EXPERIMENTAL and MOLECULAR MEDICINE, Vol. 35, No. 6, 586-590, December 2003

An improved method for constructing a full-length enriched cDNA library using small amounts of total RNA as a starting material

Jung-Hwa Oh[1], Yong Sung Kim[1] and Nam-Soon Kim[1,2]

[1]Laboratory of Human Genomics
Division of Genomics and Proteomics
Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-333, Korea
[2]Corresponding author: Tel, 82-42-879-8112;
Fax, 82-42-879-8119; E-mail, nskim37@kribb.re.kr

Accepted 30 September 2003

Abbreviations: BAP, Bacterial alkaline phosphatase; ESTs, Expressed sequence tags; TAP, Tobacco acid pyrophosphatase

Abstract

We have developed an improved method for constructing a full-length cDNA library using small quantity of material by modifying the original oligo-capping method. In our devised method, total RNAs are used in sequential oligo-capping steps directly without preliminary mRNA purification. Using this method, we constructed full-length cDNA libraries from 100 μg of total RNA. These libraries contained 8×10^5 to 8×10^6 independent clones with average insert sizes of 2.0 kb. Moreover, the number of full-length cDNAs containing the translation initiation codon ATG in the constructed libraries was estimated to 60-70%. In addition, 54% of the known cDNAs had a longer 5' end than the corresponding genes in the public database. Our results show that the method can be effectively used to construct full-length enriched cDNA libraries, especially, if starting material is limited.

Keywords: cDNA library; full-length; oligo-capping; total RNA

Introduction

As a result of the extensive efforts put into the human genome project, a steady stream of ever-larger and more complex genomic data sets have been generated and stored in public databases (Lander *et al.*, 2001; Venter *et al.*, 2001). Identifying genes from these genome sequences by bioinformatic analysis alone is still not a trivial task, because of our limited knowledge of the rules of transcription and the processing of the transcripts. Since full-length cDNAs carry complete protein coding sequences and untranslated regions, they are indispensable for the identification of genes and for the determination of protein primary structure. In particular, full-length cDNAs represent a valuable resource for functional gene studies.

The methods for effectively producing full-length cDNA mainly involve selecting the cap structure, which is the specific structure of the 5'-end mRNA in eukaryotic cells. Methods such as, oligo-capping (Kato *et al.*, 1994; Maruyama and Sugano, 1994; Suzuki and Sugano, 2001), Biotinylated CAP Trapper (Carninci *et al.*, 1996) and CAPture (Edery *et al.*, 1995) have been known. However, all of them require a large amount of starting material, at least 5 to 10 μg of poly $(A)^+$ RNA, due to necessary multiple purification or precipitation steps to recover the mRNA cap structure. Thus, it is difficult to construct full-length enriched cDNA libraries by these methods if only small amount of biological materials is available as starting material. In many instances, availability of starting material for molecular analysis is limited. To construct cDNA libraries from such small amount, several PCR-based methods have been reported (Belyavsky *et al.*, 1989; Bertiol *et al.*, 1994; Lambert *et al.*, 1997; Peterson *et al.*, 1998). However, these methods are unsuitable for obtaining full-length cDNA clones because the cap structure of the mRNAs is not selected.

In this study, a method was developed for constructing a full-length cDNA library using small quantity of material by modifying the original oligo- capping method (Maruyama and Sugano, 1994) using total RNA instead of mRNA as a starting material. Using this method, a number of full-length enriched cDNA libraries were constructed successfully, and adopted as the method for in the 21C Functional Human Genome Project in Korea. We are hopeful that this method will become a useful tool for generating resources for the study of gene function and transcriptional regulation, especially, where an accurate data acquisition from a limited amounts of clinical sample will be a tremendous advantage.

Materials and Methods

Sample Source and Total RNA Preparation

Two different cell lines, SNU475-a hepatocellular carcinoma cell line, and SNU17-a cervical carcinoma cell line, were obtained from the Korean Cell Line Bank (http://cellbank.snu.ac.kr/, Korea). In addition, a normal liver tissue, N779227, was obtained from the College of Medicine, ChungNam National University, Korea. Total RNAs were isolated from cultured cells and tissues using a commercially available RNA isolation kit (ISOGEN, QIAGEN).

Construction of a full-length enriched cDNA library

Libraries were constructed by using a modification of Maruyama and Sugano's method (Maruyama and Sugano, 1994). 100 μg of total RNA was treated with 3 units of bacterial alkaline phosphatase (TaKaRa) in 100 μl of 100 mM Tris-HCl (pH 7.5), 2 mM DTT and 80 units of RNasin (Promega) at 37°C for 60 min. After phenol extraction and ethanol precipitation, the total RNA was treated with 100 units of tobacco acid pyrophosphatase (Waco) in 100 μl of 50 mM sodium acetate (pH 5.5), 5 mM EDTA, 10 mM 2-mercaptoehanol and 80 units of RNasin at 37°C for 60 min. Pre-treated total RNA was then ligated with 0.4 μg of 5'-oligoribonucleotide (5'-oligo: 5'-AGC AUC GAG UCG GCC UUG UUG GCC UAC UGG-3') using 250 units of RNA ligase (TaKaRa) in 100 μl of 50 mM Tris-HCl (pH7.5), 5 mM $MgCl_2$, 5 mM 2-mercaptoethanol, 0.5 mM ATP, 25% PEG 8,000 and 100 units of RNasin at 20°C for 3 h. After completing these oligo-capping reactions, mRNA was isolated using a commercial kit, QIAGEN Oligotex™. The synthesis of first-strand cDNA from the purified mRNA and cDNA amplification were performed as described by Maruyama and Sugano (1994). The amplified PCR Products were then digested with *Sfi*I, and cDNAs longer than 1.3 kb were ligated into *Dra*III-digested pCNS-D2 in an orientation-defined manner. The pCNS-D2 vector contains 5' *Eco*RI-*Dra*III-*Eco*RV-*Dra*III sites at multi cloning sites, which was achieved by modifying pCNS vector (GenBank Accession no. AF416744). The ligated cDNA was then transformed into *E. coli* Top10F' (Invitrogen) by electroporation (Gene Pulser II, BioRad).

Sequencing and sequence analysis

Plasmid DNAs were extracted using a MWG plasmidprep 96 (MWG AG Biotech). Sequencing reactions were performed on a GeneAmp PCR System 9700 thermal reactor (Perkin-Elmer) using a BigDye Terminator Sequencing kit and an ABI prism 3700 DNA analyzer (PE Applied Biosystems). Base-calling and quality assessment were performed using the phred program (Ewing *et al.*, 1998). ESTs with at least 100 bp after both vector and low-quality trimming were regarded as "high-quality" ESTs. Sequence similarities were searched for in the NCBI GenBank database and in the RefSeq database using BLASTN.

Results and Discussion

To construct a full-length enriched cDNA library using small amounts of biological material, an improved oligo-capping method was developed using total RNA as a starting material. The method strategy is illustrated in Figure 1. Total RNA extracted from biological samples, is subjected to sequential oligo-capping steps directly without preliminary mRNA purification. The cap structure of the intact mRNA in the total RNA is replaced by a 5'-oligomer in the oligo-capping steps, which consist of three enzymatic reaction steps. BAP hydrolyzes the 5'-end phosphate of the truncated mRNA, rRNA and tRNA in total RNA. TAP then removes the cap structure, leaving a phosphate at the 5'-end of intact mRNA. RNA ligase then ligates the 5'-oligomer to the 5'-end phosphate of the intact mRNA, which originally had the cap structure. The pretreated intact mRNA is purified from the total RNA and used directly as a template for the first-strand cDNA synthesis. The first-strand cDNA is amplified by PCR using specific primers containing the sequence of the 5'-end oligomer and of the dT adapter linker. Thus, amplified PCR products originate from only the mRNA containing the 5'-end oligomer, *i.e.*, ligated intact mRNA with the cap structure. The PCR products are then digested with restriction enzyme *SfiI* and cloned into *Dra*III digested pCNS-D2 vector. Finally, a full-length enriched cDNA library was constructed by transforming the ligated cDNA into *E. coli* Top10F'.

Based on this strategy, three full-length cDNA libraries-L4SNU475, L19N779227, and C1SNU17 was made using 100 μg of total RNA. From each library, $8\times10^5 \sim 8\times10^6$ independent clones were obtained from 100 μg of total RNA. 100 μg of total RNA assumed to contain about 0.5-5 μg of mRNA, based on the earlier study where total RNA was found to contain 0.5-5% of mRNA (Maniatis *et al.*, 1982; Albert *et al.*, 1994). In the originally described oligo-capping method, the size of the cDNA library was reported to be about 2.0×10^4/μg of mRNA with the use of over 5 μg of mRNA (Maruyama and Sugano, 1994; Suzuki *et al.*, 1997). As compared with the original method, our results indicate that 100 μg of total RNA is sufficient to produce a full-length enriched cDNA library, and that the sequential enzymatic reactions required for oligo-capping are performed efficiently even when total RNA is used. Furthermore, when 50 μg of starting total RNA was used, about 5×10^5

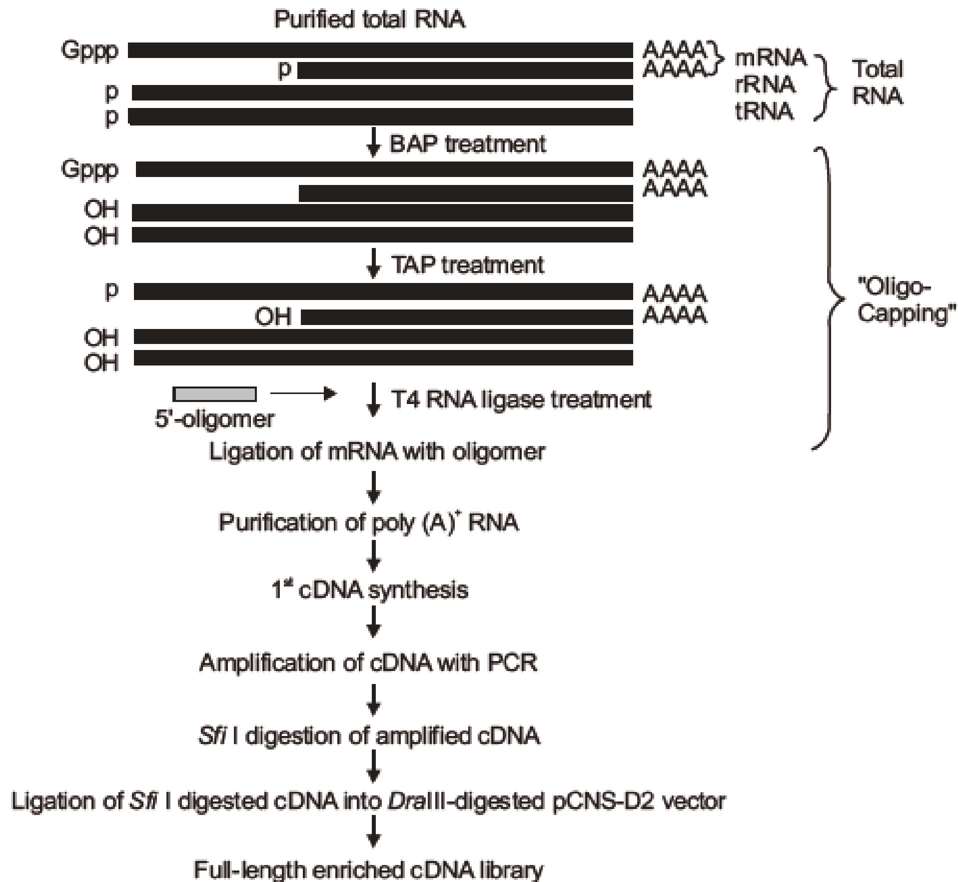

Figure 1. Overall strategy for constructing a full-length enriched cDNA library using small amount of total RNA. RNA molecules are represented by solid lines and 5'-oligo is represented by a gray box. Gppp, cap structure; p, phosphate; OH, hydroxyl; BAP, bacterial alkaline phosphatase; TAP, tobacco acid pyrophosphatase.

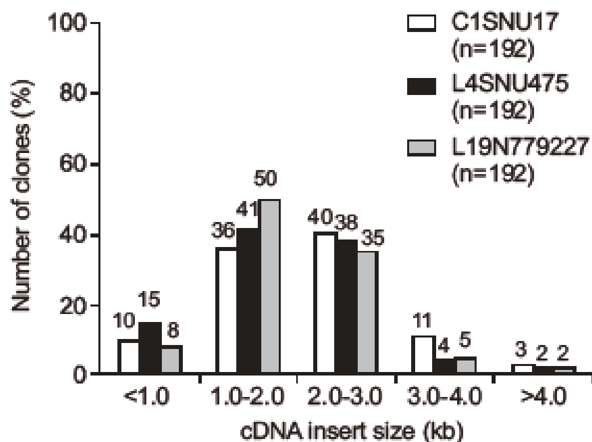

Figure 2. Size distributions of isolated cDNA inserts. 192 cDNA clones from each library were analyzed. The percentage of clones in each class is shown above each bar.

independent clones were obtained from each library (data not shown). These results demonstrate that the developed method can efficiently generate a full-length enriched cDNA library from small amounts of starting material.

To assess the quality of the constructed cDNA libraries, the lengths of cDNAs and the fullness ratios of the three constructed cDNA libraries were examined. At first, plasmids were prepared from randomly chosen 192 cDNA clones in each library, and digested with *Eco*RI and *Not*I. The cDNAs were inserted in the region between the *Eco*RI and *Not*I sites. As shown in Figure 2, most of the cDNA in the three cDNA libraries ranged from 1 kb to 4 kb, and averaged 2.0 kb. Longer cDNAs of over 4 kb, can be obtained using the developed method, though their frequencies are very low. These cDNA size distributions have been found in most cDNA libraries constructed using the previously reported methods, which is probably explained by the low quality of the starting material, an inability of these methods to

A. IFIT1

NM_001548
CB121237

B. IFIT4

NM_001549
CB150565

C. PPP4R2

NM_019853
CB150870

Figure 3. Sequence comparisons between the 5'-end sequences of our cDNA clones and those of corresponding genes in the human RefSeq database. The upper sequence represents the known gene in the existing database and the lower the corresponding cDNA sequence selected from our libraries. The first nucleotide in uppercase, which is located in the known gene, is numbered 1. Identical nucleotides obtained by aligning two sequences are marked with asterisks (*) below the pair of nucleotides. The middle sections of the novel sequences shown by only our cDNA clones, have been omitted and are shown as double slashes (//). The lengths of novel sequences are represented under the horizontal arrow.

Table 1. Estimation of the fullness ratio of the full-length enriched cDNA libraries.

	C1SNU17	L4SNU475	L19N779227
EST[a]	184	182	173
Known gene[b] (%)	174 (95)	172 (95)	164 (95)
5'-Full[c] (%)	112 (64)	104 (60)	118 (72)
Not Full[d] (%)	62 (36)	68 (40)	46 (28)

The total number of analyzed clones is shown in the appropriate column and the percentages of clones are shown in parentheses. [a]Clones that were regarded as "high-quality" ESTs of at least 100 bp, after both vector and low-quality trimming. [b]Clones that were identified as known genes by having 95% identity over 90 bp versus the GenBank and human RefSeq databases. [c]The clones that contain the translation initiation codon ATG and had homology of at least 90% over the first 50 bp of the coding sequence with human RefSeq. [d]The clones that lacked the ATG translation initiation codon.

clone longer cDNAs, or by the instability of the long plasmid clones.

To determine the fullness ratios of these libraries, randomly selected 192 cDNA clones in each library were subjected to single-pass sequencing from their 5'-ends. The sequences obtained were then matched to the GenBank and RefSeq databases. An EST scored as a "Full-length cDNA" if it contained the translation initiation codon ATG and had a homology of at least 90% over the first 50 bp of its coding sequence with the human RefSeq or mRNA. As shown in Table 1, this analysis showed that, of the known gene clones, 112 clones (64%) in the C1SNU17 library, 104 clones (60%) in the L4SNU475 library and 118 clones (72%) in the L19N779227 library contained full-length cDNAs. Of these 334 full-length cDNAs, 181 cDNAs (54%) had a longer 5' end than the human RefSeq (data not shown). Some examples of this are shown in Figure 3. In cDNA libraries constructed using conventional methods of selecting the cap structure, the fullness ratio was reported to be ca. 60-70% (Kato *et al.,* 1994; Maruyama and Sugano, 1994; Edery *et al.,* 1995; Carninci *et al.,* 1996; Suzuki *et al.,* 2000; Suzuki and Sugano, 2001). These results indicate that the libraries constructed by our method have similar or higher fullness ratios than those constructed using methods involving lower amount of starting RNA, or selecting the cap structure of the starting poly $(A)^+$ RNA.

In this report, we described an improved method for constructing full-length enriched cDNA libraries from small amount of mRNA. Large-scale cDNA collection from cDNA libraries of the different organs, tissues, or cells has been performed to elucidate gene functioning in the cells. Of the collected cDNAs, the novel genes rarely or specifically expressed in only a few cells of a tissue have been a few included, although these genes are of great interest in studies of cell differentiation, neurobiology, developmental biology and related fields. The collection of these genes,

especially of full-length cDNA, for functional studies, is often difficult, because the construction of full-length cDNA libraries requires over 5 μg of mRNA when conventional methods, *e.g.*, original oligo-capping method, are used. However, the quality of full-length cDNA libraries constructed using our method seems to be comparable or better than those constructed using conventional methods. The majority of the conventional methods require multiple purification or precipitation steps to select the cap structure of the starting mRNA, *e.g.*, oligo-capping steps, and significant mRNA degradation and loss is likely to occur. Thus, it is not possible to construct a full-length cDNA library of high quality using small amounts of mRNA as start materials. In the devised method, a large amount of total RNA containing a small amount of mRNA is used in the oligo-capping steps. This large reservoir of RNA seems to act as a carrier and to protect the smaller amount of mRNA from degradation. The remaining mRNA is then purified from the total RNA, directly used as a template for first-strand cDNA synthesis, and amplified by PCR. Using the process, we think that the degradation and loss of small amounts of mRNA are avoided. The full-length cDNA libraries constructed using the devised method were found to contain sufficient colony numbers and a high ratio of full-length cDNA. These results indicate that the developed method is efficient for generating high contents of full-length cDNA clones, from limited amounts of starting material.

Acknowledgement

This work was supported by grant numbers FG-5-1-01 and FG-5-4 from the 21C Frontier Functional Human Genome Project from the Ministry of Science & Technology of Korea. We thank Dr. Sumio Sugano, Institute of Medical Science, University of Tokyo, Japan for technical support during the construction of a full-length cDNA library using the original oligo-capping method.

References

Albert B, Bray D, Lewis J, et al. Molecular Biology of the Cell, 3rdEd, 1994, pp. 369-70, Garland Publishing inc., NY.

Belyavsky A, Vinogradova T, Rajewsky K. PCR-based cDNA library construction: General cDNA libraries at the level of a few cells. Nucleic Acids Res 1989;17:2919-32

Bertiol DJ, Smoker M, Brown AC, Jones MG, Burrows PR. A method based on PCR for the construction of cDNA libraries and probes from small amounts of tissue. Biotechniques 1994;16:1054-8

Carninci P, Kvam C, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, Kamiya M, Shibata K, Sasaki N, Izawa M, Muramatsu M, Hayashizaki Y, Schneider C. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. Genomics 1996; 37:327-36

Edery I, Chu LL, Sonenberg N, Pelletier J. An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture). Mol Cell Biol 1995;15:3363-71

Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred I. Accuracy assessment. Genome Res 1998;8:175-85

Kato S, Sekine S, Oh SW, Kim NS, Umezawa Y, Abe N, Yokoyama-Kobayashi M, Aoki T. Construction of a human full-length cDNA bank. Gene 1994;150:243-50

Lambert KN, Williamson VM. cDNA library construction from small amounts of RNA using paramagnetic beads and PCR. Methods Mol Biol 1997;69:1-12

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. Nature 2001;409:860-921

Maniatis T, Fritsch EF, Sambrook J. *Molecular Cloning:* A Laboratory Manual, 1982, Cold Spring Harbor Laboratory press, NY.

Maruyama K, Sugano S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene 1994;138:171-4

Peterson LA, Brown MR, Carlisle AJ, Kohn EC, Liotta LA, Emmert-Buck MR, Krizman DB. An improved method for construction of directionally cloned cDNA libraries from microdissected cells. Cancer Res 1998;58:5326-8

Suzuki Y, Yoshitomo-Nakagawa K, Maruyama K, Suyama A, Sugano S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 1997; 200:149-56

Suzuki Y, Ishihara D, Sasaki M, Nakagawa H, Hata H, Tsunoda T, Watanabe M, Komatsu T, Ota T, Isogai T, Suyama A, Sugano S. Statistical analysis of the 5' untranslated region of human mRNA using "Oligo-Capped" cDNA libraries. Genomics 2000;64:286-97

Suzuki Y, Sugano S. Construction of full-length-enriched cDNA libraries. The oligo-capping method. Methods Mol Biol 2001;175:143-53

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, et al. The sequence of the human genome. Science 2001;291:1304-51