- Generation of finished sequence for a 240,000-base telomeric region of human chromosome 7q. From initial sequences generated by SASE, oligonucleotides were synthesized and used for primer walking directly from cosmids comprising the contig map. Complete sequencing was performed to determine what genes, if any, are near the 7q terminus. This intriguing region lacks significant blocks of subtelomeric repeat DNA typically present near eukaryotic telomeres.

- Complete single-pass sequencing of 2018 exon clones generated from LANL's flow-sorted human chromosome 16 cosmid library. About 950 discrete sequences were identified by sequence analysis. Nearly 800 appear to represent expressed sequences from chromosome 16.

- Development of Sequence Viewer to display ABI sequences with trace data on any computer having an Internet connection and a Netscape World Wide Web browser.

- Sequencing and analysis of a novel pericentromeric duplication of a gene-rich cluster between 16p11.1 and Xq28 (in collaboration with Baylor College of Medicine).

Technology Development

Technology development encompasses a variety of activities, both short and long term, including novel vectors for library construction and physical mapping; automation and robotics tools for physical mapping and sequencing; novel approaches to DNA sequencing involving single-molecule detection; and novel approaches to informatics tools for gene identification.

Accomplishments

- Development of SCAN program for large-scale sequence analysis and annotation, including a translator converting SCAN data to GIO format for submission to Genome Sequence DataBase.

- Application of flow-cytometric approach to DNA sizing of P1 artificial chromosome (PAC) clones. Less than one picogram of linear or supercoiled DNA is analyzed in under 3 minutes. Sizing range has been extended down to 287 base pairs. Efforts continue to extend the upper limit beyond 167,000 bases.

- Characterization of the detection of single, fluorescently tagged nucleotides cleaved from multiple DNA fragments suspended in the flow stream of a flow cytometer (see picture, p. 70). The cleavage rate for Exo III at 37°C was measured to be about 5 base pairs per second per M13 DNA fragment. To achieve a single-color sequencing demonstration, either the background burst rate (currently about 5 bursts per second) must be reduced or the exonuclease cleavage rate must be increased significantly. Techniques to achieve both are being explored.

- Construction of a simple and compact apparatus, based on a diode-pumped Nd:YAG laser, for routine DNA fragment sizing.

- Development of a new approach to detect coding sequences in DNA. This complete spectral analysis of coding and noncoding sequences is as sensitive in its first implementations as the best existing techniques.

- Use of phylogenetic relationships to generate new profiles of amino acid usage in conserved domains. The profiles are particularly useful for classification of distantly related sequences.