NOTICE

**CERTAIN DATA
CONTAINED IN THIS
DOCUMENT MAY BE
DIFFICULT TO READ
IN MICROFICHE
PRODUCTS.**

CONF-9109321-1
CONF-9110302-1

DOE/ER/61233--1

DE92 004022

CHROMOSOME-SPECIFIC cDNAs/STSs

Progress Report

For Period September 1, 1991 - February 23, 1992

Marcelo Bento Soares (PI) and Argiris Efstratiadis (Co-PI)

Columbia University

722W 168th St. Box=41, New York, N.Y. 10032

November 1991

Prepared for

THE U.S. DEPARTMENT OF ENERGY
AGREEMENT No. DE-FG02-91EP61233

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

pa

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product or process disclosed or represents that its use would not infringe privately-owned rights.

ABSTRACT

This project has three main goals: (1) to construct high-quality normalized cDNA libraries from human tissues, (2) to develop methods for isolation of chromosome-specific cDNAs, (3) to contribute sequence information to the expanding cDNA/EST database.

Our short term goal has been to construct a very high quality cDNA library from a 3 month post-natal total human brain. Several libraries were constructed and carefully characterized by DNA sequencing and restriction analysis while protocols were being streamlined. We have now generated a library which has the following main features: (a) very high complexity (10^7 recombinants), (b) short poly (A) tails; (c) long size inserts (1 kb average); (d) undetectable co-cloning events; (e) low background of non-recombinants.

The next step will be to generate and test a normalized version of this brain cDNA library. Subsequently, single-stranded versions of the normalized library will be utilized for isolation of chromosome 13-specific cDNAs.

Percent efforts: The PI and Co-PI devoted 80% and 10% of their times to this project, respectively. The same percent efforts are expected to be devoted by both the PI and Co-PI during the second budget period.

Introduction

It is now widely acknowledged that the quality of the existing human brain cDNA libraries is not satisfactory. For example, while cataloguing sequences from human brain cDNAs, Adams et al (1991) reported that about 22% (134 sequences out of the 600 total) of the sequenced cDNA clones contained either no insert between the Eco RI cloning sites or consisted entirely of polyadenylate. In addition, 32% of the clones (197 sequences) consisted of matches to one of the following human sequences: mitochondrial genes (12.8%), repetitive elements (10.4%), ribosomal RNA genes (2.7%) and other nuclear genes (8.6%). In other words, 54% of their effort was abortive.

An additional serious problem that we have observed in several cDNA libraries regards co-cloning of cDNAs that correspond to different mRNAs. Whenever sequencing is done from both ends of a cDNA clone, one must make sure that both ends belong to the same mRNA. This is not always easy to assess. However, if neglected, a substantial amount of misleading information could be generated.

Our short term goal was to generate a high quality human brain cDNA library, which should have the following features:

- . 10^6 - 10^7 recombinants;
- . large size inserts (1 kb averaged);
- . short poly(A) tails;
- . no incidence of co-cloning events;
- . low background of non-recombinant clones (< 5%)

Results

1) Construction of a human brain cDNA library in the λ afmid vectors.

As described in the original grant application, we (P. Jelenc, M. B. Soares, A. Murphy and A. Efstratiadis) have developed compound cloning vectors (" λ afmids"; λ mbda-filamentous phage-plasmid) consisting of both phage λ mbda and plasmid components. The plasmid component includes both the plasmid and filamentous phage origins of replication, an antibiotic resistance gene and the cloning sites.

There are two λ afmids: A1 (ampicillin resistance; cloning sites Hind III-Bam HI-Not I-Eco RI) and K2 (kanamycin resistance; cloning sites Hind III-Not I-Bam HI-Sna BI- Eco RI).

During the last two years, we produced four generations of A1 and K2 vectors and constructed over 20 different cDNA libraries, while improving and streamlining protocols.

cDNAs are directionally cloned in both orientations into the A1 and K2 λ afmid vectors. Thus, two libraries are generated from the same cDNA preparation. More specifically, one half of the synthesized blunt-ended cDNA is ligated to Eco RI adaptors and the other half to Hind III adaptors. Since the primer utilized for first strand synthesis contains the sequence for the Not I restriction endonuclease, after Not I digestion

cDNAs will have two different sticky ends: Not I-Eco RI ends or Not I-Hind III ends, respectively. Not I-Eco RI cDNAs are cloned into the lacmid K2 and Not I-Hind III cDNAs are cloned into the lacmid A1 vector. Thus, the two resulting libraries, in the form of single-stranded circles, will have complementary strands. A1 single-stranded circles carry the sense (mRNA-like) strand, and K2 single-stranded circles carry the anti-sense strand. After hybridization of the two complementary libraries to a moderate Cot, the unhybridized, normalized, single-stranded library can be purified from the hybridized material over hydroxyapatite and electroporated into bacteria for propagation under the appropriate antibiotic selection.

We have successfully constructed a human brain cDNA library which has been documented not to present any of the undesired features so commonly encountered in the available cDNA libraries. The following is a brief description of the main features of this library.

a) Number of recombinants

Lacmids are both lambda and plasmid cloning vectors, and we are obtaining equally high cloning efficiencies (10^6 to 10^7 recombinants per 50 ng cDNA) by in vitro packaging or electroporation. The brain library that we constructed recently, was cloned directly into the plasmid component. We obtained a total of 10^7 recombinants from 50 ng of size-selected cDNA.

b) Insert lengths

We have tested several gel filtration columns for size selection of cDNAs. The very best results were obtained with a Bio-gel A50am column, which we run exactly as recommended by Huynh et al (1985). Insert lengths in the brain library are 500 bp or larger.

c) Length of Poly(A) tails in cDNA clones

Most protocols recommend addition of inadequate amounts of oligo-dT primer for 1st strand synthesis. We reasoned that if saturating amounts of primer were used, since reverse transcriptase cannot strand displace efficiently, we should be able to avoid synthesis of long stretches of tail. Indeed this strategy works very satisfactorily. Using a two-fold masswise excess of primers, we have been able to generate cDNA clones with tails that are most of the time just the length of the primer itself. We never fail to obtain a sequence from the 3' end of our clones due to presence of long tails. In summary, this seems to be a rather serious problem that we have completely solved.

d) Co-cloning events

Because of the very nature of our cloning strategy we have a very easy assay to check for co-cloning events in our cDNA libraries. 1st strand cDNA is primed with a Not I-oligo-dT primer. After second strand synthesis and appropriate adaptor ligation, cDNAs are digested with Not I

and directionally cloned into the lamid vectors as described above. Therefore, digestion of any cDNA clone with Not I should result in the linearization of the plasmid. No fragments should ever be released. The release of a Not I fragment is an indication of the occurrence of co-cloning. Using this assay, we did not detect any co-cloning event in over 100 clones that were analyzed.

2) Sequencing of cDNAs

We have recently acquired a Biomek workstation to further expedite the sequencing endeavor of the cDNAs. We have sequenced over 200 cDNAs which are currently being analyzed.

3) Conclusion

In summary, we feel that we have successfully accomplished the short term goal of this project. We have generate a high quality brain cDNA library, and we have also obtained sequence from over 200 clones, which are currently being analyzed for submission to GenBank. During the next period, we expect to be able to generate and test a normalized version of this brain library and to start obtaining chromosome 13-specific cDNAs.

Literature cited

Huynh, T.V., Young, R.A. & Davis, R.W. (1985). Constructing and screening cDNA libraries in lambda-gt10 and lambda-gt11. In "DNA Cloning, Vol 1, D.M.Glover, Ed., IRL Press.

December 3rd, 1991

Marcelo Basto Soares, Ph.D. (PI)
Assistant Professor of Neurogenetics
Department of Psychiatry, Columbia University

**DATE
FILMED
314192**

I