Survey of transcripts in the adult Drosophila brain

ABSTRACT

Our analysis of this unique Drosophila brain library suggests that the number of genes may be underestimated in this organism. This work complements the Drosophila genome project by providing information that facilitates more complete annotation of the genomic sequence. This library should be a useful resource that will help in determining how basic brain functions operate at the molecular level.

INTRODUCTION

Background Drosophila melanogaster is an important model organism. After more than 50 years of study, the anatomy of the brain is well described and many brain functions have been mapped to particular substructures. The adult brain is composed of approximately 200,000 neurons which are organized into discrete substructures. The optic lobe (composed of the lamina, medulla, lobula and lobula plate) is primarily involved in the processing of visual information from the photoreceptors and sending that information to the central brain. The antennal lobes are chiefly responsible for the processing of olfactory information. The mushroom bodies are involved in olfactory learning and memory and other complex behaviors. A group of approximately six neurons in the lateral protocerebrum are sufficient to drive circadian rhythms in locomoter activity. The central complex, although poorly understood, appears to be involved in motor coordination. Despite our increasing knowledge of Drosophila brain anatomy and function, relatively little information is available concerning the molecules expressed in the brain that coordinate function and manifest behavior. Classic methods of identifying genes involved in neural function include behavioral screening of mutagenized flies, then rescreening candidate lines for pleiotropic effects due to developmental defects. This process is both laborious and time consuming. To augment this genetic approach, sequencing of random cDNAs is proving effective in identifying genes expressed in a specific cell type. Much information has been collected through the analysis of expressed sequence tags (ESTs). Using this approach, sequence information is gathered from one or both ends of a cDNA and cataloged to determine the complexity of an mRNA population. Here, we use a modified EST approach and completely sequence novel cDNAs. Others have used a similar approach by shotgun sequencing concatenated cDNA inserts. One goal of our work was to begin to develop a catalog of transcripts expressed in the brain. These transcripts, because of the location of their expression, are expected to contain a higher proportion of clones that are involved in neuronal function. Many Drosophila head libraries have been used to isolate cDNAs that correspond to genes identified by genetic screens for their involvement in brain function. Several transcripts identified in this manner are expressed at a relatively low level (dunce, CREB, dco, period, timeless, dissonance). The Drosophila brain makes up only a small part of head tissue (approximately 14% dry weight). By eliminating non-brain tissues, we increase the relative representation of rare neural transcripts in this unique library. We began a catalog of the genes expressed in the brain of adult Drosophila in support of more conventional methods of understanding brain function. Cataloging sequence information and publishing the data through electronic databases has enriched molecular science in general. In a matter of a few minutes, one can use information from a single sequencing reaction to identify a gene that was sequenced by another laboratory, and one maybe able to deduce the function of the

isolated clone. This set of tools facilitates molecular work in virtually every branch of biological sciences. This report details construction, quality analysis and initial characterization of a unique library created from adult Drosophila brains. Surprisingly, we discovered that 11% (29 clones) of the Drosophila brain cDNA clones that were randomly chosen for analysis are not matched with any EST sequence generated in support of the Drosophila genome project (as of 10 October 2000). Further, the genes encoding 59% of these novel ESTs are not predicted by algorithms used for fly genome annotation. From our analysis of ESTs that do not correspond to one of the 13,601 annotated genes, we predict that the number of genes in the Drosophila genome may be underestimated by 10-15% (approximately 1,300 to 2,000 genes).

CONCLUSION

Conclusions The initial analysis of an adult Drosophila brain library is presented here. Somewhat surprisingly, we observe no clear connection between the abundance of a transcript and its appearance in a sequence data bank. However, molecular screens that are directed towards isolating rare transcripts may skew the transcript-related data in sequence banks towards less abundant molecules. As shown in Figure 1 and Table 2, we have identified and sequenced 29 novel clones that do not match with other known expressed sequences (but do match with fly genomic sequence information), 85 clones that are matched with EST data, 71 clones that were previously reported Drosophila sequences, 39 clones that contain ribosomal protein sequences, 27 clones that are matched with genes previously reported for other organisms (isologs, Table 3) and 20 clones corresponding to mitochondrial sequences. Why did we recover such a high percentage of novel sequences? Libraries made from brain tissue are proposed to have a higher complexity of transcripts than libraries made from other tissues. Therefore, EST screens of brain libraries should yield larger numbers of independent transcripts, as a result of the increased transcript complexity within brain tissues. Another possible explanation for the surprisingly large number of novel cDNAs identified in our analysis is that our library is not normalized. It has been proposed that hybrids form between poly(dA) and poly(dT) sequences during the hybridization/subtraction reaction and that these sequences are subsequently lost. An ultimate goal of this project is to create a database of all the transcripts expressed in the Drosophila brain and to correlate this information with their patterns of expression in the brain. This type of a database would be a valuable resource and could be used in comparative studies with other organisms. Comparisons of transcripts from organisms with relatively simple brains (Drosophila) to organisms with more complex neural function (humans) may offer insights into basic brain function and aid in the identification of transcripts involved in higher-order brain functions. The 35 clones that appear enriched in the brain may identify proteins or RNAs that are involved in a brain-specific function. Transcripts identified in this library can be directly tested for protein-protein interaction using the yeast two-hybrid capability of the library, making it a good resource for many areas of study. Our analysis of this unique brain library demonstrates that many transcribed regions of the Drosophila genome remain undiscovered, and that approximately 2,000 more genes may be identified. Genome annotation efforts emphasize identifying protein-coding regions. Thus, it is possible that some of the ESTs lacking a corresponding predicted gene were missed during genome annotation because an open reading frame (or one of sufficient size) was not predicted. Complete genomic sequences are excellent resources, and extensive annotation of a genome makes the sequence information even more powerful. Current

software is not sufficient to identify all transcribed regions within the genome. As of the year 2000, EST data for 24,193 clones from adult Drosophila head libraries is reported and estimated to represent over 40% of all Drosophila genes. Our results confirm that not all transcribed regions of the genome are identified and that EST analyses are essential for accurate and complete genome annotation.