Derivation of Clones Close to *met* by Preparative Field Inversion Gel Electrophoresis

FRANK MICHIELS, MARGIT BURMEISTER, HANS LEHRACH*

The molecular analysis of genes identified by mutations is a major problem in mammalian genetics. As a step toward this goal, preparative field inversion gel electrophoresis (FIGE) was used to selectively isolate clones from the environment of genetically linked markers, and to select a subset of these clones containing sequences next to specific restriction sites rare in mammalian DNA. This approach has been used to generate a library highly enriched in sequences closely linked to the cystic fibrosis marker *met*. One clone derived from the end of a Not I restriction fragment containing the *met* sequence was analyzed in detail and localized within a long range map to a position 300 kilobase pairs 5′ of the *met*D sequence.

MAMMALIAN MOLECULAR GENETICS has long been limited to the study of genes whose products are known. This is mainly due to the difficulty of combining genetic information, typically with a resolution of millions of base pairs, with molecular techniques best suited to the analysis and cloning of, at best, hundreds of kilobase pairs (kbp). A major step in overcoming this limitation has been the development of pulsed field gradient (PFG) gel electrophoresis, which can separate DNA molecules of hundreds to thousands of kilobase pairs in length (*1*). Both the orthogonal field alteration (OFAGE) (*1, 2*) and field inversion gel electrophoresis (FIGE) (*3*) types of PFG allow restriction mapping over more than 1000 kbp and can bridge genetically linked markers (*4, 5*). Analogously, the range of cloning techniques has been expanded by the development of chromosome jumping and linking libraries (*6*).

As an alternative and complementary approach, we introduce here the preparative use of FIGE to (i) generate libraries highly enriched in subchromosomal fragments, and (ii) to selectively clone sequences bordering restriction sites that are recognized by enzymes cleaving rarely in mammalian genomes. Experiments are described to generate additional probes of potential use in the analysis of the region of the human cystic fibrosis mutation.

Cystic fibrosis is one of the most common autosomal hereditary diseases in Caucasians, caused by a single-locus recessive mutation on chromosome *7*. Since the primary biochemical defect in this disease is not known, this mutation has been the target of intense efforts to use molecular genetics as a means of identifying the mutated gene. This has led to the isolation and characterization of a number of DNA markers, including the markers *met* and pJ3.11, that probably flank the mutation, and are separated from it by less than 1 centiMorgan. On average, 1 centiMorgan corresponds to 1000 kbp in humans (*7, 8*). The orientation of cystic fibrosis relative to *met* has not been firmly established, although there have been some indications that the disease gene is 3′ of *met*.

To contribute to a genetic analysis of this region, and as a possible step toward a molecular analysis of the gene, we have used preparative FIGE of Not I-digested DNA to isolate a size fraction containing a 450-kbp Not I fragment that hybridizes to *met* probes. To allow a second enrichment by selection of clones containing inserted human DNA, a hamster cell line containing one copy of chromosome 7 as the only human chromosome (cell line 4AF) (*9*) was used as source of DNA.

After electrophoresis, a region of the gel containing the 450-kbp region was sliced and sequences hybridizing to the probe

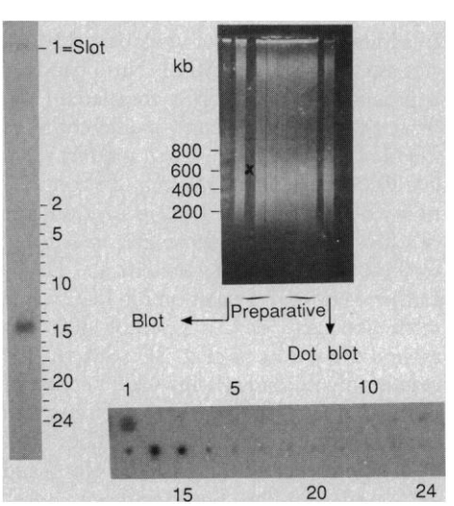

Fig. 1. Isolation of the *met* Not I band. After the preparative FIGE (ethidium bromide picture is shown with yeast strain GY22 as size marker), one lane was Southern blotted, the rest were sliced and slices from one half-lane were analyzed by dot blot hybridization. A library was constructed from slices 14 and 15. The lane marked by X separates the lane used for the Southern blot from the lanes used preparatively.

*met*D (*7*) were identified by Southern blot and dot blot analysis (Fig. 1) (*10*). DNA isolated from these gel slices was then digested partially with Sau 3AI (*11*). Two types of clonable DNA fragments were generated. First, fragments from the ends of the original Not I fragment were identified, which can be used as entry points for chromosome jumping and other walking strategies. By means of such fragments it is possible to move from a linked marker to a gene and to derive physical maps. Second, internal fragments terminating on both sides in Sau 3AI restriction sites were isolated that cover the entire Not I fragment and can be used to identify a sequence of interest within it.

To allow parallel cloning of both types of fragments and yet permit discrimination of the different types of clones, a mixture of two genetically distinguishable λ cloning vectors was used in the ligation step (Fig. 2A). Alternatively, either type of fragment could be cloned from the corresponding single ligation. One vector, a *Sam7* derivative of EMBL3 named EMBL3S (*12*), can be digested with Bam HI, and will ligate to the fragments derived from the interior of the DNA fragment. Because of the amber mutation in the *S* gene on the right arm, phage will only be able to grow on a suppressor-providing host. The other vector, a derivative of EMBL3 named EMBL6, in which a Bam HI site next to the right arm has been replaced by a Not I site (*13*), is cleaved with both Bam HI and Not I. EMBL6 will therefore selectively ligate with fragments carrying both Sau 3AI and Not I ends that originate from the ends of the Not I fragment. This vector carries a wild-type allele of the *S* gene, and can therefore grow on suppressor-free host strains. To prepare a total library (both internal and end fragments) 20% of the ligated and in vitro packaged material was plated on the *Escherichia coli* host NM539 (rK^-, mK^+, *supF*, P2) (*12*), combining a selection for inserts (Spi^- phages, which can grow on bacteria containing the P2 phage) with the suppressor function required to allow EMBL3S clones to grow. The remaining 80% were plated on *E. coli* W3110 (rK^-, mK^+, sup^0, P2) (*14*), producing an end clone library of 6000 clones, all of which are expected to contain Sau 3AI–Not I inserts.

To derive an estimate for the enrichment by the gel purification step, the total library of 30,000 plaques was hybridized to *met*D. The average insert size in EMBL3 is 15 ×

European Molecular Biology Laboratory, D-69 Heidelberg, Meyerhofstrasse 1, Federal Republic of Germany.

*Present address: Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, England.

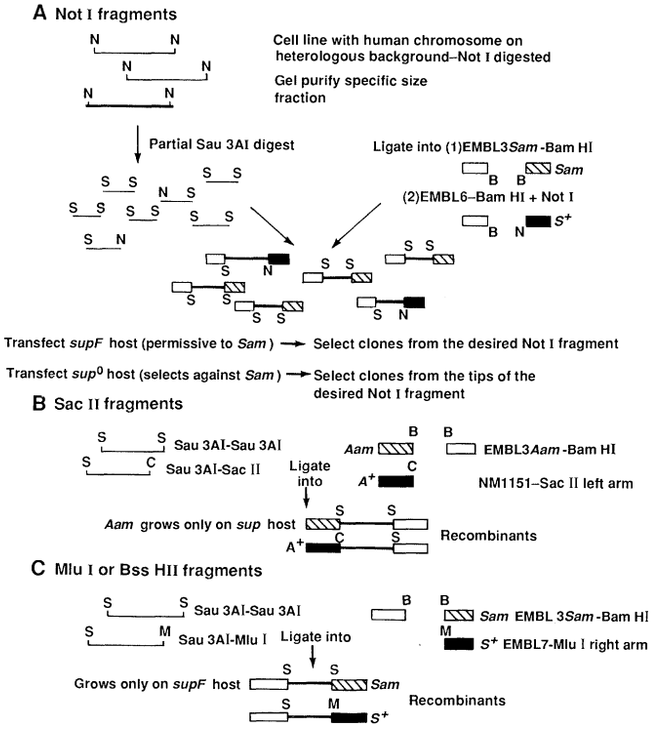

Fig. 2. (**A**) Selective cloning of internal and end fragments of the 450-kbp *met* Not I band. Crosshatched rectangle symbolizes a λ right arm with *Sam* mutation; the dotted rectangle indicates an arm with a wild-type *S* gene. (**B**) Parallel cloning of internal and Sac II end fragments by means of the 20-kbp Sac II fragment from λ NM1151 (*20*) as a left arm for Sac II end clones, and Bam HI–cleaved EMBL3*Sam* as the corresponding right arm and as a vector for internal fragments. Since NM1151 contains multiple Sac II sites, it is advisable to purify the 20-kbp vector arm (for example, on agarose gels) (*10*). (**C**) Cloning of internal and Mlu I or Bss HII end fragments. This uses a gel-purified 9-kbp Mlu I fragment from λ EMBL7 as a right arm, in combination with Bam HI–cleaved EMBL3*Sam*. (Bss HII ends are compatible with Mlu I ends.) For the construction of EMBL7, pEMBL18 DNA (*21*) was cleaved with Xba I; the single-stranded ends were filled in with DNA polymerase and recircularized. The *lacZ* function is subsequently restored by inserting a Mlu I linker (5′ GACGCGTC3′; Biolabs) into the Hinc II site, yielding pEMBL18.Mlu I. This plasmid was digested with Eco RI, ligated to EMBL3–Eco RI, and plated on NM539 (*12*). A phage was selected that contains a tandem dimer insert of pEMBL18.Mlu I, with the Mlu I site adjacent to the EMBL3 right arm (EMBL7). Sal I and Xho I end fragments can be cloned analogously by means of an arm from a EMBL3–Sal I digest.

10^3 bp. For a diploid genome size of 6×10^9 bp, the expected frequency of *met*D-hybridizing clones should be 1 in 400,000. Six positive clones were detected, suggesting an 80-fold enrichment. This enrichment factor agrees with an independently derived estimate, which was based on the comparison of the hybridization signal of *met*D in Southern blots of gel-enriched and total human DNA. Since chromosome 7 represents about 5% of total human DNA (*15*), this level of enrichment predicts that one-fourth of the human clones in both libraries were derived from the 450-kbp Not I fragment containing the *met* marker sequence.

On the basis of this calculation, 35 clones hybridizing to human repeat sequences (*16*) (approximately four times the number of expected end clones) were isolated from the end clone library. Restriction analysis showed that the clones could be divided into groups: two end fragments (Lcn2 and Lcn6) were represented nine times, three end fragments (Lcn1, Lcn8, and Lcn33) were represented twice, and the other ends were cloned only once. All clones had the expected structure, with a Not I site adjacent to the right arm, indicating the efficiency of the selection procedure.

Next, unique subclones representing all groups with at least two members were hybridized to Southern blots of DNA from the cell line 1EF (*9*), which carries a fraction of chromosome 7 surrounding the markers pJ3.11 and *met* (*17*). Lcn1, Lcn2, Lcn4, and Lcn33 were present in this cell line; Lcn6, Lcn8, and Lcn11 were not, (*18*) so they cannot have originated from the *met* Not I fragment.

As a second, more stringent test, Southern blots were made of high molecular weight DNA (4AF cell line), that had been separated on OFAGE and FIGE (*19*) after cleavage with restriction enzymes that cut infrequently in mammalian DNA. These were hybridized to *met*D and unique fragments from each group. In this analysis, Lcn2 and *met* hybridized to identical bands in Not I and Sfi I single digests, as well as in a Not I–Sfi I double digest (Fig. 3, A to E). In addition, a number of weak partial bands (for example, Sfi I and Mlu I–Nru I bands of approximately 1000 kbp) were shared (*18*). Since the Not I–Sfi I band is different from the Not I or Sfi I band, Lcn2 and *met* share five PFG restriction fragments. In conjunction with the fact that Lcn2 is on the same chromosome as *met*, these gel data firmly establish their tight physical linkage. This is further verified by positioning Lcn2 on a restriction map in a region of 800 kbp around the *met* locus (Fig. 3F). Strictly, the map is only valid in cell line 4AF, because of possible polymorphisms and differences in methylation among different cell lines (Fig. 4). The presence of the 450-kbp Not I and 450- and 300-kbp Sfi I fragments was, however, verified in other tissues and cell lines in order to exclude major DNA rearrangements in this region of 4AF and 1EF (*18*).

To orient the map and the derived probe relative to the marker pJ3.11 and the genetic map of the chromosome, we tested the hybridization of Lcn2 to DNA [from two cell lines, Cll and Hlll) (*8*)] expected to break at the *met* locus while retaining regions distal of *met*. No hybridization was found, suggesting the position of Lcn2 proximal to the *met* marker–5′ to the *met* gene. As expected from the construction of the clones, the Not I site adjacent to Lcn2 was cleaved in the cell line 4AF used in the cloning experiment. Differences in cleavability (or presence) of this site in DNA from different cell line or tissue samples are shown in Fig. 4.

As expected, most other clones tested (Lcn8, Lcn6, Lcn4, and Lcn1) also hybridized to Not I bands of approximately 450 kbp, but their PFG hybridization pattern for other enzymes (such as Nru I) excludes the possibility that they originated from the *met* Not I fragment (*18*). Lcn33 hybridized to a Not I fragment larger than 1000 kbp. Since Lcn2 occurred nine times in 35 randomly picked end clones and the other abundant end clones did not originate from the other end of the *met* fragment, the 3′ end was either underrepresented by at least a factor of 4, or was not represented in the library. Though possibly due to chance, this could be due to the absence of an Alu repeat sequence sufficiently close to the Not I site, or due to reduced clonability of this fragment. It might be possible to clone in cosmid vectors, in order to reduce the frequency of non-Alu–containing clones.

To optimize enrichment, which reduces the effort needed for clone analysis, a number of aspects have to be considered. Enrichment factors can be optimized, for example, by considering both the position of the band relative to the overall size distribution created by the enzyme used, and the exact separation characteristics of the gel [which can be selected by varying the running conditions (*2*)]. In general, only fragments flanked by restriction sites that are cleaved well in genomic DNA will give satisfactory enrichment.

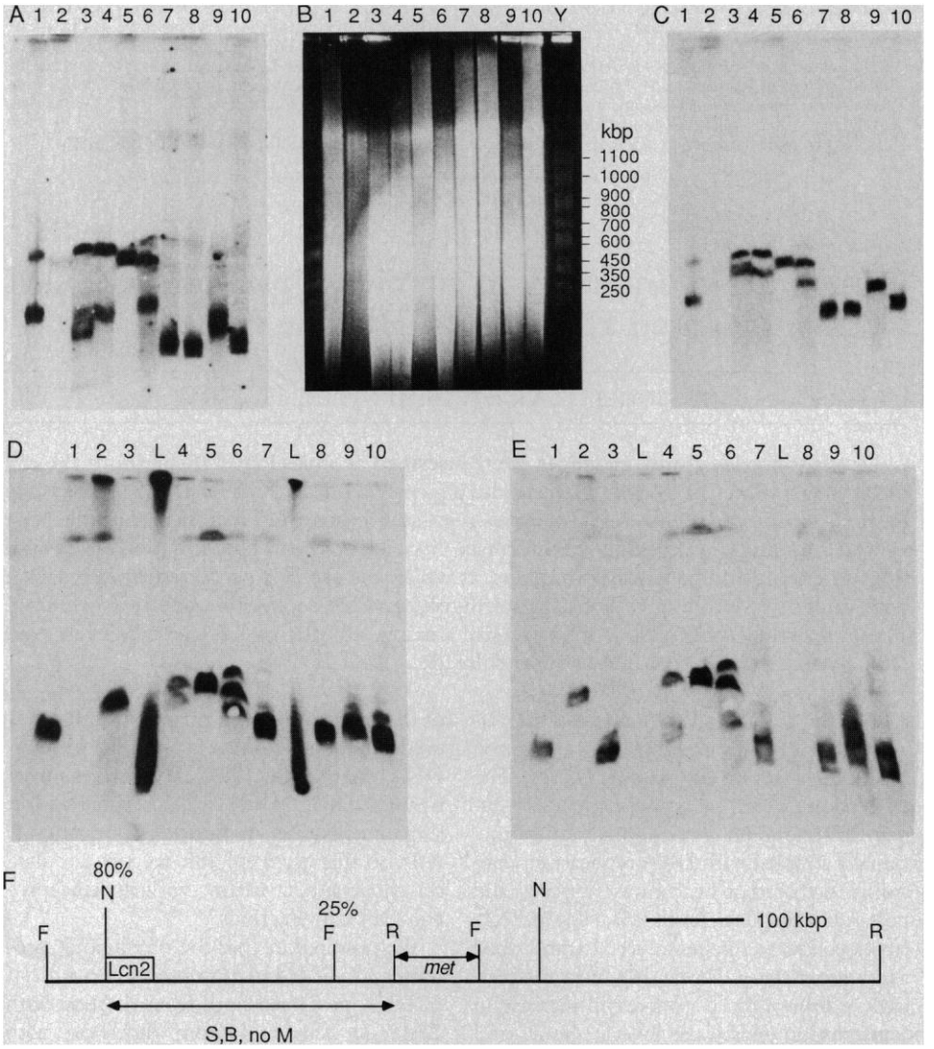

Fig. 3. PFG analysis of *met* and Lcn2. DNA (10 μg) of cell line 4AF was digested with Not I+Nru I (lane 1), Nru I (lane 2), Sfi I+Nru I (lane 3), Sfi I (lane 4), Not I (lane 5), Not I+Sfi I (lane 6), Not I+Sal I (lane 7), Sal I (lane 8), Sal I+Sfi I (lane 9), Sal I+Nru I (lane 10). It was run on FIGE (**A**, **B**, and **C**) with yeast strain IR26-19D as size marker (Y), and on OFAGE (**D** and **E**) with λ EMBL3 (42 kbp) polymers as size marker (L). Nru I cut partially in all FIGE digests, and Sfi I cut partially in the OFAGE Not I+Sfi I digest. (A) and (C) are hybridizations of the gel shown in (B). (A) and (D) are hybridizations to Lcn2, (C) and (E) are hybridizations to *met*D. In PFG, migration rates are influenced by factors besides the molecular weight (such as DNA purity, and local DNA concentration). Fragment lengths and relative positions of restriction sites are therefore only approximate. As can be seen in lanes 8 to 10, because of higher local DNA concentration, Sal I–Sfi I fragments migrate more slowly than Sal I fragments although they are the same length. This effect is more pronounced in FIGE than OFAGE. Both Lcn2 and *met* recognize a weak Sfi I band of different size. Because the size of both weak bands adds up to the common strong band, this is probably due to a partially cleavable Sfi I site between both probes. (**F**) is the restriction map derived from the digests, where N is Not I; F, Sfi I; R, Nru I; S, Sal I; B, Bss HII; and M, Mlu I. The Bss HII and Mlu I data are not shown, incomplete cleavage of a Sfi I and Nru I site is indicated in percent.

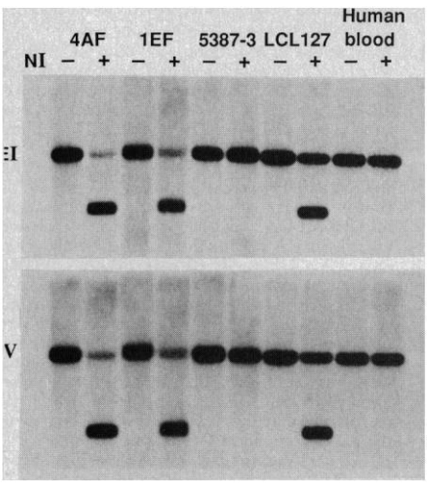

Fig. 4. Presence or absence of the Not I site in different cells. Lcn2 is bordered by a Not I site in 4AF, which is not cleavable in some other cell lines, as tested by single and double digests with two enzymes. LCL127 is a lymphoblastoid cell line; 5387-3 is a mouse cell line carrying human chromosome 7. EI, Eco RI; EV, Eco RV. The – and + refer to cleavages in the absence or presence of Not I (NI).

To allow the specific cloning of ends created by other enzymes, λ vectors in which Sac II, Mlu I or Bss HII, Sal I or Xho I can be used are available (Fig. 2, B and C). These vectors have been used successfully in the selective cloning of end fragments from mixtures. Further sites can be provided by the use of appropriate cosmid vectors.

Although the experiments described here take advantage of an interspecies cell hybrid to provide a second enrichment step, the use of alternative enrichment procedures, such as recleavage of enriched DNA followed by a second purification step on gel, or the use of sorted chromosomes (with care to avoid shearing) as starting material should be quite feasible.

Since the cloning of DNA fragments enriched by PFG techniques is independent of the availability of a starting point next to a rare cutting site, but will often result in such a clone, it is a valuable complement to chromosome jumping (*6*) and especially to rare cutter jumping libraries. Since preparative PFG requires a library construction for every cloning step, chromosome jumping will generally be considerably faster. On the other hand, the cloning of gel-enriched large DNA fragments should be less sensitive to fragment size, and allows parallel cloning of internal fragments. Further increase of the DNA size range accessible to PFG separation should further increase the attractiveness of this approach.

Collins *et al.* (*6*) have derived a clone 80 to 130 kbp 3′ to the *met*G sequence by chromosome jumping. Their PFG mapping data are consistent with the results described here.

REFERENCES AND NOTES

1. D. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984).
2. C. L. Smith and C. R. Cantor, *Methods Enzymol.*, in press; C. L. Smith, P. W. Warburton, A. Gaal, C. R. Cantor, in *Genetic Engineering*, J. K. Setlow and A. Hollaender, Eds. (Plenum, New York, 1986), vol. 8, pp. 45–70; G. F. Carle and M. V. Olson, *Nucleic Acids Res.* **12**, 5647 (1984).
3. G. F. Carle, M. Frank, M. V. Olson, *Science* **232**, 65 (1986).
4. M. Burmeister and H. Lehrach, *Nature (London)* **324**, 582 (1986).
5. D. P. Barlow, M. Bucan, H. Lehrach, B. L. M. Hogan, N. Gough, *EMBO J.* **6**, 617 (1987); B. G. Herrmann, D. P. Barlow, H. Lehrach, *Cell* **48**, 813 (1987).
6. A. Poustka, T. Pohl, D. P. Barlow, A.-M. Frischauf, H. Lehrach, *Nature (London)* **325**, 353 (1987); F. S. Collins *et al.*, *Science* **235**, 1046 (1987); A. Poustka and H. Lehrach, *Trends Genet.* **2**, 174 (1986); A. Poustka *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, in press; C. L. Smith *et al.*, *Methods Enzymol.*, in press; A.-M. Frischauf and T. Pohl, in preparation.
7. R. White *et al.*, *Nature (London)* **318**, 382 (1985); B. J. Wainwright *et al.*, *ibid.*, p. 384; L.-C. Tsui *et al.*, *Science* **230**, 1054 (1985); A. Beaudet *et al.*, *Am. J. Hum. Genet.* **39**, 681 (1986).
8. P. J. Scambler, H.-Y. Law, R. Williamson, C. S. Cooper, *Nucleic Acids Res.* **14**, 7159 (1986).

9. S. M. Arfin, R. E. Cirullo, Arredondo-Vega, F. X. M. Smith, *Somatic Cell Genet.* **9**, 517 (1983).
10. Agarose blocks containing 4AF cells (10^6 cells per 80-μl block) were prepared and digested with Not I as described (*5*). Eight solid agarose blocks were loaded on a 20 by 20 cm 1% low melting point agarose gel (BRL, Ultra Pure), in 0.2 TBE (18 m*M* tris-borate, 0.4 m*M* EDTA, *p*H 8.3) and placed in an electrophoresis chamber (*3*). Preparative FIGE was run for 24 hours at 200 volts, 15°C, 45 mA. The voltage gradient was inverted periodically, with a time ratio of 3/1 and the longer time increasing linearly from 3 to 120 seconds. After running it was stained for 1 hour in 10 m*M* EDTA+ ethidium bromide (1 μg/ml), and destained for 1 hour in 10 m*M* EDTA, and photographed under 360-nm ultraviolet light (Fig. 1). One lane of the gel was blotted onto GeneScreen membranes. The other lanes were transferred to a millimeter grid, and sliced manually in 2.5-mm slices perpendicular to the running direction. From these gel slices (approximately 3 g) a piece corresponding to half of a lane was cut off, and the rest was stored at 4°C in 1 ml of 0.5*M* EDTA (*p*H 8.0). Alternatively, to avoid a bias due to uneven distribution of DNA, slices can be melted at 65°C in 100 m*M* NaCl and a small volume removed. The small gel pieces were melted at 95°C, brought to 0.3–0.5*M* NaOH by adding 2*M* NaOH stock, and left at 65°C for 10 minutes. This step denatures the DNA and hydrolyzes the agarose. They were then filtered through GeneScreen transfer membrane by means of a commercial microtiter filtration manifold (Schleicher & Schuell) that had been prewarmed to 37°C. Southern blots and dot blots were hybridized (*5*) to *met*D to identify slices containing the *met* Not I fragment (Fig. 1). Slices 14 and 15 were brought to 100 m*M* NaCl, melted at 65°C, and incubated overnight with 50 units of agarase (Calbiochem, deoxyribonuclease-free) per milliliter. High molecular weight DNA was recovered by phenol extraction and ethanol precipitation.
11. The gel-eluted DNA (estimated 1 μg) was resuspended in 80 μl TE (10 m*M* tris, 1 m*M* EDTA) of which 5 aliquots (10 μl each) were digested in a volume of 30 μl with 0.05, 0.017, 0.005, 0.0017, 0.0005 unit of Sau 3AI (Biolabs) for 30 minutes at 37°C. The reaction was stopped by EDTA and heat inactivation, and 1 μl was electrophoresed on a 0.3% agarose gel, blotted, and hybridized to radioactively labeled hamster DNA in order to evaluate the extent of the partial digests. The remaining 30 μl of DNA were then digested under optimal conditions (0.007 unit, 30 minutes, 100 μl, 37°C), and all suitable fractions (10 to 30 kbp) were phenol-extracted and precipitated. This DNA (estimated at 0.5 μg) was ligated in one reaction to 7 μg of EMBL3*Sam*–Bam HI–Eco RI (*12*) and 0.7 μg of EMBL6–Bam HI–Eco RI–Not I (*13*), in a volume of 170 μl. The large molar excess of vector DNA ensured high cloning efficiency from small amounts of DNA, and reduces the probability of coligation of insert fragments.
12. A.-M. Frischauf, N. E. Murray, H. Lehrach, *Methods Enzymol.*, in press; A.-M. Frischauf, H. Lehrach, A. Poustka, N. Murray, *J. Mol. Biol.* **170**, 827 (1983).
13. λ EMBL5A is derived from EMBL3A by replacing both Bam HI cloning sites by a Not I linker (*12*). An Eco RI digest of EMBL5A and EMBL3S was mixed, ligated, and plated on *E. coli* MC1061, a *sup*^{0} (nonpermissive) host (*20*). One bacteriophage containing a Bam HI cloning site at the left arm and a Not I site at the right arm was identified (EMBL6).
14. B. Seed, *Nucleic Acids Res.* **11**, 2427 (1983).
15. D. L. Van Dyke, M. J. Worsham, L. J. Fisher, L. Weiss, *Hum. Genet.* **73**, 130 (1986).
16. Human clones were identified by screening plaques with a human Alu repeat probe, which recognizes 1 in 50 plaques in the total library and a total of 120 plaques in the end-clone library.
17. S. Arfin and J. J. Wasmuth, personal communication.
18. F. Michiels *et al.*, unpublished data.
19. Agarose blocks of cell line 4AF were prepared and digested as described (*5*). For FIGE, blocks were loaded on a 1% agarose gel and electrophoresed as described (*10*), but for 30 hours at 180 volts, 35 mA, at 15°C. OFAGE was as described (*5*).
20. N. E. Murray, in *Lambda II*, R. W. Hendrix, J. W. Roberts, F. W. Stahl, R. A. Weisberg, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983), pp. 395–432.
21. L. Dente, M. Sollazzo, C. Baldari, G. Cesarini, R. Cortese, in *DNA Cloning*, D. M. Glover, Ed. (IRL Press, McLean, VA, 1986), vol. 1, pp. 101–106.
22. We are grateful to A. Poustka and M. Bucan for PFG data of the *met* region, D. Barlow for help in PFG analysis techniques, A.-M. Frischauf for discussions and comments on this manuscript, and B. Williamson and colleagues for probes and discussions on cystic fibrosis. In addition we thank S. Arfin, J. Wasmuth, K.-H. Grzeschick, U. Francke, and B. Williamson for cell lines used in this study. Supported by DFG grant Le 292/2-1 and the Hereditary Disease Foundation. F.M. is supported by an EMBO postdoctoral fellowship. M.B. is an EMBL predoctoral fellow.

18 February 1987; accepted 10 April 1987

Protein-DNA Interactions in Vivo Upstream of a Cell Cycle–Regulated Human H4 Histone Gene

U. Pauli, S. Chrysogelos, G. Stein, J. Stein, H. Nick

Cell cycle–dependent histone genes are transcribed at a basal level throughout the cell cycle, with a three- to fivefold increase during early S phase. Protein-DNA interactions in the 5′ promoter region of a cell cycle–regulated human H4 histone gene have been analyzed at single-nucleotide resolution in vivo. This region contains two sites, with four potential protein-binding domains, at which the DNA is protected from reaction with dimethyl sulfate in cells and from digestion with deoxyribonuclease I in nuclei. These protein-DNA interactions persist during all phases of the cell cycle and dissociate with 0.16 to 0.2*M* sodium chloride.

The human histone genes constitute a multigene family organized as a series of clusters (*1, 2*). The expression of the cell cycle–dependent human histone genes is temporally and functionally coupled with DNA replication; control is mediated at both transcriptional and post-transcriptional levels (*3*). The cell cycle–dependent genes are transcribed throughout the cell cycle, but most interestingly exhibit a three- to fivefold increase in transcription during the initial 2 hours of S phase (*4, 5*). This increase is accompanied by modifications in chromatin structure (*6, 7*), including changes in deoxyribonuclease I (DNase I) and S1 nuclease–sensitive regions.

DNA-mediated transfection of human histone genes into cells of both primate and murine origin has demonstrated the existence of trans-acting regulatory factors (*8, 9*). For at least three cell cycle–dependent human histone genes, approximately 200 nucleotides of 5′ flanking sequence are sufficient to support specific initiation of transcription both in vitro (*10, 11*) and in vivo (*12*). In addition, more distal elements may influence the level of transcription (*12*). However, our understanding of proteins that interact with specific 5′ flanking sequences and influence the structure and transcription of human histone genes is still minimal.

To study in vivo protein-DNA interactions in the 5′ promoter of the cell cycle–regulated H4 histone gene, F0108, we used the genomic sequencing method of Church and Gilbert (*13*). Intact cells were treated with dimethyl sulfate (DMS), and the extent of methylation of each guanine residue was used to identify sites of specific protein-DNA interactions (*14*). DNA from nuclei treated with DNase I was also analyzed at single-nucleotide resolution (footprinting) (*15*), so that we were able to correlate data on chromatin structure with specific protein-DNA interactions.

We now report that the proximal 5′ promoter of this H4 gene contains two regions of DNA protected both from reaction with DMS in cells and from digestion with DNase I in nuclei. These protein-DNA interactions can be dissociated with 0.16 to 0.2*M* NaCl. Because these sites persist at all times during the cell cycle, they are thought to be important for at least the basal level transcription of this H4 histone gene.

The structure of the cell cycle–regulated F0108 H4 histone gene and the probes used for its analysis in vivo are shown (Fig. 1). The probes were designed to cover a region of 1000 nucleotides that includes the coding region. We initially focused on sites of protein-DNA interaction in vivo in the proximal 5′ flanking region of this gene. This region exhibits modifications in chromatin structure, which correlate with changes in the level of transcription (*6, 7*). In addition, deletion analysis of this region indicated that sequences between the translation start site and −120 bp were sufficient to support correct initiation of transcription (*12*).

Our initial studies in exponentially growing cells indicated strong protein-DNA contacts within the upstream proximal 200 nu-

College of Medicine, University of Florida, Gainesville, FL 32610.