

ARGENT™
Regulated Heterodimerization Kit

Version 2.0

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

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

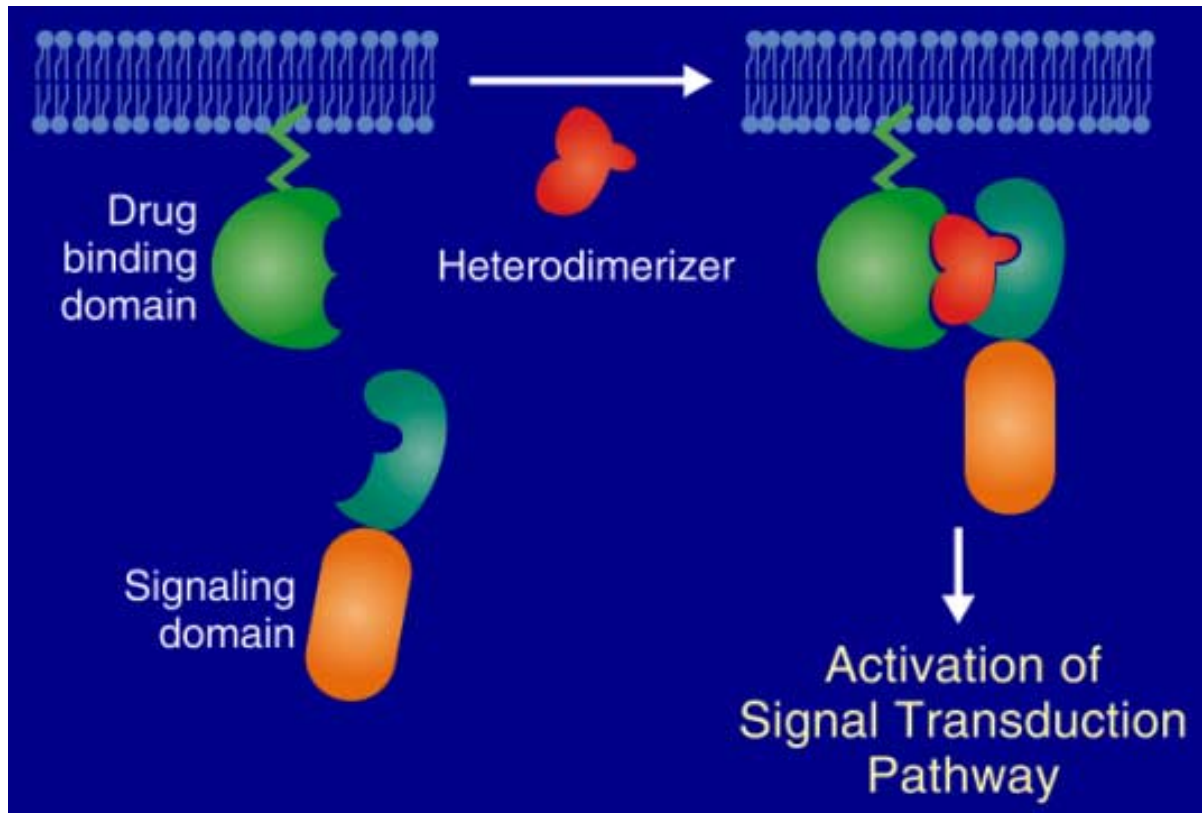
09/09/02

Table of Contents

ARGENT™ Regulated Heterodimerization Kit.....	1
Table of Contents	2
Overview	3
Applications of the ARGENT Heterodimerization Kit.....	4
Design of the kit components	4
Notes on the use of this kit	5
Kit contents	5
Expression plasmids.....	6
pC ₄ EN-F1	6
pC ₄ M-F2E	6
pC ₄ -R _H E.....	7
General vector information.....	8
Creating fusion proteins.....	8
Controlling the localization of fusion proteins.....	8
How many FKBP and FRB domains should I use?.....	8
Antibodies to detect fusion proteins.....	8
Addition of a selectable marker	8
Additional pC ₄ expression vector information	9
AP21967	9
Description.....	9
Reconstituting AP21967.....	10
Storage and handling of AP21967	10
Using AP21967 <i>in vitro</i>	10
Use of AP21967 in animals.....	10
Expected results	10
Conditions of use	11
References	11
Appendix.....	12
pC ₄ EN-F1E Annotated Sequence.....	13
pC ₄ M-F2E Annotated Sequence	14
pC ₄ -R _H E Annotated Sequence.....	15

ARGENT™ Regulated Heterodimerization Kit

The ARGENT™ Regulated Heterodimerization Kit contains reagents for bringing together two different engineered fusion proteins by adding a small molecule "dimerizer". The kit can be used to create conditional alleles of receptors, signaling molecules, and any other protein normally regulated by interactions between two different proteins, allowing complex cellular events to be brought under small molecule control.



[Controlling Signal Transduction Using Regulated Heterodimerization](#)

Overview

Many cellular processes are triggered by the induced interaction, or "dimerization", of signaling proteins (1). Examples include the clustering of cell surface receptors by extracellular growth factors, and the subsequent stepwise recruitment and activation of intracellular signaling proteins. A chemical inducer of dimerization, or "dimerizer", is a cell-permeant organic molecule with two separate motifs that each bind with high affinity to a specific protein module. Any cellular process activated by protein-protein interactions can in principle be brought under dimerizer control, by fusing the protein(s) of interest to the binding protein(s) recognized by the dimerizer. Addition of the dimerizer then links together the chimeric signaling protein(s), activating the cellular event that it controls (see the figure).

There are two classes of dimerizers. **Homodimerizers** incorporate two identical binding motifs, and can therefore be used to induce self-association of a single signaling domain. **Heterodimerizers** (as shown in the figure) have two different binding motifs, allowing the specific dimerization of two

different signaling domains when they are fused to the two appropriate ligand binding domains. The ARGENT™ Regulated Heterodimerization Kit described here provides a heterodimerizer and DNA vectors for making appropriate fusion proteins. For applications requiring homodimerization, we offer a separate regulation kit that includes a homodimerizer (see [ARGENT Regulated Homodimerization Kit](#)).

Applications of the ARGENT Heterodimerization Kit

The ARGENT Heterodimerization Kit, like the Homodimerization Kit, can be used to control any signaling process that involves regulated protein-protein interactions. The Heterodimerization Kit is particularly useful for creating specific interactions between two different proteins, in cases where the directionality of dimerization is critical. Published examples of such applications include:

- Inducible activation of heterodimeric cell surface receptors or other signaling proteins (2, 3);
- Inducible activation or inactivation of signaling proteins by recruiting them to new subcellular locations, by driving their interaction with appropriately anchored fusion proteins (e.g. those in the plasma membrane or nucleus) (4-9); and
- Inducible reconstitution of the activity of a bifunctional protein, in which two functional domains have been dissociated (10).

These studies have allowed a wide variety of cellular processes to be chemically controlled, including proliferation, differentiation and apoptosis. A complete list of publications describing the use of heterodimerizing reagents can be found in the [Regulation Kit Bibliography](#).

Regulated dimerization, using either of the ARGENT kits, has applications in many areas of functional genomics research and drug discovery. Inducible alleles of orphan receptors or other signaling proteins can be created with no knowledge of the natural ligand. These systems can be used for functional analysis of the signaling pathway in multiple cell types, potentially identifying downstream target proteins, or genes whose expression is modulated by the signaling event. Inducible animal models can be established of disease states associated with an activated signaling protein. In addition, cell lines in which a specific signal can be chemically induced may be useful in the configuration of targeted cell-based assays for small molecule drugs.

The combined use of the ARGENT Homodimerization and Heterodimerization Kits is particularly powerful for analysis of complex families of receptors or signaling proteins, in which multiple combinations can occur. The kit components can be used to specifically induce predefined homotypic or heterotypic combinations to probe their cellular and physiologic function, allowing complex issues such as receptor crosstalk to be addressed. Because no knowledge of natural receptor ligands is required, this approach is equally applicable to families of orphan receptors. Such studies provide a means to bypass the broad activity of natural ligands and uniquely activate a single receptor complex. For example, this strategy has been used to dissect the relative roles of the ErbB1 and ErbB2 receptors in EGF ligand family signaling (3).

The reagents of the ARGENT Heterodimerization Kit can be used for both *in vitro* and *in vivo* studies. The dimerizer provided with the kit, AP21967, is suitable for *in vivo* use and has been used successfully in mice. In addition, the plasmid vectors provided with the kit have been adapted for delivery by retroviral vectors (3).

Design of the kit components

The reagents in the ARGENT Regulated Heterodimerization Kit, like those of the Homodimerization Kit, are based on the human protein FKBP12 (FKBP, for FK506 binding protein) and its small molecule ligands. FKBP is an abundant cytoplasmic protein that serves as the initial intracellular target for the natural product immunosuppressive drugs FK506 and rapamycin. Both these drugs naturally act as heterodimerizers, and both have been used as the basis for heterodimerization systems (11, 12), as has FK-CsA, a cyclosporin-FK506 hybrid molecule (4). We have focused on the

use of rapamycin, because it has well understood chemistry and has favorable pharmacokinetic properties in mammals. 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 (13). To use rapamycin to induce heterodimers between proteins of interest, one of the proteins is fused to FKBP domains, and the other to a 93 amino acid portion of FRAP, termed FRB, that is sufficient for binding the FKBP-rapamycin complex (14).

In some cases, the use of rapamycin as a heterodimerizing reagent 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 used to make protein fusions allows a rapalog to be used to specifically heterodimerize engineered proteins without interfering with the activity of endogenous FRAP.

This redesigned rapamycin system forms the basis of the ARGENT Heterodimerization Kit, which provides constructs containing FKBP and the mutant FRB, and a non-immunosuppressive rapalog called AP21967. These and related reagents have been used to control the localization and activity of signaling domains as described above (3, 5, 8, 9). It is important to note that the redesigned system retains the ability to respond to rapamycin itself, as well as AP21967. Therefore experiments can be carried out with either ligand, as appropriate.

Notes on the use of this kit

Use of the Homodimerization Kit to induce heterodimers

In addition to the specific Heterodimerization Kit described here, it is also possible to use the [ARGENT Regulated Homodimerization Kit](#) to induce heterodimerization, by fusing two different signaling proteins to the same ligand binding domain. Addition of the homodimerizer creates a mixture of homodimeric and heterodimeric complexes. However, to induce heterodimers specifically and exclusively, the ARGENT Regulated Heterodimerization Kit described here should be used.

Use to control transcription

A particular case of the use of controlled heterodimerization to reconstitute protein function is in the regulation of transcription factor activity. We have developed a generic system for regulating the transcription of any target gene that makes use of the heterodimerizer system. For these reagents please see the [ARGENT Regulated Transcription Kits](#) (see also the list of [publications](#) describing the use of the Regulated Transcription system).

Kit contents

To control the activity or localization of signaling domains, one of the domains of interest is fused to one or more copies of an FKBP domain and the other to a mutant FRB domain. This allows the dimerization state to be controlled by administration of the rapalog AP21967.

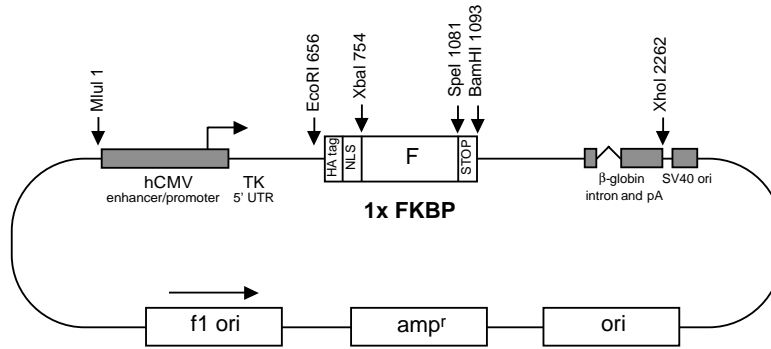
The ARGENT Regulated Heterodimerization Kit contains three plasmids, pC₄EN-F1, pC₄M-F2E, pC₄-R_HE, and an aliquot of AP21967. As described below, the plasmids in this kit provide an assortment of components (i.e. mutant FRB domain, multiple FKBP domains, an epitope tag, localization sequences) that can be easily manipulated to generate protein fusions whose activity and localization can be controlled by dimerizer.

Expression plasmids

pC₄EN-F1

Description

pC₄EN-F1 (5410 bp)



Not drawn to scale

- In pC₄EN-F1, a chimeric fusion protein containing an amino terminal epitope tag (E, from the influenza hemagglutinin [HA] gene) and nuclear localization signal (N, from SV40 large T antigen), followed by a single copy of FKBP12 (F1), is expressed under control of the human CMV enhancer/promoter (C).
- The FKBP domain is flanked by XbaI and SpeI sites. To fuse the protein of interest to a single FKBP domain clone it into the adjacent XbaI or SpeI sites as described below.
- Fusion proteins will be targeted to the nucleus due to the presence of the amino-terminal nuclear localization signal.

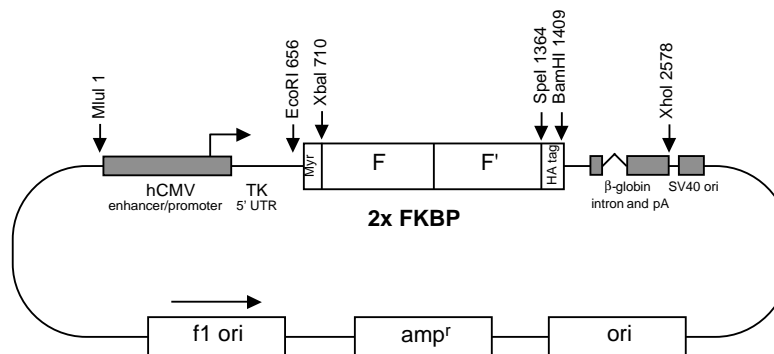
[Annotated Sequence](#)

[Raw sequence](#)

pC₄M-F2E

Description

pC₄M-F2E (5726 bp)



Not drawn to scale

- In pC₄M-F2E, a chimeric fusion protein containing an amino-terminal myristoylation signal (M), two copies of FKBP (F2), followed by a carboxy-terminal epitope tag (E, from the influenza hemagglutinin [HA] gene) is expressed under control of the human CMV enhancer/promoter (C).
- The two FKBP domains are flanked by XbaI and SpeI sites. To fuse the protein of interest to two FKBP domains clone it into the adjacent XbaI or SpeI sites as described below.
- One of the FKBP domains has changes in the codons used that do not change the amino acid sequence, but which dramatically reduce the match between the FKBP domains at the nucleotide level. We have found that this “wobble” reduces potential for recombination (i.e. making these constructs fit for use in retroviral vectors).
- Due to the presence of the amino-terminal myristoylation signal, fusion proteins will be targeted to the cytoplasmic face of membranes.

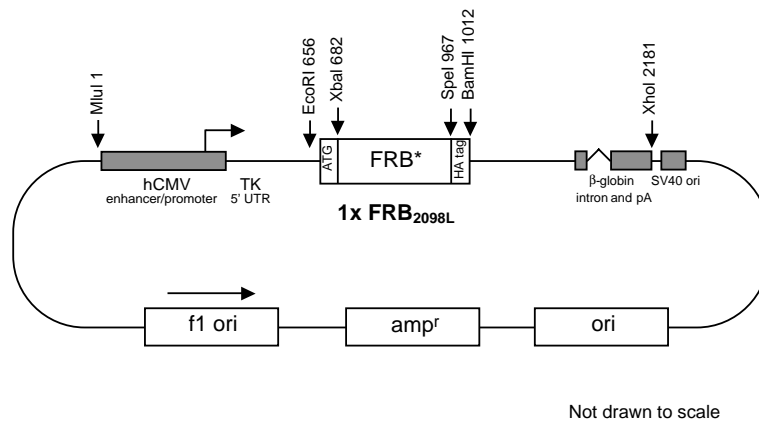
[Annotated Sequence](#)

[Raw sequence](#)

pC₄-R_HE

Description

pC₄-R_HE (5329 bp)



- In pC₄-R_HE, a chimeric fusion protein containing a single copy of the modified FRB (R_H), followed by a carboxy-terminal epitope tag (E, from the influenza hemagglutinin [HA] gene) is expressed under control of the human CMV enhancer/promoter (C).
- R_H consists of amino acids 2021-2113 of human FRAP in which the threonine at amino acid 2098 was mutated to leucine, to accommodate the chemical substitution that blocks AP21967 binding to wild type FRAP.
- The R_H domain is flanked by XbaI and SpeI sites. To fuse the protein of interest to a single R_H domain clone it into the adjacent XbaI or SpeI sites as described below.
- Unless the domain fused to R_H contains a signal that targets it to another location, fusion proteins should be localized to the cytoplasm by default as there is no targeting signal in this vector (the amino terminus of this fusion protein, upstream of the XbaI site, consists only of a methionine).

[Annotated Sequence](#)

[Raw sequence](#)

General vector information

Creating fusion proteins

The creation of fusion proteins using vectors in any of the ARIAD Regulation Kits is based on a standard cloning strategy involving the stepwise addition of compatible XbaI-SpeI fragments. To do this, amplify the coding sequence of interest by PCR so that it contains the six nucleotides specifying an XbaI site immediately 5' to the first codon (take care not to create an overlapping Dam methylation sequence, GATC, on either strand), and the six nucleotides specifying a SpeI site immediately 3' to the last codon. Then, for example, to fuse the protein of interest amino terminal to 2 FKBP_s, clone the XbaI-SpeI fragment into the XbaI site of pC₄M-F2E (XbaI and SpeI have compatible cohesive ends). If inserted in the proper orientation, the XbaI and SpeI sites, now flanking the new fusion protein, will be maintained, with the junction of the two peptides consisting of the two amino acids specified by the SpeI and XbaI sites that were fused. Or to fuse the XbaI-SpeI fragment carboxy-terminal to 2 FKBP_s, insert it into the SpeI site of pC₄M-F2E. In both cases, since the flanking XbaI and SpeI sites are maintained, additional fragments can still be fused at the amino- and carboxy- terminal ends if desired.

If the sequence to be fused contains internal XbaI or SpeI sites, fusions can still be made either by using XbaI or SpeI at both ends, or by using NheI or AvrII which also generate ends that are compatible with XbaI and SpeI. Note, though, that in these cases unique flanking XbaI and SpeI sites will not be regenerated.

The sequence between the SpeI and BamHI sites of pC₄EN-F1, pC₄M-F2E and pC₄-R_HE contains an in frame stop codon (in some cases preceded by an HA epitope tag). Therefore, stop codons should not be included in the fused sequences.

Controlling the localization of fusion proteins

The region between the EcoRI and XbaI sites in all pC₄- vectors contains a leader sequence which determines the default localization of fusion proteins. In pC₄EN-F1, pC₄M-F2E and pC₄-R_HE the default localization is to the nucleus, inner face of plasma membranes and cytoplasm, respectively. Fusion proteins can be cloned as XbaI-SpeI or XbaI-BamHI fragments (note the location of the epitope tag) into the appropriate vector to target them to the desired location.

How many FKBP and FRB domains should I use?

The number of FKBP and FRB domains best suited for each application varies. We have generally found that fusing 1 FKBP domain and one FRB domain to each signaling protein works well, although in some cases using 2 FKBP domains was preferable.

Antibodies to detect fusion proteins

Anti-HA (Babco #MMS101R-500) and anti-FKBP12 (Affinity Bioreagents #PA1-026) antibodies are available commercially. Each FKBP domain is ~12 kDa.

Addition of a selectable marker

Stable integration of plasmid vectors into cells is greatly facilitated by co-expression of the fusion protein of interest and the selectable marker gene on the same mRNA transcript. Such bicistronic mRNAs can be created by expressing the coding region of the fusion protein downstream of an enhancer/promoter and upstream of an internal ribosome entry sequence (IRES) which drives expression of a selectable marker gene. Using "pIRES-" vectors available from Clontech (e.g. pIRESneo2, pIREShyg2, pIRESpuro2, pIRESbleo), two alternative approaches can be used to generate such a vector.

- The BamHI-XhoI fragment from one of the vectors in this kit, can be replaced by a BamHI-XhoI fragment from a pIRES vector. This replaces the intron and polyA sequence from the vector in this kit with an intron, an IRES element, the selectable marker gene and a polyA sequence.

- Alternatively, an EcoRI-BamHI fragment containing the protein fusion of interest can be cloned into the MCS of a pIRES vector.

The advantage of using this approach to introduce a selectable marker is that essentially all drug resistant cells will express the protein of interest, since any mRNA that expresses the selectable marker as its second cistron should also express the protein of interest as its first cistron.

Additional pC₄ expression vector information

Origin of vector

pC₄ expression vectors are derived from the vector pCGNN (15). To create pC₄, several existing restriction sites were eliminated and several others added in order to have all functional regions of the plasmid be flanked by unique restriction sites (i.e. MluI, EcoRI, XbaI, SpeI, BamHI and XhoI).

Configuration of vectors

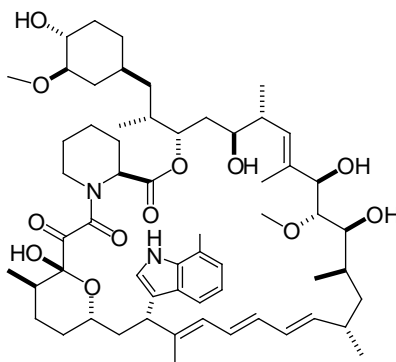
- **MluI-EcoRI:** Contains the enhancer/promoter from human CMV and the 5' UTR from the Herpes simplex virus TK gene. This fragment can be replaced with an alternate enhancer/promoter of choice (i.e. to promote tissue-specific expression)
- **EcoRI-XbaI-SpeI-BamHI:** Contains the coding region broken down as follows:
 - *EcoRI-XbaI:* leader sequence (e.g. a myristoylation sequence or simply an ATG).
 - *XbaI-SpeI:* ligand binding domains and signaling domain.
 - *SpeI-BamHI:* Carboxy-terminal sequence. In addition to the stop codon may contain an epitope tag.
- **BamHI to XhoI:** contains an 830-bp portion of the 3'UTR of the rabbit β -globin gene that includes its final intron and poly(A) signal.

Production of single stranded DNA for mutagenesis

All pC₄ plasmids contain f1 origins for rescue of single stranded DNA. The strand generated upon rescue is indicated by the arrow in plasmid maps. For example, in pC₄EN-F1 the antisense strand is generated, therefore oligonucleotides used for mutagenesis should correspond to the sense stand of the vector (the strand shown in the vector sequences).

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 μM 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

Peak effects of induced heterodimerization of two proteins using rapamycin or AP21967 are generally seen at concentrations of 50-100 and 100-500 nM, respectively (3, 8, unpublished data).

In initial experiments we recommend that AP21967 be tested across a broad range of concentrations (e.g. 0.5 nM to 500 nM) to provide a complete dose-response profile.

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 heterodimerizers can be found in the [Regulation Kits Bibliography](#).

1. Klemm, J. D., Schreiber, S. L. & Crabtree, G. R. (1998) [Dimerization as a regulatory mechanism in signal transduction](#). Annu Rev Immunol 16: 569-92.
2. Stockwell, B. R. & Schreiber, S. L. (1998) [Probing the role of homomeric and heteromeric receptor interactions in TGF-beta signaling using small molecule dimerizers](#). Curr Biol 8: 761-70.
3. Muthuswamy, S. K., Gilman, M. & Brugge, J. S. (1999) [Controlled dimerization of ErbB receptors provides evidence for differential signaling by homo- and heterodimers](#). Mol Cell Biol 19: 6845-57.
4. 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.
5. Graef, I. A., Holsinger, L. J., Diver, S., Schreiber, S. L. & Crabtree, G. R. (1997) [Proximity and orientation underlie signaling by the non-receptor tyrosine kinase ZAP70](#). Embo J 16: 5618-28.
6. Klemm, J. D., Beals, C. R. & Crabtree, G. R. (1997) [Rapid targeting of nuclear proteins to the cytoplasm](#). Curr Biol 7: 638-44.
7. 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.
8. Castellano, F., Montcourrier, P., Guillemot, J. C., Gouin, E., Machesky, L., Cossart, P. & Chavrier, P. (1999) [Inducible recruitment of Cdc42 or WASP to a cell-surface receptor triggers actin polymerization and filopodium formation](#). Curr Biol 9: 351-60.

9. Castellano, F., Montcourrier, P. & Chavrier, P. (2000) [***Membrane recruitment of rac1 triggers phagocytosis.***](#) J Cell Sci 113: 2955-61.
 10. Rossi, F., Charlton, C. A. & Blau, H. M. (1997) [***Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation.***](#) Proc Natl Acad Sci U S A 94: 8405-10.
 11. 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.
 12. 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.
 13. 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.
 14. 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.
 15. Attar, R. M. & Gilman, M. Z. (1992) [***Expression cloning of a novel zinc-finger protein that binds to the c-fos serum response element.***](#) Mol. Cell. Biol. 12: 2432-2443.
-

Appendix

pC₄EN-F1E Annotated Sequence

```

MluI <--
1  accgcttcgagctcgccccgttacataacttacggtaaatggccccgctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatggt 100
101 cccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaacgcccacttggcagtagacatcaagtgtatcatatgcca 200
201 gtacgccccctattgacgtcaatgacggtaaatggccccgctggcattatgcccagtagacattatgggactttcctacttggcagtagacatctacgt 300
301 attagtcatcgctattaccatggatgacggttttggcagtagacatcaatggcgtagcggtttgactcacgggatttccaagtctccaccccatt 400
401 gacgtcaatgggagtttggcaccacaaatcaacgggactttccaaaatgctgtaaacactccgccccattgacgcaaatggcggtaggcgtgtac 500
501 ggtgggaggtctatataagcagagctcgttttagtgaaccgtagatcgctggagacgccatccacgtgttttgacctccatagaagacaccgggaccg 600
601 atccagcctccgggggatcttgggtggcgtgaaactcccgcacctcttcggccagcaaatgcagaagcgcgt ATG GCT TCT AGC TAT CCT TAT 693
1  --> <-- HSV TK 5' UTR --> EcoRI <--
1  HA epitope tag --> <-- SV40 NLS --> XbaI <--
694 GAC GTG CCT GAC TAT GCC AGC CTG GGA GGA CCT TCT AGT CCT AAG AAG AAG AGA AAG GTG TCT AGA GGA GTG CAG 768
8  D V P D Y A S L G G P S S P K K K R K V S R G V Q 32
769 GTG GAA ACC ATC TCC CCA GGA GAC GGG CGC ACC TTC CCC AAG CGC GGC CAG ACC TGC GTG GTG CAC TAC ACC GGG 843
33 V E T I S P G D G R T F P K R G G Q T C V V H Y T G 57
844 ATG CTT GAA GAT GGA AAG AAA TTT GAT TCC TCC CGG GAC AGA AAC AAG CCC TTT AAG TTT ATG CTA GGC AAG CAG 918
58 M L E D G K K F D S R D R N K P F K F M L G K Q 82
919 GAG GTG ATC CGA GGC TGG GAA GAA GGG GTT GCC CAG ATG AGT GTG GGT CAG AGA GCC AAA CTG ACT ATA TCT CCA 993
83 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 107
994 GAT TAT GCC TAT GGT GCC ACT GGG CAC CCA GGC ATC ATC CCA CCA CAT GCC ACT CTC GTC TTC GAT GTG GAG CTT 1068
108 D Y A Y G A T G H P G I I P P H A T L V F D V E L 132
1069 CTA AAA CTG GAA ACT AGT TAT TAA ggatcctggagaacttcagggtgagtttggggacccttgattgttctttctttctgctattgtaaaat 1160
133 L K L E T S Y *
1161 tcatgttataatggaggggcaaaagttttcagggtgtgtttagaatgggaagatgctccctgtatcaccatggaccctcatgataaattttgtttcttca 1260
1261 ctttctactctgttgacaaccattgtctcctctattttctttcattttctgtaactttttcgtaaacttttagcttgcatgtgtaacgaatttttaa 1360
1361 ttcaactttgtttatttgcagattgtaagtactttctctaatcacttttttcaaggcaatcagggatattatattgtacttcagcacagttttaga 1460
1461 gaacaattgttataatataatgataaggtagaatattttctgcatataaattctggctggcgtggaaatattcttattggtagaacaactacatcctggt 1560
1561 catcatcctgctctttctctttatggttacaatgatatacactgtttgagatgaggataaaaactctgagtcacaacccgggcccctctgctaaccatggt 1660
1661 catgcctctctttttctctacagctcctgggcaacgtgctggttgggtgctgtctcatcattttggcaaaggattcactcctcaggtgcaggtgct 1760
1761 atcagaaggtggtggctggtggtgccaatgccctggctcacaataccactgagatcttttccctctgccaaaaatattggggacatcatgaagcccct 1860
1861 tgagcatctgactctctgctataaaggaaatttatttctgcaatagtgtgttggaaatttttgtgtctctcactcggaaggacatattggagggca 1960
1961 aatcatttaaacatcagaatgagattttggttttagagtttggcaacatagccatgctggtgcccataaacaagggtggctataaagaggtcatcag 2060
2061 tatatgaaacagccccctgctgtccattccttattccatagaaaagccttgacttgaggttagatttttttatatttggtttgggttattttttctt 2160
2161 taacatccctaaaattttccttacatgttttactagccagatttttccctcctcctgactactcccagtcatagtctgctcctctctcttatgaagatc 2260
2261 ctcagaggagctttttgcaaaagcctaggcctccaaaaagcctcttcaactcttctggaatagctcagaggccgagggcctcggcctctgcataaa 2360

```

pC₄M-F2E Annotated Sequence

```

MluI <--
1  acgcgttccgagctcgccttacataactacggtaaatggccgcctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatggt 100
101 cccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaacgcccacttggcagtagacatcaagtgtatcatatgcca 200
201 gtacgccccctattgacgtcaatgacggtaaatggccgcctggcattatgcccagtagacatgaccttatgggactttcctacttggcagtagacatcactgt 300
301 attagtcacgtcgtattaccatggatgacgggtttggcagtagacatcaatggcgtggatagcgggtttgactcacggggatttccaagtctccccccatt 400
401 gacgtcaatgggagtttgggtttggcaccacaaatcaacgggactttccaaaatgctgtaacaaactcgcgccattgacgcaaatgggcggtaggcgtgtac 500
501 ggtgggaggtctatataagcagagctcgttttagtgaaccgtcagatgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccg 600
601 atccagcctccgggggatcttgggtggcgtgaaactcccgcacctcttggccagcgaattcgcgcgt ATG GGG AGT AGC AAG AGC AAG CCT 691
1 M G S S K S K P 8
--> XbaI <--
692 AAG GAC CCC AGC CAG CGC TCT AGA GGA GTG CAG GTG GAA ACC ATC TCC CCA GGA GAC GGG CGC ACC TTC CCC AAG 766
9 K D P S Q R S R G V Q V E T I S P G D G R T F P K 33
767 CGC GGC CAG ACC TGC GTG GTG CAC TAC ACC GGG ATG CTT GAA GAT GGA AAG AAA TTT GAT TCC TCC CGG GAC AGA 841
34 R E Q T C V V H Y T G M L E D G K K F D S S R D R 58
842 AAC AAG CCC TTT AAG TTT ATG CTA GGC AAG CAG GAG GTG ATC CGA GGC TGG GAA GAA GGG GTT GCC CAG ATG AGT 916
59 N K P F K F M L G K Q E V I R G W E E G V A Q M S 83
917 GTG GGT CAG AGA GCC AAA CTG ACT ATA TCT CCA GAT TAT GCC TAT GGT GCC ACT GGG CAC CCA GGC ATC ATC CCA 991
84 V G Q R A K L T I S P D Y A Y G A T G H P G I I P 108
992 CCA CAT GCC ACT CTC GTC TTC GAT GTG GAG CTT CTA AAA CTG GAA ACT AGA GGC GTG CAG GTC GAG ACC ATC AGC 1066
109 P H A T L V F D V E L L K L E T R G V Q V E T I S 133
1067 CCC GGC GAC GGC CGC ACC TTT CCC AAG AGA GGC CAG ACT TGC GTG GTC CAC TAC ACC GGC ATG CTG GAG GAC GGC 1141
134 P G D G R T F P K R G Q T C V V H Y T G M L E D G 158
1142 AAG AAG TTC GAC AGC AGC CGC GAC CGC AAC AAG CCC TTC AAG TTC ATG CTG GGC AAA CAG GAA GTG ATC CGC GGC 1216
159 K K F D S S R D R N K P F K F M L G K Q E V I R G 183
1217 TGG GAG GAA GGC GTG GCT CAG ATG AGC GTG GGG CAG CGG GCC AAG CTG ACC ATC AGC CCC GAC TAT GCC TAC GGC 1291
184 W E E G V A Q M S V G Q R A K L T I S P D Y A Y G 208
1292 GCC ACC GGC CAC CCC GGC ATC ATC CCC CCC CAC GCC ACC CTG GTG TTC GAC GTG GAG CTG CTG AAG CTG GAG ACT 1366
209 A T G H P G I I P P H A T L V F D V E L L K L E T 233
1367 AGT TAT CCG TAC GAC GTA CCA GAC TAC GCA TAA gaaaagtgaggatcctgagaacttcagggtgagtttggggacccttgattgttctt 1455
234 S Y P Y D V P D Y A * 244
1456 tctttttcgctattgtaaaattcatgttatatggagggggcaagttttcagggtgtgttttagaatgggaagatgtcccttgatcaccatggaccctc 1555
1556 atgataattttgtttctttcactttctactctgttgacaaccattgtctctcttattttcttttcatttttctgtaactttttctgtaaaccttagcttg 1655
1656 cttttgtaacgaatttttaaattcactttttgtttattttgtcagattgtaagtacttttctaatcactttttttcaaggcaatcagggtatattatatt 1755
1756 gtacttcagcacagtttttagagaacaattgttataataataatgataaggtagaatatttctgcatataaattctggctggcgtggaaatattcttattgg 1855
1856 tagaaacaactacatcctggatcatcctgcctttctctttatggttacaatgatatacactgtttgagatgaggataaaatactctgagtcctcaaacg 1955
1956 ggccccctctgtaaacatgttcatgctctcttttttctacagctcctgggcaactgctggtgtgtgtgctgtctcatcttttggcaaaaggattca 2055
2056 ctctcaggtgacaggctgctatcagaaggtggctggctggccaatgccctggctgcaaaataaccatgagatctttttccctctgccccaaaatta 2155
2156 tggggacatcatgaagcccttgagcatctgacttctggctataaaagaaatttattttcatgcaatagtggttgaattttttgtgtctctcactc 2255
2256 ggaaggacatattgggagggcaaatcatttaaaacatcagaatgagttatgtgttagagtttggcaacatagccatagctggtgcatgaacaaagg 2355
2356 tggctataaagaggtcatcagtatatgaaacagccccctgctgctcattcttattccatagaaaagccttgacttgaggttagatttttttatatttt 2455
2456 gttttgtgttattttttcttaaacatccctaaaattttccttacatgttttactagccagatttttctcctcctcctgactactcccagtcagtagctgt 2555
2556 ccctcttctctatgaagatccctcgaggagcctttttgcaaaagccctaggcctccaaaaagcctcttactacttctggaatagctcagaggccgagg 2655

```

pC₄-R_HE Annotated Sequence

```

MluI <--
1  acgcgttcgagctcgcgcccttacataacttacggtaaatggccccctggctgaccgcccaacgacccccgccattgacgtcaataatgacgtatggt 100
101 cccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaacgcccacttggcagtagacatcaagtgtatcatatgcca 200
201 gtacgccccctattgacgtcaatgacggtaaatggccccctggcattatgccagtagacatgacctatgggactttcctacttggcagtagacatctacgt 300
301 attagtcacgctattaccatgggtgatcgcggttttggcagtagacatcaatggcgtagcgggtttgactcacggggattccaagtctccccccatt 400
401 gacgtcaatgggagtttgttttggcaccacaaatcaacgggactttccaaaatgctcgttaacaactccgccccattgacgcaaatgggcgtaggcgtgtac 500
501 ggtgggaggtctatataagcagagctcgtttagtgaaccgtcagatcgctggagacgccatccacgctgtttgacctccatagaagacaccgggaccg 600
601 atccagcctccgggggatcttgggtggcgtgaaactcccgcacctcttggccagcgaattccagaagccgccacc  ATG  GCT  TCT AGA  <-- XbaI
1  M  A  S  R  I  L  6
694 TGG CAT GAG ATG TGG CAT GAA GGC CTG GAA GAG GCA TCT CGT TTG TAC TTT GGG GAA AGG AAC GTG AAA GGC ATG 768
7  W  H  E  M  W  H  E  G  L  E  E  A  S  R  L  Y  F  G  E  R  N  V  K  G  M  31
769 TTT GAG GTG CTG GAG CCC TTG CAT GCT ATG ATG GAA CGG GGC CCC CAG ACT CTG AAG GAA ACA TCC TTT AAT CAG 843
32 F  E  V  L  E  P  L  H  A  M  M  E  R  G  P  Q  T  L  K  E  T  S  F  N  Q  56
844 GCC TAT GGT CGA GAT TTA ATG GAG GCC CAA GAG TGG TGC AGG AAG TAC ATG AAA TCA GGG AAT GTC AAG GAC CTC 918
57 A  Y  G  R  D  L  M  E  A  Q  E  W  C  R  K  Y  M  K  S  G  N  V  K  D  L  81
919 CTC CAA GCC TGG GAC CTC TAT TAT CAT GTG TTC CGA CGA ATC TCA AAG ACT AGT TAT CCG TAC GAC GTA CCA GAC 993
82 L  Q  A  W  D  L  Y  Y  H  V  F  R  R  I  S  K  T  S  Y  P  Y  D  V  P  D  106
994 TAC GCA TAA gaaaagtgaggatcctgagaacttcagggtgagtttggggacccttgattgttcttcttttctgctattgtaaaattcatgttatat 1090
107 Y  A  *  BamHI <--
1091 ggagggggcaaaagtttccaggggtgtgttagaatgggaagatgtccctgtatcaccatggaccctcatgataatttgtttcttctacttctactct 1190
1191 gttgacaaccattgtctcctcttattttcttttcttttctgtaacttttctgtaaaccttttagcttgcatttgttaacgaatttttaaatcacttttgt 1290
1291 ttatttgtcagattgtaagtaacttttctcctaatcacttttttcaaggcaatcagggtatattatattgtacttcagcacagtttttagagaacaattgtt 1390
1391 ataattaaatgataaggtagaataatttctgcatataaatctctggctggcgtggaataattcttatttgtagaacaactacatcctgggtcatcatcctgc 1490
1491 ctttctctttatgggtacaatgatatacactgtttgagatgaggataaaactctgagtcctaaacggggccctctgtaaacatgttcatgcctctt 1590
1591 ctttttctacagctcctgggcaacgtgctggtgtgtgtgctgtctcatcttttggcaaggattcactcctcaggtgcagggtgcctatcagaagggtg 1690
1691 gtggctggtgtggccaatgccctggctcacaataaccactgagatcttttccctctgcacaaaattatggggacatcatgaagcccttgagcatctga 1790
1791 cttctggctataaaaggaaattttatttcttgcaatagtgtgttgaatttttgtgtctctcactcggaaggacatattgggaggcaaatcattttaa 1890
1891 acatcagaatgagtatttggttttagagtttggaacatagccatagctggctgccatgaacaaagggtggctataaagggtcatcagtatatgaaaca 1990
1991 gccccctgctgtccattccttattccatagaaaagccttgacttgaggttagatttttttatattttgtttgtgtattttttctttaaaccctta 2090
2091 aaatttctctacatgttttactagccagattttctcctctcctgactactccagtcacatagctgtccctctctcttattgaagatcctcgagggac 2190
XhoI

```