Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 271–275, January 1991
Biochemistry

Lack of correlation between DNA methylation and transcriptional inactivation: The chicken lysozyme gene

STEFAN WÖLFL, MAGDALENA SCHRÄDER, AND BURGHARDT WITTIG

Institut für Molekularbiologie und Biochemie, Freie Universität Berlin, Arnimallee 22, D-1000 Berlin 33, Federal Republic of Germany

Communicated by John M. Buchanan, October 1, 1990

ABSTRACT **We have analyzed the methylation state of all nine CpG sites in the transcription start region (−420 to +250 base pairs) of the chicken lysozyme gene by genomic sequencing. One of these sites, at −81, lies within the promoter, seven are clustered within the first exon, and the last is in the first intron. Five cell types and tissues have been investigated to study the relationship between methylation level and gene expression. For each cell type used, the majority of CpG sites showed a similar level of methylation. Of two gene-nonexpressing tissues, erythrocytes are hypomethylated, whereas liver is methylated at most of its CpG sites. For gene-expressing tissues, oviduct is completely unmethylated, whereas HD-11 culture cells are methylated. Thus no correlation is observed between degree of CpG methylation and level of expression of the lysozyme gene. The observed methylation patterns are discussed in terms of possible features of the local chromatin structure.**

Several studies have led to the generalization that eukaryotic genes are undermethylated in gene-expressing and methylated in gene-nonexpressing tissues. Cytosine methylation of various genes has been investigated in the transcription start region and in 3′-flanking regions (1–5). Most data were obtained by comparing cleavage patterns with pairs of methylation-sensitive and -insensitive restriction endonucleases, an approach that addresses only a subset of the CpG sites actually present. To investigate all methylated cytosine residues within a transcription start region we have directly sequenced this region in the genomic DNA of the chicken lysozyme gene (6, 7).

For this genomic sequencing, restricted genomic DNA is subjected to chemical sequencing reactions according to Maxam and Gilbert (8), electrophoresed on sequencing gels, and transferred to nylon membranes. The sequence of interest among the total population of DNA fragments is visualized by hybridization. A radioactively labeled probe complementary to one of the restricted ends of the target sequence is used. Methylated cytosine residues are not susceptible to the cytosine-specific cleavage reaction. Therefore, when a cytosine residue in the target sequence is methylated, the respective cytosine band is missing in the sequencing band pattern (6, 7).

The published sequence of a cloned fragment includes two CpG sites upstream and multiple sites downstream of the transcription start (9). A total of 10 sites are located between nucleotides −420 and +250, a region that includes two of the hormone-responsive elements, one of the DNase I-hypersensitive sites, several A + T-rich sequence motifs, and putative alternative start sites (9–11). In addition to this functionally important region, a number of cis-acting regulatory upstream elements are required for the tissue-specific expression of the lysozyme gene (for review, see refs. 12 and 13). We, nevertheless, focused our investigation on the transcription start region because it contains several functional elements that are also found in many other eukaryotic genes. The results described here may, therefore, be of general importance for understanding the relationship between gene expression and methylation in the transcription start region.

All known expression states of the chicken lysozyme gene are represented by the five cell types and tissues investigated here. The gene is transcribed under hormonal control in the oviduct of laying hens (14). In mature macrophages it is transcribed at a lower rate from the same promoter but independently of steroid hormones. A cell line derived from chicken macrophages (HD-11) expresses the lysozyme gene as in normal macrophages but at a lower level (15, 16). Liver does not express lysozyme. Nucleated erythrocytes of chicken are in a terminally differentiated state, in which most chromatin, including that of the lysozyme gene, is in a nontranscribable state (12).

MATERIALS AND METHODS

Tissues and Cells. Erythrocytes, liver, and tubular gland cells of oviduct were obtained from a laying hen. Macrophages were derived from bone-marrow cells of 3-month-old chickens [SPF (specific pathogen free) chickens VALO; Lohmann, Cuxhaven, F.R.G.], as described by Rossi and Himmelhoch (17). HD-11 is a MC29-transformed chicken cell line established by Beug and coworkers (15, 18, 19) and provided by W. Strätling (Universität Hamburg).

Genomic Sequencing. Genomic DNA was prepared as described by Saluz and Jost (7) and digested with *Pst* I or *Bst*NI to obtain suitable fragments for genomic sequencing. The genomic sequencing reactions and the plasmid sequencing reactions were also performed according to Saluz and Jost. After gel electrophoreses, DNA was transferred to nylon membranes by electro-blot (6). The gel was placed directly on a dry sponge-like material from air-condition filters (Bartel, Berlin). The prewetted GeneScreen membrane was placed on top of the gel, and the "sandwich" was topped with another dry sponge; it was slowly submerged in buffer avoiding air bubbles. Only GeneScreen (not GeneScreen-*Plus*) membranes were suitable. After transfer, DNA was immediately bound to the membrane by UV irradiation (254 nm) and subsequent heat treatment (80°C, 10 min). Conditions of UV irradiation are important for the generation of hybridization signals; therefore, optimal conditions have to be established for each batch of nylon membrane. Single-stranded DNA hybridization probes for either upper or lower strand were prepared from the lysozyme *Pst* I fragment cloned in M13mp18 and M13mp19, respectively (see Fig. 1) (7). Hybridization was performed in rotating cylinders (Bachofer, Reutlingen, F.R.G.) at 58°C, and the membranes were washed separately in Melamine trays at 48°C using the SDS/phosphate buffer system described by Church and Gilbert (6).

Densitometric Evaluation of Autoradiographs and Data Processing. Autoradiographs were analyzed by laser densitometry (model 300 A; Molecular Dynamics, Sunnyvale,

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.

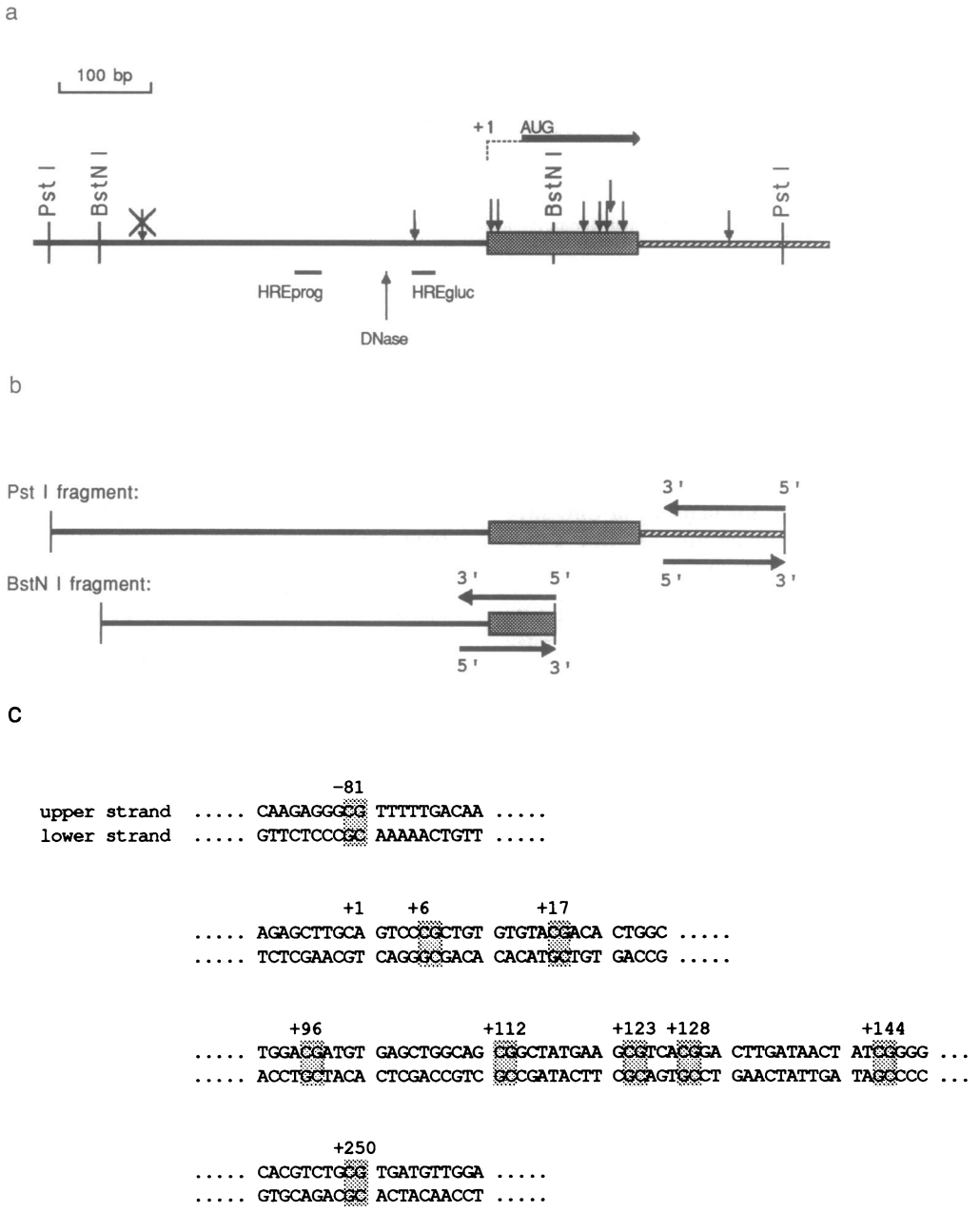

FIG. 1. Region of the chicken lysozyme gene containing regulatory sites, transcription start, and translation start (9, 20). (*a*) Fragment of ≈770 bp (−450 to +320); first exon (+1 to +165) is shown as a cross-hatched box; beginning of the first intron is shown as hatched line. Positions +1 and AUG (at +30) denote transcription and translation starts, respectively. Arrows pointing downward mark all CpG sites in this region (−81, +6, +17, +96, +112, +123, +128, +144, and +250). Single arrow pointing upward marks DNase I-hypersensitive site at −0.1 kb, only present in oviduct and macrophages (10). Arrow marked with X indicates a CpG site found in the published sequence but not in the genomic as well as in the plasmid DNA used here. Bars labeled HREprog and HREgluc represent hormone-responsive elements for glucocorticoid (−80 to −54) and progesterone receptors (−180), respectively (11). (*b*) Genomic sequencing was done on the *Pst* I and the *Bst*NI fragment of genomic DNA. Arrows show single-stranded M13 DNA hybridization probes used for indirect end-labeling of upper [⟶] and lower [⟵] strands, respectively. (*c*) Sequence details of the region analyzed. Shaded areas represent CpG sites. This sequence is found in the EMBL data base (release 23) (accession no. V00429).

CA). To overcome the problems of background and of varying loads of genomic DNA, we applied the following corrections: (*i*) Background correction: for each electrophoretic lane, background was measured individually in an area close to the region of interest but not containing bands; (*ii*) baseline used for integration: line extended from peak start to peak end; (*iii*) evaluation: the area under each peak was compared with that of the nearest neighboring peak; (*iv*) comparison: for inter-tissue comparison of individual CpG sites the highest peak area in each lane was set to 1, and the others were calculated relatively. In cases of discrepancies between densitometry and visual inspection, data were corrected according to visual impression.

RESULTS

Fig. 1 shows the region analyzed. The CpG site at position −355, present in the published plasmid sequence (9) as well as in the reference plasmid of our experiments, is not observed either in the genomic DNA of the chicken strain used here or in the HD-11 cell line (S.W., unpublished data) indicate an A/T pair at this position; strand orientation is not

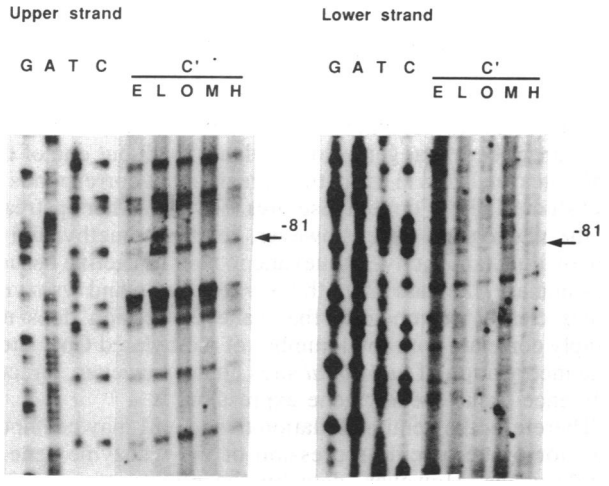

FIG. 2. Genomic sequencing autoradiograph from the chicken lysozyme *Bst*NI fragment displaying position −81. Lanes: G, A, T, and C, plasmid sequencing reactions of cloned chicken lysozyme *Bst*NI fragment; C′, genomic sequencing reactions using the cytosine-specific hydrazine reaction. Genomic DNA was from erythrocytes (lanes E), liver (lanes L), oviduct (lanes O), macrophages (lanes M), and HD-11 cells (lanes H). Each lane contains 25 μg of genomic DNA. Arrows indicate the cytosine positions of the CpG at position −81; numbering refers to transcription start site as +1. Areas below bands at −81 lane L (upper strand) and at about −68 lane M (lower strand) were manually altered to facilitate densitometric evaluation.

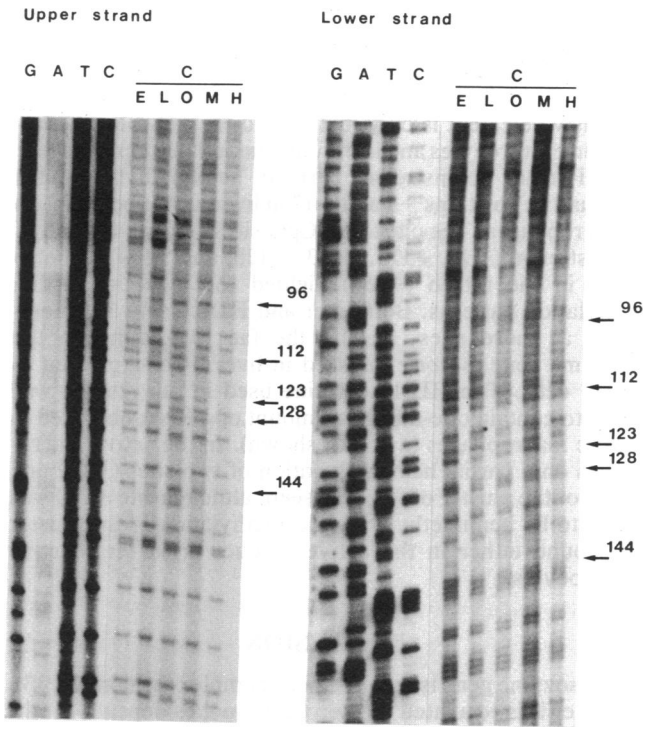

FIG. 3. Genomic sequencing autoradiograph from the chicken lysozyme *Pst* I fragment; missing bands at positions +96, +112, +123, +128, and +144 reflect methylation state of the respective cytosine residues. Lanes: G, A, T, and C, plasmid sequencing reactions of cloned chicken lysozyme *Pst* I fragment; C′, genomic sequencing reactions using the cytosine-specific hydrazine reaction. Genomic DNA was from erythrocytes (lanes E), liver (lanes L), oviduct (lanes O), macrophages (lanes M), and HD-11 cells (lanes H). Each lane contains 25 μg of genomic DNA. Arrows indicate cytosine positions of the respective CpG sites; numbering refers to transcription start site as +1.

known). A single CpG site at position −81 is thus the only site of the lysozyme gene promoter for methylation. This CpG site (−81) overlaps the glucocorticoid receptor-binding site (11) and is adjacent to the DNase I-hypersensitive site at −0.1 kilobases (kb) described for oviduct and macrophages (10). Two CpG sites are situated at positions +6 and +17 downstream from the transcription start, within the untranslated leader sequence. Five sites are clustered in the coding region near the end of the first exon (+96, +112, +123, +128, and +144). A single site is found at position +250 within the first intron. Nine sites in all have therefore been studied. Figs. 2 and 3 show examples of our genomic sequencing data. They display the CpG at position −81 (Fig. 2) and the region downstream of the transcription start where five CpG sites are clustered (Fig. 3). As examples, position +123 (Fig. 3) and position −81 (Fig. 2) of the lower strand are unmethylated in all tissues investigated. In the upper strand, site −81 (Fig. 2) is methylated in erythrocytes and unmethylated in the other tissues. The cytosine at position +128 of the upper strand (Fig. 3) is methylated in liver and HD-11 cells but unmethylated in erythrocytes, oviduct, and macrophages. Where differences in intensities seemed relevant, like those of the cytosine residue at position +112 in the macrophage DNA, bands were analyzed by laser densitometry (see *Materials and Methods*). Dark bands in all lanes of the upper strand at position −74 may reflect an unexplained hyperreactivity with respect to chemical sequencing. However, the additional band in lane O (oviduct) at position +148 of the upper strand may either be a mutation in the oviduct DNA or a site-specific cleavage not related to the sequencing reaction.

Table 1 summarizes the complex methylation state of the transcription start region for the five cell types and tissues investigated. A critical evaluation of visual impression and

Table 1. Methylation state of all CpG sites for the five cell types investigated

Tissue	Strand	Position of CpG site									Gene activity
		−81	+6	+17	+96	+112	+123	+128	+144	+250	
Erythrocyte	Upper	+	−	−	−	−	−	−	−	−	Inactive
	Lower	−	?	−	−	−	−	−	−	−	
Liver (hormone sensitive)	Upper	−	−	−	p	+	+	+	+	+	Inactive
	Lower	−	p	p	+	+	−	+	+	−	
Oviduct (hormone sensitive)	Upper	−	−	−	−	−	−	−	−	−	Active
	Lower	−	−	−	−	−	−	−	−	−	
Macrophage	Upper	−	−	−	−	p	p	−	−	+	Active
	Lower	−	p	−	−	−	−	−	−	−	
HD-11 cell	Upper	−	−	−	p	+	+	+	+	+	Active
	Lower	−	−	−	+	+	−	+	+	p	

Data were derived from autoradiographs as shown in Fig. 2 by laser densitometry (model 300 A; Molecular Dynamics, Sunnyvale CA). In cases of discrepancies between densitometry and visual inspection, data were corrected according to visual impression. −, unmethylated; p, partially methylated; +, methylated; ?, no decision. Numbering of CpG sites refers to transcription start as +1.

densitometric measurements led to the decision to reduce all data to the three statements: unmethylated (−), partially methylated (p), and methylated (+).

The cytosine at −81 that abuts the glucocorticoid hormone-responsive element is only methylated (in fact, hemimethylated) in erythrocytes and is the only methylated CpG in these cells. Further downstream, partially hemimethylated sites are found at positions +6 and +17 in liver and at position +6 in macrophages. The cluster of CpG sites close to the end of the first exon (+96, +112, +123, +128, and +144) and also a site in the intron (+250) yielded the most interesting methylation patterns. Both liver and HD-11 cells are methylated at these sites, despite the fact that although the lysozyme gene is *not* expressed in liver, it *is* expressed in HD-11 cells. The HD-11 cell line used in our studies was found to contain one-third the amount of mRNA observed in mature macrophages (data not shown). In the active macrophages only partial hemimethylation of two sites (+112 and +123) of the cluster of five was seen, although site +250 was found to be hemimethylated. No methylation of the cluster was found either in the active oviduct or in the inactive erythrocytes.

DISCUSSION

The lysozyme gene is regulated in a complex fashion (12, 13). Whether it is regulated by steroid hormones depends on the cell type: in mature oviduct, where the gene is active, as well as in liver, where it is inactive, both the hormone and the receptor proteins are available. In mature macrophages and several derived cell lines, such as HD-11, the lysozyme gene is expressed independently of the hormone. The patterns of DNase I-hypersensitive sites reflect more clearly its tissue-specific expression (10, 12). In mature oviduct four DNase I-hypersensitive sites have been noted at −0.1 kb, −1.9 kb, −2.4 kb, and −6.1 kb. The site at −1.9 kb disappears when steroid hormones are withdrawn. In liver only the site at −2.4 kb is present. In macrophages two of the same DNase I-hypersensitive sites are present (−0.1 kb, −6.1 kb), and two further sites have been found at different positions (−0.7 kb, −2.7 kb). Thus, only two DNase I-hypersensitive sites at −0.1 kb and −6.1 kb are present in all cases where the lysozyme gene is transcribed. Of these, only the −0.1-kb site lies in the region that encompasses the proximal promoter elements found in most genes. At this site methylation might directly interfere with the transcriptionally active RNA polymerase complex, and it is in this region that a number of genes have shown an inverse correlation of CpG methylation with gene expression.

It is for these reasons that we chose to study the transcription start region and also included the CpG cluster within the first exon. Although the latter is not a typical CpG "island" (2), it may be involved in regulation of the lysozyme gene.

When oviduct and liver are compared as expressing and nonexpressing tissues (21–23), one obtains the generally accepted indirect correlation between methylation and gene expression—i.e., methylated CpG residues indicate the inactive state (liver), and unmethylated CpGs indicate the active state (oviduct). When macrophages and erythrocytes are compared as hormone-insensitive tissues having an active and inactive lysozyme locus, respectively (10), the gene is seen to be undermethylated in both expression states. The DNA of mature macrophages expressing the gene even tends to be more methylated than that in erythrocytes. Moreover, the HD-11 cell line that expresses lysozyme is methylated to a similar extent to that in inactive liver. This last observation may be related to the general finding that cell lines have become more methylated upon passaging. Thus, the HD-11 cells may not be a reliable guide as to how methylation states relate to transcriptional activity (24, 25).

The most surprising result observed here is the virtual absence of methylation in erythrocytes, cells that are often taken as a standard tissue for inactive genes (10, 23). In contrast to the present data Saluz *et al.* (23, 26) have shown that in erythrocytes, where the chicken vitellogenin gene is not transcribed, that gene is methylated in all but one of the CpG sites analyzed in a functional region that corresponds to that studied here. Four of these sites located further upstream of the vitellogenin gene, however, remain unmethylated in hen oviduct (an inactive tissue) and in liver (an active tissue). In summary, the data for both the vitellogenin and lysozyme genes clearly demonstrate that gene expression does not simply correlate with total number of methylated CpG sites. The methylation of *individual* sites could, however, directly influence tissue-specific gene expression.

Therefore, the hemimethylation of site −81 may be important for tissue-specific repression of the lysozyme gene in erythrocytes. However, data on the effects of single-site methylation on transcription do not give a straightforward picture. For example, although a direct influence on binding of transcription factors CTF and SP1 by methylation was not seen *in vitro* (27–30), the methylation of a single CpG site in the promoter of the herpes simplex virus thymidine kinase gene was shown to reduce transcriptional activity by two orders of magnitude (31).

To what extent could overall chromatin structure influence the levels of methylation, bearing in mind our observation (32) that methylation affected transcription of DNA templates microinjected into mammalian cells only when the DNA was packaged into nucleosomes? A simple explanation would be that the compact chromatin structure of inactive genes leads to masking of most CpG sites, as observed for the lysozyme gene in erythrocytes. In active genes, such as lysozyme in oviduct and macrophages, the proteins of the transcriptional and regulatory machinery might shield *individual* CpG sites. Occurrence of hemimethylated sites, like the one at position −81 in erythrocytes, may reflect the proposed mechanism. Only *de novo* methylation can lead to a hemimethylated CpG site after replication (see also ref. 26). Chromatin structure allowing for *de novo* methylation of only one strand could be a possible explanation. This effect would result in reproducible, tissue-specific methylation states as described in this paper and by others (2, 3, 5, 25) but not in an overall correlation with the state of expression.

We are grateful to C. Crane-Robinson (Portsmouth Polytechnic) for a critical discussion and revision of our manuscript. We thank G. Rossi for the preparation of mature macrophages, W. Strätling for kindly providing the HD-11 cell line, G. Schütz for plasmid containing the lysozyme *Pst* I fragment, and H. P. Saluz for comments on the manuscript. Our work was supported by grants to B.W. from the following institutions: Deutsche Forschungsgemeinschaft, Bundesministerium für Forschung und Technologie (Genzentrum Berlin), and Freie Universität Berlin (Forschungsgebietsschwerpunkt "Signalstrukturen von Nukleinsäuren und Proteinen" and "Trude Goerke Stiftung für Krebsforschung"). Generous and unbureaucratic help came from the "Fonds der Chemischen Industrie in Kooperation mit dem Bundesministerium für Forschung und Technologie (BMFT)."

1. Razin, A. & Riggs, A. (1980) *Science* **210**, 604–610.
2. Bird, A. P. (1986) *Nature (London)* **321**, 209–213.
3. Becker, P. B., Ruppert, S. & Schütz, G. (1987) *Cell* **51**, 435–443.
4. Doerfler, W., Langner, K.-D., Knebel, D., Hoeveler, A., Müller, U., Lichtenberg, U., Weishaar, B. & Renz, D. (1988) in *Architecture of Eukaryotic Genes*, ed. Kahl, G. (VCH, Weinheim), pp. 409–417.
5. Cedar, H. (1988) *Cell* **53**, 3–4.
6. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
7. Saluz, H. P. & Jost, J. P. (1987) *A Laboratory Guide to Genomic Sequencing* (Birkhaeuser, Basel).

8. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
9. Grez, M., Land, H., Giesecke, K., Schütz, G., Jung, A. & Sippel, A. E. (1981) *Cell* **25,** 743–752.
10. Fritton, H. P., Igo-Kemenes, T., Nowock, J., Strech-Jurk, U., Theisen, M. & Sippel, A. E. (1984) *Nature (London)* **311,** 163–165.
11. von der Ahe, D., Renoir, J.-M., Buchou, T., Baulieu, E.-E. & Beato, M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 2817–2821.
12. Fritton, H. P., Jantzen, K., Igo-Kemenes, T., Nowock, J., Strech-Jurk, U., Theisen, M. & Sippel, A. E. (1988) in *Architecture of Eukaryotic Genes*, ed. Kahl, G. (VCH, Weinheim), pp. 333–353.
13. Sippel, A. E., Theisen, M., Borgmeyer, U., Strech-Jurk, U., Rupp, R. A. W., Püschel, A. W., Müller, A., Hecht, A., Stief, A. & Grussenmeyer, T. (1988) in *Architecture of Eukaryotic Genes*, ed. Kahl, G. (VCH, Weinheim, F.R.G.), pp. 355–369.
14. Palmiter, R. D. (1972) *J. Biol. Chem.* **247,** 6450–6461.
15. Hauser, H., Graf, T., Beug, H., Greiser-Wilke, I., Lindenmaier, W., Grez, M., Land, H., Giesecke, K. & Schütz, G. (1981) in *Haematology and Blood Transfusion*, eds. Neth, R., Gallo, R. C., Graf, T., Mannweiler, K. & Winkler, K. (Springer, Berlin), Vol. 26, pp. 175–178.
16. Steiner, C., Muller, M., Baniahmad, A. & Renkawitz, R. (1987) *Nucleic Acids Res.* **15,** 4163–4178.
17. Rossi, G. & Himmelhoch, S. (1985) *Eur. J. Cell Biol.* **38,** 280–291.
18. Beug, H., von Kirchbach, A., Döderlein, G., Conscience, J.-F. & Graf, T. (1979) *Cell* **18,** 375–390.
19. Leutz, A., Beug, H. & Graf, T. (1984) *EMBO J.* **3,** 3191–3197.
20. Jung, A., Sippel, A. E., Grez, M. & Schütz, G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5759–5763.
21. Dorbic, T. & Wittig, B. (1987) *EMBO J.* **6,** 2393–2399.
22. Dierich, A., Gaub, M.-P., LePennec, J.-P., Astinotti, D. & Chambon, P. (1987) *EMBO J.* **6,** 2305–2312.
23. Saluz, H. P., Feavers, I. M., Jiricny, J. & Jost, J. P. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6697–6700.
24. Kuhlmann, I., Achten, S., Rudolph, R. & Doerfler, W. (1982) *EMBO J.* **1,** 79–86.
25. Toth, M., Lichtenberg, U. & Doerfler, W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3728–3732.
26. Saluz, H. P., Jiricny, J. & Jost, J. P. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7167–7171.
27. Harrington, M. A., Jones, P. A., Imagawa, M. & Karin, M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2066–2070.
28. Holler, M., Westin, G., Jiricny, J. & Schaffner, W. (1988) *Genes Dev.* **2,** 1127–1135.
29. Dynan, W. S. (1989) *Trends Genet.* **5,** 35–36.
30. Ben-Hattar, J., Beard, P. & Jiricny, J. (1989) *Nucleic Acids Res.* **17,** 10179–10190.
31. Ben-Hattar, J. & Jiricny, J. (1988) *Gene* **65,** 219–227.
32. Buschhausen, G., Wittig, B., Grässmann, M. & Grässmann, A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1177–1181.