

Firefly Luciferase Engineered for Improved Genetic Reporting

Bruce A. Sherf and Keith V. Wood
Promega Corporation

*Although firefly luciferase is widely popular as a reporter of genetic function, the structure of the native enzyme and its encoding cDNA are not necessarily optimal for unbiased expression in foreign hosts. To create a genetic reporter more generally suitable and convenient for diverse applications, we have designed a modified form of the luciferase gene, designated *luc+*, which contains multiple new features. Chief among these is removal of the peroxisomal translocation sequence to yield a cytoplasmic form of the enzyme. Other changes include removal of potentially interfering restriction sites and genetic regulatory sites from the gene, improvement of the codon usage for mammalian cells, and elimination of consensus glycosylation sites from the enzyme. A related gene, *luc+NF*, has also been developed to allow optimal creation of N-terminal fusion proteins.*

Introduction

Firefly luciferase has become highly valuable as a genetic reporter due to the convenience and performance of its luminescence assay. Today it is used in virtually every experimental biological system, including prokaryotic and eukaryotic cell culture, transgenic plants and animals, and cell-free expression systems. Although referred to simply as firefly luciferase, the enzyme is derived from a specific North American firefly, *Photinus pyralis*. The enzyme is a monomeric protein (62kDa) which generates light through monooxygenation of beetle luciferin utilizing ATP and O₂.

To use luciferase as a genetic reporter, extracts of cells containing the luciferase gene (*luc*) are mixed with substrates and luminescence is measured immediately. The assay is very rapid and sensitive, providing gene expression data with little effort. Promega has further improved the conventional luciferase assay by including coenzyme A in the assay reagent to yield greater enzyme turnover and thus greater luminescence intensity (Luciferase Assay Reagent, Cat.# E1483). Using this reagent, luciferase activity can be readily measured in luminometers or scintillation counters without the need of a reagent injection device. Luciferase activity can also be detected in living cells by adding luciferin to the growth medium. This *in vivo* luminescence relies on the ability of luciferin to diffuse through cellular membranes and on the intracellular availability of ATP and O₂ (see article on page 22 of this issue).

Despite its utility as a reporter, however, luciferase has naturally evolved for the nocturnal mating behavior of beetles and not for the convenience of experimental molecular biologists. Thus, it is not necessarily optimized for the wide variety of host organisms in which it is presently being used. The major limitation may be that luciferase is a peroxisomal enzyme which may impact cellular physiology in some hosts. To improve the general suitability of luciferase as a genetic reporter, we have developed a modified form of the luciferase gene called *luc+*. The purpose of these modifications was to minimize potential biological interferences that may complicate the interpretation of reporter data. In general, *luc+* and the related *luc+NF* were designed to provide a biologically neutral form of the luciferase reporter gene.

Improved reporter gene: *luc+*

Normally, in the firefly light organ, luciferase is located in specialized peroxisomes of the photocytic cells. When expressed in foreign hosts, a conserved translocation signal within the enzyme structure causes it to accumulate in peroxisomes and glyoxisomes. In moderate to high levels of expression, the peroxisomes typically become saturated with luciferase, and much of the reporter is found in the cytoplasm (1). Localization to the peroxisomes, however, could interfere with normal cellular physiology in two ways. First, large amounts of a foreign protein in the peroxisomes could impair their normal function. Second, many other peroxisomal proteins utilize the same translocation signals (2), and thus, saturation with luciferase implies competition for the import of other peroxisomal proteins.

Peroxisomal location of luciferase may also interfere with the performance of the genetic reporter. For instance, the luciferase accumulation in the cell could be differentially affected if it is distributed into two different subcellular compartments. The stability of luciferase in peroxisomes is not known, but will be different than its stability in the cytosol. If so, the apparent expression of luciferase could be affected by changes in the distribution of luciferase between peroxisomes and the cytosol. Measurements of *in vivo* luminescence could also be affected since the availability of ATP, O₂, and luciferin within peroxisomes is not known. In particular, the concentration of luciferin could be limited by the need to diffuse across *both* the cytoplasmic and peroxisomal membranes. This partitioning of luciferase between different intracellular compartments may add unforeseen variability when comparing luciferase activities derived from genetic expression elements with different transcriptional activities.

The peroxisomal translocation signal in firefly luciferase has been identified as the C-terminal tripeptide sequence, -Ser-Lys-Leu. Removal of this sequence abolishes import into peroxisomes (3). However, the relative specific activity of this modified luciferase has not been determined. To develop an optimal cytoplasmic form of the luciferase gene, we followed two strategies. First, we designed a new C-terminal tripeptide sequence based on the available data to minimize peroxisomal import, -Gly-Lys-Thr (4). Second, we applied random mutagenesis to the C-terminal region and selected brightly luminescent colonies of *E. coli* transformed with the mutagenized luciferase genes. From sequence data of these selected mutants, we chose a clone with the sequence -Ile-Ala-Val. Consistently, both modified luciferases yielded about 4- to 5-fold greater luminescence than the native enzyme when expressed in NIH3T3 cells. We chose the luciferase containing -Ile-Ala-Val sequence for the cytoplasmic form because it usually yielded slightly greater luminescence than the luciferase with -Gly-Lys-Thr.

In addition to removing the peroxisomal translocation signal, we made several other modifications to enhance the reliability and convenience of luciferase as a genetic reporter. [Table 1](#) presents the complete list of sequence modifications to *luc*, which can be divided into four classes:

i) Restriction endonuclease sites. The native luciferase gene contains restriction sites for *Xba* I, *Eco*R I, *Bst*E II, *Eco*R V, and *Cla* I. To facilitate subcloning of the gene into diverse genetic constructs, we removed these sites by changing nucleotides in the DNA sequence without affecting the amino acid sequence. An *Nco* I site was also added at the initiating methionine codon (ATG) to aid in subcloning into many vectors which contain this site. The *Nco* I site may also be used to create N-terminal fusion proteins with luciferase, however, for this purpose we recommend using *luc*+NF (see below).

ii) Regulatory sequences. Any reporter gene may contain regulatory sequences imbedded within its coding region, which could mediate genetic activity either through the gene's native regulatory function or as a consequence of spurious recognition by transcription factors in a foreign host. In either case, these sequences would interfere with the "genetically neutral" behavior expected of a reporter gene. To minimize this possibility, we scanned the luciferase gene sequence using a database of consensus

sequences for transcription factor binding sites (5) and removed many sites which could potentially interact with common factors. In some cases where it was convenient in our modification strategy, we also removed less common potential regulatory sites. As with the removal of restriction sites, the potential regulatory sites were removed through changes in the DNA sequence that do not affect the encoded amino acid sequence. We also removed three palindromic sequences which could spuriously affect expression (22bp, 18bp, and 16bp, each with one mismatch).

iii) Codon usage. In general, codon usage presumably reflects the availability of tRNA isoforms in different organisms; efficiently expressed genes utilize the most abundant tRNA isoforms. Codon usage in mammalian cells reveals a preference for cytosine (C) or guanine (G) in the third codon position; many codons containing adenine (A) or thymidine (T) occur rarely. However, beetle luciferases generally have a high A/T content, biasing the codon frequency significantly from that of mammals. To achieve a codon usage in the firefly luciferase gene that is more congruent with mammalian genetics, the sequence modifications described above were designed wherever possible to yield more common codons (6). Also, where possible in our modifications strategy, we changed codons of ATA to ATC (Ile), GTA to GTG (Val), and TTA to CTG or CTC (Leu). These changes were chosen because they convert particularly infrequent codons to ones which are highly frequent. In total, of the 69 codons that were modified to create *luc+*, 54 represent more common codons in mammalian cells. The average usage frequency of all codons increases from 14.5 in *luc* to 25.7 in modified *luc+* (usage frequency is the occurrence per 1000 codons).

iv) Glycosylation sites. Native luciferase expressed in the peroxisomes or the cytosol normally does not contain any post-translational modifications. However, gene fusions may be made of *luc+* (or *luc+*NF) which are intended to direct a hybrid protein into the endoplasmic reticulum or Golgi apparatus. In these cellular compartments, N-linked glycosylation is known to occur at -Asn-X-(Ser/Thr)- sequences, which would have an unknown effect on luciferase enzymatic activity. To prevent the potential occurrence of N-linked glycosylation, we have altered two consensus glycosylation sites within the luciferase sequence. The modified enzyme does not exhibit any apparent change to its chemical activity. A third consensus glycosylation site is highly conserved among beetle luciferases and could not be altered in firefly luciferase without affecting enzyme performance. Most likely this site is not near the protein surface, possibly making it inaccessible for glycosylation.

Table 1. Summary of Luciferase Gene Modifications in *luc+*.

Purpose of Modification	Sequence Modification in <i>luc+</i>
Introduce <i>Nco</i> I site for the construction of N-terminal fusions with <i>luc+</i> .	<i>luc</i> : AAA Met, Glu, ATG GAA
	<i>luc+</i> : TCC Met, Glu, ATG GAA <i>Nco</i> I #47 <i>Xba</i> I
Remove internal <i>Xba</i> I site; disrupt extended palindrome.	<i>luc</i> : CTCTAGAGG
	<i>luc+</i> : C6CTGGAAAG
Eliminate potential glycosylation and ATF sites.	<i>luc</i> : AAC Ile ₁₀₀ Thr ₁₀₀ TACGGGAA ATC ACG
	<i>luc+</i> : GAC Ile ₁₀₀ Thr ₁₀₀ TACGGTGA ATC ACT
	<i>luc</i> : Asn ₁₀₀ Ile ₁₀₀ Ser ₁₀₀ TACGGTGA ATC TCG
	<i>luc+</i> : Gly ₁₀₀ Ile ₁₀₀ Ser ₁₀₀ TACGGTGA ATT TCG
Remove potential TGT-3 site; improve codon usage.	<i>luc</i> : #373 GTAGTGTGGTT
	<i>luc+</i> : GTGGTGTGGTT
Improve codon usage.	<i>luc</i> : #426 ATTACCAATAATCCAG
	<i>luc+</i> : GCTCCCAATCATCCAA
Improve codon usage.	<i>luc</i> : #546 ACCAGAGTCCTTTGATCGTGACAAA
	<i>luc+</i> : GOCAGAGTCCTTCGATAGGGACAAAG
Remove internal <i>Eco</i> R I site; improve codon usage.	<i>luc</i> : #583 <i>Eco</i> R I ATAATGAATCC
	<i>luc+</i> : ATCATGAACCTCC
Remove internal <i>Bst</i> E II site and potential AP2 and LF-A1 sites; improve codon usage.	<i>luc</i> : #609 <i>Bst</i> E II GGTACCTAAGGGTGTGGCCCTCCG
	<i>luc+</i> : GTCTGCCAAAGGTGCGCTCGCT
Remove potential AP1 site.	<i>luc</i> : #646 TGCCTCAG
	<i>luc+</i> : TGCCTGAG
Improve codon usage.	<i>luc</i> : #820 TTACGATCCCTCAGGATTACAAA
	<i>luc+</i> : CTGAGGAGCCTTCAGGATTACAAAG
Improve codon usage.	<i>luc</i> : #858 TTGCTAGTACCAACCCATTTCAC
	<i>luc+</i> : CTGCTGGTGCCAACCCATTCTCC
Eliminate internal palindrome; improve codon usage.	<i>luc</i> : #945 GGGCGCACCTCTTCGAAA
	<i>luc+</i> : TGGCGTCCCTCTCTAAG
Improve codon usage.	<i>luc</i> : #984 AAAACGCTTCATCTCCAGGGATACGA
	<i>luc+</i> : CAAGAGGTTCCATCTGCCAGGATCAGG
Eliminate potential AP1 site; improve codon usage.	<i>luc</i> : #1158 GAGAGCGGAATTATGTGTCAGAGGA
	<i>luc+</i> : AAGAGCGCAACTGTGTGAGAGGT
Eliminate palindrome structure; improve codon usage.	<i>luc</i> : #1302 AGTTGACCGCTTGAAGTCTTTAATTAATAC
	<i>luc+</i> : CGTTGACCGCTTGAAGTCTCTGATTAAGTAC
Remove internal <i>Eco</i> R V site; improve codon usage.	<i>luc</i> : #1333 <i>Eco</i> R V AAAGGATACAGGTGGCC
	<i>luc+</i> : AAAGGCTATCAGGTGGCT
Remove internal <i>Cla</i> I site; improve codon usage.	<i>luc</i> : #1365 <i>Cla</i> I ATCGATATTGTTA
	<i>luc+</i> : ATCCATCTTGCTC
Remove potential Sp1 and AP2 sites.	<i>luc</i> : #1400 CGGGCGTGCC
	<i>luc+</i> : CAGGTGTCCG
Remove peroxisome targeting sequence.	<i>luc</i> : Gly ₁₀₀ Lys ₁₀₀ Ser ₁₀₀ Lys ₁₀₀ Leu ₁₀₀ stop GGA AAG TCC AAA TTG TAA
	<i>luc+</i> : Gly ₁₀₀ Lys ₁₀₀ Ile ₁₀₀ Ala ₁₀₀ Val ₁₀₀ stop GGA AAG ATC GCC GTG TAA

The numbering scheme for amino acids is relative to the first amino acid, Met #1, of luciferase. The numbering of nucleotides is relative to the first base, "A", of the luciferase open reading frame.

Performance of *luc+* in mammalian cells

The changes incorporated into *luc+* are intended to minimize the potential for unexpected interferences with reporter performance under specific experimental conditions, providing a more reliable indicator of genetic activity than the native luciferase cDNA clone. On the basis of our design strategies, we expect the performance of *luc+* to be comparable to *luc* in most circumstances.

To test this, we substituted *luc+* for *luc* in the pGL2 Luciferase Reporter Vectors previously developed at Promega (Cat.# E1641, E1631, E1621, E1611). These vectors constitute a uniform set of genetic constructions with different combinations of SV40 promoter and enhancer sequences. The pGL2-Basic Vector contains the luciferase gene with a small-t antigen intron and mRNA polyadenylation site from SV40. The pGL2-Promoter and pGL2-Enhancer Vectors are identical to the pGL2-Basic Vector except they contain the SV40 late promoter and enhancer, respectively. The pGL2-Control Vector contains both the promoter and enhancer.

[Figure 1](#) presents the results of the comparison between *luc* and *luc+* activity in CHO cells. These results

show that the performance of *luc+* as a genetic reporter is analogous to *luc* under all combinations of genetic elements in the pGL2 Vectors. Enzyme performance in the assay procedure was also identical for all extracts.

In other cell types, differences between the performance of *luc+* and *luc* are apparent. To show this, the pGL2-Control Vector containing *luc* or *luc+* was transformed into NIH3T3 cells, HeLa cells, CHO cells, and CV-1 cells. Cells were grown and harvested as described in [Figure 1](#). The results showed that in cell lines expressing relatively high levels of luciferase activity, the performance of *luc+* and *luc* were comparable ([Table 2](#)). With the lower levels of expression, however, *luc+* supported relatively higher levels of luciferase. From our earlier experiments in NIH3T3 cells, this greater level of expression is due predominantly to the removal of the peroxisomal translocation signal, suggesting the difference in luciferase expression here is apparently associated with its translocation into peroxisomes.

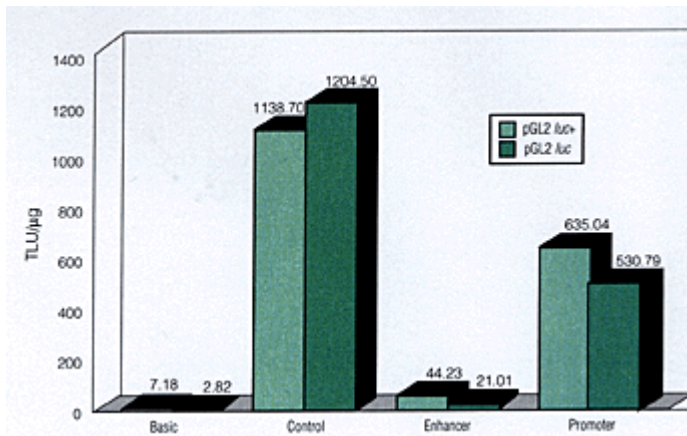


Figure 1. Comparison of *luc* and *luc+* activity in CHO cells. The pGL2-Vectors containing *luc+* or *luc* were transformed into CHO cells using calcium phosphate (ProFection[®] Mammalian Transfection System, Cat.# E1200) according to procedures supplied (7) and the cells were grown 30 hours. Extracts of the cells were made using Reporter Lysis Buffer (Cat.# E3971), and enzyme activity was measured using Luciferase Assay Reagent following the manufacturer’s recommendations (8).

Table 2. Relative Expression of *luc+* and *luc* in Various Mammalian Cell Lines.

Relative Luminescence Expression			
Cell Type	<i>luc</i>	<i>luc+</i>	Ratio
NIH3T3	0.0011	0.0050	4.7
HeLa	0.0033	0.0062	1.9
CHO	0.18	0.20	1.1
CV-1	0.90	1.00	1.1

These results are consistent with the hypothesis that reporter stability may differ between the peroxisomes and cytosol. Since higher levels of expression lead to saturation of the peroxisomes, a large portion of the native luciferase synthesized in CHO and CV-1 cells may be in the cytosol. These

conditions should minimize differences between the cytoplasmic form of luciferase from *luc+* and the native luciferase from *luc*. Under lower levels of expression, however, the differences between the luciferase forms would become more apparent as they are increasingly segregated into their respective subcellular compartments. It is interesting that the increased luminescence supported by *luc+* in NIH3T3 cells is equally evident for both *in vivo* and *in vitro* measurements. This suggests that peroxisomes do not restrict the substrate availability to the sequestered enzyme.

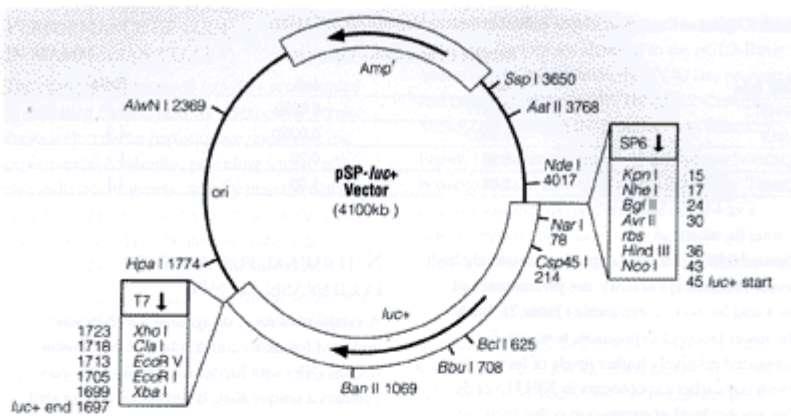
N-terminal fusions with luciferase: *luc*+NF

A variation of *luc+*, designated *luc*+NF, was designed for easier construction of N-terminal fusions (NF) with luciferase. The *luc*+NF gene contains a unique *Bst*E II restriction site located immediately downstream of the luciferase translational initiation codon (ATG). The *Bst*E II site allows construction of N-terminal fusions that replace the original ATG codon by the newly introduced DNA; sequences also may be placed between the luciferase gene and its initiation codon. By removing the ATG codon from the fusion site, spurious internal initiation at this codon can be confidently avoided, preventing the possibility of coexpressing full-length, unfused luciferase. Internal initiation from the next available ATG codon generates a polypeptide which is too short to support luminescence (unpublished data).

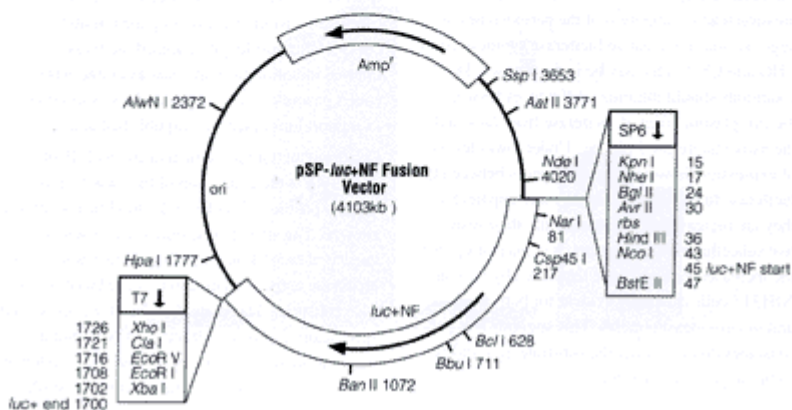
Concomitant with placement of the *Bst*E II site in *luc*+NF is the generation of two new amino acids at positions 2 and 3 of the modified luciferase enzyme. The altered N-terminal amino acid sequence of *luc*+NF may result in 2- to 3-fold lower luciferase activity relative to that produced by the *luc*+ construct. Therefore, *luc*+NF is recommended specifically for the construction of N-terminal fusion proteins devoid of an internal ATG codon at the luciferase juncture, or that require the resident ATG for translational initiation.

New cassette vectors: pSP-*luc*+ and pSP-*luc*+NF

pSP-*luc*+ and pSP-*luc*+NF are cassette plasmids containing the improved firefly luciferase genes, *luc*+ and *luc*+NF (Figure 2). These cassette vectors are not themselves intended for the eukaryotic expression of luciferase because they do not contain eukaryotic genetic regulatory elements. The *luc*+ and *luc*+NF gene are positioned downstream of an SP6 promoter and minimal ribosome binding site which, in the presence of SP6 polymerase, drive *in vivo* and *in vitro* expression of the modified luciferases. An opposing T7 promoter is also located immediately downstream of *luc*+ and *luc*+NF. These promoters allow for the convenient synthesis of sense and antisense *luc*+ or *luc*+NF transcripts for studies involving *in situ* hybridization, RNA processing, RNA transfection, or coupled *in vitro* transcription/translation and protein folding. Multiple cloning sites containing recognition sequences for a number of commonly used restriction enzymes are positioned 5' and 3' of *luc*+ and *luc*+NF.



pSP-*luc+* Vector circle map.



pSP-*luc+NF* Fusion Vector circle map.

Figure 2. Circle maps of pSP-*luc+* and pSP-*luc+NF*.

Acknowledgment: The authors would like to thank Shauna Navarro for her assistance in performing and analyzing experiments in luciferase expression.

Editor's note: Watch for a future *Promega Notes* article covering the use of the modified *luc+* in new reporter vectors, the pGL3 Vectors. More on the pGL3 Luciferase Reporter Vectors will also be presented as a poster by D. Groskreutz, B. Sherf, E. Schenborn, K. Wood and B. Brondyk at the annual meeting for the American Society of Cell Biology.

References

1. Keller, G.-A. *et al.* (1987) *Cell Biol.* **84**, 3264.
2. Gould, S. J. *et al.* (1990) *J. Cell Biol.* **110**, 27.
3. Gould, S. J. *et al.* (1989) *J. Cell Biol.* **108**, 1657.

4. Sommer, J. M. *et al.* (1992) *Mol. Biol. Cell.* **3**, 749.
5. Faisst, S. and Meyer, S. (1992) *Nucleic Acids Res.* **20**, 3.
6. Wada, K. *et al.* (1992) *Nucleic Acids Res.* **20**, 2111.
7. ProFection[®] Mammalian Transfection Systems Technical Bulletin #TM012, Promega Corporation.
8. Luciferase Assay Systems Technical Bulletin #TB101, Promega Corporation.

Ordering Information

Product	Size	Cat.#
pSP- <i>luc</i> + Vector	20mg	E1781
pSP- <i>luc</i> +NF Fusion Vector	20mg	E4471
pGL3-Control DNA	20mg	E1741
pGL3-Enhancer DNA	20mg	E1771
pGL3-Promoter DNA	20mg	E1761
pGL3-Basic DNA	20mg	E1751