CRC REVIVALS

Molecular Biology of the Hepatitis B Virus

Edited by Alan McLachlan



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 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

© 1991 by CRC Press, Inc. CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

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 13
Chromosomal Changes Associated with Hepatitis B Virus DNA Integration and
Hepatocarcinogenesis
Kenichi Matsubara
Chapter 14
The Molecular Biology of Cauliflower Mosaic Virus
Robert J. Shepherd
Chapter 15
Use of Recombinant HBV Proteins in Immunological Studies
David R. Milich
Chapter 16
Recombinant Hepatitis B Vaccines
Ronald W. Ellis and Peter J. Kniskern
Index

Chapter 1

THE BIOLOGY OF HEPATITIS B VIRUS

Anneke K. Raney and Alan McLachlan

TABLE OF CONTENTS

I.	Introc	luction	2
II.	The H of He	History of the Discovery, Isolation, and Characterization patitis B Virus	2
III.	Organ A. B.	nization of Hepatitis B Virus	4
IV.	Intrac A. B. C. D.	ellular Life Cycle of HBV in the Human Hepatocyte	7 7 7 8
	E. F.	Replication of HBV	20 21
V.	HBV A. B. C.	Infection 2 Host Range and Tissue Tropism 2 Routes of Infection 2 Clinical Consequences of Infection 2	!2 !2 !3
VI.	HBV	-Related Infectious Agents2	25
VII.	Sumn	nary and Conclusions2	25
Ackn	lowled	gments2	26
Refe	rences	2	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/cm³ 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/cm³ ²⁹⁻³³ and a sedimentation coefficient of approximately 110.³⁴ The density of the Dane particle, 1.24 to 1.27 g/cm³,^{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 singlestranded 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 strands⁴¹ (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}

		10				30					5	0			
adw2 :	AATTCO	CACTGO	CCTTC	CACC.	АААСІ	CTGCAGGA	гссс	AGA	GTC.	AGG	GGTC	TG:	[A]	CI	TCCT
adw :		-	G		G	2.54									
adr(1):	С	AA	A		G	TA			G		С	Α		т	
adr(2):	-	AA	A		G	TA			G		С	Α		т	
ayr :	С	AA	A		G	TA			G		С	Α		т	
ayw(1):		AA				A			G	A	С			т	С
ayw(2):	С	AA				A			G	Α	С			т	С
		70				90					11	0			
adw2 :	GCTGGT	rggcTo	CCAGT	TCAG	GAACA	GTAAACCC	TGCI	CCG	AAT	ATT	GCCT	CTC	CAC	CAT	CTCG
adw :															
adr(1):				С			т		С	С	т	Α	С		A
adr(2):				С			т		С	С		Α	С		Α
ayr :				С			т		C	С		Α	С		А
ayw(1):							\mathbf{T}	т	С	С			С	т	Α
ayw(2):							т		С	С	т		С		A
		130				150		4			17	0			
			~ ~ ~ ~ ~ ~					S	•			•			:
adw2 :	TCAAT	CTCCGG	GAGG	ACTG	GGGAC	CCTGTGAC	GAAC	ATG	GAG.	AAC.	ATCA	CA'	rc4	AGG	A'I''I'C
adw :						~ ~ ~	T			-					
adr(1):		TT				CAC				G	CA				
adr(2):		TT				CAC				-	CA				
ayr :		TT		-		CAC				G	CA				
ayw(1):		TT		T		C CT									
ayw(2):		тт		т		ССТ									
		190				210					23	0			
adw2 ·	CTACCI	Acces	rcere	GTGT	TACAG	CCCCCCCTTT	ኮጥጥ	יידירכ	• ጥጥር	404	AGAA	• ••••	יידיר	יאר	גיד ג גי
adw :	0111001			0101								10.			
adr(1):															
adr(2):															
avr :															
avw(1):			Т												
ayw(2):			-												
/		250				270					20	^			
		250				270			12		29				
adw2 :	CCGCA	GAGTC	FAGAC	TCGT	GGTGG	ACTTCTCT	СААТ	TTT	CTA	GGG	GGAT	CTO	cco	GT	GTGT
adw :												Α			
adr(1):	А										G	Α		AC	2
adr(2):	A										G	Α		AC	2
ayr :	Α										G	Α		AC	2
ayw(1):											A	2	A		
ayw(2):											A	2	A		
		310				330					35	0			
					•				•			•			•
adw2 :	CTTGG	CCAAA	ATTCG	CAGT	CCCCA	ACCTCCAA	TCAC	TCA	CCA	ACC	TCCT	GT	CCI	CC	AATT
adw :	~										_				
adr(1):	0										T				
adr(2):	C										T				_
ayr :	C										T				C
ayw(1):											т				С
ayw(2):															С

FIGURE 2. Comparison of the HBV DNA sequences of the major subtypes. The HBV genomes are subtypes $adw_2^{5^2}$ $adw_2^{5^4}$ adr (sequences 1⁶¹ and 2⁵⁴), ayr_1^{60} and ayw (sequences 1⁵⁰ and 2⁵⁶). 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^{ia} and PS1^{ib}, initiation codons for the 119- and 108-amino acid preS1 regions, respectively; PS2ⁱ, initiation codon for preS2 region; Sⁱ and Sⁱ, initiation and termination codon for the major HBsAg open reading frame; PCⁱ, initiation codon for the precore region; Cⁱ and Cⁱ, initiation and termination codon for the core open reading frame; Xⁱ and Xⁱ, initiation and termination codon for the X gene open reading frame; Pⁱ and Pⁱ, initiation and termination codon for the polymerase open reading frame.

	370			390					410		
adw2 :	TGTCCTGGTTA	TCGCTGGA	TGTGTC	GCGGG	CGTT	FTT.	АТС	ATATTC	CTCTTC	АТССТС	GCTG
adr(1): adr(2): ayr : ayw(1):	C C							с			
ayw(2):								С			
	430			450					470		
adw2 :	CTATGCCTCAT	CTTCTTAT	TGGTTC	TCTG	GATT	TAT	CAA	GGTATG	TTGCCC	GTTTGI	сст
adw : adr(1): adr(2): ayr : ayw(1):		9 9 9 9			с с с с с	C C C					
ayw(2):		G			С						
	490			510					530		
adw2 :	CTAATTCCAGO	АТСААСАА	CAACCAG	TACGO	GGAC	CA	TGC	аааасс	TGCACG	ACTCCI	GCT
adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	C C C	A T A T A T CT TT	т С т т	00000	G			G G G G G G G G	т	T T T A	
	550			570					590		
adw2 :	550 CAAGGCAACTC	TATGTTTC	CCTCATO	570 TTGCT	ÍGT <i>I</i>	ACA	AAA	CCTACG	590 GATGGA	AATTGO	CACC
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	550 CAAGGCAACTC A C A C A C A C A C A C	TATGTTTC A A A A	CCTCATC T T C C	570 STTGCI	ŤGT₽	ACA) C C	AAA	CCTACG T T T T T	590 GATGGA C C C C C C	AATTGC C C C	ACC T T
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	550 CAAGGCAACTC A C A C A C A C A C A C A C A C	TATGTTTC A A A A	T T C C	570 TTGC 630	∱GT₽	ACA. C C	AAA	CCTACG T T T T	590 GATGGA C C C C C C 650	AATTGO C C C	T T T
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 :	550 CAAGGCAACTO A C A C A C A C A C A C 610 TGTATTCCCAT	TATGTTTC A A A A CCCCATCGI	CCTCATC T T C C C C C C C C C C C C C C	570 STTGCT 630 CTTTTCC	¢gτ≠ GCA≯	C C C	AAA FAC	CCTACG T T T T T CCTATGG	590 GATGGA C C C C C 650 GAGTGG	AATTGO C C C GCCTCZ	CACC T T T
adw2 : adw : adr(1): adr(2): ayw (1): ayw (2): adw2 : adw : adw : adr(1): adr(2): ayw (1): ayw (1):	550 CAAGGCAACTO A C A C A C A C A C A C 610 TGTATTCCCAT	TATGTTTC A A A CCCCATCGT A A A A A A A A A A	CCTCATC T T C C C C C C C	570 STTGCT 630 CTTTCC	fgtz GCA	CAJ C C Q Q G G G	AAA TAC T T T T	CCTACG T T T T T CTATGG	590 GATGGA C C C C 650 GAGTGG	AATTGC C C C	CACC T T T AGTC
adw2 : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adw : adr(1): ayr : ayr : ayw(1): ayw(2):	550 CAAGGCAACTC A C A C A C A C A C 610 TGTATTCCCAT	TATGTTTC A A A CCCCATCGI A A A A A A A A A	CCTCATC T T C C C C C C	570 STTGCT 630 CTTTCC	GCA/ G	C C C VAAA G G G	TAC T T T T T	CCTACG T T T T CTATGG	590 GATGGA C C C 650 GAGTGG	AATTGC C C C	CACC T T T AGTC C C
adw2 : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adr(1): adr(2): ayr : ayw(1): ayw(2):	550 CAAGGCAACTO A C A C A C A C A C 610 TGTATTCCCAT	TATGTTTC A A A CCCCATCGI A A A A A A A A	CCTTGGGG	570 STTGCT 630 CTTTCC 690	G G G	C C C AAAA' G G G G	TAC T T T T T	CCTACG T T T T CCTATGG	590 GATGGA C C C C GAGTGG 710	AATTGO C C GCCTCZ	CACC TTTT T AGTC C C
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adr(1): ayr (2): ayw(1): adw2 : adw : adw : adw : adr(1): adw : adw : adw : ayw(2):	550 CAAGGCAACTO A C A C A C A C A C 610 TGTATTCCCAT	TATGTTTC A A A CCCCATCGI A A A A A A A A A	CCTTGGGC	570 STTGCT 630 CTTTCC 690 FGCCA	G G G FTTC	C C Q Q Q Q Q Q G G G G G G G G G	TAC T T T CAG	CCTACG T T T T CCTATGG	590 GATGGA C C C GAGTGG 710 GTAGGG C	AATTGC C C GCCTCZ	CACC TT TT T AGTC C C C C

		130						750						77	0		
adw2 : adw :	ACTGTTI	GGCTT	TCAG	ста	TAT	[GG]	ATGI	ATGI	GGI	TAT	rgg	GGGC	CA	AGTC	TGTA	CAGCA	Атс
adr(1):				T T												A A	
ayr :																A	
ayw(1): ayw(2):				т Т													
		790					8	B10				_		83	0	st	
adw2 :	GTGAGT	CCTTT	ΑΤΑΟ	CGC	TG	TTA	CCAI	ATTI	TC:	[TT]	CGT	CTCI	[GG	GTAT	ACAT	TTAA	ACC
adw : adr(1):	т		т	т	A							т				G	
adr(2):	T T		T	Т т	A A							T T				G	
ayw(l):	T		Ť	-	••							Ť				Q	
ayw(2):	Т		т							С		т					
		850					1	870						89	90		
adw2 :	CTAACA	AACAA	AAA	GATO	GGG	GTT.	ATT	ccci	raa:	ACT	FCA	TGGG	GCT.	ACAI	TAATI	GGAA	GTT
adw : adr(1):	т	с	с	т		С	С		т				A	TG			
adr(2): avr :	T T	C C	C C	Т Т		C	C C		Т Т				A A	TG TG			
ayw(1):	-		G	-		Ū	č	Т	-	Т	т		T	TG	С	т	
ayw(2):							С	TT	С	т				ΤG	с	т	
		910					ł	930						95	50		
adw2 :	GGGGAA	CTTTGC	CAC	AGGZ	ATC.	АТА	TTG	TAC	AAA	AGA	ГСА	AAC	ACT	GTTI	TTAGI	AAAC	TTC
adw : adr(1):	т	A			A			T'	r	AC		G	A		с		G
adr(2):	T	A	G		A			ç	г	AC		G	A		С	T	G
ayw(l):	AT GT	A	G		~					-			А		0		G
ayw(2):		С		Α	A	С	CA			A		G	Α			1	
	AT GT	C A		A A	A	c c	CA CA		G	A A		G G G	A A	с		1	
	AT GT	C A 970		A A	A	C C	CA CA	990	G	A A		G G	A A	C 10:	10	1	
adw2 :	AT GT CTGTTA	C A 970 ACAGGO	CTA	A A ITGI	A A TT	C C GGA	CA CA AAG	990 TAT	G GTC	A A AAA	GAA	G G TTG	A A TGG	C 10: GTC:	10 TTTTC	GGGCT	TTG
adw2 : adw : adr(1):	AT GT CTGTTA A	C A 970 ACAGGC T A	CTA	A A ITGI	A ATT	C C GGA	CA CA AAG	990 TAT	G GTC	A A AAA	gaa	G G TTG	A A TGG	C 10: GTC:	10	GGGCT	TTG
adw2 : adw : adr(1): adr(2):	AT GT CTGTTA A A	C A 970 ACAGGC T A T C	CCTA	A A ITGI	A ATT	C C GGA	CA CA AAG	990 TAT(G GTC	A A AAA	GAA	G G TTG	A A TGG	C 10: GTC:	10 	GGGCT	TTG
adw2 : adw : adr(1): adr(2): ayr : ayw(1):	AT GT CTGTTA A A A A	C A 970 ACAGGO T A T C T A	CCTA	A A ITG2	A	C C GGA	CA CA	990 TAT	G GTC	A A AAAA C	GAA	G G TTG	A A TGG	C 10: GTC:	10 	GGGCT' T	TTG
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	AT GT CTGTTAJ A A A A	C A 970 ACAGGC T A T C T A	CCTA	A A ITTGA	A ATT	C C GGA	CA CA	990 TAT C	G GTC	A A AAA C C	gaa T	G G TTG	A A TGG	C 10: GTC:	10	GGGCT T T	TTG
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	AT GT CTGTTA A A A A	C A 970 ACAGGO T A T C T A 1030	CCTA	A A TTG2	A	C C GGA	CA CA AAG	990 TAT C .050	G GTC	A A AAA C C	gaa T	G G TTG	A A TGG	C 10: GTC: 10	10 	GGGCT T T	TTG
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 :	AT GT CTGTTA A A A A CTGCTC	C A 970 ACAGGC T A T C T A 1030 CATTT/	CTA:	A A ITGA	A ATT	C C GGA	CA CA AAG 1	990 TAT C .050	G GTC CCT	A A AAAA C C C	GAA T TGC	G G TTG	A A TGG	C 10: GTC: 10 ⁻ PATG	10 TTTTC 70 CATG	GGGCT T T	TTG
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw : adw : adr(1):	AT GT CTGTTAJ A A A CTGCTC C	C A 970 ACAGGO T A T C T A 1030 CATTT/ T	CTA:	A A ITTG <i>i</i>	A ATT GTG	C C GGA GAI C	CA CA AAG 1	990 TAT C 050 C	g gtc cct	A A AAAA C C TAA G	GAA T TGC	G G G	A A TGG	C 10: GTC: 10 PATG	10 FTTTO 70 CATG	GGGCT T T	TTG AAG T
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adw : adr(1): adr(2):	AT GT CTGTTAJ A A A CTGCTC C C C	C A 970 ACAGGO T A T C T A 1030 CATTTA T T T	CTA:	A A ITTGI	A ATT GTG C	C C GGA GAI C C	CA CA AAG 1	990 TAT C 050	g gtc. cct	AAAA AAAA C C TAAA G G G G	GAA T TGC	G G TTG	A A TGG TGT A	C 10: GTC: 10 PATG	10 FTTTC 70 CATG	GGGCT T T FATAC	TTG AAG T
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adw : adr(1): adr(2): ayr : ayr : ayr (1):	AT GT CTGTTAL A A A A CTGCTC C C C C C	C A 970 ACAGGC T A T C T A 1030 CATTT/ T T T	CCTA!	A A TTGA	A A GTG C	C C GGA C C C T	CA CA AAG 1 VATC	990 TAT C 050 CCTG	G GTC CCT G	A A A A A A A A A A A A A C C C C C C C	GAA T TGC	G G TTG	A A TGG TGT A A	C 10: GTC: 10 ZATG	10 TTTTC 70 CATG	T T T T T T T T	TTG AAG T T T T

7

	1090		1110	1130	
adw2 :	CTAAACAGGCTTTC	ACTTTCTCGCC	CAACTTACAAGG	CCTTTCTAAGTAAACAGTACAT	GA
adr(1): adr(2): ayr : ayw(1): ayw(2):	G G G G G G T		T T	GT C A C GT A TC GT A TC GT A C GT A C	C A
	1150		1170	1190	
adw2 :	ACCTTTACCCCGTTC	CTCGGCAACG	GCCTGGTCTGT	GCCAAGTGTTTGCTGACGCAAC	ċċ
adw : adr(1): adr(2): ayr : ayw(1): ayw(2):		с с с с с с с	TAC TAC TAC A A		
	1210		1230	1250	
adw2 : adw :	CCACTGGCTGGGGCT	TGGCCATAGG A	CCATCAGCGCA	TGCGTGGAACCTTTGTGGCTCC	гс
adr(1): adr(2): ayr :	A A G T	.	G G G	1	
ayw(1): ayw(2):		T G		C G	
	1270		1290	1310	
adw2 : adw :	TGCCGATCCATACTO	CGGAACTCCT	AGCCGCTTGTT	TTGCTCGCAGCCGGTCTGGAGC	AA
<pre>adr(1): adr(2): ayr : avw(1):</pre>			A A A	GA	G
avw(2) .				А	
uyw(2).				A A	
uyw(2).	1330		1350	A A 1370 X ⁱ	
adw2 : adw : adr(1):	1330 AGCTCATCGGAACTC	ACAATTCTGT	1350 CGTCCTCTCGC	A A 1370 X ⁱ GGAAATATACATCGTTTCC <u>ATG</u> A C C C C	GC
adw2 : adw : adr(1): adr(2): ayr :	1330 AGCTCATCGGAACTC A T G A T C A T C	GACAATTCTGT C G C A C A	1350 CCGTCCTCTCGC T T T T T T	A A 1370 X ⁱ GGAAATATACATCGTTTCC <u>ATG</u> C C C C C C C C C C	GC
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	1330 AGCTCATCGGAACTC A T G A T C A T C CA T G CA T G G	SACAATTCTGT C G C A T C T C	1350 CCGTCCTCTCGC T T T T T T T A C T T C	A A 1370 X ⁱ GGAAATATACATCGTTTCC <u>ATG</u> C C C C C C C C C C C A	GC
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	1330 AGCTCATCGGAACTO A T G A T C A T C CA T C CA T G CA TC G G 1390	GACAATTCTGT C G C A T C T C T C	1350 CCGTCCTCTCGC T T T T T A T A T T C 1410	A A 1370 X ⁱ GGAAATATACATCGTTTCC <u>ATG</u> C C C C C C C C C C C C A 1430	GC
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 :	1330 AGCTCATCGGAACTO A T G A T C A T C CA T C CA T C CA T C CA TC G G 1390 TGCTAGGCTGTACTO	GCCAACTGGAT	1350 CCGTCCTCTCGC T T T T T T T A C T T C 1410 CCCTTCGCGGGA	A A 1370 X ⁱ GGAAATATACATCGTTTCC <u>ATG</u> C C C C C C C C C C C C C C C A 1430 CGTCCTTTGTTTACGTCCCGTC	GC .
adw2 : adw2 : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adw2 : adw2 : adw2 : adr(1): adr(2): ayr(1): ayw(1): ayw(1):	1330 AGCTCATCGGAACTO A T G A T C A T C CA T G CA T G CA TC G G 1390 TGCTAGGCTGTACTO C G G G G G G G G G G	SACAATTCTGI C G C A T C T C SCCAACTGGAT	1350 CCGTCCTCTCGC T T T T T T T A C T T C 1410 CCCTTCGCCGGGA G G G G G G G G G G G G G	A A 1370 x ⁱ GGAAATATACATCGTTTCC <u>ATG</u> C C C C C C C C C C C C A 1430 CGTCCTTTGTTTACGTCCCGTCC C C C C	GG .

		1450				14	70			1490		
adw2 :	CGCTG.	ААТСССО	GCGG	GACG	ACCO	CTCTCG	GGGCC	GCTTO	GGACTO	TCTCGTC	CCCTTO	стсс
adw : adr(1): adr(2): ayr : ayw(1): ayw(2):		т				G G T T	T T	T T T	C G C	AC AC AC	ΤG	TT T T
		1510				15	30			1550		
adw2 : adw : adr(1): adr(2): avr :	GTCTG C T A	CCGTTC	CAG G G	CCGA	.ccac	GGGGCG	CACCT	CTCTT	TACGCG	GTCTCCC	CGTCTO	STGC
ayw(1): ayw(2):		Т	GA GA							A A		
		1570				15	90			1610		
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	CTTCT	CATCTG	CCG	GTCC A A A A A A	GTG	IGCACTT	CGCTT	CACC	rctgcad	CGTTGCAT C C A C C C	CGGAGA(C	CCAC
	t	1630				16	50			1670		
adw2 :	CG <u>TGA</u>	ACGCCC	ATC.	AGAT	· CCTC	GCCCAAG	GTCTT	ACAT	AAGAGG <i>I</i>	CTCTTG	ACTCC	CAGC
adw : adr(1): adr(2): ayr : ayw(1): ayw(2):		G A	C C C C A	G G A CAT	T T T AT T		с	с			T T T T	тт
		1690				17	10			1730		
adw2 : adw :	AATGT	CAACGA	CCG.	ACCI	TGAC	GCCTAC	TTCAA	AGAC	IGTGTG1	TTAAGG	ACTGGG.	AGGA
adr(1): adr(2): ayr : ayw(1): ayw(2):	С	A				А А А А А			T T T	А А А А А		
		1750				17	70			1790		
adw2 : adw : adr(1): adr(2): ayr : ayw(1):	GCTGG T T T T T	GGGAGG	AGA	OATT	GTT	AAAGGTC T A	TTTGT	C C C C	GGAGGC	IGTAGGCI	ACAAAT T T T	rggt
ayw(2):	Ť				A			~			Ť	

PC-1X°. X°. X°. X°.adw : adr (1): TTTTATTadr (1): adr (2): TTTTATTadr (2): ayw (2):TTCATTadw (2): adw (2):TTCATTadw (1): adr (2): adw (2):ACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGCGCATGGACCTTAT adw (2):ACTGTTCAAGCCTCCAAGCTGCCTAGGTGGCTTGGGCGCATGGACCTTAT GGadw (1): adr (2): adw (2):ACTGTTCAAGCTCCTGTGGGGGTTACTCTGTGGTTTTTGCCTTCTGACCTTCTTTCCTTCC		181	0	i	1830		+			1850)			
adwiCredecadeacadactariaATTadr(1):TTATTadr(2):TTCATTayw(1):TTCATTayw(2):187018901910adw :adr(2):TTTTadw :adr(2):CGG.adw :adr(2):CGG.adw (2):193019501970adw (2):TTTGGadw (2):TTTGTadw (2):AAAGAATTGGAGCTACTGTGGAGTACTCTCGTTTTGCCTTCTGACTTCTTTCT	- dr. 2	amagaga ag	. I			mamaaa	Х ^с .	men	mon			• • •	-	~.
adr (1):TTATTadr (2):	adw :	CIGCGCACC.	AGCACC	<u>IIG</u> CAACI	ITTICACC		TAA	II CA	ICI	CIIG	TAC	AI	310	CC
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	adr(1):	ጥጥ								Α	т			т
ay T : TT C A T T T ayw(1): TT C A T T T ayw(2): TT T T T ayw(2): CATGGTCAAGCCTCCAAGCTGTGCCTTGGGGCATGGACATGACCCTTAT adw : adr(1): C G G ayr(2): C T T G G G G ayw(2): C T T G G G G A ayw(2): C T T G G G G A ayw(2): C T T A G C A G C C A A C A C ayw(2): T A G C A G C C C AT A C ayw(2): C C A G C C A C A C A C A C A ayw(2): C C T G G C A C A C A C A C A C A C A C A C A	adr(2):			T						A	Ť			T
ajw(1): ajw(2): $T T T$ $yw(2): 1870 1890 1910 c^{i} 1910 c^{i} T T adw : adr(1): adw : adr(2): adr(2): G C G G G G G G G G G G G G G G G G G $	ayr :	TT			С					Α	т			т
ayw(2):TT187018901910 $adw2:$ ACTGTTCAAGCCTCCAAGCTGTGCCTTGGGGGCATGGGCATGGACATGACCCTTAT $adw:$ $adr(1):$ C $adr(2):$ G $adr(1):$ C $ayw(2):$ T19301950 $adw2:$ AAAGAATTTGGAGCTACTGTGGAGTTACTCTGGTTTTGCCTTCTGACTTCTTTCCTTCC	ayw(1):										т			т
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ayw(2):										т			Т
$\begin{array}{cccccccccccccccccccccccccccccccccccc$											5			
adw2 : ACTGTTCAAGCCTCCAAGCTGTGCCTTGGGGGGTTTGGGGCATGGACATTGACCCTTAT adw : adr(1): G adr(2): G adr(2): G adw(2): C 1930 1950 1970 adw2 : AAAGAATTTGGAGCTACTGTGGGGGTTACTCTGGTTTTTGCCTTCTGACTTCTTTCCTTCC		187	0		1890			~	i	1910)			
adv : adv : adr (1): adr (2): dr (2): G G $gr (2):$ G G $gr (2):$ G G $gr (2):$ G G $gr (2):$ G G G $gr (2):$ G	adw2 :	ACTGTTCAA		AGCTGTG	CCTTGGGT	GGCTT	rgge	GCA	TGG	ACAT	TG	ACC	CTT	АТ
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	adw :													
adr(2):Gayw(1):Cayw(2):T19301950adv2:AAAGAATTTGGAGGTACTGTGGAGTTACTCGTGTTTTTGCTTCCTGACTTCTTTCCTTCC	adr(1):												G	
ayr:Gayw(1):Cayw(2):T19301950adw2:AAAGAATTTGGAGCTACTGTGGGAGTTACTCTCGTTTTTGCCTTCTGACTTCTTCCTTC	adr(2):												G	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ayr :												G	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ayw(l):										С	_		
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ayw(2):											T		
adw2 : AAAGAATTTĞGAĞCTACTĞTĞGAĞTTACTCTCĞTTTTĞCCTTCTĞACTTCTTTCCTTCC adw : adr(1): T T T G T adr(2): AT T T G T ayr : T T T A ayw(1): ATC C T G G G A adw2 : GTCAĞAĞATCTCCTAĞACACCĞCCTCAĞCTGTĞTACĞAĞAAĞCCTTAĞAĞTCTCCTAĞA adw : AC 1990 2010 2030 adw2 : GTCAĞAĞATCTCCTAĞACACCĞCCCTCAĞCTCTĞTATCĞAĞAAĞCCTTAĞAĞTCTCCTAĞA adw : AC adr(1): A TC C C T G G G A adr(2): A TC T T G G G G A ayr : A TC C T T G G G G A ayr : A TC C T T G A ayw(1): AC T T G adw2 : CATTĞCTCAČCTCAČCATCTCĞAĞCAĞCCATTCTCTĞCTĞĞĞĞĞAATTĞATĞ adw : AC 2050 2070 2090 adw2 : CATTĞCTCAČCTCAČCTCAČCATCTĞĞCAĞCCATTCTCTĞCTĞĞĞĞĞĞAATTĞATĞ adw : T A T G T T G adr(2): T A T G T T G A ayr(2): T A T G T T G A ayr(2): T A T G T T G adr(2): T A A T G T T G adr(2): T A A T C C A ayr(2): T A A T C C A 2110 2130 2150 adw2 : ACTCTAĞCTÄCCTĞĞĞTAĞTATATTTĞĞAAĞATCCAĞCATCTAĞĞĞATCTTĞTAĞTA adw : C C C A T A C adr(2): A T G C A G C C A T A C ayr(1): A T G C A G C C A T A C ayr(2): GT G C C A C C A T A C ayr(2): GT G C A C A C C ayr(2): C C A C A C C C A C A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C ayr(2): C C A C A C C A C A C C C A T A C ayr(2): C C A C A C C A C A C C C A T A C ayr(2): C C A C A C C C A T A C C A C A C C C A C A		193	0		1950					1970)			
adw2:AAAGAATTTGGAGCTACTGTGGAGTTACTCTCGTTTTGCCTTCTGACTTCTTTCCTTCC			•		•									
adw : adr (1):TTTTadr (2):ATTGTayw (2):TTTTayw (2):IATTGadw :ACCTGGadw :ACCTGGadw :ACCTGGadw :ACCTGGadw :ACCTGGadw :ACTTGGadw (2):ACTTAG205020702090adw :CATGAadr (1):TATGTadw :ATGTTGadw :ATGTTGadw :ATGTTGadw :ATGTTGadw :ATGTTGadw :ATCACAadw :AGCAACadw :AGCAACadw :AGCAACadw :AGCAACadw :AGCCAAadw :AGCCAAadw :A <td>adw2 :</td> <td>AAAGAATTT</td> <td>GGAGCTA</td> <td>CTGTGGA</td> <td>GTTACTCI</td> <td>CGTTT</td> <td>TTGC</td> <td>CTT</td> <td>CTG</td> <td>ACTI</td> <td>CT</td> <td>TTC</td> <td>CTT</td> <td>cc</td>	adw2 :	AAAGAATTT	GGAGCTA	CTGTGGA	GTTACTCI	CGTTT	TTGC	CTT	CTG	ACTI	CT	TTC	CTT	cc
adr(1):TTTTadr(2):ATTGTayr:TTTayw(1):TTTayw(2):I99020102030adw2:GTCAGAGAATCTCCTAGACACCGCCTCAGCTCTGTATCGAGAAGCCTTAGAGTCTCCTAGAAadw2:GTCAGAGAATCTCCTAGACACCGCCTCAGCTCTGTATCGAGAAGCCTTAGAGTCTCCTGAGAadw2:GTCAGAGAATCTCCTAGACACCGCCTCAGCTCTGTAGAGCCTTAGAGTCTCCTGAGGadw1:ATGadw1:ATGadw2:CATTGCTCACCTCACCTACTGCATCGCAGCCAAGCCATTCTCTGGCGGGGGAATTGATGadw1:TATadr(2):TATadr(2):TATadr(2):TATadr(2):TATadw2:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCTACCTGGGTGGGTGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCTACCTGGGTGGGTGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCTACCTGGGTGGGTGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCTACCTGGGTGGGTGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCCACCTGGGTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCCACCTGGGTGGGTGGTGGTAATAATTTGGAAGATCCAGCCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCCACCTGGGTGGGTGGTGGTAATAATTGGAAGATCCACCGCCACCTGAGGATCTTGTAGTAadw2:ACTCTAGCCACCTGGGGGGGGGGGGAATTAATTGGAAGGATCCAGCACTTGGGGATCTTGTAGTA<	adw :		-			-								-
dur (z):A11GIayrTTTTayw(1):TTTayw(2):I99020102030adw2:GTCAGAGAGATCTCCTAGACACCGCCTCAGCTCTGTATCGAGAGGCCTTAGAGTCTCCTGAGadw1:AACadr(1):ATCCadr(2):ATCTAdr(2):ATCTAdr(2):ATCTAdr(2):AA205020702090adw2:CATTGCTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGGAATTGATGadw2:CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGGGAATTGATGadw2:CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTGCTGGGGGGGAATTGATGadw2:TATadr(2):TAtdr(2):TAtdr(2):TAtdr(2):TAtdr(2):TAtdr(2):TAtdr(2):TAtdr(2):TAtdr(2):ACadw2:ACTCTAGCTACCTGGGTGGGTGATATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCTACCTGGGTGGGTGATATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCTACCTGGGTGGGTGATATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCTACCTGGGTGGGTGAGATATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCTACCTGGGTGGGTGAGATATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw2:ACTCTAGCtdr(2):Atdr(2):Atdr(2): <t< td=""><td>adr(1):</td><td></td><td>ני</td><td></td><td></td><td>T</td><td></td><td></td><td></td><td></td><td></td><td></td><td>c</td><td>T</td></t<>	adr(1):		ני			T							c	T
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	aur (2):		л. Т			Ť							9	T
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	avw(1):			-		-								Ā
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ayw(2):													A
199020102030adw2:GTCAGAGATCTCCTAGACACCGCCTCAGCTCTGTATCGAGAAGCCTTAGAGTCTCCTGAGadw:ACadr(1):A TCCTadr(2):A TCTTGGGayr:A TCCTGGayw(2):ACTTAG205020702090adw2:CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGGGGGGGG														
adw2 :GTCAGAGAGATCTCCTAGACCGCCTCAGCTCTGTATCGAGAGCCTTAGAGTCTCCTGAG adr(1):adr (1):A TCCTGGAadr (2):A TCCTTGGGAayr :A TCCTTGGAAayr :A TCCTTGGAayr :A TCCTTGAAayw(2):ACTTAGCAadr (1):TATGTTGadr (2):TATGTTGadw ::adr (2):TATCAayw(1):TATGCAadw2 :ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAGTAATCadw2 :ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTACACadw2 :ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTACACadw2 :ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTACACadw2 :ACTCTAGCCA GCCAACadw2 :ACTCAGGTGCCA GCACAadw2 :ACTCAGGTGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG		199	0		2010					2030)			
adw:ACadr(1):A TCCTGGGadr(2):A TCTTGGGAayr(2):A TCCTTGGGAayw(1):ACTTGGaAAadw2:CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGGCTGG	adw2 :	GTCAGAGAT	· CTCCTAG	ACACCCC	CTCAGCTC	TGTAT	CGAG	AAG	ССТ	י. ערעי	GTO	TC	CTG	AG
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	adw :	AC	01001110		010110010									
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	adr(1):	A TC	С		T		G	G					G	A
ayr:A TCCTGGGGAayw(1):ACTTTGGAayw(2):ACTTAGC205020702090adw2:CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGGAATTGATGadw1:TATGadr(1):TATGGadr(2):TATGTadr(2):TATGAayw(1):TATCAayw(2):TAAGC211021302150IIIadw2:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAIadw2:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAIadw2:ACTCTTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAIadw2:ACTCTTAGCT ACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAIadw2:ACTCTTAGCT ACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAIadw2:ACTCTTAGCT ACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAIadw2:ACTCTTAGC A GCCadw2:ACTCTTAGC A GCCadw2:A G CA GCadw2:A G CA G </td <td>adr(2):</td> <td>A TC</td> <td>т</td> <td></td> <td>т</td> <td></td> <td>G</td> <td>G</td> <td></td> <td></td> <td></td> <td></td> <td>G</td> <td>A</td>	adr(2):	A TC	т		т		G	G					G	A
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ayr :	A TC	С		т		G	G					G	A
ayw(2):ACTTTG205020702090 $adw2$:CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGGAATTGATG $adw1$: $adr(1)$:TATGTTG $adr(2)$:TATGTTGA ayr :TATGTTGA $ayr(1)$:TATCAAGCA $ayw(2)$:TAAGCAACAA $adw2$:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAACCCAAC $adw2$:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTACACAC $adw2$:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAACAC $adw2$:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAACAC $adw2$:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAACAC $adw2$:ACTA GCCA TCAC $adw2$:AA GCA GCCA TC $adw2$:A GCA GCCA TCAC $adw2$:A GCA GCCA TCAC $adw2$:A GCA GCC<	ayw(l):	AC	Т	T			G							
205020702090adw2 :CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGGGAATTGATGadw :adr(1):TAadr(2):TATTATGadr(2):TATATGTTATGadyw(1):TATATCAayw(2):TAACCAACCACCCACCCACCCACCCACCCACCCACCCACCAC<	ayw(2):	AC	T	'I' A			G							
adw2 : CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGGGAATTGATG 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 A ayw(1): T A T G T T G ayw(2): T A A G C A 2110 2130 $2150adw2 : ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAadw :adr(1): A T G C A G C C A A Cadr(2): A G C A G C C AT A Cayr : A G C A G C C AT A Cayr : A G C A G C C AT A Cayw(2): GT G A C A Cayw(2): GT G A C A C$		205	0		2070					2090)			
adw2:CATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGGGAATTGATGadw1:TATGTGadr(2):TATGTTGAadr(2):TATGTTGAayrTATGTTGAayw(1):TATCAACCA211021302150	-		•	•						•				•
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	adw2 :	CATTGCTCA	CCTCACC	CATACTGC.	ACTCAGGC	CAAGCC	ATTC	TCT	'GC'I	GGGG	GGI	AA.I.	IGA	TG
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	adr(1):	Ţ		۵		T		Ġ	т		т	G		
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	adr(2):	Ť		A		Ť		G	Ť		Ť	G	A	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	avr :	Ŧ		A		Ť		G	Ŧ		Ŧ	G		
ayw(2):TA A GC A211021302150 $adw2$:ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTA adw :CC A A C $adw(1)$:A T G CA G $adr(2)$:A G CA G $ayw(1)$:GTG A C A C $ayw(2)$:GGAT C	ayw(1):	т				A		т				С	A	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ayw(2):	Т				Α	A	G				С	A	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$		211	0		2130					2150)			
$\begin{array}{cccccccccccccccccccccccccccccccccccc$			•	•							r			
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	adw2 :	ACTCTAGCT	ACCTGGG	TGGGTAA	TAATTTGO	AAGAT	CCAG	CAT	CTA	AGGGZ	TC	TTG	ΓAG	TA
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	adw :			0 — 14	C	2			C	Α	• -	A		C
aor(2):AGCAGCAC ayr :AGAGCCAAC $ayw(1)$:GTGACAC $ayw(2)$:GGACCAC	adr(1):	ATG C		AG		C			C		AT	A		C
ayr: A G C ATA C ayw(1): GT G A C A C ayw(2): GG AT C C A C	adr(2):	A G C		AG		c			C		AT	A		C
ayw(2): GG AT C C A C	ayr :	AGC		A G		C		C	C	Δ	C	A		C
	ayw(1):			GG			Z	T	С	А	č	A		č

		2170						2190	D					2210			
adw2 : adw :	AATTZ	ATGTTA	ΑΤΑ	TA	ACGI A	GG	GTT	(AAA)	GA:	FCA	GGCA	АСТА	TTG	rggttt	САТА	TAT	ст
adr(1):	GC	С	GI	C	TA		CC	j.	A		Α					т	С
adr(2):	GC	С	GT	C	TA		CC	2	A		A				С	т	С
ayr :	GC	С	GI	C	TA		CC	1	A	т	Α				C	\mathbf{T}	С
ayw(l):	G	С	С		TA		CC		т			C			С	т	
ayw(2):	G	С	С		TA		CC	2	АТ						С	т	
		2230						225	0					2270			
adw2 :	TGCC	TTACTT	TTG	GAA	GAG	GA	CTG	FACT'	TG	AAT	ATTT	rggtc	TCT	TTCGGA	GTGT	GGA	TT
adw :																	
adr(1):	т					А		Т		G		G	;	т			
adr(2):	2.0					A		TT	G	G		A		Ť			
avr :						Δ		C	-	G		ē		т Т			
ayw(1)	ጥ	C				Δ	Δ	TA	Δ	G		e e					
avw(2):	Ť	č				A	A	TA	Α	G		č		T			
~ <u>}</u> .(2).	-	U				••	••					12		-			
		2290			,	•		231 P	0 1					2330			
adw2 :	CGCA	CTCCTC	CAG	ССТ	ATA	GAC	CAC	CAAA	TG	ccc	CTA	TCTT/	ATCA	ACACTI	CCGG	AAA	CT
adw :																	
adr(1):			С	т	С												
adr(2):			С	т	С												
ayr :			С	т	С												
ayw(l):				т								С				G	
ayw(2):				т												G	
		2250							•								
		2350	I					231	0					2390			
adw2 :						•			•			•					•
admin a	ACTG	ጥጥርጥጥል	GAC	GAC	GGG	ACC	CAG	GCAG	GT	CCC	CTA	CAACI	ADGA	ACTCCC	ידרכנ	CTC	CC
adw :	ACTG	TTGTTA	GAC	GAC	GGG	ACC	GAG	GCAG	GT	ccc	CTA	GAAG	AAGA	ACTCCC	TCGC	ССТС	CGC
adw :	ACTG	TTGTTA	GAC	GAC	GGG	ACC	GAG	GCAG	GT	ccc	CTA	GAAG	AAGA	ACTCCC	TCGC	ССТС	CGC
adw : adr(1):	ACTG	TTGTTA	GAC	GAC •	GGG	АСС • •	GAG	GCAG	GT	ccc	CTA	GAAG	AAGA	ACTCCC	TCG	ССТС	CGC
adw : adr(1): adr(2):	ACTG	TTGTTA	GAC	GAC •	GGG	ACC	GAG	GCAG	GT	ccc	CTA	GAAG	AAGA	ACTCCC	TCGO	ССТС	CGC
adw : adr(1): adr(2): ayr : ayw(1):	ACTG	TTGTŦA	GAC	GAC • •	GGG)	ACC 	GAG	GCAG	GT	ccc	CTA	GAAG	AAGA	ACTCCC	TCG	ССТС	CGC
adw : adr(1): adr(2): ayr : ayw(1):	ACTG	TTGTTA	GAC	GAC	GGG2	ACC	GAG	GCAG	GT	cec	CTA	GAAG	AAGA	ACTCCC	TCG	ссто	CGC
adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	ACTG	TTGTTA	GAC	GAC • • •	GGG	ACC	GAG	GCAG	GT	ccc	CTA	GAAG	AAGA	ACTCCC	TCG	CCTC	GC
adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	ACTG	TTGTTA 2410	(GAC)	GAC	GGG	ACC	GAG	GCAG 243	GT	ccc	CTA	GAAG	A AGA	ACTCCC 2450	TCG	ссто	GC
adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	ACTG	TTGTTA 2410	(GAC)	GAC • • •	GGG	ACC	GAG	GCAG 243	0	ccc	CTA	GAAG	AAGA	АСТССС 2450	TCG	CCTC	ct.
adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 :	ACTG AGAC	TTGTTA 2410 GCAGAT	GAC	GAC		ACC	GAG	GCAG 243 GCAG	O AA	.GA1	CTA	GAAG AATC	AAGA FCGG	ACTCCC 2450 GAATCI	TCGO	CCTC	ct.
adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 :	ACTG AGAC	2410 GCAGAT	GAC	GAC		ACC	GAG	GCAG 243 GCAG	GT ;AA	.GA1	CTA	GAAG AATC	AAGA TCGG	2450 GAATCI	TCGO	CCTC	ct.
adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adw : adr(1):	ACTG AGAC	2410 GCAGAI A G	GAC	GAC	GGG)	ACC	GAG	GCAG 243 GCAG	GT	.GA1	CTA	GAAG AATC	AAGA TCGG	2450 GAATCI	TCGO	CCTC	cGC
adw : adr(1): adr(2): ayw(1): ayw(2): adw2 : adw2 : adw : adr(1): adr(2):	ACTG	2410 GCAGAI A G A G	GAC	GAC		ACC	GAG	GCAG 243 GCAG	O	.GA1	CTA	GAAG AATC	AAGA TCGG	2450 GAATCI	TCGC	CCTC C FGT	cGC
adw : adr(1): adr(2): ayw(2): ayw(2): adw2 : adw2 : adw : adr(1): adr(2): ayr :	ACTG	241C GCAGAI A G A G A G A G A G	GAC	GAC • • • • • • • • • • • • • • • • • • •	GGG 	ACC	GAG	GCAG 243 GCAG	GT	.GA1	CTA	GAAG/	AAGA FCGG	2450 GAATCT	TCGC	CCTC C FGT	ct.
adw : adr(1): adr(2): ayw(1): ayw(1): ayw(2): adw2 : adw2 : adr(1): adr(2): ayr : ayw(1):	ACTG	241C GCAGAI A G A G A G A G A G A G	GAC	GAC • • • • • • • • • • • • • • • • • • •		ACC	CGTC	GCAG 243 GCAG	O ,	.GA1	CTA	GAAG	AAGA FCGG	ACTCCC 2450 GAATCI	TCAA	CCTC C FGT	ct.
adw : adr(1): adr(2): ayw(2): ayw(1): ayw(2): adw2 : adw2 : adr(1): adr(2): ayr(2):	ACTG	241C GCAGAI A G A G A G A G A G A G A G A G	GAC	GAC CAT A A A A A A A A A A		ACC	GAG	GCAG 243 GCAG	O	GAI	CTA	GAAG	AAGA TCGG	2450 GAATCI	TCAA	CCTC C FGT	ct. FAG
adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adw : adr(1): adr(2): ayr : ayr(1): ayw(1): ayw(2):	ACTG	2410 GCAGAI A G A G A G A G A G A G A G A G A G) CCTC	GAC CAT A A A A A A A A A	GGG	ACC	GAG	GCAG 243 GCAG 249	O AA	GAI	CTA	GAAG	Y AGA	2450 GAATCT 2510	TCAA	CCTC C IGT	ct. TAG
adw : adr(1): adr(2): ayw(2): ayw(1): ayw(2): adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	ACTG	2410 GCAGAI A G A G A G A G A G A G A G C TTTC		CAT A A A A A A A A A A A A A A A A A A	GGGG	ACC	CGTC	243 2243 GCAG 249	0	GAT	CTA	GAAG	TCGG	2450 GAATCT 2510	TCAA	CCTC CTGT TGT	TAG
adw : adr(1): adr(2): ayw(2): ayw(1): ayw(2): adw2 : adr(1): adr(2): ayr(2): ayw(1): ayw(2): adw2 : adw2 : adw2 :	ACTG AGAC TATT	2410 GCAGAI A G A G A G A G A G A G A G A G CCTTG	GAC CCTC	CAT A A A A A A A A A A A A A A A A A A	GGG,	ACC	GGGG	243 GCAG 249 249 AACT	O AA	GAT	CTA	GAAG AATC	AAGA TCGG ATTC	2450 2450 2510 2510	TCAA'	C G G G G G G G C C C C C C C C C C C C	cgc ct. FAG
adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw : adr(1): ayr : ayr(1): ayw(2): adw2 : adw2 : adw2 : adw :	ACTG AGAC TATT	2410 GCAGAI A G A G A G A G A G A G A G A G A G A G	GAC	GAC	GGG,	ACC	GAG CGTC GGGA C	243 GCAG 249 249 AACT	O AA	GAT	CTA CCTC	GAAG AATC	NAGA FCGG	2450 2450 2510 2510	TCAA CAAT	C C FGT ACC	TAT
adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw : adr(1): ayr (1): ayr(2): adw2 : adw2	ACTG AGAC TATT	2410 GCAGAT A G A G A G A G A G C CTTGO	GAC	GAC	GGGG	ACC	GAG CGTC GGGA C	243 GCAG 249 249 249	O AA PO	GAT	CTA CTC GGGG	GAAG AATC	NAGA TCGG	2450 GAATCI 2510 CCTCTAC	TCAA CAAT T	(FGT)	cGC ct. IAG
adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw : adr(1): ayr (2): ayr (2): adw2 : adw2 : adw	ACTG AGAC TATT C	2410 GCAGAI A G A G A G A G A G A G A G C CTTGO) CCTC J	GAC 	GGG	ACC	GAG CGTC GGGA	243 GCAG 245 249 249	GT AA	GAT	CTA CCTC GGGG	GAAG AATC	NAGA ICGG ATTC	2450 GAATCH 2510 CCTCTAC T	CAGT. T T T	CCTC (FGT? ACC?	TAT GGC
adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw : adr(1): ayr : ayw(1): ayr : ayw(1): ayr (2): adw : adr(1): adr(2): adw : adr(1): adr(2): adr	ACTG AGAC TATT C C C	2410 GCAGAI A G A G A G A G A G A G C CCTTGO) CCTC) ;	GAC	GGG,	ACC	GAG CGTC	243 GCAG 249 AACT	GT AA	GAT GAT	CTA CCTC GGGG F F F	GAAG AATC	NAGA FCGG	2450 GAATCT 2510 CCTCTAC T T	CCAA CCAA T T T T	CCTC (FGT ACC	TAT G G G
adw : adr(1): adr(2): ayw(2): ayw(1): ayw(2): adw : adr(1): adr(2): ayw(1): ayw(2): adw : adr(1): adr(2): adw : adr(1): adr(2): ayr : ayw(1): ayr : ayr : ayw(2):	ACTG AGAC TATT C C C	2410 GCAGAI A G A G A G A G A G C CTTGO) CCTC J	CATI CATI A A A A A A A A A A A	GGG		GAG CGTC GGGA	243 GCAG 249 249 AACT	O AA	GA1	CTA CCTC GGGG F F F F	GAAG AATC	NAGA TCGG	2450 2450 2510 2510 2510 2517 7 T	CAGT. T T T T	C C FGT ACC	TAG GGGG

	2530								25	50						2	570			
adw2 :	CTTT	AATCO	CTG	AAT	GGC	AAA	CTC	CT.	FCC	TTT	CC	CTA	AGA	TTC	ATT	TAC	AAG	AGG	ACA	TTAT
adw :	c			c	00			c					c				c			
adr(1):	C		С	G				c				С	c				G			
ayr :				т	A			С				С	С				G			
ayw(1):		~	c	T	A		A	A	T				T	A			CC	A		
ayw(2):		C	C	т	А		A	C	Т				т	A			CC	A		
		259	90						26	LO						2	630			
adw2 :	TAAT	AGGT	TC	AAC	AAT	TTG	TGG	GCO	ссто	TC	AC	CTG	ГАА	ATG	AAA	AGA	GAA	GAT	TGA	AATT
adw :		۵				Δ				m		Δ	m			Δ	G		Δ	
adr(2):		Â				A				Ť		G	Ť			A	G		A	
ayr :		A				A				G		A	т			Α	G		Α	
ayw(1):	CA	AA	G		G		A		A			A	T		G	A		~	C	
ayw(2):	CA	AA	G				A		A			A	C		G	A		U	C	•
		265	50 •						267	70 •				•		2	690			
adw2 : adw :	AATT	ATGCO	CTG	CTA	GAT	TCT.	ATC	CTI	ACCO	CAC	AC	TA	AAT.	ATT	FGC	ССТ	TAG.	ACA	AAG	GAAT
adr(l):					G				Α	TT		С					G	т		С
adr(2):					G				Α	TT		_					-			С
ayr :	~			~	G	m		х	A	TT		C			2	A	G	m	c	C
ayw(1): avw(2):	G		A	C	G	T		A	AGG	GTT		c			A	A	G	T	G	T
		271	LO						27:	30						2	750			
			•		~ ~ ~		100				~ ~					~~`				
adw2 : adw :	TAAA	CCLLL	VI'I'	ATC	CAG	ATC.	AGG'	TAC	51"17	AT (E	Υ.L.Lλ	AC.L.	rcci	AAA	CCA	GAC.	ATT	A'1''1'	TACA
adr(1):					т	A	т	С						A		т	G			
adr(2):		G			т	Α	т	С						A		т	G			
ayr :		G			т	AT	T	С						A		Т	G	~		
ayw(1): ayw(2):						A AT	TT									T T		C		
		277	70						279	90						2	810			
- d	ma cm /	~~~~~			000		mmor			in	20				~~~	C m2		~~~	~ m	•
adw :	IACI		GA	AGG	CIG	GIA	IIC.	IAI	CA12	AGG	A	3995	IAA	CCA	LAC	GIA	GCG	CHI	CAT	1116
adr(1):		G				С				AZ	Ā	Α		т		С		т		
adr(2):		G				С				2	A	Α		т		С		С		
ayr :	<u>c</u>	G			~	С	2 111			2	A	A		T		°,		c		
ayw(1): ayw(2):	c	A			G		AT		с	2	A	A		T		A		c		
		283	20						285	50						2	870			
		202							20:]	PS1	ia						•
adw2 : adw :	CGGG	I'CAC(CAT	ATT	CTT	GGG.	AAC	AAC	SAG	CTA	27	AGC <u>1</u>	<u>ATG</u>	GGA	GGT	TGG	TCA	I'CA	AAA	CCTC
adr(1):	т										G	3	•		• • •	• • •	т	ċ	• • •	••
adr(2):	т											rani					Ť	c		
ayr :	Т																т	С		
avw(1):									in the second											
	T								Т			•••	•••	•••	•••	•••	•••	• • •	•••	• • • •

		289	0 2.1.1.	-				2	910					293	30			
adura .	CC33	ACCON	51-1		CAA	• •••	ጥጥጥ	ישמי	TTCCC	סידי בי	• • •	car	ייידי ג	CTTT	• ••••	rcaz	ە ت ىس	TC.
adw	GCAA	AGGC <u>A</u>	1990	GGAC	GAA	IC.	1110	-19	11000	C	CICIC	3993	77 1	C	100	.cgr	11CA	10
adr(1):	AC									C				C				c
adr(2).	AC																	č
aur (2).	AC																	č
ayw(1):	AC			CA				CA	CCAG								С	č
avw(2):		••		CA				CA	CCAG								č	č
-1.(-)-		••															-	-
		295	0					2	970		-			299	90			
adw2 :	AGTT	GGACC	CTG	CATT	CGG	AG	CCA	ACT	CAAA	CAATC	CAGA	rTG	GGI	ACT:	rca/	CCC	CCGI	'CA
adw :									CI	ł							A	
adr(1):				G													AA	
adr(2):				G													AA	
ayr :		-	-	G	-		-		~~~~							-	AA	•
ayw(1):		T	A	C	A		A	A	CGCZ	7						T	AA	<u>.</u>
ayw(2):		т	A	C	A		A	A	CGCI	7						т	AA	
		301	0					3	030					30	50			
adw2 :	AGGA	CGACT	GGC	CAGC	AGC	ĊA	ACC	AAG	TAGG	AGTGG	GAGC	ATT	CGG	GGC	CAAC	GC	TCAC	cc
adw :		С						G							G			
adr(1):		TC		A	G	А	ТА	G		С	A	С 🗌			G	Т		
adr(2):		TC A		A	.G	A	т	G		С					G	Т		
ayr :		TC		A	G	A	T	G		С			Г	т	G	т		
ayw(1):		AC		A	C		A	G		CT				1	rgg	TT		
ayw(2):		AC		A	C		A	G		CT					ГGG	AT		
		307	0					3	090					31	10			
adw2 :	стес	307 ACACG	0 GCG	GTAT	TTT	GG	GGT	3 GGA	090 GCCC	ICAGG	CTCA	GGG	CA	31 TAT	10 	CCA	CAGI	GT
adw2 : adw :	стес	307 ACACG	0 GCG	GTAT	TTT	GG	GGT	3 GGA	090 GCCC	ICAGG	CTCA	GGG	CA	31 ГАТ	10 IGA	CCA	CAGI	GT
adw2 : adw : adr(1):	стсс	307 ACACG	0 GCG	GTAT C	TTT	GG	GGT	3 GGA	090 GCCC	ICAGG	SCTCA	GGG	CA	31 TAT	10 IGA	CCA A	CAGI	GT C
adw2 : adw : adr(1): adr(2):	СТСС А А	307 ACACG	0 GCG	GTAT C C	TTT	GG	GGT	3 GGA	090 GCCC	ICAGG	CTCA	GGG	CA	31 TAT'	10 IGA	CCA A A	CAGI	C C C
adw2 : adw : adr(1): adr(2): ayr :	CTCC A A A	307 CACACG	0 GCG A	GTAT C C CC	TTT	GG	GGT	3 GGA	090 GCCC	ICAGG	CTCA	GGG	CA	31 FAT	10 TGA	CCA A A A	CAGI	C C C C
adw2 : adw : adr(1): adr(2): ayr : ayw(1):	CTCC A A A A A	307 ACACG G	0 GCG A A	GTAT C C CC CC	TTT	GG	GGT	3 GGA	090 GCCC	TCAGG	CTCA	GGG	CA	31 TAT C	10 IGA AC	CCA A A A AA	CAGI C TT	CGT C C C C C
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	CTCC A A A A A A	307 ACACG G	0 GCG A A A	GTAT C C CC CC CC	TTT	GG	GGT	3 GGA	090 GCCC	ICAGG	CTCA	GGG	CA	31 TAT C C	10 IGA AC AC	A A A A A A A A A A	CAGI C TT GT	C C C C C C C C C C
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	CTCC A A A A A A	307 ACACG G 313	O GCG A A A O	GTAT C C CC CC CC	TTT	GG	GGT	3 GGA 3	090 GCCC 150	ICAGG	CTCA	GGG	CA	31 TAT C C 31	LO IGA AC AC	CCA A A A A A A A A	CAGI C TT GT	C C C C C C C C
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 :	CTCC A A A A A A CAAC	307 ACACG G 313 AATTC	O GCG A A A O	GTAT C CC CC CC CC	TGC	GG	GGT	3 GGA 3 CCA	090 GCCC 150 ATCG	ICAGG	CTCA	GGG AAG	GC.	31 TAT C 31 AGC	LO TGA AC AG 70 CTA	CCA A A A A A A A A A A A A A A A A A A	CAGI C TT GT CCAJ	CCCCCCCCC.
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 :	CTCC A A A A A CAAC	307 ACACG G 313 AATTC	O GCG A A A CTC	GTAT C CC CC CC CC	TTTT TGC	GG	GGT	3 GGA 3 CCA	090 GCCC 150 ATCG	ICAGG GCAGT	CTCA	GGG AAG	GC.	31 TAT C 31 AGC	IO IGA AC AG 70 CTA	CCA A A A A A A A A A A A A A	CAGI C TT GT CCAI	CC CC CC CC CC CC CC
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adw : adw :	CTCC A A A A A CAAC	307 CACACG G 313 CAATTC GCG	O GCG A A A A C TC	GTAT C CC CC CC CC	TTTT TGC T	GG CI	GGT	3 GGA 3 CCA	090 GCCC 150 ATCG	ICAGG GCAGI	CAGG	GGG	GC.	31 FAT [,] C C 31 AGC	LO IGAO AC AG 70 CTA	CCA A A A A A A A A A A A A A A A A A	CAGT C TT GT CCAT	CGT CCCCC CCT
adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw : adr(1): adr(2):	CTCC A A A A A CAAC G G	307 ACACG G 313 CAATTC GCG GCA	O GCG A A A O CTC	GTAT C CC CC CC CC	TTTT TGC T	GG	GGT	3 GGA 3 CCA	090 GCCC 150 ATCG	ICAGG GCAG1	CAGG	3GG	GC.	31 FAT C C 31 AGC	LO IGA AC AG	CCA A A A A A A A CTC	CAGT C TT GT CCAT	CGT CCCCC CCT
adw2 : adw : adr(1): adr(2): ayw (1): ayw(2): adw2 : adw2 : adw : adr(1): adr(2): ayr :	CTCC A A A A A CAAC G G G G G G G	307 ACACG G 313 CAATTC GCG GCA GCA	0 GCG A A A O	GTAT C CC CC CC CTCC	TTTT TGC T	GG CI	GGT	3 GGA 3 CCA	090 GCCC 150 ATCG	ICAGG GCAG1	CTCA	GGG	GC. A A	31 TAT C C 31 AGC	LO IGA AC AG	CCA A A A A A A A A CTC	C AGT C TT GT CCAT	CT CC CT
adw2 : adw : adr(1): adr(2): ayw(1): ayw(1): ayw(2): adw2 : adw : adr(1): adr(2): ayr : ayw(1):	CTCC A A A A A CAAC G G G G G G G G G G G	307 ACACG G 313 CAATTC GCG GCA GCA A A	0 GCG A A A O CTC	GTAT C CC CC CC CTCC	TTT TGC T		GGT	3 GGA 3 CCA	090 GCCC 150 ATCG	ICAGG GCAGT	CTCA	GGG AAG	GC. A A A	31 TAT C C 31 AGC	LO IGA AC AG	CCA A A A A A A A A CTC C C	C AGT TT GT CCAT	GT CCCCC .TT GC
adw2 : adw : adr(1): adr(2): ayw(1): ayw(2): adw2 : adw2 : adw2 : adw : adr(1): ayr (2): ayw(1): ayw(2):	CTCC A A A A A CAAC G G G G G G G G G G G	307 CACACG G SATTC GCG GCA GCA A A A	0 GCG A A A O CTC	GTAT C CC CC CC CC CTCC	TTT TGC T	GG T	GGT CCA	3 GGA 3 CCA	090 GCCC 150 ATCG	ICAGG GCAGI C	CTCA	GGG AAG	GC. A A	31 IAT C C 31 AGC	LO FGA AC: AG 70 CTA	CCA A A A A A A A CTC C C C	C C C C T T G T C C C C C C C C C C C C	CGT CCCCC .TT
adw2 : adw : adr(1): adr(2): ayw(1): ayw(1): ayw(2): adw2 : adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	CTCC A A A A A CAAC G G G G G G G G	307 CACACG G 313 CAATTC GCG GCA A A A 319	0 GCG A A A O CTC G G	GTAT C CC CC CC CC CC CC	TTT TGC T		GGT CCA T	3 GGA 3 CCA 3	090 GCCC 150 ATCG 210	ICAGG GCAGI C PS2 ⁱ	CTCA CCAGG	3GG	GC. A A	31 IAT C C 31 AGC	LO IGA AC AG	CCAA A A A A A A A CTCC C C C	C AGI TT GT CCAN GC GC	GT CCCCC .T
adw2 : adw : adr(1): adr(2): ayw(1): ayw(1): ayw(2): adw2 : adw : adr(1): adr(2): ayr : ayw(1): ayw(2):	CTCC A A A A A CAAC G G G G G G CTCC	307 CACACG G 313 CAATTC GCG GCA GCA A A 319 CACCTC	0 GCG A A A O CTC G G O TAA	GTAT C CC CC CC CC CC CC CC CC CC	TTT TGC I		GGT CCA T CATC	3 GGA 3 CCA 3 CCA	090 GCCC 150 ATCG 210 AGGC	ICAGG GCAGI C PS2 ⁱ C <mark>ATG</mark> C	CAGG	G G G G G G G G G G G G G G G G G G G	GC. A A	31 IAT C C 31 AGC	LO IGAO AC. AG. 70 CTA	CCAA A A A A A A A CTCC C C C	C CAGT TT GT CCAT GC GC	GT CCCCC .TT GG
adw2 : adw : adr(1): ayr(2): ayw(1): ayw(2): adw2 : adw : adr(1): ayr : ayr(2): ayr(2): ayw(2):	CTCC A A A A A CAAC G G G G G G G CTCC	307 CACACG G 313 CAATTC GCG GCA GCA A A 319 CACCTC	0 GCG A A A O CTC G G G O CTAA	GTAT C CC CC CC CC CC CC CC	TTT TGC I	CT T	GGT CCA T CATC	3 GGA 3 CCA 3 CTC	090 GCCC 150 ATCG 210 AGGC	ICAGG GCAGI C PS2 ⁱ CATGC	CAGG	G G G G G G G G G G G G G G G G G G G	GC. A A	31 FAT C C 31 AGC	LO FGA AC AG	CCA A A A A A A A A CTC C C	C C C C T T G T C C C C C C C C C C C C	GT CCCCC .TCT GG
adw2 : adw : adr(1): adr(2): ayw(1): ayw(2): adw2 : adw2 : adr(1): ayw(2): adr(2): ayw(2): adw2 : adw2 : adw2 : adw2 : adw2 : adw2 :	CTCC A A A A A CAAC G G G G G G G G CTCC	307 CACACG G 313 CAATTC GCG GCA GCA A A 319 CACCTC	0 GCG A A A A O CTC G G O · TAA	GTAT C CC CC CC CC CTCC	TTT TGC I	.CI	GGT PCCA T EATC	3 GGA 3 CCA 3 CTC	090 GCCC 150 ATCG 210 2210	ICAGG GCAGI C PS2i C <u>ATG</u> C	CAGG	G G G G G G G G G G G G G G G G G G G	GC. A A	31 FAT C C 31 AGC	LO IGAO AG	CCA	C C A G T T G T C C A G T C C A G T G C A G C G C G C G C G C	GT CCCCC .T
<pre>adw2 : adw : adr(1): adr(2): ayw(1): ayw(1): ayw(2): adw2 : adr(1): adr(2): ayw(2): adw2 : adw2 : adw2 : adw2 : adw2 : adw2 : adw2 :</pre>	CTCC A A A A A CAAC G G G G G G G G CTCC	307 CACACG G SATTC GCG GCA A A 319 CACCTC	0 GCG A A A A O CTCC G G G	GTAT C CC CC CC CC CTCC	TTT TGC I	·GG	GGT CCA T SATC	3 GGA 3 CCA 3 CTC	090 GCCC 150 ATCG 210 240 240	ICAGG GCAGI C PS2 ⁱ C <u>ATG</u> C	CAGG	G G G G G G G G G G G G G G G G G G G	GC. A A	31 FAT C C 31 AGC	LO IGAO AG	CCA A A A A A A A CTC C C C	C CAGI TT GT CCAI GC GC	GT CCCCC .TCT GG
adw2 : adr(1): adr(2): ayr : ayr(1): ayr(2): adw2 : adw2 : adr(1): ayr(2): ayr(1): ayr(2): adw2 : adw2 : adr(1): adr(2	CTCC A A A A A CAAC G G G G G G G G CTCC	307 CACACG G CAATTC GCG GCA A A 319 CACCTC	0 GCG A A A A CTC G G G O · TAA	GTAT C CC CC CC CC CC CC CC CC CC CC CC	TTT TGC I	.CI	GGT CCA T CATC	3 GGA 3 CCA 3 CTC	090 GCCC 150 ATCG 210 AGGC	ICAGG GCAGT C PS2 ⁱ C	CAGG	G G G G G G G G G G G G G G G G G G G	GC. A A A	31 FAT C C 31 AGC	LO IGA AC: AG: 70 CTA	CCA A A A A A A A CTC C C C	C AGJ TT GT CCAJ	GT CCCCC.
adw2 : adr(1): adr(2): ayr : ayw(1): ayw(2): adw2 : adw2 : adr(1): ayr (2): ayr : ayr(1): ayw(2): adw2 : adw2 : adr(1): adr(2): adr(2): ayr : ayr(1):	CTCC A A A A A CAAC G G G G G G G G CTCC	307 CACACG G 313 CAATTC GCG GCA A A 319 CACCTC	0 GCG A A A A O CTC G G G G	GTAT C CC CC CC CC CC CC CC CC CC CC CC CC	TTT TGC I LCAG	·GG	GGT CCA T CATC	3 GGA 3 CCA 3 CTC	090 GCCC 150 ATCG 210 AGGC	ICAGG GCAGI C PS2 ⁱ C	CAGG	G G G G G G G G G G G G G G G G G G G	GC. A A	31 FAT C C 31 AGC	LO FGA AC. AG. 70 CTA	CCA A A A A A A A A A C C C C	C AGT TT GT CCAT GC GC	GT CCCC .T

13



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 DNAs⁴⁷⁻⁶¹ 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 product⁷⁸⁻⁸⁰ 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, respectively⁷⁵ (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 sequence⁷⁵ (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 Asn₁₄₆-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 Asn₁₄₆ 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 22nm 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 polypeptide 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 reported¹⁰² 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 strand⁴⁴ 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 HBVspecific transcripts of 2.1 and 3.5 kb¹³⁰⁻¹³⁴ (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/