Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector

(electroporation/physical mapping/human genome)

HIROAKI SHIZUYA*, BRUCE BIRREN, UNG-JIN KIM, VALERIA MANCINO, TATIANA SLEPAK, YOSHIAKI TACHIIRI, AND MELVIN SIMON[†]

Division of Biology, 147-75, California Institute of Technology, Pasadena, CA 91125

Contributed by Melvin Simon, June 30, 1992

ABSTRACT   A bacterial cloning system for mapping and analysis of complex genomes has been developed. The BAC system (for bacterial artificial chromosome) is based on *Escherichia coli* and its single-copy plasmid F factor. It is capable of maintaining human genomic DNA fragments of >300 kilobase pairs. Individual clones of human DNA appear to be maintained with a high degree of structural stability in the host, even after 100 generations of serial growth. Because of high cloning efficiency, easy manipulation of the cloned DNA, and stable maintenance of inserted DNA, the BAC system may facilitate construction of DNA libraries of complex genomes with fuller representation and subsequent rapid analysis of complex genomic structure.

There is currently underway an intense effort to construct a high-resolution physical map of each of the human chromosomes. Eventually, these maps will be composed of overlapping fragments of human DNA and will allow the direct acquisition of DNA fragments that correspond to specific genes. Completion of the physical map requires the availability of comprehensive libraries of DNA clones in appropriate vectors. Furthermore, the accuracy and efficiency of physical mapping increase progressively with the size of the clone fragments in these libraries. Thus, the construction of libraries using yeast artificial chromosomes (YACs), which permit cloning of fragments of ≥500 kilobase pairs (kb), represents a fundamental advance in our ability to generate physical maps that order DNA over multi-megabase distances (1). However, some difficulties have been encountered with the manipulation of YAC libraries (2–4). Thus, for example, in various libraries a fraction of the clones result from co-cloning events; i.e., they include in a single clone noncontiguous DNA fragments. We describe here a bacterial cloning system that we refer to as BACs, bacterial artificial chromosomes. This system may provide a supplement and alternative to the YAC system for some applications requiring cloning of large fragments. The BAC system is based on the well-studied *Escherichia coli* F factor. Replication of the F factor in *E. coli* is strictly controlled (5). The F plasmid is maintained in low copy number (one or two copies per cell), thus reducing the potential for recombination between DNA fragments carried by the plasmid. Furthermore, F factors carrying inserted bacterial DNA are capable of maintaining fragments as large as 1 megabase pair, suggesting that the F factor is suitable for cloning of large DNA fragments (6). Other bacterial systems for cloning large DNA have been developed. For example, the system based on bacteriophage P1 is in use (7). However, the P1 vector has a maximum cloning capacity of 100 kb. A bacterial system based on F factors has been reported (8). However, in this system,

human DNA inserts >120 kb have not been cloned and characterized. The BAC system allows us to clone large DNA from a variety of complex genomic sources into bacteria, where the DNA is stable, easy to manipulate, and represents a single foreign DNA source.

MATERIALS AND METHODS

Construction of BAC Vector. A 97-base-pair (bp) synthetic oligonucleotide containing nucleotide sequences for T7/SP6 promoters, two cloning sites (*Hin*dIII and *Bam*HI), and sites for rare-cutter restriction enzymes (*Not* I, *Eag* I, *Xma* I, *Sma* I, *Bgl* I, and *Sfi* I) were inserted into the single *Sal* I site of pMBO131 (9). A 400-bp *Sac* I fragment of bacteriophage λ carrying the *cosN* site was then inserted into the *Sac* I site of the plasmid. A 42-bp oligonucleotide having a bacteriophage P1 *loxP* site was inserted at a *Cla* I site between *cosN* and the cloning site. The resulting plasmid is called pBAC108L.

Preparation of High Molecular Weight DNA. We prepared human DNA for cloning in agarose from MOLT-4 cells (a leukemic cell line). Cells were harvested by low-speed centrifugation and suspended in phosphate-buffered saline (PBS) to a final cell density of 10^8 per ml. An equal volume of liquefied 1% low-melting agarose in PBS was mixed with the cell suspension and then the whole agarose/cell mixture was poured into a 7 mm × 7 mm plastic mold as described by Birren and Lai (10). After the mold was placed at 0°C for 10 min to solidify, the agarose block was pushed out into 50 ml of digestion buffer (1% sodium N-lauroylsarcosine/0.5 M EDTA, pH 8.0, with proteinase K at 2 mg/ml). Cells were digested at 50°C for 2 days with one change of an additional 50 ml of lysis buffer, and the digest was dialyzed against four changes of 50 ml of 50 mM EDTA, pH 8.0.

Partial Digestion of DNA. For *Hin*dIII partial digestion, agarose plugs were first dialyzed extensively against 10 mM Tris·HCl/1 mM EDTA, pH 8.0, with 50 mM NaCl, and then equilibrated with *Hin*dIII digestion buffer (50 mM NaCl/10 mM Tris·HCl, pH 8.0/10 mM MgCl₂) containing nuclease-free bovine serum albumin (100 μg/ml) at 4°C for 30 min. Serially diluted *Hin*dIII (1–10 units) was added to the plugs and incubated at 37°C for 30 min with gentle shaking. Optimal digestion conditions were determined by analyzing the size range of digested DNA by pulsed-field gel electrophoresis.

Cloning Human DNA into pBAC. DNA ranging from 100 to 300 kb was excised from low-melting agarose pulsed-field gels. The agarose slice was melted at 65°C for 5 min and digested with GELase (Epicentre, Madison, WI) by adding 1 unit of enzyme, mixing gently, and incubating at 40°C for 1 hr. One to three hundred nanograms of the human DNA was

ligated to HindIII-digested pBAC108L (molar ratio of 10 to 1 in pBAC excess) with 400 units of T4 DNA ligase at 16°C overnight. To prevent religation of the vector, HindIII-cut pBAC108L was treated with HK phosphatase (Epicentre). Transformation of E. coli DH10B cells (BRL) was carried out by electroporation. Competent cells were prepared and electroporation was performed according to the manufacturer's instructions (Bio-Rad). After electroporation, cells were incubated at 37°C with gentle shaking for 45 min prior to spreading on LB plates containing chloramphenicol at 12.5 μg/ml. Plates were incubated at 37°C overnight. To identify BAC transformants carrying human inserts, colonies were lifted onto nylon filters (Hybond-N, Amersham) for hybridization with radioactively labeled total human DNA.

RESULTS

The BAC Vector. Many bacterial vectors based on high- to medium-copy-number replicons often show structural instability of inserts, deleting or rearranging portions of the cloned DNA, particularly DNA inserts of eukaryotic organisms that may contain families of repeated sequences (11–13). Therefore, it is difficult to clone and maintain large intact DNA in bacteria. However, the F factor not only codes for genes that are essential to regulate its own replication but also controls its copy number. The regulatory genes include oriS, repE, parA, and parB. The oriS and repE genes mediate the unidirectional replication of the F factor while parA and parB maintain copy number at a level of one or two per E. coli genome. The BAC vector (pBAC) incorporates these essential genes as well as a chloramphenicol resistance marker and a cloning segment (Fig. 1). The cloning segment includes (i) the bacteriophage λ cosN and P1 loxP sites, (ii) two cloning sites (HindIII and BamHI), and (iii) several C+G-rich restriction enzyme sites (Not I, Eag I, Xma I, Sma I, Bgl I, and
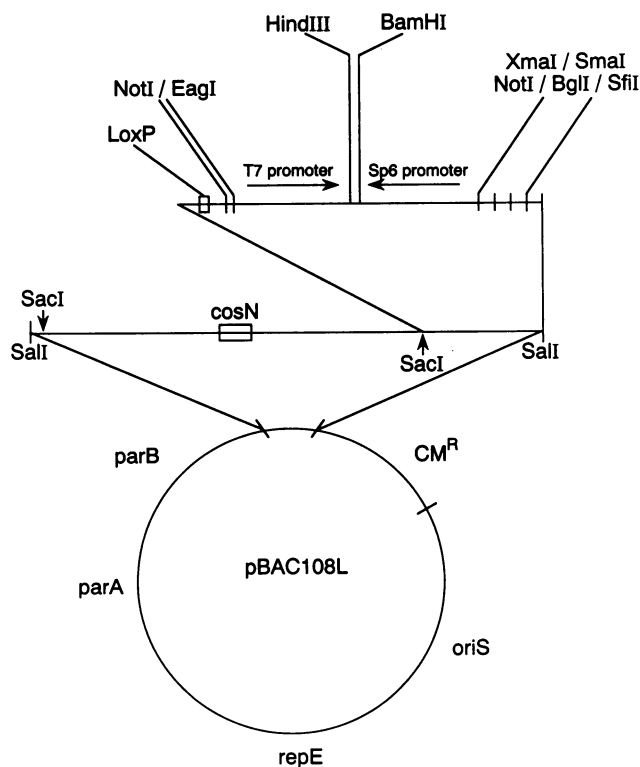

FIG. 1. Construction of pBAC vector. The plasmid is based on a mini-F plasmid, pMBO131 (9). A detailed description of the construction is provided in Materials and Methods. CM^R, chloramphenicol resistance marker.

Sfi I) for potential excision of the inserts. The cloning site is flanked by T7 and SP6 promoters for generating RNA probes for chromosome walking, and for DNA sequencing of the inserted segment at the vector–insert junction. The cosN site provides a fixed position for specific cleavage with the bacteriophage λ terminase (14). The loxP site can be utilized similarly. In this case P1 Cre protein catalyzes the cleavage reaction in the presence of the loxP oligonucleotide (ref. 15 and unpublished result). These sites (cosN and loxP) allow the convenient generation of ends that can be used for restriction-site mapping to arrange the clones in an ordered array. Restriction maps of the individual clones can be determined by indirect end-labeling and subsequent partial digestion (14, 16, 17).

Cloning Human DNA in the BAC Vector. The ligated BAC DNA molecules are expected to be much larger than those traditionally used in bacterial cloning. As one expects, transformation becomes increasingly more difficult as the size of DNA molecules increases (unpublished observation). We have tested several procedures to transform large circular F'lac DNA (180 kb) into different E. coli strains and found that standard electroporation gives high-efficiency transformation, about 10^6 transformants per μg of DNA (data not shown). The frequency of transformation with F'lac DNA is about 1/40th that found with pBAC plasmid with no insert. Among E. coli strains commonly used for electroporation, we found a few strains that are capable of transformation by F'lac DNA, and we chose DH10B as host. The exact nature of the genetic basis of the competence is not known.

Typically, when 2 μl of the ligation mixture containing 50 ng of input human DNA was transformed into highly concentrated competent cells (10^8), 100–1000 colonies appeared on the plate. Colonies appearing on selective plates were hybridized with whole human DNA to identify clones with human inserts. Subsequent analysis indicated high variability with respect to insert frequency from experiment to experiment; 10–50% of the transformants carried human inserts, depending upon the batch of the vector and insert DNA used.

Analysis of the Human Inserts. Although the amount of DNA obtained from bacterial cells carrying the single-copy plasmid is expected to be low, routine minipreps from 1.5 ml of overnight culture yielded enough DNA for several experiments. Twenty randomly picked clones having human inserts were analyzed by pulsed-field gel electrophoresis after Not I digestion (Fig. 2A). All lanes contain the 6.7-kb vector band and human insert DNA of various molecular weights. The size distribution of the inserts in Fig. 2A ranges from 10 to 215 kb, with an average size of 100 kb. Fig. 2C shows an additional five large clones ranging in size up to 300 kb, which were obtained from separate transformations. The DNA in the gel of Fig. 2A was transferred to a nylon filter and probed with total human DNA. As expected, only the insert hybridized to the human DNA probe and no additional minor bands appeared even after prolonged exposure (Fig. 2B). Deletions occurring within the BACs in the E. coli host would yield additional minor bands on restriction digestion. These were not detected by ethidium bromide staining or by the more sensitive hybridization assay. This suggests that human inserts on the BAC vector are structurally quite stable within the resolution of the pulsed-field gel electrophoresis analysis.

To further examine the structural stability of such large regions of human DNA cloned in E. coli, the restriction patterns of DNA taken before and after long-term culture were compared. Overnight cultures of two clones (125-kb and 175-kb inserts) were diluted and used to inoculate fresh broth for overnight growth. This process was repeated for 5 days (about 100 generations). The sizes of the restriction fragments produced by digestion of the DNA from the final cultures (Nhe I, Sac I, and Xho I) were compared with those from the original day 1 DNA. No visible difference between the two
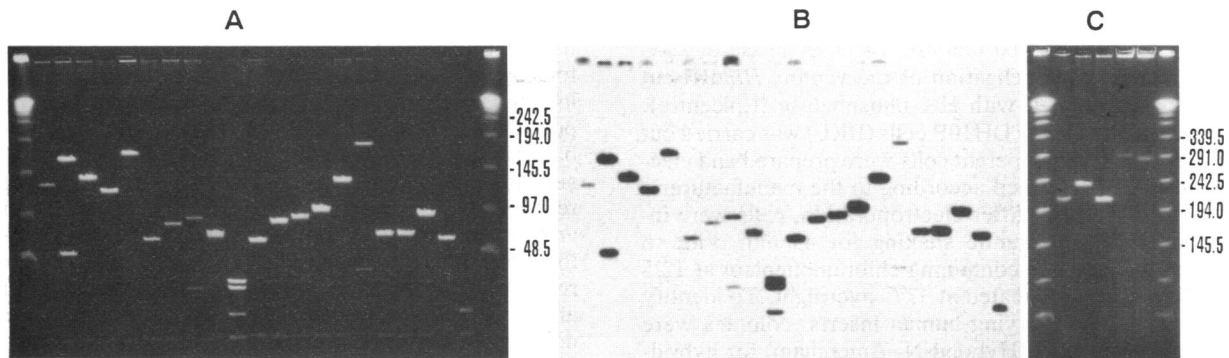
FIG. 2. Analysis of human BAC clones by pulsed-field gel electrophoresis. pBAC with inserts was prepared from small cultures (1.5 ml) of transformed colonies by a standard alkaline lysis procedure. (*A*) Ethidium bromide-stained BAC DNA analyzed by pulsed-field gel electrophoresis. DNA minipreparations were digested with *Not* I. Separation of the digested DNA was carried out on a Bio-Rad CHEF Mapper apparatus for 20 hr at a field strength of 6 V/cm in a 1% agarose gel in 45 mM Tris/45 mM boric acid/1 mM EDTA at 16°C with a linear pulse time ramping from 5 to 15 sec. (*B*) Southern blot showing hybridization to radioactive total human DNA probe. DNA from *A* was transferred to a nylon filter (Hybond-N, Amersham). DNA hybridization was performed at 42°C in 50% formamide for 16 hr and the filter was washed three times with 0.3 M NaCl/0.03 M sodium citrate, pH 7.0, at 47°C. (*C*) Sizing of five large BAC DNA. For separation of DNA fragments of 50–500 kb, switch times are ramped from 7 sec to 45 sec over 29 hr.

samples after digestion and pulsed-field gel electrophoresis was observed (Fig. 3). Moreover, *Not* I-digested DNA from 12 individual colonies derived from the day 5 culture showed no visible differences when compared with each other or with DNA from the same number of colonies from the day 1 culture (data not shown). Comparable experiments with cosmid clones (35- to 45-kb inserts) in the Super-cos or Lawrist vectors showed that up to 40% of the clones rearranged after a shorter period of serial growth (13).

We have obtained BAC clones larger than 300 kb; however, this is probably not the upper limit for BAC cloning. In
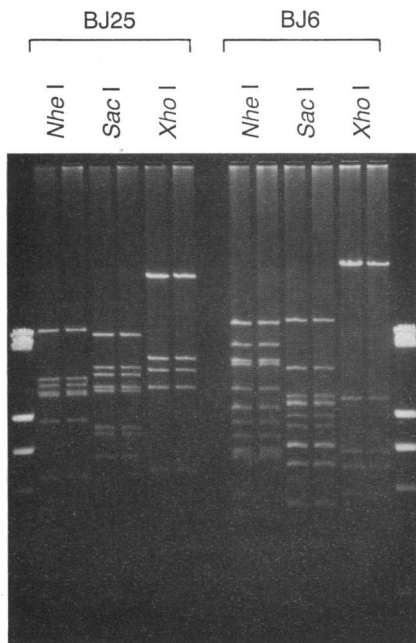

FIG. 3. Comparison of restriction fragments of two clones before and after 100 generations. Two clones carrying BACs of 125 kb (BJ25) and 175 kb (BL6) were grown in 5 ml of LB containing chloramphenicol (12.5 μg/ml) and grown at 37°C overnight (day 1). Cultures were diluted 100-fold and 0.1 ml of this dilution was inoculated into 500 ml of LB with chloramphenicol (12.5 μg/ml) and incubated overnight. Serial cultures were made in this way for an additional 3 days. Plasmid DNAs were prepared from day 1 and day 5 cultures, digested with three different restriction enzymes (*Nhe* I, *Sac* I, and *Xho* I), and analyzed by pulsed-field gel electrophoresis (day 1, left lane; day 5, right lane).

the case of YACs, the size of the inserts can be larger than 300 kb. However, a high percentage of YAC clones, particularly higher molecular weight inserts, represent chimeric molecules. These chimeric clones may hamper the progress of mapping and analysis of chromosomes. To evaluate the frequency of co-cloning events among the BAC clones, we have employed fluorescence *in situ* hybridization with BAC DNA minipreps. We have thus far analyzed 28 randomly picked human BAC clones (all between 200 and 300 kb) and identified only one clone that potentially represents a chimeric clone (data not shown). Comparable experiments with YAC clones carrying large human DNA inserts showed that a substantial number of clones mapped to multiple sites, suggesting that many of the clones were chimeric (2). The co-cloning frequency in the BAC system appears to be significantly lower than that in the YAC. Another possible source of chimeric YAC clones is that multiple artificial chromosomes may be found in a single yeast cell (3, 4). In the BAC system, the two genes *parA* and *parB* are involved in the exclusion of extraneous F factors (two BACs in a single cell). In one experiment we analyzed more than 300 clones and have not found multiple BACs in a single cell.

DISCUSSION

Several lines of evidence suggest the utility of the BAC vector as a general cloning vehicle to construct comprehensive libraries of higher organisms. (*i*) Large plasmids can be transfected to *E. coli* by electroporation. The transformation of bacteria by electroporation is 10–100 times more efficient than yeast spheroplast transformation. This reduces the amount of DNA necessary for library construction. This feature is particularly suited to creating specialized libraries where the source DNA is limited. (*ii*) Traditional bacterial colony lifts and hybridization methods can be used for library screening even though only a single copy of pBAC plasmid exists in cells. (*iii*) Unlike YACs, which are linear, BACs with inserts, like large F' factors, exist as supercoiled circular plasmids in *E. coli*. This permits easy isolation and manipulation of the large DNA in solution with minimal breaking, whereas yeast chromosomes are more difficult to isolate intact since the linear DNA is more susceptible to shear. Furthermore, the ability to visualize the cloned DNA by ethidium bromide staining in pulsed-field gels vastly increases the speed of structural analysis of the clones. (*iv*) DNA from minipreps is easily reintroduced into bacterial cells by the same electroporation procedure. (*v*) A set of

genes for F' maintenance is sufficient to sustain large foreign DNA in the BACs. (*vi*) A low frequency of co-cloning in the BACs has significant advantage in physical mapping by "bottom-up" approaches.

One of the important aspects in any cloning system is the genetic stability of cloned DNA. Results showing a stable structure for human DNA in the BAC vector suggests that the cloned DNA should represent a stable source of specific fragments. However, more experiments including long-range restriction-site mapping with BACs will be required to demonstrate that cloned DNA fragments are indeed faithful copies of the source DNA. Some clones in any genomic library may be expected to show instability in a particular host because of the recombinogenic activity of certain specific DNA sequences. A bacterial system may offer advantages in this respect because the pathways of genetic recombination in *E. coli* have been described in detail, and the use of appropriate bacterial mutants may reduce instability problems. For this purpose we have constructed several candidate repository strains that lack many recombination functions. A set of new strains are based on DH10B. They are capable of transfection of the large DNA and deficient in host-controlled restriction (*hsdMRS*), methylation-sensitive restriction (*mcrABC, mrr*), and various types of general recombination systems (*recBC, sbcBC*). Among the strains constructed are HS986, having a *recJ* mutation, and HS979, having a *recA* deletion mutation, as well as the mutations described above.

In any cloning system, analysis and fingerprinting of many clones with large inserts require rapid and convenient methods. One fingerprinting strategy that takes advantage of the vector involves the use of λ terminase to generate cohesive ends at the *cosN* site, which may then be used as reference points to establish a restriction map by partial digestion. The restriction pattern of each clone can be compared and used to recognize overlapping clones. Independent labeling of both allows us to carry out the fingerprinting analysis with greater speed and accuracy. The other benefit of having *cosN* on the vector is that the BAC vector can be used as a highly efficient cloning vector for DNA of 40-kb average size after packaging and infection as in conventional cosmid cloning. Extensive libraries of fosmid (F-based cosmid) clones are readily constructed and offer increased insert stability (13). The clones can be easily analyzed and used in physical mapping.

Recently, we have begun construction of a BAC library of total human DNA. In addition we have prepared a chromosome 22-specific fosmid library with 7-fold redundancy (unpublished results). The library together with fosmids will augment physical chromosome mapping with YACs and with conventional cosmids.

1. Burke, D. T., Carle, G. F. & Olson, M. V. (1987) *Science* **236**, 806–812.
2. Green, E. D., Riethman, H. C., Dutchik, J. E. & Olson, M. V. (1991) *Genomics* **11**, 658–669.
3. Wada, M., Little, R. D., Abidi, F., Porta, G., Labella, T., Cooper, T., Della Valle, G., D'Urso, M. & Schlessinger, D. (1990) *Am. J. Hum. Genet.* **46**, 95–106.
4. Bellis, M., Gerard, A., Charlieu, J. P., Marcais, B., Brun, M. E., Viegas-Pequignot, E., Carter, D. A. & Roizes, G. (1991) *DNA Cell Biol.* **10**, 301–310.
5. Willetts, N. & Skurray, R. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. in chief Neidhardt, F. C. (Am. Soc. Microbiol., Washington), Vol. 2, pp. 1110–1133.
6. Low, K. B. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. in chief Neidhardt, F. C. (Am. Soc. Microbiol., Washington), Vol. 2, pp. 1134–1137.
7. Sternberg, N. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 103–107.
8. Hosoda, F., Nishimura, S., Uchida, H. & Ohki, M. (1990) *Nucleic Acids Res.* **18**, 3863–3869.
9. O'Connor, M., Pfeifer, M. & Bender, W. (1989) *Science* **244**, 1307–1312.
10. Birren, B. & Lai, E., in *Pulsed Field Gel Electrophoresis: A Practical Guide* (Academic, San Diego), in press.
11. Wyman, A. R., Wolfe, L. B. & Botstein, D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2880–2884.
12. Ishiura, M., Hazumi, N., Koide, T., Uchida, T. & Okada, Y. J. (1989) *J. Bacteriol.* **171**, 1068–1074.
13. Kim, U.-J., Shizuya, H., de Jong, P. J., Birren, B. & Simon, M. I. (1992) *Nucleic Acids Res.* **20**, 1083–1085.
14. Rackwitz, H. R., Zehetner, G., Murialdo, H., Delius, H., Chai, J. H., Poustka, A., Frischauf, A.-M. & Lehrach, H. (1985) *Gene* **40**, 259–266.
15. Abremski, K., Hoess, R. & Sternberg, R. (1983) *Cell* **32**, 1301–1311.
16. Rackwitz, H. R., Zehetner, G., Frischauf, A.-M. & Lehrach, H. (1984) *Gene* **30**, 195–200.
17. Kohara, Y., Akiyama, K. & Isono, K. (1987) *Cell* **50**, 495–508.