Making the Libraries

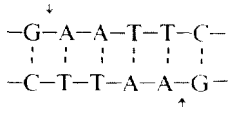
Cloning the DNA

The cells have been cultured, the chromosomes isolated and sorted. Now can begin the final act in the making of a library—that of multiplying the DNA in the sorted chromosomes by a factor of more than a million. Some tricks of recombinant DNA technology make such great amplification possible, and the magicians who perform these tricks are lowly but accommodating bacteria and viruses.

Briefly, four steps are involved in cloning the DNA: preparation of DNA fragments from the sorted chromosomes and from bacterial viruses (bacteriophages, or simply phages); formation of recombinant DNA molecules including both chromosomal and viral DNA fragments: assembly of phages containing the recombinant DNA molecules: and cloning of the phages (and hence replication of the chromosomal DNA fragments) in a bacterial host. For concreteness we describe, and illustrate in the accompanying figure, the making of what we call a Phase I library, one based on relatively small chromosomal DNA fragments. (Our Phase 11 libraries, which will serve a different purpose, will be based on larger fragments.)

First the DNA molecules in the sorted chromosomes must be extracted from the other chromosomal components. For reasons that will become clear later, this extraction must be performed very gently to minimize breaking the very long and very fragile DNA molecules.

The extracted DNA is then digested with the restriction enzyme *EcoRI*, which recognizes the sequence of six bases –G–A–A–T–T–C– and, as shown below, cuts the DNA at each occurrence of' this sequence by breaking one covalent bond on each strand:

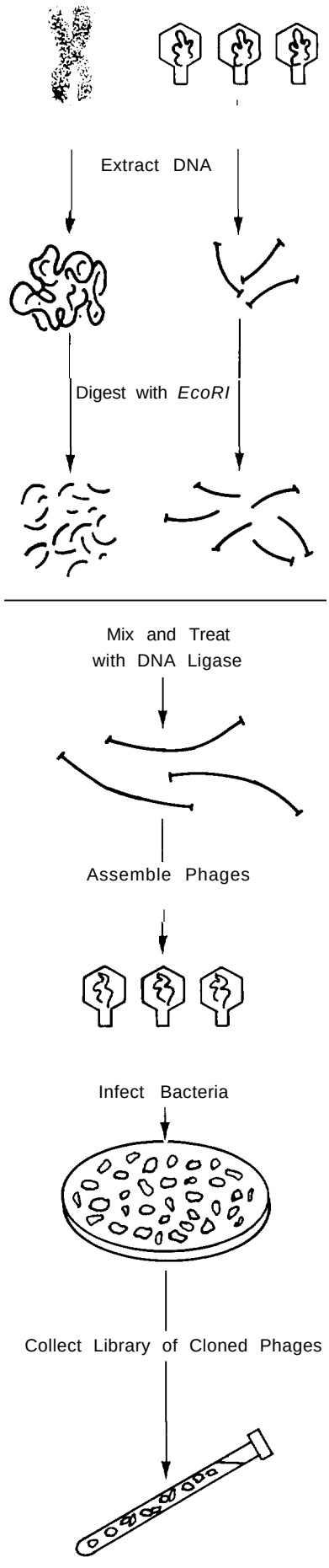


EcoRI is produced by certain strains of the bacterium *Escherichia coli* as a defense against infection by the DNA of bacteriophages. The enzyme severs the viral DNA and thus prevents its propagation in the bacterium. (To avoid self-destruction, the bacterium also produces another enzyme that modifies its own DNA so that it cannot be severed by *EcoRI.)* other bacterial species produce restriction enzymes that recognize other sequences. For example, *HindIII,* the restriction enzyme used by our collaborators at Livermore for their phase I libraries, is produced by the bacterium *Hemophilus influenzae.* Like *EcoRI, HindIII* recognizes a six-base sequence (–A–A–G-T–T--T–) and makes staggered cuts.

Complete digestion of the chromosomal DNA with *EcoRI* (or *HindIII)* produces a set of relatively small flagments. We can estimate the average length of these fragments by assuming a random distribution along the chromosomal DNA of equal numbers of the four bases A, T, G, and C. Then the probability of any six successive bases being the *EcoRI* (or *HindIII)* recognition sequence is $(1/4)^6$, which implies an average of about 4100 base pairs between cleavage sites, (An intact DNA molecule in the largest human chromosome (number 1) is about 2.5×10^8 base pairs in length.)

We, and our collaborators at Livermore, use DNA from the bacteriopbage Charon 2 I A as a cloning vector. that is. as a medium capable of accommodating foreign DNA segments for replication. Charon 21A is a modified version of the λ bacteriophage; its linear DNA molecule is about 42,000 base pairs long, terminates with short stretches of single-stranded DNA, and includes a single *EcoRI* cleavage site.) The DNA molecules extracted from the phages are cut by *EcoRI* into two "arms." The cut ends of these arms are then stripped of a phosphate group (for a reason to be explained later) by treatment with the enzyme alkaline phosphatase.

The next step is fairly simple, at least in theory. We mix the two DNA preparations and let nature take its course—nature in this case meaning the tendency for single strands of DNA to "stick together" (become linked by hydrogen bonds) at regions with complementary bases, Since the cutting action of *EcoRI* has equipped both the chromosomal DNA fragments and the vector arms with just such sticky ends. aggregates consisting of' two arms linked by a (single) chromosomal DNA fragment will form. (Herein lies the reason for gentle extraction of the DNA from the chromosomes: physical breaks do not result in sticky ends.) These aggregates arc then converted into intact DNA molecules by treatment with the enzyme DNA ligase, which establishes a covalent bond at the break along each strand. The result is a set of recombinant DNA molecules—that is, molecules of phage DNA carrying human DNA fragments. (To increase the probability of forming recombinant aggregates. the chromosomal fragments are mixed with an excess of arms. Therefore, many of the aggregates formed will consist simply of two arms. But DNA ligase cannot repair the



III. CLONING THE: DNA

DNA in such aggregates because of the missing phosphate groups, So our foresight in the previous step reduces the background of nonrecombinant DNA molecules in the final library.)

Next we assemble Charon 21A phages, drawing on proteins produced by mutant varieties of the phage for the necessary structural components and on our supply of recombinant DNA molecules for the genetic component. This step unavoidably introduces a deficiency in the resulting library: those recombinant DNA molecules containing fragments longer than about 9100 base pairs are too large to be encased in the heads of the phages, and therefore the library will lack that portion (about 33 mass percent) of the chromosomal DNA residing on such fragments. It is for this reason that we and our collaborators at Livermore use different restriction enzymes and thus produce libraries based on different sets of chromosomal DNA fragments. The expectation is that the DNA missing from one laboratory's version of a library will be found in the other,

The final step in making the libraries, that of cloning the recombinant DNA molecules, is accomplished by seeding the phages onto a "lawn," or monolayer, of an appropriate strain of *E. coli,* When a phage injects its recombinant DNA molecule into such a cell, the biological apparatus of the cell is subverted to production of new phages. The cell ruptures when between one and two hundred new phages have been produced, and these then infect surrounding cells. The process continues. resulting in formation of a plaque, or clear area, in the translucent lawn of bacteria. In each plaque arc some one to ten million infective phages, each containing an exact copy of the recombinant DNA molecule carried by the original infecting phage.

These vastly multiplied recombinant DNA molecules, then, constitute a library-conveniently and safely packaged within the proteinaceous armor of phages. This construction of a library puts us at the starting line. Its application to problems of clinical and fundamental significance lies ahead. ■

what genes or combinations of genes are involved; in fact, we haven't come close to settling even the nature-versus-nurture issue. And even when a single defective gene is known to be the culprit—as it is in certain kinds of mental retardation—developing the techniques for introducing the normal gene and getting it expressed properly will take a long time.

HILDEBRAYD: Yes, routine clinical application of gene therapy is a long way off, especially in light of the ethical and legal questions it raises, But in the process of ironing out those questions, society should focus on the extraordinary benefits to be derived from the technology rather than its "Brave New World" implications.

MOYZIS: I think other developments in molecular genetics pose more immediate concerns. For instance. the availability of probes specific to the X and Y chromosomes has made it very easy to determine the sex of a fetus. Since even in "sexually liberatcd" western countries the preference for having a male child first is still strong. routine determination of fetal sex could conceivably lead to dramatic changes in the ratio of the sexes. The ethical. social, and economic implications of this possibility need immediate consideration.

SCIENCE: *Are issues such as these heavy on your mind as you go about your research?*

CRAM: I don't think so. The goals of basic research arc information and understanding. It's up to society to determine how those are applied. Take the photoelectric effect, for example. Its applications are mainly viewed as beneficial, although a few may be regarded by some as deplorable. But a scientist doesn't hide his head in the sand in the face of information that can possibly be applied in a harmful way.

MOYZIS: I would second that. You can't be paralyzed because a bit of knowledge can be used inappropriately. And 1 repeat that most of the worrisome aspects of gene therapy are a long way down the road. A major problem that must he solved before it can be applied to many diseases is how to introduce the good genes into only the appropriate cells of