RESEARCH NEWS

Dysfunctional genomics: BACs to the rescue

Melvin I. Simon

The Human Genome Project has provided potential access to all of our genes and their DNA sequences. The challenge now is to understand what each gene does, the mechanisms underlying the specificity of gene expression in appropriate cells at particular times in development, and the nature of the interactions among groups of genes that control cellular activity. This emerging area of study—the post-Human Genome Project field of inquiry—has been referred to as functional genomics.

A variety of shortcuts to establish gene function are being explored. Sequence comparisons and sophisticated computational manipulations (informatics) can sometimes provide clues to gene function. Studies of cell-specific or tissue-specific patterns of transcription using sophisticated "DNA-chips" can be used to determine how gene expression is controlled during development in response to hormones, growth factors, or disease. However, the ultimate answers to questions about function require that we be able to modify genes and their controlling elements (e.g., promoters, silencers, and splice sites) and directly demonstrate their activity in cells and in organisms. Because eukaryotic genes encompass large segments of DNA often exceeding 100 kb pairs, these kinds of manipulations are difficult. In addition, large fragments of DNA cloned into yeast vectors (YACs) have been found to undergo rearrangement as a result of aberrant recombination, making them difficult to use. In this issue, Yang et al.[1] provide an approach to the problem.

Starting with a mouse gene cloned into a bacterial artificial chromosome (BAC)[2], these authors adapted some of the techniques and manipulations invented over the past 40 years in studies of bacterial genetics to insert into the mouse gene a reporter that can signal tissue-specific expression. They introduced this ~100 kb construct into the germline of a mouse and demonstrated all of the appropriate tissue-specific expression. Previous work with a smaller, more conventional construct did not show the complete pattern of expression found in normal mice. Presumably, this was due to the absence of remote regulatory sequences.

The approach that Yang et al. used involved the construction of a DNA fragment that had segments of sequences derived from the cloned gene flanking a reporter gene (*lacZ*). The homologous segments were used to guide recombination of the modifying element into the cloned gene. However, in the BAC system, the integrity of complex cloned genes is maintained in *Escherichia coli* by having the genes present on a single copy vector and by eliminating recombination functions from the bacterium. The absence of the major recombination pathway allows genes that have repeated sequence elements to be archived in a stable way in the bacterial cell over many generations without the scrambling that results from intergenic and intragenic recombination.

In order to generate the recombinant gene, the authors introduced recombination function into the bacterium in a transient manner. It persisted long enough to allow integration and substitution of the resident mouse sequences to occur and then the recombination function was lost. There was some concern that even the transient presence of recombination functions might rearrange the cloned mouse gene; however, it is clear that this did not occur. These results do not rule out the possibility that in some fraction of BAC clones scrambling might occur in genes that carry multiple repeat sequences. Nonetheless, the potential of the method remains enormous.

There are many systems where regulatory elements are very distant from coding regions. The size of the BAC clones and the ease with which these techniques permit the clone to be manipulated will allow the entire unit to be modified and to be introduced directly into the mouse by simple transgene injection. The large clone size could provide "buffering" around regulatory regions so that nonhomologous recombination could be used to insert transgenes into the germline without loss in fidelity of transcriptional control. Furthermore, rather than preparing new libraries and novel vectors, existing, well-characterized BACs can be retrofitted with appropriate features, including specific reporter constructs, promoters for ectopic expression, deletions, mutations, and novel regulatory sites.

Indeed, the initial BAC vectors were invented[2] with an eye to subsequent modification and are equipped with integration sites for site-specific recombination mediated by the Cre recombinase. In fact, the approach can be further extended by using a variety of sophisticated bacterial genetic techniques to modify specific BACs without interfering with homologous recombination. Thus, transposons constructed to carry functionalizing sequences can be introduced either by site-specific recombination or at random into the cloned gene. These may be used in transgenic mice so that the timing and level of gene expression can be controlled or varied in vivo.

Finally, BAC technology itself is continuing to develop. Projects are currently under way to further annotate BAC libraries by identifying groups of BACs that correspond to all of the expressed sequence tags (ESTs) and thus to all of the expressed genes that have been identified. A variety of laboratories are engaged in BAC-end sequencing projects[3]. Thus, in the next two years, the ends of hundreds of thousands of BAC clones will be sequenced. In addition, similar projects using murine libraries will allow cross-referencing so that BACs corresponding to specific human genes can be compared with corresponding syntenic regions of the mouse chromosome, and the effects of gene modification in the mouse can be compared with gene function in human populations. These annotated BACs will provide resources that will make every gene or controlling element available as a reliable set of DNA clones.

The challenge of functional genomics will keep biologists employed for the next generation. The ability to easily retrofit BAC clones and reintroduce them into cells and animals is just a small first step in this enormous challenge. The work by Yang et al. will surely be followed by a variety of other clever uses of microbial systems to increase the utility and efficiency of the processes needed to define gene function. It is clear that having a trustworthy and versatile cloning system that can be manipulated in a simple and well defined host will be of great value.

Melvin I. Simon is Anne P. and Benjamin F. Biaggini professor of biological sciences, division of biology, California Institute of Technology, 1201 E. California Blvd., Pasadena, CA 91125 (simonm@starbase1.caltech.edu).

1. Yang, X.W., Model, P., and Heintz, N. 1997. *Nature Biotechnology* **15**:859–865.
2. Kim, U.-J., et al. 1996. *Genomics* **34**:213–218.
3. Venter, J.C., et al. 1996. *Nature* **381**:364–366.