Selective enrichment of a large size genomic DNA fragment by affinity capture: an approach for genome mapping

Rajendra P.Kandpal, David C.Ward and Sherman M.Weissman

Department of Human Genetics, Yale University School of Medicine, New Haven, CT 06510, USA

Received December 5, 1989; Revised and Accepted March 2, 1990

ABSTRACT

A method to enrich large size DNA fragments obtained by digestion with rare cutting restriction endonucleases was developed and applied for the isolation of a 150 kb *Sfi*I fragment containing the β -globin gene cluster. The digested DNA is rendered single stranded at the ends by diffusing a strand specific exonuclease into an agarose plug containing DNA. The plug is melted and solution hybridization is then performed with a bridge RNA containing specific sequences from the end of a desired fragment linked to a common probe sequence. The common probe sequence is annealed to a biotinylated RNA and the resulting tripartite hybrid is retained onto a solid matrix containing avidin and specifically released by ribonuclease action. Enrichments of greater than 350 fold have been achieved consistently. Such directed purification of large DNA fragments without cloning can considerably expedite mapping and gene localization in a complex genome and facilitate the construction of sublibraries from defined regions of the genome.

INTRODUCTION

A megascale restriction map of the human genome, placement of known cDNA clones on the map and sequencing of the genome are desirable goals in human genetics and molecular biology. The availability of linking libraries (1,2) and the feasibility of separating large size fragments obtained by partial digestion (3,4,5,6) have made large scale restriction maps an attainable objective. Cosmid (7) and yeast artificial chromosome (YAC) (8) cloning techniques are useful for ordering successively larger pieces of the genome. Cosmid clustering, although reasonably rapid, is time consuming due to the short distances (<40kb) it permits one to proceed in each screening cycle. YAC cloning provides larger segments of genomic DNA (>100kb), but does not permit direct isolation of specific DNA sequences. Also, the possibilities of rearrangements and deletions in YAC clones, and of sequences resistant to cloning have not been fully evaluated. To circumvent potential problems of bias of sequence representation in a library, rearrangements and screening difficulty, and to supplement existing techniques of YAC cloning, we have developed an approach of chromosome 'fishing'. The aim of this technique is to isolate large genomic DNA fragments

(>100kb) without biological selection and then use the isolated fragment either for constructing highly enriched sublibraries or recovery of cDNAs corresponding to the isolated fragment. In this report, we describe a method that utilizes a biotin-avidin capture system and a highly specific ribonuclease release reaction for the selective enrichment of large human genomic DNA fragments, and suggest how partially purified fragments could be used for further characterization of the genome.

To select intact DNA fragments of interest, it is necessary to form a hybrid between target and probe sequences without denaturing the target sequences. A Rec A protein mediated search for homologous sequences (9) has been used in our laboratories to isolate desired plasmid clones (10). This approach is advantageous because a probe may be used from anywhere within the large DNA target. However, limited stability of D loop complexes at low salt and high temperatures restrict its applicability to certain kinds of selection procedures. We chose a different approach which involves resection of the genomic DNA, previously digested with a rare cutter (*Sfi*I, in the present experiments), and hybridization to a RNA probe specific to the resected end and annealed to a second biotinylated RNA partly complementary to the probe. This strategy requires a probe from sequences adjacent to a rare cutting site in the region of interest, and therefore relies on a linking library. Such linking libraries, consisting of DNA fragments containing rare cutting sites, are becoming increasingly available for various human chromosomes.

MATERIALS AND METHODS

Preparation and restriction digestion of genomic DNA

The large size human genomic DNA was prepared by direct lysis of 3.1.0 lymphoblastoid cells (11) in agarose blocks (12). The agarose plug containing DNA was digested to completion by incubating with *Sfi*I (20 units/ μ g DNA) in 250 μ l buffer for 6 to 8 hours.

Pulsed field gel electrophoresis and Southern blotting

The *Sfi*I digested DNA was electrophoresed in a CHEF electrophoretic apparatus at 250 V with a pulse time of 60 sec (5). The gel was stained with ethidium bromide and blotted onto a Hybond membrane. The membrane was hybridized with a [³²P] labeled (13) β -globin specific probe in 0.5 M Na-Pi pH7.2, 7% SDS, 1% BSA and 1mM EDTA (14). The filter was

washed with $0.1 \times \text{SSC}$ at 65° for 30 minutes and exposed to an X-ray film. The autoradiogram confirmed that the β -globin cluster lies in a ≈ 150 kb *Sfi*I fragment, as estimated by comparison with concatamers of wild type λ DNA (data not shown).

Resection of Restriction Digested DNA with λ Exonuclease
*Sfi*I digested genomic DNA in an agarose plug was incubated with λ exonuclease buffer (70 mM Glycine-KOH, pH 9.4) without MgCl_2 and 25 units of λ exonuclease (Bethesda Research Laboratories) for 20 minutes to let the enzyme diffuse into the plugs. The reaction was started by adding MgCl_2 to a final concentration of 3 mM and the incubation continued at 15°C for 10 minutes. The reaction was stopped by adding EDTA to a final concentration of 20 mM and the λ exonuclease was digested with $100\mu\text{g}$ proteinase K at 37°C for 30 minutes. Proteinase K was inactivated by incubating the plug in a buffer containing 10 mM Tris, 1 mM EDTA and 1 mM phenylmethylsulfonyl fluoride.

Preparation of RNA transcripts

A typical RNA transcription reaction mixture contained 40mM Tris-HCl pH 8.0, 25mM NaCl, 0.5mM each of ATP, CTP, GTP, UTP, 8mM MgCl_2 , 2mM Spermidine, 50 units of ribonuclease inhibitor (Promega), 10mM dithiothreitol, 200 units of T3 or T7 RNA polymerase and 10 to 20 μg of appropriately digested DNA in a 200 μl volume. Reactions were incubated at 37°C for 60 minutes. The DNA template was digested with pancreatic DNase (1 unit/ μg DNA) at 37°C for 15 minutes, and the DNase was inactivated by heating at 70°C for 10 minutes. The reaction mixture was passed through a Sephadex G-50 spin column and the eluate was precipitated with ethanol. Biotinylated RNA was synthesized by replacing UTP with 1mM bio-11-UTP (Enzo Biochemicals) in the reaction mixture.

Polymerase chain reaction (PCR)

A typical PCR assay was carried out in a reaction mixture containing 50mM KCl, 10mM Tris-HCl, pH 8.6, 1mM MgCl_2 , 0.2 mM each of dATP, dCTP, dGTP and dTTP, 50–100 ng primers, 0.5 unit Taq polymerase and DNA template in a reaction volume of 50 μl . Reactions were run in a Cetus Perkin Elmer Thermocycler (15) using the following conditions: 1 min. at 94°C , 2 min. at 55°C , 1 min. at 72°C and repeating the same parameters for 30 to 60 cycles. The amplified samples were run on a 1.2% agarose gel, stained with ethidium bromide and the DNA was blotted onto a Hybond membrane. The sequences of the primer pairs for various genes and fragments are as follows: β -globin:

[TGGTGGTCTACCTTGGACCCAGA; ACGTGCAGCTTGTCACAGTGCAGC]

*Hind*III fragment:

[CTGCTTAATATGTGTCATAC; GGGCCTAGGGTAAGAGGCTG]

HLA-B*7:

[GGGTGGTCCTAGGGTGTCCC; GTCCTGAGTTTGGTCTCGCC]

OTC:

[TTCAATATTCCTCATCACTTTTG; AGGCCTTCCTGGTGATGACTGTGAC]

N-ras:

[ATGACTGAGTACAACTGGT; CTCTATGGTGGGATCATATT]

TNF:

[CAGCCACTGGAGCTGCCCCCTC; GGA CTGCTGAGCTCAAGGG]

FISHING PROTOCOL

Figure 1. Principle of the affinity purification procedure. A. A general description showing the digestion of the genomic DNA with a rare cutter, limited digestion with λ exonuclease, hybridization with a bridge RNA and biotinylated probe RNA, and retention of hybrid molecule onto a Vectrex-avidin matrix. B. Structure of hybridizing species. A pBSM13(+) plasmid containing a probe DNA subcloned from the end of a 150 kb *Sfi*I fragment and two SV40 DNA fragments A and B (plasmid 1). The plasmid 2 contains probe SV40 DNA fragment B in a pBSM13(+) vector. The RNA transcript of plasmid 1 with T3 RNA polymerase and biotinylated RNA transcript of plasmid 2 with T7 RNA polymerase, when hybridized together yield a RNA-biotinylated RNA hybrid. This hybrid can be hybridized to λ exonuclease resected genomic DNA and retained onto an avidin matrix. The stretch of single-stranded RNA in the tripartite hybrid molecule facilitated the release of the captured large fragment by its susceptibility to ribonuclease action.

Figure 2. Structure of the 150 kb SfiI genomic fragment and Time course and RNA concentration dependence of Hybridization with genomic DNA. A. The relative position of β -like globin genes in the 150 kb SfiI fragment. Arrows indicate the primers which are specific for the β -globin gene and from within a 500 bp HindIII fragment located toward the extreme 3' end of the large fragment. Both the sets of primers amplify 200 bp long DNA segments. (?) indicates a stretch of DNA which had been difficult to clone by conventional techniques. B. The 1 kb DNA piece from which a 452 bp SfiI to SacI fragment was isolated and subcloned into a pBSM13(+) plasmid vector containing a 215 bp long HindIII DNA fragment from SV40. C. PCR amplification of ribonuclease eluates from experiments in which the amount of RNA and the time of hybridization were changed. 40–60 μ g of SfiI digested 3.1.0 DNA was resected with λ exonuclease and hybridized to varying amounts of RNA transcripts for either 1 hour or 15 hours at 65°C as described in the Methods section.

The numbers in parentheses represent the amount of bridge RNA and time of hybridization. Lanes 1, 10 (5 μ g, 1hr); 2, 11 (5 μ g, 15hrs); 3, 12 (40 μ g, 1hr); 4, 13 (40 μ g, 15hrs); 5, 14 (60 μ g, 1hr); 6, 15 (60 μ g, 15hrs); 7, 16 (150 μ g, 1hr); 8, 17 initial DNA amplified and 1/10 fraction applied to gel. Lanes 1–8, primers specific for β -globin; Lanes 11–18, primers specific for 3' downstream HindIII fragment. Amplification was carried out for 30 cycles as described in the Methods section.

The size of the amplified product for these genes and DNA fragments is approximately 200 bp except for *N-ras*, which gives a 100 bp product. Some of these primer pairs yield artifactual bands representing primer dimers when PCR amplification is carried out for greater than 40 cycles.

Protocol for affinity enrichment of the 150 kb SfiI fragment containing the human β -globin gene locus

The agarose plug containing 40 to 60 μ g of SfiI digested and λ exonuclease resected genomic DNA was melted at 70°C and mixed with a probe RNA-biotinylated RNA hybrid in a hybridization buffer containing 0.5 M Na-Pi, pH 7.2, 4% SDS and 5 mM EDTA, in a total volume of 1.5 ml. Hybridization was carried out at 65°C for 1 hour. Biotinylated RNA was transcribed by T7 RNA polymerase from a HindIII digested pBSM13(+) plasmid containing a 135bp HindIII to AclI fragment from SV40 DNA. 20 μ g of bridge RNA, as determined by absorbance at 260nm, was annealed to an 8 fold molar excess of biotinylated RNA in the hybridization buffer at 65°C for 45 minutes before hybridizing to the λ exonuclease resected genomic DNA. The probe RNA-biotinylated RNA hybrid is schematically drawn in Fig. 1B, where the bridge RNA is a T3 transcript of EcoRI digested pBSM13(+) plasmid containing a 452bp left end piece from the SfiI fragment (Figs. 2A & B) and a 215bp HindIII DNA segment from SV40. The hybridization mixture was diluted

to 10 ml to a final concentration of 75 mM Na-Pi and 100 mM Tris, pH 7.5, and gently mixed with 350 mg Vectrex-avidin suspended in buffer. The Vectrex-avidin was previously incubated in 150 mM NaCl, 100 mM Tris, pH 7.5, 25 μ g/ml of sonicated and denatured salmon sperm DNA and then washed with buffer devoid of salmon sperm DNA. The diluted hybridization mixture was incubated with Vectrex-avidin at room temperature for 60 minutes with intermittent gentle mixing. The matrix was allowed to settle under gravity and the supernatant collected. The matrix was then washed successively with 10 ml of buffer A (150 mM NaCl, 100 mM Tris, pH 7.5, 0.5% SDS, 1mM EDTA) at room temperature [23°–25°]; 10 ml of buffer B (15 mM NaCl, 100 mM Tris, pH 7.5, 1 mM EDTA, 0.5% SDS) at room temperature; 10 ml of buffer B at 65°C for 20 minutes; 10 ml of buffer C (80 mM NaCl, 10 mM Tris, pH 7.5) at room temperature; and finally the specific DNA was eluted by incubating the matrix with 2 ml of buffer C containing 100 μ g/ml ribonuclease at 37°C for 60 minutes.

Construction of a fragment specific sublibrary in λ gt10 phage vector

DNA in the ribonuclease eluate of the Vectrex-avidin matrix was precipitated by ethanol in the presence of 20 μ g of glycogen. A small fraction of the precipitated DNA was saved for PCR assays and the remaining DNA was digested with 10 units of

EcoRI at 37°C for 15 minutes. The enzyme was inactivated by heating at 65°C. The digested DNA was precipitated by ethanol and ligated to 100 ng of *EcoRI* digested and dephosphorylated λ gt 10 vector arms in a 4 μ l reaction volume at 14°C for 12 hours. The ligated mixture was packaged into phage heads with Giga-pack Gold packaging extract purchased from Stratagene.

RESULTS

Affinity capture protocol

The principle of the method is schematically represented in Figs. 1A & B and has been used for the isolation of the human β -globin gene locus. The human β -globin cluster is encompassed in an approximately $\approx$ 150kb *SfiI* fragment (Fig. 2A), and is very well characterized (16, 17). This genomic region is also known to contain sequences which have been refractory to cloning (18). These retrovirus-like sequences are unstable in phage when propagated in a variety of strains of *E. coli* due to recombination. In the fishing experiment to isolate the 150kb genomic DNA fragment reported here, we used a 450 bp probe (Fig. 2B) subcloned from the 5' upstream region adjacent to the *SfiI* site. This large fragment was selected because of the sequence information available which facilitated synthesis of a set of PCR primers originating from within the globin gene and another set from a 500bp *HindIII* fragment sequence located some 10–15 kb upstream of the distal *SfiI* site. These PCR primers from the regions about 50 kb and 140 kb away from the fishing probe were helpful in assessing the integrity of the isolated fragment. A detailed protocol of the experiments is given in the Methods section.

Resection of genomic DNA

Control experiments revealed that commercial preparations of lambda exonuclease produced no detectable internal cleavage of large size genomic DNA. Although other enzymes such as exonuclease III are sufficiently pure for many experiments, we chose lambda exonuclease for resecting genomic DNA, because it causes stepwise removal of nucleotides from the 5' end of the DNA in a highly processive manner (19). Thus, an excess of enzyme can be used for resection to obtain a nearly homogeneous species of resected DNA. While it is reported that λ exonuclease utilizes DNA containing blunt ends as a preferred substrate (20), we found that DNA containing 5' recessed or 3' recessed ends can also be used. λ exonuclease resection can be carried out by diffusing the enzyme into agarose plugs containing digested genomic DNA, and the reaction can be stopped by diffusion of EDTA solution into the plug.

Hybridization of probe and target sequences, retention of hybrid molecules onto avidin matrix and specific release with ribonuclease

In the experiments described here, we have melted the agarose plug containing *SfiI* digested human DNA after λ exonuclease resection but prior to hybridization with a bridge RNA probe. The structure of hybridizing RNA species is shown in Fig. 1B. A vector was constructed in which a probe sequence could be cloned downstream of a 215bp *HindIII* DNA fragment from SV40, which in turn was immediately downstream of a T3 RNA polymerase promoter. After cleavage with *EcoRI* and transcription with T3 RNA polymerase, an RNA probe is generated that has at its 5' end SV40 sequences and at its 3' end sequences complementary to the region of target DNA exposed

by λ exonuclease. This transcript is termed the bridge RNA because it serves to connect the target DNA to a second segment of biotinylated RNA (Fig. 1B) complementary to the 5' end of the SV40 DNA. This RNA-biotinylated RNA hybrid and any target DNA sequence annealed to it at the other end will be retained on avidin immobilized to a solid matrix. In control experiments when [³²P]-labeled bridge RNA was annealed to biotinylated probe RNA, greater than 80% of the bridge RNA was retained on the avidin matrix. Binding of biotinylated RNA to an avidin containing matrix could be conducted in the presence of 0.5% SDS, which inhibited any traces of ribonuclease. Because of the extremely high affinity of biotin for avidin (21), the matrix could be washed repeatedly under stringent conditions (65°C, 0.1 $\times$ SSC, 0.5% SDS). Such washings were essential for reducing non-specific binding of DNA to the matrix and networking of DNA due to exposure of repetitive sequences at the ends consequent to λ exonuclease resection and subsequent hybridization.

As shown in Fig. 1B, the bridge and probe RNAs were so constructed that a segment of 80bp of single stranded RNA separated the duplex between the bridge RNA and the target DNA on one side and the RNA-biotinylated RNA hybrid on the other side. This strategy was planned to facilitate enzymatic release of the captured target DNA by the highly specific action of ribonuclease on single stranded RNA and the bridge was sufficiently long to avoid any problems due to steric hinderance.

RNA concentration dependance and time course of hybridization for enrichment of a 150 kb *SfiI* genomic fragment

The ribonuclease eluted DNA from various experiments, performed with different RNA concentrations and different times of hybridization, was amplified with both globin primers and the *HindIII* fragment primers. As indicated in Fig. 2C, the recovery of the isolated fragment increased from around 2–3% (at 1hr hybridization and 5 μ g RNA concentration) to about 10–15% (at 15 hrs. hybridization and 150 μ g RNA concentration). The recovery plateaued at around 40–60 μ g RNA concentration. The ribonuclease eluate yielded comparable amounts of amplified DNA with both β -globin and *HindIII* primers; this indicates that the hybridization and other manipulations are not causing any detectable internal breakage in the fragment. PCR amplification of the genomic DNA template with β -globin primers and *HindIII* primers was carried out with varying amounts of template. A proportional response of amplified product was observed in the range of template concentration used. The estimates of recovery were made by comparing PCR amplification of the test samples with a standard curve.

Assessment of recovery and purity of the isolated fragment

Two methods were employed to estimate the recovery and purity of the isolated fragment. One was to amplify the flow-through, wash fractions and the ribonuclease eluate from the avidin matrix using primers specific for β -globin genes and other randomly chosen genes. In the second method, the affinity captured *SfiI* fragment was completely digested with *EcoRI* and cloned into a λ gt10 vector. The library was then screened with the probes specific for globin genes.

A typical experiment with 40 μ g of genomic DNA, 20 μ g of bridge RNA and a 1 hour hybridization time was subjected to PCR analysis to assess the efficiency of the protocol. As shown in Fig. 3A, B, C and D, when various fractions were amplified,

Figure 3. PCR amplification of various fractions of a typical hybridization experiment with β -globin and HLA-B7 primers. 40 μ g of *Sfi*I digested 3.1.0 DNA was resected with λ exonuclease and hybridized to 20 μ g of appropriate probe transcripts for 1 hour at 65°C, as described in the Methods section. An aliquot each from initial DNA, flow-through, first wash, second wash, 65°C wash, Pre-ribonuclease wash and ribonuclease eluate were amplified by PCR. Panels A and B are ethidium bromide stained gels of samples amplified with primers specific for β -globin gene and HLA-B7 gene, respectively. Panels C and D are autoradiograms of Southern blots of panels A and B probed with β -globin cDNA and HLA-B7 cDNA, respectively. Lanes 1 to 6 in panels A and C represent initial DNA, flow-through, first wash, 65°C wash, Pre-RNase wash and RNase eluate, respectively. Lanes 1–6 in panels B and D represent flow-through, first wash, second wash, 65°C wash, pre RNase wash and RNase eluate, respectively.

Arrows indicate the size of expected bands which are $\approx$ 200 bp for β -globin and $\approx$ 800 bp for HLA-B7.

both the flow-through and wash fractions contained DNA detectable with the globin as well as control HLA-B7 primers, however, the ribonuclease eluate contained only the globin specific amplified product. No HLA-B7 specific sequences were detected in the ribonuclease eluate even after DNA blotting and hybridization to a specific probe, indicating specific retention and release of the globin fragment. PCR was carried out on various fractions to quantitate the recovery in this experiment (Fig. 4A). These results indicate that the recovery in the typical experiment is less than 5%. This can be increased to about 15% by longer hybridization times and using higher RNA concentrations (Fig. 2B). For comparison of recovery we used optimized hybridization conditions to purify a 6 Kb *Mlu*I-*Eco*RI fragment containing the HLA-B7 gene and embedded within a 40 kb cosmid. As shown in Fig. 4B, the recovery of this fragment ranged between 15–20%, and did not seem to be affected by the presence of melted agarose or genomic DNA. These results suggest that the recovery is not markedly affected by the size of DNA or any other discernable parameter. The nanogram amounts of isolated fragment are sufficient to make a deep library. However, the most important consideration in this approach is the purity of the isolated fragment. Most PCR experiments did not show amplification of any of the single copy sequences with primer pairs which did not lie on the *Sfi*I fragment. The results of one such experiment are shown in Fig. 5. The ribonuclease eluate

Figure 4. Recovery of the large *Sfi*I fragment in a typical experiment and its comparison with HLA-B7 cosmid DNA recovery. A. Equal aliquots of each of the fractions from initial DNA (lanes 1, 2), flow-through (lanes 3, 4), first wash (lanes 5, 6), second wash (lanes 7, 8), 65°C wash (lanes 9, 10), pre RNase wash (lanes 11, 12) and RNase eluate (lanes 13, 14, 15) from the experiment described in the legend to Fig. 3, were amplified with β -globin primers in a 50 μ l reaction volume for 25 cycles (lanes 1, 3, 5, 13), 30 cycles (lanes 2, 4, 6, 7, 14), 40 cycles (lanes 8, 9, 11, 15) and 50 cycles (lanes 10, 12), and applied to a 1.2% gel. The gel was ethidium bromide stained and photographed. B. Hybridizations were performed between *Mlu*I digested and λ exonuclease resected cosmid, containing the HLA-B7 gene, and appropriate RNA-biotinylated RNA transcripts, as described in the Methods section, under various conditions, viz. hybridization in 100 μ l (lanes 9, 10), 1.3ml (lanes 5, 6); 1.3ml, in the presence of *Mlu*I digested genomic DNA (lanes 3, 4); 1.3ml, in the presence of 0.15% agarose (lanes 7, 8). The ribonuclease eluate from these experiments were amplified with B7 specific primers for 30 cycles (even number lanes) and 40 cycles (odd number lanes) and applied to a 1.2% agarose gel for electrophoresis. Lanes 1 and 2 represent initial DNA, lane 11 is primers alone and lane 12 contains size markers.

and initial DNA were amplified with primers specific for β -globin and the 500bp *Hind*III fragment, as well as other genes such as TNF, OTC, HLA-B7 and N-ras which do not lie on the *Sfi*I fragment. None of these genes unrelated to the isolated fragment were detectable even after 60 cycles of PCR. This result would suggest that the isolated fragment is absolutely pure. However, as discussed in the following section, the isolated fragment is only 5% pure. The aliquot of DNA used for PCR assay in these experiments contained approximately 1–2 thousand copies of the *Sfi*I fragment, and the contaminating sequences at an average of one molecule per tube. The contaminating sequences are present, but below the level of sensitivity of detection by PCR assay.

Enrichment of β -globin genes in the fragment specific sublibrary

To check the quality of the recovered fragment, a λ gt10 phage library was constructed from the *Sfi*I fragment purified from about 40 micrograms of genomic DNA. The library had about 330,000 plaques of which 330 were specific for β - or δ -globin genes and 100 plaques were positive for the γ -globin gene as determined

Figure 5. Assessment of the purity of the DNA recovered by ribonuclease elution. Aliquots of the RNase eluate of the hybridization experiment described in the legend to Fig. 3, were amplified with the primers specific for the genes or DNA segments identified as β -globin, tumor necrosis factor (TNF), HLA-B7, Ornithine transcarbamylase (OTC), 500 bp *Hind*III fragment toward extreme 3' end of the *Sfi*I fragment (Hin) and N-ras, respectively. Lanes 1–4 represent amplification of the initial DNA and lanes 5–10 represent amplification of ribonuclease eluate. Amplification was carried out for 60 cycles under the conditions mentioned in the Methods section.

by hybridization to specific probes. Hybridization to [³²P]-labeled total genomic DNA suggested that about half of the clones in the library had DNA of human origin. The average abundance of β -globin gene in a genomic library is one in a million clones (assuming that average size of an *Eco*RI fragment is between 4–5 kb). The frequencies of β -globin and γ -globin genes in the sublibrary are 1/1000 and 1/3300 respectively. Thus the minimum estimated purification of the *Sfi*I fragment was between 350 and 1000 fold. The efficiency of cloning a known amount of an *Eco*RI insert DNA suggests that the recombinant plaques represented in this sublibrary will have arisen from 1–2 ng of DNA. Since the recovery is about 5% in this experiment, the actual fragment will represent about 100 pg, and hence by this calculation the purity of the fragment is greater than 5%. This is 1000 fold greater than its abundance in the complete human genome. While these calculations are only approximate, the data clearly support the contention that a substantial enrichment of the desired *Sfi*I fragment was achieved.

DISCUSSION

This method is applicable for isolating DNA fragments obtained by digestion with any restriction endonuclease, provided a suitable probe is available from the end of the fragment. To apply the method systematically to extensive regions of the genome, it would be necessary to obtain linking clones devoid of any repetitive DNA, but a probe of 200–300 nucleotides devoid of repeats would suffice for the fishing protocol and these can be isolated from most linking clones. The results described in this paper do not address the upper size limit of the DNA fragment that can be isolated. At present we are exploring the feasibility of isolating DNA fragments a million basepairs long. Preliminary experiments, in which agarose plugs containing yeast chromosomes were melted, diluted and electrophoresed, indicate that these manipulations do not cause any detectable breakage of such large DNA fragments and point toward the feasibility of isolating them by the protocol described here. Currently, we

are exploring this method for the isolation of fragments which are larger than routinely clonable in YAC vectors.

The approach described in this paper does not involve any cloning in the selection step and hence, facilitates isolation of regions of the chromosome with sequences which are difficult to clone due to their tendency to undergo rearrangement or exhibit negative biological selection in *E. coli* or yeast. These isolated fragments can then be used as probes to recover cDNAs encoded in the region containing such sequences.

The level of purity obtained in the experiments reported here is sufficient for some screening purposes. Size selection of digested DNA by PFG before hybridization or use of DNA from somatic cell hybrids could decrease the non-specific background of human DNA in the sublibrary and consequently increase the enrichment. Alternatively, sequences within the enriched DNA fraction can be PCR amplified using *Alu* or Line primers containing cloning sites and subsequently cloned as previously described (22,23). We are attempting further enrichment of the sublibrary to near purity by denaturation and limited renaturation procedures as mentioned in the following section.

The present protocol provides a systematic and relatively simple way to proceed from a library of linking clones to a set of sublibraries enriched for desired sequences adjacent to linking clones. Such libraries can be useful in several ways for genomic mapping and sequencing. For example, denaturation and limited renaturation of non-repetitive sequences in a restriction digest of a 'fished' fragment or of inserts from an enriched library would yield further enrichment proportional to the second power of the original purification. The clones from such a nearly pure library can be used directly for sequencing and/or as probes to isolate relevant clones from YAC, cosmid or phage libraries. In addition, the 'fished' large fragment can be immobilized to nylon membrane and used to select for cDNAs by hybridization and the resulting clones could be further purified by denaturation and renaturation to obtain a library of expressed sequences corresponding to the initial large fragment. The ease of processing and minimum manipulations make it possible to simultaneously handle multiple samples and potentially automate the entire procedure.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. O. Smithies for providing λ phage clones from globin region, and Mr. Leif Madsen for the excellent technical preparation of this manuscript. This work was supported by an Outstanding Investigator Grant: CA 42556, from the National Cancer Institute to Sherman M. Weissman; and a National Institutes of Health Grant: GM 40633, to David C. Ward.

REFERENCES

1. Smith, C.L., Lawrance, S.K., Gillespie, G.A., Cantor, C.R., Weissman, S.M. and Collins, F.S. (1987) *Methods Enzymol.* **151**, 461–489.
2. Wallace, M.R., Fountain, J.W., Brereton, A.M. and Collins, F.S. (1989) *Nucleic Acids Res.* **17**, 1665–1677.
3. Schwartz, D.C. and Cantor, C.R. (1984) *Cell* **37**, 67–75.
4. Carle, G.F., Frank, M. and Olson, M.V. (1986) *Science* **232**, 65–68.
5. Chu, G., Vollrath, D. and Davis, R.W. (1986) *Science* **234**, 1982–1985.
6. Smith, C.L., Econome, J.G., Schutt, A., Klco, S. and Cantor, C.R. (1987) *Science* **236**, 1448–1453.
7. Collins, J. and Hohn, B. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4242–4246.
8. Burke, D.T., Carle, G.F. and Olson, M.V. (1987) *Science* **235**, 806–812.
9. Radding, C.M. (1986) *Annu. Rev. Genet.* **16**, 405–437.
10. Rigas, B., Welcher, A.A., Ward, D. and Weissman, S.M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9591–9595.
11. Gladstone, P., Fueresz, L. and Pious, D. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1235–1239.
12. Smith, C.L., Warburton, P.E., Gaal, A. and Cantor, C.R. (1986) In J.K. Setlow and A. Hollander (Eds.), *Genetic Engineering*. Plenum, New York, Vol. 8, pp. 45–70.
13. Feinberg, A.P. and Vogelstein, B. (1984) *Analyt. Biochem.* **137**, 266–267.
14. Church, G.M. and Gilbert, W. (1988) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
15. Saiki, R.K., Gelfand, D.H., Stoffel, S., Clark, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. (1988) *Science* **239**, 487–491.
16. Collins, F.S. and Weissman, S.M. (1984) *Prog. Nucleic Acid Res. Mol. Biol.* **31**, 315–458.
17. Li, Q., Powers, P.A. and Smithies, O. (1985) *J. Biol. Chem.* **260**, 14901–14910.
18. Mager, D.L. and Henthorn, P.S. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7510–7514.
19. Radding, C.M. and Carter, D.M. (1971) *J. Biol. Chem.* **246**, 2513–2518.
20. Little, J.W. (1981) In J.G. Chirikjian and T.S. Papas (Eds.), *Gene Amplification and Analysis*. Elsevier, Vol. 2, pp. 136–147.
21. Green, N.M. (1975) *Adv. Protein Chem.* **29**, 85–133.
22. Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H. and Caskey, C.T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6686–6690.
23. Ledbetter, S.A., Nelson, D.L., Warren, S.T. and Ledbetter, D.H. (1990) *Genomics*, in press.