b) We will use the chromosome 22 specific Fosmid and BAC libraries to generate contigs by conventional restriction fragment fingerprinting approaches. We will continue to develop new fingerprinting techniques by using the "built-in" lambda cosN site and the P1-loxp site on the vector and we will introduce unique sites designed for triple strand protection cleavage. Our prediction is that the stability of the Fosmid and BAC clones will allow us to prepare larger contigs with fewer gaps.

c) We will use a sensitive CCD camera and multi-colored fluorescent probes to rapidly map markers on metaphase, interphase, and meiotic chromosomes in order to correlate Fosmid and BAC contigs with the chromosome and to localize new probes relative to genetically mapped markers. These experiments will be done in collaboration with Dr. Julie Korenberg (Cedars-Sinai Hospital, Los Angeles, California). FISH will position contigs on the map at relatively low resolution and will be useful to confirm contig assignments.

d) We will continue to use pulsed field gel electrophoresis to confirm contig assignments and to place contigs in the context of larger well defined fragments of chromosome 22. We will use Fosmid and BAC clones that carry NotI sites to generate probes for linking NotI fragments on the chromosome. We will develop a NotI fragment map of chromosome 22. We will prepare and characterize fragments excised from chromosome 22 by a variety of other methods.

e) We will apply the triple strand protection - Achilles heel method of cutting DNA (in collaboration with Dr. Peter Dervan, Chemistry Division, Caltech) in order to introduce specific breaks in chromosome 22 and to accurately determine the physical distance between cutting sites on mapped Fosmid clones. This approach will also allow us to correlate the physical distance measured by gel electrophoresis with the map generated by cloning techniques and more importantly, to verify the relative distances within contigs and test the additivity of the map distances. It will provide probes for closing gaps from specific regions of chromosome 22.

2. Progress Report

Recent technical progress in molecular biology has made the mapping of entire mammalian chromosomes an attainable goal. However, a number of problems must still be overcome before genome mapping becomes rapid, efficient, and reliable. The limited size of cosmid inserts, as well as their tendency to rearrange, necessitates construction of very large libraries for mapping, due to the many gaps encountered in aligning cosmid contigs. Larger fragments can be cloned using the phage P1, but the maximum size of cloned inserts is fixed at only twice that of cosmids. The power of YACs has been demonstrated in isolating large regions of human DNA, recombining them to build up even larger regions and closing gaps in cosmid based maps (Burke et al., 1987). However, existing YAC libraries contain a high proportion of chimeric clones, and YACs are difficult to use for detailed mapping, often requiring recloning into cosmid sized pieces. Our work has addressed some of these issues by creating an alternative and complementary approach to cloning and mapping large DNA.

A. The BAC System

The BAC vector pBAC108L (Figure 1), consists of a 5.5 Kb segment of EcoRI fragment 5 of F-factor DNA, and the cloning module. The F-portion includes the selectable marker (chloramphenicol resistance), and the basic replication unit which includes oriS, parA, parB, and repE. These genes are essential for replication and copy control of the plasmid, and provide incompatibility (two F-factors cannot occupy one
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.
potential site in the cell). The cloning module includes 1) the bacteriophage lambda cosN site, 2) the bacteriophage P1 loxP site, 3) the cloning sites, and 4) sites for several "rare cutting" restriction enzymes NotI, EagI, XmaI, SmaI, BglII, and SfiI) for excising the inserts. The cosN and loxP sites are used to create fixed points in the circular molecule for mapping, by cleavage with lambda terminase or P1 Cre protein, respectively. In addition, the multiple cloning site is flanked by T7 and SP6 promoters, for generating probes when necessary.

![Physical map of pBAC108L](image)

Figure 1. Physical map of pBAC108L

It is essential to eliminate as much of the recombination function as possible from *E. coli* so that cloned DNA containing numerous repeated sequences can be stably maintained. However, the host strain must also be able to grow well enough to permit easy handling. We have constructed a set of new strains carrying deficiencies in 1) host controlled restriction, 2) methylation sensitive restriction, and 3) various recombination systems. Three strains have been used as hosts, HS986 (mcrA-B-C-mrr-hsd-recB-C-J-sbcB-C-endA-), HS979 (mcrA-B-C-mrr-hsd-recA-B-C-sbcB-C-endA-), and DH10B (Life Technologies, Bethesda, Maryland).

In order to construct BAC clones, DNA from a human cell line is prepared in agarose and subjected to partial HindIII digestion and size selection in pulsed field gels. DNA of the desired molecular weight (100-300 Kb) is excised from the gel and the gel is digested with agarase (Epicentre). This material (which runs with the expected size on re-electrophoresis) is ligated overnight with an equal amount of dephosphorylated pBAC vector. We found that by electroporation 200 Kb supercoiled F' molecules can transform *E. coli* with a frequency of 1 x 10^6 per µg, in our hands. This is about 35 times lower than the frequency with pBAC itself when molecular weights are considered. Colonies arising on selective plates are screened by hybridization with total human DNA to identify those with human inserts. NotI digestion of miniprep DNA excises the insert and is used for size analysis. Figure 2A shows clones from a recent transformation. Figure 2B shows that these inserts are of human origin by virtue of hybridization. Each lane contains the 6.5 Kb pBAC vector band as well as additional human DNA. The average size of these inserts is 110 Kb, with sizes ranging from 10 Kb to 215 Kb (n=21).
B. Fosmid Cloning

A valuable additional product of our work in developing the BAC system has been the generation of a new single copy vector based on the F factor replicon for efficient cloning and stable maintenance of 40 Kb average insert sized DNA. We call these clones Fosmids. As shown in Figure 3, in order to prepare the pFOS1 vector, a pUC plasmid derivative with the lambda cosN site was fused with the plasmid pBAC. Cutting the resulting plasmid at two unique sites generates two separate arms for ligation. In packaging, the fragment between the two cosN sites bearing the pUC origin is eliminated, rendering the ligation product single copy in E. coli. The two cosN sites permit equimolar recovery of arms which can be ligated to dephosphorylated mammalian DNA. In this way self ligation of the insert DNA is avoided and the material need not be size selected, permitting efficient cloning from small amounts of materials, such as sorted chromosomal DNA. Thus, Fosmid libraries can be prepared in much the same way as with conventional double cosN site vectors (Bates, 1987).

Figure 3. Construction of pFOS1
Our first Fosmid library employed partially digested, dephosphorylated DNA from a CHO-human hybrid cell line ligated to pFOS vector arms. This cell line, KG-1 (ATCC-GM10888), contains chromosome 22 as the sole human material. Packaged particles were used to transfect either DH10B or DH5a (mcr-)(Life Technologies, Bethesda, Maryland). Colonies containing human DNA were identified by colony hybridization using labelled total human DNA at moderate stringency. Colonies containing human DNA in single copy Fosmids could be detected and distinguished from the background of clones containing hamster DNA using ordinary colony hybridization procedures. Approximately 0.5% of the resulting colonies were clearly positive for human DNA, yielding thus far approximately 1,300 clones (about 1X coverage of chromosome 22). We can, of course, extend the size of this library. By continuing to pick clones we can obtain any desired level of coverage of chromosome 22. We are currently using FISH to verify that these clones are evenly distributed along the chromosome.

Recently, in collaboration with Pieter deJong (Lawrence Livermore National Laboratory, LLNL) we have constructed a larger chromosome 22 Fosmid library directly from flow sorted chromosome 22 prepared at LLNL from a CHO-human hybrid cell line containing human chromosomes 9, 22 and Y. Eighty-two percent of the clones in this library were positive for human sequences, meaning this library of approximately 12,000 clones represents about 7X coverage of chromosome 22. We are in the process of gridding and evaluating this library.

Unlike cosmids which are present in multiple copies in each cell, Fosmids are maintained at single copy level. However, this reduction in copy number does not present any barrier to the convenient use of Fosmids in cloning or analysis of mammalian DNA. As described above, standard colony hybridization techniques can be used with Fosmid libraries. More importantly, Fosmid clones can be rapidly prepared and analyzed using the same simple procedures that are used for cosmids. Figure 4 shows a comparison of the recovery of DNA for a pWE-15 cosmid (Wahl et al., 1987) with that of two Fosmids using either a boiling or alkaline lysis miniprep procedure for 1 ml of culture. DNA from an equivalent volume of the culture of each clone has been cut either with EcoRI or NotI. Sufficient DNA is obtained using either of these procedures to permit analysis of the clones. In the original photograph of this gel, bands as small as 0.5 Kb could be easily detected in the Fosmid lanes. There is also sufficient DNA for generating probes for FISH and for colony hybridization.

Figure 4. Recovery and digestion of miniprepped cosmid and Fosmid DNA. Clones containing Chromosome 22 DNA were prepped using either a boiling (left) or alkaline (right) protocol and digested with NotI (N) and EcoRI (E).
To obtain preliminary data regarding the KG-1 Fosmid library and verify that our techniques for preparation and analysis could be applied to random Fosmid clones on a routine basis, we prepared DNA from all clones in the first 96-well plate. These samples, representing approximately 3.7 Mb of chromosome 22, were screened for NotI and MluI restriction sites. Of the 95 clones grown in the dish, minipreps of 1 ml yielded DNA of sufficient amount and quality to be successfully scored for cleavage on the first try for all clones. Fifteen clones were identified as containing at least one NotI site and are used to develop probes for a NotI pulsed field map. Forty-five clones contained at least one MluI site.

C. Stability of Fosmid and BAC Clones

The frequency of rearrangements involving mammalian DNA cloned in cosmid vectors, although widely acknowledged and referred to in anecdotal fashion, has not been quantitatively reported in the literature. This in part is because 1) stability of cloned DNA reflects a complex interplay of both the vector and the host used. 2) an appropriate test should be able to detect rearrangements throughout the entire insert. 3) a high degree of tedious is involved in performing statistically significant sampling and analysis while the work itself is not central to the cloning projects. and 4) the suspected answer is not one that anyone generating libraries is eager to confirm. However, we feel that the stability of cloned human DNA during propagation is important to the successful use of libraries for mapping, so we have completed a rigorous test of the stability of human DNA cloned in our BAC and Fosmid vectors prior to our full reliance on the system. In addition, we have performed similar tests on cosmid vectors currently in use. Our initial findings indicate an exceptionally high degree of stability for our BAC and Fosmid clones, greatly exceeding that observed with conventional cosmids.

We have applied two different tests to a random sampling of Fosmids containing DNA from human chromosome 22 to ask how frequently rearrangements take place. In the first, we selected 20 clones from an original microtiter plate of the hybrid cell chromosome 22 Fosmid library and grew them under selection through serial passage for 160 generations. Restriction patterns of the DNA with three different 6 base recognition enzymes were compared before and after passage. Figure 5 shows representative data from two of these clones. For each of the 20 clones, identical patterns were found indicating that no rearrangements or deletion which altered any of the restriction fragments had occurred during 160 generations of serial growth.

![Figure 5. Stability of Fosmids containing human DNA.](image-url) DNA from two chromosome 22 Fosmids was prepared on day 1 (d1) and day 8 (d8) of serial passage and digested with EcoRI (E), HindIII (H) or BamHI (B). Sizes in Kb are indicated.
To examine the stability of even larger regions of human DNA cloned in *E. coli*, overnight cultures of two clones containing 170 Kb and 125 Kb inserts, were diluted and about 10⁶ bacteria were inoculated into 500 ml of fresh broth to grow to late log phase. This process was repeated for about 100 generations. Pulsed field gel electrophoresis shows no visible difference between samples taken before or after passage after digestion with several different restriction enzymes (NotI, NheI, SacI, and XhoI) (Figure 6). Twelve individual colonies isolated from the first and fifth day liquid culture were also analyzed separately by NotI digestion. No difference between any of these colonies was detected. Detecting rearrangements with the above test of serial passage assumes either that rearranged clones will have a growth advantage over the parental clone or that rearrangements with similar growth rates must arise at a frequency of greater than 0.1% per generation. We have also used a second test of clone stability in which we pick random BAC clones and re-streak. DNA from a minimum of ten individual colonies for each clone was examined. In each case, all ten colonies were identical to each other (and to the parental clone).

These results with our Fosmid and BAC clones stand in sharp contrast to the results obtained with human DNA cloned in two different cosmid vectors. Lawrist 16 (Pieter de Jong, personal communication) (which uses a lambda replication origin) and Supercos (which uses a pBR origin) (Stratagene, San Diego, California), like pFOS I both have double cosN sites permitting us to test the stability of clones generated in parallel from the same DNA source using the same cloning procedure for each vector. Random colonies were picked and diluted daily, and DNA was prepared after 20, 40, 60, and 80 generations of growth under the same conditions used for propagating the Fosmids. Eco RI restriction digests were compared for the cultures. For Lawrist, 7 out of 18 showed large deletions after just 40 generations of growth. By 60 generations these had completely overgrown the parental clone. For the Supercos, 7 out of 16 clones tested showed large deletions in the same time period as the Lawrist 16 clones. Thus, using less stringent test conditions than we had applied to the Fosmid clones, roughly 40% of these cosmid clones underwent massive rearrangement of the inserted human DNA. Figure 7 shows typical rearrangements occurring with these vectors.
Figure 7. Instability of human DNA in conventional cosmid vectors. Lawrist 16 (L) and supercos (S) cosmids were grown by serial passage and DNAs prepared on day 1, day 2, day 3, and day 4 were digested with EcoRI and analyzed by agarose gel electrophoresis.

While a simple conclusion from the above is that serial passage of cosmids is to be avoided, we feel the greater significance comes from recognizing that when generating and characterizing libraries, by the time a new clone has been grown after transfection, picked and grown in a microtiter plate, and grown for preparation of DNA, a minimum 30-40 fully unavoidable generations have passed. In our experiments cosmid subpopulations showing rearrangements were often apparent after just 20 generations. Complete overgrowth of the deleted form often occurred in 40 generations. Even if intact cosmids can be generated and screened, this tendency to rearrange makes further study and propagation of specific regions difficult. In the well studied C. elegans system (Coulson et al., 1988) cosmid instability was shown to play a role in all the gaps in the continuity of the map. For mammalian genomes the problem of unstable regions present an even greater obstacle. Thus, Fosmid stability will provide an advantage that could lead to fuller representation.

D. Mapping by FISH

In the past year we have become proficient in the methods of probe localization by fluorescent in situ hybridization. As an initial test of both our facility with the technique and of our chromosome 22 specific Fosmid library, 10 Fosmids were randomly picked and mapped on metaphase spreads of human chromosomes. In spite of the smaller amount of DNA obtained from minipreps of Fosmid clones, all of these clones gave strong fluorescent signals. Figure 8 shows some of these results and demonstrates that Fosmid oriented mapping can clearly exploit the power of FISH. Further, all of these clones mapped to chromosome 22. These clones were distributed along chromosome 22, without obvious clustering that might indicate uneven representation in the library. Figure 8 shows that we can distinguish the relative positions of Fosmids A3, A5, and A6 in the telomeric end of the q arm. Figure 9 shows that we are also able to perform multi-color FISH. This enables us to simultaneously localize an unknown probe labelled with one fluorescent marker relative to a set of "standards", consisting of probes from known locations labelled with a separate fluorescent marker. Multi-band pass filters and our image processing capability permit distances between the signals to be measured and statistics compiled without encountering registration problems. The hybridization signal from three different Fosmids can be distinguished on chromosome 22, each being recognized by its color. Based on our experience with multiple probes we believe that we can achieve resolution of 4 Mb on chromosome 22 by FISH with metaphase spreads.
Figure 8. *In situ* hybridization of chromosome 22 Fosmids. Three Fosmids were mapped to distinct areas near the telomere of 22q by FISH with human metaphase chromosomes. Left to right clone 1A3, 1A5, and 1A6.

Figure 9. Multicolor *in situ* hybridization to chromosome 22. These separate probes were labelled and simultaneously hybridized to metaphase human chromosomes. The relative position of these probes are shown in the drawing on the left. Signals from the Fosmid 1H8 (green), GNAZ (red), and Fosmid 1A5 (green) are visible in the middle right. DAPI counter-staining of the chromosomes is shown on the right.
Figure 10. *In situ* hybridization of BAC clones to fluorescence banded chromosomes. Hybridization signals from BL2 (175 Kb human insert) above, and BL6 (215 Kb human insert) below.
We are also able to use our larger BAC clones as probes for in situ hybridization. Figure 10 shows hybridization signals obtained from labelled BAC clones in collaboration with Dr. Julie Korenberg. These BACs were derived from a library that we are preparing with total human DNA. To date we have found no evidence for co-cloning events in our BAC clones by FISH. Two of the initial series of BACs localized by in situ hybridization (Figure 10) appear to warrant special attention. They map at loci currently undergoing intense investigation to find probes linked to diseases in this region. On the basis of FISH and chromosome banding, BL6 maps to chromosome 8-q24.13, the location of probes for Langer-Gideon Syndrome. BL2 maps to 10p11.21, the band that has been localized for Multiple Endocrine Neoplasia Type IIb. Since both BACs have inserts in the range of 200 Kb they may be of use to workers trying to localize markers in these regions.

E. Transposon Facilitated DNA Sequencing

In anticipation of the need for methods to rapidly sequence contiguous regions of the chromosome we (together with our collaborators in the Meyerowitz laboratory) developed a method for introducing and mapping minitransposons into target DNA. The ends of a progressive set of insertions act as primers for DNA sequencing. The method is rapid and adaptable to obtaining sequence from a targeted region of a large insert or for complete automated sequencing of long DNA segments.

F. Publications


3. Proposed Experiments - Design and Methods

In the twenty-two months that we have been working on this project we have accomplished approximately two-thirds of the goals that we designated in our original proposal. We have in hand two chromosome 22 specific Fosmid libraries; one that is seven-fold redundant made from flow sorted human chromosome 22 together with Pieter deJong at Lawrence Livermore Laboratory and the other that was derived in our laboratory directly from the KG-1 strain and can be readily extended to any desired level of coverage. We have demonstrated using FISH that markers chosen randomly from this library mapped to chromosome 22 and these Fosmids are extremely stable. We have also demonstrated that we can prepare clones containing fragments as large as 225 Kb. These BAC clones show remarkable stability and while we are continuing to check them by FISH, the early