

# Substantially enhanced cloning efficiency of SAGE (Serial Analysis of Gene Expression) by adding a heating step to the original protocol

M. Kenzelmann\* and K. Mühlemann

Institute of Medical Microbiology, University of Bern, Friedbühlstrasse 51, 3010 Bern, Switzerland

Received October 16, 1998; Revised and Accepted December 1, 1998

## ABSTRACT

**The efficiency of the original SAGE (Serial Analysis of Gene Expression) protocol was limited by a small average size of cloned concatemers. We describe a modification of the technique that overcomes this problem. Ligation of ditags yields concatemers of various sizes. Small concatemers may aggregate and migrate with large ones during gel electrophoresis. A heating step introduced before gel electrophoresis breaks such contaminating aggregates. This modification yields cloned concatemers with an average size of 67 tags as compared to 22 tags by the original protocol. It enhances the length of cloned concatemers substantially and reduces the costs of SAGE.**

The Serial Analysis of Gene Expression (SAGE) method (1) reduces cDNA molecules that have been reverse transcribed from polyadenylated mRNA by a series of enzymatic manipulations to tags of 10–13 bp. Each tag represents one mRNA. Tags are ligated to form concatemers that are cloned and sequenced. Comparing the sequence information from the tags with the GenBank database provides qualitative information about transcribed genes. The frequency of a specific tag within the SAGE tag population correlates with its relative abundance in the cell and gives quantitative information about expressed genes. The SAGE method therefore allows for cataloging and comparison of expressed genes under various physiological conditions (1–7).

The statistical analysis of the expression of thousands of different genes requires the sequencing of a large number of tags. Despite its elegant approach, the efficiency of the original SAGE protocol was limited by a restriction of the length of cloned concatemers to an average of ~22 tags. Powell recently presented a modified protocol which increased the number of tags per clone by 43% from an average of 21 to 35 tags by the elimination of contaminating linker molecules (8). Instead of using unlabeled primers as in the original SAGE protocol by Velculescu *et al.* (9), she used biotinylated primers to prepare bulk PCR reactions of

SAGE ditags. Ditags were then efficiently separated from the PCR primers by streptavidin coated magnetic beads instead of a gel purification step. Here we describe another modification of the original SAGE protocol that increases the length of cloned concatemers substantially.

Referring to steps 10, 11 and 12 of the detailed SAGE protocol (9) bulk amounts of tags are obtained by PCR amplification of tag dimers, so called ditags (1 ditag = 26 bp). The linkers used to create these amplicons are removed after PCR amplification by a gel purification step. Purified ditags are then ligated to form concatemers. The concatemers were separated according to size on an 8% polyacrylamide gel where they form a smear ranging from molecules of ~300 bp to several kb in size. Cloning of selectively cut out bands in the range of large size molecules nevertheless yields inserts with an average size of 300 bp (~22 tags). The reason for this size limit of clonable concatemers has remained unclear. In our hands it was unaffected by the use of the DNA condensing agent hexammine cobalt(III) chloride (data not shown).

We modified the protocol as follows. The concatemerization reaction of ditags (~300 ng of DNA per reaction) was performed as described in the original protocol (step 11) with the exception that we incubated the ligation reaction overnight. The ligation sample was then heated for 15 min at 65°C and quickly chilled on ice for 10 min. Concatemers were separated on an 8% polyacrylamide gel as described above. This heating step shifted the banding pattern in the gel towards concatemers of smaller sizes (Fig. 1). We compared the effect of the heating step on the cloning efficiency by dividing the sample into three size fractions: fraction 1 ranging from 700 to 1000 bp; fraction 2 comprising molecules of 1000 to 1600 bp; and fraction 3 ranging from 1600 to 2500 bp. Following the protocol of Velculescu *et al.* (9), we obtained clones with average insert sizes of ~300 bp (~22 tags per clone; Table 1). These results are in good agreement with those obtained by other groups (7,8). As demonstrated in Table 1, heating of concatemers before gel separation led to a substantial increase in the length of cloned concatemers from an average of 287 bp (22 tags) to 873 bp (67 tags).

\*To whom correspondence should be addressed. Tel: +41 31 632 3251; Fax: +41 31 632 3550; Email: kenzel@imm.unibe.ch

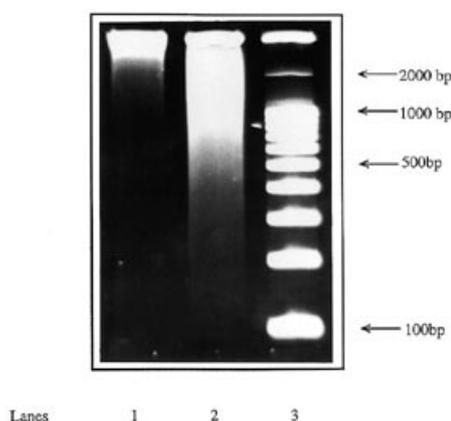
**Table 1.** Comparison of length of cloned concatemers obtained by the original and the modified SAGE protocol

	Average concatemer size [number of tags per clone]			
	Non-heated sample <sup>a</sup>		Heated sample <sup>b</sup>	
Fraction 1 (700–1000 bp)	137 (± 47)	[11 (± 3)]	669 (± 119)	[51 (± 10)]
Fraction 2 (1000–1600 bp)	208 (± 100)	[16 (± 8)]	993 (± 342)	[76 (± 27)]
Fraction 3 (1600–2500 bp)	542 (± 105)	[42 (± 8)]	1455 (± 225)	[112 (± 17)]
Average of fractions 1–3	287 (± 201)	[22 (± 16)]	873 (± 346)	[67 (± 27)]

Average results from two independent cloning experiments. Fractions 1–3 represent concatemer size ranges that were selectively cut out of the 8% polyacrylamide gel (step 11 of the detailed protocol) as outlined in the text. The average concatemer size reflects the length of the cloned concatemers without vector sequence (226 bp; vector pZerO, Invitrogen). Numbers in round brackets are standard deviations.

<sup>a</sup>Results were obtained by following the detailed SAGE protocol of Velculescu *et al.* (9). The average percentage of clones that contained no insert or an inappropriately small insert size was ~30% in all three fractions.

<sup>b</sup>Results were obtained using the modified protocol described in the text. The average percentage of clones that contained no insert or an inappropriately small insert size was 20–30% for fraction 1, 40–50% for fraction 2 and ~70% for fraction 3. The increasing proportion of false positives that parallels the increasing length of cloned DNA fragments is a common observation in cloning experiments and is most probably due to an increasing rate of vector self ligation.



**Figure 1.** Separation of concatemers by 8% polyacrylamide gel electrophoresis before cloning [step 11 of the original protocol by Velculescu *et al.* (9)]. Approximately 300 ng of gel purified ditags were used for the concatemerization reaction. Lane 1, non-heated concatemer sample according to the original protocol of Velculescu *et al.* (9); lane 2, heated concatemer sample; lane 3, 100 bp ladder molecular marker (GenSura). The DNA was stained with SYBR™ Green I (FMC BioProducts). Heating the concatemer sample leads to a shift in the banding pattern towards molecules of smaller size.

Our results suggest that the ligation of ditags to concatemers may be incomplete, yielding a large fraction of fragments of small size. These fragments may aggregate by hydrogen bonds between the 4 bp overhangs created by cleavage with the anchoring restriction enzyme *Nla*III. The hydrogen bonds may be further stabilized by the high amount of  $Mg^{2+}$  ions in the ligation buffer (final concentration: 10 mM  $MgCl_2$ ; 5× ligase buffer, Gibco BRL Life Technologies) and by inactivated enzyme molecules binding to the overhang sites of the concatemer fragments. These aggregates migrate through the gel as apparently large concatemer molecules. During cloning they disaggregate to their original smaller size. Heating the ligation sample before gel electrophoresis breaks hydrogen bonds and prevents contamination of the large size fraction by aggregates of smaller molecules. The same effect may

also be obtained in other experiments that involve ligation reactions and cloning.

For an optimal performance of SAGE (high throughput of tags with minimal sequencing costs) it is essential to obtain maximum information per clone. Powell (8) recently described a modification that overcomes the problem of contaminating linker molecules and increases the amount of tags per clone from an average of 21 to 35. Our modification of the SAGE protocol addresses the incomplete ligation during the concatemerization process and leads to an even greater increase of the number of tags per clone to 67. Both modifications may act synergistically if combined. The integration of Powell's and our modification into the original protocol of Velculescu *et al.* (9) optimizes the efficiency of SAGE by reducing the number of sequencing reactions.

## ACKNOWLEDGEMENTS

We thank B. Fartmann, J. Graf, T. Seebeck, R. Friis and M. Täuber for technical advice, valuable discussion, encouragement and support.

## REFERENCES

- Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) *Science*, **270**, 484–487.
- 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.
- 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.
- Madden, S.L., Galella, E.A., Zhu, J., Bertelsen, A.H. and Beaudry, G.A. (1997) *Oncogene*, **15**, 1079–1085.
- Polyak, K., Xia, Y., Zweier, J.L., Kinzler, K.W. and Vogelstein, B. (1997) *Nature*, **389**, 300–305.
- Hermeking, H., Lengauer, C., Polyak, K., He, T., Zhang, L., Thiagalingam, S., Kinzler, K.W. and Vogelstein, B. (1997) *Mol. Cell*, **1**, 3–11.
- Bertelsen, A.H. and Velculescu, V.E. (1998) *Drug Discov. Today*, **3**, 152–159.
- Powell, J. (1998) *Nucleic Acids Res.*, **26**, 3445–3446.
- Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1997) *Serial Analysis of Gene Expression: Detailed Protocol* (Version 1.0c), September 1997. Available from Johns Hopkins Oncology Center and Howard Hughes Medical Institute, 424 North Bond Street, Baltimore, MD 21231, USA; Fax: +1 410 955 0548.