MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue

N. A. Datson*, J. van der Perk-de Jong, M. P. van den Berg, E. R. de Kloet and E. Vreugdenhil

Division of Medical Pharmacology, Leiden/Amsterdam Center for Drug Research, Leiden University, PO Box 9503, 2300 RA Leiden, The Netherlands

ABSTRACT

Serial Analysis of Gene Expression (SAGE) is a powerful expression profiling method, allowing the analysis of the expression of thousands of transcripts simultaneously. A disadvantage of the method, however, is the relatively high amount of input RNA required. Consequently, SAGE cannot be used for the generation of expression profiles when RNA is limited, i.e. in small biological samples such as tissue biopsies or micro-dissected material. Here we describe a modification of SAGE, named microSAGE, which requires 500- to 5000-fold less starting material. Compared with SAGE, microSAGE is simplified due to incorporation of a 'single-tube' procedure for all steps from RNA isolation to tag release. Furthermore, a limited number of additional PCR cycles are performed. Using micro-SAGE gene expression profiles can be obtained from minute quantities of tissue such as a single hippocampal punch from a rat brain slice of 325 μm thickness, estimated to contain, at most, 10^5 cells. This method opens up a multitude of new possibilities for the application of SAGE, for example the characterization of expression profiles in tissue biopsies, tumor metastases or in other cases where tissue is scarce and the generation of region-specific expression profiles of complex heterogeneous tissues.

INTRODUCTION

Nearly all biological events like cell division and differentiation, responsiveness to hormones or growth factors and ultimately cell death are associated with changes in expression of key genes. In addition, extensive changes in gene expression occur during the onset and progression of disease. By comparing expression profiles under different conditions, individual genes or groups of genes can be identified that play an important role in a particular signalling cascade or process or in disease etiology, perhaps providing clues to the underlying molecular mechanism or even to their function.

Several methods have been developed to identify changes in expression profiles, including subtractive hybridisation (1,2), comparative EST analysis (3–6) and differential display (7–11). However, most of these methods are only capable of analysing limited numbers of transcript species simultaneously and do not provide quantitative data on expression levels. In addition, only changes in expression of abundant mRNAs can be detected.

The Serial Analysis of Gene Expression (SAGE) method, in contrast, allows qualitative and quantitative analysis of thousands of transcripts simultaneously (12). In SAGE, short sequence tags (~10 bp) are isolated from mRNA at a defined position, ligated to long multimers, cloned and sequenced. The frequency of each tag in the cloned multimers directly reflects transcript abundancy. In addition, the short tags are long enough to uniquely identify the corresponding transcript in database searches. Thus, SAGE results in an accurate picture of gene expression at both the qualitative and the quantitative level. In a single sequencing reaction over 30 tags can be read serially, an improvement of efficiency of at least 30-fold compared with conventional EST analysis (4,13,14). Depending on the number of tags sequenced, changes in expression levels of rare transcripts can be detected. The power of SAGE for use in expression profiling has been nicely demonstrated in a number of studies, including character-isation of the entire yeast transcriptome (15), identification of p53-regulated genes (16,17) and analysis of expression profiles in normal versus cancer cells (18).

A major drawback of SAGE is the requirement of a large amount of input RNA [2.5–5 μg poly(A)$^+$ RNA]. Although SAGE potentially has applications in many fields of research, its use is thus restricted to situations in which the amount of starting material is not limiting, such as yeast cultures, cell lines or large solid tumors. Analysis of changes in expression profiles in small or scarce biological samples, e.g. biopsies or post-mortem material, is not possible simply due to the fact that these tissue samples do not contain the required 2.5–5 μg mRNA. In addition, the analysis of expression profiles in complex tissues composed of highly heterogeneous cell populations is rather difficult, since transcriptional changes in a specific subtype of cells will be diluted by the expression profiles of other cell types present in the tissue, thus perhaps masking relevant changes in expression. In such cases it is preferable to specifically isolate the cell population of interest for expression profiling, rather than using the complex tissue as a whole. Although microdissection technology has

Table 1. Main differences between SAGE procedure and modified procedure for limited amounts of tissue

	SAGE (Velculescu et al., 1995)	microSAGE procedure
Amount of input material:	$2.5 - 5$ µg polyA$^+$ RNA	Total RNA from a single punch of a 300 µm tissue slice (10^5 cells; ± 1-5 ng polyA$^+$ RNA)
Capture of cDNA by:	Streptavidin-coated magnetic beads	Streptavidin-coated PCR tube
Multiple-tube reaction vs. single- tube reaction:	♦ Subsequent reactions performed in multiple tubes ♦ Multiple phenol-chloroform extraction and ethanol precipitation steps between cDNA synthesis and tag release	♦ Single-tube reaction from RNA isolation to tag release ♦ easy change of buffers between subsequent steps ♦ no phenol-chloroform extraction or ethanol precipitation step until after tag release ♦ fewer manipulations reduces loss of material
PCR:	25-28 cycles of PCR	28 cycles followed by a limited number of cycles of re-PCR on excised ditag (8-15 cycles)

improved significantly over the last years (19), obtaining sufficient RNA from a specific subpopulation of cells is a laborious task. Finally, a disadvantage of SAGE is that it is characterised by a large number of sequential reactions and purifications, which can give rise to a significant loss of material.

To overcome some of the above-mentioned problems, we have developed a modified SAGE procedure, microSAGE, which allows use of very limited amounts of starting material (Table 1). MicroSAGE is simplified due to the incorporation of a 'single-tube' procedure replacing several of the many steps in SAGE. The single-tube procedure is not only easier to perform, but is also accompanied by less loss of material between subsequent steps. In addition, in microSAGE a limited number of additional PCR cycles are performed to generate sufficient ditag. We demonstrate that using microSAGE it is possible to zoom in on a highly specialised brain region and to obtain a region-specific SAGE expression profile. The present modified SAGE procedure opens up a multitude of new possibilities for expression profiling, for example when combined with microdissection to generate region-specific expression profiles of complex heterogeneous tissues. Moreover, it can be used for expression profiling in tissue biopsies, tumor metastases or in cases where tissue is scarce, i.e. post-mortem tissue.

MATERIALS AND METHODS

Punches of rat brain tissue sections

After decapitation, the rat brain was removed from the skull and rapidly frozen in isopentane on a mixture of dry ice and ethanol and stored at –80°C until further use. Alternating coronal sections of ~75 and 325 µm were prepared using a cryostat at –18°C, thaw-mounted on poly-L-lysine-coated slides and stored at –80°C. The 75 µm sections were Nissl-stained with cresylviolet, while the 325 µm unstained sections were used for punching out the dentate gyrus of the hippocampus according to the Palkovits punch out technique (20). Immediately before punching, the sections were removed from –80°C and placed in the cryostat at –18°C. The stained 75 µm sections, containing tissue which had been present on either side immediately adjacent to the 325 µm slice, served as landmarks to facilitate punching out of the correct region. Using a hollow needle (0.3 mm in diameter) chilled by dipping in liquid nitrogen, part of the inner blade of the dentate

gyrus was removed, transferred to a tube containing 20 µl of TRIzol (Gibco BRL) and stored at 4°C until further use. After punching, the 325 µm sections were also Nissl-stained with cresylviolet to confirm removal of the correct region.

RNA isolation and cDNA synthesis

Punches stored in TRIzol were homogenised using a micropestle, which was rinsed with 180 µl of TRIzol, making the total volume 200 µl, and incubated for 5 min at room temperature. Alternatively, instead of adding 180 µl TRIzol, multiple homogenised punches can be pooled after homogenisation. RNA isolation with TRIzol was performed according to the manufacturer's instructions, using 1 µl of glycogen (20 mg/ml; Boehringer Mannheim) as a carrier in the precipitation. The washed RNA pellet was resuspended in 20 µl of lysis buffer (mRNA Capture Kit; Boehringer Mannheim). The RNA was enriched for polyadenylated RNA molecules using the mRNA Capture Kit (Boehringer Mannheim). First, 4 µl of a biotinylated oligo(dT)$_{20}$ primer (5 pmol/µl) was added to the RNA and annealed for 5 min at 37°C. Subsequently, the RNA was transferred to a streptavidin-coated PCR tube and incubated for another 3 min at 37°C, thus immobilizing the mRNA fraction to the wall of the tube. Non-bound RNA was removed by gently washing three times with 50 µl washing solution (mRNA Capture Kit). After the final wash, the washing solution was removed and bound RNA was rinsed once with 50 µl 1× first strand buffer (50 mM Tris–HCl pH 8.3, 75 mM KCl, 3 mM MgCl$_2$; Gibco BRL). First strand cDNA synthesis was performed in the same tube and primed from the bound oligo(dT)$_{20}$ for 2 h at 42°C in a 20 µl reaction containing 4 µl 5× first strand buffer, 2 µl 0.1 M DTT, 1 µl 10 mM dNTPs, 1 µl SuperScript II RT (200 U/µl; Gibco BRL) and 12 µl DEPC-treated H$_2$O. After removal of the first strand synthesis solution, the single-stranded cDNA bound to the PCR tube was rinsed once with 50 µl washing solution and then with 50 µl 1× second strand buffer (100 mM KCl, 10 mM (NH$_4$)$_2$SO$_4$, 5 mM MgCl$_2$, 0.15 mM β-NAD, 20 mM Tris–HCl pH 7.5, 0.05 mg/ml BSA). Double-stranded cDNA was synthesized in a 20 µl reaction volume containing 4 µl 5× second strand buffer, 0.4 µl 10 mM dNTPs, 1 µl DNA polymerase I (10 U/µl; Gibco BRL), 0.5 µl T4 DNA ligase (5 U/µl; Gibco BRL), 0.5 µl RNase H (1 U/µl; Boehringer Mannheim) and 13.6 µl H$_2$O for 2 h at 16°C. The double-stranded cDNA was stored at –20°C until further use.

Anchoring and tagging of cDNA

After removal of the second strand reaction mixture, the double-stranded cDNA bound to the PCR tube was rinsed once with 50 µl washing solution and then with 50 µl 1× restriction buffer (50 mM potassium acetate, 20 mM Tris acetate, 10 mM magnesium acetate, 1 mM DTT pH 7.9; NEBuffer 4; New England Biolabs). The cDNA was digested with 20 U of the anchoring enzyme *Nla*III (New England Biolabs) for 1 h at 37°C in a 25 µl reaction volume, followed by heat inactivation at 65°C for 20 min. After rinsing with 50 µl washing solution and 50 µl 1× ligase buffer (50 mM Tris–HCl pH 7.6, 10 mM MgCl₂, 1 mM ATP, 1 mM DTT, 5% w/v PEG-8000; Gibco BRL), linkers 1 and 2 were added to the tube in a total volume of 25 µl consisting of 2.5 µl of each linker (100 ng/µl), 5 µl 5× ligase buffer and 15 µl LoTE (3 mM Tris–HCl pH 7.5, 0.2 mM EDTA pH 7.5). The linkers were annealed by heating for 2 min at 50°C followed by 15 min at room temperature and ligated after addition of 1 µl T4 DNA ligase (5 U/µl; Gibco BRL) for 2 h at 16°C. After ligation the reaction mixture was removed and the bound cDNA was rinsed with 50 µl washing solution and 50 µl 1× restriction buffer (NEBuffer 4; New England Biolabs). The cDNA tags were released by digestion with 2 U of the tagging enzyme *Bsm*FI (2 U/µl; New England Biolabs) for 1 h at 65°C in a 25 µl reaction volume. After digestion, the reaction mixture was transferred to a new 1.5 ml tube and the volume was raised to 200 µl with LoTE. The mixture was extracted with an equal volume of phenol–chloroform–isoamyl alcohol (25:24:1) (PCI), ethanol precipitated (200 µl sample, 3 µl glycogen, 100 µl 10 M ammonium acetate and 700 µl ethanol) by centrifugation at 13 000 r.p.m. for 15 min at 4°C. The pellet was washed twice with 70% ethanol and resuspended in 21.5 µl LoTE.

Ligation to ditags and PCR amplification

The released cDNA tags were blunt-ended at 37°C for 30 min in a 30 µl reaction containing 21.5 µl cDNA tags, 6 µl 5× second strand buffer, 0.5 µl BSA (10 mg/ml), 0.5 µl 25 mM dNTPs, 1.5 µl Klenow (1 U/µl; Amersham). PCI extraction and ethanol precipitation was performed as described above and the pellet was resuspended in 4 µl LoTE. Ligation to ditags was performed overnight at 16°C in a 6 µl reaction using 4 U T4 DNA ligase (5 U/µl; Gibco BRL). After ligation, the volume was raised to 20 µl by addition of 14 µl LoTE and 1 µl was diluted 100-fold. One microliter of the diluted ligation mixture was used as input in a 50 µl PCR reaction containing 8 mM MgCl₂, 6% DMSO, 1 mM dNTPs and 350 ng of both SAGE primers (12,15) in PCR buffer II (Perkin Elmer) using 5 U of AmpliTaq Gold (5 U/µl; Perkin Elmer) and amplified for 28 cycles of 30 s at 95°C, 1 min at 55°C and 1 min at 70°C with an initial heat activation of the enzyme for 15 min at 95°C and a final extension of 5 min at 70°C (Fig. 2A). The products derived from nine parallel reactions were pooled, extracted with PCI and ethanol precipitated. The pellet was washed with 70% ethanol, air-dried and resuspended in 100 µl LoTE. The entire sample was loaded on four lanes of a 12% polyacrylamide gel with a 10 bp ladder (Gibco BRL) as a marker. The region of the gel around 100 bp was excised across all four lanes of the gel and the gel was fragmented by spinning through a 0.5 ml tube, pierced with a 21 gauge needle, inserted in a 1.5 ml tube. The DNA was eluted from the gel fragments by adding 300 µl LoTE and incubating for 15 min at 65°C, followed by removal of the polyacrylamide on SpinX columns (Costar). After PCI

extraction and ethanol precipitation, the pellet was resuspended in 1 ml LoTE and 1 µl was used as input in a 50 µl re-PCR using the same primers as described above. A series of PCR reactions was performed to determine the optimal number of cycles of re-PCR, ranging from 6 to 18 cycles. Subsequently, a large-scale PCR amplification was performed consisting of 96 100 µl PCRs to generate sufficient material for ditag isolation.

Ditag isolation and concatenation

The 96 parallel PCR reactions were pooled, extracted with PCI, ethanol precipitated and resuspended in 250 µl LoTE. The material was loaded on a total of 12 lanes of a 12% polyacrylamide gel. After ethidium bromide staining the upper band of ~100 bp was excised and purified from the gel as described above and resuspended in 170 µl LoTE. The linkers were cleaved off by digestion with 100 U of *Nla*III in a 200 µl reaction volume for 1 h at 37°C. After digestion, the material was PCI extracted at 4°C and subsequently ethanol precipitated by chilling for 10 min in a dry ice/ethanol bath and centrifugation for 15 min at 13 000 r.p.m. in a microcentrifuge at 4°C. The pellet was resuspended in 15 µl LoTE and loaded on two lanes of a 12% polyacrylamide gel with a 10 bp ladder as marker (Fig. 2B). The ditag band running at 22–26 bp was excised and eluted as described above, except that the incubation was performed at 37 instead of 65°C. The pellet was resuspended in 7.5 µl LoTE. Purified ditags were ligated to concatemers by addition of 5 U T4 DNA ligase in a total volume of 10 µl for 30 min at 16°C and run in a single lane on an 8% polyacrylamide gel with a 100 bp ladder as a marker (Fig. 2C). After ethidium bromide staining, the gel regions between 400 and 800 bp and >800 bp were excised and the concatemers were purified as described above with an incubation at 65°C. Purified concatemers were subsequently cloned in the *Sph*I site of pZero (Invitrogen).

Sequencing and analysis of clones

PCR with vector-specific primers was performed on individual bacterial colonies containing cloned concatemers to determine insert length. Only PCR products >500 bp, which should contain at least 15 tags, were selected for sequence analysis. Direct sequencing of PCR products was performed using the BigDye Primer Kit (Perkin Elmer) and analysed using a 377 ABI automated sequencer (Perkin Elmer) according to the manufacturer's instructions. Sequence files were analysed using the SAGE program group (12,15).

Reverse northern blot analysis

RT–PCR products of six chosen genes corresponding to SAGE tags (novel G protein-coupled receptor, myosin light chain, cofilin, Stat5b, GAPDH and melatonin-related receptor) were generated. The primer sequences used to generate the RT–PCR products are listed below: novel G protein-coupled receptor, 5′-CTGAACGTC-TGTGTCATCGC-3′ and 5′-AACACATTGCAGCCAGTGC-3′; myosin light chain, 5′-TCTCCTCTTCGACAGAACCG-3′ and 5′-TCAACCTGATGTGTGTGCC-3′; cofilin, 5′-TTCGCAAGTC-TTCAACGCC-3′ and 5′-TGACCTCCTCGTAGCAGTTAGC-3′; Stat5b, 5′ -CTCCAGAACACGTATGACCG 3′ and 5′-CTTCTC-GATGATGAACGTGC-3′; GAPDH, 5′-ATTGTTGCCATCAA-CGACC-3′ and 5′-ATTGAGAGCAATGCCAGCC-3′; melatonin-related receptor, 5′-ACTGTTCTGGATGTCCTGCC-3′ and 5′-TCAGGATT CTGTCCAGCTGG-3′.
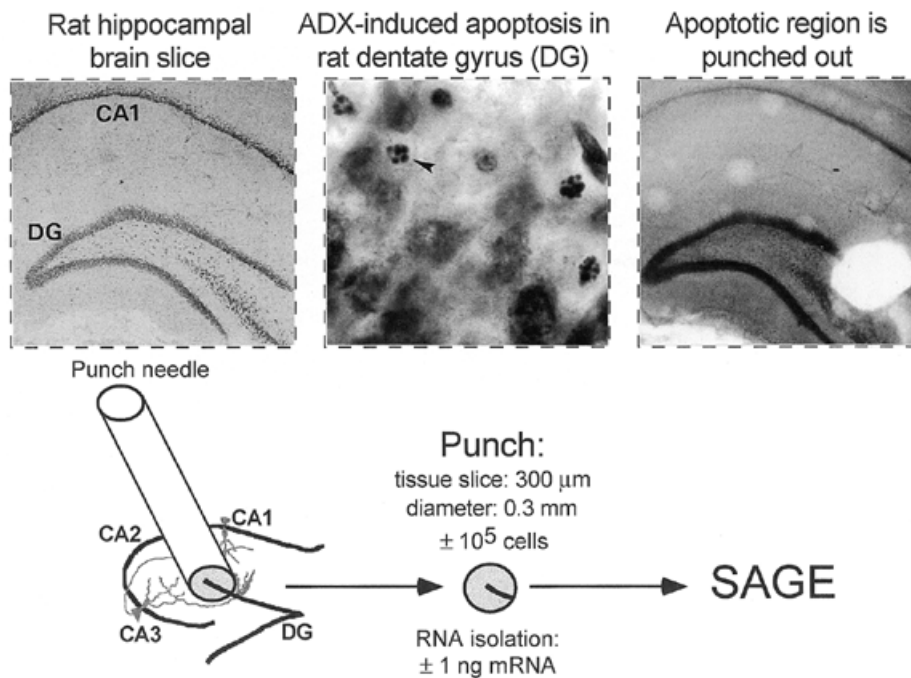
Figure 1. Cresylviolet-stained tissue sections of rat hippocampus. The 75 μm section (left) localized immediately adjacent to the 325 μm section was stained prior to punching and facilitates identification of the correct anatomical region within the unstained 325 μm section. A higher magnification of the inner blade of the dentate gyrus of an adrenalectomised rat shows the presence of multiple neurons with a clearly apoptotic morphology, an example of which is marked with an arrow (middle). After removal of this region with a punch needle, the 325 μm slice was stained to check the location of the punch (right). RNA isolated from correctly localised punches was subsequently used as input in the microSAGE procedure.

Equimolar amounts of the PCR products (500 ng of a 1 kb product) were denatured by adding 0.4 M NaOH/10 mM EDTA and heating to 100°C for 10 min. The denatured DNA samples were subsequently applied to a Bio-Dot Microfiltration apparatus (Bio-Rad) and dot-blotted onto Hybond N+ according to the manufacturer's recommendations. Membranes were hybridised with [α-^{32}P]dCTP-labelled cDNA derived from a single dentate gyrus punch using the Multiprime DNA labelling system (Amersham) and standard protocols (21). Hybridised and washed dot-blots were analysed using a PhosphorImager (Molecular Dynamics).

RESULTS

Isolated removal of a specific brain region for expression profiling

We are interested in how glucocorticoids affect the morphology and function of the hippocampus, in particular with respect to their role in adrenalectomy-induced apoptosis in the rat dentate gyrus (22). Since these effects are subfield-specific, e.g. the effects in the CA1 region are fundamentally different than in the dentate gyrus (23), we have developed a method which allows expression profiling in specifically removed apoptotic subfields of the inner blade of the dentate gyrus (Fig. 1, middle). The region of interest is specifically removed from a 325 μm rat brain slice by punching it out with a hollow needle (0.3 mm in diameter) according to the Palkovits punch out method (20). We chose the latter to microdissect the hippocampus, since it is relatively simple to perform and does not require any specialised microdissection equipment. Punching is performed on frozen unfixed and

unstained material. Therefore, prior to punching, tissue sections cut on either side of the slice are stained to facilitate recognition of the dentate gyrus and thus removal of the correct region. In addition, the stained sections are checked for the presence of apoptosis. Combined with post-staining of the remainder of the slice after punching, only punches containing the region of interest and with a sufficient degree of apoptosis are included in the further procedure. Figure 1 shows an example of a correctly punched region of the dentate gyrus (Fig. 1, right).

Total RNA isolated from a single punch was used as input material in a modified SAGE procedure. We estimate that a dentate gyrus punch contains a maximum of 10^5 cells. The amount of mRNA present is at most 1–5 ng, which is a factor of 500–5000 less than has been described so far as required input in the SAGE procedure.

Ditag amplification and isolation

In microSAGE (in contrast to SAGE) a limited number of additional PCR cycles (re-PCR) are performed on the excised ditag using the same primers in order to generate sufficient ditag for subsequent manipulations, a consequence of using minute amounts of input RNA. After an initial PCR of 25–28 cycles, the PCR products are size-separated on a gel and the region around 100 bp containing the amplified ditag (Fig. 2) is excised. After purification of the DNA from the gel slice, the required number of cycles of re-PCR is empirically determined, but restricted to a minimum, since the probability of PCR-based artefacts increases with the number of PCR cycles performed. An example of the initial PCR amplification and the re-PCR are shown in Figure 2A.
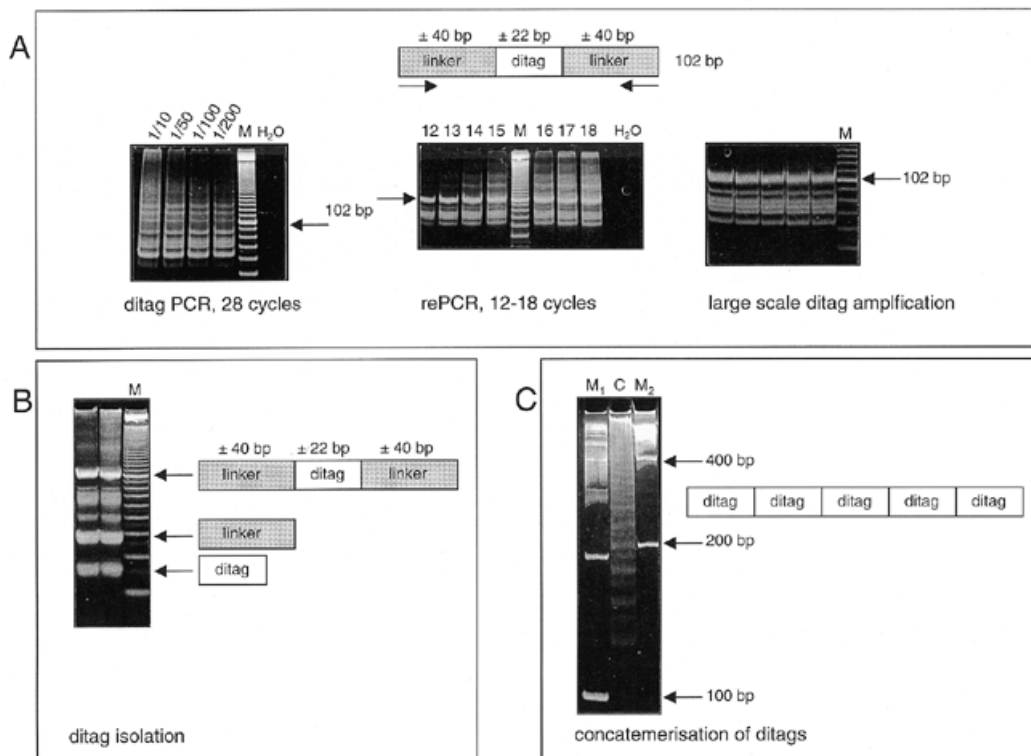
Figure 2. Ethidium bromide-stained polyacrylamide gels [12% in (A) and (B), 8% in (C)] showing examples of several steps in the (micro)SAGE procedure. (**A**) PCR amplification of the ditag. Shown are 28 cycles of PCR of various dilutions (1/10, 1/50/ 1/100 and 1/200) of 1 µl of the ligated ditag derived from punch material and a negative control performed on H_2O (left). The 102 bp band corresponding to the amplified ditag is faintly visible among several other background bands. After excision of the ditag band and extraction of the DNA, a series of PCRs with varying number of cycles (in this case 12–18 cycles) is performed to determine the optimal number of cycles of re-PCR (middle). The negative control performed on H_2O is amplified for 30 cycles. After large-scale re-PCR (in this particular case 12 cycles of re-PCR were considered optimal) the PCR products are concentrated and run on a preparative gel from which the 100 bp ditag band is excised (right). (**B**) After digestion with *Nla*III to cleave off the linkers, the small ditag of 22–26 bp is excised and purified. (**C**) The isolated ditags are ligated to concatemers that are size-separated on a polyacrylamide gel. The regions of the gel containing concatemers ranging from 400 to 800 bp and >800 bp are excised, after which the purified concatemers are cloned in pZero. M, 10 bp ladder; M_1, 100 bp ladder; M_2, 200 bp ladder; C, concatemers.

Examples of the subsequent steps of the procedure, i.e. ditag isolation and concatenation, are given in Figure 2B and C, respectively.

A microSAGE expression profile of a single dentate gyrus punch

Using the procedure described above, a partial expression profile was obtained from a single dentate gyrus punch. After determination of the length of cloned ditag concatemers by PCR, clones containing an expected minimum of 15 tags were selected for sequence analysis. After sequencing 128 selected clones, a total of 1497 ditags was obtained. Of these 1497 ditags, 924 were unique, while 176 were encountered more than once. The repeated ditags were only included once in the analysis by the software to avoid bias due to PCR-based artefacts derived from preferential amplification of particular ditag species. Of the remaining 1100 ditags, 2200 tags could be extracted. 109 tags were discarded because they corresponded to linker sequences (see Discussion), whereas 299 tags were removed from the collection due to ambiguities in the sequence, leaving a total of 1792 tags. Among the remaining 1792 tags there were 1242 different species, ranging in frequency from 1 to 116. Sixteen percent of the tags were encountered more than once, while the

vast majority of the tags (84%) were encountered only once (Table 2).

Comparison with rat sequences in GenBank (Release 106.0, April 1998) gave hits for 55 of the 1242 tags (4%) (Table 3). Of these, 31 were hits with known rat genes and 24 with ESTs or other cDNA clones. This relatively low percentage of hits is due to under-representation of rat sequences compared with human and mouse sequences in public databases. For example, comparison with mouse sequences in GenBank (Release 99.0, February 1997) gave 206 hits (17%), almost four times as many as with rat but still considerably lower than reported previously in another SAGE study, where 54% of human tags matched GenBank entries (18). Twenty-three tags had hits with both rat and mouse sequences, and in 13 cases (57%) the hit was with the exact mouse homolog of the rat gene. Therefore, the lack of rat GenBank entries can be partly compensated by comparing rat tags to mouse sequences, since there is an ~50% chance of having a hit with the mouse homolog of the missing rat entry.

Some caution is due when a SAGE tag matches with an EST. In contrast to mRNA sequences, many of the EST sequences deposited in GenBank are numbered from the 3′-end towards the 5′-end. The tag sequences extracted by the SAGE software will in this case be adjacent to the most 5′ *Nla*III site in a cDNA sequence instead of the most 3′ *Nla*III site. In addition, the

Table 2. SAGE abundancy classes

Frequency	Total number of tags		Number of different tags		GenBank hits (rat)		GenBank hits (mouse)	
>100	116	(7%)	1	(0.1%)	0	(0%)	0	(0%)
50 and ≤ 100	0	(0%)	0	(0%)	0	(0%)	0	(0%)
10 and ≤ 50	85	(5%)	6	(0.5%)	1	(0.1%)	1	(0.1%)
> 5 and ≤ 10	99	(6%)	13	(1%)	4	(0.3%)	1	(0.1%)
>1 and ≤ 5	451	(25%)	181	(15%)	8	(0.6%)	28	(2%)
1	1041	(58%)	1041	(84%)	42	(3%)	176	(14%)
Total	1792	(100%)	1242	(100%)	55	(4%)	206	(17%)

The tags are divided into different groups according to frequency of appearance among the 1792 tags comprising the expression profile. The number of tags giving a hit with an entry in GenBank (rat or mouse) is listed per abundancy class.

Table 3. Hits of tags with rat GenBank entries

Tag	Gene (Accession no.)	Frequency	Percent
CCTGTAATCC *	IFNgamma-induced astrocyte EST (H39351); EST (AA859834)	12	0.67%
GTGAAACCCC	novel G protein-coupled P2 receptor (D63665)	10	0.56%
GACAAGGTCG	growth hormone (E00002)	7	0.39%
TACCACCTTT	myosin light chain (M12022)	7	0.39%
AAAAAAAAA *	calcineurin A (M29275) ; aldehyde dehydrogenase (AF001898); orexin receptor-2 (AF041246); glutamic acid decarboxylase (X57573)	6	0.33%
AAGGAGATGG *	ribosomal protein L31 (X04809); hypoxanthine guanine phosphoribosyltransferase (AF009655)	3	0.17%
GAAGCAGGAC	cofilin (X62908)	3	0.17%
AACGAGGAAT ■	18S rRNA (M11188)	2	0.11%
CACAGAGTCC	EST (AA817914)	2	0.11%
GAAAAAAAAA *	EST (AA859996); EST (AA899266)	2	0.11%
CCTGTAACCC	unknown protein PIPPin (X89962)	2	0.11%
GAGGAGGTGG	EST (AA875282)	2	0.11%
TATATATAGT	Taq I-298 repetitive sequence,clone ID 4 (X06261)	2	0.11%
AAATACAAGA	EST (AA818063)	1	0.05%
AATCAACCCG ■	mitochondrial genes for 16S rRNA and tRNAs (V00681)	1	0.05%
AATCGCTTCT	transcription factor Stat5b (X97541)	1	0.05%
ACAAATGACA	EST (AA858803)	1	0.05%
ACTGTGCCTG	e17-28.1.3-5' Lambda Zap Express Library (R82902)	1	0.05%
AGAAGGATGT	fibrinogen B-beta-chain (M27220)	1	0.05%
AGGACAAATA	pancreatic Islet cDNA similar to cytochrome B (C06585)	1	0.05%
AGTTTTACAA	proteasomal ATPase (D50695)	1	0.05%
ATCAGTATAG	EST (AA819519)	1	0.05%
ATGAGCTGTC	EST (AA819034)	1	0.05%
ATGTATGGGG	EST (AA819181)	1	0.05%
CACATTGGAG	vitamin D binding protein (M60203)	1	0.05%
CACTTTTTTT	EST (AA817788)	1	0.05%
CCAATGGTCC	cysteine proteinase inhibitor cystatin C (X16957)	1	0.05%
CCAGAGGCTG	EST (AA899981)	1	0.05%
CCTGTGGTTC	vasopressin V1b receptor (D45400)	1	0.05%
CCTTTGCCCA	UV73 (U12266)	1	0.05%
CTTCTGATAA	triosephosphate isomerase (L36250)	1	0.05%
GAAGTCGGAA ■	28 S ribosomal RNA (M63916)	1	0.05%
GATGCCCCCC	pancreatic Islet cDNA similar to cytocrome oxidase I (C06520)	1	0.05%
GATGTGGTTT	EST (AA901228)	1	0.05%
GATTCCGTGA	ribosomal protein L37 (X66369)	1	0.05%
GCAACAACAC	EST (AA900974)	1	0.05%
GCCACCCCCT	EST (AA859401)	1	0.05%
GCCTCCAAGG	glyceraldehyde-3-phosphate dehydrogenase (M29341)	1	0.05%
GGATTCGGTC	acidic ribosomal protein P0 (Z29530)	1	0.05%
GGCTGGGGGC	profilin (X96967)	1	0.05%
GGGCTGGGGT	EST (AA875618)	1	0.05%
GTACTCTACT	leukocyte common antigen related protein (M60103)	1	0.05%
GTGGTGTGCT	polymeric immunoglobulin receptor (X15741)	1	0.05%
GTTTAAAATT	K$^+$ channel (M81784)	1	0.05%
TAAAAAAAAA *	trachea (Gang Li) cDNA clone 7A3 (U82289); intrinsic factor-B12 receptor precursor (CUBILIN) (AF022247)	1	0.05%
TAACTTTTGG	cDNA similar to fisp-12 (C06571)	1	0.05%
TATATATAGC	EST (AA818148)	1	0.05%
TGCAGATGAG	Ig germline V-lambda chain gene V-region (M17092)	1	0.05%
TGGTTGTGGG	EST (AA859565)	1	0.05%
TGTCTGTGGA	16 kDa subunit of vacuolar H(+)-ATPases (D10874)	1	0.05%
TGTTATTTAA	cell surface protein (X01785)	1	0.05%
TTAAAAAAAA *	connexin 40 (AF021806); EST (AA859029)	1	0.05%
TTCTCTACCA	EST (AA819729)	1	0.05%
TTTCCAGTTA	EST (AA818486)	1	0.05%
TTTGTGATCT	melatonin-related receptor (U52218)	1	0.05%

Listed are the tag sequence, the absolute tag frequency and frequency as a percentage of all 1792 tags and the matching GenBank entry with accession number. The tags matching with more than one gene are marked with an asterisk. The tags derived from rRNA are marked with a black square. The tags used for validation of the profile with reverse northern are underlined.
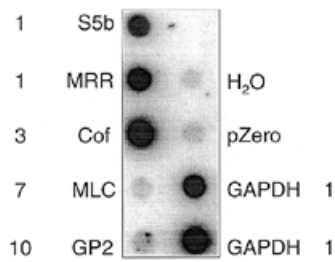
Figure 3. Reverse northern blot of RT–PCR products corresponding to transcripts identified by microSAGE, hybridised with ^{32}P-labelled cDNA of a single dentate gyrus punch. S5b, Stat5b (729 bp); MRR, melatonin-related receptor (215 bp); Cof, cofilin (372 bp); MLC, myosin light chain (453 bp); GP2, G protein-coupled P2 receptor (901 bp); GAPDH, glyceralde-hyde-3-phosphate dehydrogenase, both GAPDH probes overlap but have different lengths of 825 bp (lower spot) and 366 bp (upper spot), respectively; pZero, plasmid DNA, negative control for non-specific binding of probe to DNA; H$_2$O, H$_2$O as negative control for non-specific binding to membrane. The absolute frequencies of the tags are listed next to the abbreviated gene.

extracted tag will be on the wrong side of the *Nla*III site, resulting in a false match.

Of the 55 tags with GenBank matches, six matched perfectly with more than one gene (Table 3, tags marked with an asterisk). In four of these cases the tag consisted of a low-complexity sequence with a high percentage of A residues which could be part of a poly(A) tail, explaining the match with multiple genes.

Since 1242 tags represent the mere tip of the iceberg of transcripts present in a cell or tissue, these tags are most likely derived from abundant transcripts encoding the basic structural cell elements. Accordingly, the matched tags included the housekeeping gene GAPDH and genes for several cytoskeleton-associated proteins such as myosin and the actin-binding proteins cofilin, profilin and vitamin D-binding protein (24). In addition, genes encoding proteins involved in energy metabolism and protein synthesis were encountered among the tags, for example cytochrome B, cytochrome oxidase I and various ribosomal proteins. Also, some of the identified tags corresponded with genes known to be abundantly expressed in hippocampus, or more generally in brain, like the V1b vasopressin receptor (25), a brain-specific K$^+$ channel (26) and a melatonin-related receptor (27). In addition, some of the tags were derived from rRNAs (Table 3, tags marked with a square), most likely caused by incomplete enrichment for polyadenylated RNAs.

Validation of the microSAGE expression profile

We used reverse northern blotting (28–30) to validate the data obtained with microSAGE. Six SAGE tags with GenBank hits (Table 3, underlined tags: novel G protein-coupled receptor, myosin light chain, cofilin, Stat5b, GAPDH and melatonin-related receptor) were selected, the corresponding sequences retrieved and primer pairs designed for RT–PCR. Dot blots containing equimolar amounts of the six RT–PCR products were hybridised with radiolabelled cDNA derived from a single dentate gyrus punch. Of the selected six genes, four hybridised strongly, including GAPDH, a housekeeping gene known to be abundantly expressed (Fig. 3).

DISCUSSION

The ability to look into a cell or tissue at a given time point to see which set of genes is expressed and to assess their relative levels, has enormous potential for the enhancement of our understanding of how cells work under normal conditions or how they become diseased. The SAGE method is a powerful expression profiling tool, allowing qualitative and quantitative analysis of thousands of transcripts simultaneously. However, a disadvantage of SAGE is the relatively large amount of input RNA that is necessary, making the method unsuitable for analysis of gene expression profiles in small tissue samples or microdissected parts of complex heterogeneous tissues consisting of multiple cell types.

To enable application of SAGE to small quantities of tissue, we have made several modifications to the original procedure (12) (Table 1). The modifications mostly involve the first steps of the procedure, from RNA isolation to PCR, but leave the basic principles of SAGE unaltered. The original SAGE protocol is characterised by many sequential steps, each followed by a phenol–chloroform extraction and an ethanol precipitation to inactivate and remove enzymes and to purify and concentrate the material for use in the following reaction. However, these extraction and precipitation steps are renowned for the concomitant loss of material that can occur, especially if the amount of starting material is low. In our modified protocol, all steps from RNA isolation to tag release are performed in a single tube in which the RNA, and later the cDNA, remains immobilised to the wall of the tube by means of streptavidin–biotin binding (Table 1). This obviates the need to perform a phenol–chloroform extraction and ethanol precipitation between each subsequent step. Enzymes of previous reactions are now simply removed by heat inactivation and disposal of the solution and after washing and a change of buffer the next reaction can be performed in the same tube. Consequently, the most important advantage of this single-tube procedure is a reduction of the number of manipulations and reduction of the accompanying loss of material. Furthermore, total RNA is used rather than poly(A)$^+$ RNA, obviating the need for an additional mRNA extraction step. Instead, the poly(A)$^+$ fraction is directly bound to the streptavidin-coated wall of the tube via annealing to a biotinylated oligo(dT) primer, which also serves as a primer in the subsequent cDNA synthesis. Another difference with the original protocol is that a limited number of cycles of re-PCR are performed on the gel-purified ditag band to generate sufficient ditag. From the PCR onward the protocol is essentially identical to the original. Using this modified procedure, called microSAGE, an expression profile can be obtained from as little as 1–5 ng of mRNA, allowing expression profiling in small tissue specimens or microdissected parts of complex heterogeneous tissues.

Quantitative validation of the obtained SAGE profile is difficult when using limited amounts of tissue, since insufficient RNA can be isolated from a single punch to perform, for example, northern blot analysis. Reverse northern blotting is a sensitive method to compare expression levels of different transcripts between two or more mRNA pools, but is less suitable for comparing expression levels within a single mRNA pool because of differences in length and hybridisation efficiency between probes. The observation that only four of the six tags validated with reverse northern gave a hybridisation signal could be due to the fact that the intensity of hybridisation is not proportional to the tag frequency in the microSAGE analysis or that it is not justified to directly compare hybridisation signals within a mRNA pool. The two overlapping

GAPDH probes with different lengths, which do not exhibit exactly the same hybridisation intensity, exemplify this. Comparison of hybridisation intensities of the same probe under different conditions, e.g. in punch material from ADX rats compared with normal controls, using optimised hybridisation conditions is preferable. Another factor complicating the validation is that the 1792 tags characterised here form a random coincidental selection from the entire expression pool, representing only a small fraction of the estimated 50 000 different transcripts in a dentate gyrus punch. Therefore, it is most unlikely that the obtained distribution of tags paints an accurate quantitative picture of gene expression, since this would require sequencing of at least 10-fold more tags.

The use of minute amounts of starting material demands an amplification step to enable experimental manipulation. In the original SAGE procedure 25–28 cycles of PCR are performed to amplify the pool of ditags. Although this PCR step should be relatively free of bias since all ditags are of approximately equal length, it cannot be completely prevented that some ditag species are still preferentially amplified. This still does not jeopardise the quantitative aspect of the SAGE data, because the software, which counts each exclusive ditag combination only once, excludes any PCR artefacts. A high percentage of ditags excluded from analysis for this reason, however, does cause a reduction in the average number of analysed tags obtained per clone. The incidence of these artefacts increases exponentially with the number of PCR cycles performed. In particular, a single ditag species consisting of linker sequences had a high frequency among the excluded ditags (TCCCCGTACANNNTTAATAGGGA) (data not shown). At present it is not clear how this ditag is generated. It is therefore advisable to perform multiple parallel PCR reactions of fewer cycles each in order to generate sufficient amounts of ditag, rather than to push the PCR amplification to the limit of maximum yield.

The high percentage of different tags encountered only once (84%) is an indication of the high complexity of gene expression in the brain, even in a dentate gyrus punch which has a much reduced heterogeneity of cell types compared with the whole hippocampus. Striking is the fact that the most abundant tags mostly represent unknown genes. In general, the more abundant genes are well represented in GenBank. Explanations for this could be that these tags perhaps are derived from the so far unsequenced 3′-untranslated region of known genes or represent genes specifically expressed at a high level in dentate gyrus, since little is known about the overall gene expression in this subfield of the hippocampus.

In conclusion, we describe a modified SAGE procedure, microSAGE, suitable for expression profiling in limited amounts of tissue. We demonstrate the feasibility of microSAGE by obtaining an expression profile of a single dentate gyrus punch, containing a factor of at least 500–5000 less RNA than normally required for SAGE. This broadens the applications of SAGE enormously.

ACKNOWLEDGEMENTS

REFERENCES

1 Watson,J.B., Battenberg,E.F., Wong,K.K., Bloom,F.E. and Sutcliffe,J.G. (1990) *J. Neurosci. Res.*, **26**, 397–408.
2 Travis,G.H. and Sutcliffe,J.G. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 1696–1700.
3 Adams,M.D., Kelley,J.M., Gocayne,J.D., Dubnick,M., Polymeropoulos,M.H., Xiao,H., Merril,C.R., Wu,A., Olde,B., Moreno,R.F. *et al.* (1991) *Science*, **252**, 1651–1656.
4 Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M., Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C. (1992) *Nature*, **355**, 632–634.
5 Becker,K.G., Mattson,D.H., Powers,J.M., Gado,A.M. and Biddison,W.E. (1997) *J. Neuroimmunol.*, **77**, 27–38.
6 Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A., Fuldner,R.A.,Marmaros,S., Glodek,A., Gocayne,J.D., Adams,M.D., Kerlavage,A.R., Fraser,C.M. and Venter,J.C. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 8303–8307.
7 Liang,P. and Pardee,A.B. (1992) *Science*, **257**, 967–971.
8 Liang,P., Averboukh,L. and Pardee,A.B. (1993) *Nucleic Acids Res.*, **21**, 3269–3275.
9 McClelland,M., Mathieu-Daude,F. and Welsh,J. (1995) *Trends Genet.*, **11**, 242–246.
10 Vreugdenhil,E., de Jong,J., Schaaf,M.J., Meijer,O.C., Busscher,J., Vuijst,C. and de Kloet,E.R. (1996) *J. Mol. Neurosci.*, **7**, 135–146.
11 Welsh,J., Chada,K., Dalal,S.S., Cheng,R., Ralph,D. and McClelland,M. (1992) *Nucleic Acids Res.*, **20**, 4965–4970.
12 Velculescu,V.E., Zhang,L., Vogelstein,B. and Kinzler,K.W. (1995) *Science*, **270**, 484–487.
13 Adams,M.D. (1996) *Bioessays*, **18**, 261–262.
14 Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C. (1993) *Nature Genet.*, **4**, 256–267.
15 Velculescu,V.E., Zhang,L., Zhou,W., Vogelstein,J., Basrai,M.A., Bassett,D.E., Hieter,P., Vogelstein,B. and Kinzler,K.W. (1997) *Cell*, **88**, 243–251.
16 Madden,S.L., Galella,E.A., Zhu,J., Bertelsen,A.H. and Beaudry,G.A. (1997) *Oncogene*, **15**, 1079–1085.
17 Polyak,K., Xia,Y., Zweier,J.L., Kinzler,K.W. and Vogelstein,B. (1997) *Nature*, **389**, 300–305.
18 Zhang,L., Zhou,W., Velculescu,V.E., Kern,S.E., Hruban,R.H., Hamilton,S.R., Vogelstein,B. and Kinzler,K.W. (1997) *Science*, **276**, 1268–1272.
19 Simone,N.L., Bonner,R.F., Gillespie,J.W., Emmert-Buck,M.R. and Liotta,L.A. (1998) *Trends Genet.*, **14**, 272–276.
20 Palkovits,M. (1973) *Brain Res.*, **59**, 449–450.
21 Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22 De Kloet,E.R., Vreugdenhil,E., Oitzl,M.S. and Joëls,M. (1998) *Endocrine Rev.*, **19**, 269–301.
23 Sloviter,R.S., Valiquette,G., Abrams,G.M., Ronk,E.C., Sollas,A.L., Paul,L.A. and Neubort,S. (1989) *Science*, **243**, 535–538.
24 Nishida,E., Maekawa,S. and Sakai,H. (1984) *Biochemistry*, **23**, 5307–5313.
25 Lolait,S.J., O'Carroll,A.M., Mahan,L.C., Felder,C.C., Button,D.C., Young,W.S., Mezey,E. and Brownstein,M.J. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 6783–6787.
26 Drewe,J.A., Verma,S., Frech,G. and Joho,R.H. (1992) *J. Neurosci.*, **12**, 538–548.
27 Laudon,M., Nir,I. and Zisapel,N. (1988) *Neuroendocrinology*, **48**, 577–583.
28 Francia,G., Mitchell,S.D., Moss,S.E., Hanby,A.M., Marshall,J.F. and Hart,I.R. (1996) *Cancer Res.*, **56**, 3855–3858.
29 Ghosh,P.K., Baskaran,N. and van den Pol,A.N. (1997) *Dev. Brain Res.*, **102**, 1–12.
30 Lanahan,A., Lyford,G., Stevenson,G.S., Worley,P.F. and Barnes,C.A. (1997) *J. Neurosci.*, **17**, 2876–2885.