CRC REVIVALS

Molecular Biology of the Hepatitis B Virus

Edited by
Alan McLachlan

CRC Press
Taylor & Francis Group

Molecular Biology of the Hepatitis B Virus

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Molecular Biology of the Hepatitis B Virus

Editor
Alan McLachlan, Ph.D.
Associate Member
Division of Biochemistry
Department of Molecular and Experimental Medicine
The Scripps Research Institute
La Jolla, California

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

First published 1991 by CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Reissued 2018 by CRC Press

Library of Congress Cataloging-in-Publication Data

McLachlan, Alan
Molecular biology of the hepatitis B virus / Alan McLachlan
p. cm.
Includes bibilographical references and index.
ISBN 0-8493-5516-8
1. Hepatitis B virus. 2. Biology—molecular. I. McLachlan, Alan. II. Title.
[DNLM: 1. Hepatitis B virus. QW 710 G289h]
QR749 .H64G78
616' .0149—dc20
DNLM/DLC
for Library of Congress 91-32641

A Library of Congress record exists under LC control number: 91208209

Publisher's Note
The publisher has gone to great lengths to ensure the quality of this reprint but points out that some imperfections in the original copies may be apparent.

Disclaimer
The publisher has made every effort to trace copyright holders and welcomes correspondence from those they have been unable to contact.

ISBN 13: 978-1-315-89565-9 (hbk)
ISBN 13: 978-1-351-07475-9 (ebk)

Visit the Taylor & Francis Web site at http://www.taylorandfrancis.com and the
CRC Press Web site at http://www.crcpress.com

PREFACE

The principal aim of this volume, *Molecular Biology of the Hepatitis B Virus*, is to present a comprehensive and precise account of the current state of knowledge regarding the various molecular aspects of the life cycle of the hepatitis B virus (HBV). The areas of the molecular biology of HBV covered include the animal model systems, sequence data on the hepadnavirus genomes, the transcripts coded for by the viral genome and the sequence elements involved in regulating their expression, hepadnavirus replication, analysis of the various HBV gene products and their role in virion synthesis and assembly, a description of the consequences of long-term exposure to hepadnavirus infection and its association with hepatocellular carcinoma, the use of recombinant technologies in the generation of second generation vaccines, and the utilization of recombinant technologies to analyze an immune mediated disease. The volume, therefore, serves as a detailed source of information on the molecular aspects of hepadnavirus biology and contains only enough clinical and immunological data to place the molecular data in the appropriate context for an immunologically mediated disease.

THE EDITOR

Alan McLachlan, Ph.D., is an Associate Member in the Division of Biochemistry, Department of Molecular and Experimental Medicine at The Scripps Research Institute in La Jolla, California.

Dr. McLachlan received his B.Sc. and Ph.D. degrees in 1977 and 1980, respectively, from the Department of Biochemistry, University of Aberdeen, Aberdeen, Scotland. After doing postdoctoral work at Harvard University, Cambridge, Massachusetts, he was appointed an Assistant Member at the Research Institute of Scripps Clinic in 1986.

Dr. McLachlan is the recipient of research grants from the National Institutes of Health. He has published more than 50 papers. His current research interests are in gene regulation and the biogenesis of hepatitis B virus.

DEDICATION

This book is dedicated to my father and mother, John and Margaret McLachlan, for the sacrifices they made to guarantee that I was given the best possible educational opportunities.

ACKNOWLEDGMENTS

I am grateful to all of the authors for their contributions. Their efforts are greatly appreciated. In addition, I am especially grateful to Anneke Raney for her assistance with every stage and aspect of the generation of this book. I am indebted to Judith Preston and Lynn Keyes for their dedication in preparing the final manuscript.

CONTRIBUTORS

Ronald W. Ellis, Ph.D.
Senior Director and Head
Department of Cellular and
Molecular Biology
Merck Sharpe and Dohme Research Labs
West Point, Pennsylvania

Doris Fernholz, M.S.
Max-Planck-Institut für Biochemie
Martinsried, Germany

Markus Fischer
ZMBH University of Heidelberg
Heidelberg, Germany

Wolfram H. Gerlich, Ph.D.
Professor
Department of Medical Microbiology
Georg-August-University
Göttingen, Germany

Klaus-Hinrich Heermann, M.D.
Professor
Department of Medical Microbiology
Georg-August-University
Göttingen, Germany

Peter J. Kniskern, Ph.D.
Associate Director
Department of Cellular and
Molecular Biology
Merck Sharpe and Dohme Research Labs
West Point, Pennsylvania

Patricia L. Marion, Ph.D.
Senior Research Scientist
Department of Infectious Disease
Stanford University
Stanford, California

Kenichi Matsubara, Ph.D.
Institute for Molecular and Cellular Biology
Osaka University
Suita, Osaka, Japan

Alan McLachlan, Ph.D.
Associate Member
Department of Molecular
and Experimental Medicine
The Scripps Research Institute
La Jolla, California

David R. Milich, Ph.D.
Associate Member
Department of Molecular Biology
The Scripps Research Institute
La Jolla, California

Roger H. Miller, Ph.D.
Senior Staff Fellow
Hepatitis Viruses Section
National Institutes of Health
Bethesda, Maryland

Christine Pourcel, Ph.D.
Institut Pasteur
Paris, France

Anneke K. Raney
Research Assistant
Department of Molecular and
Experimental Medicine
The Scripps Research Institute
La Jolla, California

Heinz Schaller, Dr.rer.nat.
Professor
ZMBH University of Heidelberg
Heidelberg, Germany

Nancy Schek, Ph.D.
Professor
ZMBH University of Heidelberg
Heidelberg, Germany

H.-J. Schlicht, Ph.D.
Department of Virology
University of Ulm
Ulm, Germany

Ralf Schneider, M.S.
Max-Planck-Institut für Biochemie
Martinsried, Germany

Florian Schödel, M.D.
Max-Planck-Institut für Biochemie
Martinsried, Germany

Christoph Seeger, Ph.D.
Fox Chase Cancer Center
Philadelphia, Pennsylvania

Yosef Shaul, Ph.D.
Senior Scientist
Department of Molecular Genetics
and Virology
The Weizmann Institute of Science
Rehovot, Israel

Robert J. Shepherd, Ph.D.
Department of Plant Pathology
University of Kentucky
Lexington, Kentucky

Aleem Siddiqui, Ph.D.
Associate Professor
Department of Microbiology
University of Colorado Medical School
Denver, Colorado

Rolf Sprengel, Ph.D.
Max-Planck-Institut für Biochemie
Martinsried, Germany

David N. Standring, Ph.D.
Assistant Professor
Hormone Research Institute
University of California School of Medicine
San Francisco, California

Thomas Weimer, Ph.D.
Max-Planck-Institut für Biochemie
Martinsried, Germany

Gerhild Wildner, Ph.D.
Max-Planck-Institut für Biochemie
Martinsried, Germany

Hans Will, Ph.D.
Max-Planck-Institut für Biochemie
Martinsried, Germany

TABLE OF CONTENTS

Chapter 1
The Biology of Hepatitis B Virus 1
Anneke K. Raney and Alan McLachlan

Chapter 2
Ground Squirrel Hepatitis Virus 39
Patricia L. Marion

Chapter 3
The Biology of Avian Hepatitis B Viruses 53
Florian Schödel, Thomas Weimer, Doris Fernholz, Ralf Schneider, Rolf Sprengel, Gerhild Wildner, and Hans Will

Chapter 4
Hepatitis B Virus Transgenic Mouse Model 81
Christine Pourcel

Chapter 5
Transcription of Hepadnaviruses 95
Aleem Siddiqui

Chapter 6
Surface Proteins of Hepatitis B Viruses 109
K.-H. Heermann and W.H. Gerlich

Chapter 7
The Molecular Biology of the Hepatitis B Virus Core Protein 145
David N. Standring

Chapter 8
Biosynthesis and Enzymatic Functions of the Hepadnaviral Reverse Transcriptase 171
Hans-Jürgen Schlicht, Ralf Bartenschlager, and Heinz Schaller

Chapter 9
The Hepadnaviral X Protein 181
Nancy Schek, Markus Fischer, and Heinz Schaller

Chapter 10
Regulation of Hepadnavirus Transcription 193
Yosef Shaul

Chapter 11
Hepadnavirus Replication 213
Christoph Seeger

Chapter 12
Evolutionary Relationship Between Hepadnaviruses and Retroviruses 227
Roger H. Miller

Chapter 13
Chromosomal Changes Associated with Hepatitis B Virus DNA Integration and Hepatocarcinogenesis 245
Kenichi Matsubara

Chapter 14
The Molecular Biology of Cauliflower Mosaic Virus 263
Robert J. Shepherd

Chapter 15
Use of Recombinant HBV Proteins in Immunological Studies 283
David R. Milich

Chapter 16
Recombinant Hepatitis B Vaccines 307
Ronald W. Ellis and Peter J. Kniskern

Index 323

Chapter 1

THE BIOLOGY OF HEPATITIS B VIRUS

Anneke K. Raney and Alan McLachlan

TABLE OF CONTENTS

I. Introduction 2

II. The History of the Discovery, Isolation, and Characterization of Hepatitis B Virus 2

III. Organization of Hepatitis B Virus 14
- A. Viral Coding Capacity 14
- B. Structure of the HBV Particle 15

IV. Intracellular Life Cycle of HBV in the Human Hepatocyte 17
- A. Virus Entry into Hepatocytes 17
- B. Generation of a Transcriptionally Active HBV Template 17
- C. Transcription of the HBV Genome 18
- D. Expression and Secretion of the HBV Surface and Nucleocapsid Polypeptides 19
- E. Replication of HBV 20
- F. Maintenance of the Nuclear Covalently Closed Circular HBV DNA Pool 21

V. HBV Infection 22
- A. Host Range and Tissue Tropism 22
- B. Routes of Infection 23
- C. Clinical Consequences of Infection 24

VI. HBV-Related Infectious Agents 25

VII. Summary and Conclusions 25

Acknowledgments 26

References 26

I. INTRODUCTION

Hepatitis B virus (HBV) is the prototype member of a family of DNA viruses that primarily infect the liver and share a similar viral morphology and cellular life cycle.[1] The other identified and characterized members of this family are the woodchuck hepatitis virus (WHV),[2] the ground squirrel hepatitis virus (GSHV),[3] the duck hepatitis B virus (DHBV),[4] and the heron hepatitis B virus (HHBV).[5] In addition to these five viruses, the tree squirrel hepatitis B virus (TSHV) possibly represents an additional member of this virus family.[6] As a consequence of the unique features of this group of viruses, they have been classified as a separate family of viruses known as the hepadnaviruses.[1] The name *hepadna*virus reflects the *hepa*totropism of these *DNA* viruses.

II. THE HISTORY OF THE DISCOVERY, ISOLATION, AND CHARACTERIZATION OF HEPATITIS B VIRUS

The path that ultimately led to the discovery of one of the viruses, hepatitis B virus (HBV), responsible for parenterally transmitted hepatitis (serum or type B hepatitis), began in 1965 with the observation by Dr. Baruch Blumberg and colleagues of a precipitin reaction between the sera from an Australian aborigine and a frequently transfused hemophilia patient from New York city.[7] The lipoprotein present in the serum of the aborigine responsible for the precipitin reaction was called "Australia antigen." The subsequent observation that Australia (Au) antigen, or serum hepatitis (SH) antigen as it was also named,[8] was present at a much higher frequency in the sera of acute and chronic hepatitis patients than in control subjects led to the hypothesis that this antigen may be associated with an infectious agent responsible for "viral hepatitis."[9-15] With a view to testing this idea, electron microscopic analysis of Au antigen-positive sera revealed particles that reacted with antibodies against Au antigen.[16] These particles were predominantly spheres and filaments approximately 22 nm in diameter (Figure 1). The length of the filaments varied from less than 50 nm up to 1000 nm.[16-19] In addition to these forms of Au antigen, a larger particle, the Dane particle, which was much less abundant than the smaller particles,[20] was found subsequently in the sera of serum hepatitis patients.[17,21,22] This particle is 42 nm in diameter and comprises a 28-nm diameter inner body, the nucleocapsid or core, surrounded by a 7-nm outer coat (Figure 1). Since the 22-nm spheres and filaments aggregated with the larger particles in the presence of antibodies against Au antigen, it was suggested that all of these particles shared a common surface or envelope antigen,[17] subsequently called hepatitis B surface antigen (HBsAg).[23]

The identification of the various particulate forms in the sera of hepatitis patients did not resolve which, if any of these particles, represented the infectious agent. On the basis of morphology, it was suggested that the 42-nm Dane particle represented the agent responsible for serum hepatitis and the 22-nm spheres and filaments represented excess virus coat material.[17] Support for this suggestion came from the observation that the 22-nm spheres appeared to lack nucleic acid.[24] Further analysis of Dane particle structure was achieved by detergent treatment that released the inner body, the nucleocapsid, as a spherical 28-nm component that can form aggregates in the presence of posthepatitis but not prehepatitis sera.[25] This represented the identification of an additional antigen–antibody system specific for Au antigen-positive hepatitis. It also permitted further physical and biochemical characterization of the various viral and subviral components present in the sera of serum hepatitis patients.

The physical characterization of the 22-nm spheres demonstrated that these particles had an estimated buoyant density in cesium chloride (CsCl) of 1.18 to 1.22 g/cm^3 and a sedimentation coefficient in the range from 40 to 54.[24,26-28] From further analysis, the approximate

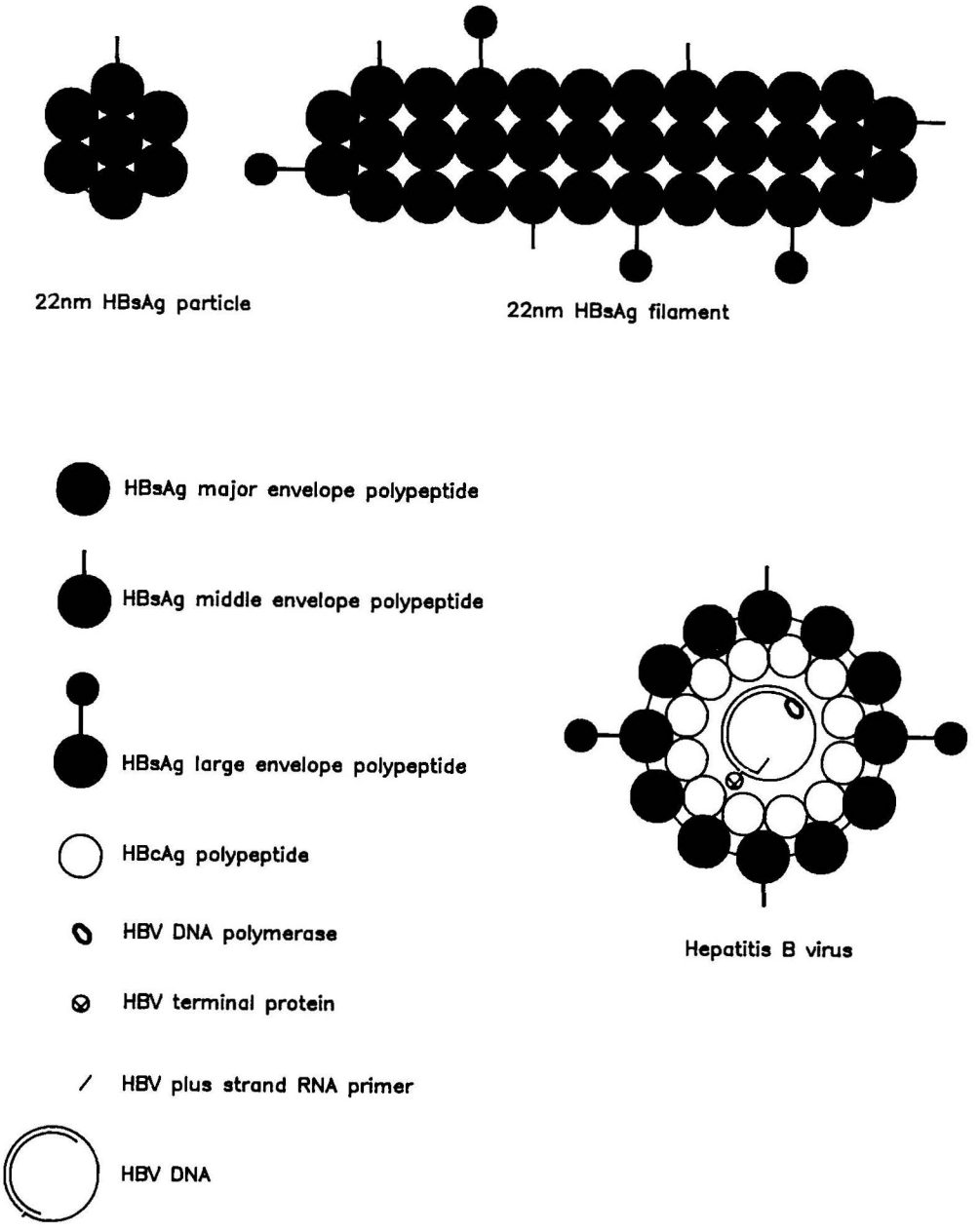

FIGURE 1. Diagrammatic representation of the structure of the 22-nm hepatitis B surface antigen (HBsAg) sphere, 22-nm HBsAg filament, and hepatitis B virus (Dane particle). HBcAg, hepatitis B core antigen.

molecular weight of 2 to 4 $\times$ 10^6 was estimated for the 22-nm sphere.[26,28] The density of the 22-nm filaments is similar to that of the 22-nm spheres.[19,28] The 28-nm nucleocapsid of the Dane particle was shown to have a density in CsCl of 1.30 to 1.36 g/cm^3 [29-33] and a sedimentation coefficient of approximately 110.[34] The density of the Dane particle, 1.24 to 1.27 g/cm^3,[31,33] was found to be intermediate between those of the 22-nm spheres and the 28-nm inner body, as might be expected based on its composition. Dane particles and nucleocapsids have

been shown to exist as two subpopulations with different densities.[31] The less dense populations lack the HBV DNA and represent defective particles, whereas the more dense populations contain HBV DNA.[31]

Characterization of the physical and biochemical properties of the various viral and subviral particles demonstrated that there was a DNA polymerase activity and endogenous primer–template complex tightly associated with the nucleocapsid of the Dane particle.[34,35] The DNA polymerase activity was not associated with the 22-nm surface antigen particles.[34] The DNA template that represented the substrate for this polymerase was isolated from the nucleocapsid of the Dane particle and by electron microscopy was shown to be a circular molecule of 0.78 μm.[36] A double-stranded DNA molecule of this length has a molecular weight of 1.6×10^6, which corresponds to approximately 2450 nucleotide pairs.[36,37] This was the first evidence that the genome of HBV was very small. Restriction enzyme and electron microscopic analyses of the HBV genome before and after the Dane particle DNA polymerase reaction demonstrated that the majority of circular DNA molecules within the virion possessed a single-stranded region encompassing 15 to 50% of the length of the genome.[38-40] The endogenous viral DNA polymerase reaction is responsible for the conversion of this single-stranded region to double-stranded DNA.[38-40] Based on these studies,[38-40] the size of the HBV genome after modification by the endogenous DNA polymerase reaction was reestimated to be approximately 3200 nucleotide pairs. Therefore, the HBV genome consists of a long strand of 3200 nucleotides and a short strand of variable length. Subsequent analysis demonstrated that the long strand of the viral DNA also contained a nick or short gap.[41] In addition, it was shown that this discontinuity in the long strand and the nick or short gap remaining in the short strand after completion of the endogenous DNA polymerase reaction are located at unique positions approximately 226 nucleotides apart in the HBV DNA.[41-43] Hence, the circularity of the genome is maintained by the approximately 226-nucleotide 5′-terminal cohesive overlap between the ends of the long and short DNA strands[41] (Figure 1). Additional studies of the HBV genome demonstrated that there is a protein, the terminal protein, covalently attached to the 5′ terminus of the long strand.[44] There is probably also a short oligoribonucleotide attached to the 5′ end of the short strand of the HBV genome in the virion,[43] as observed in DHBV,[45] GSHV,[46] and WHV.[46]

The inability to infect permanent cell lines in culture or standard laboratory animals with HBV restricted the source of virus to patients' sera. This limited the amount of HBV DNA that was available for study. However, the detailed analysis of the structure of the viral DNA permitted the cloning[47,48] and subsequent sequencing of the complete HBV genome.[49-61] The HBV genome sizes varied between 3182 and 3221 nucleotides (Figure 2). As a consequence of cloning the HBV genome, large amounts of pure HBV DNA of defined sequence could be obtained, analyzed, and manipulated. In addition, the infectivity and functionality of cloned DNA was tested by inoculating the complete HBV genome into the livers of chimpanzees.[62,63] The observation of a typical, self-limited, acute hepatitis with hepatitis B surface antigenemia indicated that a nondefective HBV genome had been introduced into the livers of the chimpanzees, and therefore the cloned DNA encoded all of the essential information to complete the viral life cycle.[62,63] This critical experiment verified that the 3200-nucleotide, partially single-stranded DNA present in the Dane particle represents the HBV genome and that the Dane particle is almost certainly the infectious agent responsible for serum or type B hepatitis. Further confirmation of the biological activity of cloned HBV came from transfection experiments where the complete HBV genome was introduced into various hepatoma cell lines or transgenic mice and Dane particles were subsequently produced.[33,64-69] In two cases, the Dane particles secreted by the cell lines were shown to be infectious in chimpanzees.[70,71]

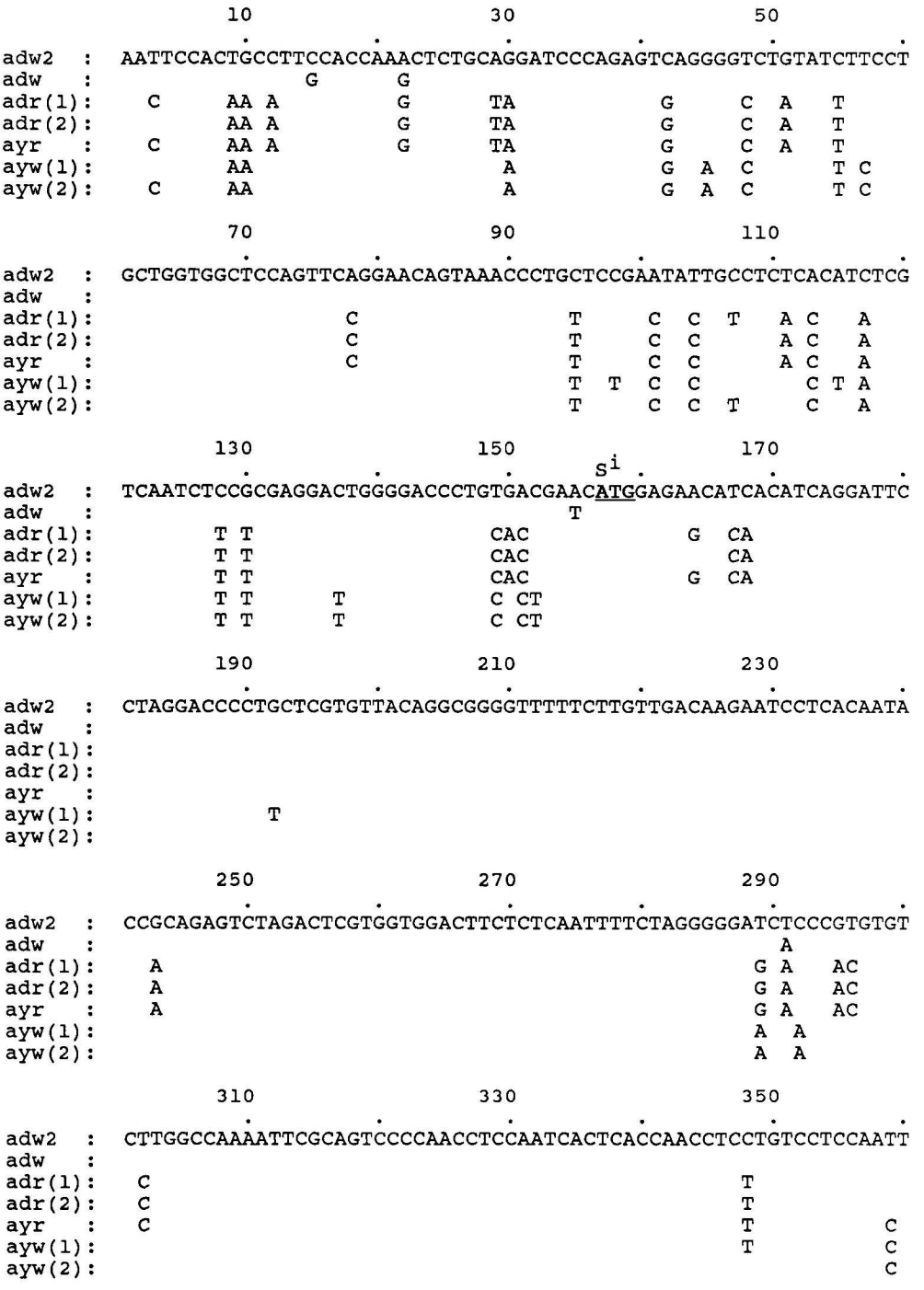

FIGURE 2. Comparison of the HBV DNA sequences of the major subtypes. The HBV genomes are subtypes *adw*$_2$,[52] *adw*,[54] *adr* (sequences 1[61] and 2[54]), *ayr*,[60] and *ayw* (sequences 1[50] and 2[56]). The sequences were aligned with the *adw*$_2$ sequence using the first A residue of the *Eco*RI site as nucleotide 1. The differences between the *adw*$_2$ sequence and the other sequences are indicated. Dots indicate the location of gaps necessary to permit maximum alignment of the nucleotide sequences. PS1^{i}a and PS1^{i}b, initiation codons for the 119- and 108-amino acid preS1 regions, respectively; PS2^i, initiation codon for preS2 region; S^i and S^t, initiation and termination codon for the major HBsAg open reading frame; PC^i, initiation codon for the precore region; C^i and C^t, initiation and termination codon for the core open reading frame; X^i and X^t, initiation and termination codon for the X gene open reading frame; P^i and P^t, initiation and termination codon for the polymerase open reading frame.

```
                370                 390                 410
                 .                   .                   .
adw2  :  TGTCCTGGTTATCGCTGGATGTGTCTGCGGCGTTTTATCATATTCCTCTTCATCCTGCTG
adw   :
adr(1):
adr(2):          C
ayr   :          C
ayw(1):                                          C
ayw(2):                                          C

                430                 450                 470
                 .                   .                   .
adw2  :  CTATGCCTCATCTTCTTATTGGTTCTTCTGGATTATCAAGGTATGTTGCCCGTTTGTCCT
adw   :
adr(1):                  G                C  C
adr(2):                  G                C  C
ayr   :                  G                C  C
ayw(1):                  G                C
ayw(2):                  G                C

                490                 510                 530
                 .                   .                   .
adw2  :  CTAATTCCAGGATCAACAACAACCAGTACGGGACCATGCAAAACCTGCACGACTCCTGCT
adw   :
adr(1):     C        A  T    T     C              G          T
adr(2):     C        A  T    C     C     G        G          T
ayr   :     C        A  T    T     C              G          T
ayw(1):                CT         C             CGG       T    A
ayw(2):                TT    T    C              G

                550                 570                 590
                 .                   .                   .
adw2  :  CAAGGCAACTCTATGTTTCCCTCATGTTGCTGTACAAAACCTACGGATGGAAATTGCACC
adw   :              A
adr(1):       A C                T                   T    C     C     T
adr(2):       A C                T                   T    C     C     T
ayr   :       A C                T                   T    C     C     T
ayw(1):       A C        A       C          C        T    C
ayw(2):       A C        A       C          C        T    C

                610                 630                 650
                 .                   .                   .
adw2  :  TGTATTCCCATCCCATCGTCCTGGGCTTTCGCAAAATACCTATGGGAGTGGGCCTCAGTC
adw   :
adr(1):                  A               G  T
adr(2):                  A               G  T
ayr   :                  A               G  T
ayw(1):                  A            G     T                      C
ayw(2):                  A            G     T                      C

                670                 690                 710
                 .                   .                   .
adw2  :  CGTTTCTCTTGGCTCAGTTTACTAGTGCCATTTGTTCAGTGGTTCGTAGGGCTTTCCCCC
adw   :
adr(1):          C                                   C
adr(2):          C
ayr   :          C
ayw(1):          C
ayw(2):          C
```

FIGURE 2 (continued).

```
                730                 750                 770
                .                   .                   .                   .
adw2  :  ACTGTTTGGCTTTCAGCTATATGGATGATGTGGTATTGGGGGCCAAGTCTGTACAGCATC
adw   :
adr(1):                  T                                               A
adr(2):                  T                                               A
ayr   :                                                                  A
ayw(1):                  T
ayw(2):                  T

                790                 810                 830
                .                   .                   .                   .
                                                                        st
adw2  :  GTGAGTCCCTTTATACCGCTGTTACCAATTTTCTTTTGTCTCTGGGTATACATTTAAACC
adw   :
adr(1):  T           T    T  A                        T              G
adr(2):  T           T    T  A                        T
ayr   :  T           T    T  A                        T              G
ayw(1):  T           T                                T
ayw(2):  T           T                      C         T

                850                 870                 890
                .                   .                   .                   .
adw2  :  CTAACAAAACAAAAAGATGGGGTTATTCCCTAAACTTCATGGGCTACATAATTGGAAGTT
adw   :
adr(1):     T     C   C T     C  C     T           A  TG
adr(2):     T     C   C T     C  C     T           A  TG
ayr   :     T     C   C T     C  C     T           A  TG
ayw(1):              G          C  T     T  T     T  TG C       T
ayw(2):                         C  TT  C T          TG C       T

                910                 930                 950
                .                   .                   .                   .
adw2  :  GGGGAACTTTGCCACAGGATCATATTGTACAAAAGATCAAACACTGTTTTAGAAAACTTC
adw   :
adr(1):     T     A         A          TT   AC    G  A      C        G
adr(2):     T     A  G      A           T   AC    G  A      C     T  G
ayr   :     T     A  G      A               AC    G  A      C     T  G
ayw(1):  AT  GT C         A  A  C  CA        A      G A
ayw(2):  AT  GT A         A     C  CA    G   A      G A  C

                970                 990                 1010
                .                   .                   .                   .
adw2  :  CTGTTAACAGGCCTATTGATTGGAAAGTATGTCAAAGAATTGTGGGTCTTTTGGGCTTTG
adw   :
adr(1):     A  T  A
adr(2):     A  T  C
ayr   :     A  T  A
ayw(1):  A                                  C                   T
ayw(2):                              C      C T                 T

               1030                1050                1070
                .                   .                   .                   .
adw2  :  CTGCTCCATTTACACAATGTGGATATCCTGCCTTAATGCCTTTGTATGCATGTATACAAG
adw   :
adr(1):     C  T              C           G                          T
adr(2):     C  T           C  C           G        A                 T
ayr   :     C  T              C           G        A                 T
ayw(1):     C  T              T        G  G                    T     T
ayw(2):     C  T              T        T                       T  GT
```

FIGURE 2 (continued).

```
                1090                1110                1130
                 .                   .                   .                   .
adw2   :  CTAAACAGGCTTTCACTTTCTCGCCAACTTACAAGGCCTTTCTAAGTAAACAGTACATGA
adw    :
adr(1):       G                          T           GT  C     A   C  C
adr(2):       G                          T           GT        A  TC
ayr    :       G                                      GT        A  TC A
ayw(1):       G                                      GT        A   C
ayw(2):    G  G         T                             GT        A   C

                1150                1170                1190
                 .                   .                   .                   .
adw2   :  ACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGTTTGCTGACGCAACCC
adw    :
adr(1):                  C          T A      C
adr(2):                  C          T A      C
ayr    :                 C          T A      C
ayw(1):                  C            A
ayw(2):                  C            A

                1210                1230                1250
                 .                   .                   .                   .
adw2   :  CCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCATGCGTGGAACCTTTGTGGCTCCTC
adw    :                  A
adr(1):         A                        G                          T
adr(2):         A                        G
ayr    :      G  T                       G
ayw(1):                    T   G                              TC
ayw(2):                    T   G                              C      G

                1270                1290                1310
                 .                   .                   .                   .
adw2   :  TGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGCTCGCAGCCGGTCTGGAGCAA
adw    :
adr(1):                              A                GA
adr(2):                              A                                  G
ayr    :                             A                                  G
ayw(1):                                                  A
ayw(2):                                                  A

                1330                1350                1370
                 .                   .                   .                  Xi   .
adw2   :  AGCTCATCGGAACTGACAATTCTGTCGTCCTCTCGCGGAAATATACATCGTTTCCATGGC
adw    :                                                         A
adr(1):   A  T      G         C  G  T         T         C  C  C  C
adr(2):   A  T         C      C     T         T         C  C  C
ayr    :   A  T         C     C  A  T         T         C  C  C
ayw(1):   CA T      G     T  C     T     A   C  C
ayw(2):   CA TC     G  G  T  C     T  T      C  C               A

                1390                1410                1430
                 .                   .                   .                   .
adw2   :  TGCTAGGCTGTACTGCCAACTGGATCCTTCGCGGGACGTCCTTTGTTTACGTCCCGTCGG
adw    :
adr(1):      C  G   G                  G                   C
adr(2):         G   G                  G                   C
ayr    :            G                  G                   C
ayw(1):            G                  G
ayw(2):            G                  G
```

FIGURE 2 (continued).

```
                1450                1470                1490
                 .         .         .         .         .         .
adw2   :  CGCTGAATCCCGCGGACGACCCCTCTCGGGGCCGCTTGGGACTCTCTCGTCCCCTTCTCC
adw    :
adr(1):                         G         T     C    AC      T G  TT
adr(2):                         G         T     G    AC           T
ayr    :                        G         T     C    AC           T
ayw(1):           T             T       T
ayw(2):                         T       T

                1510                1530                1550
                 .         .         .         .         .         .
adw2   :  GTCTGCCGTTCCAGCCGACCACGGGGCGCACCTCTCTTTACGCGGTCTCCCCGTCTGTGC
adw    :
adr(1):  C
adr(2):  T          G
ayr    :  A          G
ayw(1):             GA                          A
ayw(2):           T GA                          A

                1570                1590                1610
                 .         .         .         .         .         .
adw2   :  CTTCTCATCTGCCGGTCCGTGTGCACTTCGCTTCACCTCTGCACGTTGCATGGAGACCAC
adw    :                                                       C
adr(1):                 A                           C
adr(2):                 A                           C
ayr    :                A                           A
ayw(1):                 A                           C
ayw(2):                 A                           C

                1630                1650                1670
              pt
                 .         .         .         .         .         .
adw2   :  CGTGAACGCCCATCAGATCCTGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCCCAGC
adw    :
adr(1):          G   C   G  T        C                          T
adr(2):              C   G  T                                   T
ayr    :             C   G  T              C                    T
ayw(1):             C  A  AT                                    T
ayw(2):       A     A CAT  T                                    T T T

                1690                1710                1730
                 .         .         .         .         .         .
adw2   :  AATGTCAACGACCGACCTTGAGGCCTACTTCAAAGACTGTGTGTTTAAGGACTGGGAGGA
adw    :
adr(1):           A              A              T       A
adr(2):  C                       A                      A
ayr    :                         A              T       A
ayw(1):                          A              T       A
ayw(2):                          A              T       A

                1750                1770                1790
                 .         .         .         .         .         .
adw2   :  GCTGGGGGAGGAGATTAGGTTAAAGGTCTTTGTATTAGGAGGCTGTAGGCACAAATTGGT
adw    :  T                       T A                          T
adr(1):  T                                  C                 T
adr(2):  T                                  C                 .........
ayr    :  T                                  C                 T
ayw(1):  T                                  C                 T
ayw(2):  T                 A                                  T
```

FIGURE 2 (continued).

```
               1810                1830                1850
                  .     PCi .         .       Xt.         .         .
adw2   : CTGCGCACCAGCACCATGCAACTTTTTCACCTCTGCCTAATCATCTCTTGTACATGTCCC
adw    :
adr(1) :    TT                                          A   T       T
adr(2) : ..................T                            A   T       T
ayr    :    TT                       C                  A   T       T
ayw(1) :                                                    T       T
ayw(2) :                                                    T       T

               1870                1890                1910
                  .         .         .         .  Ci     .         .
adw2   : ACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTAT
adw    :
adr(1) :                                                         G
adr(2) :                                                         G
ayr    :                                                         G
ayw(1) :                                                   C
ayw(2) :                                                      T

               1930                1950                1970
                  .         .         .         .         .         .
adw2   : AAAGAATTTGGAGCTACTGTGGAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCC
adw    :
adr(1) :                T                T                          T
adr(2) :               AT                T                       G  T
ayr    :                T                T                          T
ayw(1) :                                                            A
ayw(2) :                                                            A

               1990                2010                2030
                  .         .         .         .         .         .
adw2   : GTCAGAGATCTCCTAGACACCGCCTCAGCTCTGTATCGAGAAGCCTTAGAGTCTCCTGAG
adw    :   AC
adr(1) : A TC          C           T           G  G              G  A
adr(2) : A TC          T           T           G  G              G  A
ayr    : A TC          C           T           G  G              G  A
ayw(1) :   AC       T     T                    G
ayw(2) :   AC       T     T A                  G

               2050                2070                2090
                  .         .         .         .         .         .
adw2   : CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGGAATTGATG
adw    :
adr(1) :      T              A              T     G  T     T  G
adr(2) :      T              A              T     G  T     T  G  A
ayr    :      T              A              T     G  T     T  G
ayw(1) :      T                             A     T            C A
ayw(2) :      T                             A  A  G            C A

               2110                2130                2150
                  .         .         .         .         .         .
adw2   : ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTA
adw    :                               C             C  A     A     C
adr(1) :  A T G  C           A G            C        C     AT A     C
adr(2) :  A   G  C           A G            C        C     AT A     C
ayr    :  A   G  C           A G            C        C     AT A     C
ayw(1) :                      GT                  G     A  C  A     C
ayw(2) :                      GG                AT   C     C  A     C
```

FIGURE 2 (continued).

```
                2170                2190                2210
                 .                   .                   .
adw2   :  AATTATGTTAATACTAACGTGGGTTTAAAGATCAGGCAACTATTGTGGTTTCATATATCT
adw    :             A
adr(1):   GC     C   GT   TA    CC    A     A                    T  C
adr(2):   GC     C   GT   TA    CC    A     A                 C  T  C
ayr    :   GC    C   GT   TA    CC    A  T  A                 C  T  C
ayw(1):   G      C  C     TA    CC     T           C          C  T
ayw(2):   G      C  C     TA    CC    AT                      C  T

                2230                2250                2270
                 .                   .                   .
adw2   :  TGCCTTACTTTTGGAAGAGAGACTGTACTTGAATATTTGGTCTCTTTCGGAGTGTGGATT
adw    :
adr(1):    T                A     T     G        G     T
adr(2):                     A     TT G  G        A     T
ayr    :                    A     C     G        G     T
ayw(1):    T  C             A  A  TA A  G        G
ayw(2):    T  C             A  A  TA A  G        G     T

                2290                2310                2330
                 .                   .                   .
                                     pi
adw2   :  CGCACTCCTCCAGCCTATAGACCACCAAATGCCCCTATCTTATCAACACTTCCGGAAACT
adw    :
adr(1):             C  T  C
adr(2):             C  T  C
ayr    :             C  T  C
ayw(1):                T                      C              G
ayw(2):                T                                     G

                2350                2370                2390
                 .                   .                   .
adw2   :  ACTGTTGTTAGACGACGGGACCGAGGCAGGTCCCCTAGAAGAAGAACTCCCTCGCCTCGC
adw    :
adr(1):                  ......
adr(2):                  ......
ayr    :                 ......
ayw(1):                  ......
ayw(2):                  ......

                2410                2430                2450
                 .                   .                   .          ct.
adw2   :  AGACGCAGATCTCCATCGCCGCGTCGCAGAAGATCTCAATCTCGGGAATCTCAATGTTAG
adw    :               A
adr(1):       A  G      A
adr(2):       A  G      A
ayr    :      A  G      A
ayw(1):       A  G      A
ayw(2):       A  G      A

                2470                2490                2510
                 .                   .                   .
adw2   :  TATTCCTTGGACTCATAAGGTGGGAAACTTTACGGGGCTTTATTCCTCTACAGTACCTAT
adw    :                        C
adr(1):     C                          T           T     T     G
adr(2):     C                          T           T     T     G
ayr    :    C                          T           T     T     G
ayw(1):                          G     T           T     T     G
ayw(2):                                            T     T     G
```

FIGURE 2 (continued).

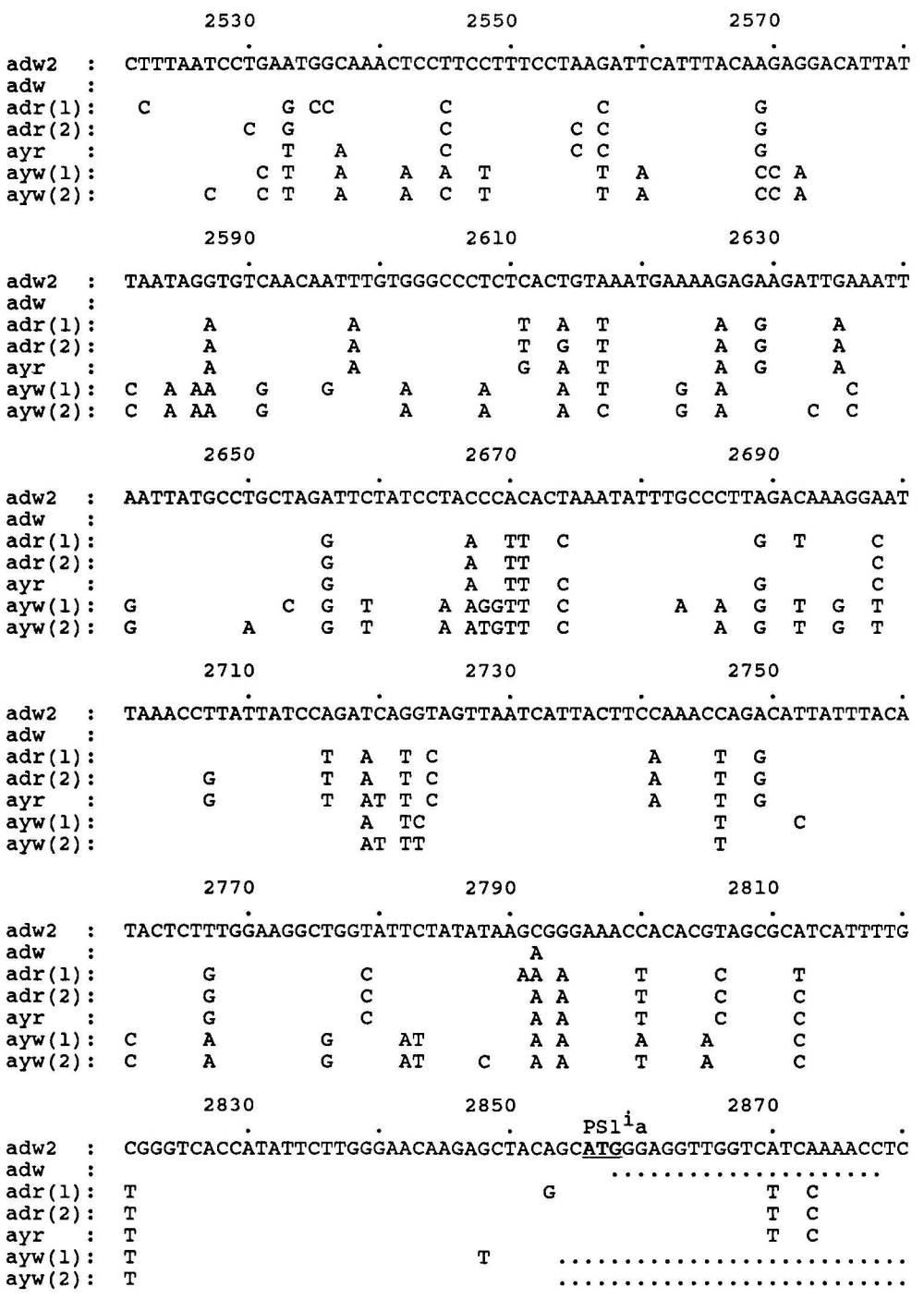

FIGURE 2 (continued).

```
               2890                2910                2930
                 PS1ib      .         .         .         .
adw2   : GCAAAGGCATGGGGACGAATCTTTCTGTTCCCAATCCTCTGGGATTCTTTCCCGATCATC
adw    :                                   C            C
adr(1):   AC                                                       C
adr(2):   AC                                                       C
ayr    :  AC                                                       C
ayw(1):  ......        CA         CACCAG                        C  C
ayw(2):  ......        CA         CACCAG                        C  C

               2950                2970                2990
                  .         .         .         .         .         .
adw2   : AGTTGGACCCTGCATTCGGAGCCAACTCAAACAATCCAGATTGGGACTTCAACCCCGTCA
adw    :                               CA                        A
adr(1):               G                                          AA
adr(2):               G                                          AA
ayr    :              G                                          AA
ayw(1):         T  A  C   A    A   A CGCA                    T   AA
ayw(2):         T  A  C   A    A   A CGCA                    T   AA

               3010                3030                3050
                  .         .         .         .         .         .
adw2   : AGGACGACTGGCCAGCAGCCAACCAAGTAGGAGTGGGAGCATTCGGGCCAAGGCTCACCC
adw    :      C                   G                        G
adr(1):      TC         AG  A T A G       C     AC         G  T
adr(2):      TC A       AG  A  T  G       C                G  T
ayr    :     TC         AG  A  T  G       C         T  T   G  T
ayw(1):       AC        AC      A G       CT            TGG TT
ayw(2):       AC        AC      A G       CT            TGG AT

               3070                3090                3110
                  .         .         .         .         .         .
adw2   : CTCCACACGGCGGTATTTTGGGGTGGAGCCCTCAGGCTCAGGGCATATTGACCACAGTGT
adw    :
adr(1):   A            C                                     A      C
adr(2):   A            C                                     A      C
ayr    :  A        A  CC                                     A   C  C
ayw(1):   A  G     A  CC                                C ACAA  TT  C
ayw(2):   A        A  CC                                C AGAA  GT  C

               3130                3150                3170
                  .         .         .         .         .         .
adw2   : CAACAATTCCTCCTCCTGCCTCCACCAATCGGCAGTCAGGAAGGCAGCCTACTCCCATCT
adw    :
adr(1):    G  GCG          TT                       A
adr(2):    G  GCA                                   A
ayr    :   G  GCA                                   A
ayw(1):    G   A   G                    C                    C  GC G
ayw(2):    G   A   G           T        C                    C  GC G

               3190                3210
                  .         .         .  PS2i   .
adw2   : CTCCACCTCTAAGAGACAGTCATCCTCAGGCCATGCAGTGG
adw    :
adr(1):
adr(2):
ayr    :
ayw(1):          T G   A   C
ayw(2):          T G   A   C
```

FIGURE 2 (continued).

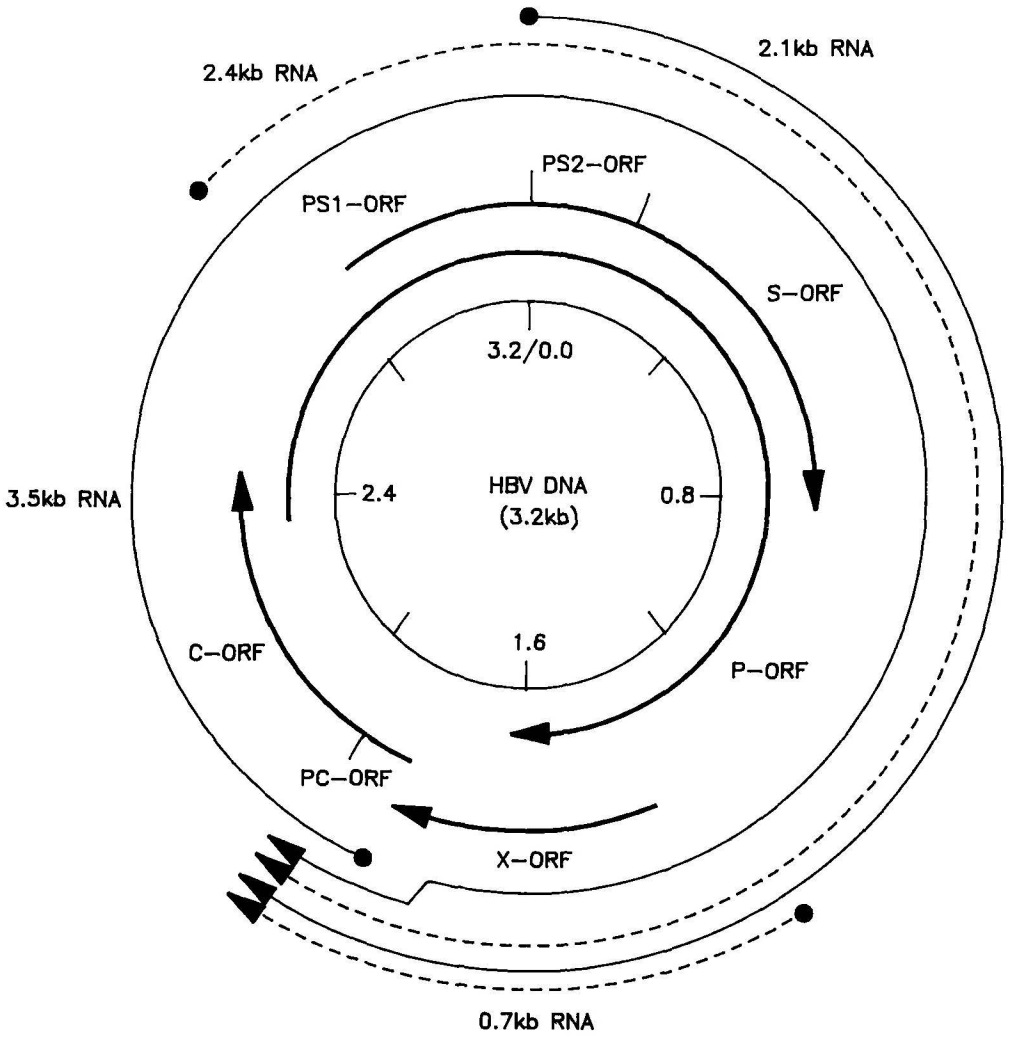

FIGURE 3. Organization of the HBV genome. The coordinates of the HBV DNA are the same as in Figure 2. PS, presurface; S, surface; P, polymerase; X, X gene; PC, precore; C, core; ORF, open reading frame.

III. ORGANIZATION OF HEPATITIS B VIRUS

A. VIRAL CODING CAPACITY

The sequences of the various HBV DNAs[47-61] revealed that there are four long open reading frames conserved between all the viral genomes (Figures 2 and 3). These encode the envelope or surface antigens (HBsAg),[72-75] the nucleocapsid antigens (HBcAg and HBeAg),[51,76,77] the DNA polymerase (P) gene product[78-80] and the X gene product.[66,81,82] The surface antigen (subtype *ayw*) open reading frame contains three in-frame translational initiation codons that permit the synthesis of HBsAg polypeptides of 25, 31, and 43 kDa.[50] These polypeptides are variably glycosylated, giving rise to the six polypeptides HBsAg/P25:GP28, HBsAg/GP33:GP36, and HBsAg/P43:GP46, respectively[75] (Figure 4). These surface antigen polypeptides are also known as the major, middle, and large HBsAg polypeptides, respectively.

The nucleocapsid antigen open reading frame contains two in-frame translation initiation codons that permit the synthesis of polypeptides of 18 and 21 kDa (Figure 5). Synthesis of the

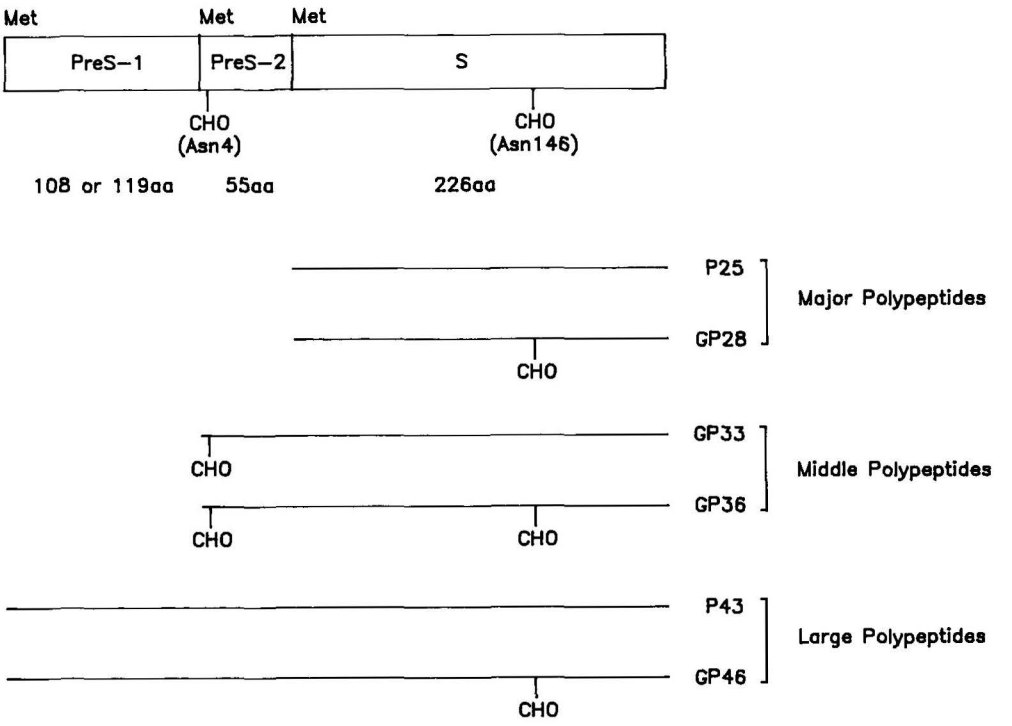

FIGURE 4. Organization of the HBsAg open reading frame. The methionine (Met) residues encoded by the initiation codons for the large, middle, and major surface antigen polypeptides are indicated. The asparagine (Asn) residues at which carbohydrate (CHO) can modify the HBsAg are shown.

HBeAg/P18 polypeptide results from the proteolytic processing of both the amino and carboxy termini of the primary translation product of the complete nucleocapsid open reading frame, which has the capacity to encode a polypeptide of 24 kDa.[76,77] The polypeptide product synthesized from the second in-frame translation initiation codon is the 21-kDa core antigen (HBcAg/P21) polypeptide.[51]

The DNA polymerase and X gene open reading frames have the capacity to code for polypeptides of 94 and 17 kDa, respectively. The molecular masses of the HBV polypeptides are derived from their predicted amino acid sequences determined from the nucleotide sequence of the HBV genome (subtype *ayw*).[50] A variation in the molecular mass of the large HBsAg polypeptide has been observed and ascribed to surface antigen subtype variation.[75] This difference in size results from an additional 11 amino-terminal amino acids present in the large envelope polypeptide, HBsAg/P44:GP47 (subtype adw_2)[52] (Figure 2).

B. STRUCTURE OF THE HBV PARTICLE

The envelope of the HBV Dane particle possesses the same antigenic determinants as are found on the 22-nm diameter subviral spheres and filaments present in the serum of infected individuals.[17,22] Characterization of the polypeptide composition of HBV and subviral particles demonstrated that the common antigenic determinants were located within the HBsAg polypeptides.[74,75] The HBsAg polypeptides compose a set of six coterminal polypeptides that differ by the extent of their glycosylation and amino-terminal sequence[75] (Figure 4). The major HBsAg (HBsAg/P25:GP28) comprises 226 amino acids and the HBsAg/P25 polypeptide differs from the HBsAg/GP28 polypeptide by the addition of Asn_{146}-linked complex glycan.[75,83,84] The glycan, representing approximately 75 μg carbohydrate per milligram HBsAg,

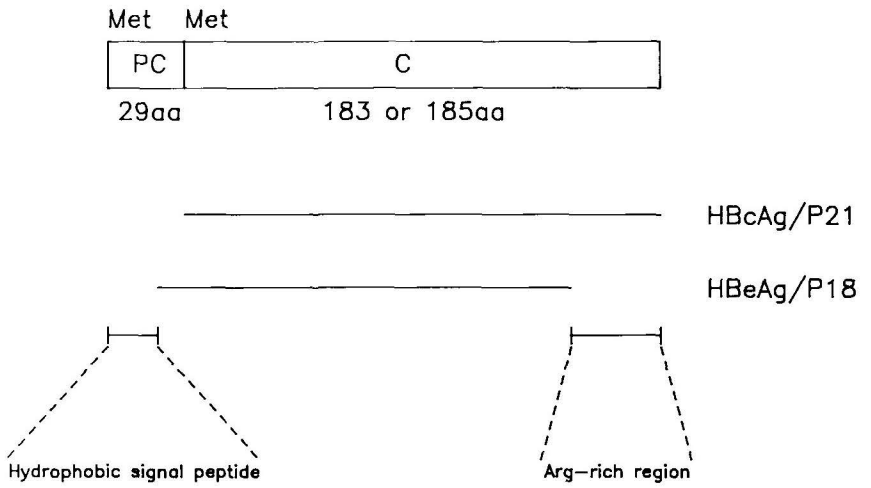

FIGURE 5. Organization of the nucleocapsid open reading frame. The methionine (Met) residues encoded by the precore (PC) and core (C) initiation codons are indicated. The 19-amino acid hydrophobic signal peptide and the 34-amino acid arginine-rich carboxy-terminal region cleaved from the HBeAg/P24 precursor to produce the secreted HBeAg/P18 are shown.

is composed of *N*-acetylglucosamine, mannose, galactose, and sialic acid residues plus fucose as a minor component.[83,85] The middle HBsAg (HBsAg/GP33:GP36) differs from the major HBsAg by an additional 55 amino-terminal amino acids encoded by the preS2 region.[74] In addition, there is evidence that Asn_4-linked (Asn_4 of the preS2 region) mannose-rich glycan is always found added to the middle HBsAg polypeptides, accounting for the absence of a nonglycosylated form of this polypeptide.[74,86] The large HBsAg (HBsAg/P43:GP46) differs from the middle HBsAg by an additional 108 or 119 amino-terminal amino acids, depending on the viral strain, encoded by the preS1 region. In addition, the preS2 glycosylation site is not substituted in the large HBsAg polypeptide, and the presence or absence of Asn_{146}-linked complex glycan in the major HBsAg domain of the large HBsAg polypeptide is considered to account for the two forms of this polypeptide.[75]

Characterization of the surface antigen composition of serum-derived HBV particles, filaments, and spheres demonstrated that they possessed different ratios of the various envelope polypeptides.[75] The composition of the filaments and Dane particles is approximately 10 to 20% of each of the middle and large polypeptides, with the remaining HBsAg contributed by the major surface antigen polypeptides. In contrast, the 22-nm spheres comprise only 1 to 2% large envelope polypeptide, 10 to 20% middle polypeptide, and the remaining envelope antigen is contributed by the major surface antigen polypeptide.[75] The level of glycosylation of Asn_{146} among the three envelope polypeptides was approximately the same, with about half of the molecules modified.[75] The contribution of carbohydrate to the total mass of the 22-nm spheres has been estimated to be between 3 and 8%,[85,87] and based on the relative abundance of the larger envelope polypeptides and the extent of their glycosylation, it is probable that the contribution of carbohydrate to the mass of the HBsAg filaments and envelope of the virion is similar. The HBV envelope contains approximately 400 subunits,[75] including 40 to 80 molecules of each of the middle and large polypeptides in the virion. In comparison, the 22-nm sphere consists of approximately 100 polypeptide subunits,[75,88] of which very few are the large envelope polypeptide. Eukaryotic expression studies of the three surface antigen open reading frames demonstrated that the middle and major envelope polypeptides can assemble into spherical 22-nm HBsAg particles,[89-99] whereas production of the large envelope polypep-

tide is necessary for the synthesis and assembly of HBsAg filaments.[94,100] These observations indicate that the large envelope polypeptide influences the nature of the assembled surface antigen particles and is an important component of the HBV envelope.[75] In addition to the surface antigen polypeptides, the 22-nm HBsAg particles contain cellular lipid.[101-103] The contribution of the lipid to the mass of the 22-nm spheres was determined to be approximately 25% and its composition is similar to other normal human serum lipoproteins.[101] The presence of glycolipid in HBsAg particles has been reported[102] but has not been observed in subsequent analyses.[85,101]

The predominant polypeptide of the nucleocapsid of HBV is the 21-kDa hepatitis B core antigen, HBcAg.[51,75,104-109] The nucleocapsid appears to be composed of approximately 180 of these polypeptide subunits, which have been proposed to be assembled with icosahedral symmetry.[110] The HBcAg is a phosphoprotein that possesses autophosphorylating serine protein kinase activity.[111-113] The significance of this activity is currently unclear. It has been shown that there is endogenous DNA polymerase activity within the nucleocapsid of HBV,[34,114] which is presumably encoded by the HBV DNA polymerase open reading frame.[78-80] As in the DHBV system, it is likely that the amino-terminal portion of the HBV DNA polymerase open reading frame encodes a polypeptide, the terminal protein, which is attached to the 5′ end of the HBV DNA minus strand[44] and probably serves as the primer for reverse transcription of the viral pregenomic RNA.[114-116] The HBV DNA polymerase open reading frame also encodes RNase H activity, which is likely to be responsible for the degradation of the pregenomic RNA as the minus strand of HBV DNA is being synthesized.[114]

The three envelope polypeptides, the P21 core polypeptide, and the HBV DNA polymerase polypeptide represent all of the protein components of the virion that have been identified as viral structural polypeptides. In addition to the protein, carbohydrate, and lipid, the virion also contains a 3.2-kb partially double-stranded circular DNA genome that is encapsidated within the nucleocapsid.[36,38-41]

IV. INTRACELLULAR LIFE CYCLE OF HBV IN THE HUMAN HEPATOCYTE

A. VIRUS ENTRY INTO HEPATOCYTES

The inability to infect permanent tissue culture cell lines with HBV is a major obstacle to identifying the mechanism of viral entry into hepatocytes.[117-119] This has prevented the identification of putative viral receptors and their ligands. However, evidence has been presented suggesting there may be selective interactions between surface antigen polypeptides and cell surface elements of several hepatic and nonhepatic tissue culture cell lines and also primary human hepatocyte plasma membrane components.[120-124] These interactions may be occurring directly through receptor–ligand interactions or indirectly through polyalbumin, the albumin receptors on the surface of the human hepatocytes,[125-127] and the receptor for polymerized human serum albumin present in the preS2 region of the envelope polypeptides.[92,93,128] The role of these interactions in the infection of hepatocytes and possibly other cell types is currently unclear.

B. GENERATION OF A TRANSCRIPTIONALLY ACTIVE HBV TEMPLATE

Once HBV has entered the hepatocyte, whether by a receptor-mediated process or by fusion with the plasma membrane, it must release the nucleocapsid from the surface antigen envelope (Figure 6). This process may be intimately linked with virus entry into the cell or may represent a separate step in the viral life cycle. The viral genome must then be translocated to the nucleus, presumably within the nucleocapsid and mediated by the nuclear translocation signal present in the HBcAg polypeptide.[94] Since the presumed template for the

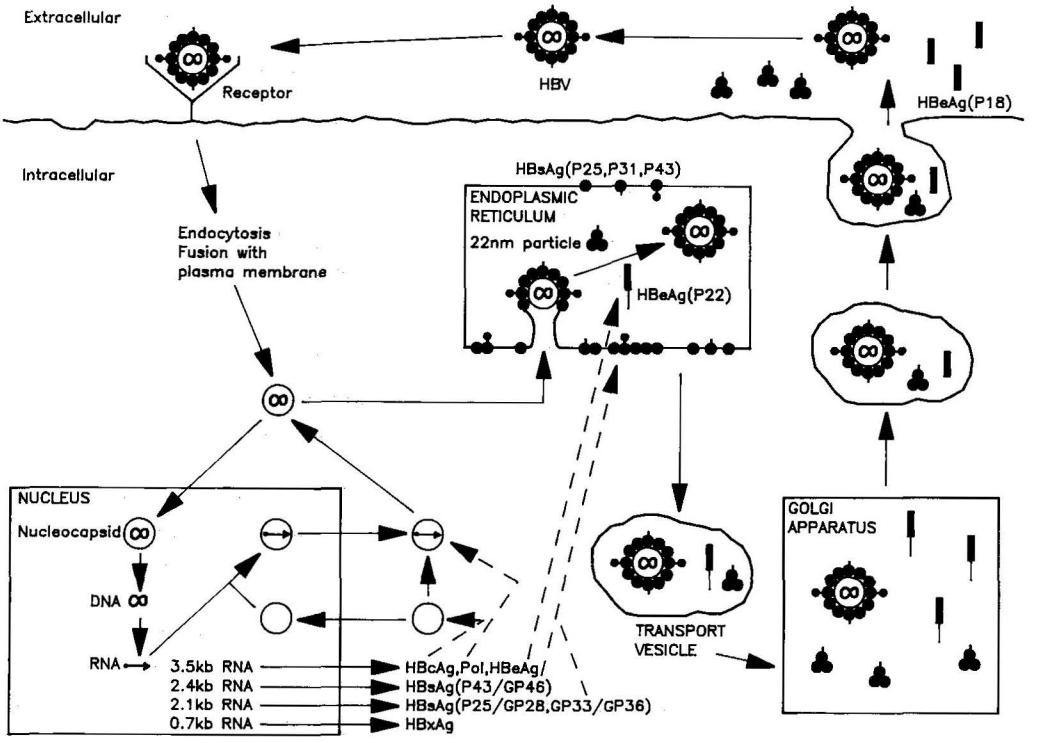

FIGURE 6. The intracellular pathway for the synthesis and secretion of HBV, HBsAg subviral particles, and HBeAg polypeptides.

transcription of HBV mRNAs is a covalently closed circular molecule of the HBV genome, several modifications of the infecting viral genome must be made before transcription can occur.[43] The single-stranded region of the HBV genome must be converted to double-stranded DNA. As it is known that the synthesis of this region of DNA can be completed *in vitro* by the endogenous HBV DNA polymerase present in the nucleocapsid of the virus,[35,37,38] it is likely that during infection the same enzyme may convert the partially double-stranded DNA into nicked circular double-stranded DNA. This event could occur within the nucleocapsid or might occur to the HBV genome after it has been released from the nucleocapsid into the nucleoplasm. In addition to the synthesis of DNA, removal of the terminal protein and oligoribonucleotide attached to the 5′ ends of the viral DNA strands, and ligation of the nicks in the double-stranded DNA must occur to generate a covalently closed, double-stranded HBV genome.[129]

C. TRANSCRIPTION OF THE HBV GENOME

HBV-infected liver in both acute and chronic disease expresses two predominant HBV-specific transcripts of 2.1 and 3.5 kb[130-134] (Figures 3 and 6). The 2.1-kb RNA has several transcription initiation sites that span the translation initiation codon for the middle envelope polypeptide,[67,135,136] and terminates at the single polyadenylation site in the HBV genome.[135-137] These transcripts encode the middle and major surface antigen polypeptides. The 3.5-kb RNA also has several transcription initiation sites that span the initiation codon for the precore signal sequence which precedes the HBcAg open reading frame.[67,135] The transcripts initiating before the precore initiation codon encode the secreted HBeAg/P18 polypeptide, whereas the transcripts initiating after the precore initiation codon encode the HBcAg/