A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones

ICI Pharmaceuticals, Biotechnology Department, Mereside, Alderley Park, Macclesfield, Cheshire
SK10 4TG, UK

Received February 15, 1990, Revised and Accepted April 18, 1990
EMBL accession nos X52166, X52167

ABSTRACT

The recent development of yeast artificial chromosome (YAC) vectors has provided a system for cloning fragments that are over ten times larger than those that can be cloned in more established systems. We have developed a method for the rapid isolation of terminal sequences from YAC clones. The YAC clone is digested with a range of restriction enzymes, a common linker is ligated to the DNA fragments and terminal sequences are amplified using a vector specific primer and a linker specific primer. Sequence data derived from these terminal specific products can be used to design primers for a further round of screening to isolate overlapping clones. The method also provides a convenient method of generating Sequence Tagged Sites for the mapping of complex genomes.

INTRODUCTION

The recent development of yeast artificial chromosome (YAC) cloning systems enables the cloning of fragments that are several hundred kilobases in size (1). We have constructed a 3.5 human genome equivalent library from size fractionated DNA. The library has an average insert size of >350 kb and has been prepared as a gridded array (2, 3). Individual YAC clones can be isolated by conventional hybridisation techniques (4) or by PCR screening (5). Restriction mapping of YAC clones is facilitated by the use of probes derived from the vector arms (1). Elegant fingerprinting methods have been used to establish cosmid contigs (6) but these methods have not yet been applied to YACs (7). The construction of YAC contigs would be simplified if probes derived from the termini of the insert could be isolated. The isolation of such probes by subcloning is time consuming (8) and although PCR amplification between Alu consensus sequences and vector sequences has been described (9), this technique is not universally applicable since it relies entirely on the presence of an Alu repeat within a short distance of the end.

The analysis of YAC terminal sequences by inverse PCR (10,11) has been described (12). The technique of inverse PCR relies on circularisation and depends on the presence of two restriction sites within the size range of PCR amplification. We have developed a linker specific amplification technique for walking and sequencing from a defined starting point which can be used to generate terminus-specific probes from YAC clones.

MATERIALS AND METHODS

Construction of Vectorette Libraries

The construction of the gridded YAC library and isolation of YAC colonies has been described previously (2, 3). Individual YAC colonies were used to inoculate 10 ml SD medium (7 g/litre Bacto yeast nitrogen base without amino acids, 20 g/litre glucose, 55 mg/litre adenine and tyrosine, 14 g/litre casamino acids) and shaken at 200rpm overnight at 30°C. Cells were harvested by centrifugation, washed once in 50 mM EDTA and resuspended to 3—5×10^9/ml in Yeast Resuspension Buffer (YRB; 1.2 M sorbitol, 10 mM Tris HCl pH 7.5, 20 mM EDTA, 14 mM B mercaptoethanol and 20 U/ml Lyticase (Sigma). The suspension was incubated at 37°C for 1 hour by which time ~80% of the cells were spheroplasts. The cells were gently mixed with an equal volume of a 1 % solution of LGT agarose in YRB at 37°C and transferred to a plug mould to set (13). Plugs were transferred to 2—3 volumes of Yeast Lysis Buffer (YLB; 0.1 M EDTA, 10 mm Tris HCl pH 8.0, 1 % lithium dodecyl sulphate) and gently shaken at room temperature for 1—2 hours. The solution was replaced with 10 volumes of YLB and incubated at 40—50°C overnight. The plugs were rinsed once in YLB and stored in YLB at room temperature.

Oligonucleotides were synthesised by phosphoramidite chemistry on an Applied Biosystems 380B instrument and following deprotection, the oligonucleotides were phosphorylated with γATP and polynucleotide kinase and purified as described previously (14).

Each complete plug was cut into 3 equal slices. All slices to be used for the construction of Vectorette Libraries were immersed together in 50 ml TE (10 mM Tris HCl pH 7.6, 1 mM EDTA) and gently agitated for two hours at 4°C. The slices were then immersed in fresh TE and agitation continued at 4°C for 16 hours. Two slices were used for each restriction digest and these were then placed in the appropriate restriction enzyme buffer (500 μl) and incubated for 2 hours at 4°C. The buffer was removed and replaced by 100 μl fresh buffer containing restriction enzyme (20 units) and the slices incubated for 2 hours at 37°C. The solution was then removed and replaced with 1 ml 1× ligation buffer (50 mM Tris HCl pH7.6, 10 mM MgCl2, 1 mM DTT). The slices were then incubated for 2 hours at 4°C.
Fig 1. (a) Sequence of vectorette oligonucleotides, showing position of universal amplification primer, 224, and sequencing primer, 368. The sequence N1, N2, N3, N4 is complementary to the overhang created by the restriction enzyme used to prepare the library. The vectorette overhang used to construct the Hinfl library was ANT where N indicates the presence of all 4 bases. (b) Schematic showing specific amplification of terminal sequences.

The buffer was removed and replaced by 100 μl fresh ligation buffer containing 2 pmole (6 pmole for frequent cutters) annealed Vectorette oligonucleotides. The contents of the tube were heated to 65°C for 15 minutes to melt the plugs. The solution was cooled to 37°C, ATP was added to a final concentration of 1 mM followed by 9 units of T4 DNA ligase (Boehringer Mannheim). After one hour at 37°C, the volume was adjusted to 500 μl with water. Vectorette libraries were stored in aliquots at -20°C.

Amplification of Libraries

A 5 μl aliquot of each Vectorette Library was amplified using either primers 1089 and 224 or 1091 and 224 (Figure 2) in a total volume of 100 μl containing 10 mM Tris HCl pH 8.3, 50 mM KCl, 1 mM MgCl2, 0.1% gelatin, 100 μM dNTP and 100 pmole of each primer. Reaction mixtures were denatured at 96°C for 10 minutes and then cooled to 92°C before adding Taq DNA polymerase (Perkin-Elmer Cetus, 2 units). Subsequent cycles consisted of denaturation for 2 minutes at 92°C, annealing for 2 minutes at 60°C and extension for 3 minutes at 72°C. A total of 38 cycles were performed. Aliquots of the reaction were analysed on 1.4% agarose gels. Amplification products were eluted from the gels on to NA45 paper (Schleicher and Schull) and sequenced using primers 1207, 1208, or 368 as described previously (14). Hybridisation probes were generated by digestion of reaction mixtures with EcoRI to remove the amplified vector segment and subsequent isolation of insert DNA on 1.4% low gelling temperature (LGT) agarose gels.

RESULTS AND DISCUSSION

The general concept of the method and its application to the analysis of human genomic DNA will be described elsewhere (manuscript in preparation). The specific application of the method to the analysis of YAC clones is outlined in Figure 1. Spheroplasts are prepared and immobilised in agarose blocks which are equilibrated in buffer to remove low molecular weight impurities prior to restriction digest. We have not evaluated the
use of miniprep DNA for this purpose. After digestion with a restriction enzyme, the resulting DNA fragments are ligated to a synthetic oligonucleotide duplex containing an appropriate 4bp overhang which will anneal to the specific 4bp overhang generated by the restriction enzyme. These synthetic duplexes are termed vectorettes and the construct resulting from ligation of a vectorette duplex to an appropriate restriction digest is known as a vectorette library. Suitable oligonucleotides can be generated for a range of restriction enzymes including those which create a 3' overhang or blunt ends. The only enzymes which cannot be used with pYAC4 libraries are EcoRI which cuts at the cloning site or enzymes which cut between the cloning site and the primer annealing region of the YAC vector eg Smal. The duplex contains regions of non-complementarity (Figure 1).

When amplification is performed between primers derived from the yeast vector and primer 224, only terminal fragments containing vector sequences will be amplified. Products can only be synthesized from primer 224 if an initial round of synthesis has taken place from the yeast vector primers. In this way, specific amplification products can be synthesized from the termini of the YAC insert. The sequence of the YAC vector and the relative position of the primers are shown in Figure 2. We have not investigated the use of alternative sequences as linkers or primers. Provided the substrate is thoroughly denatured, we do not normally encounter background products.

We have tested this system with YAC clone 14IB10 which contains a 460 kb insert. This clone was identified by PCR screening of the YAC library with primers from the D7S8 (J3.11) locus. Aliquots of 14IB10 were digested with Hinfl and Rsal and Vectorette Libraries were constructed as described in the Methods section. Amplification of individual Vectorette Libraries was performed with primer combinations 1089 and 224 or 1091 and 224. Amplification between 1089 and 224 gave a product of 0.65 kb from a Hinfl library, while amplification between primers 1091 and 224 gave a product of 0.6 kb from a Rsal library. Digestion of these products with EcoRI gave the expected insert products of 0.35 kb and 0.4 kb respectively confirming that amplification had occurred across the EcoRI cloning site at both ends of the insert (Figure 3). To further confirm the localization of the region being amplified, duplicate Southern blots of the partial digests of 14IB10 with 3 infrequently cutting enzymes were made. These blots were first hybridized to EcoRI digested and purified vectorette products. An identical hybridization pattern was obtained providing confirmation that the amplification products were indeed from the insert ends of the YAC (data not shown).

The Hinfl amplification product was purified by gel
electrophoresis, isolated and sequenced using primer 1207. Approximately 200 bp of sequence was obtained. Additional data was obtained by sequence analysis of a 0.75 kb product obtained by amplification of a Rsal Vectorette Library using primers 1089 and 224 (Figure 4). When the sequences were searched against the GenBank database (Release 56.0), no homologies were found to any previously described sequences or to families of repetitive sequences (Alu, LINE). Oligonucleotide primers (1422, 1423) were designed to amplify a 300 bp product from the ‘left hand’ terminus of clone 14IB10. The position of the primers is shown in Figure 4. DNA from pools of 864 clones followed by sub-pools of 96 clones from the YAC library were screened by PCR with these primers. The predicted product was observed in genomic controls, pool 141 (which contained clone 14IB10) and in 2 additional pools. Positive clones were localised to individual microtirte plates (19E, 311) by a further round of amplification of sub-pools of 96 colonies. Individual colonies were identified by hybridisation of the labelled PCR product to a replica filter prepared from the master 96 well plates. These positive clones were not detected by the J3.11 specific primers in a PCR reaction.

The approach described above provides a rapid method of obtaining sequence information from the termini of YAC inserts. The information is sufficient to design amplification primers which can then be used to isolate further overlapping clones. These can be used in turn for further chromosome walking. Alternatively, hybridisation probes can be isolated by EcoRI digestion of the amplification products. The technique has been successful in 17/18 attempts to date and we have used the approach to isolate several overlapping YAC clones. An approach employing amplification between vector and Alu consensus sequences has been described recently. However, in this method amplification could often occur between Alu sequences within any YAC insert giving rise to multiple products which cannot be unambiguously assigned to the termini of the insert. Recently, Olson et al (15) have proposed the use of Sequence Tagged Sites (STSs) to generate landmarks for the mapping of complex genomes. The technique described in this paper provides a convenient method of generating such markers.

The material described in this paper is the subject of UK Patent GB 8917143.3

REFERENCES