BIOCHEMISTRY AND BIOLOGY OF DNA METHYLATION

Proceedings of a Fogarty International Center Conference
held in Bethesda, Maryland
April 17–19, 1985

Editors

Giulio L. Cantoni
Laboratory of General
and Comparative Biochemistry
National Institute of Mental Health
Bethesda, Maryland

Aharon Razin
Department of
Cellular Biochemistry
The Hebrew University
Hadassah Medical School
Jerusalem, Israel

ALAN R. LISS, INC. • NEW YORK

Address all Inquiries to the Publisher
Alan R. Liss, Inc., 41 East 11th Street, New York, NY 10003

Printed in the United States of America

Library of Congress Cataloging in Publication Data

Main entry under title:

Biochemistry and biology of DNA methylation.

Includes index.
1. Deoxyribonucleic acid—Metabolism—Congresses.
2. Methylation—Congresses. I. Cantoni, Giulio L.
II. Razin, Aharon, 1935- . III. John E. Fogarty
International Center for Advanced Study in the Health
Sciences. [DNLM: 1. DNA—metabolism—congresses.
2. Methylation—congresses. [W1 PR668E v.198 /
QU 58 B6145 1985]
QP624.B56 1985 574.87'3282 85-14807
ISBN 0-8451-5048-0

Contributors

George Acs, Departments of Biochemistry and Pediatrics, Mount Sinai School of Medicine, New York, NY **[105]**

Nissim Benvenisty, Developmental Biochemistry Research Unit, Institute of Biochemistry, Hebrew University, Hadassah Medical School, Jerusalem, Israel **[185]**

Giorgio Bernardi, Laboratoire de Génétique Moléculaire, Institut Jacques Monod, Paris, France **[3]**

Timothy Bestor, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA **[95]**

Marcia A. Blackman, Department of Microbiology and Immunology, University of California, Berkeley, CA **[201]**

Thomas L.J. Boehm, Frankfurt Medical School, Center of Biological Chemistry, Frankfurt, Federal Republic of Germany **[271]**

Arthur H. Bolden, Department of Cell Biology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ **[79]**

Maria Jolanta Brzoska, Frankfurt Medical School, Center of Biological Chemistry, Frankfurt, Federal Republic of Germany **[67, 271]**

Giulio L. Cantoni, Laboratory of General and Comparative Biochemistry, National Institute of Mental Health, Bethesda, MD **[xv, 47]**

Doris Cassio, Centre de Génétique Moléculaire du C.N.R.S., Gif-Sur-Yvette, France **[177]**

Sara Chaffee, Department of Pediatrics, Duke University Medical Center, Durham, NC **[305]**

Judith K. Christman, Departments of Biochemistry and Pediatrics, Mount Sinai School of Medicine, New York, NY **[105]**

Pawel Dobrzanski, Institute of Genetics, University of Cologne, Cologne, Federal Republic of Germany **[133]**

Walter Doerfler, Institute of Genetics, University of Cologne, Cologne, Federal Republic of Germany **[133]**

Dusan Drahovsky, Frankfurt Medical School, Center of Biological Chemistry, Frankfurt, Federal Republic of Germany **[67, 271]**

Karen A. Dyer, Departments of Medicine and Genetics, and the Center for Inherited Diseases, University of Washington, Seattle, WA **[223]**

Melanie Ehrlich, Department of Biochemistry, Tulane Medical School, New Orleans, LA **[255]**

Andrew P. Feinberg, Oncology Center and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD **[279]**

The number in brackets is the opening page number of the contributor's article.

Ester Feldmesser, Department of Cellular Biochemistry, The Hebrew University, Hadassah Medical School, Jerusalem, Israel **[239]**

Stanley M. Gartler, Departments of Medicine and Genetics, and the Center for Inherited Diseases, University of Washington, Seattle, WA **[223]**

Walter Gilbert, The Biological Laboratories, Harvard University, Cambridge, MA **[313]**

Jennifer A.M. Graves, Department of Genetics and Human Variation, La Trobe University, Melbourne, Australia **[223]**

Michael L. Greenberg, Department of Medicine, Duke University Medical Center, Durham, NC **[305]**

F. Grosveld, Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London **[157]**

Stefan Grünwald, Frankfurt Medical School, Center of Biological Chemistry, Frankfurt, Federal Republic of Germany **[67]**

Joan Hare, Institute of Molecular Biophysics, Florida State University, Tallahassee, FL **[37]**

Janie J. Harrison, Division of Cancer Genetics, Dana-Farber Cancer Institute, Boston, MA **[293]**

Michael S. Hershfield, Department of Medicine, Duke University Medical Center, Durham, NC **[305]**

Hans-Peter Hirth, Frankfurt Medical School, Center of Biological Chemistry, Frankfurt, Federal Republic of Germany **[67]**

Vernon M. Ingram, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA **[95]**

Tal Kafri, Department of Cellular Biochemistry, The Hebrew University, Hadassah Medical School, Jerusalem, Israel **[239]**

Douglas H. Keith, Beckman Research Institute of the City of Hope, Division of Biology, Duarte, CA **[211]**

Dagmar Knebel, Institute of Genetics, University of Cologne, Cologne, Federal Republic of Germany **[133]**

Birgitt Knust-Kron, Institute of Genetics, University of Cologne, Cologne, Federal Republic of Germany **[133]**

Marian Elliott Koshland, Department of Microbiology and Immunology, University of California, Berkeley, CA **[201]**

Joanne Kurtzberg, Department of Pediatrics, Duke University Medical Center, Durham, NC **[305]**

Klaus-Dieter Langner, Institute of Genetics, University of Cologne, Cologne, Federal Republic of Germany **[133]**

Jacqueline Levilliers, Centre de Génétique Moléculaire du C.N.R.S., Gif-Sur-Yvette, France **[177]**

E. Murray, Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London **[157]**

Carlo M. Nalin, Department of Cell Biology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ **[79]**

Marie-Odile Ott, Department of Molecular Biology, Institut Pasteur, Paris, France **[177]**

Franco Palitti, Frankfurt Medical School, Center of Biological Chemistry, Frankfurt, Federal Republic of Germany **[67]**

Gerd P. Pfeifer, Frankfurt Medical School, Center of Biological Chemistry, Frankfurt, Federal Republic of Germany **[67]**

Aharon Razin, Department of Cellular Biochemistry, The Hebrew University, Hadassah Medical School, Jerusalem, Israel **[xv, 185, 239]**

Lea Reshef, Developmental Biochemistry Research Unit, Institute of Biochemistry, Hebrew University, Hadassah Medical School, Jerusalem, Israel **[185]**

Alexander Rich, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA **[23]**

Arthur D. Riggs, Beckman Research Institute of the City of Hope, Division of Biology, Duarte, CA **[211]**

Mariano Rocchi, Laboratorio Genetica Molecolare, Istituto Giannina Gaslini, Genova, Italy **[223]**

Ruth Sager, Division of Cancer Genetics, Dana-Farber Cancer Institute, Boston, MA **[293]**

Daniel V. Santi, Department of Biochemistry and Biophysics and Department of Pharmaceutical Chemistry, University of California, San Francisco, CA **[119]**

Natalie Schneiderman, Departments of Biochemistry and Pediatrics, Mount Sinai School of Medicine, New York, NY **[105]**

Carole H. Sellem, Centre de Génétique Moléculaire du C.N.R.S., Gif-Sur-Yvette, France **[177]**

Judith Singer-Sam, Beckman Research Institute of the City of Hope, Division of Biology, Duarte, CA **[211]**

Steven S. Smith, Beckman Research Institute of the City of Hope, Divisions of Biology and Surgery, Duarte, CA **[11]**

Linda Sperling, Centre de Génétique Moléculaire du C.N.R.S., Gif-Sur-Yvette, France **[177]**

Moshe Szyf, Department of Cellular Biochemistry, The Hebrew University, Hadassah Medical School, Jerusalem, Israel **[239]**

J. Herbert Taylor, Institute of Molecular Biophysics, Florida State University, Tallahassee, FL **[37]**

Mary Ellen Tolberg, Beckman Research Institute of the City of Hope, Divisions of Biology and Surgery, Duarte, CA **[11]**

Lily Vardimon, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA **[23]**

Martina Vogel, Frankfurt Medical School, Center of Biological Chemistry, Frankfurt, Federal Republic of Germany **[67]**

Richard Y.-H. Wang, Department of Biochemistry, Tulane Medical School, New Orleans, LA **[255]**

Cheryl A. Ward, Department of Cell Biology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ **[79]**

Mary C. Weiss, Department of Molecular Biology, Institut Pasteur, Paris, France **[177]**

Arthur Weissbach, Department of Cell Biology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ **[79]**

Andreas Werle, Frankfurt Medical School, Center of Biological Chemistry, Frankfurt, Federal Republic of Germany **[271]**

Ulrike Weyer, Institute of Genetics, University of Cologne, Cologne, Federal Republic of Germany **[133]**

John C. Wu, Department of Biochemistry and Biophysics and Department of Pharmaceutical Chemistry, University of California, San Francisco, CA **[119]**

Xian-Yang Zhang, Department of Biochemistry, Tulane Medical School, New Orleans, LA **[255]**

Preface

The first reports about the presence of methylated bases in DNA appeared almost 30 years ago, but the biological significance of these findings was not appreciated until much later.

In bacteria, the classical work of Meselson and Yuan established the role of DNA methylation in the modification and restriction phenomenon. More recent studies suggested that strand discrimination in post-replicative mismatch repair is also achieved by DNA methylation.

In vertebrates exciting progress has been made in our understanding of the biological significance of DNA methylation: it has been found that patterns of methylation of mammalian DNA are tissue specific and clonally inherited. Furthermore, it has been established that DNA methylation patterns are frequently correlated with gene expression. The hypothesis that gene activity is correlated with specific patterns of methylation is supported by the resulting experiments where the activity of a gene is modulated by the extent of its methylation: it was found that it is possible to inhibit gene activity by methylating a specific gene *in vitro*, while activation of specific genes can be achieved by inhibition of DNA methylation *in vivo*.

Much further work will be needed for full understanding of the biological function of DNA methylation: the enzymology of DNA methylation and its control is still inadequately understood and the mechanism of DNA demethylation is entirely unknown at the present time. A better understanding of the relationship between cell differentiation and DNA methylation may shed light on the possibility that malignant transformation, that is an abnormal form of differentiation, may somehow be related to changes in DNA methylation patterns.

This volume includes reports representing various aspects of research in the field of DNA methylation. The studies upon which these reports are based have been presented at a workshop on the Chemistry, Biochemistry and Biology of DNA Methylation, held on April 17–19, 1985, in Bethesda, Maryland, under the sponsorship of the Fogarty International Center.

The financial support of the Fogarty Center and the generous gifts by Biogen Research Corporation, Hoffman-La Roche Inc., and Smith, Kline and French Laboratories are gratefully acknowledged. Special thanks are also due to the staff of the Fogarty Center and especially to Mrs. Nancy Shapiro for the administrative assistance and expert help in the organization of the workshop.

Giulio L. Cantoni, M.D.
Aharon Razin, Ph.D.

Contents

Contributors xi

Preface
Giulio L. Cantoni and Aharon Razin xv

METHYLATION AND STRUCTURE OF EUKARYOTIC DNA

The Organization of the Vertebrate Genome and the Problem of the CpG Shortage
Giorgio Bernardi 3

Concerted Modification of Middle Repetitive Sequences
Steven S. Smith and Mary Ellen Tolberg 11

In Vitro Methylation of B-DNA and Z-DNA
Lily Vardimon and Alexander Rich 23

Methylation Directed Strand Discrimination in Mismatch Repair
Joan Hare and J. Herbert Taylor 37

ENZYMOLOGY AND REGULATION OF DNA METHYLATION IN EUKARYOTES

The Role of S-Adenosylhomocysteine in the Biological Utilization of S-Adenosylmethionine
Giulio L. Cantoni 47

Characterization of DNA Methylating Enzymes by Use of Monoclonal Antibodies
Dusan Drahovsky, Gerd P. Pfeifer, Stefan Grünwald, Hans-Peter Hirth, Martina Vogel, Maria Jolanta Brzoska, and Franco Palitti 67

The Effect of Flanking Sequences on the *De Novo* Methylation of C-G Pairs by the Human DNA Methylase
Arthur Weissbach, Carlo M. Nalin, Cheryl A. Ward, and Arthur H. Bolden. . . . 79

The DNA-Methylating System of Murine Erythroleukemia Cells
Timothy Bestor and Vernon M. Ingram 95

Interaction of DNA Methyltransferase and Other Non-Histone Proteins Isolated From Friend Erythroleukemia Cell Nuclei With 5-Azacytosine Residues in DNA
Judith K. Christman, Natalie Schneiderman, and George Acs 105

On the Mechanism and Inhibition of DNA Cytosine Methyltransferases
John C. Wu and Daniel V. Santi 119

DNA METHYLATION AND GENE ACTIVITY

Site-Specific Promoter Methylations and Gene Inactivation
Walter Doerfler, Klaus-Dieter Langner, Dagmar Knebel, Ulrike Weyer, Pawel Dobrzanski, and Birgitt Knust-Kron 133

Methylation and γ-Globin Expression
E. Murray and F. Grosveld 157

Relationship Between Expression of the Albumin Gene and Its State of Methylation
Mary C. Weiss, Carole H. Sellem, Marie-Odile Ott, Jacqueline Levilliers, Doris Cassio, and Linda Sperling 177

Developmental Changes in the Methylation Pattern, Chromatin Conformation and Expression of the Rat Phosphoenolpyruvate Carboxykinase Gene
Nissim Benvenisty, Aharon Razin, and Lea Reshef 185

Differentiation-Specific Methylation of the Immunoglobulin Heavy Chain Locus
Marcia A. Blackman and Marian Elliott Koshland 201

DNA METHYLATION AND X CHROMOSOME INACTIVATION

Methylation of the PGK Promoter Region and an Enhancer Way-Station Model for X-Chromosome Inactivation
Arthur D. Riggs, Judith Singer-Sam, and Douglas H. Keith 211

A Two Step Model for Mammalian X-Chromosome Inactivation
Stanley M. Gartler, Karen A. Dyer, Jennifer A. Marshall Graves, and Mariano Rocchi 223

DNA METHYLATION, DIFFERENTIATION AND TUMORIGENESIS

Cell Specific DNA Methylation Patterns; Formation and a Nucleosome Locking Model for Their Function
Aharon Razin, Ester Feldmesser, Tal Kafri, and Moshe Szyf 239

Human DNA Methylation: Methylated DNA-Binding Protein, Differentiation and Cancer
Melanie Ehrlich, Xian-Yang Zhang, and Richard Y.-H. Wang 255

Suppression of Malignant Phenotype, Changes in Gene Expression and DNA Hypermethylation in Human Melanoma Cells Treated With Arabinosyl Cytosine
Dusan Drahovsky, Maria Jolanta Brzoska, Andreas Werle, and Thomas L.J. Boehm 271

The Molecular Biology of Human Cancer
Andrew P. Feinberg 279

Adipocyte Differentiation in CHEF Cells
Ruth Sager and Janie J. Harrison 293

Altered Regulation of S-Adenosylmethionine Dependent Methylation in Adenosine Deaminase Deficiency
Michael S. Hershfield, Joanne Kurtzberg, Sara Chaffee, and Michael L. Greenberg 305

Concluding Remarks
Walter Gilbert 313

Index 317

METHYLATION AND STRUCTURE OF EUKARYOTIC DNA

Biochemistry and Biology of DNA Methylation, pages 3–10
© 1985 Alan R. Liss, Inc.

THE ORGANIZATION OF THE VERTEBRATE GENOME AND THE PROBLEM OF THE CpG SHORTAGE

Giorgio Bernardi

Laboratoire de Génétique Moléculaire,
Institut Jacques Monod
2, Place Jussieu, 75005 Paris, France.

Investigations carried out in our laboratory (see Bernardi et al., 1985 for a recent report) have shown that nuclear DNA from warm-blooded vertebrates can be separated, by using density gradient centrifugation in the presence of DNA ligands, into a small number of major components and into several satellite and minor components. The major components comprise : (i) two light components, L1 and L2, poorly or not resolved in some genomes; and (ii) two or three heavy components, H1, H2 and H3; the latter represent about one third of main-band DNA from warm-blooded vertebrates and account for its strong heterogeneity and CsCl band asymmetry. In contrast, main-band DNAs from most cold-blooded vertebrates show weak heterogeneities, only slightly skewed CsCl bands, and major components having buoyant densities that are only or mainly in the range of the light components of warm-blooded vertebrates. The families of DNA molecules (30-100 kb) forming the major components are derived, by the unavoidable breakage which accompanies DNA preparation, from corresponding families of much longer segments, the isochores, which have an average size well above 200 kb and a fairly homogeneous base composition.

The heavy DNA components of warm-blooded vertebrates : (i) have mainly arisen by regional increases in GC, which accompanied the transition from cold-blooded to warm-blooded vertebrates; these increases in GC seem to be mainly due to structural requirements at the DNA, chromatin and chromosome level; (ii) contain over half of the genes tested so far, a particularly high concentration being found in the heaviest components which only represent 4-8 % of nuclear DNA; and (iii) are responsible for an alternating pattern of heavy and light isochores which underlies the Giemsa and Reverse chromosomal patterns.

Major components can be used to study the genome distribution of any sequence which can be probed. This approach has revealed : (i) that the distribution of genes, of families of interspersed repeated sequences, and of integrated viral sequences is highly non-uniform in the genome of warm-blooded vertebrates; (ii) that the GC content and CpG/GpC ratio of both coding and non-coding sequences, as well as the GC levels of codon third positions show a direct linear relationship with the GC contents of the isochores harboring the sequences.

Here I would like to discuss in more detail the distribution of CpG doublets in the vertebrate genome and the problem of the CpG shortage. It is well known that vertebrate DNAs show a characteristic deficiency in the CpG doublet (Josse et al., 1961; Swartz et al., 1962). The C residue in this doublet is 55-90 % methylated in vertebrate DNAs (Van der Ploeg and Flavell, 1980; Gruenbaum et al., 1981; Kunnath and Locker, 1982).

CpG shortage and polypeptide-specifying DNA

The first explanation proposed to account for this outstanding feature of the vertebrate genome (Subak-Sharpe et al., 1966) was based on the observation that a CpG shortage is also exhibited by the genomes of small vertebrate viruses (like polyoma and SV40), which use essentially all of their genetic information for directing protein synthesis. It was proposed (Subak-Sharpe, 1967; 1974) that the genomes of these viruses derived from polypeptide-specifying DNA stretches of the ancestral host cells and therefore exhibited the same CpG shortage, which was visualized as reflecting constraints from the translation apparatus. In agreement with this proposal, CpG shortage was absent in genes not coding for polypeptides, like the tRNA, 5S RNA and rRNA genes. The lack of a CpG shortage in the genomes of intermediate and large vertebrate viruses (like adenovirus and herpes symplex, respectively) was attributed to their capacity of modifying the host translation apparatus either by virus-encoded tRNAs, or by modification of pre-existing cell-specified tRNAs through virus-encoded enzymes, or by differential stimulation of the host cell to produce those tRNAs needed to redress the balance. Since only a small proportion of vertebrate DNA is thought to be actively involved in protein coding and since the same CpG shortage was found in fractions of guinea pig DNA, as obtained by density gradient centrifugation, by differential renaturation, or from dispersed and condensed chromatin, it was further assumed that "the bulk of vertebrate DNA derives from and maintains the gross sequence characteristics of polypeptide-specifying DNA " (Russell et al., 1976).

Two subsequent observations from other laboratories did not support the explanation just outlined, namely (i) the presence of a large number of CpG doublets in rabbit (Salser, 1977) and human (Forget et al., 1979) α-globin genes; (ii) the uniform levels of CpG in both coding and non-coding sequences from the human α- and β-globin gene clusters (Lennon and Fraser, 1983). Since these observations concerned such a minute fraction of the genome, it could be argued, however, that the sequences studied were just exceptions to the rule.

Our investigations (Bernardi et al., 1985) have provided an unequivocal demonstration that CpG shortage is not dependent upon polypeptide coding. In fact (see Fig. 1), (i) CpG shortage usually is strong in the "light" genes from warm-

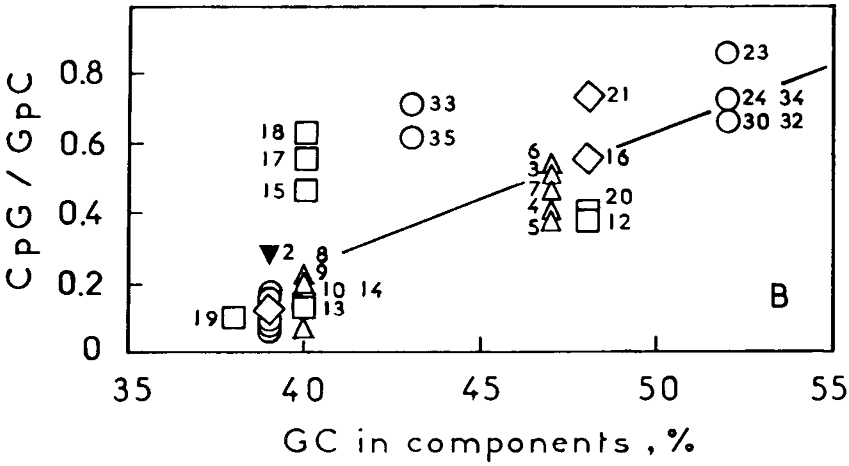

Fig. 1. Plot of CpG to GpC ratios for vertebrate genes and exons against the GC contents of DNA components harboring the genes. From Bernardi et al.(1985); the cluster of unnumbered genes comprises gene 22 and genes 25 to 29 (the key for the genes investigated can be found in the reference).

blooded vertebrates and, in all likelihood, in most genes from cold-blooded vertebrates, but becomes increasingly weaker and almost disappears in "heavy" genes from warm-blooded vertebrates; (ii) CpG shortage affects both coding and non-coding (intronic and intergenic) sequences from a given isochore family essentially to the same extent. In other words, CpG levels in both coding and non-coding sequences from warm-blooded vertebrate depend upon the GC levels of the corresponding isochores. This conclusion relies not only on the set of genes shown in Fig. 1, but also on a larger sample from data banks (not shown; Bernardi et al., 1985). In fact, our findings not only contradict the explanation proposed by Subak-Sharpe and co-workers for the CpG shortage, but also demonstrate that CpG shortage is not uniform throughout the genome of at least warm-blooded vertebrates in contrast with the conclusion of Russell et al. (1976). Finally, as far as the different CpG levels exhibited by the genomes of small and large vertebrate viruses are concerned, we have shown that these differences are not related to the size, but simply to the GC levels of the corresponding genomes. Indeed, a plot of CpG *versus* genome GC (not shown) exhibits a direct linear relationship, as previously found for vertebrate genes (Fig. 1).

CpG shortage and methylation

An alternative explanation (Salser, 1977) for the CpG shortage was that this doublet is highly methylated in vertebrates and that mCpG is a hot spot for mutation since it can be deaminated to TpG. Indeed, m5C residues in single stranded DNA are deaminated at 95°C at three times the rate of C residues and thermophylic bacteria with optimal growth temperatures equal to or higher than 60°C generally avoid having m5C in their genome (Ehrlich et al., 1985).

As far as this explanation is concerned, it is of interest to consider our observations : (i) that the transition from cold-blooded to warm-blooded vertebrates is accompanied by the appearance, or by a strong increase, of GC-rich segments in the genome (the "heavy" isochores) and of "heavy" genes; and (ii) that "heavy" isochores and "heavy genes" show a much decreased discrimination against the doublet CpG.

The latter point generalizes the previous observations that the rabbit and human α-globin genes show a much reduced CpG shortage. This situation was explained by assuming that in these cases CpG was not counterselected because required for mRNA structure and function (Salser, 1977). This explanation is, however, contradicted by the fact that CpG levels are comparable in both genes and pseudogenes from the human α-globin cluster (Lennon and Fraser, 1983) and, more generally, by our observation that the same relationship holds between CpG levels of