Libraries from Flow-sorted Chromosomes

Larry L. Deaven

For any kind of genomic-DNA library, subdividing the DNA of the entire genome before library construction is almost always advantageous. The resulting set of libraries includes all of the genomic DNA, but each library is less complex than a single library containing all of the cellular DNA. A natural way to make subsets of human DNA is to make a separate library for each chromosome. To include all of the nuclear DNA in human cells, 24 different libraries are necessary (22 autosomes plus the X and Y chromosomes). The libraries vary in size, the largest (for chromosome 1) being five times as large as the smallest (for chromosome 21).

The most efficient way to make chromosome-specific libraries is to start with flow-sorted chromosomes. Los Alamos scientists pioneered the technology of flow sorting chromosomes as a direct result of the invention and development of flow cytometers at the Laboratory during the 1970s. Figure 1 diagrams flow sorting as we use it in making DNA libraries.

The first libraries made from sorted chromosomes at Los Alamos were from Chinese hamster chromosomes. Those chromosomes are larger and better differentiated from one another by base-pair content than are human chromosomes, properties which make them relatively easy to sort on a flow cytometer. On the basis of that success, we thought it would be feasible to construct certain types of libraries from sorted human chromosomes. The Department of Energy agreed to support the work, and because the scope of the envisioned project was large, we asked our colleagues at Lawrence Livermore National Laboratory if they would join in an effort to make a complete set of chromosome-specific libraries. Our initial discussions in 1983 led to the National Laboratory Gene Library Project, which continues to be a component of the Human Genome Centers at the two laboratories.

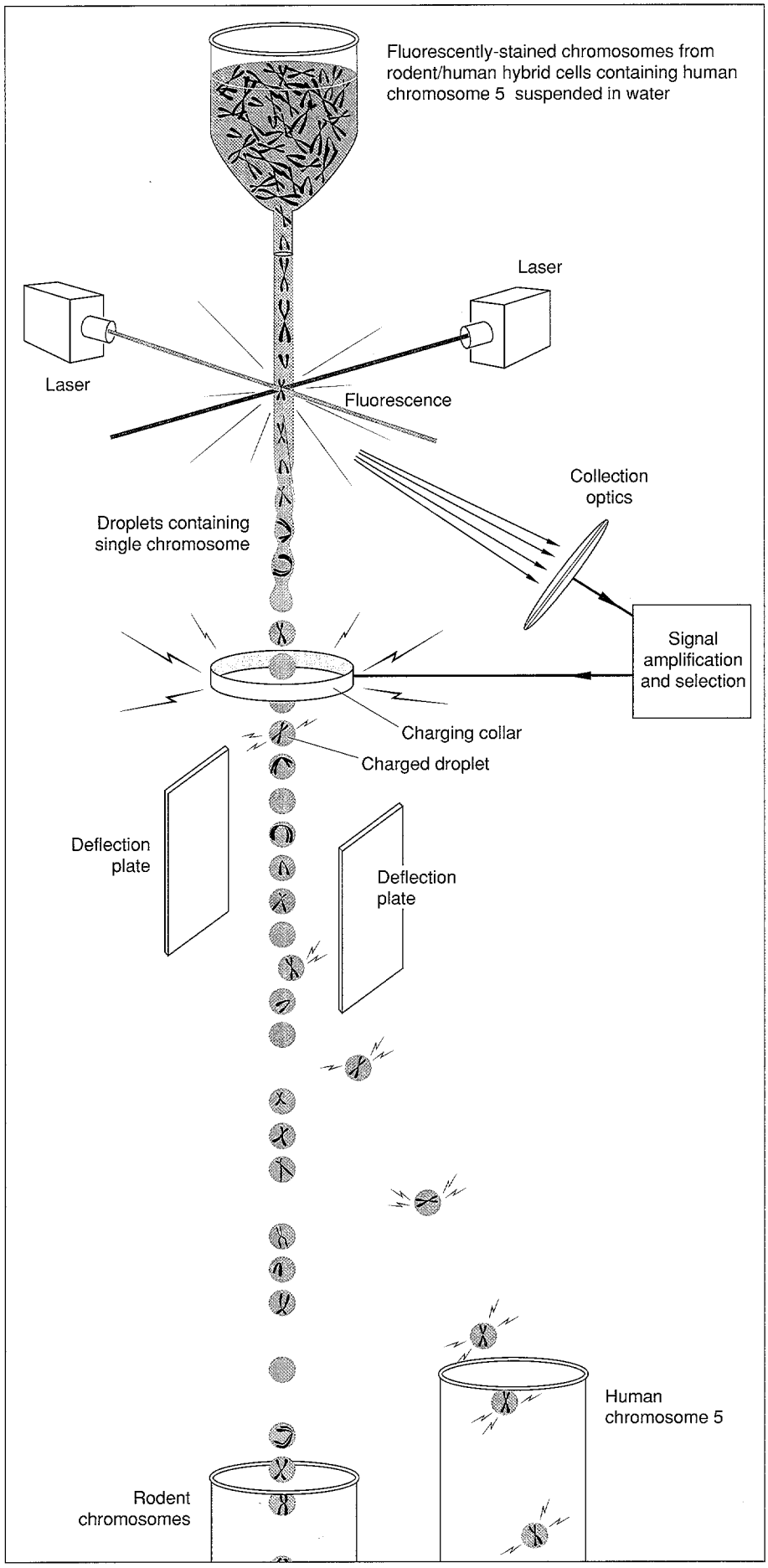
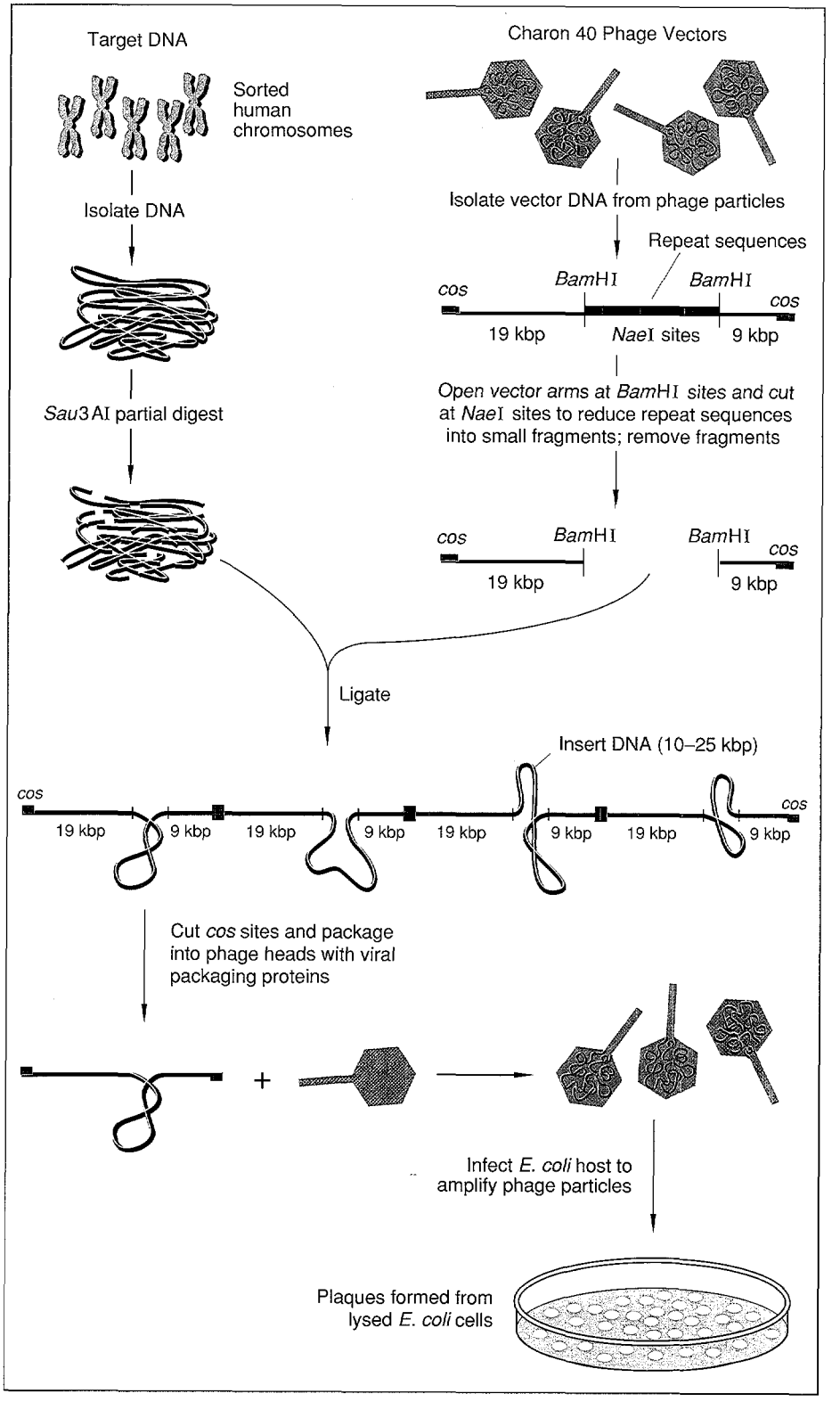


Figure 1. Purifying Chromosomes through Flow Sorting

Flow sorting provides a way of separating chromosomes of one type from a mixture. The example in the illustration is the separation of human chromosome 5 from rodent chromosomes all isolated from a rodent/human hybrid cell line. A liquid suspension of metaphase chromosomes is carried through the flow sorter in a narrow stream. The chromosomes have been stained with two fluorescent dyes: Hoechst 33258, which binds preferentially to AT-rich DNA, and chromomycin A_3, which binds preferentially to GC-rich DNA. The stained chromosomes pass through a point on which two laser beams are focused, one beam to excite the fluorescence of each dye. Each chromosome type has characteristic numbers of AT and GC base pairs, so chromosomes can be identified by the intensities of the fluorescence emissions from the two dyes. If the fluorescence intensities indicate that the chromosome illuminated by the lasers is the one desired, the charging collar puts an electric charge on the stream shortly before it breaks into droplets. When droplets containing the desired chromosome pass between charged deflection plates, they are deflected into a collection vessel. Uncharged droplets lacking the desired chromosome go into a waste-collection vessel. The flow instruments used at Los Alamos can analyze 1000 to 2000 chromosomes per second and sort approximately 50 chromosomes per second.

Figure 2. Phage Cloning Using Sorted Human Chromosomes as Target DNA

The phage vector (Charon 40) used to construct libraries from flow-sorted human chromosomes at Los Alamos contains a *cos* site, a large number of restriction sites, and a removable section consisting of repeat sequences (see Figure 6 in the main text). When the vector is used for cloning, the section of repeat DNA is cut into small pieces and discarded. The removal provides space for insert DNA. The vector consists of a 19-kbp arm and a 9-kbp arm, leaving room for inserts of 10 to 25 kpb. After the vector DNA has been isolated from phage particles, it is digested with the restriction enzymes *Bam*HI and *Nae*I. The eightyfold-repeated sequence constituting the central portion of Charon 40 contains an *Nae*I site, so the central portion is cut into small pieces by the *Nae*I digestion. The *Bam*HI digestion provides cloning sites on one end of each vector arm. Because *Bam*HI and *Sau*3AI produce identical sticky ends, the cloning sites are compatible with the *Sau*3AI sites on the ends of each fragment of partially digested target DNA. When the vector arms are ligated with fragments of target DNA, a concatamer forms that is cut at the *cos* sites to form individual recombinant phage chromosomes. These chromosomes are packaged into phage particles which then infect *E. coli* cells.



In 1983 the Human Genome Project did not exist. It was too early to seriously consider the construction of a physical map and the sequencing of the entire genome. Genetic mapping, on the other hand, was enjoying a period of unprecedented growth. The theory and methodology of finding genes using DNA markers had been developed, and major efforts were under way to locate human disease genes and to develop high-resolution genetic maps (see "Modern Linkage Mapping" in "Mapping our Genes"). Accordingly our first aim for the library project was to construct a phage library of small DNA inserts for each human chromosome. Small inserts were desirable for two reasons. First, the major challenge in making libraries from sorted chromosomes is to maximize the efficiency of each step in the cloning procedure in order to be able to make large libraries from small amounts of sorted DNA. In 1983, the technology for making small-insert (complete-digest) libraries was more reliable and could start with smaller amounts of target DNA than that for large-insert (partial-digest) libraries, which require cosmids. The second reason was the utility of small-insert libraries to genetic mappers. Repetitive DNA sequences are dispersed throughout the human genome, and the larger the insert, the more likely it is to contain at least one sequence repeated elsewhere. Probes containing repetitive sequences hybridize to many sites in the genome unless the repeat sequence is blocked. Single-copy probes identify only one site, a useful step in genetic mapping.

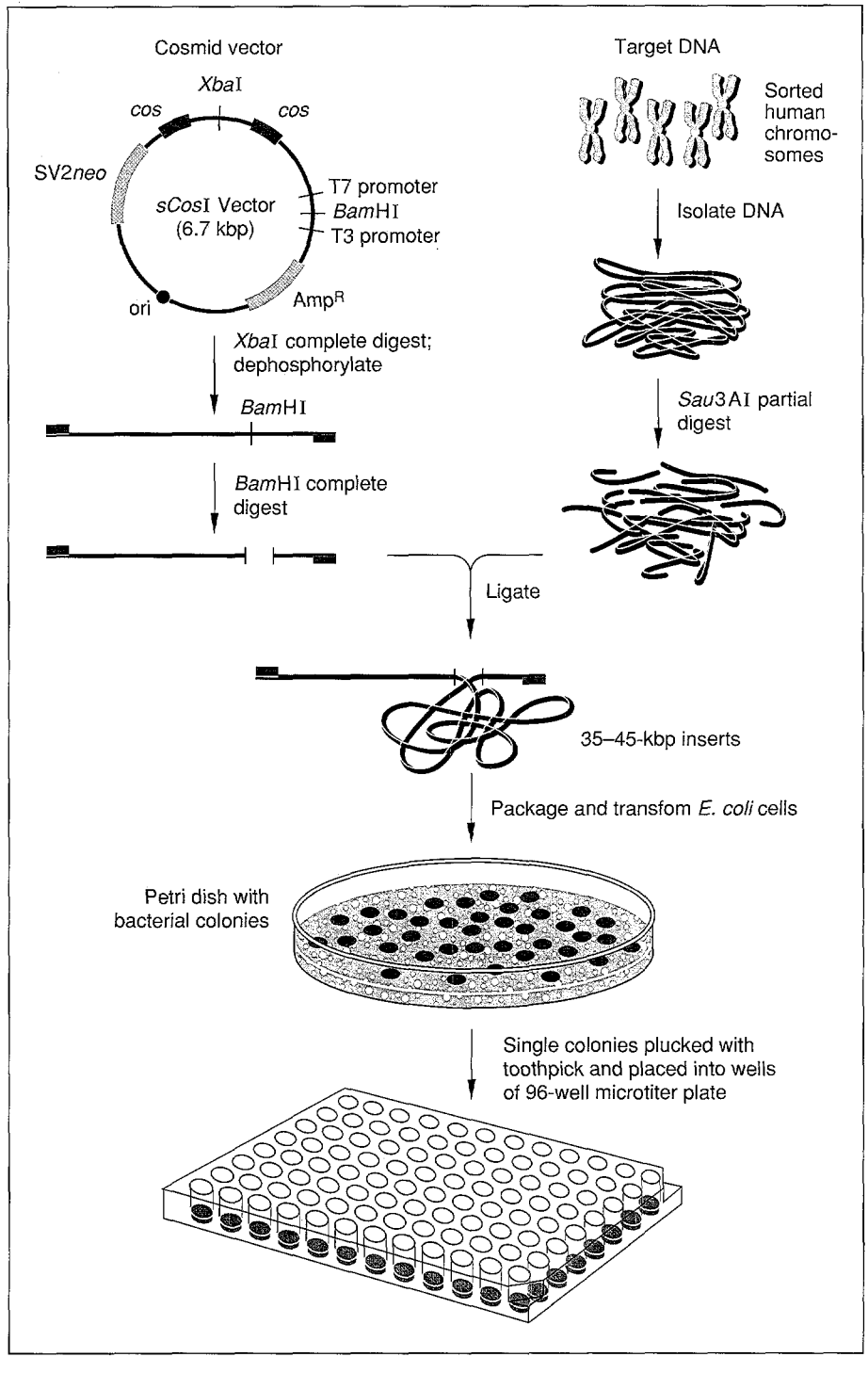
Our strategy for the first set of libraries made from sorted chromosomes was to digest the chromosomal DNA completely with a six-base cutter and to clone the fragments into a λ-phage vector called Charon 21A. Such a restriction enzyme reduces DNA to fragments having an average length of 4 kbp. However, approximately a third of the DNA is in fragments larger than 9 kbp, the upper limit for acceptance by Charon 21A. To reduce the amount of uncloned DNA, we constructed for each chromosome two libraries using different restriction enzymes; the Los Alamos project used *Eco*RI, while the Livermore project used *Hin*dIII. We estimate that at least 90 percent of the chromosomal DNA is contained in the two libraries together.

Our small-insert libraries were amplified one time, then sent to the American Type Culture Collection in Rockville, Maryland, where they are stored in liquid nitrogen. Samples from the original libraries are available to research groups throughout the world. They have been used extensively as a source of probes for polymorphic markers used in mapping genes, especially genes that can cause diseases. For example, as part of the searches for the defects responsible for cystic fibrosis and Huntington's disease, several hundred probes have been isolated from the chromosome-4 and chromosome-7 libraries and mapped to those chromosomes. Although improved methods now permit the construction of larger-insert libraries, the Los Alamos and Livermore complete-digest libraries are still useful. Over 4000 samples have been sent to research laboratories.

As we were finishing construction of the complete-digest libraries, it became obvious that chromosome-specific libraries with larger inserts were highly desirable. For molecular studies of gene structure and expression, they would have the advantage of

Figure 3. Cosmid Cloning of DNA from Sorted Human Chromosomes

The cosmid vector (sCos 1) contains two *cos* sites for rejoining the linear recombinant molecule after transformation. It also contains two selectable markers [resistance to ampicillin (amp^R) and to neomycin (SV2*neo*)], a number of restriction sites, a plasmid replicon including an origin of replication (ori), and promoter sequences from the T3 and T7 phage. The T3 and T7 promoters are used to generate end probes, as discussed in the section on YACs in the main article. The vector molecule is linearized by cutting with the restriction enzyme *Xba*I, then separated into two cloning arms by cutting with *Bam*HI. After fragments between 35 and 45 kbp in length are ligated to the vector arms, the recombinant DNA molecules thus produced are packaged into phage protein coats. The resulting infectious phage particles insert the recombinant molecules into *E. coli* cells where the molecules cyclize and live as plasmids. To prevent the faster-growing *E. coli* cells from overwhelming the slower ones, each colony is placed in a separate well of a microtiter plate.



containing whole genes or even groups of genes in a single cloned insert. Moreover, molecular biologists were then discussing and planning the mapping and sequencing of the entire human genome. Large-insert libraries from each chromosome would be a valuable resource for those massive tasks. The entire human genome in a cosmid library can be thought of as a jigsaw puzzle of 75,000 pieces; the chromosome-specific libraries would be 24 puzzles with an average of 3125 pieces in each.

During the years we spent constructing small-insert libraries, significant improvements were made in the efficiency of vector systems capable of carrying large inserts. The most important improvement for our large-insert project was the construction of cosmid vectors with two *cos* sites instead of one. Such cosmid vectors can be cleaved into two cloning arms, each with a *cos* site at one end. The cosmid arms can then be ligated to the partially digested human target-DNA fragments, much as in phage cloning. Each resulting recombinant molecule consists of two cloning arms each ligated to an end of a fragment of human DNA. If the *cos* sites are between 30 kbp and 52 kbp apart, the recombinant molecule can be packaged in vitro to produce infectious phage particles. Using this cloning system, a cosmid library with inserts 35 to 45 kbp in length can be made from less than a microgram of DNA.

The laboratories' joint strategy for the construction of a second set of libraries with larger inserts was to divide the human chromosomes between Los Alamos and Livermore. Each laboratory would construct a partial-digest phage and cosmid library for the chromosomes assigned to it. Los Alamos has made libraries for chromosomes 4, 5, 6, 8, 11, 13, 16, and 17; Livermore, for chromosomes 19, 22, and Y.

Our current work incorporates several changes in the construction and handling of libraries. All chromosomes are sorted from hybrid-cell lines rather than from human cells because of the advantages discussed in the main text. The phage libraries, illustrated in Figure 2, have inserts 10 to 25 kbp long. They are stored as pools of clones in a liquid medium and distributed as samples like the small-insert libraries.

As illustrated in Figure 3, the cosmid libraries are seeded on Petri plates. The libraries are then arrayed, that is, single colonies are transferred to 96-well microtiter plates. Enough colonies are isolated to cover the chromosome five times. A chromosome of average size requires about 5×3125 or 15,625 colonies. The inserts have not yet been characterized, so we do not know whether the DNA in the inserts covers the entire chromosome. When all the colonies have been transferred, we make five to ten copies of each microtiter plate. Sets of microtiter plates are sent to laboratories involved in projects to map the entire chromosome or a major portion of it. In addition, the colonies in one set of plates are allowed to grow to high density, and then the bacteria are removed from each well and pooled. Laboratories interested in isolating one or a few genes on the chromosome can obtain portions of the pooled library.

An advantage of storing a library in a set of microtiter plates is that each clone has an alphanumerically labeled location. The labeling permits all the data on the

clones from different laboratories to be combined for analysis. Ideally, all interested laboratories should have copies of the plates; however, distribution of so many copies would be too expensive.

The partial-digest libraries that have been completed are major resources for laboratories constructing physical maps for chromosomes 4, 5, 8, 11, 16, 17, and 19. The libraries are used directly in assembling contigs of cosmids and also contribute to physical mapping with YACs. In order to make a high-resolution map from YAC contigs, each YAC must be subcloned into cosmid or phage vectors, a time-consuming process. A more rapid way to find cosmids that are part of a YAC is to screen an arrayed cosmid library with DNA from the YAC insert. A second major use of the partial-digest libraries is in the isolation of genes for detailed studies of normal and abnormal structure and expression. A third use is the identification of specific chromosomes or parts of chromosomes. Each library is very pure, and the inserts in it can be labeled with fluorescent stains and hybridized in situ to cells or metaphase chromosomes. In interphase cells hybridization reveals the nuclear location of the chromosome represented in the library. In metaphase chromosomes hybridization identifies only the pair of chromosomes that the library represents. If a piece of a labeled chromosome has been broken and has translocated to another chromosome, the translocation is easily visible. The latter application is revolutionizing the detection of chromosomal rearrangements induced by substances that break chromosomes and by diseases like cancer, in which rearranged chromosomes are common.

Although our cosmid libraries are not yet complete, during the past two years we have devoted a substantial portion of our library-construction effort to YAC cloning. We were fortunate in having Mary Kay McCormick join our Center in 1989. Before coming to Los Alamos, she had demonstrated the feasibility of using sorted chromosomes as the source of target DNA in making YACs. To construct a YAC library, we had to overcome two major obstacles. Long pieces of human DNA had to be obtained from sorted chromosomes, and YAC-cloning techniques had to be optimized in order to use the small amounts of DNA available after sorting. Solutions to both problems were found through the skills of dedicated investigators. Chromosome isolation and flow sorting must be accomplished without delay because DNA degradation begins as soon as the chromosomes are extracted from the cells. To sort 1-microgram samples of DNA in a limited time, sorting continues around the clock. The sorted chromosomes are collected in agarose plugs which hold the DNA in the stable agarose matrix and protect it from shear stresses during isolation from the chromosome and digestion with restriction enzymes. The agarose is then melted so that the vector arms and DNA ligase can be mixed in. After ligation the recombinant molecules are fractionated by preparative pulsed-field gel electrophoresis, which concentrates all the DNA fragments longer than 200 kpb into a single band in the gel.

To facilitate transformation, the walls of yeast cells are removed. (Yeast cells without walls are called spheroplasts.) The long recombinant DNA molecules are added to

the spheroplasts in the presence of the polyamines spermine and spermidine, which are believed to bind to and condense DNA. To obtain large numbers of recombinant yeast colonies (as many as 2400 have been obtained from 1 microgram of target DNA), all of the above steps must work well. Probably the most frustrating step is transforming the yeast cells. It is difficult to control, it sometimes fails, and because it is the last cloning step, failure means that all the previous work must be repeated.

We have completed two YAC libraries, one for chromosome 16 and one for chromosome 21. Both libraries were made from target DNA completely digested with restriction enzymes that have infrequent cleavage sites. Therefore, how completely the libraries represent chromosomal DNA depends on how uniformly the cleavage sites are distributed along the chromosomes. We will not know the completeness of the representation until we have generated a considerable amount of data on each library. Preliminary results suggest that the YACs made from digests with *Eag*I or with a combination of *Not*I and *Nhe* are clustered near certain chromosomal regions such as the centromere, but that YACs made from *Cla*I digests may be more uniformly distributed. We are attempting to ensure that future YAC libraries have unbiased distributions by making them from partial digests. Other studies of the libraries suggest that the frequency of chimeric inserts is quite low. Fifty-three YAC inserts have been hybridized in situ to chromosome 21. None of them hybridized to more than one region of the chromosome, which would have been evidence of a chimera.

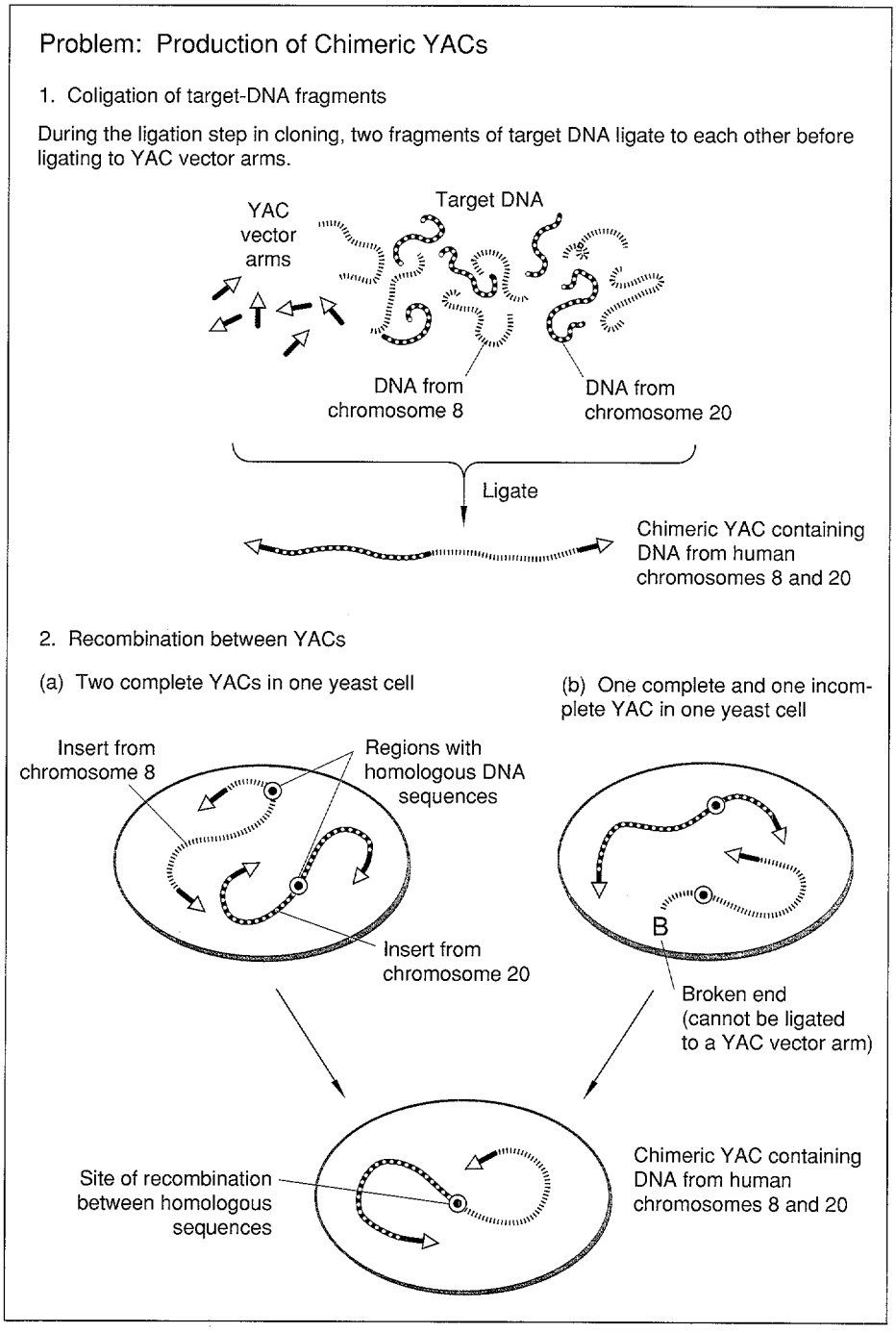
The reasons for the absence of chimeric inserts are not completely clear. We took a number of steps intended to reduce their frequency. As illustrated in Figure 4, chimeric YACs are believed to originate either from ligation of two pieces of target DNA or from recombination between two YACs after they have both transformed the same yeast cell, especially when at least one YAC is incomplete.

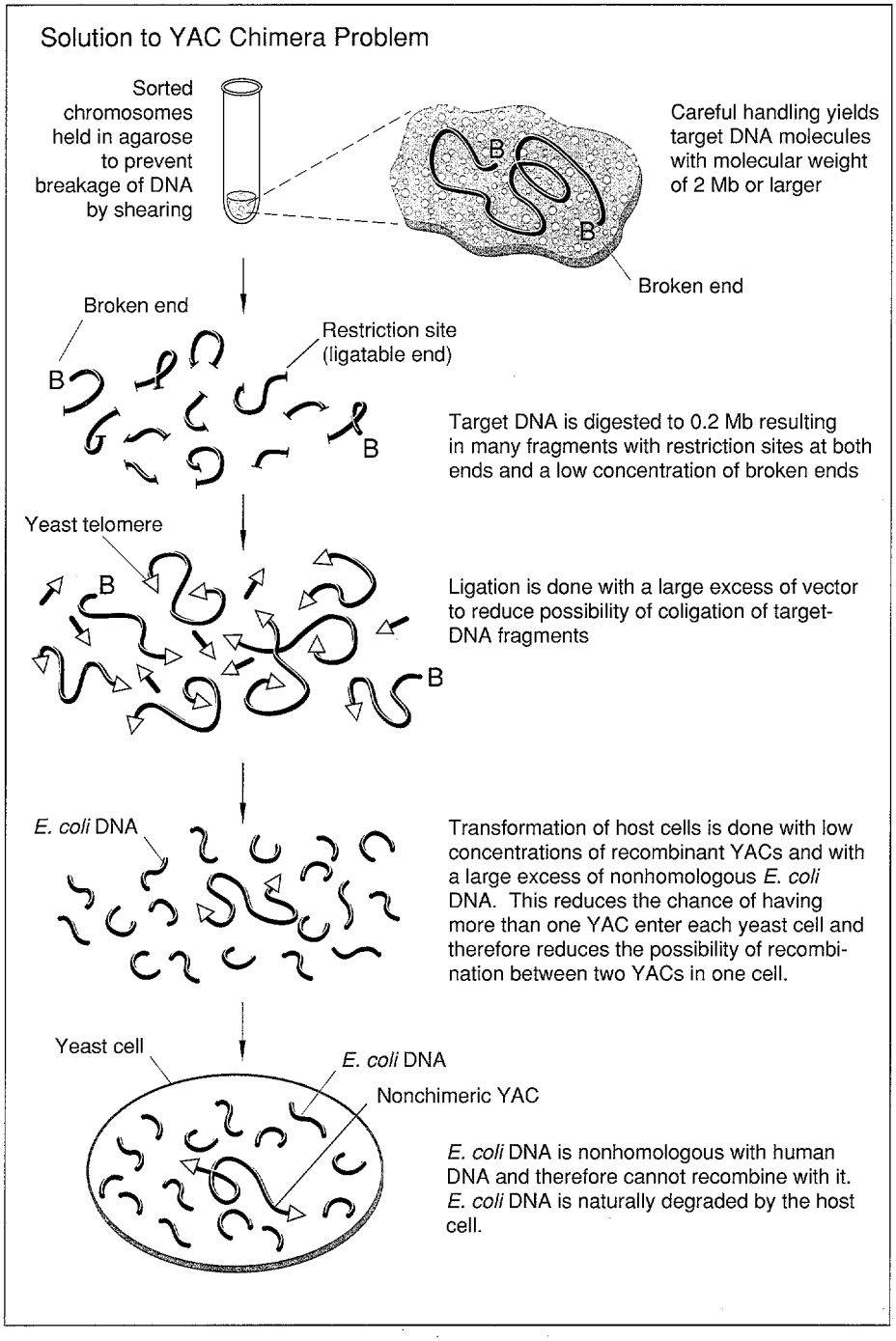
To minimize the coligation of target DNA, we added much more vector DNA to the ligation mixture than the restriction fragments could react with. To reduce the possibility of recombination inside yeast cells, we took two precautions. The first was to handle target-DNA restriction fragments so as to minimize breakage. The second was to attempt to limit the possibility that more than one YAC would enter a single spheroplast by diluting the YACs to the point where it was unlikely that two YACs would enter the same spheroplast.

Although sufficient data are not yet available to thoroughly evaluate the chromosome-specific YAC libraries, all evidence to date suggests that they will be a valuable resource for constructing physical maps of chromosomes. The libraries combine the advantages of large insert size and division into subsets of the genome to provide the least complex mapping elements available. They are being used to close the gaps between cosmid contigs in the Los Alamos chromosome-16 map, and they should prove to be excellent sources of fragments for the initiation of maps of other chromosomes. We expect the Library Project now to focus on the construction of large-insert libraries in YACs and other cloning systems under development.

Figure 4. Chimeric YACs

Part A shows two causes of chimeric YACs. The first is that, since target DNAs are all cut with the same restriction enzyme, they can ligate to each other. The resulting chimeric insert can then ligate to vector arms. The second is that if two YACs enter the same yeast cell and their inserts have homologous sequences, they can recombine with each other, producing a chimeric YAC. Recombination is especially likely if one or both of the YACs is incomplete, either because the insert is broken or because it ligated to only one vector arm. Part B shows our solutions to the problem. To limit breaking we keep sorted chromosomes in agarose and handle the DNA carefully. Our target DNA molecules are typically longer than 2000 kbp. Since the restriction digest produces fragments averaging 200 kpb, few fragments have broken ends. Then we add many more vector molecules than insert molecules to the ligation reaction, making ligation between two insert molecules unlikely. During transformation, to reduce the probability that two YACs enter a yeast cell, we add *E. coli* DNA, which is not homologous with human DNA. That step greatly dilutes the YACs while keeping the total DNA concentration high enough to induce transformation.




Solution to YAC Chimera Problem
Sorted chromosomes held in agarose to prevent breakage of DNA by shearing
B
B
Careful handling yields target DNA molecules with molecular weight of 2 Mb or larger
Broken end
Broken end
Restriction site (ligatable end)
B
B
Target DNA is digested to 0.2 Mb resulting in many fragments with restriction sites at both ends and a low concentration of broken ends
Yeast telomere
B
B
Ligation is done with a large excess of vector to reduce possibility of coligation of target-DNA fragments
E. coli DNA
Transformation of host cells is done with low concentrations of recombinant YACs and with a large excess of nonhomologous E. coli DNA. This reduces the chance of having more than one YAC enter each yeast cell and therefore reduces the possibility of recombination between two YACs in one cell.
Yeast cell
E. coli DNA
Nonchimeric YAC
E. coli DNA is nonhomologous with human DNA and therefore cannot recombine with it. E. coli DNA is naturally degraded by the host cell.

Library Distribution

The success of the people working in the Library Project has created a need for large-scale duplication of clones in microtiter plates. The Los Alamos portion of the cosmid-library project will require copying over 200,000 clones six to ten times, and our future work in YAC-library construction will produce more clones to be copied. As important as duplicating clones in microtiter plates is making replicas of microtiter plates as spots on nylon membranes, a procedure that provides a convenient way to screen an entire library. A 96-prong stamp is inserted into the wells of a microtiter plate and then gently placed on a membrane. The bacteria collected on each prong are transferred to the membrane. The membrane rests on an agar culture medium from which the bacteria absorb nutrients. The resulting 96 colonies in the form of spots on the membrane can then be screened with a DNA probe. Any spots that hybridize with the probe DNA can be identified and the corresponding clones can be located in the microtiter plate. Those clones can then be selected and regrown for further analysis. We use this screening procedure extensively in our construction of a map of chromosome 16, and we currently use it to share our libraries with other laboratories. For example, an investigator at the Institute of Cytology of the Russian Academy of Sciences is interested in finding inserts that come from a region of chromosome 5. We sent her a set of membranes containing spots from each microtiter well in the arrayed chromosome-5 library. She probed the membranes with her collection of probes from the region she was interested in, and we selected and shipped colonies corresponding to each of the spots that tested positive. Duplicating and shipping copies of the library in microtiter plates is expensive, and we hope that the use of membranes will prove to be a useful alternative.

To help us meet the demands of library duplication, a group of robotics engineers at Los Alamos has constructed a robot capable of accomplishing that task. The robot can choose a microtiter plate from a dispenser, scan the barcode label on the plate, and insert a 96-prong tool into the wells in the plate. The robot then presses the tool against a membrane, transferring spots of bacteria from the prongs to the membrane. Finally it sterilizes the tool, replaces the lid on the microtiter plate, and returns the plate to a stacker. The robot can transfer colonies to the same membrane up to 96 times, each time shifting the position of the tool slightly, and thus can vary spot densities from 576 to 9216 per 22-cm^2 membrane. The robot's versatility is valuable; because denser packing of spots is more efficient but may be harder to read, different densities are suitable for different applications. ■

Further Reading

L. Scott Cram, Dale M. Holm, and Paul F. Mullaney. 1980. Flow cytometry: A new tool for quantitative cell biology. *Los Alamos Science*, volume 1, number 1.

L. Scott Cram, Larry L. Deaven, Carl E. Hildebrand, Robert K. Moyzis, and Marvin Van Dilla. 1985. Genes by mail. *Los Alamos Science*, number 12.