Chapter 3

Construction of targeting vectors and a novel vector system for recombination-based DNA cloning

This chapter describes the construction of three different targeting vectors for the introduction of a constitutive knockout, a lacZ knock-in and a conditional mutation into the SAP102 locus. It also presents the construction and validation of a novel set of plasmids to facilitate rapid and flexible targeting vector construction using homologous recombination in bacteria.

3.1 SAP102 constitutive targeting vector

Advantages of a constitutive SAP102 knockout mouse

A constitutive knockout strategy for SAP102 was initially used because a germline deletion results in every cell of the targeted mouse lacking SAP102 throughout its lifetime, the most comprehensive method of determining SAP102 function in the whole organism. This is also the situation for human males with inherited mutations in SAP102, thus a constitutive knockout makes the best model of the human disorder. A simple deletion of part of the gene also means rapid generation of the mutant mouse and an uncomplicated targeting vector for testing recombineering-based construction strategies.

SAP102 is not expressed in ES cells

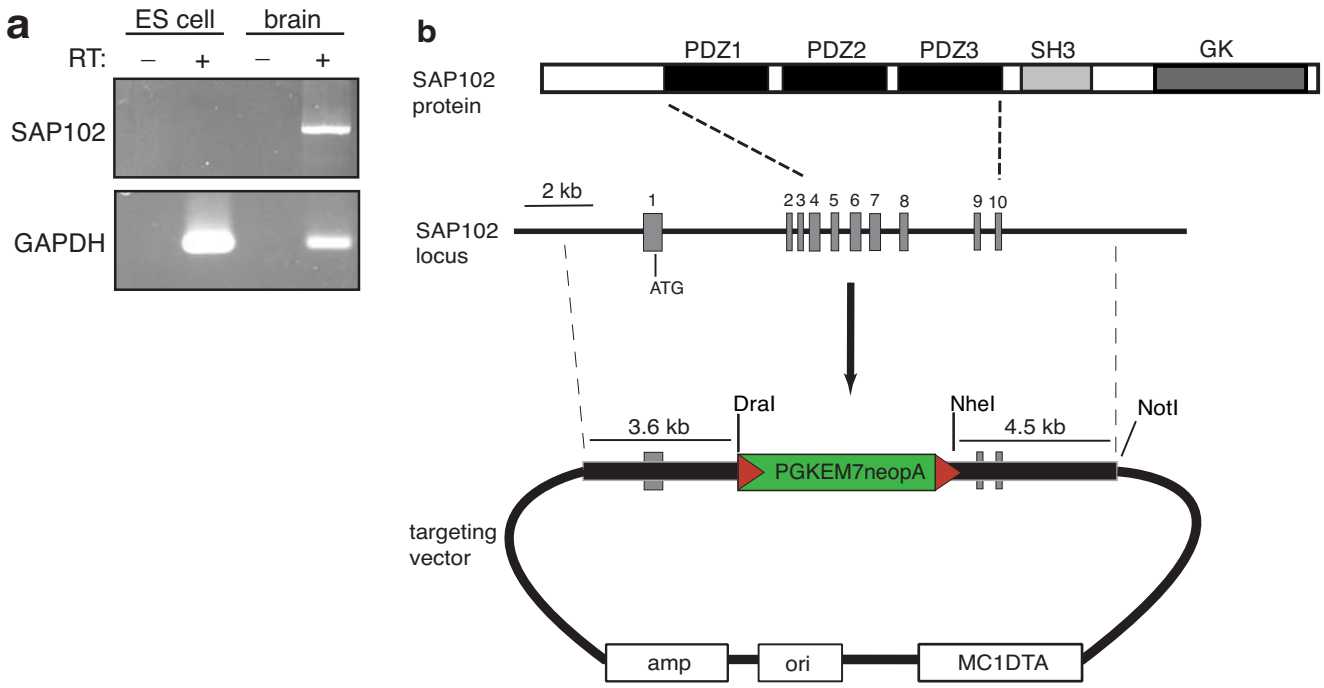
To design a targeting strategy, the mouse HM-1 ES cells to be used for targeting were first tested for SAP102 expression using RT-PCR. If a gene is expressed in ES cells it can be targeted by insertion of a positive selectable marker with no promoter but only a splice acceptor sequence and an internal ribosome entry site (IRES) to drive independent translation, so that the marker is expressed only when integrated into an active locus. This enhances targeting efficiency by greatly reducing random genomic integration of the targeting construct (Hasty et al., 2000). Figure 3.1a shows that SAP102 mRNA is present in total RNA extract from whole brain but not from ES cells, therefore a promoter needed to be included in the selection cassette for targeting.

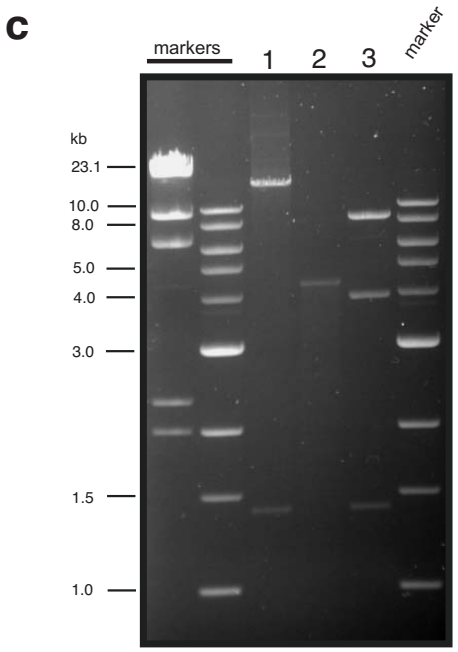
Targeting vector construction

The constitutive targeting vector was constructed using SAP102 genomic sequence ENSMUSG00000000881 in the Ensembl repository (Birney et al., 2004) and cDNA sequence NM_016747 in the NCBI Entrez Nucleotide database (www.ncbi.nlm.nih.gov). The accuracy of the sequence was verified by PCR and Southern blot analyses. Multiple primer pairs through the target region amplified PCR products of the expected sizes and on Southern blots three unique, intronic, DNA probes from the target region hybridised to genomic DNA fragments of the expected sizes following a variety of restriction enzyme digestions (figures 4.1, 4.2 and data not shown).

To obtain a genomic SAP102 clone, a 129Sv genomic BAC library was screened using PCR primer pairs at each end of the targeting region. The identity and structure of the SAP102 locus in the clone was verified by end-sequencing, restriction digestion, PCR and Southern blot using the known genomic sequence and the experimental results from analyses of the genomic locus above (data not shown).

The targeting vector was constructed by first excising a SAP102 genomic fragment containing the targeting region from the BAC clone into a pBluescript vector containing a DTA negative selectable marker (pSKDTA). This was performed by recombination-based cloning in JC9604 strain *E.coli*, using short homology arms of 70 bp each which were attached to the ends of pSKDTA by PCR amplification (Zhang et al., 2000). This recombineering step was successful but inefficient, producing eight antibiotic-resistant colonies of which only one carried the correct clone (table 3.1, row 1).





Expected NheI fragment sizes (kb)

1	2	3
15.3	4.3	8.3
1.4		3.9
		1.4

Figure 3.1 SAP102 constitutive targeting vector. (a) SAP102 is not expressed in ES cells. RT-PCR using primers for SAP102 cDNA produces a 1 kb band from RNA from whole brain but not from mouse ES cells. Amplification of GAPDH shows the integrity of the cDNA in each sample. **(b)** Targeting vector design and construction. The PDZ domains of SAP102 are encoded by exons 4 to 10 inclusive. The targeting construct consists of a genomic fragment containing this region with exons 2 to 8 inclusive replaced by a loxP-flanked selection cassette (green box with red triangles). A unique NotI restriction site allows for linearisation of the vector and DraI and NheI restriction sites are included at the ends of the cassette to facilitate genotyping. The vector backbone carries a diphtheria toxin A (DTA) fragment negative marker to discourage random integration. **(c)** Restriction digestion confirms the structure of the targeting vector. Shown are NheI digestions of the targeting vector backbone with SAP102 genomic subclone only (1), the parental plasmid containing the selection cassette (2) and the completed targeting vector with the cassette inserted (3).

An attempt was then made to produce a lacZ knock-in targeting vector by inserting an T3-IRES-lacZ-polyA-*loxP-neo-loxP* cassette into the cloned SAP102 genomic fragment by recombineering. Insertion of this cassette into the coding region of a gene results in a fusion mRNA containing both the endogenous cDNA and cassette sequences. The IRES allows cap-independent ribosome binding and lacZ translation independently of the remainder of the mRNA transcript. This experiment was repeatedly unsuccessful even after increasing the length of the short homology arms from 70 to 800 bp (Liu et al., 2003), shortening the cassette by removing the IRES-lacZ marker and trying a number of different recombination-competent *E.coli* strains (table 3.1, rows 2-6). Further experiments attempting to repeat the original BAC excision or to excise different genomic fragments from the same BAC also failed (data not shown).

To facilitate rapid construction of the targeted mouse a traditional restriction-ligation strategy was then used to introduce a mutation into the SAP102 genomic fragment. No suitable exonic restriction sites were available for insertion of the lacZ cassette so instead a simple *loxP*-flanked antibiotic resistance (*neo*) cassette was used. Figure 3.1b shows the targeting strategy and completed constitutive targeting vector. The cluster of exons encoding the PDZ domains were targeted since they mediate interaction with NMDARs and are close to the 5' end of the gene, minimising the possibility of a functional N-terminal peptide being produced upstream of the mutation. A deletion of exons 2-8 inclusive was engineered as shown (figure 3.1b), removing the coding sequence for amino acids 120 – 400, deleting PDZ domains 1 and 2 and introducing a frameshift mutation between exons 1 and 9. The deleted section was replaced with a selection cassette containing the mouse neomycin phosphotransferase (*neo*) gene driven by a compound phosphoglycerate kinase (PGK) and EM7 promoter for kanamycin resistance in bacterial and G418 resistance in vertebrate cells respectively. The integrity of the completed targeting vector was confirmed by extensive restriction digestion and by sequencing of the junctions of ligated

DNAs. Figure 3.1c shows a restriction digest of the completed targeting vector and its two parental plasmids.

3.2 Construction of an improved system for flexible and efficient targeting vector production

Because of the power and flexibility offered by recombineering as a method for building targeting vectors and because of its successful use by others (Copeland et al., 2001; Liu et al., 2003; Zhang et al., 2000), attempts were made to devise a reliable recombineering strategy that would be effective in producing many different types of mutations in a large number of different mouse loci.

To improve recombineering efficiency and facilitate engineering of a variety of types of mutations in many different genes, a new set of plasmids for recombineering-based targeting vector construction was constructed (figure 3.2). The vectors were designed to minimise intra- and inter-plasmid sequence homology, thereby reducing the risk of aberrant recombination between plasmid backbones and selection cassettes during homologous recombination in bacteria. Donor and recipient vectors carry different antibiotic markers to allow independent selection and further minimise inter-plasmid sequence homology. Each vector contains multiple cloning sites (MCS) for insertion of PCR-amplified homology arms to mediate recombination. All insertion cassettes carry the *neo* gene driven by dual promoters for positive selection in both *E.coli* and ES cells. Each plasmid was verified by extensive restriction digestion followed by full sequencing using a transposon-mediated shotgun method. The verified sequence of each plasmid is shown in appendix 3.

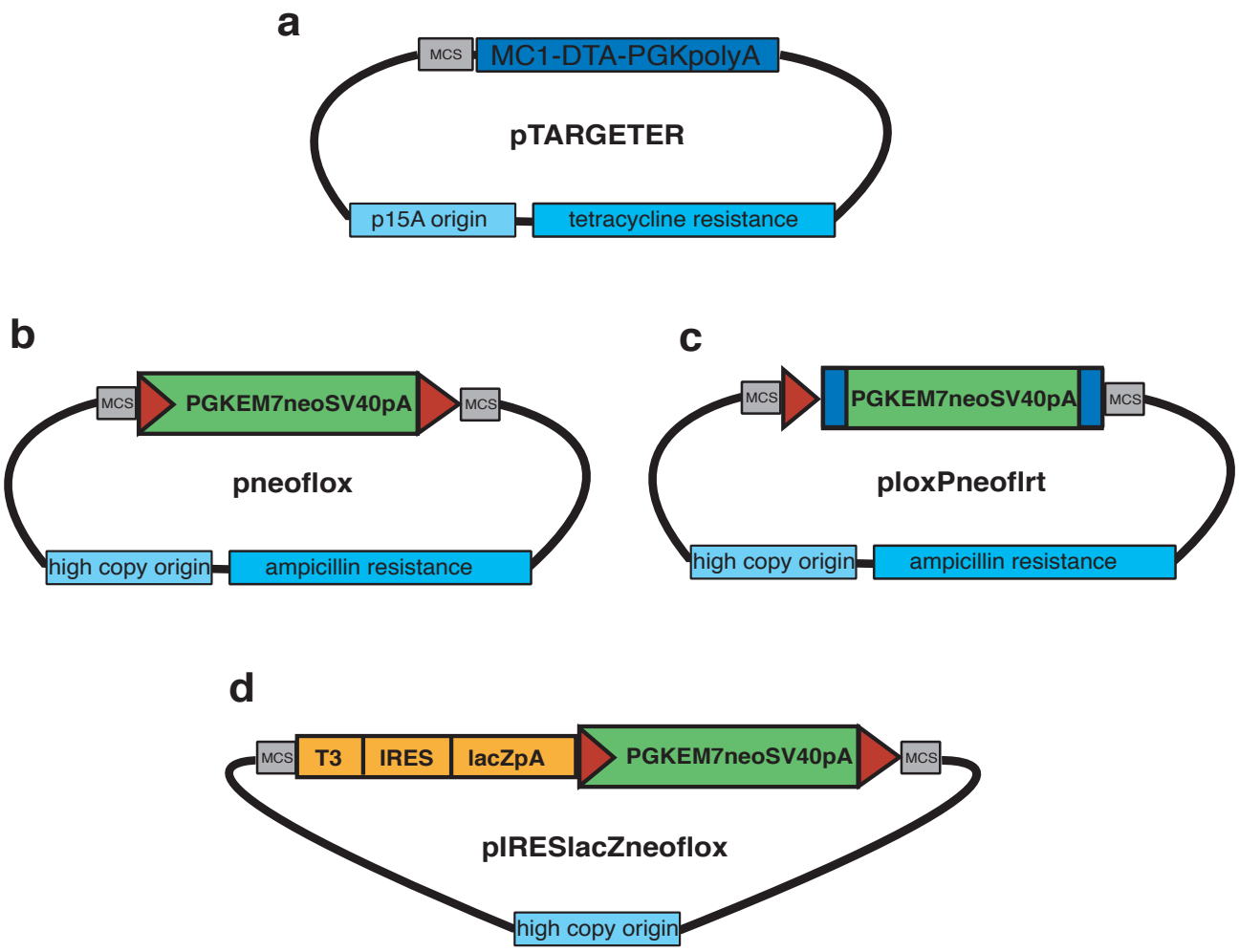
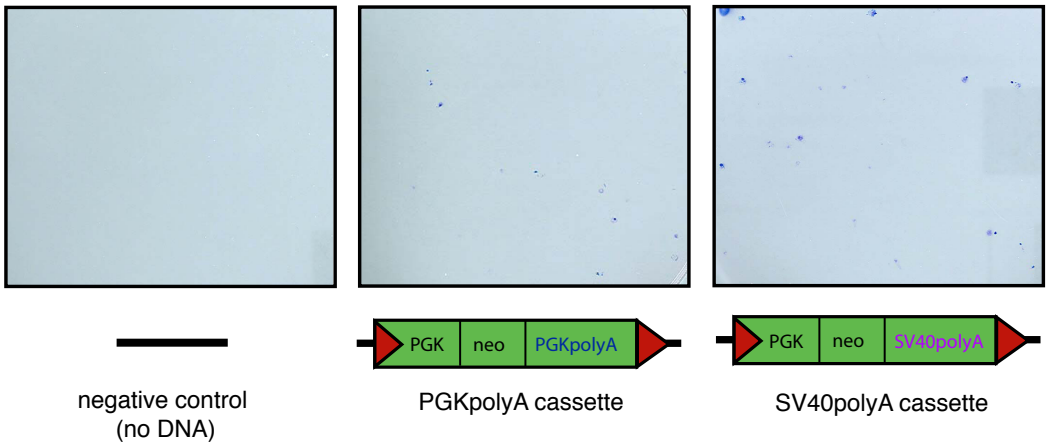


Figure 3.2 The TARGETER vector system for recombination-based construction of targeting vectors. (a) pTARGETER is the targeting vector backbone. It carries a low-copy p15A replication origin to facilitate insertion of large genomic fragments and a DTA negative marker with PGK polyadenylation signal to discourage random integration of the vector into the ES cell genome. **(b)** pneoflox allows introduction of a single loxP site via insertion of a loxP-flanked antibiotic resistance (neo) cassette followed by Cre recombinase-mediated excision of the cassette. **(c)** ploxPneoflrt facilitates insertion of a second loxP site into the same genomic fragment using a loxP site adjacent to an FRT-flanked antibiotic resistance cassette which is subsequently removed by Flp recombinase-mediated excision. **(d)** pIRESlacZneoflox allows insertion of a lacZ marker to track transcriptional activity. The cassette carries termination codons in all three frames (T3) to interrupt translation of the endogenous protein. The selection cassette can be subsequently excised by Cre recombinase. All vectors carry multiple cloning sites (MCS) to facilitate the insertion of short homology arms for recombineering. Red triangles - loxP sites; dark blue boxes - FRT sites.

a



b

plasmid	G48-resistant colonies
none	0
PGKpolyA	33
SV40polyA 1	47
SV40polyA 2	51
SV40polyA 3	33

Figure 3.3 The SV40 polyA signal sequence in TARGETER cassettes mediates efficient expression of a drug resistance marker in mouse ES cells. (a) ES cells were electroporated with linearised plasmid vectors containing neo antibiotic resistance gene with a PGK promoter and either a PGK or SV40 polyA signal sequence. Electroporated cells were subjected to G418 selection for five days, after which drug-resistant colonies were stained with Giemsa for counting. Representative images of Giemsa-stained colonies are shown. **(b)** Colony counting shows that the SV40 polyA in three different plasmids is at least as effective at conferring drug resistance as the PGK polyA.

The targeting vector backbone is p*TARGETER* (figure 3.2a), with a low-copy p15A replication origin to allow incorporation of large genomic fragments. As a negative selection marker for gene targeting, the plasmid carries a Diphtheria Toxin A (DTA) fragment driven by an MC1 promoter with a mouse phosphoglycerate kinase (PGK) polyA signal. DTA is more effective and convenient than thymidine kinase in reducing random genomic integration of the targeting construct (Yagi et al., 1993; Yanagawa et al., 1999).

p*neoflox* (figure 3.2b) is a vehicle for insertion of a *loxP*-flanked selection cassette into a genomic fragment. Exposure to Cre recombinase excises the cassette, leaving a single *loxP* site behind. p*loxPneoflrt* (figure 3.2c) enables insertion of a second *loxP* site as its selection cassette can be excised by Flp recombinase without interference from the *loxP* sites. These plasmids can be used to construct targeting vectors for simple deletions, conditional mutations and point mutations.

p*IRESlacZneoflox* carries a knock-in cassette consisting of the lacZ reporter gene preceded by an IRES to drive reporter expression following integration into a target exon (figure 3.2d). Translation termination codons in three different frames are included at the front of the cassette to ensure truncation of the endogenous protein. *LoxP* sites flanking the selection cassette allow its removal following homologous recombination, preventing interference between the strong PGK promoter in the cassette and that of the endogenous SAP102 gene (Dymecki, 2000).

Since the DTA negative selection marker in *pTARGETER* uses the PGK polyA signal, this sequence was removed from the insertion cassettes and replaced with the SV40 polyA signal to further minimise sequence homology. The efficacy of the SV40 polyA in mediating expression of the selectable marker was tested by electroporation of the new plasmids into mouse ES cells, shown in figure 3.3. SV40 polyA-containing selection cassettes produced at least as many G418-

resistant colonies as the original, PGK polyA-containing vectors, demonstrating that the SV40 polyA is effective for use in gene targeting in ES cells.

3.3 SAP102 targeting vector construction using the *TARGETER* vector system

To test the efficacy of the *TARGETER* system and to produce additional useful SAP102 mutations in mice, lacZ knock-in and conditional targeting vectors for SAP102 were constructed. Each step was performed by recombination-based cloning in EL350 *E.coli* (Liu et al., 2003), using PCR-amplified homology arms ligated into multiple cloning sites in the *TARGETER* vectors.

SAP102 lacZ knock-in targeting vector

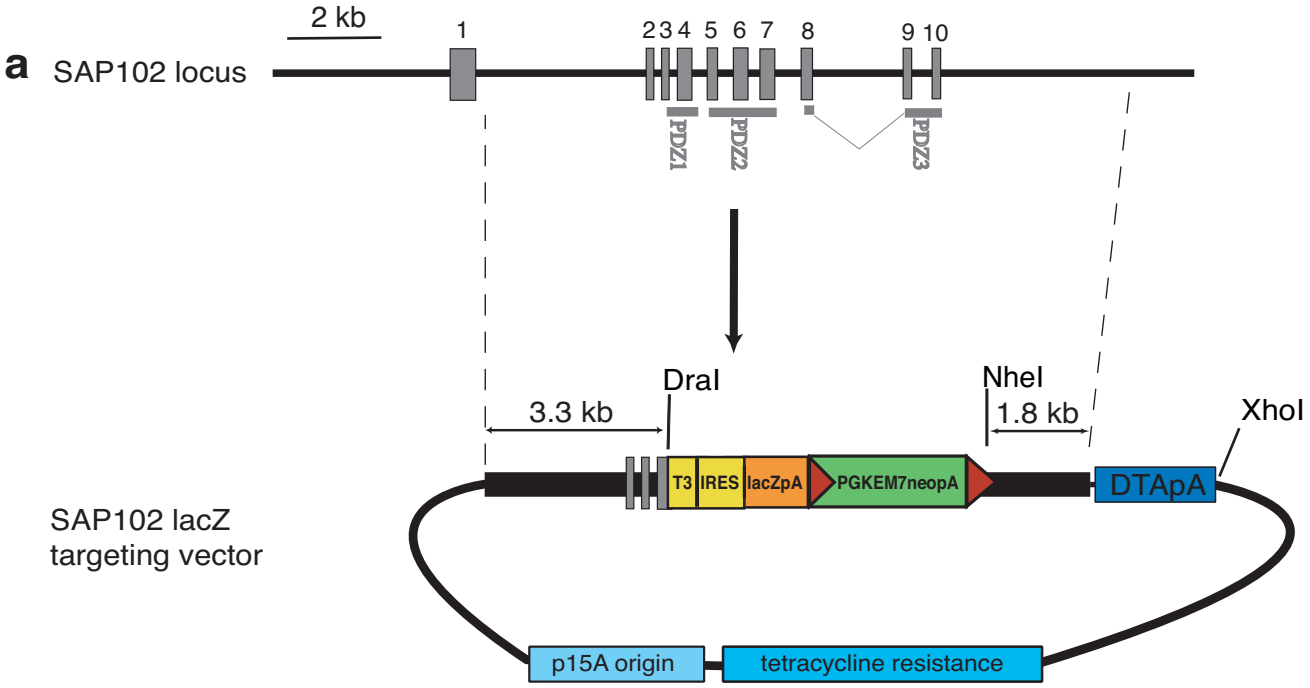
This vector inserts a lacZ reporter gene into the SAP102 locus, allowing cell-specific analysis of SAP102 transcriptional activity along with a constitutive knockout mutation. In addition to its use in analysing spatial and temporal SAP102 expression patterns, it provides a simple, high-resolution means to track X-inactivation in heterozygous targeted female mice. Skewed X-inactivation is often associated with mental retardation in heterozgous human females (Ropers and Hamel, 2005) and there is some evidence for impaired cognitive function in female carriers of SAP102 mutations (Tarpey et al., 2004).

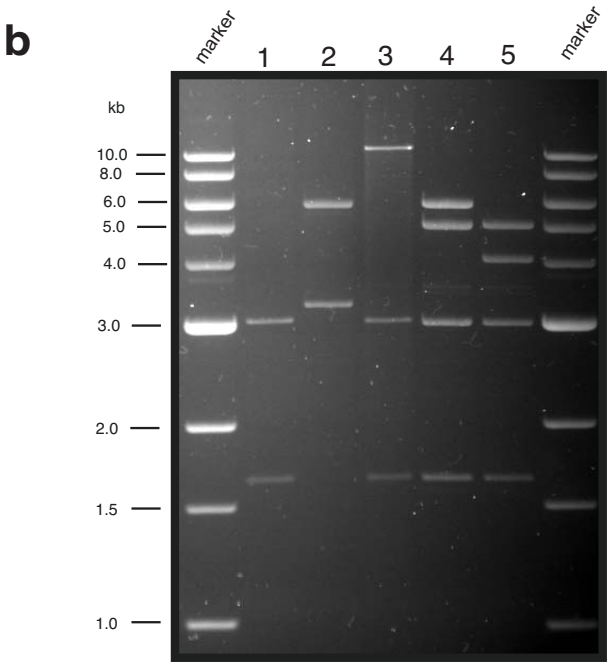
To construct the targeting vector, a 10.9 kb genomic fragment encompassing the SAP102 targeting region was excised from the SAP102 BAC clone into p*TARGETER* by recombineering. The IRES-lacZ-polyA-neoflox cassette from p*IRESlacZneoflox* was then inserted into this fragment, again by recombineering, deleting part of exon 4 and all of exons 5-10 inclusive. An in-frame myc epitope was incorporated in front of the cassette to allow tracking of any truncated SAP102 peptide produced using an anti-myc antibody (van der Weyden et al., 2002). Figure 3.4a shows the SAP102 lacZ knock-in targeting vector, the integrity of which was confirmed by

extensive restriction digestion and by DNA sequencing across each recombination junction. The functionality of the *loxP* sites was confirmed by Cre-mediated recombination in arabinose-induced EL350 *E.coli.* Figure 3.4b shows a restriction digest confirming the structure of the parental plasmids, the completed targeting vector and the same vector following Cre-mediated cassette excision.

SAP102 conditional targeting vector

Flanking essential exons of SAP102 with *loxP* sites by gene targeting allows the use of Cre recombinase-expressing mouse lines to generate spatially- and temporally-restricted ablation of SAP102 function (Dymecki, 2000). SAP102 begins to be expressed soon after birth in mice (Sans et al., 2000) and may be required for postnatal viability. Mutations in SAP97, another early-expressing PSD-95 family protein, are lethal. If SAP102 constitutive knockout mice do not survive then a conditional mutation will be necessary to generate live adult mice for experimental analyses. The fact that SAP102 is expressed early postatally and that humans carrying SAP102 mutations experience developmental delay (Tarpey et al., 2004), as well as the known developmental role of SAP102-interacting proteins such as NMDARs (Sprengel and Single, 1999) strongly suggests that SAP102 plays a role in mouse postnatal development. Ablation of SAP102 in adulthood would provide a means for distinguishing between its developmental function and an acute role in synaptic transmission, synaptic plasticity and cognitive function in the adult. Restriction of the mutation to specific brain regions, such as the hippocampus or cerebellum, would make possible a specific analysis of SAP102 function in those areas.





Expected BamHI fragment sizes (kb)

1	2	3	4	5
3.0	5.9	10.9	5.9	5.0
1.6	3.3	3.0	5.0	4.1
		1.6	3.0	3.0
			1.6	1.6

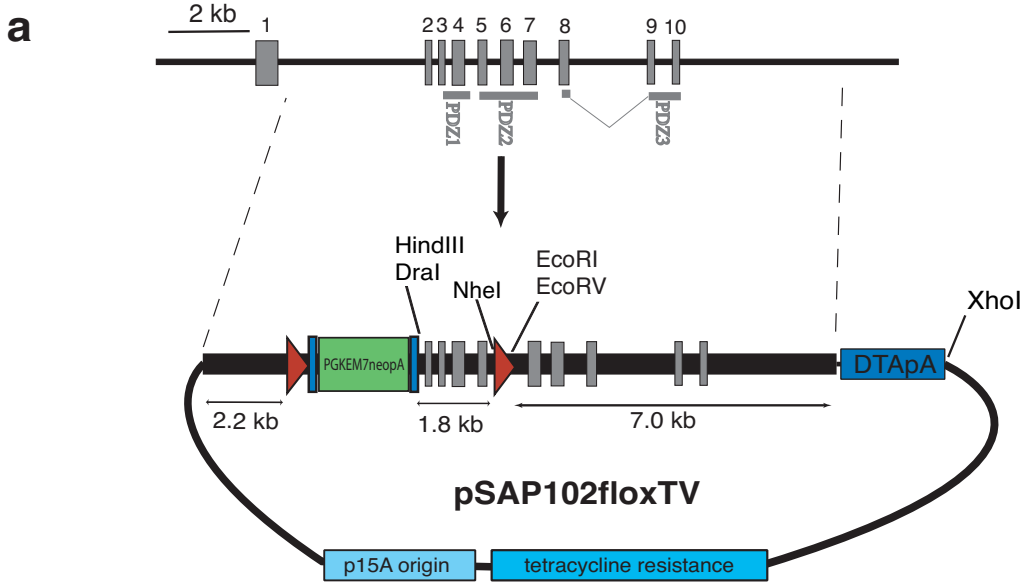
Figure 3.4 SAP102 lacZ knock-in targeting vector. (a) The T3-IRES-lacZ-neo cassette replaces part of exon 4 and all of exons 5-10 inclusive. SAP102 genomic fragments of 3.3 and 1.8 kb mediate homologous recombination. The vector backbone is pTARGETER, containing the DTApolyA negative marker against random integration. DraI and NheI sites in the selection cassette allow genotyping and a unique XhoI site is included for vector linearisation. **(b)** Restriction digestion confirms the structure of the targeting vector. Shown are BamHI digestions of pTARGETER (1), pIRESlacZneoflox with short homology arms for recombination (2), pTARGETER with SAP102 genomic subclone (3), the completed lacZ targeting vector (4) and the same vector following Cre recombinase-mediated recombination to remove the selection cassette (5).

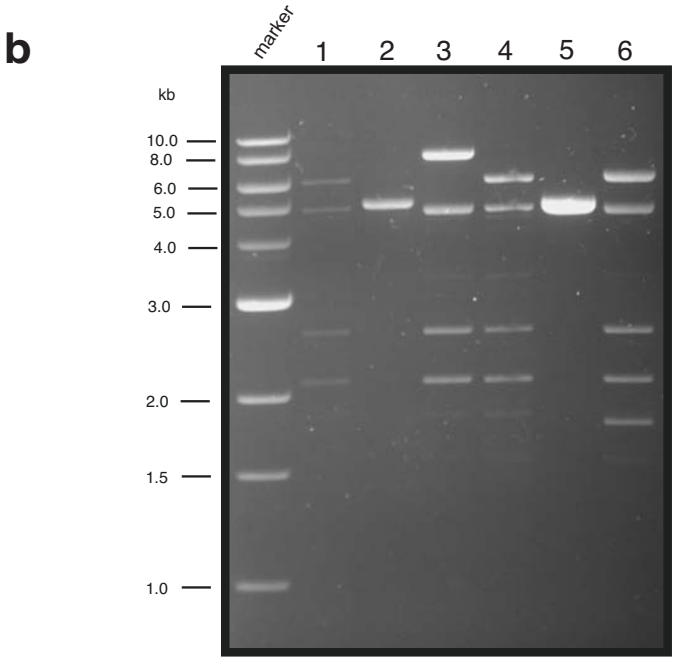
The conditional targeting vector was built using the same SAP102 genomic fragment in p*TARGETER* as was used in constructing the lacZ knock-in vector. Into intron 5 of SAP102 in this subclone was inserted the *loxP*-flanked selection cassette from p*neoflox*, which was then excised in arabinose-induced, Cre-expressing, EL350 *E.coli*, leaving a single *loxP* site behind. A second *loxP* with an adjacent, *FRT*-flanked selection cassette was then inserted from p*loxPneoflrt* into SAP102 intron 1 in the same fragment, completing the targeting vector (figure 3.5a). Extensive restriction digestion and sequencing across the recombination junctions confirmed the vector's integrity (figure 3.5b) and the functionality of the *loxP* and *FRT* sites was confirmed by Cre- and Flp-mediated recombination in arabinose-induced EL350 and EL250 *E.coli* respectively.

Recombineering efficiency using *TARGETER* vectors

In contrast to previous attempts, each recombineering step performed using the *TARGETER* vectors worked with striking efficiency, as shown in table 3.1, rows 7-10. The system worked effectively in excising and inserting a variety of DNA fragments of sizes from 1.9 to 10.9 kb. Sequencing across junctions confirmed that this was accomplished with precise control, to a single nucleotide, over the sites of recombination and without the need for restriction enzyme sites in the genomic sequence. Each step produced at least 1,000 antibiotic-resistant bacterial colonies of which, in the majority of cases, 90 % or more carried the desired recombinant plasmid. The lowest proportion of correct recombinants achieved was 25 %, in the BAC excision step (table 3.1).

These vectors have now been used to produce targeting vectors, exclusively by recombineering, for six other brain-expressed genes with recombination efficiencies of 59.5 +/- 33% for the BAC excision and 51.2 +/- 45 % for the selection cassette insertion step (Noboru Komiyama, personal communication), demonstrating the reliability of the system and its applicability to different loci. One of these vectors has been used for targeting ES cells with a targeting efficiency of 8 %.





Expected HindIII fragment sizes (kb)

1	2	3	4	5	6
6.0	5.1	7.9	6.1	4.6	6.2
4.9		4.9	4.9	0.5	4.9
2.6		2.6	2.6		2.6
2.1		2.1	2.1		2.1
					1.8

Figure 3.5 SAP102 conditional targeting vector. (a) The vector consists of a loxP site and FRT-flanked selection cassette in intron 1 and a second loxP site in intron 5 of a SAP102 fragment in the pTARGETER backbone. Restriction sites were introduced along with the mutations as shown to facilitate genotyping. Arms for homologous recombination are 7.0 and 2.2 kb. Red triangles - loxP sites; blue boxes - FRT sites; green box - selection cassette. **(b)** Restriction digestion confirms the structure of the targeting vector. Shown are HindIII digests of pTARGETER with the SAP102 genomic subclone (1), pneoflox with short homology arms for recombination (2), pTARGETER.SAP102 with neoflox cassette inserted (3), the same plasmid after Cre recombinase-mediated removal of the cassette, leaving a single loxP site (4), ploxPneoflrt with short homology arms for recombination (5) and the completed targeting vector with the loxPneoflrt cassette inserted (6).

Table 3.1 Efficiency of recombineering in construction of SAP102 targeting vectors

	Cloning step	Size of excision/ insertion fragment (kb)	Donor vector	Recipient vector	*E.coli* strain	length per homology arm (bp)	Antibiotic -resistant colonies	% correct (from at least 20 analysed)
1.	BAC excision	13.0	SAP102 BAC clone	pBSSK.DTA	HS996 + pBADαβγ	70	8	12.5
2.	IRESlacZneoflox insertion	6.0	pSP76IRESlacZneoflox	pBSSK.DTA.SAP102	JC9604	70	120	0
3.	IRESlacZneoflox insertion	6.0	pSP76IRESlacZneoflox	pBSSK.DTA.SAP102	HS996 + pBADαβγ	70	75	0
4.	IRESlacZneoflox insertion	6.0	pSP76IRESlacZneoflox	pBSSK.DTA.SAP102	JC9604	70	150	0
5.	IRESlacZneoflox insertion	6.0	pSP76IRESlacZneoflox	pBSSK.DTA.SAP102	HS996 + pBADαβγ	800	160	0
6.	neoflox insertion	1.9	pSP77neoflox	pBSSK.DTA.SAP102	EL350	800	100	0
7.	BAC excision	10.9	SAP102 BAC clone	pTARGETER	EL350	650	1,000	25
8.	neoflox insertion	1.9	pneoflox	pTARGETER.SAP102	EL350	400	3,000	100
9.	neoflrt insertion	1.9	ploxPneoflrt	pTARGETER.SAP102.loxP	EL350	500	1,000	90
10.	IRESlacZneoflox insertion	6.0	pIRESlacZneoflox	pTARGETER.SAP102	EL350	800	8,000	100

3.4 Discussion

These results show that the *TARGETER* vectors combined with recombineering technology provide a rapid and powerful method for constructing targeting vectors for introducting of a variety of mutations into the mouse genome, including simple deletions, knock-ins and conditional mutations. This strategy works with high efficiency and allows precise control over insertion and excision points in the DNA without the need for restriction sites. The system is effective for constructing targeting vectors for numerous loci and should be broadly useful for targeting almost any location in the genome.

The construction of three different targeting vectors for the SAP102 locus provides a basis for a detailed genetic analysis of SAP102 *in vivo* in the mouse. The constitutive knockout vector will allow analysis of SAP102 function across all organ systems and will provide a model for NS-XLMR in humans. The lacZ knock-in vector will facilitate high-resolution examination of SAP102 transcriptional activity and X-inactivation in heterozygous female mice. The conditional targeting vector will help determine the role of SAP102 in specific brain regions during distinct developmental periods.