation of *Caenorhabditis elegans*. Mol. Cell. Biol. 5, 3484-3496.
18. Jaussi, R., Cotton, B., Juretic, N., Christen, P., and Schumperli, D. (1985) The primary structure of the precursor of chicken mitochondrial aspartate aminotransferase. Cloning and sequence analysis of cDNA. J. Biol Chem. 260,16060-160633
19. Inouye, S., and Tsuji, F.I. (1994) Evidence for redox forms of Aequorea green fluorescent protein. Febs Letters 351, 211-214

Acknowledgments:

We would like to thank Roger Heim, Andrew Cubitt, and Roger Tsien, Jeff Way, Marty Chalfie, Jim Waddle, Craig Mello, Meagan Jacoby, Susan Berget, Nick Marsh Armstrong, and our colleagues at this institution for their help and suggestions. Work in this lab is supported by the National Institutes of Health (Grant R01-GM37706), the Carnegie Institution and the Rita Allen Foundation.

A. Intron insertion leads to more efficient *lacZ* vectors

Construct	Introns												X-GAL Staining
	A	B	C	D	E	F	G	H	I	J	K	L	(subjective measure)
pPD87.53	-	-	-	-	-	-	-	-	-	-	-	-	+/-
pPD87.18	+	-	-	-	-	-	-	-	-	-	-	-	+
pPD87.69	-	-	-	-	-	-	-	-	-	-	-	+	+
pPD88.22	+	-	-	-	-	-	-	-	-	-	-	+	+++
pPD87.50	+	-	-	-	+	-	-	+	-	-	-	+	+++++
PD88.27	+	-	+	-	+	+	-	+	-	+	+	+	+++++++++
pPD88.75	+	+	+	+	+	+	+	+	+	+	+	+	+++++++++

basal (weak) *unc-54* promoter driving *lacZ* constructs with multiple introns

Quantitative activity assays using a soluble substrate (CPRG) suggest a several hundred fold activity difference between 1-intron (+) and 12-introns (+++++++++).

B. The background also increases as introns are added

Construct	Introns												X-GAL staining
	A	B	C	D	E	F	G	H	I	J	K	L	(in gut+pharynx)
pPD90.16	-	-	-	-	-	-	-	-	-	-	-	-	none
pPD21.28	+	-	-	-	-	-	-	-	-	-	-	-	extremely rare
pPD90.23	+	-	-	-	-	-	-	-	-	-	-	+	extremely rare
pPD90.28	+	-	-	-	+	-	-	+	-	-	-	+	extremely rare
pPD90.39	+	+	+	+	+	+	-	-	-	-	-	-	occaisional
pPD90.18	-	-	-	-	-	-	+	+	+	+	+	+	rare
pPD90.31	+	-	+	-	+	+	-	+	-	+	+	+	frequent
pPD89.03	+	+	+	+	+	+	+	+	+	+	+	+	very frequent

lacZ vectors with NO PROMOTER and with multiple introns

C. Intron insertion also gives more efficient *gfp* vectors

Construct	Introns					Fluorescence Activity
	A	å	ß	∂	L	(subjective measure)
pPD74.95	-	-	-	-		None
pPD91.14	+	-	-	-		+
pPD91.15	+	-	-	+		++
pPD91.19	+	+	-	+		++
pPD96.12	+	+	+	+		+++

basal (weak) *unc-54* promoter driving *gfp* constructs with multiple introns

Revision history of documentation

1.0 version distributed at 1995 C. elegans meeting (June 4, 1995)

1.02 version posted online June 12, 1995

includes clarification to documentation of vector pPD95.16

Note that pPD95.16 contains the decoy, but that this promoter segment still gives considerable signal in midstage embryos and in later stages, even with the synthetic "decoy" sequence.

includes correction to documentation of vector pPD95.85

Note that pPD95.85 does not contain the decoy sequence.

1.021 version posted online June 27, 1995

includes correction to documentation of vectors pPD95.25, pPD95.27

Original documentation stated that pPD95.25 and pPD95.27 had 5 and 8 introns respectively.
in reality : pPD95.27 has 5 introns (AEHKL)
pPD95.25 has 8 introns (ACEFHJKL)
The documentation and vector archives have been updated to indicate the correct structures