*394–395* *Nucleic Acids Research, 1996, Vol. 24, No. 2* 

# RNA fingerprinting by molecular indexing

**Kikuya Kato**

Okayama Cell Switching Project, ERATO, JRDC, Pasteur Building 4F, 103-5 Tanakamonzencho, Sakyo-ku, Kyoto 606, Japan

Received November 2, 1995; Accepted November 27, 1995

Class IIS restriction enzymes, a subgroup of class II, cleave DNA at a precise location outside their recognition sites, and produce overhangs of unknown sequences (1). Molecular indexing is a series of techniques designed to characterize DNA fragments by these unknown sequences (2–4). I applied this principle for description of the total mRNA population using a 3′ end cDNA fragment generated by class IIS restiction enzymes (5). The method is based on the finding that *Escherichia coli* DNA ligase discriminates three nucleotides adjacent to the joining site. Fragments are discriminated by a library of 64 adaptors for all possible overhangs, and selected fragments are PCR-amplified using an adaptor–primer and an anchored oligo-dT primer. They are separated and displayed by a denaturing polyacrylamide gel electrophoresis. Comparing electropherograms from various sources of RNA, differentially expressed genes can be easily identified.

This method has several advantages over display techniques based on arbitrarily primed PCR (6,7). In particular, the method can display most genes with low redundancy. However, amplified fragments correspond to 3′ ends of mRNA, and there is not much chance of them to containing coding regions. Additional experiments are required to obtain encoded protein sequences. Here, I describe an alternative protocol for amplifying fragments from upstream regions.

The outline of the technique is as follows. The cDNA was at first digested by a class II restriction enzyme that generates an overhang of a defined sequence, and an adaptor cohesive to the end was ligated to it. The 5′ end of the cohesive end of the adaptor must be phosphorylated so that the adaptor sequence attaches the PCR template afterwards. The cDNA was digested by a class IIS restriction enzyme that produces a four nucleotide 5′ overhang. A total of 64 biotinylated adaptors were prepared for all possible overhangs. Each adaptor had a 5′ four nucleotide overhang, of which the outermost base was a mixture of A, C, G and T, and the three inner bases were one of all possible sequences. The cDNA was ligated to one of the 64 adaptors. Repeating the ligation with all the adaptors, restriction fragments that had ends created by both of the enzymes were divided into 64 subpopulations. After recovery with streptavidin-coated paramagnetic beads, the cDNA was treated with a dilute alkali. PCR amplification proceeded with this cDNA sample, using the adaptor–primers. The products were separated by denaturing polyacrylamide gel electrophoresis and the sizes of fragments were automatically recorded by a 373A sequencer (Perkin-Elmer), then analyzed by the 672 GeneScan software (Perkin-Elmer). Comparing electropherograms with different sources of RNA, specific expressed genes were easily

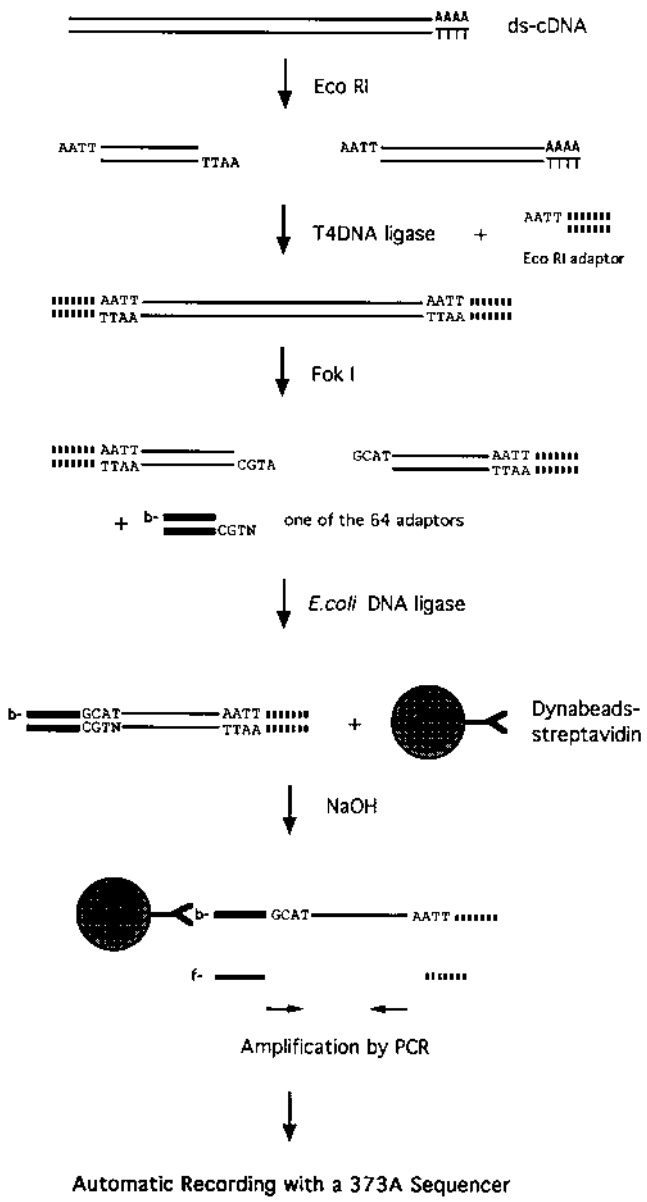


**Figure 1.** Outline of the technique. This figure shows a situation where *Eco*RI and *Fok*I were used as restriction enzymes.

* To whom correspondence should be addressed

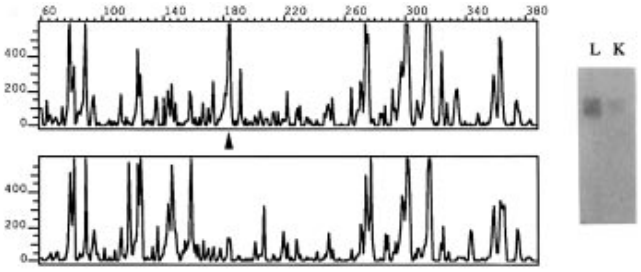


**Figure 2.** Example of electropherograms. (left) The top electropherogram is of mouse liver RNA, and the bottom is of mouse kidney RNA. The sequence of the overhang of the biotinylated adaptor is NCCC. (right) Northern hybridization. The probe corresponds to the peak marked by an arrow head. Lane L, total mouse liver RNA; lane K, total mouse kidney RNA. Northern hybridization experiment was proceeded as described (5). (Materials and Methods) The *Eco*RI adaptor consisted of 5′-P-AATTCTTAACCAGGCTGAACTTGC-TC-3′ and 5′-OH-GAGCAAGTTCAGCCTGGTTAAG-3′. The 5′ end of the cohesive end must be phosphorylated by T4 polynucleotide kinase. The adaptor library for class IIS restriction enzyme was the same as previously described (5). The double-stranded cDNA was synthesized as described (10). Usually, 3 µg RNA was converted to double-stranded cDNA with 200 U reverse transcriptase. It was digested by 5 U *Eco*RI for 1 h, and ligated to 5 pmol of the *Eco*RI adaptor in 20 µl 66 mM Tris–HCl (pH 7.5) containing 6.6 mM $MgCl_2$, 10 mM DTT, 0.1 mM ATP with 150 U T4 DNA ligase for at least 14 h. Following digestion by 5 U *Fok*I (TaKaRa) for 50 min, the sample was dissolved in 70 µl distilled water. One microliter of the cDNA sample was ligated to 1 pmol of one from the adaptor library in 10 µl 10 mM Tris–HCl (pH 8.0) containing 4 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 0.5 mM NAD, 1.2 mM EDTA, 0.005% BSA, 3 U *E.coli* DNA ligase (TaKaRa) at 16°C for at least 14 h. After recovering the ligated molecules with 0.05 mg of Dynabeads–streptavidin (Dynal) and treatment with 0.1 N NaOH, they were mixed with 10 µl of a reaction mixture consisting of 1× PCR buffer for the Stoffel fragment, 2.5 mM $MgCl_2$, 2 mM deoxynucleotide triphosphates, 5 pmol each of the adaptor primers (ClS, 5′-OH-GTACATATTGTCGTTAGAACGC-3′; λgt 10 forward primer, 5′-OH-GAGCAAGTTCAGCCTGGTTAAG-3′) and 1 U of Stoffel fragment (Perkin-Elmer). TAMRA-labelled ClS primer was used for fluorescent detection. PCR amplification consisted of 35 cycles of 94°C 30 s, 55°C 1 min and 72°C 1 min, followed by soaking for 20 min at 72°C. Three microliters of the amplified product was mixed with 5 µl of 10 mM Tris–HCl (pH 8.0), 10 mM $MgCl_2$, 50 mM NaCl, 0.2 mM dNTP and 0.5 U T4 DNA polymerase, and incubated at 37°C for 50 min. One fifth of the sample was loaded to a 6% denaturing polyacrylamide gel for electrophretic separation monitored by a 373A sequencer (Perkin-Elmer).

identified. Various combinations of enzymes and adaptors generate enough fingerprinting profiles to identify specific expressed genes. The strategy is schematically summarized in Figure 1.

As a trial, fingerprinting patterns from mouse liver and kidney were compared with some adaptors from the biotinylated adaptor library. In this case, *Eco*RI and *Fok*I were the class II and class IIS enzyme respectively. An example is shown in Figure 2 (left). The expression status of the fragment marked by an arrow head was confirmed by Northern hybridization (Fig. 2, right).

The method described here divides cDNA fragments with ends generated by both of enzymes into 64 subpopulations, and theoretically it displays all the fragments of this category without redundancy. Amplified fragments are not restricted to 3′ end regions, and have an increased likelihood of containing coding regions. This contrasts with the original indexing procedure (5) or differential display (6), which are intended to amplify 3′ end cDNA fragments. It also shares other advantages with the original indexing method over arbitrarily primed PCR. However, multiple sets of class II and class IIS enzymes have to be used when the entire mRNA population is to be examined. This is also true of differential display which needs serial sets of primers, and the major disadvantage compared the original indexing method, which divides the entire mRNA population into 576 subpopulations only with three class IIS restriction enzymes. Thus this technique is suited for quick sampling and the characterization of differentially expressed genes. The original method should be used for the complete description of expressed genes.

Recently, another display tehcnique named gene expresion fingerprinting (GEF) was introduced (8). This is also based on ligation mediated PCR, and reaction is not sensitive to amplification condition. The major disadavantage is that only abundant species are displayed because the number of amplified fragments applied to a single lane of a gel is ~2000.

The application of this method is not restricted to cDNA. Although similar techniques using T4 DNA ligase have been already proposed (4), the three nucleotide recognition by *E.coli* DNA ligase offers more precise base discrimination. There should be many applications for this technique in genome studies, for example, isolation of clones linked to restriction sites generated by rare cutting enzymes such as *Not*I (9).

## ACKNOWLEDGEMENT

The author thanks Prof. Hiroto Okayama for support.

## REFERENCES

1 Szybalski, W., Kim, S.C., Hasan, N. and Podhajska, A.J. (1991) *Gene*, **100**, 13–26.
2 Brenner, S. and Livak, K.J. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 8902–8906.
3 Smith, D.R. (1992) *PCR Meth. Appl.*, **2**, 21–27.
4 Unrau, P. and Deugau, K.V. (1994) *Gene*, **145**, 163–169.
5 Kato,K. (1995) *Nucleic Acids Res.* **23**, 3685–3690.
6 Liang, P. and Pardee, A.B. (1992) *Science*, **257**, 967–971.
7 Welsh, J., Chada, K., Dalal, S.S., Cheng, R., Ralph, D. and McClelland, M. (1992) *Nucleic Acids Res.*, **20**, 4965–4970.
8 Ivanova, N.B. and Belyavsky, A.V. (1995) *Nucleic Acids Res.*, **23**, 2954–2958.
9 Zabarovsky, E.R., Allikmets, R., Kholodnyuk, I., Zabarovska, V.I., Paulsson, N., Bannikov, V.M., Kashuba, V.I., Dean, M., Kisselev, L.L. and Klein, G. (1994) *Genomics*, **20**, 312–316.
10 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.