Direct construction of a chromosome-specific *NotI* linking library from flow-sorted chromosomes

---

Margaret R.Wallace, Jane W.Fountain, Anne M.Brereton and Francis S.Collins*

Howard Hughes Medical Institute and Departments of Internal Medicine and Human Genetics,
University of Michigan, Ann Arbor, MI, USA

Received October 13, 1988; Revised and Accepted January 16, 1989

ABSTRACT

A linking library consists of genomic DNA fragments which contain a specific rare restriction enzyme site; such clones are very useful as probes in pulsed field gel electrophoresis and in mapping and cloning large regions of DNA. However, identifying those linking clones which map to a certain chromosomal region can be laborious. Therefore, we have developed a straightforward procedure for constructing a linking library directly from flow-sorted chromosomes. As a test of the approach, a *NotI* linking library was constructed from the chromosome 17 fraction of a flow-sort of human chromosomes, using only 70 ng of DNA. Thirteen of sixteen linking clones were mapped to chromosome 17, suggesting that the library is highly enriched for this chromosome. This method should be generally applicable to other chromosomes and enzymes as well.

INTRODUCTION

Pulsed field gel electrophoresis (PFGE) is a relatively new and powerful technique which allows the mapping of rare restriction enzyme sites across large chromosomal regions (e.g. 100-10,000 kb) (1,2,3,4,5). Once a coarse PFGE map of the region has been developed, specific subdivisions of interest can be studied in finer detail. However, PFGE mapping strategy requires that a sufficient number of clones be spaced within reasonable distances of each other in the region in order to physically connect probes. This requirement is often difficult to satisfy, resulting in incomplete maps. Such a problem can be minimized by using probes which contain the rare sites useful in PFGE studies. These genomic clones, called linking clones, are particularly valuable for PFGE analysis because they can detect PFGE fragments which are physically adjacent on the chromosome (6,7,8). Surrounding rare sites (of the same enzyme) can be

readily mapped with linking clones, using partial digests with PFGE. Linking clones are also valuable in that they often mark the locations of transcribed genes (6).

There are a number of strategies for the construction of linking libraries (6,7,8,9,10,11). However, with total genomic linking libraries, many random clones have to be mapped in order to find those from the region of interest. For many purposes, it would be more expedient to generate a chromosome-specific linking library. While such libraries can be constructed from already-existing chromosome-specific genomic libraries (6,9), we have devised an alternative strategy which has certain advantages. This alternative method of library construction uses flow-sorted chromosomes as the source DNA. In this paper we report the generation of a NotI linking library from a human chromosome 17 flow-sort fraction. Analysis of multiple clones has shown that this library is greatly enriched for chromosome 17 linking clones and should represent most of the commonly unmethylated chromosome 17 NotI sites in human lymphoblast DNA. The method described is straightforward and requires very little DNA, and ought to be applicable to other enzymes and chromosomes. It provides a powerful tool for obtaining a set of linking clones from a particular chromosome.

MATERIALS AND METHODS

Library construction

A chromosome 17 fraction was obtained from a flow-sort of human lymphoblastoid chromosomes, and contained 1.4×10^6 chromosomes, representing approximately 140 ng DNA. The DNA was extracted from the chromosome material by ultracentrifugation at 40,000 g followed by proteinase K treatment, phenol extractions, and precipitation (12). Following the strategy outlined in Figure 1, the DNA was digested to completion with BglIII, and half (70 ng) was ligated at low concentration to promote self-circularization in the presence of an excess of BamHI-ended suppressor tRNA genes (supF) (13). Specifically, the chromosome 17 DNA concentration in the ligation was 1.8 ug/ml, and a 25-fold molar excess of BamHI-ended supF was present (concentration 1 ug/ml). The ligated DNA was precipitated and

Figure 1. Schematic of strategy for constructing the chromosome 17 *NotI* (N) linking library. Represented are the two kinds of fragments generated by *Bgl* III (B) digest: those which contain *NotI* sites and those which do not. The 240 bp *supF* gene is indicated by the hollow sector in each circle. The *Bam* HI ends (the *supF* ends) and the *Bgl* III ends are complementary, and when ligated to each other result in the elimination of both sites. Only those circles containing unmethylated *NotI* sites (approximately 1 per 100-200 circles) linearize with *NotI* and ligate into the *NotI* site of the vector (λ Ch3A Δ lacXN). The phage arms are designated by the hatched boxes in the bottom part of the figure.

digested with *NotI*. The vector for the final ligation was bacteriophage λ Ch3A Δ lacXN, a λ Ch3A derivative containing a *NotI* cloning site which accepts inserts of size 0-12 kb. This vector

was constructed as follows: λ Ch3A DNA was digested to completion with EcoRI and partially digested with BamHI. The EcoRI-BglII polylinker of π vx (14) was ligated to the phage DNA, resulting in a phage called λ Ch3A Δ lac. λ Ch3A Δ lac was then digested with EcoRI and the ends phosphatased prior to ligation with the NotI-containing linker

5' pAATTGCGGCCGC 3'
3' CGCCGGCGTTAAp 5'

which resulted in the final vector λ Ch3A Δ lacXN. The NotI-digested DNA was ligated into the NotI site of the vector, packaged, and plated on the sup⁻ E.coli strain MC1061 to select for supF-containing phage (14). The resulting library was screened with total human DNA, and sixteen positive plaques were grown and DNA purified (14).

Southern analyses

For the purpose of mapping and analyzing linking clones, Southern blots (15) were made with human and rodent DNA as well as DNA from several chromosome 17 rodent-human hybrids. The DNAs (10 ug) were digested with PstI or PvuII, electrophoresed through a 1% agarose gel, and transferred to Hybond N membranes as previously described (16). Probes from the linking clones were prepared by digestion of phage DNA with NotI, followed by electrophoretic separation of the NotI inserts on a gel and purification of the inserts with the Geneclean procedure (Bio 101, LaJolla, CA). The inserts were radiolabelled by the random primer method (17). Hybridizations were carried out as previously described (16), except in the case of probes containing highly repetitive sequences, where filters were prehybridized with 250 ug/ml denatured sonicated human placental DNA at 65°C for 1-2 hours. Repetitive probes were also annealed to placental DNA (same concentration) for 1-2 hours prior to being placed on the filters.

Pulsed field gel analyses

Human and hybrid DNAs (5 ug in agarose blocks) were digested, electrophoresed on field inversion gels, and blotted as previously described (16). The same hybridization conditions were used as for Southern analyses.

RESULTS

Six thousand plaques were obtained from the 70 ng of chromosome 17 DNA. From a random sampling of 30, we unexpectedly found that a very high proportion (80%) of the clones contained supF with no genomic DNA evident, apparently due to nonhomologous ligation of BamHI ends to NotI ends. Therefore we screened the library with total human DNA as a means of identifying legitimate linking clones. Approximately 10% of the original clones hybridized to total human DNA (i.e., contained highly repetitive sequences). Sixteen of these repetitive clones were mapped against the chromosome 17 hybrids, with three proving to be from other chromosomes while the remaining 13 (81%) resided on 17q. The average insert size of the sixteen clones was 5.5 kb. As an example, Figure 2 shows the resulting autoradiographs from two clones, R8 and R9. Both probes revealed human-specific bands in lanes H (human DNA), 17 (a mouse-human hybrid containing only human chromosome 17), and 17q (a mouse-human hybrid carrying an isochromosome 17q), thereby mapping both clones to the q arm of 17. Clone R9 also hybridized with mouse DNA (lane M), and was one of six clones which showed homology with mouse and/or rat DNA.

One representative clone, 17L1, was further characterized to establish that its structure was correct. Hybrid mapping indicated that this clone represents a single locus between the centromere and 17q12. Restriction digests showed that the 7.2 kb insert contained the 240 bp supF fragment and spanned a NotI site, with 4.2 kb of genomic DNA on one side of the site and 2.8 kb on the other. To verify this, fragments isolated from each side of the NotI site were used to probe human DNA digested with BglII alone or BglII and NotI together. The resulting autoradiographs (Figure 3) showed that 17L1 is derived from one contiguous 7.0 kb genomic BglII fragment containing an unmethylated NotI site. Analysis of 17L1 and preliminary digests of the other 15 clones showed that all possessed correct linking clone structure, with one clone containing a double insert.

To test the applicability of these linking clones in large-scale restriction mapping, PFGE blots prepared from

Figure 2. Results of hybridization of two linking clones to chromosome 17 somatic cell hybrids and control DNAs digested with *Pvu*II. M represents the mouse DNA lane, H is human DNA, 17 is a mouse-human hybrid containing only human chromosome 17, and 17q is a mouse-human hybrid containing a human isochromosome 17q. The linking clones R8 and R9 have inserts of sizes 5 and 6 kb respectively. Both clones are positive in the 17q lanes and therefore reside on the q arm. Only clone R9 (right panel) shows homology with mouse DNA, with the human (h) and mouse (m) bands indicated. Final washes were in 0.1xSSC/0.1% SDS at 65°C for 20 minutes.

complete and partial *Not*I digests were probed with two 17L1 fragments from opposite sides of the *Not*I site. As shown in Figure 4, the probes (A and B) each detected a unique smaller *Not*I fragment, and, as well, hybridized to the same 430 kb larger fragment. Since neither probe spanned the *Not*I site and the upper band was much fainter than the lower bands, this suggested that the 430 kb fragment was a partial digestion product arising from incomplete digestion at the *Not*I site between the two probes.

DISCUSSION

The use of linking clones in PFGE analysis is a powerful tool for mapping large regions of DNA (4,6,7,8,10,13). Most of the restriction enzymes used in generating such libraries (e.g. *Not*I) contain the commonly-methylated sequence CpG in their

Figure 3. Analysis of the structure of the linking clone 17L1. The panels show autoradiograph results from human leukocyte DNA digested with *Bgl*II (Bg) or *Bgl*II and *Not*I (BgN (a double digest)), blotted, and probed with two fragments of 17L1. The left two lanes show results from probe A (open box in the schematic below), an 800 bp *Not*I/*Xho*I fragment from one side of the *Not*I site in 17L1. Probe B (dark box in schematic) is a 700 bp *Not*I/*Pvu*II fragment from the other side of the *Not*I site, and results with this probe are in the right two lanes. Neither probe contains highly repetitive sequences. As summarized in the bottom schematic, both probes hybridize to the same 7.0 kb genomic fragment, and the *Not*I site between them divides the fragment into segments of 4.2 and 2.8 kb.

recognition sites, and only when unmethylated can these sites be cut. Unmethylated rare sites tend to be clustered in HTF islands, which are thought to be associated with actively transcribed regions (18). Therefore linking clones are especially interesting when searching for a specific locus such as a disease gene, as they may identify candidate genes. In support of this notion, six of the sixteen linking clones shared sequence homology with other species, which is a common property

Figure 4. Illustration of *NotI* mapping using the two halves of chromosome 17 linking clone 17L1. Lanes 1 and 2 contain human lymphoblastoid and fibroblast DNA samples respectively, which were partially digested with *NotI* and separated by field inversion gel electrophoresis (5). This blot was hybridized with the 700 bp probe B from 17L1 (see Figure 3), resulting in identification of a primary band at 150 kb and a faint secondary band at 430 kb (lanes 3 and 4, which are the same as lanes 1 and 2 respectively). The filter was then stripped and hybridized with probe A (800 bp fragment, Figure 3), which recognized a primary band of 280 kb and also showed the same faint 430 kb band (lanes 1 and 2). The derived *NotI* map of the region is shown in the schematic at the bottom.

of transcribed genes. Thus, the chromosome 17 linking library may be particularly useful in cloning disease genes such as those associated with von Recklinghausen neurofibromatosis (NF1) (17q11.2) (16), Miller-Dieker syndrome (17p13) (19), colon cancer (17p) (20), and acute promyelocytic leukemia (17q11-12) (21). An important advantage with linking clones is that they can also be used in conjunction with jumping libraries (based on the same rare-cutting enzyme) to move rapidly in a directed

fashion along a chromosome (6,8,10,22). In a recent study, linking and jumping clones were effectively used in characterizing the region of the Huntington Disease locus (10).

Figure 4 illustrates the utility of linking clones in rare-site restriction mapping. The use of probes from both sides of 17L1's NotI site allowed us to map the adjacent NotI sites. This was based on the use of partial restriction digestion and is a simple example of the Smith-Birnstiel mapping technique (13). Using both halves of linking clones as probes for partial digestion PFGE, one can map in both directions (i.e. "double-barreled" Smith-Birnstiel mapping (6)). By expanding this strategy, a large region of DNA could be mapped with just a few linking clones.

A shortcoming of the library described here is the presence of background clones which are phage containing only the supF gene. When these clones were digested with NotI and examined by Southern analysis, the inserts were found to be composed solely of supF fragments with no obvious genomic DNA present. The background clones apparently arose from aberrant ligation of supF (BamHI ends) directly into the vector (NotI ends). Such ligation may have been a normal rare event that was magnified during linking library construction, which selects for rare events. We have seen an identical background problem in another library constructed in the same fashion, but not in a partial-digest linking library made with a different vector (one that requires larger inserts for viability). Upon further examination, we found that the vector contained multiple tandem NotI sites rather than the expected single site, apparently due to failure to remove the excess linkers during vector construction. This explains the apparent conservation of the NotI site in the supF clones when aberrant ligation should have destroyed the site. In order to avoid these background clones, we screened the library with human DNA to identify insert-bearing phage. This screening step was not a major inconvenience because the library contained so few clones. A disadvantage, however, is that screening with human DNA fails to detect non-repetitive legitimate clones, which may constitute a substantial fraction. Alternatively, the library could have

been screened by analyzing the insert sizes of random clone minilysates, which easily distinguishes supF from genomic inserts. With this alternative method, unique and low-copy as well as repetitive clones could be identified with no bias toward one or the other category.

Screening of the library with human DNA revealed that 10% of the 6000 plaques contained repetitive sequences, with approximately 80% of clones being background (supF-only) clones. This would seem to indicate that about half of the legitimate NotI linking clones contain repeats, in contrast to the study by Buiting et al. (9) in which 95% of their SacII linking clones were unique or low-copy. However their library contained clones of both methylated and unmethylated genomic sites, which may represent DNA with differing composition relative to repeats. In addition, their insert size averaged 3.4 kb, whereas the legitimate flow-sort clones in our library were of average size 5.5 kb. On the assumption that about 50% of the linking clones in our library do not contain repeats, it is reasonable to infer that the library contains at least 1200 legitimate clones. With 3000 Mb in the haploid human genome and assuming that the average NotI PFGE genomic fragment is 500-1000 kb in size (6,23), one expects to find 3000-6000 unmethylated NotI sites per haploid genome, or about 130-260 per average chromosome. Since chromosome 17 is somewhat smaller than average, we estimate that it contains 90-200 unmethylated NotI sites. From the calculations above, the 1200 or so linking clones should represent most or all of the unmethylated chromosome 17 NotI sites which are clonable with this method. This library and a NotI jumping library (6,8,10,13,22) could be conveniently used together to map consecutive unmethylated NotI sites on chromosome 17 (6,22).

Ideally, a linking library should be constructed from size-selected, partially-digested DNA to maximize the chances of cloning all of the specific rare sites and to obtain larger stretches of DNA on both sides of the sites. However, this is not feasible when starting with flow-sorted chromosomes because there is insufficient DNA for size selection (and cloning from a partial digest is generally too inefficient without size

selection). Thus we chose the complete BglII digest approach, realizing that only BglII fragments of size less than 12 kb would be cloned. This may well result in the absence of a small proportion of NotI linking clones from this library. A possible solution to this problem would be the parallel construction of a complementary linking library using a different enzyme such as BamHI, which could be carried out precisely as in Figure 1.

Several other approaches for construction of chromosome-enriched linking libraries have been suggested, each with advantages and disadvantages. For example, linking libraries can be created from previously-existing chromosome-specific genomic libraries (which were initially made from flow-sorted chromosomes) (6,9). An advantage of such a "secondary" approach is the availability of large quantities of source DNA. However, most of those initial libraries originated from complete HindIII or EcoRI digests with average insert sizes of 4-5 kb (24), so resulting linking clones may not be especially large (9). Also, as with the approach reported here, those original chromosome-specific libraries lack some sequences due to size limitations of their cloning vector, and so resulting linking libraries may be incomplete. Such "secondary" linking libraries have an additional disadvantage: through the processes of cloning and amplification, chromosome-specific genomic libraries have lost the source DNA's original methylation pattern and thus the rare sites which are normally methylated and unrestrictable in vivo will be cloned in "secondary" linking libraries. Since these sites do not usually cut in the genome or mark HTF islands, such clones are less valuable for PFGE mapping or for linking and jumping strategies. The flow-sort linking library approach clones only the unmethylated sites, which are more likely to be useful in such studies. Additionally, amplification of the chromosome-specific genomic libraries may have resulted in under- or over-representation of various sequences, which would be reflected and perhaps magnified in "secondary" linking libraries.

Alternatively, somatic cell hybrid DNA could be employed as the DNA source for a linking library, with desired linking

clones identified with a species-specific probe (6) as was done by Pohl et al. for chromosome 4p (10). This is an especially useful approach in that specific sub-chromosomal regions can be targeted, provided that for each purpose the hybrids exist and do not contain excessive rearrangements. The preferred partial-digest method of linking library construction mentioned above can also be used with the hybrid approach. However, a somatic cell hybrid linking library will contain many more clones than a similar flow-sort library, thus requiring a greater screening effort. Also, as discussed previously, screening with a species-specific probe will miss non-repetitive clones, although this problem is minimized in the partial-digest technique since the inserts are larger. There is also a slight risk of cross-hybridization of the probe to rodent DNA (10).

Considering the small quantity of starting material, the flow-sort method of linking library construction is quite efficient and has provided clones which will be valuable in NotI mapping of chromosome 17. The specificity of other libraries constructed in this fashion will of course depend on the particular chromosome and the quality and purity of the flow-sorted material (24). The method of library construction from flow-sorted chromosomes has both advantages and disadvantages relative to other techniques, but in many cases will be an attractive approach for obtaining region-specific linking clones. This method should be applicable to other human chromosomes as well as chromosomes from other species. Linking libraries generated by this and other approaches should prove valuable in mapping and characterizing complex genomes.

ACKNOWLEDGEMENTS

We are grateful to Jeffery Cole and Janice Weber for technical help and advice. We thank Dr. Joe Gray (Lawrence Livermore Laboratories) for the flow-sorted material, Dr. David Ledbetter (Baylor University) for somatic cell hybrids, and Drs. Fred Blattner (Wisconsin) and Jungsuh Kim and Kirk Fry (Genelabs Inc.) for the vector. This work was supported by the Howard Hughes Medical Institute and National Institutes of Health (NIH) grants GM34960, NS23427 and NS23410 to F.S.C., and NIH-5-T32-GM07544-10 to J.W.F.

---

*To whom correspondence should be addressed at: Howard Hughes Medical Institute, University of Michigan, 5518 MSRB I, 1150 W. Medical Center Drive, Ann Arbor, MI 48109-0650, USA

REFERENCES

1. Schwartz, D.C. and Cantor, C.R. (1984) *Cell* 37:67-75.
2. Carle, G.F. and Olson, M.V. (1984) *Nucleic Acids Res.* 12:5647-5664.
3. Chu, G., Vollrath, D. and Davis, R.W. (1986) *Science* 234:1582-1585.
4. Smith, C.L. and Cantor, C.R. (1987) *Methods Enzym.* 155: 449-467.
5. Carle, G.F., Frank, M. and Olson, M.V. (1986) *Science* 232:65-68.
6. Smith, C.L., Lawrance, S.K., Gillespie, G.A., Cantor, C.R., Weissman, S.M. and Collins, F.S. (1987) *Methods Enzym.* 151: 461-489.
7. Poustka, A., Pohl, T., Barlow, D.P., Zehetner, G., Craig, A., Michiels, F., Ehrlich, E., Frischauf, A.-M. and Lehrach, H. (1986) *Cold Spring Harbor Symp. Quant. Biol.* 15:131-139.
8. Poustka, A. and Lehrach, H. (1986) *Trends Genet.* 2:174-179.
9. Buiting, K., Passarge, E. and Horsthemke, B. (1988) *Genomics* 3:143-149.
10. Pohl, T.M., Zimmer, M., MacDonald, M.E., Smith, B., Bucan, M., Poustka, A., Volinia, S., Searle, S., Zehetner, G., Wasmuth, J.J., Gusella, J., Lehrach, H. and Frischauf, A.-M. (1988) *Nucleic Acids Res.* 16:9185-9198.
11. Ito, T. and Sakaki, Y. (1988) *Nucleic Acids Res.* 16:9177-9184.
12. Fuscoe, J.C., Clark, L.M. and Van Dilla, M.A. (1986) *Cytogenet. Cell Genet.* 43:79-86.
13. Collins, F.S. (1988) In Davies, K. (ed), *Genome Analysis, A Practical Approach*, IRL Press, London.
14. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor.
15. Southern, E.M. (1975) *J. Mol. Biol.* 98:503-517.
16. Fountain, J.W., Wallace, M.R., Brereton, A.M., O'Connell, P., White, R.L., Rich, D.C., Ledbetter, D.H., Leach, R.J., Fournier, R.E.K., Menon, A.G., Gusella, J.F., Barker, D., Stephens, K. and Collins, F.S. (1989) *Am. J. Hum. Genet.* 44:58-67.
17. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* 132: 6-13.
18. Bird, A.P. (1986) *Nature* 321:209-213.
19. Stratton, R.F., Dobyns, W.B., Airhart, S.D. and Ledbetter, D.H. (1984) *Hum. Genet.* 67:193-200.
20. Monpezat, J.P., Muleris, M., Lefrancois, D., Grunwald, D., Salmon, R.J., Dutrillaux, B. and Thomas, G. (1987) *Cytogenet. Cell Genet.* 46:663.
21. Bloomfield, C.D., Trent, J.M. and Van Den Berghe, H. (1987) *Cytogenet. Cell Genet.* 46:344-345.
22. Poustka, A., Pohl, T.M., Barlow, D.P., Frischauf, A.-M. and Lehrach, H. (1987) *Nature* 325:353-355.
23. Lindsay, S. and Bird, A.P. (1987) *Nature* 327:336-338.
24. Gray, J.W., Dean, P.N., Fuscoe, J.C., Peters, D.C., Trask, B.J., Van Den Engh, G.J. and Van Dilla, M.A. (1987) *Science* 238:323-329.