460

707490

DOE F 1332.16 (10-84)
(Formerly RA-427)

OMB Approval
No. 1910-1400

U. S. DEPARTMENT OF ENERGY

UNIVERSITY CONTRACTOR, GRANTEE, AND COOPERATIVE AGREEMENT RECOMMENDATIONS FOR ANNOUNCEMENT AND DISTRIBUTION OF DOCUMENTS

See Instructions on Reverse Side

1. DOE Report No. DOE/ER/60692-1

2. DOE Contract No. DE-FG05-88ER60692

3. Title "Mapping and ordered cloning of the human X chromosome

4. Type of Document ("x" one)
   - [x] a. Scientific and technical report
   - [ ] b. Conference paper:
     - Title of conference _____
     - Date of conference _____
     - Exact location of conference _____
     - Sponsoring organization _____
   - [ ] c. Other (Specify) _____

5. Recommended Announcement and Distribution ("x" one)
   - [ ] a. Unrestricted unlimited distribution.
   - [x] b. Make available only within DOE and to DOE contractors and other U. S. Government agencies and their contractors.
   - [ ] c. Other (Specify) _____

6. Reason for Recommended Restrictions

   Pre-publication results are described.

7. Patent and Copyright Information:
   Does this information product disclose any new equipment, process, or material? [ ] No [x] Yes If so, identify page nos. 4-5 (Alu PCR)
   Has an invention disclosure been submitted to DOE covering any aspect of this information product? [x] No [ ] Yes
   If so, identify the DOE (or other) disclosure number and to whom the disclosure was submitted.
   Are there any patent-related objections to the release of this information product? [x] No [ ] Yes If so, state these objections.
   Does this information product contain copyrighted material? [x] No [ ] Yes
   If so, identify the page numbers _____ and attach the license or other authority for the government to reproduce.

8. Submitted by — Name and Position (Please print or type)
   David L. Nelson, Ph.D. Senior Research Associate

   Organization
   Baylor College of Medicine, Institute for Molecular Genetics

   Signature David L Nelson — Phone 713-799-6552 — Date 1-1-89

   REPOSITORY Oak Ridge Operations Office

   COLLECTION Procurement + Contracts Div. AD-421

   BOX No. Active Records Contract for Human Radiation Exps. Pgt.

   FOLDER

   FOR DOE OR OTHER AUTHORIZED USE ONLY

1021437

AD-44

9. Patent Clearance ("x" one)
   - [ ] a. DOE patent clearance has been granted by responsible DOE patent group.
   - [ ] b. Report has been sent to responsible DOE patent group for clearance.

APPROVED FOR RELEASE OR PUBLICATION - O.R. PATENT GROUP
BY DM DATE 5/29/89

A. Progress 7/1/88-12/1/88

Progress on this project has been made in several areas. These will each be discussed in detail below. Overall, the progress has been quite impressive, with some of the major goals of the application nearly achieved. The areas currently underway are: 1) The establishment of a large number of cloned DNA segments from the region of the X chromosome retained in the X3000-11.1 cell line, 2) The development and use of a sub-chromosomal localization panel of somatic cell hybrids for mapping cloned probes from the region, 3) The development of *Alu* PCR as a technique for isolating human-specific DNA from a variety of sources, 4) The construction of a long-range restriction map of Xq28, and 5) The establishment of methods for identification and isolation of expressed sequences.

Isolation and Mapping of Clones from Xq24-ter

A recombinant library in the λ DASH vector was constructed from DNA derived from the X3000-11.1 cell line. Screening a portion of this library with radiolabeled total human DNA has provided a total of 450 purified clones. This represents approximately one-fourth of the human sequences in the cell line. These clones have been stored in 8 by 12 arrayed "titer tubes" for easy storage and access. DNA has been prepared from approximately 200 of these clones, and fifty of these have been mapped on a panel of somatic cell hybrids. Results from this mapping of clones are included in Table 1. Several clones were apparently of hamster origin. Additionally, a disproportionate number of clones was found to reside in the interval Xq26-Xq27.1. In fact, over one-half of the mapped clones were derived from this region. Inspection of the cell line by cytogenetics revealed no abnormalities, and use of several of the clones mapping to this region demonstrates no visible amplification of the region in the hybrid. This remains a surprising and unexplained result. We have constructed new libraries in both λ and cosmid vectors. Analysis of a like number of clones from each of these new libraries will allow us to determine whether this phenomenon is related to the original library.

In order to determine the localization on the X chromosome of each of the λ clones described above, we established a mapping panel containing DNA from several hybrid cell lines with varying amounts of the X chromosome in a hamster background. This panel is shown in Figure 1. Creation of the panel required both the establishment of new hybrids and the collection of previously established lines. Additional hybrids are also being pursued in order to further refine the mapping panel (see Table 2).

Analysis of λ clones was carried out using DNA derived from human positive clones selected at random. DNA from the entire phage clone was radiolabeled with ^{32}P by random hexamer primed labelling with Klenow. Each probe was applied to a Southern filter carrying *Eco RI* digested DNA from each member of the mapping panel. Probes and filters were pre-associated with an excess of human placental DNA in order to eliminate hybridization to repeated sequences in the probe. Establishment of this capability has been extremely valuable in this and other areas of this project. We have developed methods for the isolation of human placental DNA in large scale, and established a working relationship with a local maternity ward in order to obtain the materials. Typical yields from a single placenta are 800-900 mg of DNA. It is possible with these methods to unambiguously map 85-90% of the λ clones. Some clones have required second hybridizations, and the usual difficulties related to Southern hybridization are also noted. The results of mapping studies through late November are shown in Table 1.

Establishment of *Alu* PCR Technique

One of the major goals of the original proposal was the establishment of the *Alu* PCR method. This goal has largely been met. Oligonucleotides derived from the consensus *Alu* repeat sequence from humans have been synthesized and tested for their ability to amplify human DNA specifically from a variety of sources. One region of the *Alu* sequence has been found to be useful in such experiments. All other regions have proven to contain sufficient homology to rodent sources to prime amplification of the rodent sequences in somatic cell hybrids when oligonucleotide primers derived from these sequences are used. Figure 2 shows the *Alu* consensus sequence and positions of oligonucleotide probes thus far synthesized. Of particular note is that the location of the sequence capable of priming

10214438

amplification in a human specific manner is in the 31bp primate-specific insert in the second repeat of the human sequence.

Conditions have been established which permit amplification of the human sequences specifically from somatic cell hybrid DNA, both bulk DNA and DNA that has been size-fractionated by pulsed-field electrophoresis, as well as from YAC and λ clones. The amount of sequence amplified appears dependent upon the particular oligonucleotide and the amount of target human DNA as a proportion of the total target DNA. Results from the original effective primer, TC-65 will be described in a manuscript in preparation. This primer used alone allows amplification between *Alu* sequences in opposite orientation and within a distance of up to 5-6 kb. It does not amplify every such pair of *Alu* sequences, demonstrating some preference for particular *Alu* repeats. This can be demonstrated in the X3000-11.1 hybrid, where amplification with TC-65 shows individual bands numbering between 25 and 50. Since this hybrid contains about 40 Mb of human DNA, and *Alu* sequences are thought to occur every 4 kb on average, there should be approximately 10,000 *Alu* repeats in this region. Of these, a reasonable estimate of copies in opposite orientation and within an amplifiable distance would be 1500. Since only 50 species appear with this primer, there must be characteristics of these 1/30 pairs of *Alu* repeats that favor amplification with this primer. Fortuitously, one of the sequences (a 3.5 kb band) amplified by this primer from the X3000-11.1 cell line hybridizes with a cDNA probe for the HPRT mRNA. While confirming that a *bona fide* human-specific sequence is amplified from this somatic cell hybrid, this sequence also affords the opportunity to investigate the sequence of the *Alu* repeats involved in this amplification. To localize the repeats within the hprt gene, we amplified DNA from 6 phage clones encompassing the entire coding region using the TC-65 primer. Two of the clones demonstrate the expected 3.5 kb band, which also hybridizes to the cDNA probe. Since these clones overlap and contain only exons 7, 8 and 9, we can predict that the *Alu* repeats priming synthesis of this sequence are located 3.5 kb from one another, in opposite orientation and spanning one or more of the last three exons in the gene. Inspection of the sequence from this region (which is being generated in a collaborative arrangement with W. Ansorge at EMBL/Heidelberg) has allowed us to identify the *Alu* sequences involved. The *Alu* repeats span exon 9. This sequence information should help us understand the selective nature of the amplification with the TC-65 oligonucleotide, as well as provide a pair of *Alu* sequences that can be amplified with other primers to determine the parameters important in this reaction.

Additional primers complementary to the same region of the *Alu* repeat have also been found to amplify human DNA specifically from hybrid cells. Primers without tails have been particularly useful in this regard, as they allow considerably more sequence to be amplified from a given DNA source. As an example, the X3000-11.1 line when amplified by these primers shows a smear of bands numbering in the hundreds, while hybrids containing smaller amounts of human sequence (1-2 Mb) show a discrete number of bands (10-20). Amplified sequence of this type has already proven useful for screening libraries as a probe specific to the region in the hybrid. These probes should allow the extension of the *Alu* PCR technique to a number of different areas, including amplification from pulsed-field gel fragments, YAC and cosmid clones. Characterization of additional oligonucleotides is important in defining the parameters important in this reaction.

Long Range Restriction Mapping in Xq24-ter

One of the aims of the original proposal was the development of a long range restriction map of the Xq24-ter region. We have initiated this in the Xq28 region utilizing previously isolated clones. The map we have developed stretches from G6PD/GdX to the color blindness locus, covering over 3 Mb of DNA. A difficulty with this map is the lack of probes in a 1.5 to 2 Mb region between Factor VIII and St14. We are attempting to develop techniques for the isolation of DNA from this region directly from pulsed-field gel slices of DNA from the X3000-11.1 hybrid. Two methods are being used. The first involves the *Alu* PCR protocol described above. The technique is not hindered by the presence of agarose in the reaction, thus PFG slices can be amplified directly. No evidence of hamster DNA amplification is found, while a considerable amount of DNA is produced. Currently, we are attempting to isolate clones from this material. The second method involves cloning the DNA from the gel slice directly. The gel fragment is dialyzed to remove the gel buffer, then equilibrated with restriction digestion buffer and cut with *Eco RI*. Digested DNA from the gel slice is eluted and cloned into a λ phage vector. Clones are screened for the presence of human repeat sequences, and selected. This method has provided two candidates for the Xq28 gap. The *Alu* PCR method is still being developed, but should

10214 39

constitute a more rapid protocol for isolation of human DNA of a given PFG fragment specifically from a somatic cell hybrid.

Development of Methods to Identify Expressed Sequences

This area represents a departure from the stated aims in the original proposal, but is of sufficient importance that we have devoted some resources to its development. Establishment of techniques for the identification and/or isolation of DNA sequences containing expressed genes will be important not only for genome projects in general, but also for the pursuit of disease loci in particular. In this regard, we have attempted to adapt the *Alu* PCR method to the isolation of expressed sequences from somatic cell hybrids, again utilizing X3000-11.1 as a test case. In this general protocol, oligonucleotides with *Alu* specificity have been used to prime synthesis of cDNA from nuclear RNA isolated from either X3000-11.1 or its rodent parent, RJK 88. Using TC-65, this procedure demonstrates cDNA can be made from X3000-11.1, but not from RJK 88. This human specificity appears to derive from transcribed sequences from the human region contained in the hybrid. These results are still preliminary, and we are attempting to use the cDNAs both as probes to identify λ clones containing the expressed region, and as substrate for cloning these sequences directly. PCR is also being used to amplify these transcripts. While this protocol may not be useful for the isolation of tissue or developmental stage specific RNAs, it should allow the identification and isolation of a number of "housekeeping" genes from a region of interest.

Our second method for identifying expressed sequences uses oligonucleotide probes to determine whether cloned fragments might contain sequences unique to expressed genes. Thus far, two oligonucleotide probes show considerable promise. The first is designed to recognize the consensus sequence for transcription initiation factor SP1. SP1 binding sites occur in a large number of gene promoter regions. This oligonucleotide recognizes a subset of clones derived from the X3000-11.1 region as well as control clones containing expressed genes. The second probe is designed to recognize the consensus sequence for a splice site acceptor. It too recognizes some of the clones from the X3000-11.1 region and the positive control clones. Additional probes are being developed, and the clones recognized by these probes are being analyzed to determine whether they contain expressed sequences.

1021440