Molecular Biology of the Hepatitis B Virus

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
Boca Raton Ann Arbor Boston London

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
CIP

Direct all inquiries to CRC Press, Inc., 2000 Corporate Blvd., N. W., Boca Raton, Florida, 33431.

International Standard Book Number 0-8493-5516-8

Library of Congress Card Number 91-32641
Printed in the United States 0 1 2 3 4 5 6 7 8 9

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.

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

Anneke K. Raney and Alan McLachlan

TABLE OF CONTENTS

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

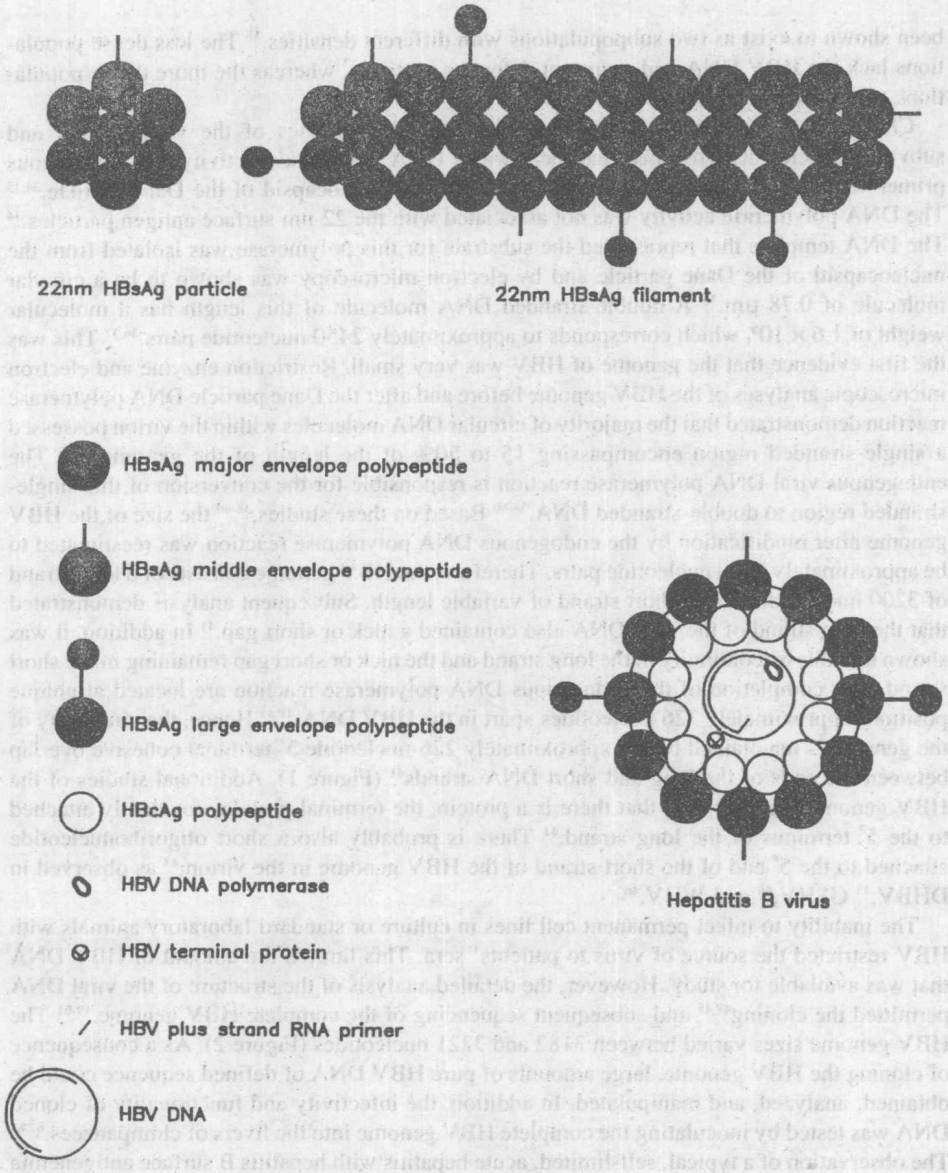

FIGURE 1. Diagrammatic representation of the structure of the 22-nm hepatitis B surface antigen (HBsAg) sphere, 22-nm HBsAg filament, and hepatitis B virus (Dane particle). HBcAg, hepatitis B core antigen.

molecular weight of 2 to 4 × 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]

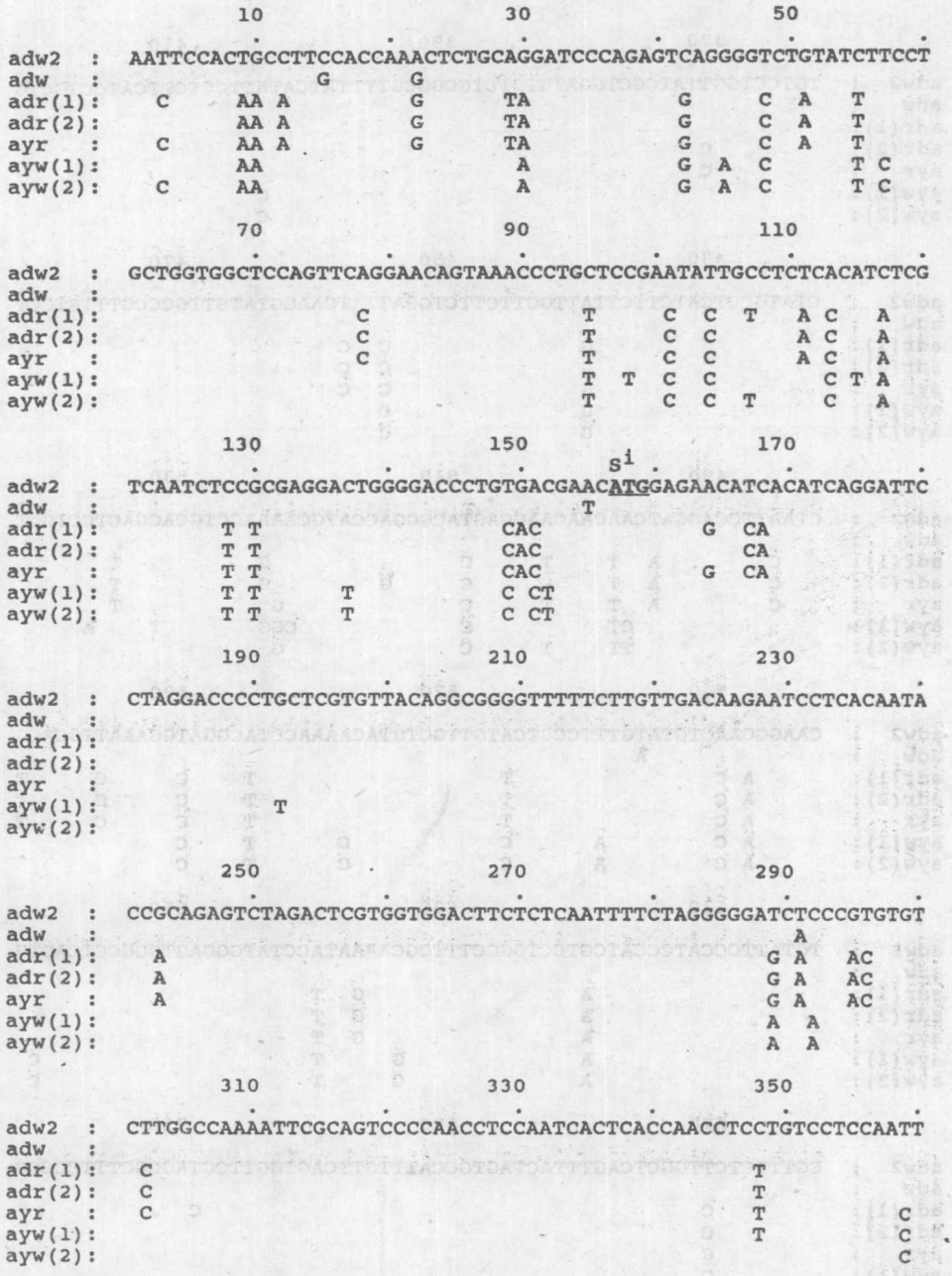

FIGURE 2. Comparison of the HBV DNA sequences of the major subtypes. The HBV genomes are subtypes *adw*$_2$,[57] *adw*,[54] *adr* (sequences 1^{61} and 2^{54}), *ayr*,[60] and *ayw* (sequences 1^{50} and 2^{56}). The sequences were aligned with the *adw*$_2$ sequence using the first A residue of the *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^{1}a and PS1^{1}b, initiation codons for the 119- and 108-amino acid preS1 regions, respectively; PS2^1, initiation codon for preS2 region; S^1 and S^1, initiation and termination codon for the major HBsAg open reading frame; PC^1, initiation codon for the precore region; C^1 and C^1, initiation and termination codon for the core open reading frame; X^1 and X^1, initiation and termination codon for the X gene open reading frame; P^1 and P^1, 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).

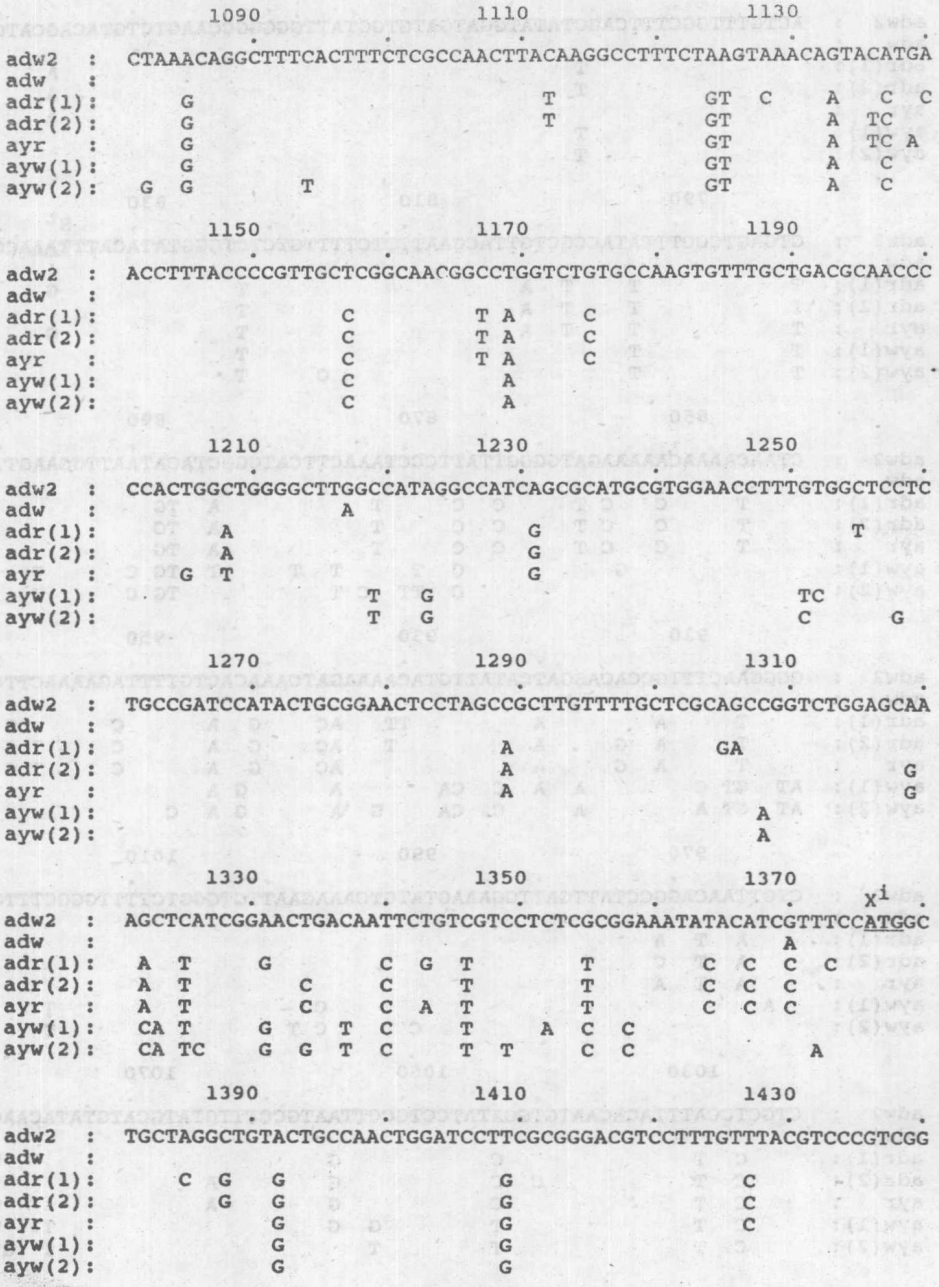

FIGURE 2 (continued).