

ARGENT™
Regulated Transcription Retrovirus Kit

Version 2.0

ARIAD Pharmaceuticals, Inc.
26 Landsdowne St
Cambridge, MA, 02139

RegulationKits@ariad.com
Tel: (617) 494-0400

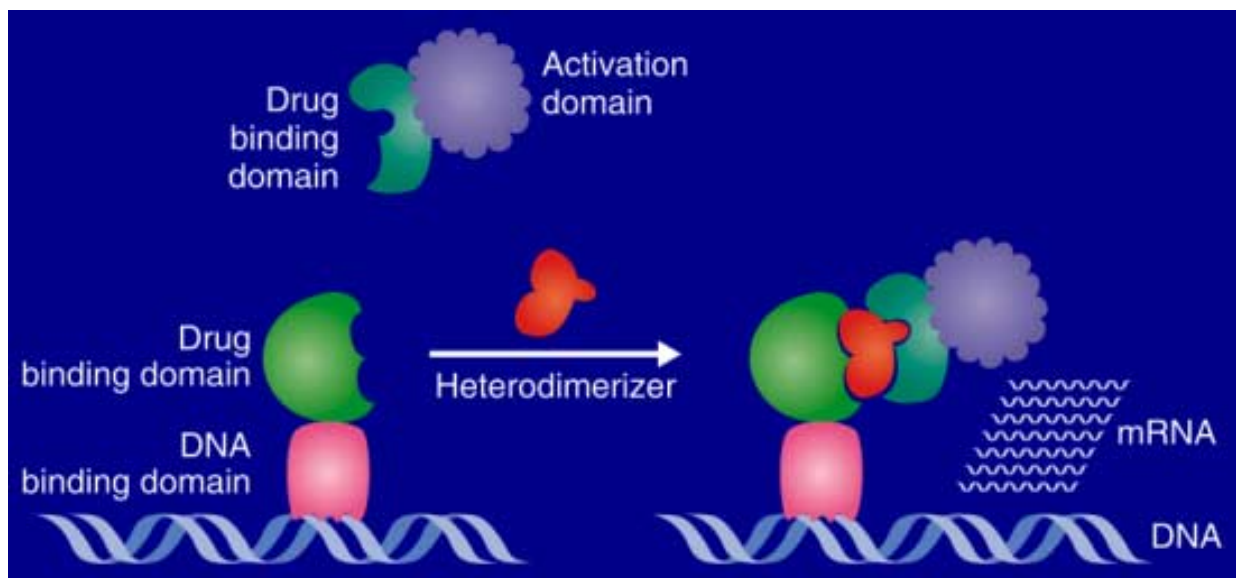
09/09/02

Table of Contents

ARGENT™ Regulated Transcription Retrovirus Kit.....	1
Table of Contents	2
Overview	3
Applications of the ARGENT Regulated Transcription Retrovirus Kit.....	4
Design of the kit components	5
Notes on the use of this kit	6
Kit contents	6
Transcription factor vector.....	7
pL ₂ N ₂ -R _H S3H/ZF3.....	7
Target gene vector.....	8
pLH-Z ₁₂ I-PL	8
All-in-one vector.....	9
pL ₆ N ₂ -R _H S3H/ZF2-PL.....	9
Shuttle vector	10
pBS-SV-PL-Ap	10
General vector information.....	10
Which vectors to use?.....	10
Generation of a human growth hormone control vector.....	10
Retrovirus Production	11
Generation of inducible clones.....	11
Creating an all-in-one vector in a new vector backbone	11
Antibodies to detect fusion proteins.....	12
AP21967	12
Description.....	12
Reconstituting AP21967.....	13
Storage and handling of AP21967	13
Using AP21967 <i>in vitro</i>	13
Use of AP21967 in animals.....	13
Expected results	13
Conditions of use	14
References	15
Appendix.....	16
pL ₂ N ₂ -R _H S3H/ZF3 Annotated Sequence.....	17
pLH-Z ₁₂ I-PL Annotated Sequence	20
pL ₆ N ₂ -R _H S3H/ZF2-PL Annotated Sequence.....	22
pBS-SV-PL-Ap Annotated Sequence	25

ARGENT™ Regulated Transcription Retrovirus Kit

ARGENT™ Regulated Transcription Kits contain reagents for placing the transcription of a target gene under the control of a small molecule “dimerizer”. The kits can be used to achieve tightly regulated conditional expression of genes of interest, allowing gene function to be investigated *in vitro* or *in vivo*. In the Retrovirus Kit described here, the regulatory system has been incorporated into retroviral vectors. A [plasmid-based](#) version is also available.



[Controlling Gene Expression Using Regulated Transcription](#)

Overview

Activation of gene expression in eukaryotes is controlled by the induced binding of transcription factor proteins to target genes. Transcription factors are bifunctional proteins that recognize specific DNA sequences near target genes and then recruit the transcriptional machinery of the cell to activate transcription. The two domains responsible for these activities, the DNA binding domain and the transcriptional activation domain, are functionally separable and can reconstitute a sequence-specific transcriptional activator even when expressed as individual proteins and brought together via a noncovalent interaction.

This modular architecture has been exploited to develop a general method for controlling gene transcription using small molecules. The strategy is based on the use of chemical inducers of dimerization, or “dimerizers”, to induce the interaction of engineered proteins (1). A dimerizer is a cell-permeant organic molecule with two separate motifs that each bind with high affinity to a specific protein module. By fusing such modules to a DNA binding domain and an activation domain, the reconstitution of a functional transcription factor, and therefore the expression of a target gene, can be made absolutely dependent on the presence of dimerizer (see the figure).

In principle, the expression of any cloned gene can be brought under dimerizer control, by equipping the gene with upstream sequences that are recognized by the engineered DNA binding domain. Following introduction of the modified gene into cells that also express the engineered transcription factor proteins, addition of dimerizer will lead to dose-dependent activation of target gene expression. Because the transcription factor fusion proteins have no affinity for one another in the absence of

dimerizer, regulation is characterized by an extremely low, usually undetectable, basal level of gene expression. In addition, the highly potent activation domains incorporated into the system typically lead to high maximal levels of induced gene expression, often in excess of levels obtained with strong constitutive promoters/enhancers.

There are two classes of dimerizers. **Homodimerizers** incorporate two identical binding motifs, whereas **heterodimerizers** have two different binding motifs, allowing the specific dimerization of two different proteins when they are fused to two appropriate ligand binding domains. Regulated transcription is usually accomplished using a heterodimerizer, as shown in the figure, since this leads to the most efficient reconstitution of an active transcription factor. The ARGENT™ Regulated Transcription Retrovirus Kit described here provides a heterodimerizer, vectors encoding the engineered transcription factor fusion proteins, and target gene vectors into which genes of interest can be inserted. The two components of the regulatory system— the engineered transcription factors and the target gene cassette— are provided both as separate retroviral vectors, and a single, “all-in-one” retroviral vector. A separate kit is available in which the constructs are instead provided on plasmid vectors (see [ARGENT Regulated Transcription Plasmid Kit](#)).

Applications of the ARGENT Regulated Transcription Retrovirus Kit

The ability to control the transcription of specific genes using small molecules has broad utility in biological research. Varying the expression level of a gene is a powerful way to study its function, allowing the creation of inducible alleles in cell culture and in transgenic animals. By precisely varying the expression level using the dose of ligand, detailed questions can be asked about the physiological role of the gene, and the protein it encodes.

The ARGENT regulated transcription system has been used to achieve regulated gene expression in a variety of contexts, including in transiently or stably transfected cells in culture (2-4), and in mice and primates when delivered using adenovirus and adeno-associated virus (AAV) vectors (5, 6). Under all conditions tested, the system has allowed tight, dose-dependent control of gene expression. A complete list of publications describing the use of ARGENT regulated transcription reagents can be found in the [Regulation Kit Bibliography](#).

The retroviral reagents provided in this kit are based on those described in a [recent publication](#) in *Proceedings of the National Academy of Sciences* (7). Incorporation of the regulatory system into retroviral vectors was shown to allow its rapid, efficient and stable delivery to a broad range of cell types, including those that are difficult to transfect. These characteristics make the system particularly useful for the rapid generation of stable cell lines for the *in vitro* analysis of gene function. However, the ligand provided with this kit, AP21967, is also suitable for *in vivo* use and has been used successfully to achieve regulated gene expression in mice.

A key feature of regulation of transcription using the dimerizer system is the extremely low level of basal expression in the absence of inducer. This is critical for applications where even minimal “leakiness” (expression in the absence of inducer) is unacceptable. In the recent *PNAS* publication, we used an all-in-one retroviral vector to generate stable cell lines inducibly expressing the highly toxic diphtheria toxin A chain gene, suggesting that the system will be particularly suitable for analysis of the many interesting genes that promote cell death, block the cell cycle, or are otherwise toxic.

Tight, dimerizer-inducible control of gene expression also has applications in many areas of drug discovery. For example, cell lines in which expression of a single gene can be chemically induced may be useful in the configuration of targeted cell-based assays for small molecule drugs. The reagents of the Retroviral Kit, and in particular the all-in-one vector, provide a rapid and efficient way to generate such cell lines.

Design of the kit components

Rapamycin and its analogs

The reagents in the ARGENT™ Regulated Transcription Retroviral Kit, like those of the other ARGENT kits, are based on the human protein FKBP12 (FKBP, for FK506 binding protein) and its small molecule ligands. FKBP is an abundant cytosolic protein that serves as the initial intracellular receptor for the natural product immunosuppressive drugs FK506 and rapamycin. Both these drugs act naturally as heterodimerizers, and both have been used to control transcription (2, 3), as has FK-CsA, a cyclosporin-FK506 hybrid molecule (8). We have focused on the use of rapamycin, because it has well understood chemistry and has favorable pharmacokinetic properties. Rapamycin functions by binding with high affinity to FKBP, and then to the large PI3K homolog FRAP (RAFT, mTOR), thereby acting as a heterodimerizer to join the two proteins together (9). To control transcription of a target gene, a DNA binding domain is fused to one or more FKBP domains, and a transcriptional activation domain is fused to a 93 amino acid portion of FRAP, termed FRB, that is sufficient for binding the FKBP-rapamycin complex (10). Only in the presence of rapamycin are the two fusion proteins dimerized and therefore capable of activating the transcription of a gene equipped with binding sites for the DNA binding domain (2).

In some cases, the use of rapamycin may be compromised by its cell cycle inhibitory effects (the result of inhibiting FRAP kinase activity, which in T cells leads to immunosuppression). To overcome this limitation, we have engineered the system to function with non-immunosuppressive analogs of rapamycin, which we call rapalogs. These compounds have been chemically modified so that they no longer can bind to wild-type endogenous FRAP, greatly reducing immunosuppressive activity. The compounds can however bind to a modified FRAP that contains a single designed amino acid change (T2098L). Incorporation of this mutation into the FRB domain fused to the activation domain allows a rapalog to be used to specifically heterodimerize the engineered transcription factor fusion proteins without interfering with the activity of endogenous FRAP.

The redesigned rapamycin system forms the basis of this kit, which includes the mutant FRB sequence, and a non-immunosuppressive rapalog, AP21967. It is important to note that the redesigned system retains the ability to respond to rapamycin itself, as well as to AP21967. Therefore experiments can be carried out with either ligand, as appropriate.

Transcription factor components

Since we are developing regulatory systems for use in human gene therapy, we have built our transcription system using only human proteins to minimize the potential for immunogenicity in clinical applications. The DNA binding domain we use is called ZFHD1, a composite human DNA binding domain with novel DNA recognition specificity (11). ZFHD1 is composed of two zinc finger domains from the human transcription factor Zif268, joined to a homeodomain derived from the human transcription factor Oct-1. ZFHD1 binds with high affinity and specificity to a unique composite DNA binding sequence, but not to Zif268 or Oct-1 binding sites. Typically, multiple copies of the ZFHD1 binding site are included upstream of target genes to obtain robust gene activation.

The activation domain used in the first version of our system consisted of the carboxy terminal 191 amino acids from the p65 subunit of human NF- κ B (12). In our system, this p65-derived domain substantially outperforms the commonly used activation domain from the herpesvirus VP16 protein. Since the level of activation of a target gene is directly proportional to the potency of the transcriptional activation domain, we have invested significant effort in trying to identify even more potent domains. We recently described a new activation domain consisting of the carboxy terminal 271 amino acids of p65 fused to the activation domain from human Heat Shock Factor 1. Use of this stronger composite activation domain, which we call S3H, led to significantly higher levels of gene activation in cells transduced with the retroviral constructs (7). For this reason the vectors provided in this kit incorporate the S3H activation domain.

Notes on the use of this kit

Use of the previous kit based on AP1510

This kit, along with the companion [plasmid-based kit](#), replaces the original Regulated Transcription kit that was based on the homodimerizer AP1510 and FKBP fusion proteins (13). We have found that the rapamycin-based reagents significantly out-perform the original reagents in all applications tested. In particular, the pharmacological properties of AP1510 preclude its use for *in vivo* studies, whereas rapamycin and AP21967 are well suited to these applications. If you are already using the AP1510-based kit, we will continue to provide dimerizer for *in vitro* experiments upon request. Please note that rapamycin and AP21967 will **not** activate transcription from constructs generated using the previous kit.

Regulating heterodimerization of proteins other than transcription factors

Regulated heterodimerization is generally applicable to the study of signaling pathway components and other proteins that function through protein-protein interactions. However the kit described here has been designed specifically for use in regulating target genes, and the genes encoding the fusion proteins cannot readily be reconfigured for other uses. For applications other than regulated transcription of target genes, please request the [ARGENT™ Regulated Heterodimerization Kit](#).

Rapamycin analog AP22565

Please note that the rapalog provided in this kit, AP21967, is different from AP22565, the analog used in our recent PNAS publication (7), although the two molecules are from the same chemical class and are highly related. AP21967 can be used equivalently in all the applications described in the paper.

Kit contents

This kit provides three retroviral vectors, each of which allows production of infectious retroviral particles following transfection into an appropriate packaging cell line. The kit includes

- A transcription factor retroviral vector
- An empty target gene retroviral vector, for inserting the gene of interest
- An all-in-one retroviral vector containing transcription factor genes and the target gene cassette
- A shuttle vector to facilitate exchange of selectable markers in the all-in-one vector
- An aliquot of the rapalog AP21967

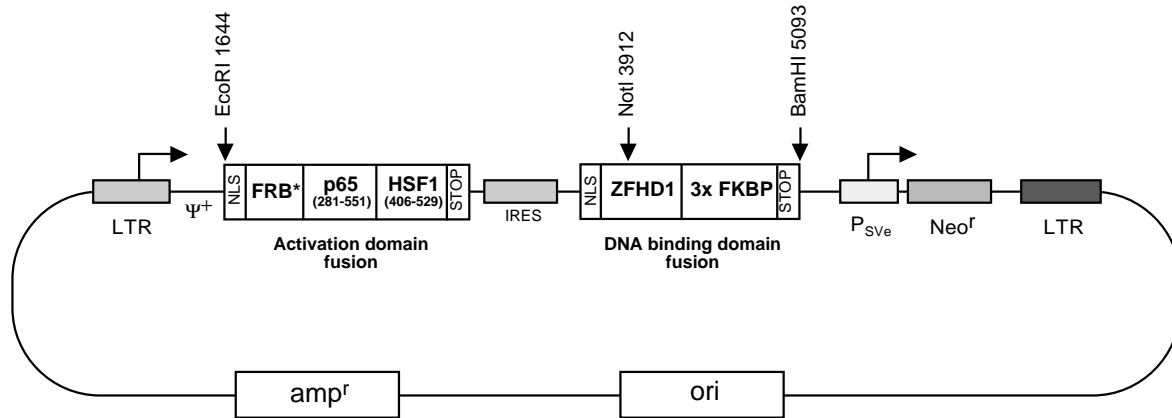
The transcription factor vector, pL₂N₂-R_HS3H/ZF3, expresses the transcription factor fusions from the retroviral LTR and also contains an internal neo selection cassette. pLH-Z₁₂I-PL is an empty target gene vector containing a hygro selectable marker, the regulated promoter, and a polylinker, allowing insertion of the target gene of interest. The all-in-one vector, L₆N₂-R_HS3H/ZF2-PL, contains the LTR-driven transcription factor fusions, a neo selectable marker and the regulated promoter with a polylinker. The neo selectable marker cassette in this vector can be replaced with a different selectable marker of interest using the fourth plasmid provided with the kit, pSV-PL-Ap.

Transcription factor vector

pL₂N₂-R_HS3H/ZF3

Description

pL₂N₂-R_HS3H/ZF3 (9000 bp)



Not drawn to scale

- The vector was constructed by inserting a transcription factor cassette between the EcoRI and BamHI sites of pLXSN2, a derivative of pLXSN (Clontech) in which the pBR322 replication origin is replaced by a pUC replication origin.
- The transcription factor cassette is expressed from the retroviral LTR on a bicistronic transcript and consists of:
 - an activation domain fusion (R_HS3H) which contains the FRB fragment of human FRAP (R_H), fused to a highly potent chimeric activation domain called S3H. S3H consists of amino acids 281 to 551 from the p65 subunit of human NFκB (S3) and amino acids 406-530 from human heat shock factor 1(H). The FRB domain consists of amino acids 2021-2113 of FRAP, in which the threonine at amino acid 2098 is mutated to leucine. This mutation allows the protein to bind to rapamycin analogs (e.g. AP21967) which no longer bind appreciably to endogenous FRAP. We have found the more potent S3H activation domain (versus the original activation domain (2)) to be required for robust dimerizer-dependent transcription in a retroviral context (7). We believe that this is due to the fact that retroviral vectors integrate at low copy number and that transcription factor expression from the retroviral LTR is not as strong as from the CMV enhancer/promoter present in plasmid based vectors.
 - a DNA binding domain fusion (ZF3) which consists of the ZFHD1 DNA binding domain (Z) and three tandemly repeated copies of human FKBP12 (F3). To reduce the likelihood of recombination during the retroviral life cycle the tandemly repeated FKBP domains are encoded by open reading frames rendered non-identical by the incorporation of silent mutations.
- Both fusion proteins contain an amino-terminal nuclear localization signal (N₂, from the human *c-myc* gene).
- The two coding regions are separated by an internal ribosome entry sequence (IRES) derived from the encephalomyocarditis virus to allow translation of the second cistron (ZF3).
- We have found the optimal configuration of the transcription factor fusions to be that in which ZFHD1 is fused to three tandemly-reiterated FKBP12 domains and the p65 activation domain to a

single FRB domain. This configuration theoretically allows recruitment of up to 3 activation domains per DNA binding site.

- pL₂N₂-R_HS3H/ZF3 also contains an internal neo selection cassette consisting of a *neo* gene (Neo^r) under the control of the simian virus 40 early promoter (PSV_e).

[Annotated Sequence](#)

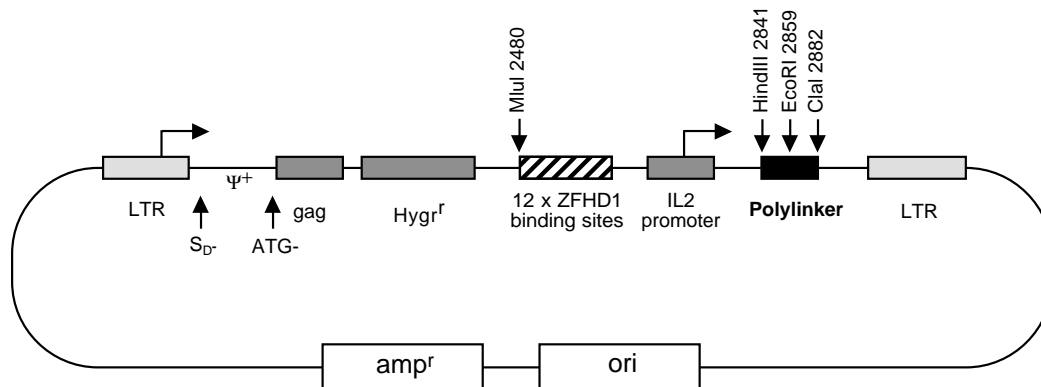
[Raw sequence](#)

Target gene vector

pLH-Z₁₂I-PL

Description

pLH-Z₁₂I-PL (5631 bp)



Not drawn to scale

- pLH-Z₁₂I-PL contains 12 ZFHD1 binding sites and a minimal human interleukin-2 gene promoter (Z₁₂I), upstream of a polylinker (PL) inserted downstream of an LTR-driven hygromycin resistance gene (Hygro^r). The Z₁₂I-PL cassette was isolated as an MluI-ClaI fragment from the pZ₁₂I-PL-2 target gene plasmid and inserted into pLH (2) such that transcription from the Z₁₂I promoter occurs in the same direction as LTR-driven transcription.
- Insertion of the gene of interest into the polylinker places its expression under control of the dimerizer-regulated transcription factors.
- We have found that the Z₁₂I promoter exhibits very low basal expression in retrovirally transduced cells.

[Annotated Sequence](#)

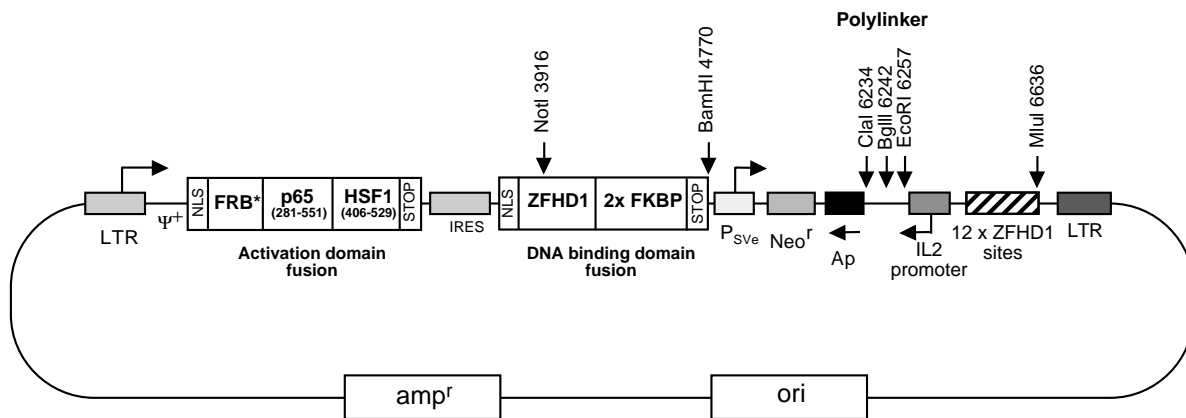
[Raw sequence](#)

All-in-one vector

pL₆N₂-R_HS3H/ZF2-PL

Description

pL₆N₂-R_HS3H/ZF2-PL (9577 bp)



Not drawn to scale

- pL₆N₂-R_HS3H/ZF2-PL allows delivery of transcription factors and a regulatable target gene in the same retroviral vector. It is essentially a modified version of pL₂N₂-R_HS3H/ZF3 in which a target gene can be placed under the control of a Z₁₂I promoter inserted between the *neo* gene and 3'LTR. The transcription factor cassette differs from that present in pL₂N₂-R_HS3H/ZF3 in that the DNA binding domain fusion contains two, rather than three FKBP repeats. The number of FKBP repeats was reduced to minimize the size and complexity of the retroviral insert without significantly affecting performance in cell types tested to date.
- Target genes are inserted into a polylinker flanked by the Z₁₂I promoter and the SV40 polyadenylation sequence (pA). An internal polyadenylation sequence is required because transcription from the Z₁₂I promoter is in the opposite direction to that of viral genomic RNA. The empty target gene expression cassette (isolated as a MluI - BamHI fragment from pZ₁₂I-PL-2 – see the ARGENT™ Regulated Transcription Plasmid Kit) is located between the *neo* gene and 3'LTR. Despite the proximity of the Z₁₂I promoter to the viral LTR, target gene expression in the absence of dimerizer is undetectable (see below). It is possible that this reflects a certain level of antisense inhibition of basal target gene expression due to the vector configuration.

[Annotated Sequence](#)

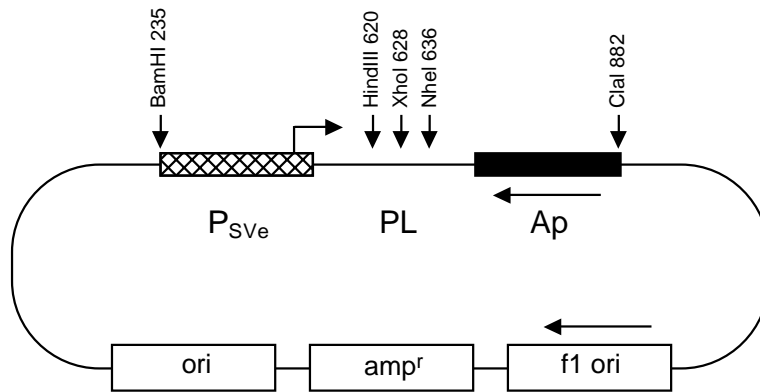
[Raw sequence](#)

Shuttle vector

pBS-SV-PL-Ap

Description

pBS-SV-PL-Ap (3545 bp)



Not drawn to scale

- pBS-SV-PL-Ap allows different selectable markers to be inserted into the all-in-one vector pL₂N₂-R_HS3H/ZF3 to replace the *neo* gene. The coding region for the new selectable marker (e.g. *hygro*^r, *zeocin*^r, CD8 or EFGP) is cloned into one of the sites in the polylinker (PL) so that its expression is driven by the SV40 enhancer (P_{SVe}). The BamHI-ClaI fragment containing the SV40 enhancer, selectable marker and polyadenylation sequence for the target gene (Ap) is then inserted into pL₂N₂-R_HS3H/ZF3 to replace the corresponding portion of that vector.

[Annotated Sequence](#)
[Raw sequence](#)

General vector information

Which vectors to use?

The ability to deliver the entire system in a single step whilst maintaining tight, highly inducible target gene regulation means that pL₆N₂-R_HS3H/ZF2-PL will often be the retroviral vector of choice. However, pL₆N₂-R_HS3H/ZF2-PL based retroviruses are large and relatively complex and this results in titers that are 100 fold lower than those of LXS_N, the vector from which they are derived. If the proposed target gene is relatively large (>2kb), or if high viral titer is an important consideration, it may be preferable to use the two virus system in which the rapamycin-responsive transcription factors and the regulatable target gene are delivered separately. We have successfully isolated inducible clones by sequential infections with pLH-Z₁₂I-PL and pL₂N₂-R_HS3H/ZF3 based retroviruses followed by double selection for the *hygro* and *neo* markers (see below).

Generation of a human growth hormone control vector

To generate vectors containing a human growth hormone (hGH) target gene, replace the MluI-ClaI fragment from pLH-Z₁₂I-PL or pL₆N₂-R_HS3H/ZF2-PL with the MluI-ClaI fragment from the plasmid

pZ₁₂I-hGH-2 (from the [ARGENT Regulation Transcription Plasmid Kit](#)). Several kits are available for the quantitation of hGH levels, including an RIA from Nichols Diagnostic (# 40-2155) and an ELISA from Roche (# 1 585 878).

Retrovirus Production

Supernatants containing infectious retroviral particles can be generated by transient transfection of retroviral plasmids into packaging cell lines (7). We have successfully used the Phoenix Ampho cell line (available from [Garry Nolan, Stanford University](#)). Packaging cell lines are also available from commercial sources such as Clontech.

In our hands, transient packaging of pL₂N₂-R_HS3H/ZF3 and pLH-Z₁₂I-PL based plasmids gives retroviral supernatants with titers of 10⁴ cfu/ml (approximately 10 fold lower than that of L₆ packaged under similar conditions). Titers of retroviral supernatants produced using pL₆N₂-R_HS3H/ZF2-PL based vectors are 10³ cfu/ml (100 fold lower than that of L₆). It is possible that titers can be improved by optimization of the transient packaging protocol or by the generation of stable producer clones.

Generation of inducible clones

G418-resistant clones can be picked for expansion approximately 10 days following infection with L₆-based retroviruses. Of these, typically half give rise to highly inducible target gene expression with negligible basal transcription (see below).

When using the two virus system, clones stably transduced with each retrovirus can be generated sequentially. For example,

1. Infect cells with pL₂N₂-R_HS3H/ZF3 to stably integrate the transcription factor expression cassette.
2. Screen individual clones by transiently transfecting the target gene of interest or an easily assayed target gene (e.g. pZ₁₂I-hGH-2 from the [ARGENT Regulated Transcription Plasmid Kit](#)).
3. Select the clone with lowest background and highest AP21967-dependent induction.
4. Infect cells with the pLH-Z₁₂I-PL based retrovirus containing the target gene.
5. Screen individual clones for the lowest background and highest levels of AP21967-dependent target gene expression.

Alternatively, a faster approach is to perform successive infections with pL₂N₂-R_HS3H/ZF3- and pLH-Z₁₂I-PL-derived retroviruses followed by double selection for hygromycin B and G418 resistance. In our experience, between 10 and 15% of the doubly transduced clones isolated in this way exhibit strongly inducible target gene expression.

Creating an all-in-one vector in a new vector backbone

For certain purposes it may be desirable to clone the transcription factor fusion proteins and target gene cassette into a new vector backbone. This is best done in a sequential procedure that involves first inserting a polylinker into the new vector backbone, followed by the transcription factor and target gene cassettes (\pm the selectable marker cassette). For example, to create a vector similar to pL₆N₂-R_HS3H/ZF2-PL in a new vector backbone:

1. Insert a polylinker containing EcoRI, BamHI and MluI sites into the new vector backbone, oriented so that the EcoRI site is immediately downstream of the promoter provided by the vector.
2. Insert the transcription factor cassette from pL₂N₂-R_HS3H/ZF3 as an EcoRI-BamHI fragment.
3. Clone the target gene into pL₆N₂-R_HS3H/ZF2-PL.
4. Insert the MluI-BamHI fragment containing the target gene cassette and the selectable marker cassette into the vector already containing the transcription factor cassette.

Note that the strategy described above creates a slightly larger vector than pL₆N₂-R_HS3H/ZF2-PL since the DNA binding domain in pL₂N₂-R_HS3H/ZF3 is fused to 3 FKBP domains rather than 2 FKBP domains as in pL₆N₂-R_HS3H/ZF2-PL. If size constraints require the use of the smaller configuration,

then a 2 domain version can easily be created by swapping the NotI-BamHI fragments of the two vectors.

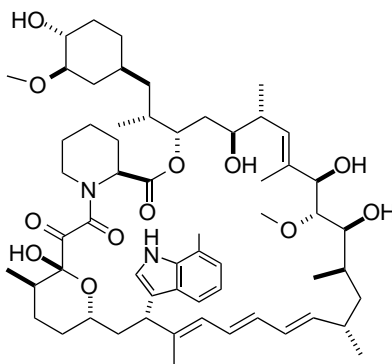
It may be necessary to make adjustments to this cloning strategy, if, for example, a downstream polyadenylation sequence is not provided by the vector or if a selectable marker is not desired. Note, however, that if adjustments are made, the optimal configuration of the components may need to be reevaluated. For example, we have found that when the selectable marker cassette is omitted, the preferred configuration is to have the target gene cassette in the same orientation as the transcription factor cassette. Such a vector has the same performance as one in which the target gene is in the opposite orientation but can be produced at significantly higher titer.

Antibodies to detect fusion proteins

Anti-p65 antibodies (Santa Cruz Biotechnology #sc-372) can be used to detect the R_HS3H (~72 kDa) activation domain fusion protein.

AP21967

Description



AP21967

AP21967 is a chemically modified derivative of rapamycin that can be used to induce heterodimerization of FKBP and FRB_{T2098L}-containing fusion proteins. AP21967 is greater than 1000-fold less immunosuppressive than rapamycin as measured in an *in vitro* splenocyte proliferation assay. In all studies to date, AP21967 is non-toxic to cells at up to 1 μ M concentrations, or mice at up to 30 mg/kg doses.

To date, AP21967 has only been tested *in vitro* and in mice. We do not yet know whether it crosses the blood-brain barrier in mice or whether it works in yeast or any other model organisms.

AP21967 cannot be used to heterodimerize proteins containing a wild type FRB domain. If you have already made constructs using the wild type FRB domain, you must use rapamycin as the heterodimerizer.

Note, however, that the presence of the T2098L mutation in FRB has little or no detrimental effect on the binding of rapamycin. Therefore, as noted earlier, rapamycin can also be used to dimerize fusion proteins made using the reagents in this kit. Rapamycin is available commercially from Sigma (cat # R0395) or Affinity BioReagents (cat # IR-022).

Reconstituting AP21967

AP21967 (molecular mass 1017.4 Da) is provided in lyophilized form which should be reconstituted as a concentrated stock in an organic solvent. We recommend dissolving the lyophilized material in absolute ethanol to make a 1 mM solution (e.g. dissolve 250 µg AP21967 in 246 µl ethanol). After adding the appropriate volume of ice-cold ethanol, seal and vortex periodically over a period of a few minutes to dissolve the compound. Keep on ice during dissolution to minimize evaporation.

Storage and handling of AP21967

Once dissolved, the stock solution can be kept at -20°C indefinitely, in a glass vial or a microfuge tube. Further dilutions in ethanol can be similarly stored. At the bench, solutions in ethanol should always be kept on ice, and opened for as short a time as possible, to prevent evaporation and consequent changes in concentration.

Using AP21967 *in vitro*

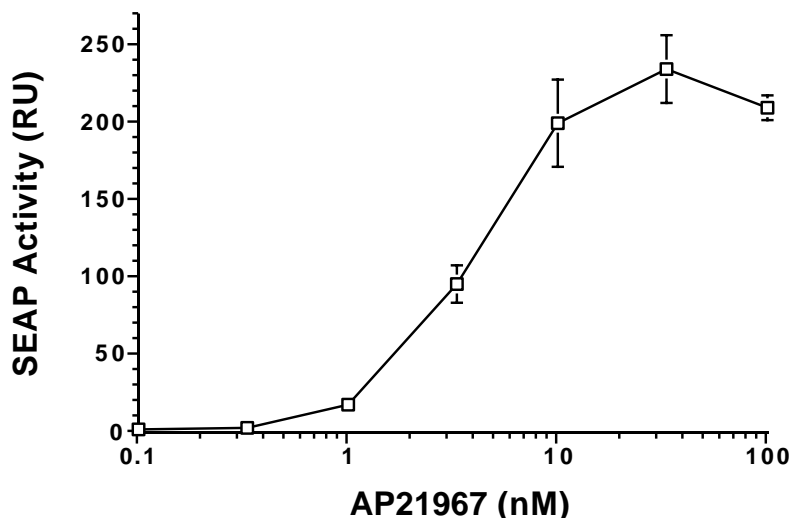
Working concentrations of AP21967 can be obtained by adding compound directly from ethanol stocks, or by diluting serially in culture medium just before use. In the latter case we recommend that the highest concentration does not exceed 5 µM, to ensure complete solubility in the (aqueous) medium. In either case, the final concentration of ethanol in the medium added to mammalian cells should be kept below 0.5% (a 200-fold dilution of a 100% ethanol solution) to prevent detrimental effects of the solvent on the cells.

Use of AP21967 in animals

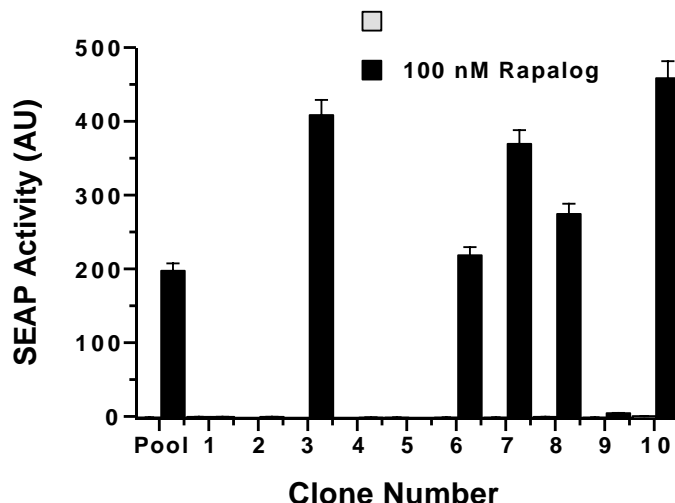
Once preliminary *in vitro* experiments have been carried out successfully we will be happy to provide quantities of AP21967 necessary for use in animals.

Expected results

The figure below shows the effects upon reporter gene expression of adding increasing amounts of AP21967 to a pool of HT1080 cells (ATCC CCL-121) stably transduced with an all-in-one retrovirus bearing a secreted alkaline phosphatase (SEAP) target gene. In the absence of dimerizer, target gene expression is undetectable. Half-maximal induction occurs at ~4 nM and peak induction at 10-30 nM AP21967. In initial experiments we recommend that AP21967 be tested across a broad range of concentrations (e.g. 0.1 to 100 nM) to provide a complete dose-response profile.



In a separate experiment, 10 G418-resistant clones were isolated from a pool of cells transduced with an all-in-one retroviral vector. Each clone was assayed for SEAP production alongside the parental pool in the presence or absence of 100 nM rapalog. The results (below) reveal that half the clones displayed inductions greater than or equal to the parental pool. Inducible clones can therefore be isolated from retrovirally transduced cells at a high frequency.



We have successfully demonstrated inducible regulation in a number of cell lines, including NIH 3T3 (murine fibroblasts), Ba/F3 (murine pre-B lymphocytes), C2C12 (murine myoblasts), and HT1080 (human fibrosarcoma cells) (7). In the presence of dimerizer, target gene expression is typically induced at least 1000-fold — to absolute levels similar to those seen when expression of the same target gene is driven directly from the retroviral LTR. The extremely low basal level of target gene expression is illustrated by the ability to make stable cell lines that inducibly express the highly toxic diphtheria toxin A chain gene (7). Using all-in-one vectors, stably transduced pools and clones exhibiting tightly regulated, highly inducible target gene expression can typically be generated in two to three weeks.

We anticipate that the performance of the system in various cell types will correlate with LTR activity since the level of transcription factor expression appears to be an important factor in the level of transcription that can be achieved. Although licensing restrictions prevent us from providing vectors bearing an enhanced Green Fluorescent Protein (EGFP) reporter gene, we have found this control to be particularly convenient for rapidly assessing the performance of the system in new cell types. Plasmids containing the EGFP coding sequence can be obtained commercially from Clontech.

Conditions of use

Please bear in mind that these materials will be provided to you pursuant to a Material Transfer Agreement (MTA). Our MTA contains, among other provisions, certain restrictions on the transfer to others of our materials and any derivatives you create using or incorporating our materials. If you wish to share the materials or derivatives with colleagues or collaborators, they must first complete our MTA. Please also be aware that our Kits are not to be used in research funded by, or conducted on behalf of, a commercial or for-profit entity. Those situations require a [commercial agreement](#).

We certainly hope that you obtain interesting results and that they are presented and published without delay. But please note that under the terms of the MTA, you need to give us advance notice of any such presentations or publications, including talks, posters, and submissions of abstracts or manuscripts for publication. Also, in the event of a patent filing, a copy of the patent application must be provided to ARIAD. Advance notice is usually 4 weeks prior to submission, but please check your MTA for specific details.

Please also be aware that the use of intellectual property or materials of others, in conjunction with the Regulation Kit, may have additional ramifications. For example, if you plan to use a Regulation Kit together with human embryonic stem cells from WiCell (WARF), we and you are required to execute an additional MTA which will be provided to you.

We appreciate your cooperation in this regard.

References

References cited here are listed below. A complete list of articles that have used dimerizers to regulate transcription can be found in the [Regulation Kits Bibliography](#).

1. Spencer, D. M., Wandless, T. J., Schreiber, S. L. & Crabtree, G. R. (1993) [Controlling signal transduction with synthetic ligands](#). *Science* 262: 1019-24.
2. Rivera, V. M., Clackson, T., Natesan, S., Pollock, R., Amara, J. F., Keenan, T., Magari, S. R., Phillips, T., Courage, N. L., Cerasoli, F., Jr., Holt, D. A. & Gilman, M. (1996) [A humanized system for pharmacologic control of gene expression](#). *Nat Med* 2: 1028-32.
3. Ho, S. N., Biggar, S. R., Spencer, D. M., Schreiber, S. L. & Crabtree, G. R. (1996) [Dimeric ligands define a role for transcriptional activation domains in reinitiation](#). *Nature* 382: 822-6.
4. Liberles, S. D., Diver, S. T., Austin, D. J. & Schreiber, S. L. (1997) [Inducible gene expression and protein translocation using nontoxic ligands identified by a mammalian three-hybrid screen](#). *Proc Natl Acad Sci U S A* 94: 7825-30.
5. Rivera, V. M., Ye, X., Courage, N. L., Sachar, J., Cerasoli, F., Jr., Wilson, J. M. & Gilman, M. (1999) [Long-term regulated expression of growth hormone in mice after intramuscular gene transfer](#). *Proc Natl Acad Sci U S A* 96: 8657-62.
6. Ye, X., Rivera, V. M., Zoltick, P., Cerasoli, F., Jr., Schnell, M. A., Gao, G., Hughes, J. V., Gilman, M. & Wilson, J. M. (1999) [Regulated delivery of therapeutic proteins after in vivo somatic cell gene transfer](#). *Science* 283: 88-91.
7. Pollock, R., Issner, R., Zoller, K., Natesan, S., Rivera, V. M. & Clackson, T. (2000) [Delivery of a stringent dimerizer-regulated gene expression system in a single retroviral vector](#). *Proc Natl Acad Sci U S A* 97: 13221-6.
8. Belshaw, P. J., Ho, S. N., Crabtree, G. R. & Schreiber, S. L. (1996) [Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins](#). *Proc Natl Acad Sci U S A* 93: 4604-7.
9. Choi, J., Chen, J., Schreiber, S. L. & Clardy, J. (1996) [Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP](#). *Science* 273: 239-42.

10. Chen, J., Zheng, X. F., Brown, E. J. & Schreiber, S. L. (1995) [*Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue.*](#) Proc Natl Acad Sci U S A 92: 4947-51.
 11. Pomerantz, J. L., Sharp, P. A. & Pabo, C. O. (1995) [*Structure-based design of transcription factors.*](#) Science 267: 93-6.
 12. Schmitz, M. L. & Baeuerle, P. A. (1991) [*The p65 subunit is responsible for the strong transcription activating potential of NF-kappa B.*](#) Embo J 10: 3805-17.
 13. Amara, J. F., Clackson, T., Rivera, V. M., Guo, T., Keenan, T., Natesan, S., Pollock, R., Yang, W., Courage, N. L., Holt, D. A. & Gilman, M. (1997) [*A versatile synthetic dimerizer for the regulation of protein-protein interactions.*](#) Proc Natl Acad Sci U S A 94: 10618-23.
-

Appendix

pL₂N₂-R_HS3H/ZF3 Annotated Sequence

```

1 gaattgctagcaattgctagcaattgctagcaattcataccagatcaccgaaaactgtcctccaaatgtgtccccctcacactcccaaatcgcgggctt 100
101 ctgctcttagaccactctaccctattccccacactcaccggagccaaagccgcggcccttcctgcttttctgctttgaaagacccaccctgaggtggc 200
201 aagctagcttaagtaacgccactttgcaagcatggaaaatacataactgagaatagaaaagttcagatcaaggtcaggaacaagaacagctgaata 300
301 ccaaacaggatattctgtggaagcgttctcgtccccggctcaggcccaagaacagatgagacagctgagtgatgggccaacaggatattctgtggaagc 400
401 agttcctgccccggctcggggccaagaacagatggtccccagatgcggtccagccctcagcagtttctagtgaatcatcagatgtttccagggtgcccc 500
501 aggacctgaaaatgacctgtaccttattgaaactaaccaatcagttcgtcttcgcttctgttcgctcgttccgctctccagactcaataaaagacc 600
601 cacaaccctcactcggcgccagctctccgtagactgcgtgccccgggtaccgctattccccataaagcctcttctgctttgcatccgaatcgtgg 700
701 ctgctgttctctggagggtctcctctgagtgattgactaccacagcgggggtctttcatttgggggctcgtcgggatttggagaccctgccagg 800
801 gaccaccgaccaccaccgggaggttaagctggccagcaacttatctgtgtctgtccgattgtctagtgtctatgtttgatgttatgcccctgcgtctgta 900
901 ctagttagctaaactagctctgtatctggcgaccctgtggtggaactgacgagttctgaacccccggcgaaccctgggagacgtcccagggactttggg 1000
1001 ggccgtttttgtggccgacctgaggaagggagtcgagtggaatccgaccccgctcaggatattggttctggtaggagacgagaaacctaaacagttcc 1100
1101 cgctcctgctgaatttttctctcgggttggaaaccgaagccgcgctcttctgctgctgcagcgtgcagcatcgttctgtgttctctgctgactgt 1200
1201 gtttctgtatttctgtaaaataggccagactgttaccactcccttaagttttagcttaggtcactggaagatgctgagcggatcgtccacaaccag 1300
1301 tcgtagatgtcaagaagagacgttgggttaccttctgctctgcagaatggccaaccttaactcggatggccgagacggcaccttaaccgagacc 1400
1401 tcatcaccaggttaagatcaaggtcttttcaactggccgcgcatggacaccagaccaggtcccctacatcgtgacctgggaagccttggcttttgacc 1500
1501 ccctcctgggtcaagccctttgtacacccctaaagcctcctcctctccatccgcccctctctccccctgaaacctcctcgttgcaccctcct 1600
1601 cgatcctccttttatccagccctcactccttctctagggcggcaattccagaagccacc ATG GAC TAT CCT GCT GCC AAG AGG GTC AAG 1690
1 M D Y P A A K R V K 10
1691 TTG GAC TCT AGA ATC CTC TGG CAT GAG ATG TGG CAT GAA GGC CTG GAA GAG GCA TCT CGT TTG TAC TTT GGG GAA 1765
11 L D S R I L W H E M W H E G L E E A S R L Y F G E 35
1766 AGG AAC GTG AAA GGC ATG TTT GAG GTG CTG GAG CCC TTG CAT GCT ATG ATG GAA CGG GGC CCC CAG ACT CTG AAG 1840
36 R N V K G M F E V L E P L H A M M E R G P Q T L K 60
1841 GAA ACA TCC TTT AAT CAG GCC TAT GGT CGA GAT TTA ATG GAG GCC CAA GAG TGG TGC AGG AAG TAC ATG AAA TCA 1915
61 E T S F N Q A Y G R D L M E A Q E W C R K Y M K S 85
1916 GGG AAT GTC AAG GAC CTC CTC CAA GCC TGG GAC CTC TAT TAT CAT GTG TTC CGA CGA ATC TCA AAG ACT AGA AGT 1990
86 G N V K D L L Q A W D L Y Y H V F R R I S K T R S 110
1991 GAG CCC ATG GAA TTT CAG TAC CTG CCA GAT ACA GAC GAT CGT CAC CGG ATT GAG GAG AAA CGT AAA AGG ACA TAT 2065
111 E P M E F Q Y L P D T D D R H R I E E K R K R T Y 135
2066 GAG ACC TTC AAG AGC ATC ATG AAG AAG AGT CCT TTC AGC GGA CCC ACC GAC CCC CGG CCT CCA CCT CGA CGC ATT 2140
136 E T S F N Q I M K K S G P T D P R P P R R I 160
2141 GCT GTG CCT TCC CGC AGC TCA GCT TCT GTC CCC AAG CCA GCA CCC CAG CCC TAT CCC TTT ACG TCA TCC CTG AGC 2215
161 A V P S R S A S V P K P A P Q P Y P F T S S L S 185
2216 ACC ATC AAC TAT GAT GAG TTT CCC ACC ATG GTG TTT CCT TCT GGG CAG ATC AGC CAG GCC TCG GCC TTG GCC CCG 2290
186 T I N Y D E F P T M V F P S G Q I S Q A S A L A P 210
2291 GCC CCT CCC CAA GTC CTG CCC CAG GCT CCA GCC CCT GCC CCT GCT CCA GCC ATG GTA TCA GCT CTG GCC CAG GCC 2365
211 A P P Q V L P Q A P A P A P A P A M V S A L A Q A 235
2366 CCA GCC CCT GTC CCA GTC CTA GCC CCA GGC CCT CCT CAG GCT GTG GCC CCA CCT GCC ACC ACC CAG GCT 2440
236 P A P V P V L A P G P P P Q A V A P P A P K P T Q A 260
2441 GGG GAA GGA ACG CTG TCA GAG GCC CTG CTG CAG CTG CAG TTT GAT GAT GAA GAC CTG GGG GCC TTG CTT GGC AAC 2515
261 G E G T L S E A L L Q L Q F D D E D L G A L L G N 285
2516 AGC ACA GAC CCA GCT GTG TTC ACA GAC CTG GCA TCC GTC GAC AAC TCC GAG TTT CAG CAG CTG CTG AAC CAG GGC 2590
286 S T D P A V F T D L A S V D N S E F Q Q L L N Q G 310
2591 ATA CCT GTG GCC CCC CAC ACA ACT GAG CCC ATG CTG ATG GAG TAC CCT GAG GCT ATA ACT CGC CTA GTG ACA GGG 2665
311 I P V A P H T T E P M L M E Y P E A I T R L V T G 335
2666 GCC CAG AGG CCC CCC GAC CCA GCT CCT GCT CCA CTG GGG GCC CCG GGG CTC CCC AAT GGC CTC CTT TCA GGA GAT 2740
336 A Q R P P D P A P A P L G A P G L P N G L L S G D 360
2741 GAA GAC TTC TCC TCC ATT GCG GAC ATG GAC TTC TCA GCC CTG CTG AGT CAG ATC AGC TCC ACT AGA GGC TTC AGC 2815
361 E D F S S I A D M D F S A L L S Q I S S T R G F S 385
2816 GTG GAC ACC AGT GCC CTG CTG GAC CTG TTC AGC CCC TCG GTG ACC GTG CCC GAC ATG AGC CTG CCT GAC CTT GAC 2890
386 V D T L F S P S V T V P D M S L P D L D 410
2891 AGC AGC CTG GCC AGT ATC CAA GAG CTC CTG TCT CCC CAG GAG CCC CCC AGG CCT CCC GAG GCA GAG AAC AGC AGC 2965
411 S S L A S I Q E L L S P Q E P P R P P E A E N S S 435
2966 CCG GAT TCA GGG AAG CAG CTG GTG CAC TAC ACA GCG CAG CCG CTG TTC CTG CTG GAC CCC GGC TCC GTG GAC ACC 3040
436 P D S G K Q L V H Y T A Q P L F L L D P G S V D T 460
3041 GGG AGC AAC GAC CTG CCG GTG CTG TTT GAG CTG GGA GAG GGC TCC TAC TTC TCC GAA GGG GAC GGC TTC GCC GAG 3115
461 G S N D L P V L F E L G E G S Y F S E G D G F A E 485
3116 GAC CCC ACC ATC TCC CTG CTG ACA GGC TCG GAG CCT CCC AAA GCC AAG GAC CCC ACT GTC TCC TAA taggatctccgg 3193
486 D P T I S L L T G S E P P K A K D P T V S * 507
3194 ttatnttccaccatattgcccgtcttttggcaatgtgagggccggaaaacctggccctgtcttctgacgagcattcctaggggcttttccctctcgcca 3293

```

3294 aaggaatgcaaggtctgtgaatgctggaaggaagcagttcctctggaagcttcttgaagacaacaacgctctgtagcaccctttgcaggcagcggaa 3393
EMCV IRES

3394 cccccacctggcgacaggtgctctgctggccaaaagccacgtgtataagatacacctgcaaaagggcgacacccccagtgccacgttgtgagttggata 3493

3494 gttgtggaaagagtcaaatggctctctcaagcgtatcaacaaggggctgaaggtgcccagaaggtacccccattgtatgggatctgatctggggcctc 3593

3594 ggtgcacatgctttacatgtgttttagtcgaggttaaaaaacgtctagggccccccgaaccacgggggacgtggttttcctttgaaaaacagatgataatac 3693
 <-- *c-myc NLS* <-->

3694 c ATG GAC TAT CCT GCT GCC AAG AGG GTC AAG TTG GAC TCT AGA GAA GCG CCA TAT GCT TGC CCT GTC GAG TCC 3766
 1 M D Y P A A K R V K L D S R E R P Y A C C P V E S 24

3767 TGC GAT CGC CGC TTT TCT CGC TCG GAT GAG CTT ACC CGC CAT ATC CGC ATC CAC ACA GGC CAG AAG CCC TTC CAG 3841
 25 C D R R F S R S D E L T R H I R I H T G Q K P F Q 49

3842 TGT CGA ATC TGC ATG CGT AAC TTC AGT CGT AGT GAC CAC CTT ACC ACC CAC ATC CGC ACC CAC ACA GGC GGC GGC 3916
 50 C R I C M R N F S R S D H L T T H I R T H T G G G 74
 Not I

3917 GGC AGG AGG AAG AAA CGC ACC AGC ATA GAG ACC AAC ATC CGT GTG GCC TTA GAG AAG AGT TTC TTG GAG AAT CAA 3991
 75 R R R K K R T S I E T N I R V A L E K S F L E N Q 99
ZFHD1

3992 AAG CCT ACC TCG GAA GAG ATC ACT ATG ATT GCT GAT CAG CTC AAT ATG GAA AAA GAG GTG ATT CGT GTT TGG TTC 4066
 100 K P T S E E I T M I A D Q L N M E K E V I R V W F 124

4067 TGT AAC CGC CGC CAG AAA GAA AAA AGA ATC AAC ACT AGA GGA GTG CAG GTG GAA ACC ATC TCC CCG GGA GAC GGG 4141
 125 C N R R Q K E K R I N T R G V Q V E T I S P G D G 149
 --> <--

4142 CGC ACC TTC CCC AAG CGC GGC CAG ACC TGC GTG GTG CAC TAC ACC GGG ATG CTT GAA GAT GGA AAG AAA TTT GAT 4216
 150 R T F P K R G Q T C V V H Y T G M L E D G K K F D 174
FKBP

4217 TCC TCC CGG GAC AGA AAC AAG CCC TTT AAG TTT ATG CTA GGC AAG CAG GAG GTG ATC CGA GGC TGG GAA GAA GGG 4291
 175 S S R D R K P R K L G K Q E V I R G W E E G 199

4292 GTT GCC CAG ATG AGT GTG GGT CAG AGA GCC AAA CTG ACT ATA TCT CCA GAT TAT GCC TAT GGT GCC ACT GGG CAC 4366
 200 V A Q M S V G Q R A K L T I S P D Y A Y G A T G H 224

4367 CCA GGC ATC ATC CCA CCA CAT GCC ACT CTC GTC TTC GAT GTG GAG CTT CTA AAA CTG GAA ACT AGA GGC GTT CAG 4441
 225 P G I I P P H A T L V F D V E L L K L E T R G V Q 249
 --> <--

4442 GTG GAA ACC ATC AGT CCA GGG GAT GGC CGA ACT TTT CCA AAG AGA GGG CAG ACT TGC GTC GTG CAT TAT ACT GGT 4516
 250 V E T I S P G G R T F P K R G Q T C V V V H Y T G 274

4517 ATG CTG GAG GAT GGG AAA AAG TTC GAC TCT TCC AGA GAT CGG AAC AAA CCA TTC AAA TTC ATG CTC GGG AAA CAG 4591
 275 M L E D G K K F D S S R D R N K P F K F M L G K Q 299

4592 GAA GTT ATC CGC GGA TGG GAG GAG GGC GTG GGC CAG ATG TCC GTG GGC CAG CGC GCC AAG CTA ACC ATC TCC CCA 4666
 300 E V I R G W E E G V A Q M S V G Q R A K L T I S P 324
FKBP' (silent mutations)

4667 GAC TAC GCC TAC GGA GCC ACC GGA CAC CCC GGT ATC ATA CCC CCA CAC GCC ACC CTT GTG TTT GAC GTG GAA CTG 4741
 325 D Y A Y G E A T G H P I I P H A T L V F E L 349
 --> <--

4742 CTT AAG CTA GAG ACT AGA GGC GTG CAG GTC GAG ACC ATC AGC CCC GGC GAC GGC CGC ACC TTT CCC AAG AGA GGC 4816
 350 L K L E T R G V Q V E T I S P G D G R T F P K R G 374

4817 CAG ACT TGC GTG GTC CAC TAC ACC GGC ATG CTG GAG GAC GGC AAG AAG TTC GAC AGC AGC CGC GAC CGC AAC AAG 4891
 375 Q T C V V H Y T G M L E D G K K F D S S R D R N K 399
FKBP'' (silent mutations)

4892 CCC TTC AAG TTC ATG CTG GGC AAA CAG GAA GTG ATC CGC GGC TGG GAG GAA GGC GTG GCT CAG ATG AGC GTG GGG 4966
 400 P F K F M L G K Q E V I R G W E E G C V A Q 424

4967 CAG CGG GCC AAG CTG ACC ATC AGC CCC GAC TAT GCC TAC GGC GCC ACC GGC CAC CCC GGC ATC ATC CCC CCC CAC 5041
 425 Q R A A K L S P D Y A Y G A T G H P G I I P P H 449
 --> <--

5042 GCC ACC CTG GTG TTC GAC GTG GAG CTG CTG AAG CTG GAG ACT AGT TAA taaggtccggctgtggaatgtgtcagttagggt 5125
 450 A T L V F D V E L L K L E T S * 465
 BamHI <-->

5126 gtggaaagtcaccaggtctccacagcaggcagaagtagtgcagaagcatctcaatttagtcagcaaccaggtgtggaagtcaccaggtctccacagcagg 5225
SV40 early promoter

5226 cagaagtagtcaaaagcatgcatctcaatttagtcagcaaccatagtcggcgccctaaactccgcccactccgcccagttccgcccattct 5325

5326 ccgccccatggctgactaattttttttatattatgagagggcggcggcctctgagctattccagaagtagtgaggaggttttttggaggcc 5425
 --> <--

5426 taggcttttgcaaaagcgtgggctgaggtcgagggcggatctgatcaagagacaggatgaggatcggtttcgc ATG ATT GAA CAA GAT GGA 5516
 1 M I E Q D G 6

5517 TTG CAC GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG GCA CAA CAG ACA ATC GGC TGC 5591
 7 L H A G S P A A W V E R L F G Y D W A Q Q T I G C 31

5592 TCT GAT GCC GCC GTG TTC CGG CTG TCA GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC 5666
 32 S D A A V F R L S A Q G R P V L F V K T D L S G A 56

5667 CTG AAT GAA CTG CAG GAG GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG ACG GGC GTT CCT TGC GCA GCT GTG CTC 5741
 57 L N E L Q D E A A R L S W L A T T G V P C A A V L 81

5742 GAC GTT GTC ACT GAA GCG GGA AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA TCT CAC 5816
 82 D V V T E A G R D W L L L G E V P G Q D L L S S H 106

5817 CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC 5891
 107 L A P A E K V S I M A D A M R R L H T L D P A T C 131

5892 CCA TTC GAC CAC CAA GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT GTC GAT CAG GAT 5966
 132 P F D H Q A K H R I E R A R T R M E A G L V D Q D 156
NcoI

5967 GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC 6041
 157 D L D E E H Q G L A P A E L F A R L K A R M P D G 181

6042 GAG GAT CTC GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT GGC CGC TTT TCT GGA TTC 6116
 182 E D L V V T H G D A C L P N I M V E N G R F S G F 206

6117 ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG 6191
207 I D C G R L G V A D R Y Q D I A L A T R D I A E E 231
6192 CTT GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT CCC GAT TCG CAG CGC ATC GCC TTC 6266
232 L G G E W A D R F L V L Y G I A A P D S Q R I A F 256
--> <--
6267 TAT CGC CTT CTT GAC GAG TTC TTC TGA gctggactctggggttcgataaaaataaaagattttattagtctccagaaaagggggaatga 6357
257 Y R L L D E F F * 265
6358 aagacccccactgtaggtttggcaagctagcttaagtaacgcattttgcaaggcatggaaaaatacataactgagaatagagaagttcagatcaaggctc 6457
6458 aggaacagatggaacagctgaatatgggccaacaggatctctgtggttaagcagttcctgccccggctcagggccaagaacagatggaacagctgaatat 6557
6558 gggccaacaggatctctgtggttaagcagttcctgccccggctcagggccaagaacagatggtccccagatgcggtccagccctcagcagtttctagaga 6657
6658 accatcagatggttccagggtgccccaggacctgaaatgacctgtgccttatttgaactaaccaatcagttcgttctctcgcttctgttcgctcgttc 6757
6758 tgctccccgagctcaataaaagagcccacaaccctcactcggggcgccagtcctccgattgactgagtcgccccgggtaccggtgatccataaacctc 6857
--> *MoMLV Poly A signal*
6858 cttgcagttgcatccgacttgtggtctcgtgttccctgggaggggtctcctctgagtgattgactaccogtcagcgggggtctttcattctgcattaatg 6957
6958 aatcggccaacgcgcggggagagggcggtttgcgatttgggctcttccgcttctcctcgtcactgactogctgcgctcgggtcgttcggctcggcgagcg 7057
7058 gtatcagctcactcaaaggcggttaacaggttatccacagaatcaggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggccaggaacc 7157
7158 gtaaaaaggccggttctgctggcggtttttccataggctccgccccctgacgagcatcacaanaatcgacgctcaagttagaggtggcgaacccgcagc 7257
7258 actataagatacaggcggtttccccctggaagctccctcgtgctctcctggtccgacctgcccgttaccggatacctgtccgctttctccctctcg 7357
7358 ggaagcgtggcgctttctcatagctcagcgtgtaggtatctcagttcgggtgtaggtcgttccgctccaagctgggctgtgtgcagcaaccccccttcagc 7457
7458 ccgaccgctgcgcttccggttaactatcgtcttgagtcacaaccggtaagacacgacttatcgcactggcagcagccactggtaacaggattagcag 7557
7558 agcaggtatgtaggcgggtctacagagttctgaaagtggtggcctaactacggctacactagaaggacagatttggtatctgcgctctgctgaagcca 7657
7658 gttaccttcggaaaaagagttggtagctcttgatccggcaaaaccaccgctggtagcgggtgttttttggttgcaagcagcagattacgcgcagaa 7757
7758 aaaaaggatctcaagaagatcctttgatctttctcaggggtctgacgctcagtggaacgaaaactcagtttaagggattttggtcatgagattatcaaa 7857
7858 aaggatcttcacctagatccttttcggccggccgcaaatcaatctaaagtatatatagtaaaacttggctgacagttaccaatgcttaacagtgagg 7957
7958 cacctatctcagcagatctgtctatttcgttcacatagttgctgactccccgctcgtgtagataactacgatacgggaggggttaccatctggccccag 8057
8058 tgctgcaatgataccgagagaccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaact 8157
8158 ttatccgctccatccagctctattaattgttgccgggaagctagagtaagtagttcgcagttaatagtttgcgcaacggttggccattgctacaggca 8257
8258 tcgtggtgtcagcctcgtcgtttgggtatgggtcattcagctccgggtcccaacgatcaaggcgagttacatgatccccatggtgtgcaaaaaagcgg 8357
8358 tagctccttcggtcctccgatcgttctgcaagtaagttggccgaggttatcactcatggttatggcagcactgcataattctcttactgtcatgcca 8457
8458 tccgtaagatgctttctgtgactggtgagtagtcaaccaagtcattctgagaatagtgtagtgcggcagccgagttgctctgcccggcgctcaacacggg 8557
8558 ataataccgcccacatagcagaactttaaagtgctcatcattggaaaaagcttctcggggcgaaaaactcctcaaggatcttaccgctggtgagatccag 8657
8658 ttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgttctcgggtgagcaaaaacaggaaggcaaaatgcccgaaaa 8757
8758 aagggataaagggcgacacggaatgtgaaactcatactcttcttttcaatattatgaaagcatttatcagggttatgtctcatgagcggatata 8857
8858 tatttgatgtatttagaaaaataaacaataggggttcgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacatt 8957
8958 aacctataaaaataggcgtatcacgaggccctttcgtcttcaa 9000

pLH-Z₁₂I-PL Annotated Sequence

<--
1 ctgcagcctgaatatgggccaacaggatatctgtggttaagcagttcctgccccgctcagggccaaagacagatggaacagctgaatatgggccaaca 100
101 ggatatctgtggttaagcagttcctgccccgctcagggccaaagacagatggtccccagatcggtccagccctcagcagttcttagagaaccatcagat 200
201 gttccagggtgccccaggacctgaatgacctgtgccttatttgaactaaccaatcagttcgcttctcgtctctgttccgcccgtctctgctccccga 300
301 gctcaataaaagagcccaaacccctcactcggggcgccagtcctccgattgactgagtcgcccgggtaccctgtatccaataaacccctcttgagcttg 400
401 catccgacttgggtctcgtgttctcttgggaggtctcctctgagtgattgactaccctcagcgggggtctttcatttgggggctcgtccgggactcg 500
501 gagaccctgcccaggaccaccagaccaccaccgggaggttaagctggccagcaacttatctgtgtctgtccgattgtctagtgcttatgactgatttta 600
601 tgcgctcgtcggtagctagttagtaactagctctgtatctggcggaccctggtggaactgacgagttctgaacaccggccgcaaccctgggagacg 700
701 tcccagggacttggggccgttcttggggccgacctgaggaagggagtcgagtggaatccgaccocctcaggatattggttctggtaggagacgag 800
801 aacctaaacagttcccgctcctgtctgaatttttgccttgggttggaaacgaagccgcgctcttctgctgcagcatgcttctggttctctctgt 900
901 ctgactgtgttctgtatttctgtctgaaaattagggccagactgttaccactcccttaagtttgaccttagatcactggaagatgctgagcggctcgtc 1000
1001 acaaccagtcggtagatgtcaagaagagacgttgggttaccttctgctctgcagaatggccaacctttaacgtcggatggccgcgagacggcaccttaa 1100
1101 ccgagacctcatcaccaggttaagatcaaggtcttttccctggcccgatggacaccagaccaggtcccctacatcgtgacctgggaagccttggtct 1200
1201 tttgacccccctcctgggtcaagcccttggtaacccctaaagcctcgcctcctctcttccatccgcccgtctctccccctgaaacctctcttctga 1300
1301 ccccgcctcaatcctcctttatccagccctcactcctctctagggccggcccgatcagcttgccagctggggccctctctggtaaggttgggaagcc 1400
1401 ctgcagactaaacacgatgataataccatgaaaaagcctgaactcaccgcgacgtctgtcgcagaagtcttctgatgaaaagttcgacagcgtctccgac 1500
1501 ctg ATG CAG CTC TCG GAG GGC GAA GAA TCT CGT GCT TTC AGC TTC GAT GTA GGA GGG CGT GGA TAT GTC CTG CGG 1575
1 M Q L S E G E E S R A F S F D V G G R G Y V L R 24
1576 GTA AAT AGC TGC GCC GAT GGT TTC TAC AAA GAT CGT TAT GTT TAT CGG CAC TTT GCA TCG GCC GCG CTC CCG ATT 1650
25 V N S C A D G F Y K D R Y Y R H F A S A A L P I 49
1651 CCG GAA GTG CTT GAC ATT GGG GAA TTT AGC GAG AGC CTG ACC TAT TGC ATC TCC CGC CGT GCA CAG GGT GTC ACG 1725
50 P E V L D I G E F S E S L T Y C I S R R A Q G V T 74
1726 TTG CAA GAC CTG CCT GAA ACC GAA CTG CCC GCT GTT CTG CAG CCG GTC GCG GAG GCC ATG GAT GCG ATC GCT GCG 1800
75 L Q D L P E T E L P A V L Q P V A E A M D A I A A 99
1801 GCC GAT CTT AGC CAG ACG AGC GGG TTC GGC CCA TTC GGA CCG CAA GGA ATC GGT CAA TAC ACT ACA TGG CGT GAT 1875
100 A D L S Q T S G F G P F G P Q G I G Q Y T T W R D 124
1876 TTC ATA TGC GCG ATT GCT GAT CCC CAT GTG TAT CAC TGG CAA ACT GTG ATG GAC GAC ACC GTC AGT GCG TCC GTC 1950
125 F I C A I A D P H V Y H W Q T V M D D T V S A S V 149
1951 GCG CAG GCT CTC GAT GAG CTG ATG CTT TGG GCC GAG GAC TGC CCC GAA GTC CGG CAC CTC GTG CAC GCG GAT TTC 2025
150 A Q A L D E L M L W A E D C P E V R H L V H A D F 174
2026 GGC TCC AAC AAT GTC CTG ACG GAC AAT GGC CGC ATA ACA GCG GTC ATT GAC TGG AGC GAG GCG ATG TTC GGG GAT 2100
175 G S N N V L T D N G R I T A V I D W S E A M F G D 199
2101 TCC CAA TAC GAG GTC GCC AAC ATC TTC TTC TGG AGG CCG TGG TTG GCT TGT ATG GAG CAG CAG ACG CGC TAC TTC 2175
200 S Q Y E V A N I F F W R P W L A C M E Q Q T R Y F 224
2176 GAG CGG AGG CAT CCG GAG CTT GCA GGA TCG CCG CGG CTC CGG GCG TAT ATG CTC CGC ATT GGT CTT GAC CAA CTC 2250
225 E R R H P E L A G S P R L R A Y M L R I G L D Q L 249
2251 TAT CAG AGC TTG GTT GAC GGC AAT TTC GAT GAT GCA GCT TGG GCG CAG GGT CGA TGC GAC GCA ATC GTC CGA TCC 2325
250 Y Q S L V D G N F D A A W A Q G R C D A I V R S 274
2326 GGA GCC GGG ACT GTC GGG CGT ACA CAA ATC GCC CGC AGA AGC GCG GCC GTC TGG ACC GAT GGC TGT GTA GAA GTA 2400
275 G A G T V G R T Q I A R R S A A V W T D G C V E V 299
2401 CTC GCC GAT AGT GGA AAC CGA CGC CCC AGC ACT CGT CCG AGG GCA AAG GAA TAG agtagatgccagccgggatctatcgacc 2482
300 L A D S G N R R P S T R P R A K E * 317
2483 cgtgctagctaatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcgctcgcag 2582
2583 aatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcg 2682
2683 gctcagtaaatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcgctcgcagtaaatgatggcgctcgcag 2782
2783 caaggttccctatcactctctttaaactactcactcacagtaacctcaactcctgccacagcttgcctgcagcgggaattccactagtcgagatctcca 2882
Clai
2883 tccgataaaataaaagatttttattttagtctccagaaaaaggggggaatgaaagacccccacctgtaggtttggcaagctagcttaagtaacgccattttgca 2982
2983 aggcattgaaaaatacataactgagaatagagaagttcagatcaaggctcaggaacagatggaacagctgaaatgggccaacaggatatctgtggttaag 3082
3083 cagttcctgccccggctcagggccaaagacagatggaacagctgaaatgggccaacaggatatctgtggttaagcagttcctgccccggctcagggccaa 3182
3183 agaacagatgggtcccagatgagggtccagccctcagcagttctagagaaccatcagatggttccagggtgccccaggacctgaaatgacctgtgct 3282
3283 ttttgaactaaccaatcagttcgttctcgtctctgttccgcccgtctctgctccccagatcaataaaagagcccacaaccctcactcggggcgccag 3382
3383 tctcctgattgactgagtcgcccgggtaccctgtatcccaataaacctctctgagttgcatccgacttgggtctcgtctgttctctgggaggtctcct 3482
3483 ctgagtgattgactaccctcagcgggggtcttccatttccgacttgggtctcgtctccttgggaggtctcctctgagtgattgactaccctcagc 3582
3583 ggggtcttcacatgcagcagatgatacaaaataatttgggttttttcttaagtatttacattaaatggccatagttgcaataatgaaatcgcccaacgcgc 3682

3683 ggggagaggcgggtttcggtattggcgctcttccgcttcctcgcctcactgactcgctgcgctcggtcggttcggtcggcgagcgggtatcagctcactcaa 3782
3783 aggcggtaatacgggtatccacagaatcaggggataaacgaggaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcggtt 3882
3883 gctggcggtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcaagggtggcgaaaccgacaggactataaagataccag 3982
3983 gcgtttccccctggaagctccctcgtgcgctctcctggtccgacctgcgcttaccggataacctgctcgctttctcccttcgggaagcgtggcgcttt 4082
4083 ctcatagctcacgctgtaggtatctcagttccggtgtaggtcgcttccgctccaagctgggctgtgtgcacgaacccccggttcagcccaccgctgcgcctt 4182
4183 atccggtaactatcgtcttgagttccaaccoggtaaagacacgaactatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggc 4282
4283 ggtgctacagagttcttgaagtgggtggcctaactacggctacactagaaggacagtatgttggtatctgcgctctgctgaagccagttacctcggaaaaa 4382
4383 gagtggtagctcttgatccggcaaaacaccaccgctggtagcgggtggttttttggttgcaagcagcagattacgcgcagaaaaaaaggatctcaaga 4482
4483 agatcctttgatctttctacggggtcagcgtcagtggaacgaaaactcacgttaagggattttgggtcatgagattatcaaaaaggatcttcacctag 4582
4583 atccttttaaatataaaatgaagtttgcggccgcaaatcaatcctaaagtatatatgagtaaaacttggctgacagttaccaatgcttaatacagtgaggca 4682
4683 cctatctcagcgatctgtctatctcgttccatcagttgctgactccccgctcgtgtagataactacgatacgggagggcttaccatctggccccagtg 4782
4783 ctgcaatgataaccgcgagaccacgctcaccggctccagatttatcagcaataaacagccagccggaaggccgagcgcagaagtggtcctgcaacttt 4882
4883 atccgcctccatccagttctattaattggtgcccgggaagctagagtaagtagttcggcagttaatagtttgcgcaacgttgttgccattgctacaggcatc 4982
4983 gtgggtcacgctcgtcgtttgggtatggcttcaatcagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcgggta 5082
5083 gctccttcggctcctccgatcgttgtcagaagtaagttggccgcagtggtatcactcatggttatggcagcactgcataaattctcttactgtcatgccatc 5182
5183 cgtaagatgcttttctgtgactgggtgagtactcaaccaagtcattctgagaatagtgatgcggcgaccgagttgctcttgcggcgctcaacacgggat 5282
5283 aataccgcgccacatagcagaactttaaagtgctcatcattggaacacggttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagtt 5382
5383 cgatgtaaccctcgtgcaccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa 5482
5483 gggataaaggcgacacggaaatgtgaatactcatactcttcttttcaatattatgaagcatttatcagggttatgtctcatgagcggatatacata 5582
5583 tttgaatgtatttagaaaaataaacaatagggttccgcgcacatttc 5631

3194 ccggttattttccaccatattgccgtcttttggcaatgtgagggcccgaaacctggccctgtcttcttgacgagcattcctaggggtctttcccctctc 3293

3294 gccaaaggaatgcaaggtctgttgaatgtogtgaaggaagcagttctctggaagcttcttgaagacaacaacgtctgtagcagccctttgcaggcagc 3393

3394 ggaacccccacactggcgacaggtgacctctgcccgaagccacgtgtataagatacacctgcaaaagggcgacacaacccccagtgccacgtgtgagttg 3493

3494 gatagttgtggaagagtcaaatggctctcctcaacgctattcaacaaggggctgaaggatgccagaaggtacccccattgtatgggatctgatctgggg 3593

3594 cctcgtgacacatgctttacatgtgttttagtcgaggttaaaaaacgtctagggcccccaaacacggggacgtggttttcccttgaaaaacacgatgata 3693

3694 atacc ATG GAC TAT CCT GCT GGC AAG AGG GTC AAG TTG GAC TCT AGA GAA CGC CCA TAT GCT TGC CCT GTC GAG 3767
 1 M D Y P A A K R V K L D S R E R P Y A C P V E 23

3768 TCC TGC GAT CGC CGC TTT TCT CGC TCG GAT GAG CTT ACC CGC CAT ATC CGC ATC CAC ACA GGC CAG AAG CCC TTC 3842
 24 S C D R R F S R S D E L T R H I R I H T G Q K P F 48

3843 CAG TGT CGA ATC TGC ATG CGT AAC TTC AGT CGT AGT GAC CAC CTT ACC ACC CAC ATC CGC ACC CAC ACA GGC GGC 3917
 49 Q C R I C M R N F S R S D H L T T H I R T T G G 73

3918 GGC CGC AGG AGG AAG AAA CGC ACC AGC ATA GAG ACC AAC ATC CGT GTG GCC TTA GAG AAG AGT TTC TTG GAG AAT 3992
 74 G R R R K K R T S I E T N I R V A L E K S F L E N 98

3993 CAA AAG CCT ACC TCG GAA GAG ATC ACT ATG ATT GCT GAT CAG CTC AAT ATG GAA AAA GAG GTG ATT CGT GTT TGG 4067
 99 Q K P T S E E I T M I A D Q L N M E K E V I R V W 123

4068 TTC TGT AAC CGC CGC CAG AAA GAA AAA AGA ATC AAC ACT AGA GGA GTG CAG GTG GAA ACC ATC TCC CCG GGA GAC 4142
 124 F C N R R Q K E K R I N T R G V Q V E T I S P G D 148

4143 GGG CGC ACC TTC CCC AAG CGC GGC CAG ACC TGC GTG GTG CAC TAC ACC GGG ATG CTT GAA GAT GGA AAG AAA TTT 4217
 149 G R T F P P K R G Q T C V V H Y T G M L E D G K K F 173

4218 GAT TCC TCC CGG GAC AGA AAC AAG CCC TTT AAG TTT ATG CTA GGC AAG CAG GAG GTG ATC CGA GGC TGG GAA GAA 4292
 174 D S S R D R N K P F K F M L G K Q E V I R G W E E 198

4293 GGG GTT GCC CAG ATG AGT GTG GGT CAG AGA GCC AAA CTG ACT ATA TCT CCA GAT TAT GCC TAT GGT GCC ACT GGG 4367
 199 G V A Q M S V G Q R A K L T I S P D Y A Y G A T G 223

4368 CAC CCA GGC ATC ATC CCA CCA CAT GCC ACT CTC GTC TTC GAT VTG GAG CTT CTA AAA CTG GAA ACT AGA GGC GTG 4442
 224 H P G I I P H A T L V E L L K L E T R G V 248

4443 CAG GTC GAG ACC ATC AGC CCC GGC GAC GGC CGC ACC TTT CCC AAG AGA GGC CAG ACT TGC GTG GTC CAC TAC ACC 4517
 249 Q V E T I S P G D G R T F P K R G Q T C V V H Y T 273

4518 GGC ATG CTG GAG GAC GGC AAG AAG TTC GAC AGC AGC CGC GAC CGC AAC AAG CCC TTC AAG TTC ATG CTG GGC AAA 4592
 274 G M L E D G K K F D S S R D R N K P F K F M L G K 298

4593 CAG GAA GTG ATC CGC GGC TGG GAG GAA GGC GTG GCT CAG ATG AGC GTG GGG CAG CGG GCC AAG CTG ACC ATC AGC 4667
 299 Q E V I R G W E E G V A G M S V G Q R A K L T I S 323

4668 CCC GAC TAT GCC TAC GGC GCC ACC GGC CAC CCC GGC ATC ATC CCC CCC CAC GCC ACC CTG GTG TTC GAC GTG GAG 4742
 324 P D Y A Y G A T G H P G I I P P H A T L V F D V E 348

4743 CTG CTG AAG CTG GAG ACT AGT TAA taaggatccggctgtggaatgtgtgtcagttaggggtgtgaaagtccccaggctccccagcagcaga 4834
 349 L L K L E T S * 356

4835 agtatgcaaaagcatgcatctcaattagtcagcaaccaggtgtggaagtccccaggctccccagcagcagcagaagtatgcaaaagcatgcatctcaattagt 4934

4935 cagcaaccatagtcgccccctaaactccgccatcccgcccctaactccgccagttccgccattctccgccatggctgactaattttttttat 5034

5035 tgcaaggccgagggccgctcggcctctgagctattccagaagttagtgaggaggcttttttggaggcctaggttttgcataaagcttgggctcgaggtc 5134

5135 gaggcggatctgatcaagagacaggatgaggatcgttttcgc ATG ATT GAA CAA GAT GGA TTG CAC GCA GGT TCT CCG GCC GCT 5217
 1 M I E Q D G L H A G S P A A 14

5218 TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG 5292
 15 W V E R L F G Y D W A Q Q T I G C S D A A V F R L 39

5293 TCA GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC CTG AAT GAA CTG CAG GAC GAG GCA 5367
 40 S A Q G R P V L F V K T D L S G A L N E L Q D E A 64

5368 GCG CGG CTA TCG TGG CTG GCC ACG ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG GGA AGG 5442
 65 A R L S W L A T T G V P C A A V L D V V T E A G R 89

5443 GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC 5517
 90 D W L L L G E V P G Q D L L S S H L A P A E K V S 114

5518 ATC ATG GCT GAT GCA ATG CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA GCG AAA CAT 5592
 115 I M A D A M R R L H T L D P A T C P F D H Q A K H 139

5593 CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG 5667
 140 R I E R A R T R M E A G L V D Q D D L D E E H Q G 164

5668 CTC GCG CCA GCC GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC GTC GTG ACC CAT GGC 5742
 165 L A P A E L F A R L K A R M P D G E D L V V T H G 189

5743 GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG 5817
 190 D A C M V E N A G R F S G F I D C G R L G V 214

5818 GCG GAC CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT GGC GGC GAA TGG GCT GAC CGC 5892
 215 A D R Y Q D I A L A T R D I A E E L G G E W A D R 239

5893 TTC CTC GTG CTT TAC GGT ATC GCC GCT CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC 5967
 240 F L V L Y G I A A P D S Q R I A F Y R L L D E F F 264

5968 TGA gcgggactctggggttcgagatccgattttaccacattgtagaggttttacttgccttaaaaaacctccccacacctccccctgaacctgaaacat 6066
 265 * 265

Poly A signal

6067 aaaatgaatgcaattgttgttgaactgtttattgcagcttataatggttacaaataaagcaatagcatcacaatttcacaaataaagcattttttt 6166
SV40 late 3'UTR ←--- --> ClaI BglII EcoRI

6167 cactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtctcgaagcggccatcgatggagatctcgactagtggattccgc 6266
 <--- IL2 promoter

6267 tgcagggcaagcttgtggcaggagttgaggttactgtgagtagtgattaaagagagtgatagggaaactcttgaacaagagatgcaatttatactgttaat 6366
 --->

6367 tctggaaaaatattatgggggtgtcaaaatgttgtaaaatcgcgttctagacgcccattactcgagcggccatcatttagtcgaccgcccatt 6466

6467 actcgagcggccatcatttagctagacgcccattactcgagcggccatcatttagtcgagcggccatcatttagtcgagcggccatcatttagctagagc 6566
 12x ZFHD1 sites MluI

6567 ccatcattactcgagcggccatcatttagtcgagcggccatcatttagctagcagcgtcgaaatgaccgaccaagcgagcgc 6666

6667 aacctgccatcacgagatttcgattccacgcgcgcttctatgaaaggttgggcttcggaatcgtttccgggacgcggctggatgatcctccagcgcg 6766

6767 gggatctcatgctggagatcccgcgattagccaatttgttaaaagacaggatatacagtggtccaggctctagttttgactcaacaatatcaccagctgaa 6866
 <---

6867 gcctatagagtacgagccatagataaaaaaaagattttatttagtctccgaaaaaggggggaatgaaagaccccactgtaggtttggcaagctagct 6966

6967 taagtaacgccattttgcaagcagtgaaaaatacataactgagaatagagaagttcagatcaaggtcaggaacagatggaacagctgaatatgggcca 7066
 MoMLV 3' UTR

7067 acaggatatactgtggaagcagttcctgccccggctcagggccaagaacagatggaacagctgaatatgggccaacaggatatactgtggaagcagttc 7166

7167 ctgccccggctcagggccaagaacagatggcctccagatgcccgtccagcctcagcagtttctagagaacctcagatgtttccaggggtccccaaggac 7266
 MoMLV Poly A signal

7267 ctgaaatgacctgtgcttatttgaactaaccaatcagttcgttctcgttctgctcgcgcgcttctgctccccgagctcaataaaagagcccacaac 7366

7367 ccctcactcggggcgcagctcctccgattgactgagtcgccccgggtaccctgtgatccaataaacctcttgcagttgcatccgacttgggtctcgcgtg 7466
 --->

7467 ttccctgggaggggtctcctctgagtgatgactaccctcagcgggggtcttctcattctgcattaatgaatcggccaacgcgcggggagagcgggttgc 7566

7567 gtattgggcgctcttccgcttccctcgtcactgactcgtcgcgtcggctcgttcggctcggcgagcgggtatcagctcaaaagcggtaatacgggt 7666

7667 atccacagaatcaggggataacgcaggaagaacatgtgagcaaaagggccagcaaaagggcaggaaacctgaaaaagggcgcggttgcggcttttccat 7766

7767 aggctccgccccctgacgagcatcacaataatcgacgctcaagttagaggtggcgaaccgcagcagactataagataaccagcgtttccccctggaa 7866

7867 gtcctcctcgtcgcctctcctgttccgacctgccgcttaccggataacctgtccgcttctcccttcgggaagcgtggcgcttctcatagctcacgctg 7966

7967 taggtatctcagttcgggtgtaggtcgttccgctccaagctgggctgtgcaacgaacccccgttcagcccagcctcgcgcttaccggttaactatcgt 8066

8067 cttgagtcacaaccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcaggtatgtaggcgggtctacagagttct 8166

8167 tgaagtgggtggcctaactacggctacactagaaggacagatttggatctcgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttg 8266

8267 atccggcaaaaccaccgctggtagcgggtgttttttgttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttt 8366

8367 tctacggggtctgacgctcagtggaacgaaaactcacgttaagggatttggctcagagattatcaaaaaggatcttccactagatccttttgcggccgg 8466

8467 ccgcaaatcaatctaaagtataatagtaaaacttggctcagacgttaccaatgcttaatcagtgaggcaacctatctcagcagatctgtctatttcgttca 8566

8567 tccatagttgcctgactccccctcgtgtagataaactacgatacgggagggcttaccatctggccccagtgctgcaatgataaccgcgagaccacgctcac 8666

8667 cggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggctcctgcaacttatccgctccatccagctctataattgttg 8766

8767 ccgggaagctagagtaagtagttcgcagtttaagtttgcgcaacgttggccattgtctacagcagcagcgtggtgtcagcctcgtcgtttggtatggct 8866

8867 tcattcagctccgggtoccaacgatcaagcaggttacatgatccccatggtgtgcaaaaaagcgggttagctccttcggctcctcagatcgttgcagaa 8966

8967 gtaagtggccgagtggtatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgctttctgtgactggtgagta 9066

9067 ctcaaccaagtcatcttgagaatagtgtatgcggcgaccgagttgctcttgcggcgctcaaacgggataataaccgcgccatagcagaacttataaa 9166

9167 gtgctcatcattggaaaacgcttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcagatgtaaccactcgtgcacccaactgat 9266

9267 cttcagcatcttttactttccaccagcgttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaaggggaataaggcgcacagcgaatgtgaa 9366

9367 actcatactcttcttttcaatattatgaagcatttatcaggggtattgtctcatgagcggatataatattgaaatgatttagaaaaataaacaata 9466

9467 ggggtccgcgacatttccccgaaaagtgccacctgacgtctaagaaaccattatcatgacattaacctataaaaataggcgtatcacgagccct 9566

9567 ttcgtcttcaa 9577

pBS-SV-PL-Ap Annotated Sequence

```

1  gactggaagcgggcagtgagcgcaacgcaattaatgtgagtttagctcactcattagggcaccocaggctttacactttatgcttcgggctcgtatgtgt 100
101 gtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaaagctcgaaattaaccctcactaaaggggaacaaaagctgga 200
201 gctccaccgcggtggcggccgctctagaactagtBamHIggatccc<--cgggctgcaggaattaatcgccagcacagtggtcgacctgtggaatgtgtgtcagtt 300
301 aggggtgtggaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccagggtgtggaagtccccaggctcccca 400
401 gcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcSV40 early promoterccgcccctaactccgcccataactccgcccagttccgccc 500
501 attctccgcccattggctgactaatttttttattttatgacagagggc▶gaggccgctcggcctctgagctattccagaagtagtgaggaggcttttttgg 600
601 aggcctaggcttttgcaaaa-->agctHindIIIgtXhoIctNheIcgagc<--agctagcPoly A signalcgattctaccattttttagtagggttttacttgctttaaaaaacctccacacctcccc 700
701 tgaacctgaaacataaaatgaatgcaattggtgtgttaactgtttattgcagcttataatggttacaataaagcaatagcatcacaatttcacaaa 800
801 taaagcattttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgctcgaagcggccSV40 late 3'UTRatcga←atgggtacccaattcg 900

```