Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 6697–6700, September 1988
Biochemistry

Genomic sequencing and *in vivo* footprinting of an expression-specific DNase I-hypersensitive site of avian vitellogenin II promoter reveal a demethylation of a mCpG and a change in specific interactions of proteins with DNA

(*in vivo* dimethyl sulfate titration/5-methylcytosine/gene regulation)

H. P. SALUZ, I. M. FEAVERS, J. JIRICNY, AND J. P. JOST

Friedrich Miescher Institut, P.O. Box 2543, CH-4002 Basel, Switzerland

Communicated by Diter von Wettstein, June 7, 1988 (received for review May 2, 1988)

ABSTRACT **Genomic sequencing was used to study the *in vivo* methylation pattern of two CpG sites in the promoter region of the avian vitellogenin gene. The CpG at position +10 was fully methylated in DNA isolated from tissues that do not express the gene but was unmethylated in the liver of mature hens and estradiol-treated roosters. In the latter tissue, this site became demethylated and DNase I hypersensitive after estradiol treatment. A second CpG (position −52) was unmethylated in all tissues examined. *In vivo* genomic footprinting with dimethyl sulfate revealed different patterns of DNA protection in silent and expressed genes. In rooster liver cells, at least 10 base pairs of DNA, including the methylated CpG, were protected by protein(s). Gel-shift assays indicated that a protein factor, present in rooster liver nuclear extract, bound at this site only when it was methylated. In hen liver cells, the same unmethylated CpG lies within a protected region of ≈20 base pairs. *In vitro* DNase I protection and gel-shift assays indicate that this sequence is bound by a protein, which binds both double- and single-stranded DNA. For the latter substrate, this factor was shown to bind solely the noncoding (i.e., mRNA-like) strand.**

Eukaryotic DNA contains 5-methylcytosine, which occurs almost exclusively in the dinucleotide CpG (1–3). Considerable evidence suggests that methylated cytosine residues in the regulatory region of both viral and eukaryotic genes reduce their transcription (2, 4, 5). In addition, DNA methylation influences chromatin structure *in vivo*, suggesting that methylation may regulate transcription by altering protein–DNA interactions (6). So far, however, the molecular mechanisms of the methylation-dependent regulation of gene expression are poorly understood. As a model system we chose the avian vitellogenin gene, which is expressed in the liver of egg-laying hens and estradiol-treated immature chicks (ref. 7 and the refs. therein). In this system we have now identified one CpG within the promoter region, the methylation state of which correlates with gene expression. We therefore decided to determine whether the modification of this specific site would prevent or alter the *in vivo* binding of specific proteins to this stretch of DNA. Such a project requires a method that can determine the methylation state *in vivo* of all CpGs in the regulatory sequence of interest, as well as an accurate procedure to detect protein–DNA interaction *in vivo*. To date only a few procedures have been used to study the interaction of proteins with DNA *in vivo* (8–14). Genomic sequencing introduced by Church and Gilbert (15) and optimized in our laboratory (16, 17) was combined with the reaction *in vivo* of unprotected guanosine with dimethyl sulfate (genomic footprinting). These methods allow us to study *in situ* with unprecedented accuracy the interactions of proteins with specific genes. In this paper we describe *in vivo* changes in the methylation pattern, as well as in the DNA–protein interactions in the promoter region of the vitellogenin gene, after activation of this gene by estradiol.

MATERIALS AND METHODS

All restriction enzymes were obtained from Boehringer Mannheim and Biofinex (Praroman, Switzerland). 3′-Deoxyadenosine 5′-[α-^{32}P]triphosphate (triethylammonium salt), 3000 Ci/mmol (1 Ci = 37 GBq) were purchased from Amersham. GeneScreen (NEF-872) membranes were from New England Nuclear.

Genomic Sequencing. The genomic sequencing procedure was done as described by Saluz and Jost (16, 17). A 181-base pair (bp) *Dra* I–*Eco*RI fragment (positions −252 to −72) cloned into M13 mp19 and mp18 was used as template for synthesizing the upper- and lower-strand probes, respectively.

***In Vivo* Genomic Footprinting.** Livers were perfused with a cold solution of 0.15 M NaCl, and the cell suspension was prepared by digesting the minced tissues with a solution of 1 mg of hyaluronidase and 0.5 mg of collagenase per ml of 0.15 M NaCl/20 mM Hepes, pH 8/1 mM $CaCl_2$/0.1 mM $ZnCl_2$/0.1 mM $CoCl_2$. After 40-min incubation at 37°C the cell suspension was filtered through a nylon grid and centrifuged at 1000 × *g* for 5 min. Then after two washes with cold 0.15 M NaCl/20 mM Hepes, pH 8.5/5 mM EDTA, cells were resuspended in Dulbecco's modified Eagle's medium at 1–2 × 10^8 cells per ml and used immediately for the *in vivo* footprinting with dimethyl sulfate. Incubation with 0.5–0.00005% dilutions of dimethyl sulfate was done at 20°C for 5 min, and the reaction was quenched with 10 vol of ice-cold 0.15 M NaCl/20 mM Hepes, pH 8/5 mM EDTA containing 1% bovine serum albumin and 100 mM mercaptoethanol. After two washes with the cold stop buffer, DNA was prepared as described (17). DNA was then digested with the restriction endonuclease *Dra* I, and the β eliminations were done with 1 M piperidine. The DNA fragments were separated, transferred to nylon membranes, and hybridized as described for genomic sequencing (17).

DNase I Protection and Gel-Shift Assay. A 5′-end-labeled synthetic DNA fragment (positions −12 to +61) was incubated in binding buffer (100 mM NaCl/10 mM Tris·HCl, pH 8/1 mM EDTA) with 2.5 μg of nonspecific competing *Escherichia coli* DNA and increasing concentrations of liver nuclear extracts or HeLa cell lysate that had been fractionated on

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: NHP1 and NHP4, nonhistone proteins type 1 and 4, respectively.

heparin-Sepharose (Pharmacia). After 30 min on ice the reaction mixtures were adjusted to 10 mM $MgCl_2$ and incubated with 1 μg of DNase I at 25°C for 1 min. The reactions were terminated by adding 90 μl of stop solution (0.2 M NaCl/30 mM EDTA/1% $NaDodSO_4$/100 μg of tRNA) followed by phenol extraction and ethanol precipitation. The resulting fragments were analyzed on denaturing 8% polyacrylamide gels.

Gel-shift assays were done as described (18).

RESULTS

Correlation of Gene Expression with the *in Vivo* Demethylation of a Single mCpG in the Promoter Region. There are two CpG sites close to the transcriptional start of the chicken vitellogenin gene, one at position +10 in the DNase I-hypersensitive site B2 and one at position −52 (Fig. 1). As neither of these CpGs lies within the recognition sequence of a methylation-sensitive restriction enzyme, genomic sequencing was used to determine the methylation state of cytosines in this region in the DNA isolated from liver, oviduct, and erythrocytes of chickens. Fig. 1 shows that the cytosine at position +10 (labeled e) is fully methylated in all tissues except the livers of hens and estradiol-treated roosters, where it was unmethylated on both strands. This CpG (e) was shown to be demethylated in a way similar to those CpGs situated in the estrogen-response element of the vitellogenin gene (17). The demethylation was also accompanied by the onset of vitellogenin mRNA synthesis, although in this case the methyl groups were removed from both strands on a similar time scale. Demethylation of this CpG, therefore, coincides with gene expression and appearance of the DNase I-hypersensitive site B2 (19, 20). The cytosine at position −52 was unmethylated in all studied tissues.

Correlation of Gene Expression with the Interaction *in Vivo* of Proteins with DNA in the DNase I-Hypersensitive Site B2 and Its Flanking Regions. As used in other laboratories (10, 11, 13), dimethyl sulfate at 0.5% reacted with guanosine residues to a different extent depending on the DNA strand and the state of gene expression. As shown in rooster liver cells (Fig. 2) where the gene is inactive, the guanosine residues were better protected on the upper than on the lower DNA strand, and for the active gene in hen liver cells the reverse situation was seen—i.e., the upper DNA strand reacted much more strongly with dimethyl sulfate than the lower strand at this dimethyl sulfate concentration. For this reason we titrated the cells with different concentrations of dimethyl sulfate; a decrease in dimethyl sulfate concentration by as much as 10,000-fold resulted in the selective protection of specific guanosines in the "over-sensitive" DNA strands. These results are not an experimental artifact from incomplete DNA transfer to the nylon membranes because several experiments were done and the same filters used for hybridization with the upper and lower DNA strand probes, respectively. Note that similar results were obtained with *in vitro* dimethyl sulfate protection experiments done with nonhistone protein type 1 (NHP1) and the estrogen-response element (21). A graphic interpretation of the results of Fig. 2 appears in Fig. 3. In the silent gene (rooster) the CpG at position 10 is protected by a footprint of at least 10 bp. In the active gene (hen), however, where the CpG is demethylated, there is a protected area of up to 20 bp. Interestingly, for both the active and inactive genes a certain periodicity occurs in the *in vivo* footprints. To determine whether the *in vivo* footprints represent binding of specific protein(s) to this DNA stretch, we conducted *in vitro* DNase I protection experiments and gel-shift assays. Double-stranded oligonucleotides (positions −12 to +61) were synthesized, and protein binding was examined by DNase I digestions. The results (Fig. 4) show

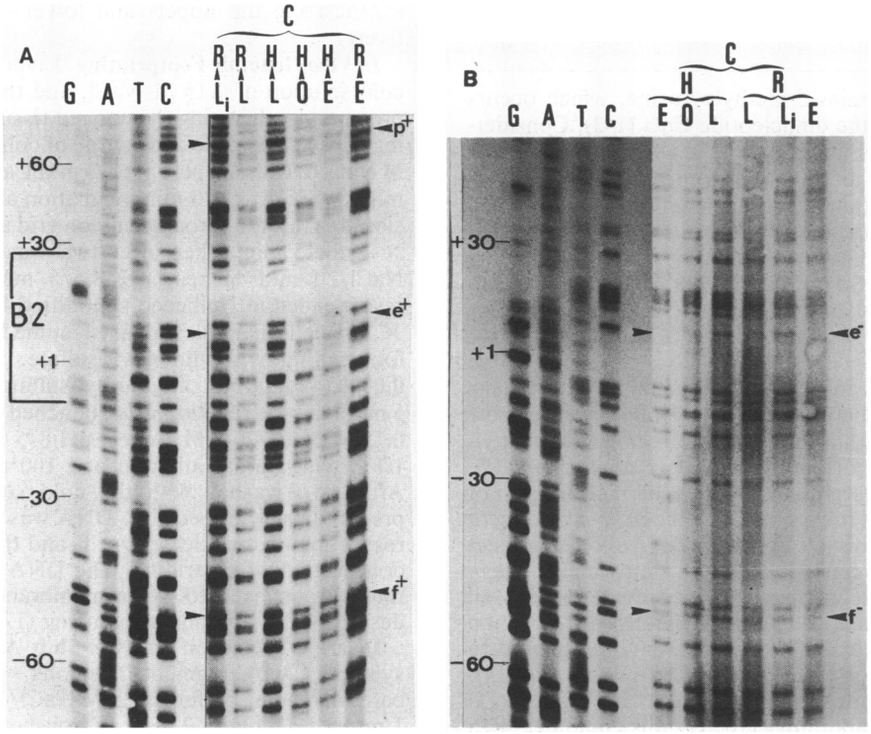

FIG. 1. Genomic sequencing of the promoter region of vitellogenin gene containing the methylation sites CpG e and f. All lanes of the experimental DNA represent the cytosine-specific reaction (C group) of the upper (*A*) and lower (*B*) DNA strand made from total DNA of estradiol-treated rooster liver (RL_i), rooster liver (RL), hen liver (HL), hen oviduct (HO), hen erythrocytes (HE), and rooster erythrocytes (RE). In *B*, H and R represent hen and rooster, respectively. Vertical bar B2, DNase I-hypersensitive site present only in hen liver or estradiol-treated immature chicken liver (19, 20).

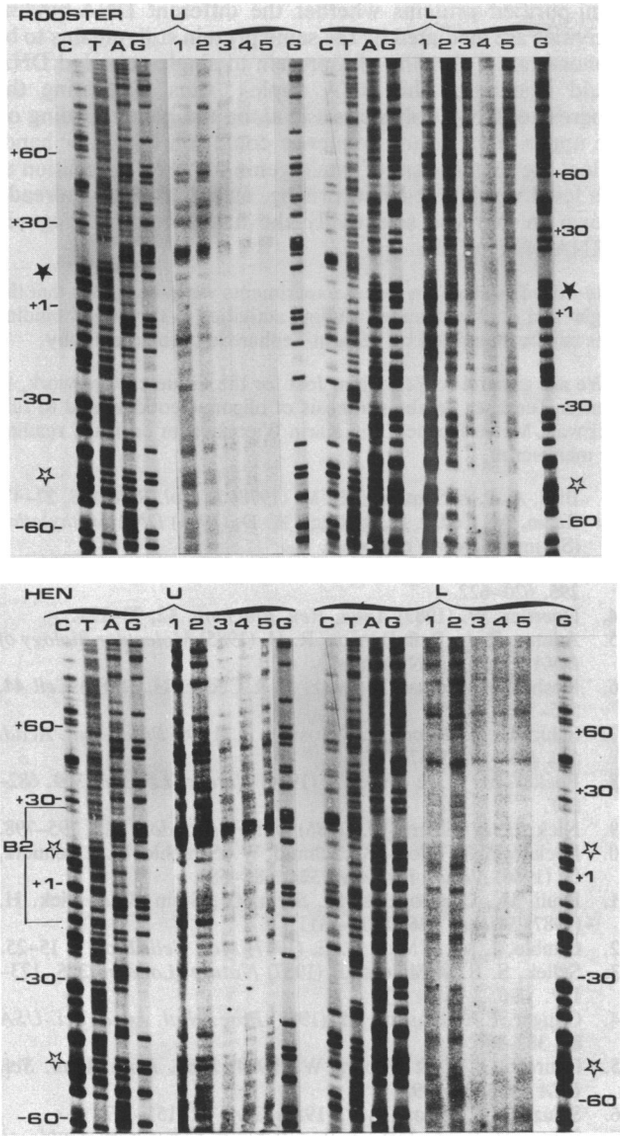

FIG. 2. Dimethyl sulfate footprints of upper (U) and lower (L) strands of rooster and hen liver cells. Control sequences (C, T, A, and G) are shown for orientation. A decreasing concentration of dimethyl sulfate (lanes 1–5: 0.5%, 0.05%, 0.005%, 0.0005%, and 0.00005%, respectively) was used to reveal details of the footprint. ★ and ☆, Methylated and demethylated cytosines, respectively. Position of the hormone and tissue-specific DNase hypersensitivity site B2 (19, 20) is shown.

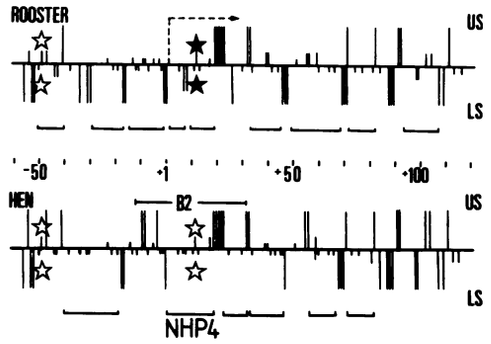

FIG. 3. Graphic representation of the results of Fig. 2. The relative extent of the reaction of dimethyl sulfate with guanosine residues was determined by trend analysis (17). Lines above each horizontal line represent the upper DNA strand (US) and lines below represent the lower DNA strand (LS). Length of the vertical lines indicates relative strength of the autoradiogram signal. ★ and ☆, Methylated and demethylated CpG, respectively. B2 is the tissue-specific and expression-specific DNase I-hypersensitive site.

that a protein factor was bound to the region of the CpG in the middle of the DNase I-hypersensitive site B2, giving a protected window between nucleotides +3 to +16 (Fig. 4). This latter footprint was obtained with both HeLa cell extracts and hen liver nuclear extracts and corresponds to the window obtained by treating hen liver cells with dimethyl sulfate *in vivo*. An oligonucleotide corresponding to the DNA between positions +3 and +16 was synthesized and used as a substrate for gel-shift assays (Fig. 4). The results confirmed that this sequence was bound by a protein factor nonhistone protein type 4 (NHP4) in a 2500-fold excess of nonspecific competing *E. coli* DNA. This protein is neither organ- nor species-specific, being found in the liver and oviduct of egg-laying hens and in human cell lines such as HeLa and MCF7 (Fig. 4, *Right Top* and *Middle*). Moreover, the ubiquitous nuclear NHP4 is more abundant in hen liver than in rooster liver (inactive gene) and binds to the upper, but does not bind to the lower DNA strand (Fig. 4, *Right Middle*).

We decided to investigate whether methylation of this site altered the protein–DNA interaction pattern *in vitro*. For this reason, we synthesized an oligonucleotide, corresponding to nucleotides +3 to +29, in which the CpG e was methylated. Upon fractionation of the rooster and hen liver nuclear extracts on heparin-Sepharose we found that the rooster liver extracts contained a protein eluting with 0.5 M KCl, which bound exclusively to the methylated oligonucleotide and not to the nonmethylated substrate (Fig. 4, *Right Lower*). This protein failed to bind the substrate corresponding to the NHP4 site +3 to +16. The results suggest that depending on the state of CpG methylation, the DNase hypersensitive site of the promoter binds different proteins.

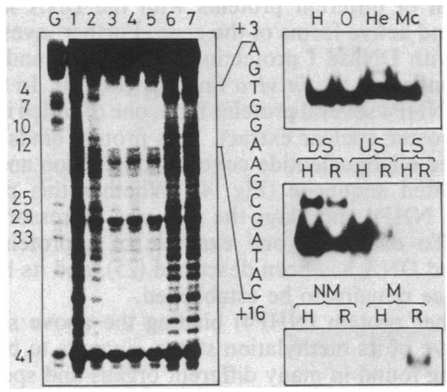

FIG. 4. DNase I protection in the region of the transcriptional start of the vitellogenin gene and gel-shift assays. (*Left*) DNase I protection experiments were done on the lower strand of oligonucleotide duplex (positions −12 to +61) using a 0.5 M heparin-Sepharose fraction of HeLa cell lysates. Lane G represents the guanosine-specific sequence reaction. DNase I digestion of the oligonucleotide without protein (lanes 1, 6, and 7) and with 6, 15, 30, and 60 μg of protein (lanes 2, 3, 4, and 5, respectively). The sequence protected by the protein on the CpG site is shown at the right. (*Right*) Gel-shift assay (only the protein–DNA complexes are shown). (*Top*) Complexes between labeled oligonucleotide (positions +3 to +16) with 5 μg of nuclear protein from hen liver (H), oviduct (O), HeLa cells (He), and MCF7 cells (Mc). (*Middle*) Complex formed between the same labeled oligonucleotide as above with hen liver (H) or rooster liver (R) nuclear extracts. DS, double-stranded DNA; US and LS, upper and lower single-stranded DNA, respectively. (*Bottom*) Complex between the labeled oligonucleotide duplexes (positions +3 to +29) with hen (H) or rooster liver (R) nuclear fraction (eluted from heparin-Sepharose with 0.5 M KCl). NM and M, nonmethylated and methylated oligonucleotides, respectively.

DISCUSSION

Recent work of several groups using a variety of experimental approaches has established that there is a negative correlation between DNA methylation and the expression of eukaryotic genes (2, 4, 6). In this study we investigated the possible mechanism for this effect using the promoter region of the avian vitellogenin gene as a model. Previous genomic sequencing data from our laboratory have shown that changes in the methylation pattern around the estradiol response element of this gene *in vivo* are estradiol dependent but not tissue specific (18). In the promoter region of the same gene, however, the CpG (e) was found to be unmethylated only in the tissue in which the gene was actively transcribed—i.e., liver. Moreover, the demethylation of this site was symmetrical and parallel to the onset of vitellogenin mRNA synthesis (H.P.S., data not shown). This demethylation could also be positively correlated with the appearance of the DNase I hypersensitivity site B2 in the promoter region of the gene in the DNA isolated from hens and immature chicks treated with estradiol (19, 20).

The sequence of the DNase I-hypersensitive site B2, presumably a region of active chromatin around the transcriptional start site, is highly conserved among the vitellogenin genes of chicken, *Xenopus*, and *Caenorhabditis elegans* (22–24) but is also present in other unrelated genes (data not shown). A precise study of changes in the pattern of protein–DNA interaction in the native chromatin of active and inactive vitellogenin gene was investigated by means of the *in vivo* genomic footprinting with dimethyl sulfate. Fig. 3 clearly shows that the DNA protection pattern in the chromatin exhibits a certain periodicity at the optimal concentration of dimethyl sulfate. In the DNase I-hypersensitive site B2 and the region surrounding CpG (e) there are marked differences between the pattern of protection found in hen and rooster liver cells. These differences could reflect the interaction of different proteins with the DNA in the repressed and active forms of the gene. Further investigations *in vitro* with DNase I protection experiments and gel-shift assays confirmed the *in vivo* findings (Fig. 4). In the same region as NHP4 several proteins bind, one of which is specific for the rooster nuclear extract. This protein binds only the methylated oligonucleotide covering this region and not the unmethylated sequence (Fig. 4). Whether this protein is related to NHP4 and plays the role of a repressor is unanswered. To date only one example of a protein-binding methylated DNA has been described (25), and its biological significance remains to be established.

The other protein (NHP4) binding the above sequence, irrespective of its methylation status, appears to be ubiquitous, being found in many different organs and species. An interesting feature of these protein(s) is the ability to bind both double- and single-stranded DNA. In the latter case, the protein binds only the mRNA-like DNA strand and not the complementary strand. It is worth noting that in the same nuclear extract we also found a trans-acting factor that binds double- and single-stranded GTCTTGTTCCAAAC sequence present in the third intron of the vitellogenin gene (18). Because in both cases the experiments were done with semi-purified proteins whether the different DNA-binding activities are mediated by the same protein still remains to be demonstrated. Binding of a protein to single-stranded DNA could destabilize the DNA duplex, thus facilitating the progress of RNA polymerase II along the gene. Binding on the upper DNA-strand sequence could, on the other hand, reflect the metabolism of vitellogenin mRNA. Regulation at this level would not be surprising, as estradiol was already shown to increase selectively the half life of vitellogenin mRNA *in vivo* (26).

Note Added in Proof. In recent experiments we have shown that the single- and double-stranded binding activities to the NHP4 binding sites can be separated by heparin-Sepharose chromatography.

We are grateful to Yan-Chim Jost for the tissue culture work, to Werner Zuercher for the synthesis of oligonucleotides, and to Iain McEwan, Melya Hughes, and Karin Wiebauer for critically reading the manuscript.

1. Bird, A. P. & Southern, E. M. (1978) *J. Mol. Biol.* **118,** 27–47.
2. Razin, A., Cedar, H. & Riggs, A. D. (1984) *DNA Methylation* (Springer, New York).
3. Gruenbaum, Y., Cedar, H. & Razin, A. (1982) *Nature (London)* **295,** 620–622.
4. Doerfler, W. (1983) *Annu. Rev. Biochem.* **52,** 93–124.
5. Adams, R. L. P. & Burdon, R. H. (1985) *Molecular Biology of DNA* (Springer, New York).
6. Keshet, I., Lieman-Hurwitz, J. & Cedar, H. (1986) *Cell* **44,** 535–543.
7. Saluz, H. P., Jiricny, J. & Jost, J. P. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7167–7171.
8. Becker, M. M. & Wang, J. (1984) *Nature (London)* **309,** 682–687.
9. Nick, H. & Gilbert, W. (1985) *Nature (London)* **313,** 795–798.
10. Becker, P. B., Gloss, B., Schmid, W., Straehle, U. & Schuetz, G. (1986) *Nature (London)* **324,** 686–688.
11. Pauli, U., Chrysogelos, S., Stein, G., Stein, J. & Nick, H. (1987) *Science* **236,** 1308–1311.
12. Gimble, J. M. & Max, E. E. (1987) *Mol. Cell. Biol.* **7,** 15–25.
13. Sellek, S. B. & Majors, J. (1987) *Nature (London)* **325,** 173–177.
14. Giniger, E. & Ptashne, M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 382–386.
15. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1991–1995.
16. Saluz, H. P. & Jost, J. P. (1986) *Gene* **42,** 151–157.
17. Saluz, H. P. & Jost, J. P. (1978) *A Laboratory Guide to Genomic Sequencing* (Birkhaeuser, Basel, Boston).
18. Jost, J. P., Saluz, H. P., Jiricny, J. & Moncharmont, B. (1987) *J. Cell. Biochem.* **35,** 69–82.
19. Burch, J. B. E. & Weintraub, H. (1983) *Cell* **33,** 65–76.
20. Burch, J. B. E. (1984) *Nucleic Acids Res.* **12,** 1117–1135.
21. Feavers, I. M., Jiricny, J., Moncharmont, B., Saluz, H. P. & Jost, J. P. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7453–7457.
22. Geiser, M., Mattaj, I. W., Wilks, A. F., Seldran, M. & Jost, J. P. (1983) *J. Biol. Chem.* **258,** 9024–9030.
23. Germond, J. E., Walker, P., Heggeler, B., Brown-Luedi, M., de Bony, E. & Wahli, W. (1984) *Nucleic Acids Res.* **12,** 8595–8609.
24. Spieth, J., Denison, K., Kirtland, S., Cone, J. & Blumenthal, T. (1985) *Nucleic Acids Res.* **13,** 5283–5295.
25. Wang, R. Y. H., Zhang, X. Y., Khan, R., Zhai, Y., Huang, L. H. & Ehrlich, M. (1986) *Nucleic Acids Res.* **14,** 9843–9860.
26. Brock, M. L. & Shapiro, D. J. (1983) *Cell* **34,** 207–214.