1991

Assembly And Transport Of Adenovirus Fiber.

Jeong Shin Hong
University of Alabama at Birmingham

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9123886

Assembly and transport of adenovirus fiber

Hong, Jeong Shin, Ph.D.

University of Alabama at Birmingham, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

ASSEMBLY AND TRANSPORT OF ADENOVIRUS FIBER

by

JEONG SHIN HONG

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School, The University of Alabama at Birmingham

BIRMINGHAM, ALABAMA

1991

ABSTRACT OF DISSERTATION
GRADUATE SCHOOL, UNIVERSITY OF ALABAMA AT BIRMINGHAM

Degree Ph.D. Major Subject Biochemistry

Name of Candidate Jeong Shin Hong

Title Assembly and Transport of Adenovirus Fiber

Adenovirus fiber is a glycoprotein important to a number of adenovirus functions, such as virion assembly and virus infection. The fact that fiber is a nuclear trimeric glycoprotein makes fiber a promising model for the study of the interrelationship between intracellular protein transport, protein assembly, and posttranslational modification. The research presented in this dissertation includes determination of the nucleotide and predicted amino acid sequence of the BamHI-D fragment of Ad7 DNA, the examination of the role of O-GlcNAc in nuclear targeting, the identification of the peptide sequence for nuclear localization of fiber, and the investigation of the role of the C-terminal end of fiber in trimer formation.

The sequence determination of the Ad7 fiber gene demonstrated that Ad7 fiber has essentially the same secondary structure as fibers from other known serotypes. Three short regions of partial homology have been identified near both N- and C-termini of the predicted fiber gene sequences from Ad2, Ad3, Ad5, and Ad7.

Fibers from Ad2 and Ad5 contain O-linked N-acetyl glucosamine. H5ts142 is a temperature-sensitive, glycosylation defective mutant of Ad5. This mutant fiber demonstrated that O-GlcNAc addition is not necessary

for nuclear targeting of fiber, since the protein continued to accumulate in the nucleus at the nonpermissive temperature. The fiber from this mutant also showed that trimerization is not required for the nuclear transport of fiber.

Fiber has to be transported to the nucleus to be assembled into the virions. The nuclear translocation of the fiber protein was studied by expressing fiber using vaccinia virus expression system. The fiber protein contains all the information required for nuclear transport in itself since it was transported to the nucleus in the absence of other adenovirus proteins. The tetrapeptide KRAR at the N-terminus of Ad2 fiber was essential for this process, although it was not sufficient to transport the nonnuclear protein, ß-galactosidase, to the nucleus. Deletions at the C-terminus of the fiber did not have any effect on the nuclear localization of the fiber. Nuclear localization defective mutants of Ad2 fiber demonstrated that nuclear translocation of fiber is not a prerequisite for trimer formation.

Fiber is a trimeric protein. The fiber of H5ts142 cannot form trimers at the nonpermissive temperature. The fiber gene of H5ts142 includes one amino acid change at the amino acid position 440: from alanine to valine. The temperature-sensitive phenotype of this mutant resulted from this single mutantion. The same mutation introduced into Ad2 fiber did not have the same effect as in H5ts142 fiber. Deletions at the C-terminal end of the protein resulted in a trimer-defective phenotype. Therefore, the C-terminus must play an important role in the trimerization of fiber.

Abstract Approved by: Committee Chairman Jeffrey A. Engler

Program Director Jeffrey A. Engler

Date _____ Dean of Graduate School Gerry L. Hickey

DEDICATION

I would like to dedicate this thesis to my parents with all my love.

ACKNOWLEDGEMENTS

The idea of coming to the graduate school in the United States was exciting and adventurous, but at the same time scary. There are many people who have helped me to get over the cultural difference and language problem. Also many have supported me towards the completion of my graduate studies academically. So many people have helped me throughout my years of graduate school that I find it impossible to name them all. To those whom I have not mentioned please know that I am extremely grateful for all your help.

First of all, I must thank my mentor, Dr. Jeffrey Engler. Without his patience, understanding, support, and guidance this would not have been possible. His encouragement, especially in my early years as a graduate student, helped me to build confidence in my work. I appreciate the opportunity that he has given me to work in his lab and to learn about science as well as life in general.

I also want to thank the members of my committee, Dr. Gillian Air, Dr. Charles Bugg, Dr. Richard Compans, and Dr. Tim Townes, for their time and assistance during the course of this work.

I would like to thank Dr. Karen G. Mullis, for her support as a colleague and a friend. She, more than anyone else, helped me to get over the cultural difference and was always kind enough to explain all of my trivial questions. Without her, graduate school would not have been the same. I thank all the members of the Engler lab who have contributed

throughout this work, especially Nancy Caterina and Barbara Mulach for helping me to make this thesis more readable.

I must thank my friends, Dr. Jeunghoi Ha and Insil Joung, for rescuing me from the deprivation of Korean language, especially through the endless phone calls at night. We have gone through the same problems at the same time and their companionship helped me to get over the difficulties that we shared.

I also would like to thank my aunt and uncle, Mr and Mrs. Donald Koh, and cousins Julie, Jeana, and Charles for providing love and support as a family.

Finally, but most of all, I thank my parents, Mr. and Mrs. Dongsam Hong, for allowing me to come this far and have opportunities to experience another world and learn. Without their love and support this would not have been possible. I also thank my brothers, Chunghoon and Janghoon, for their moral support, especially for listening to all of my complaints patiently.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xii
INTRODUCTION	1
REFERENCES FOR INTRODUCTION	29
CHARACTERIZATION OF THE EARLY REGION 3 AND FIBER GENES OF AD7	34
NUCLEAR TRANSPORT OF AN ADENOVIRUS FIBER MUTANT THAT LACKS O-LINKED N-ACETYLGLUCOSAMINE	66
THE AMINO TERMINUS OF THE ADENOVIRUS FIBER PROTEIN ENCODES THE NUCLEAR LOCALIZATION SIGNAL	97
TRIMER FORMATION DEPENDS ON AMINO ACID SEQUENCES IN THE CARBOXYL-TERMINUS OF AD2 AND AD5 FIBER	125
SUMMARY	154
REFERENCES FOR SUMMARY	163
APPENDIX A	166
APPENDIX B	189
APPENDIX C	195

LIST OF TABLES

Table		Page
	INTRODUCTION	
I	Subgroup classification of human adenoviruses	2
II	Amino acid sequences that act as signals for translocation to the cell nucleus	19
	CHARACTERIZATION OF THE EARLY REGION 3 AND FIBER GENES OF AD7	
I	Nucleotide and predicted amino acid sequence homology for ORFs in early region 3 and the fiber gene	51
	THE AMINO TERMINUS OF THE ADENOVIRUS FIBER PROTEIN ENCODES THE NUCLEAR LOCALIZATION SIGNAL	
I	Sequences of the DNA oligomers used in this study	107
	TRIMER FORMATION DEPENDS ON AMINO ACID SEQUENCES IN THE CARBOXYL-TERMINUS OF AD2 AND AD5 FIBER	
I	List of the oligonucleotides used for sequencing	135

LIST OF FIGURES

Figure		Page
	INTRODUCTION	
1	Model of adenovirus virion	5
2	Model of the adenovirus life cycle	10
3	Generation of recombinant vaccinia virus	23
	CHARACTERIZATION OF THE EARLY REGION 3 AND FIBER GENES OF AD7	
1	The nucleotide sequence of the BamHI-D fragment of Ad7 (Gomen) DNA	42
2	Comparison of the protein coding capacity of the Ad7 BamHI-D fragment with the corresponding region of Ad3	48
3	The structural organization of Ad7 fiber and homology to Ad3 fiber	55
4	S1 nuclease mapping of the 5' and 3' ends of the exon that encodes Ad7 fiber mRNA	58
	NUCLEAR TRANSPORT OF AN ADENOVIRUS FIBER MUTANT THAT LACKS O-LINKED N-ACETYLGLUCOSAMINE	
1	Characterization of monoclonal antibodies against fiber . .	74
2	Characterization of galactosylation of Ad5 and of H5ts142 fiber	77
3	Nuclear localization of Ad5 wild type and of H5ts142 fiber at 32°C and 39.5°C	81
4	Immunoblot on H5ts142 infected cell lysates after the addition of cycloheximide, followed by temperature shift	83
5	Nuclear localization of H5ts142 fiber after addition of cycloheximide and temperature shift	87

LIST OF FIGURES (Continued)

Figure		Page
6	Immunoblot on H5ts142-infected cell lysates after the addition of cycloheximide, followed by temperature shift	89
	THE AMINO TERMINUS OF THE ADENOVIRUS FIBER PROTEIN ENCODES THE NUCLEAR LOCALIZATION SIGNAL	
1	A. Construction of recombinant plasmid B. Location of possible nuclear targeting sequences within the protein	103
2	Immunoblot on cell lysates infected with recombinant vaccinia virus	109
3	Localization of fiber in recombinant virus infected HEp2 cells	112
4	Localization of fibers in 2F-TKRVRL and 2F-ΔKRAR-PKRARP infected HEp2 cells	115
5	Localization of recombinant ß-galactosidase in HeLA T4 cells	118
	TRIMER FORMATION DEPENDS ON AMINO ACID SEQUENCES IN THE CARBOXYL-TERMINUS OF AD2 AND AD5 FIBER	
1	Restriction fragments containing the fiber gene of Ad5 and H5ts142	133
2	The location of mutations in fiber gene of H5ts142	136
3	Immunofluorescence of recombinant Ad5 wild-type fiber (S3-5F) and of recombinant H5ts142 fiber (S3-5ts142)	139
4	Immunofluorescence of the recombinant H2ts142 fiber (S3-2ts142)	142
5	Construction of deletion of Ad2 fiber.	144
6	Immunofluorescence on C-terminal deletion mutants of AD2 fiber	146
7	Immunoblot on C-terminal deletion mutants infected cell lysates	148

LIST OF FIGURES (Continued)

Figure		Page
	APPENDIX A	
1	Detailed comparison of Ad3 and Ad7 nucleotide and predicted amino acid sequences in early region 3	168
2	Comparison of the predicted amino acid sequences for Ad2, Ad3, Ad5, and Ad7 fiber genes	177
3	Conserved amino acid segments at N-terminus and C-terminus of the predicted amino acid sequences for Ad2, Ad3, Ad4, Ad5, and Ad7	181
4	Detailed comparison of Ad2, Ad3,and Ad7 nucleotide and predicted amino acid sequences in comparable ORFs in early region 3	183
5	Comparison of the early region 4 sequences from Ad2 and Ad7	187
	APPENDIX B	
1	Nuclear localization of Ad7 wild type fiber	191
2	Western blot of H5ts142, Ad5, and uninfected cell lysates	193
	APPENDIX C	
1	A diagram of a series of replacement mutations in region 1 of Adenovirus DNA polymerase (Ad Pol).	202
2	The protein- coding capacity of the E2b region of Ad12	208
3	Comparison of the protein-coding potential of Ad2 (class C), Ad7 (class B), and Ad12 (class A) in the region that encodes the Mr 13600 protein (a) and the Ad Pol (b) protein products of Ad2	210
4	Comparison of predicted acceptor splice site for the Ad Pol mRNA	213
5	Comparison of the predicted pTP and adenovirus DNA polymerase aa sequences of Ad2, Ad7, and Ad12	215

LIST OF ABBREVIATIONS

A	alanine
aa	amino acid
Ab	antibody
Ad2	adenovirus serotype 2
Ad3	adenovirus serotype 3
Ad5	adenovirus serotype 5
Ad7	adenovirus serotype 7
Ad12	adenovirus serotype 12
Ad18	adenovirus serotype 18
Ad31	adenovirus serotype 31
Ala	alanine
Arg	arginine
Asn	asparagine
Asp	aspartic acid
ATP	adenosine triphosphate
bp	base pair
C-terminal	carboxy-terminal
D	aspartic acid
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
ds	double-stranded
E	glutamic acid

E1a	early region 1a
E3	early region 3
ER	endoplasmic reticulum
GlcNAc	N-acetylglucosamine
Glu	glutamic acid
Gln	glutamine
gpt	the E. coli gene which codes for the enzyme xanthine-guanine phosphoribosyl-transferase
K	lysine
kb	kilobase
KDa	kilodalton
L	leucine
L5	late region 5
Leu	leucine
Lys	lysine
μ	micro (10^{-6})
mAb	monoclonal antibody
mRNA	messenger RNA
MW	molecular weight
nm	nanometer (10^{-9})
N-linked	asparagine-linked
NPT	nonpermissive temperature
N-terminal	amino-terminal
ORF	open reading frame
O-linked	hydroxyl-linked

O-GlcNAc	hydroxyl-linked N-acetylglucosamine
P	proline
PB1	polymerase basic protein 1 of influenza virus
PB2	polymerase basic protein 2 of influenza virus
Pro	proline
PT	permissive temperature
Q	glutamine
R	arginine
RNA	ribonucleic acid
S	serine
S1	nuclease S1
SDS	sodium dodecyl sulfate
Ser	serine
Thr	threonine
TK	thymidine kinase
TK-	thymidine kinase-minus
ts	temperature sensitive
V	valine
Val	valine
WGA	wheat germ agglutinin

INTRODUCTION

Adenoviruses are nonenveloped particles with icosahedral symmetry that contain genomes of linear double-stranded (ds) DNA. Since the discovery of adenovirus by Rowe et al. (49), 41 antigenic types of adenoviruses have been found to infect humans, producing a variety of acute respiratory, ocular, gastrointestinal, and urinary diseases. Numerous other types of adenoviruses infect a wide variety of lower animal species (for review, see 17). In addition to the clinical importance of adenoviruses as human pathogens, adenoviruses have played an important role in the discovery of various features of viral structure, replication, and cell transformation. Molecular studies of adenoviruses have led to greater knowledge of a variety of cellular processes. Construction of viral mutants and the subsequent genetic studies will aid the understanding of these processes. The focus of this dissertation is to study intracellular protein trafficking, especially to the nucleus, and the inter-relationship among protein structure, posttranslational modification, and protein translocation. Adenovirus fiber, a trimeric glycoprotein, serves as an excellent model system for this purpose.

Classification of human adenoviruses

All human adenoviruses are members of the genus Mastadenovirus of the family Adenoviridae to indicate their similarity to other vertebrate (but nonavian) adenoviruses. At least 41 distinct serotypes of human adenovirus have been recognized and they have been classified into seven subgroups, denoted A to G, based on four different criteria (Table 1, 16):

Table 1: Subgroup classification of human adenoviruses

Subgroup	Serotypes	DNA homology	Oncogenicity	Cell transformation	HA group
A	12, 18, 31	48-69% with A 8-20% outside	100% tumors in 4 months	(+)	3A
B	3, 7, 14, 16, 21, 34, 35	89-94% with B 9-20% outside	10-50% tumors in 4-18 months	(+)	1A, B
C	1, 2, 5, 6	99-100% with C 10-16% outside	None	(+)	3A
D	8-10, 13, 15, 17, 19, 20, 22-30, 32, 33, 36-39	99-94% with D	None	---	2A-F
E	4	4-23% outside	None	---	3A
F	40	----	None	---	--
G	41	----	None	---	--

oncogenicity in rodents, DNA sequence homology, hemagglutination, and antigenicity of virus proteins in transformed cells.

Subgroup A viruses (Ad12, Ad18, and Ad31) are strongly oncogenic and able to rapidly induce tumors in the majority of inoculated animals. Subgroup B viruses have a very limited capacity to induce tumors. Subgroups C to G are nononcogenic and able to transform rodent cells in vitro but not in vivo. The transformed cells are free of infectious virus but contain virus specific proteins that can be recognized by immunological tests. In general, cells transformed by members of the same subgroup synthesize virus-specific proteins that share at least one common antigenic determinant, whereas cells transformed by members of different subgroups contain serologically distinct, virus-specific proteins.

The extent of DNA homology between serotypes has been measured by DNA:DNA hybridization. Generally, serotype members within a subgroup are up to 85% homologous while cross subgroup DNA homology is limited to about 10% to 15% (5, 16, 26). The differences within the groups are confined largely to two major surface antigens of the virus capsid (hexon and fiber proteins), while differences between the groups seem to extend completely along the chromosome. However, these studies underestimate the inter-subgroup homologies in some genes such as adenovirus polymerase, preterminal protein, and protein IVa2 (51, see Appendix A).

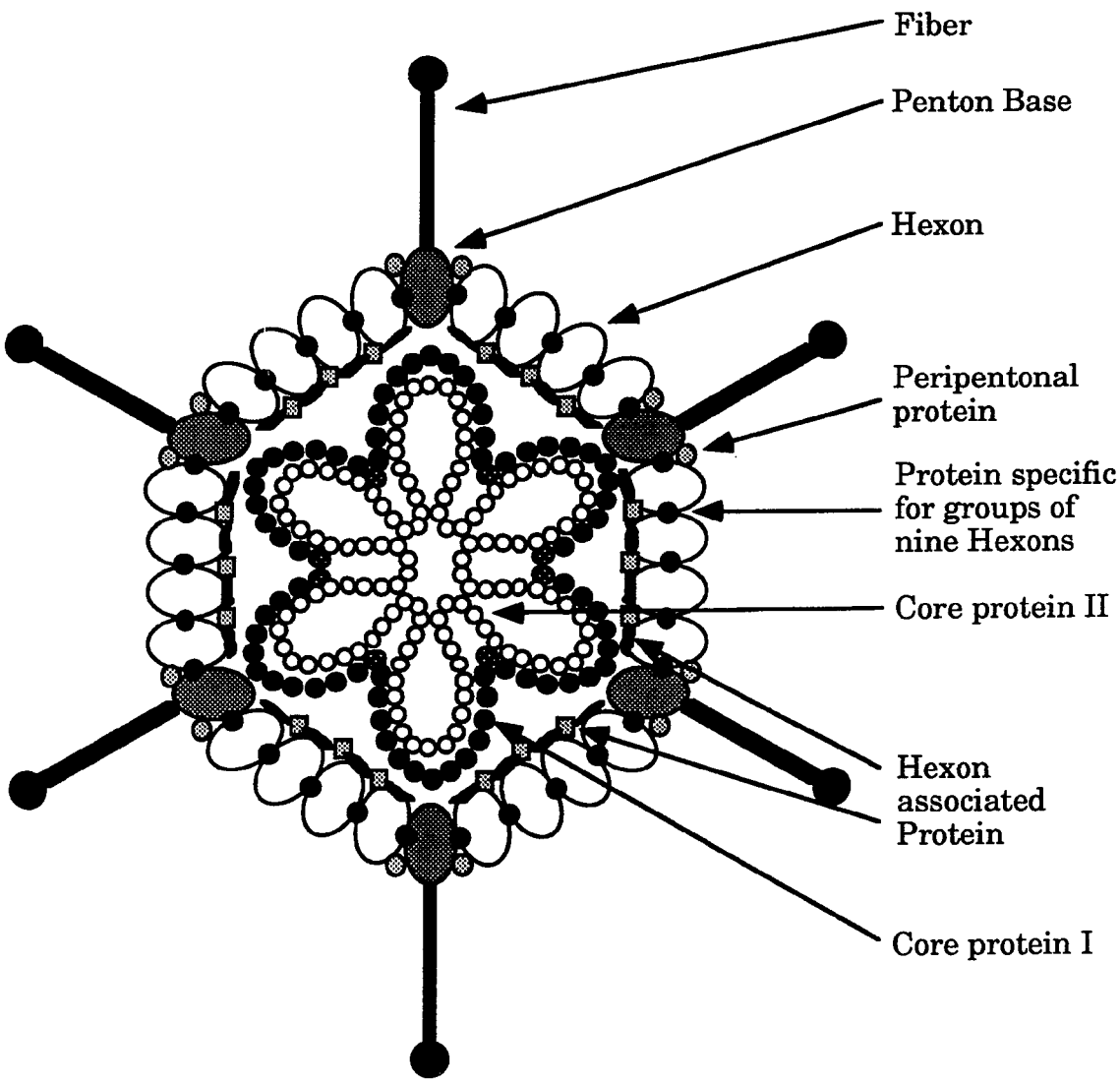
Human adenovirus serotypes have been classified into subgroups on the basis of their ability to agglutinate Rhesus monkey and rat erythrocytes. The hemagglutination reaction is inhibited by antisera against viruses of the same subgroup but not by antisera to viruses of heterologous groups (16, 17).

The composition and morphology of Adenovirus

The adenovirus genome is a linear, double-strand DNA with no single strand breaks. Double-stranded DNA comprises 11.6%-13.5% of the mass of the virion, the remainder being protein (17). Despite substantial nucleotide sequence divergence, all adenovirus serotypes studied so far show the same general genetic organization. Since the genomes of the highly homologous types Ad2 and Ad5 have been studied most extensively, they serve as models for the other serotypes. Traditionally, the adenovirus genes are subdivided into the early genes, which are expressed before the onset of viral DNA replication, and the late genes, which are transcribed after replication of adenovirus DNA has started. The adenovirus genetic information is scattered over the two complementary strands. About 69% of all genetic information is located on the rightward-transcribed strand (r-strand), while 31% of the coding sequences are present on the leftward-transcribed strand (l-strand). Adenovirus fiber gene, the major subject in this dissertation, is encoded in the L5 family of late transcripts.

The adenovirion is icosahedral and contains at least nine structural proteins (Figure 1). The structural polypeptides have been labeled II to IX and IIIa on the basis of decreasing molecular weight as revealed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (16, 17). Polypeptides II, III, and IV correspond to hexon, penton base, and fiber, which are the major constituents of the outer capsid of the virion. Polypeptide IIIa is associated with the vertex region, and the stoichiometry implies the existence of 5 molecules of polypeptide IIIa for each penton structure. The viral DNA is associated with polypeptides V and VII to form a nuclear core organized as 12 spherical subunits, each in contact with a vertex of the icosahedral outer capsid. Both polypeptide VI and VIII are associated with hexons, and polypeptide VI also has a high affinity for

Figure 1: Model of adenovirus virion. The model for adenovirus virions features a double stranded DNA genome surrounded by a proteinacious coat. This tentative model illustrates the location of proteins in the adenovirus type 2 virion. The major capsid components are hexon, penton base, and fiber. The minor capsid components include peripentonal protein, hexon associated proteins, and protein specific for groups of nine hexons. The core of the virion is composed of two proteins, core protein I and II, which form a nucleocapsid structure with the double stranded DNA viral genome.


Fiber
Penton Base
Hexon
Peripentonal protein
Protein specific for groups of nine Hexons
Core protein II
Hexon associated Protein
Core protein I

DNA. Polypeptide IX is associated with hexon and has a stabilizing effect on the capsid.

There are 240 hexon macromolecules in the virus capsid and only 12 pentons. The hexon protein was the first animal virus protein to be crystallized. Hexon (324K MW from Ad2) consists of three identical polypeptides (108K MW from Ad2) and forms the surface protein shell. The top of the hexon contains the type-specific antigenic determinants, whereas the group specificity is determined by the bottom portion of the hexon (17).

The capsomers at the vertices of the icosahedron are called pentons because they are surrounded by five neighbors, the peripentonal hexons. A penton is composed of two structural subunits, a penton base and a thin antenna-like projection called a fiber. Penton base is composed of identical polypeptide chains, each having a molecular weight of approximately 85K. Molecular weight determinations using neutron-scattering techniques have given a molecular weight of 246K for the penton base (11), which therefore suggests that it is a trimeric structure. The penton base carries a weak antigenic determinant, ß, that is common to pentons from all adenovirus serotypes. The penton base is also responsible for cytopathic changes in adenovirus-infected cells (16, 17).

Adenovirus Fiber Protein

The fiber protein of adenovirus is a highly asymmetrical structure consisting of three domains: an N-terminal tail that interacts with penton base (10), a shaft, and a knob at the C-terminus. The knob contains the antigenic determinants which determine the serotype-specificity and is responsible for binding to the cell surface receptor. The length of the fiber is a characteristic property for members of different adenovirus subgroups. Subgroup B adenoviruses have the shortest fibers, 10 nm, and subgroup C have the longest, 23-37 nm. The fibers from Ad2 and Ad5, belonging to

subgroup C are the only fibers which so far have been studied thoroughly (16,17).

The native fibers from Ad2 and Ad5 have a molecular weight of 183,000-207,000 (17). Upon denaturation, the fiber polypeptide has a molecular weight of about 62,000. This suggests that the fiber is a trimeric structure which is thought to be associated with one penton base to form a penton complex. Green et al. (18) proposed a dimeric structure for fiber which agrees with the mol. wt. of 150,000-160,000 reported by Devaux et al.(11). Green et al. (18) analyzed the amino acid sequence of the adenovirus fiber protein and discovered a 15-residue segment which repeats approximately 22 times, over a range of 300 residues. Each 15-residue segment contains two short ß-strands and two ß-bends. The 44 ß-strands together give a long narrow, amphipathic ß-sheet, which could be stabilized by dimer formation to give the shaft of the fiber. A comparison between the predicted amino acid sequences of the fiber polypeptides of Ad2, Ad5 (5) (subgroup C), Ad3 (52), and Ad7 (26) (subgroup B) revealed that they have almost identical secondary structures, consisting of a tail, a shaft, and a knob. A striking difference between the fibers from subgroup B and subgroup C adenoviruses is that the shaft of the subgroup B fiber is significantly shorter, containing only 6 repeat units compared with 22 in the subgroup C fiber (5, 26, 52). The fiber is coded by a single gene and sequencing the Ad2 fiber gene region has revealed a single open reading frame which codes for a polypeptide of molecular weight 61,925. Thus if two populations of fiber polypeptide exist, their difference is probably caused by posttranslational modification of the polypeptide chain. Recently, more evidence supporting the trimer structure of fiber has been reported from an analysis of the virion composition by Van Oostrum and Burnett (54) and the crystal structure of fiber by Devaux et al (9). They

proposed a triple-helical model for fiber in which the shaft is stabilized by inter-chain hydrogen bonds as well as hydrophobic interactions.

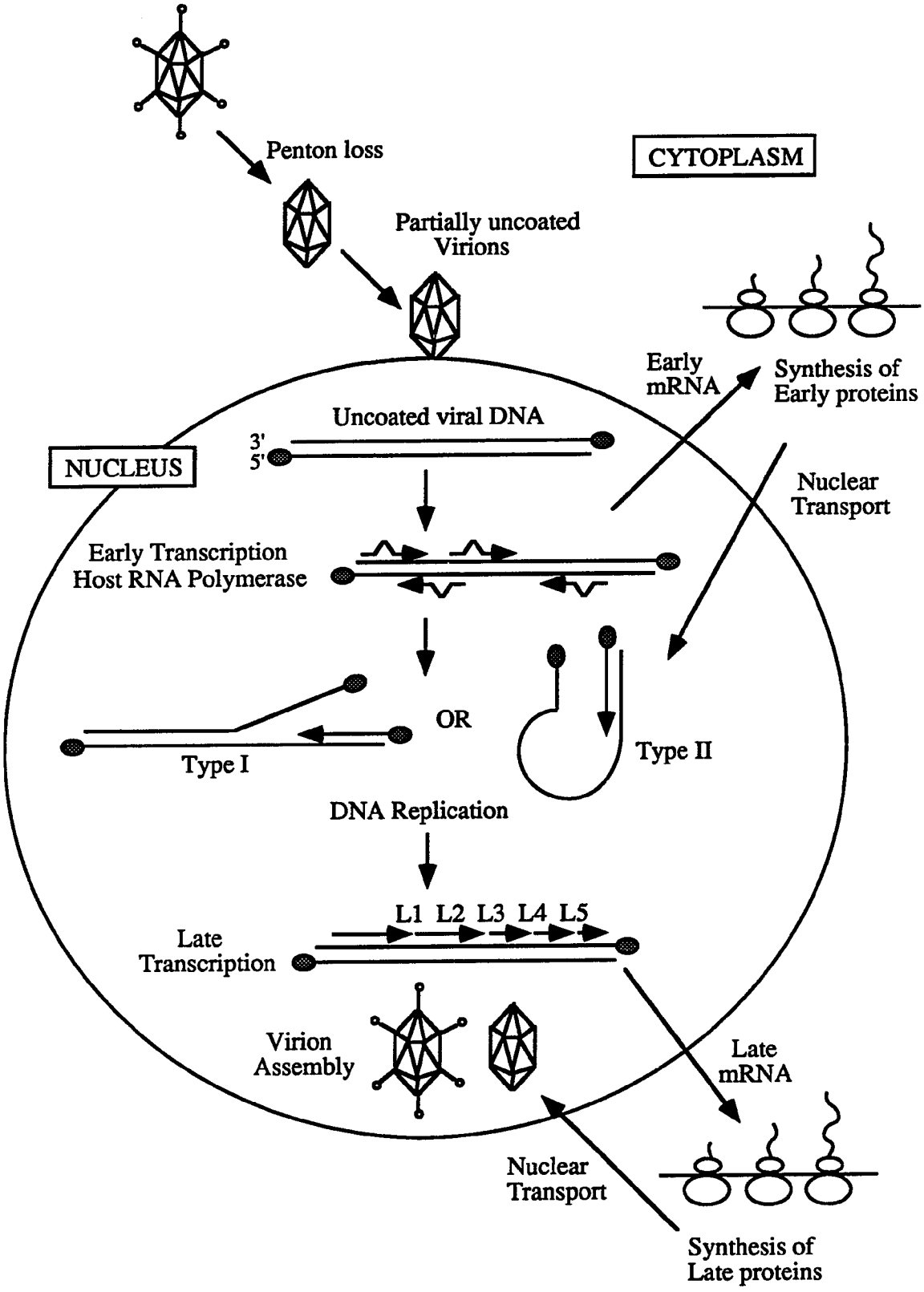
Adenovirus fiber is a glycoprotein. Ishibashi and Maizel (28) previously demonstrated that Ad2 fiber could be labeled by using radioactive glucosamine and that this label could be removed by treatment with weak alkali; this observation suggested that fiber contained an O-linked amino sugar derivative. More recently, it has been shown that both Ad2 and Ad5 fibers have monosaccharide GlcNAc in an O-glycosidic linkage, but that fiber from Ad7 does not contain this modification (4, 38). The functional significance of the carbohydrate moiety remains unknown.

Lytic infections by Adenoviruses (Figure 2)

Adenovirus infection has drastic effects on the metabolism of host cells as the virus usurps almost the entire synthetic machinery in its host to direct the production of 4,000-10,000 progeny viral particles per cell. The infected cells do not divide, but continue to increase in mass. The adenoviruses do not synthesize enzymes that lyse the cells; the infected cells are killed by a process of attrition (16, 17).

Adenovirus fiber functions to allow attachment of the virus to one of the 5,000 (Ad2) to 7,000 (Ad3) specific receptors on the plasma membrane of the cell (8). Once bound, an adenovirus particle may enter the cell by endocytosis (8). Within a few minutes the virion is partially uncoated with the pentons removed and viral DNA becomes partially accessible to deoxyribonuclease (DNase). The resulting particles, devoid of pentons but otherwise intact, appear to become associated with pores in the nuclear envelope. This transfer of partially uncoated virions across the cell's cytoplasm is believed to be facilitated by microtubules. The second step in uncoating seems to result in the removal of the remaining capsomers from the DNA-protein core, generating an intermediate that is even more

Figure 2: Model of the adenovirus life cycle. The adenovirus life cycle encompasses the time from the virion binding to the cell surface receptors until the release of infectious virions. The life cycle is divided into early and late stages. Early events include cell receptor binding, endocytosis, synthesis of early mRNA and proteins. Late events are initiated by viral DNA replication which can occur using either a type I or type II replication intermediate. During the late stage of infection, late mRNA is produced, late proteins are synthesized, and virion assembly occurs in the nucleus.


Penton loss
CYTOPLASM
Partially uncoated
Virions
Early
mRNA
Synthesis of
Early proteins
Uncoated viral DNA
3'
5'
NUCLEUS
Nuclear
Transport
Early Transcription
Host RNA Polymerase
OR
Type I
Type II
DNA Replication
L1 L2 L3 L4 L5
Late
Transcription
Virion
Assembly
Late
mRNA
Nuclear
Transport
Synthesis of
Late proteins

sensitive to DNase than are pentonless virions. Once the cores are within the nucleus, the final stage of uncoating occurs to generate viral DNA that is apparently almost free of protein.

The productive cycle of adenovirus in infected cells is divided into two phases, early and late, delineated by the onset of viral DNA replication which begins about 6 to 8 hours after infection. During the early phase, only a fraction of the information encoded by the viral genome is expressed as mRNA, which directs the synthesis of a small number of viral proteins. At this time, viral mRNA constitutes a minor proportion of the total RNA of the cell; but once replication of adenoviral DNA begins, the amount of viral mRNA increases dramatically until, by 18 hours, it comprises virtually all the mRNA bound to the cell's polyribosomes. Late mRNA, present once viral DNA synthesis has been initiated, accounts for most of the information of the viral genome and directs the synthesis of large amounts of viral structural proteins.

The structural proteins are synthesized in the cytoplasm and rapidly transported to the nucleus to be assembled into the virion, probably in the form of a trimer for the main capsid units: hexon, penton base, and fiber. Empty capsids are formed by a process of self-assembly and DNA is packaged into these capsids. It is not yet clear whether the DNA is packaged before the core protein. The young virions contain several polypeptides in a precursor form which are cleaved proteolytically in the virion structure.

Intracellular trafficking of proteins

The eukaryotic cell has localized many of its metabolic processes into specific membrane-limited compartments or organelles. Since most cellular polypeptides are synthesized on ribosomes in the cytoplasmic compartment, the cell must possess specific transport mechanisms in

order to deliver newly synthesized proteins to their site of function. Proteins destined for the extracellular space, the plasma membrane, or lysosomes are first cotranslationally inserted into or transferred across the endoplasmic reticulum, and then transported to their correct destinations via the Golgi complex. Proteins specific for the ER and the Golgi complex use only the initial parts of this transport pathway. In contrast, those proteins destined for the nucleus, mitochondria, or chloroplasts are routed directly in completed form to the proper organelle (2).

It is assumed that proteins that leave the cytoplasm must contain a signal, a specific structural determinant, that can be recognized by receptors on the target organelle (12, 42, 55). To localize and characterize those structural features of a protein that act as signals, peptide signal sequences have been investigated extensively by recombinant DNA techniques. The signal activity of a proposed peptide has been tested directly by fusing it to another protein and analysing the behavior of this hybrid protein.

Only proteins that carry a special hydrophobic signal peptide, a basic amino acid terminus followed by a stretch of hydrophobic amino acids, are imported into the endoplasmic reticulum (ER) from the cytosol. As the signal peptide appears from the ribosome, it interacts with a signal-recognition particle and further synthesis is delayed until the complex reaches the docking protein on the ER membrane (54). This binding to the membrane initiates an ATP-dependent translocation process that threads a loop of polypeptide chain across the ER membrane, usually coupled to translation, and the signal is generally cleaved at this time. Most proteins are exported from the ER but those soluble proteins that function in the lumen of the ER have a carboxy-terminal tetrapeptide (KDEL = Lys-Asp-Glu-Leu) which may serve to recover protein lost by inadvertent export.

Recovery would require a specific receptor which might act at a salvage compartment between the transitional elements of the ER and the Golgi stack (56). Proteins that enter the lumen of the ER can be N- or O-glycosylated during the transport if they contain the necessary targeting sequences or protein structure for these modifications.

The Golgi apparatus receives newly synthesized proteins from the ER and distributes them to the plasma membrane, lysosomes, and secretory vesicles (2). Transport to the Golgi occurs by bulk flow. Proteins traverse the Golgi travelling sequentially to three compartments, and undergo enzymatic modifications including both N- and O-linked oligosaccharide processing in each compartment. Signal patches, noncontiguous regions of protein brought together by protein folding, are used to direct the sorting to lysosomes and secretory granules from the Golgi. Proteins lacking signal patches are nonselectively transported to the cell surface by bulk flow (48).

Proteins intended for transport to lysosomes contain mannose-6-phosphate in their structure. Receptors for mannose-6-phosphate located on the lumenal face of the Golgi membrane bind the mannose-6-phosphate residue of the lysosomal protein and direct the protein to lysosome (2).

Mitochondria import most of their proteins from the cytosol. Only proteins that contain a specific signal peptide are translocated into mitochondria. The signal peptide, a longer, amphiphilic sequence that is rich in basic and hydroxylated amino acids but poor in acidic ones, is usually located at the amino terminus and is cleaved off after import. Translocation is driven by both ATP hydrolysis and the electrochemical gradient across the inner membrane. Proteins are usually directed into the matrix unless additional sequences interrupt transport and they cross both mitochondrial membranes at adhesion sites (2, 56).

Transport of proteins into the nucleus

In eukaryotic cells, the cytoplasm and nucleus are divided into discrete compartments by the nuclear envelope. Communication between these two compartments occurs via nuclear pores. The nuclear pores are able to selectively transport substrates ranging in size from 40-2,000 kD while still providing open channels for the passive diffusion of smaller molecules (for review, see 12). The absolute size cutoff for a passive diffusion of a spherical protein is around 60kD (27). Although it has been suggested that selective nuclear accumulation of proteins is mediated by selective retention of nuclear proteins by binding to some compartment of the nucleus after diffusion, it is becoming clear that nuclear proteins accumulate in the nucleus by a selective transport mechanism. Currently, very little is known about the proteins responsible for regulating this uniquely flexible transport apparatus.

Protein import through the nuclear pore complex has been recently described by a number of groups. Newmeyer et al. (40) have shown that the import of a nuclear protein, nucleoplasmin, into the nucleus displays the characteristics for an active transport system: transport requires a signal domain, is completely ATP-dependent, is temperature-dependent, and requires an intact nuclear envelope. It has also been shown that the lectin wheat germ agglutinin (WGA) inhibits nucleoplasmin transport by binding to sugar moieties in the nuclear pore (14). It was concluded that pore complex glycoproteins play a role in the transport mechanism. Nuclear transport of a protein can be divided into at least two different steps (1, 41, 43). The first step, rapid binding of a protein to the pore, is signal-dependent but it does not require ATP and is not inhibited by WGA. The second step, translocation through the pore, is slower, ATP-dependent, and WGA-sensitive.

Recently genes encoding several nuclear proteins have been studied and recombinant DNA techniques allowed the determination of a protein sequence responsible for nuclear localization of SV40 large T antigen (a tetramer with subunits of MW=94,000) (30, 31). SV40 large T antigen is predominantly a nuclear protein but is also found in a modified form in the plasma membrane. Within a putative DNA binding domain of SV40 large T antigen, there is a striking tract of five basic amino acids, (127)Lys-Lys-Lys-Arg-Lys(131), just down stream from a cluster of serine and threonine phosphorylation sites. Kalderon et al. (30, 31) generated a series of point mutations in SV40 large T antigen using mixed oligonucleotide mutagenesis and showed that mutation of lysine(128) to threonine abolished transport into the nucleus. This mutant protein binds DNA and is present in a soluble form in the cytoplasm. Therefore, the abolition of transport into the nucleus is not due to either failure to bind DNA in the nucleus or to irreversible binding to cytoplasmic components. Mutation of other amino acids in the region of lysine(128) influenced the nuclear localization of the large T antigen. When probed with a monoclonal antibody to large T antigen, some cells showed predominantly nuclear staining, some exclusively cytoplasmic, some exclusively nuclear, and some showed equal staining in the nucleus and cytoplasm. This result is thought to indicate that mutations in the region of lysine(128) influence the rate of accumulation of the mutant proteins in the nucleus. Kalderon et al. (30, 31) identified a minimum protein sequence from SV40 large T antigen that specifies entry into the nucleus, and they have examined its properties in detail. The shortest amino acid sequence derived from SV40 large T antigen that confers nuclear localization on a nonnuclear protein (chicken muscle pyruvate kinase) when linked to it at the amino terminus is pro-lys-lys-(128)lys-arg-lys-val (31) and this is regarded as a prototype for other

such signals (52). It is also able to target large T variants to the nucleus, whether located at its original site or appended to the amino terminus. Roberts et al. (47) have shown that a nuclear localization signal can function at a variety of positions within a protein, but that in some locations its activity is masked. When multiple copies of a partially defective signal are integrated into pyruvate kinase, multiple signals can cooperate to enhance nuclear accumulation (47). Partial cell-dependent efficiency of nuclear localization of a mutant of SV40 large T antigen (a natural deletion mutant lacking an internal large T antigen domain that includes the signal for nuclear transport) has been reported (15). When the signal-specifying sequences were inserted near the 3' end of the gene, different cell lines differed considerably in their ability to localize some variant molecules into the nucleus. The nuclear localization improved either with reiteration of the signal or with a left-flank modification of the signal amino acid context. The importance of protein context around the signal is also shown by Rihs and Peters (46). The residues 126-132, PKKKRKV, of SV40 large T were necessary and sufficient to localize large proteins to the nucleus. However, the addition of the 15 amino acids (residues 111-125) which flank the karyophilic sequence on its N-terminal border have a rather dramatic effect on the kinetics of nuclear transport even though those residues do not seem to contain a second karyophilic signal (46). Therefore, it has been concluded that while a minimal nuclear localization signal is sufficient for nuclear transport, its activity is crucially dependent on its context within the protein.

Homologous sequences have been found in other nuclear proteins; Smith et al. (53) conducted a computer search of a data base of 2511 protein sequences for homologies to the SV40 prototype sequence (PKKKRKV). The majority of proteins selected with a homologous sequence were viral

proteins, including adenovirus E1A and SV40 VP1, VP2, and VP3. The type of experimental approach described for SV40 large T antigen has been used in a number of studies to link sequences from nuclear proteins to the sequences of nonnuclear proteins. Nuclear localization signal sequences have been identified in many nuclear proteins (Table 2). The disappointing but not surprising result of these investigations is that there is no signal sequence common to all the nuclear proteins studied. This divergence among signals suggests that the shape and charge of this region rather than amino acid sequence may be of primary importance for signal recognition. In summary, the essence of a nuclear targeting sequence appears to be a short run of the basic amino acids, lysine and arginine; but possession of such a sequence, while probably necessary, is not sufficient to ensure nuclear entry. The location of the nuclear import signal on a protein does not seem to be important, however, its context within the protein is crucial. Because the signal peptide is not removed, nuclear proteins can be imported repeatedly, as is required each time the nucleus reassembles following mitosis.

It has been recently reported that some nuclear proteins contain two regions within their sequences for nuclear localization (19, 36, 39, 44). These proteins can be separated into two groups: those with two discrete signals and others with bipartite nuclear localization signals. Polyoma large T antigen contains two mutually independent sequence elements that contribute to the nuclear localization of the protein. Deleting both of those sequences from large-T abolishes nuclear accumulation, while deleting only one has less effect (44). NS1 protein of influenza virus also includes two domains containing nuclear localization signals; either region alone is sufficient to direct the NS1 protein to the nucleus (19). In these proteins, although either one of two nuclear localization signals can direct the

Table 2. Amino acid sequences that act as signals for translocation to the cell nucleus

Nuclear protein	Transported nonnuclear protein	Deduced signal sequence	Reference
Large T antigen (SV40)	Chicken muscle Pyruvate kinase	pro-lys-lys(128)-lys -arg-lys-val	30, 31
Mat alpha 2 (Yeast)	E. coli ß-galactosidase	lys(3)-ile-pro-ile-lys	20
Nucleoprotein (Influenza virus)	Chimpanzee alpha-globin	ala(336)-ala-phe-glu-asp -leu-arg-val-leu-ser	6
Large T antigen (Polyoma virus)	Chicken muscle Pyruvate kinase	val-ser-arg-lys(192)-arg -pro-arg-pro-pro-pro-lys- -lys(282)-ala-arg-glu-asp	43
E1a (Adenovirus)	E. coli galactokinase	lys-arg-pro-arg-pro	34
Nucleoplasmin (Xenopus oocyte)	Chicken muscle Pyruvate kinase	ala-val-lys-arg-pro-ala -ala-thr-lys-lys-ala-gly -glu-ala-lys-lys-lys	12
Fiber (Adenovirus)	E. coli ß-galactosidase	lys(2)-arg-ala-arg-pro -ser-glu-asp-thr-phe	the 3rd manuscript

protein to the nucleus, the existence of multiple nuclear signal sequences in one protein may effect the kinetics of entry into the nucleus. Proteins requiring bipartite nuclear localization signals also have been reported (36, 39). Xenopus protein N1 requires two signal sequences (VRKKRKT and AKKSKQE) separated by a 10 amino acid spacer (32). Adenovirus DNA-binding protein (35), polymerase basic protein 1 (PB1) (39), and polymerase basic protein 2 (PB2) (36) of influenza virus possess two discrete domains for nuclear localization, both of which are required for nuclear localization; one domain alone is not sufficient.

Glycosylation

Glycoproteins are proteins with one or several covalently linked carbohydrate groups. These carbohydrates serve many functions. They often affect the conformation of the polypeptide, making it more resistant to proteolytic digestion. Some carbohydrate residues are important in the sorting of proteins to their correct organelles within a cell. There are two major types of glycoproteins synthesized within the lumen of the endoplasmic reticulum and the Golgi apparatus; N-linked glycoproteins contain an oligosaccharide attached to the side chain of asparagine and O-linked glycoproteins contain an oligosaccharide attached through the hydroxyl group of either serine or threonine. These two major subgroups are distinct in oligosaccharide composition and mechanism of oligosaccharide processing (33).

N-linked glycosylation occurs in the endomembrane system: the endoplasmic reticulum, Golgi complex and related vesicles. Much is known about the carbohydrate structures and oligosaccharide processing of N-linked glycoproteins, which are the most common. Although Asn-X-Ser/Thr is the amino acid sequence required for N-linked glycosylation, only a small portion of these tripeptides are known to be glycosylated.

Therefore, the tripeptide Asn-X-Ser/Thr is essential but not sufficient for N-linked glycosylation.

O-linked glycosylation also occurs within the lumen of the Golgi apparatus. Many O-linked glycoproteins consist of an O-glycosydic bond from N-acetylgalactosamine to serine or threonine. A consensus amino acid sequence for O-linked glycosylation has not been identified and no specific amino acid sequence seems to be required for O-linked glycosylation. The secondary structure in the protein may influence the glycosylation of specific serine and threonine residues.

Cytoplasmic glycosylation pathways have also been described (23). These include novel forms of glycosylation in which N-acetylglucosamine (GlcNAc) is covalently attached to the protein by an O-glycosydic linkage. These novel glycoproteins are distinct from other O-glycoproteins in the types and number of saccharides involved. O-linked GlcNAc-containing proteins are localized predominantly in the nucleus and cytoplasm. Little is known about the signal sequence for O-linked GlcNAc addition although preliminary evidence from three proteins suggests that an acidic amino acid followed by a serine, a proline and a run of serine and threonine residues is required (21). The attachment site has not been identified in any other O-linked N-acetylglucosamine containing proteins.

A number of O-linked GlcNAc containing proteins have been identified in a diverse group of biological systems such as viruses (3, 4, 38), rodents (22, 25, 50), and humans (24, 29). Although little is known about the role that O-GlcNAc plays in a cell, suggested functions for this modification include: (a) the signal for targeting of some proteins to the nucleus (29, 50), (b) the proper assembly of multimeric complexes (7, 22, 24, 25), (c) the blocking of one or more regulatory (phosphorylation) sites in phosphoproteins (24, 29), (d) the prevention of proteolysis (7, 29, 49), and (e)

the activation of transcriptional function in the case of SP1 (29). The second manuscript in this dissertation shows that the first potential role listed above does not pertain to fiber.

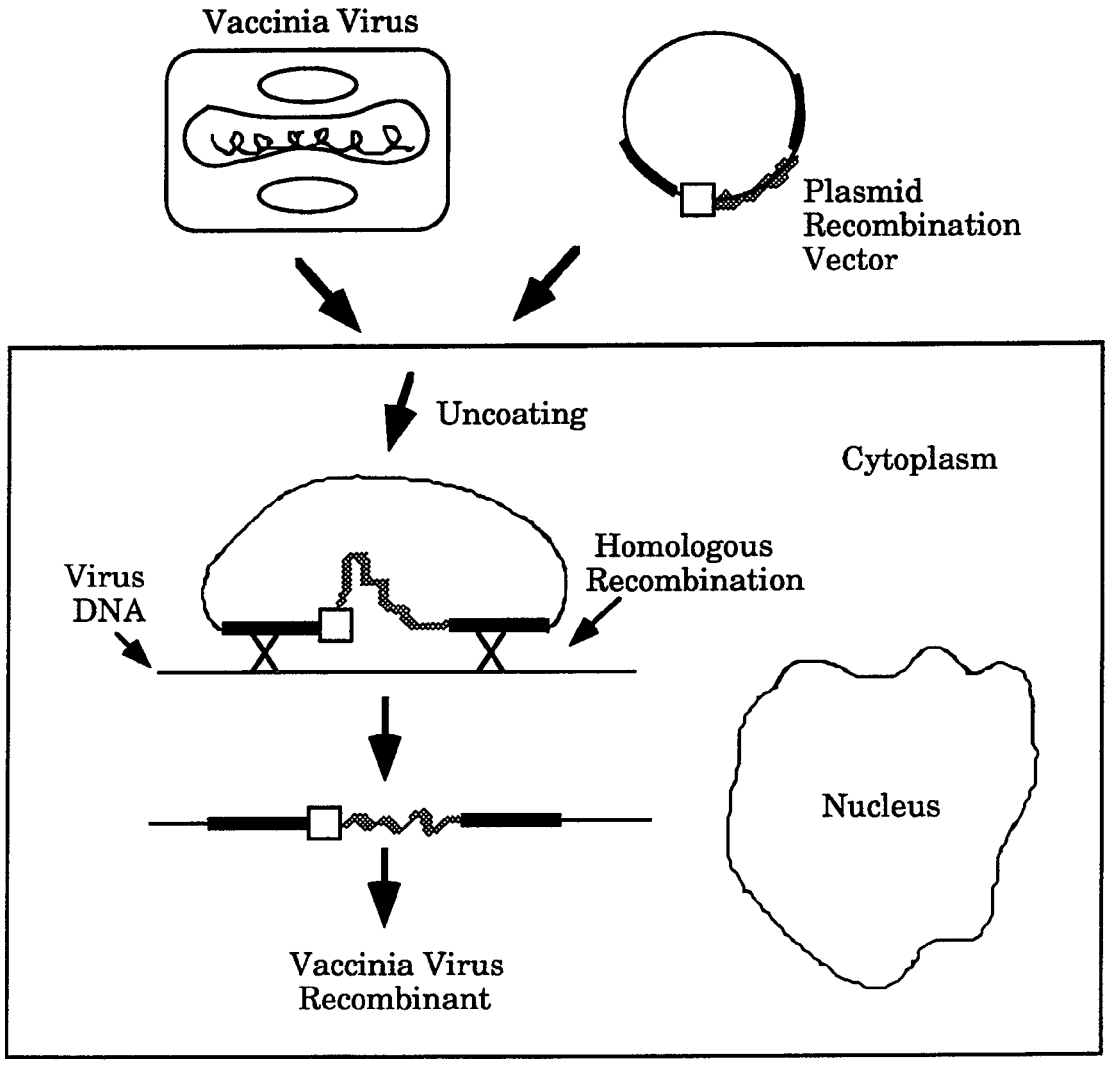
Recently, it has been shown that both Ad2 and Ad5 fibers contain a monosaccharide GlcNAc attached in an O-glycosidic linkage, but Ad7 fiber lacks this modification (4, 38). Many GlcNAc monosaccharides appear to be buried within the fiber structure. It has been suggested that GlcNAc may play a role in assembly or stabilization of Ad2 and Ad5 fiber trimers (38).

Recombinant vaccinia virus expression vector

Recombinant DNA technology has made it possible to insert and express heterologous genes in a variety of different viruses (45). Although most viral expression vectors including SV40, adenoviruses, and retroviruses offer simplicity of vector construction and relatively high level of expression, these viruses have a limited host range and/or limited capability to accommodate extra genetic information before becoming defective in replication. The large DNA viruses, such as vaccinia virus and herpes virus, are more difficult to engineer genetically than the viruses described above, but they have a wider host range and a greater capacity for a foreign DNA.

The large size of the vaccinia genome (approximately 180 kb with a coding capacity for over 150 polypeptides) has made the construction of the recombinant genomes by standard in vitro cloning techniques impossible. However, insertion of DNA into the virus genome has been achieved in infected cells by a two step process (Figure 3). In the first step, recombinant DNA techniques are used to construct a plasmid (insertion vector) that contains a chimeric gene. This plasmid contains a vaccinia virus promoter fused to a foreign gene, flanked by DNA from a

Figure 3. Generation of recombinant vaccinia virus. Cells infected with vaccinia virus are transfected with a plasmid construct. This construct contains a correctly oriented vaccinia virus promoter adjacent to a foreign gene coding sequence flanked on either side by virus DNA sequences. Homologous recombination occurs between the flanking sequences and virus genomic DNA with the resultant insertion of the foreign gene into virus DNA which can be packaged and produce a recombinant virus.


Vaccinia Virus
Plasmid
Recombination
Vector
Uncoating
Cytoplasm
Homologous
Recombination
Virus
DNA
Nucleus
Vaccinia Virus
Recombinant

nonessential region of the virus genome, usually the TK (thymidine kinase) gene. The second step is the insertion of this foreign gene into the vaccinia virus by homologous recombination in vivo. Recombinants can be isolated by selecting for the TK-minus (TK^-) phenotype.

Falkner and Moss (13) have constructed a series of vaccinia virus vectors in which coexpression of the Escherichia coli gpt gene provides efficient dominant selection of recombinants and therefore overcomes all of the deficiencies of TK^- selection such as attenuated infectivity of TK^- virus, use of TK^- cell line, use of mutagenic selective agents (e.g., 5-bromodeoxyuridine), and spontaneous development of TK^- mutants. In this selection system, mycophenolic acid, an inhibitor of purine metabolism, reversibly blocks formation of vaccinia virus plaques which can be overcome by expressing xanthine-guanine phosphoribosyl-transferase from the gpt gene and adding xanthine and hypoxanthine to the medium (37).

Use of a vaccinia virus expression vector throughout the research on fiber has offered several advantages; i) the recombinant viruses can be selected by a one step dominant selection, ii) since the virus carries out its entire life-cycle in the cytoplasm of the cell, the nucleus remains intact and nuclear translocation of a protein can be investigated without difficulty, iii) posttranslational modification such as glycosylation occurs faithfully, and iv) the recombinant virus expresses a significant amount of protein in a short time.

Problems to be studied

Viruses have often been used as a model to elucidate a number of intracellular events since they use established cellular pathways throughout their lifecycle. Adenovirus fiber is a glycoprotein which is involved in two major functions: attachment of the virion to receptors on

the cell surface and type-specific antigenicity. In this dissertation, fiber has been used to study assembly and transport of adenovirus structural proteins.

Adenovirus fiber is an excellent tool to study translocation and assembly of a nuclear, oligomeric protein for several reasons. Structural proteins of adenovirus are made in large excess in virus infected cells. The virus must contain a very efficient transport mechanism to translocate these proteins into the nucleus to be assembled into virus particles. Also fiber genes from several adenovirus serotypes have been cloned and subjected to extensive sequence analysis. Valuable information has been obtained from comparisons of these sequences. Fiber forms a very stable trimer and also contains a novel form of glycosylation. The abundance of this protein and the availability of its genes provide an opportunity to study relationships among nuclear translocation, trimerization, and glycosylation of fiber. Taken together, these characteristics make fiber an ideal system for the study of similar cellular events.

This dissertation accomplishes four goals: i) characterization of early region 3 and fiber genes of Ad7, ii) investigation of the postulated role of glycosylation on translocation of fiber to the nucleus, iii) identification of a nuclear translocation signal for adenovirus fiber, iv) investigation of the requirements for trimerization of fiber.

The first manuscript presents the nucleotide sequence and predicts amino acid sequences for open reading frames (ORF) encoded in the BamHI D fragment of Ad7, and provides comparisons of these sequences to those found in Ad2, Ad3, and Ad5. The fiber gene of late region 5 (L5) is found within this fragment. S1 mapping experiments show that the 5' and 3' boundaries of the main exon in fiber mRNA lie at each end of the proposed fiber ORF. The predicted amino acid sequence of Ad7 fiber shares

60% amino acid homology to Ad3 fiber, but only 20% to Ad2 fiber. The knowledge gained from this part of the study provides an essential basis for further studies of fiber.

The second manuscript describes nuclear transport of fiber molecules that lack O-linked N-acetylglucosamine modification. Adenovirus fiber protein is synthesized from cytoplasmic polyribosomes and transported to the nucleus for assembly into virions. The fact that O-linked GlcNAc-containing proteins are localized predominantly in the nucleus and cytoplasm of cells has led to the suggestion that O-linked GlcNAc might be a unique targeting signal for nuclear transport. Monoclonal antibodies that recognize adenovirus capsid proteins were prepared. Two specific antibodies which are important throughout this dissertation have been obtained: anti-fiber antibody, 4D2-5, which recognizes both monomeric and trimeric fibers from Ad2, Ad5, and Ad7; and anti-trimer antibody, 2A6-36, which recognizes only the trimer form of fibers from Ad2 and Ad5. The focus of this study was to investigate the role of O-linked GlcNAc on nuclear transport of adenovirus fiber protein. The results of this study on H5ts142, a trimerization and glycosylation defective fiber mutant of Ad5, demonstrate that both glycosylated and nonglycosylated Ad5 fiber can be transported efficiently to the nucleus, suggesting that the presence of O-linked GlcNAc is not obligatory for correct targeting of fiber.

The third manuscript presents the nuclear localization signal of the adenovirus fiber protein. Recently the peptide sequence which directs a protein to the nucleus has been discovered in a variety of nuclear proteins (Table 2). Vaccinia virus expression vector allows the study of the nuclear localization of fiber independently from adenovirus infection. Immunofluorescence using the monoclonal antibodies, 4D2-5 and 2A6-36,

made previously against the Ad2 virion was the major experimental technique used to visualize fiber within the cells. Although KRARP or TKRVRL (from N-terminus of Ad2 or Ad7 fibers, respectively) is an essential component of fiber for nuclear translocation, neither sequence is sufficient to accumulate a hybrid protein in the nucleus when fused to the N-terminus of ß-galactosidase; on the other hand, fusion of KRPRP (a known targeting sequence from the C-terminus of Ad2 E1A protein) or of KRARPSEDTF (from Ad2 fiber) to ß-galactosidase resulted in nuclear localization of the hybrid protein. Perhaps KRARP context is important. Changing the charge of amino acids 91 and 92 within another potential targeting sequence (LKKTK to LEETK) has little effect on nuclear targeting. Deletions at the C-terminal end of Ad2 fiber have no apparent effect on nuclear translocation of the protein. Relocation of KRARP from the N-terminus to the C-terminus of Ad2 fiber did not restore the ability of fiber to accumulate in the nucleus.

The fourth manuscript describes the experiments which map the domains important for assembly of the adenovirus fiber trimer. The fiber gene of the temperature sensitive mutant, H5ts142, which fails to make trimer at 39.5°C, was sequenced. A single amino acid change in the fiber gene is sufficient to cause this defective phenotype. In addition, the importance of the C-terminal end of fiber on assembly of trimeric structure is discussed; the extreme C-terminus of fiber seems to be essential for trimerization of fiber.

REFERENCES FOR INTRODUCTION

1. Akey, C. W. and Goldfarb, D. S. 1989. Protein import through the nuclear pore complex is a multistep process. J. Cell Biol. 109:971-982.

2. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D. 1983. "Chapter 7. Internal membranes and the synthesis of macromolecules." In Molecular biology of the cell. Garland Publishing, Inc., New York & London, p319-481.

3. Benko, D. M., Haltiwanger, R. S., Hart, G. W. and Gibson, W. 1988. Virion basic phosphoprotein from human cytomegalovirus contains O-linked N-acetylglucosamine. PNAS. 85:2573-2577.

4. Caillet-Boudin, M.-L., Strecker, G. and Michalski, J.-C. 1989. O-linked GlcNAc in serotype-2 adenovirus fibre. Eur. J. Biochem. 184:205-211.

5. Chroboczek, J. and Jacrot, B. 1987. The sequence of adenovirus fiber: similarities and differences between serotypes 2 and 5. Virol. 161:549-554.

6. Davey, J., Dimmock, N. J. and Colman, A. 1985. Identification of the sequence responsible for the nuclear accumulation of the influenza virus nucleoprotein in Xenopus oocytes. Cell. 40:667-675.

7. Davis, L. I. and Blobel, G. 1986. Identification and characterization of a nuclear pore complex protein. Cell. 45:699-709.

8. Defer, C., Marie-Thérèse, B., Caillet-Boudin, M.-L. and Boulanger, P. 1990. Human adenovirus-host cell interactions: comparative study with members of subgroups B and C. J. Virol. 64:3661-3673.

9. Devaux, C., Adrian, M., Berthet-Colominas, C., Crusack, S. and Jacrot, B. 1990. Structure of Adenovirus Fibre I. Analysis of Crystals of fibre from Adenovirus serotype 2 and 5 by electron microscopy and X-ray crystallography. J. Mol. Biol. 215:567-588.

10. Devaux, C., Caillet-Boudin, M.-L., Jacrot, B. and Boulanger, P. 1987. Crystallization, enzymatic cleavage, and the polarity of the adenovirus type 2 fiber. Virol. 161:121-128.

11. Devaux, C., Zulauf, M., Boulanger, P. and Jacrot, B. 1982. Molecular weight of adenovirus serotype 2 capsomers: A new characterization. J. Mol. Biol. 156:927-937.

12. Dingwall, C. and Laskey, R. A. 1986. Protein import into the cell nucleus. Ann. Rev. Cell Biol. 2:367-390.

13. Falkner, F. G. and Moss, B. 1988. Escherichia coli gpt gene provides dominant selection for vaccinia virus open reading frame expression vectors. J. Virol. 62:1849-1854.

14. Finlay, D. R., Newmeyer, D. D., Price, T. M. and Forbes, D. J. 1987. Inhibition of in vitro nuclear transport by a lectin that binds to nuclear pores. J. Cell Biol. 104:189-200.

15. Fischer-Fantuzzi, L. and Vesco, C. 1988. Cell-dependent efficiency of reiterated nuclear signals in a mutant simian virus 40 oncoprotein targeted to the nucleus. Mol. Cell. Biol. 8:5495-5503.

16. Flint, S. J. and Broker, T. R. 1981. "Adenoviruses." In DNA tumor viruses. J. Tooze ed. Chapters 8-10. CSH Laboratory, Cold Spring Harbor, NY.

17. Ginsberg, H. S. 1984. "The adenoviruses," Plenum Press, New York.

18. Green, N. M., Wrigley, N. G., Russell, W. C., Martin, S. R. and McLachlan, A. D. 1983. Evidence for a repeating cross-ß sheet structure in the adenovirus fibre. EMBO J. 2:1357-1365.

19. Greenspan, D., Palese, P. and Krystal, M. 1988. Two nuclear location signals in the influenza virus NS1 nonstructural protein. J. Virol. 62:33020-3026.

20. Hall, M. N., Hereford, L. and Herskowitz, I. 1984. Targeting of E.coli ß-galactosidase to the nucleus in yeast. Cell. 36:1057-1065.

21. Haltiwanger, R. S., Holt, G. D. and Hart, G. W. 1990. Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. J. Biol. Chem. 265:2563-2568.

22. Hanover, J. A., Cohen, C. K., Willingham, M. C. and Park, M. K. 1987. O-linked N-acetylglucosamine is attached to proteins of the nuclear pore: Evidence for cytoplasmic and nucleoplasmic glycoproteins. J. Biol. Chem. 262:9887-9894.

23. Hart, G. W., Haltiwanger, R. S., Holt, G. D. and Kelly, W. G. 1989. Glycosylation in the nucleus and cytoplasm. Ann.Rev. Biochem. 58:841-874.

24. Holt, G. D., Haltiwanger, R. S., Torres, C.-R. and Hart, G. W. 1987. Erythrocyte contains cytoplasmic glycoproteins. J. Biol. Chem. 262:14847-14850.

25. Holt, G. D., Snow, C. M., Senior, A., Haltiwanger, R. S., Gerace, L. and Hart, G. W. 1987. Nuclear pore complex glycoproteins contain cytoplasmically disposed O-linked N-Acetylglucosamine. J. Cell Biol. 104:1157-1164.

26. Hong, J. S., Mullis, K. G. and Engler, J. A. 1988. Characterization of the early region 3 and fiber genes of Ad7. Virol. 167:545-553.

27. Hunt, T. 1989. Cytoplasmic anchoring proteins and the control of nuclear localization. Cell. 59:949-951.

28. Ishibashi, M. and Maizel, J. V. 1974. The polypeptides of adenovirus VI. Early and late glycoproteins. Virol. 58:345-361.

29. Jackson, S. P. and Tjian, R. 1988. O-Glycosylation of eukaryotic transcription factors:Implications for mechanisms of transcriptional regulation. Cell. 55:125-133.

30. Kalderon, D., Richardson, W. D., Markham, A. F. and Smith, A. E. 1984. Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature. 311:33-38.

31. Kalderon, D., Roberts, B. L., Richardson, W. D. and Smith, A. E. 1984. A short amino acid sequence able to specify nuclear location. Cell. 39:499-509.

32. Kleinschmidt, J. A. and Seiter, A. 1988. Identification of domains involved in nuclear uptake and histone binding of protein N1 of Xenopus laevis. EMBO J. 7:1605-1614.

33. Kornfeld, R. and Kornfeld, S. 1985. Assembly of asparagine-linked oligosaccharides. Ann. Rev. Biochem. 54:631-664.

34. Lyons, R. H., Ferguson, B. Q. and Rosenberg, M. 1987. Pentapeptide nuclear localization signal in adenovirus E1a. Mol.Cell.Biol. 7:2451-2456.

35. Morin, N., Delsert, C. and Klessig, D. F. 1989. Nuclear localization of adenovirus DNA-binding protein: requirements of two signals and complementation during virus infection. Mol. Cell. Biol. 9:4372-4380.

36. Mukaigawa, J. and Nayak, D. P. 1991. Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2. J. Virol. 65:245-253.

37. Mulligan, R. and Berg, P. 1981. Selection for animal cells, that express the E. coli gene coding for xanthine-guanine phosphoribosyl transferase. Proc. Natl. Acad. Sci. USA. 78:2072-2076.

38. Mullis, K. G., Haltiwanger, R. S., Hart, G. W., Marchase, R. B. and Engler, J. A. 1990. Relative accessibility of N-acetylglucosamine in trimers of the adenovirus types 2 and 5 fiber proteins. J. Virol. 64:5317-5323.

39. Nath, S. T. and Nayak, D. P. 1990. Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). Mol. Cell Biol. 10:4139-4145.

40. Newmeyer, D. D., Finlay, D. R. and Forbes, D. J. 1986. In vitro transport of a fluorescent nuclear protein and exclusion of non-nuclear proteins. J. Cell. Biol. 103:2091-2102.

41. Newmeyer, D. D. and Forbes, D. J. 1988. Nuclear import can be separated into distinct steps in vitro: Nuclear pore binding and translocation. Cell. 52:641-653.

42. Pfeffer, S. R. and Rothman, J. E. 1987. Biosynthetic protein transport and sorting by the endoplasmic reticulum and golgi. Ann.Rev.Biochem. 56:829-852.

43. Richardson, W. D., Mills, A. D., Dilworth, S. M., Laskey, R. A. and Dingwall, C. 1988. Nuclear protein migration involves two steps: Rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. Cell. 52:655-664.

44. Richardson, W. D., Roberts, B. L. and Smith, A. E. 1986. Nuclear location signals in polyoma virus large-T. Cell. 44:77-85.

45. Rigby, P. W. J. 1983. Cloning vectors derived from animal viruses. J. Gen. Virol. 64:255-266.

46. Rihs, H.-P. and Peters, R. 1989. Nuclear transport kinetics depend on phosphorylation site-containing sequences flanking the karyophilic signal of the Simian virus 40 T-antigen. EMBO J. 8:1479-1484.

47. Roberts, B. L., Richardson, W. D. and Smith, A. E. 1987. The effect of protein context on nuclear location signal function. Cell. 50:465-475.

48. Rose, J. K. and Doms, R. W. 1988. Regulation of protein export from the endoplasmic reticulum. Ann. Rev. Cell Biol. 4:257-288.

49. Rowe, W. P., Huebner, A. J., Gilmore, L. K., Parrot, R. N., and Ward, T. G. 1953. Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Proc. Soc. Exp. Biol. Med. 84:570-573.

50. Schindler, M., Hogan, M., Miller, R. and DeGaetano, D. 1987. A nuclear specific glycoprotein representative of a unique pattern of glycosylation. J. Biol. Chem. 262:1254-1260.

51. Shu, L., Hong, J. S., Wei, Y. F. and Engler, J. A. 1986. Sequence of the genes encoded in early region 2B of human adenovirus type 12. Gene. 46:187-195.

52. Signas, C., Akusjarvi, G. and Petterson, U. 1985. Adenovirus 3 fiber polypeptide gene: Implications for the structure of the fiber protein. J. Virol. 53:672-678.

53. Smith, A. E., Kalderon, D., Roberts, B. L., Colledge, W. H., Edge, M., Gillett, P., Markham, A., Paucha, E. and Richardson, W. D. 1985. The nuclear location signal. Proc. R. Soc. Lond. B. 226:43-58.

54. Van Oostrum, J. and Burnett, R. M. 1985. Molecular composition of the adenovirus type 2 virion. J. Virol. 56:439-448.

55. Walter, P. and Lingappa, V. 1986. Mechanism of protein translocation across the endoplasmic reticulum. Ann. Rev. Cell Biol. 2:499-516.

56. Warren, G. 1987. Signals and salvage sequences. 327:17-18.

CHARACTERIZATION OF THE EARLY REGION 3 AND FIBER GENES OF AD7

by

JEONG SHIN HONG, KAREN GHEESLING MULLIS, AND JEFFREY A. ENGLER

Virology, 167:545-553, 1988

AUTHORSHIP

This work was a collaborative effort. Approximately half of the dideoxy sequencing was the work of Karen Mullis.

ABBREVIATIONS

Ad2, Ad3, Ad4, Ad5, Ad7, Ad12, Ad35, adenovirus serotypes 2, 3, 4, 5, 7, 12, and 35 (respectively); bp, base pairs; g.c., genome coordinate; GlcNAc, N-acetylglucosamine; nm, nanometers; ORF, open reading frame

ABSTRACT

The nucleotide sequence and the predicted amino acid sequences for open reading frames (ORFs) encoded in the BamHI D fragment of Ad7 (Gomen) DNA show an organization and conservation of potential polypeptides between Ad3 and Ad7. Five ORFs encoded within early region 3 (E3) and shared with the corresponding region of Ad3 can be identified; four of these potential coding regions also share homology to ORFs found in E3 of Ad2 and Ad5. The fiber gene of late region 5 (L5) is also apparent within this region; S1 mapping experiments show that the 5' and 3' boundaries of the main exon in fiber mRNA lie at each end of the proposed fiber ORF. The predicted amino acid sequence for Ad7 fiber shares 60% amino acid homology to Ad3 fiber, but only 20% to Ad2 fiber. Surprisingly, there are three regions of partial amino acid homology near the N- and C-termini of the predicted fiber gene sequences from Ad2, Ad3, Ad5, and Ad7; these conserved regions may be important for interaction with penton base, for proper folding of the shaft of the molecule, or for recognition of the cellular receptor to which adenovirus attaches during infection.

INTRODUCTION

Human adenoviruses are DNA tumor viruses that infect HeLa cells in culture, mobilize cellular and virus components for expression of viral proteins, and assemble new capsid particles in the cell nucleus. The

nonenveloped icosahedral shell of the virus is composed primarily of the hexon, penton base, and fiber proteins. Fiber is a glycoprotein important to a number of adenovirus functions, including assembly of virus particles, attachment of virus to the cell membrane during infection, and both type- and group-specific antigenicity (reviewed in Pettersson, 1984). The length of the fiber, as observed in the electron microscope, is characteristic of a given adenovirus subtype, and varies from 11 nm for Ad7 to almost 50 nm for some avian adenoviruses (Norrby, 1969). Dorsett and Ginsberg (1975) have suggested that fiber is a trimer that contains one polypeptide which is post-translationally modified in a different way from the other two subunits. The native fiber polypeptide from Ad2 has been reported to have a molecular weight of about 200 kDa (Dorsett and Ginsberg, 1975), suggesting a trimeric structure, but much lower values of 150 - 160 kDa have also been reported, consistent with a dimeric structure (Devaux et al., 1982). Green et al. (1983) have proposed that the rod part of the fiber is composed of two amphipathic b-sheets, held together by hydrophobic interactions. Signäs et al. (1985) reported that the primary sequence of the Ad3 fiber is compatible with this proposed structure.

The fiber polypeptide of Ad2 and Ad5 (group C viruses) and in Ad3 and Ad7 (group B viruses) is encoded in the L5 region of the adenovirus genome (Chow et al., 1977; Kilpatrick et al., 1979; Cladaras and Wold, 1985; Signäs et al., 1985; Chroboczek and Jacrot, 1987). In Ad2, fiber mRNA can sometimes also contain additional leader sequences (x, y, and z), in addition to the common tripartite leader. These additional leader sequences are encoded in region E3 which is upstream of the fiber gene in the adenovirus genome. The E3 region in Ad2 and Ad5, located between g.c. 76 and 86, encodes at least 9 different mRNAs which can be subdivided into two families of molecules (E3A and E3B) with the same 3' ends,

depending on which polyadenylation site is used (Chow et al., 1979; Kitchingman and Westphal, 1980; Cladaras et al., 1985). Region E3 codes for a large number of r-strand transcripts and polypeptides. Very little is known about the function of region E3. The nucleotide sequence of region E3 is well conserved between Ad2 and Ad5 (Cladaras and Wold, 1985) and, where known, Ad3 and Ad7 (this manuscript). On the other hand, this region seems to be dispensable for the virus growth in tissue culture cells, suggesting that this region might play an important role during the virus life cycle in its natural host. Analysis of the DNA sequence of the E3 region of Ad3 suggests that some ORFs are organized in a similar way in Ad2 and Ad3, but that other ORFs appear to be unique to one serotype or the other (Signäs et al., 1986).

In this paper, the nucleotide sequence of the BamHI D fragment of the class B adenovirus Ad7 (genome coordinates between 83.1 and 92.9) is presented. This DNA fragment of Ad7 includes a part of the E3 region, entire L5 region, and a portion of early region 4. The nucleotide and the predicted amino acid sequences of the genes encoded in this fragment have been compared to their apparent homologs in Ad2 (subgroup C) and Ad3 (subgroup B). Within the region compared, ORFs identified in Ad3 can also be found in Ad7; the predicted amino acid sequences of these ORFs in Ad3 and Ad7 showed homologies ranging from 60 to 100%. S1 nuclease mapping methods have been used to identify the positions of the exon that encodes the fiber polypeptide.

MATERIALS AND METHODS

A. Enzymes and Buffers. Restriction endonucleases were purchased from New England Biolabs, Bethesda Research Laboratories, or Boehringer Mannheim Biochemicals. T4 polynucleotide kinase and the Klenow fragment of DNA polymerase I from E. coli were purchased from

Pharmacia PL Biochemicals. TAE Buffer and TBE buffer used for gel electrophoresis are described in Maniatis et al. (1982).

B. Preparation of DNA and Nucleotide Sequence Analysis. Plasmid pJB740 (Engler and Kilpatrick, 1981) contains the Ad7 (Gomen) BamHI D fragment (Ad7 genome coordinates 83.1 to 92.9). This Ad7 DNA fragment was subcloned into pUC119 (Vieira and Messing, 1987) in both orientations (pUC1 and pUC7), to provide a source of single-stranded DNA containing the entire BamHI D fragment. Plasmid DNA was prepared by the alkaline lysis procedure (Birnbaum and Doly, 1979), followed by RNAse treatment and precipitation of the DNA with polyethylene glycol (PEG 8000). Single-strand DNA was made from pUC1 or pUC7 by infecting bacteria containing each plasmid with a helper M13 bacteriophage (VCM-13; Stratagene Cloning Systems, San Diego, CA); after 6 hours of infection, phage were collected by PEG precipitation, followed by several phenol extractions to release and purify single-stranded phage DNA.

The nucleotide sequence was determined by dideoxynucleotide sequencing (Sanger et al., 1977, 1980; Biggin et al., 1983). During the initial stage of sequencing, M13 phage containing sonicated pJB740 DNA fragments (cloned into the SmaI cleavage site of M13 DNA) was used as a source of single-stranded template DNA; M13 plaques containing Ad7-specific DNA were identified by making nitrocellulose filter copies of the plaques and probing the lifts with RNA probes made with the SP6 RNA polymerase and clones synthesizing RNA complementary to either the sense (pSP51) or antisense (pSP41) strands of the Ad7 DNA. At later stages of sequencing, oligonucleotide primers were synthesized on an Applied Biosystems Model 380A DNA synthesizer, in order to fill the remaining unknown regions in the DNA sequence. Either double-stranded or single-stranded pUC1 or pUC7 was used for the sequence analysis with the

synthesized primers. Each nucleotide of each strand of the sequence was determined at least twice. The sequence obtained was recorded and analyzed using computer programs developed by Staden (1977, 1978) and by Devereux et al. (1984).

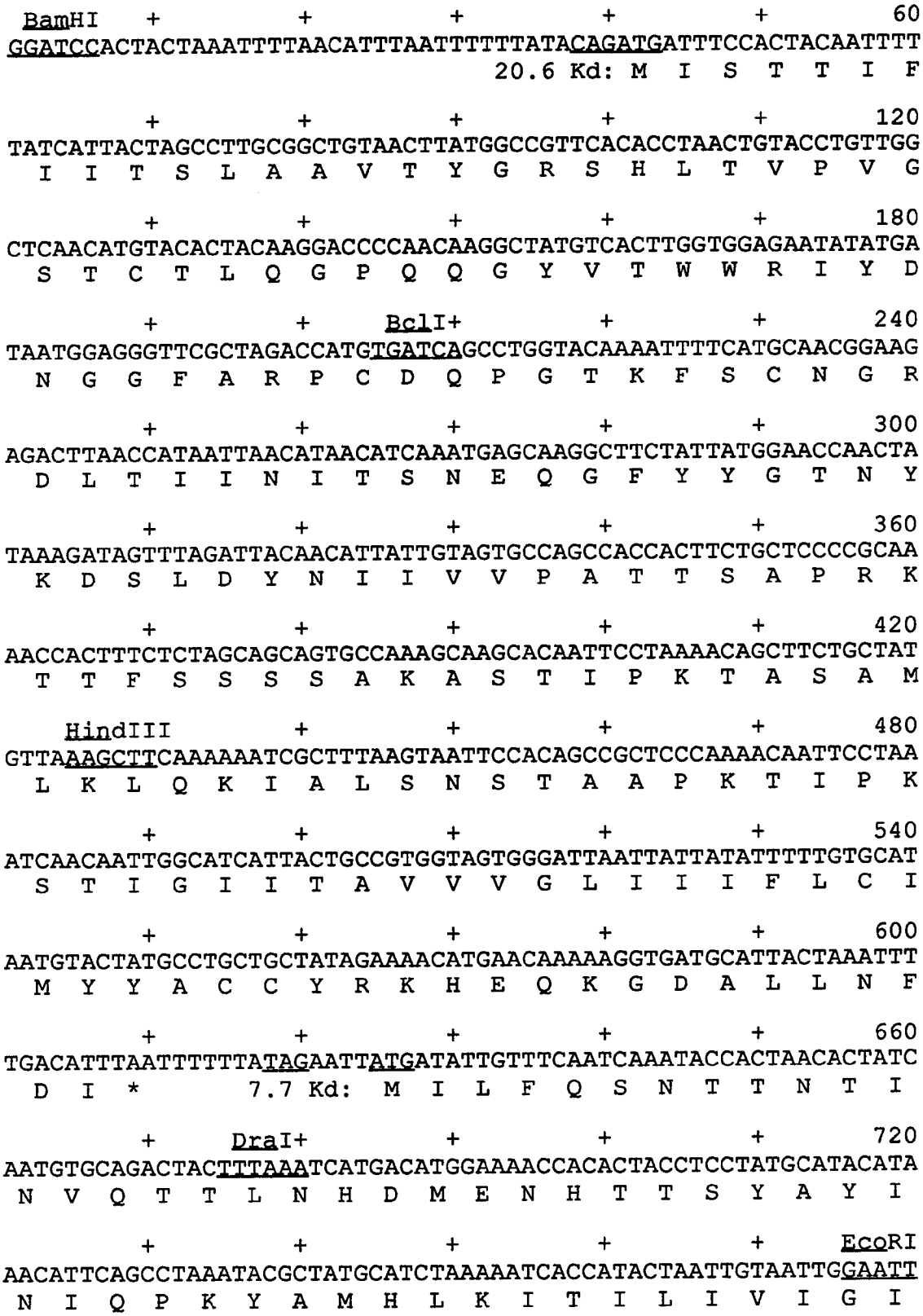
C. S1 nuclease mapping of the 5' and 3' boundaries of the exon that encodes the Ad7 fiber polypeptide. Ad7 virus and A549 cells were purchased from ATCC (American Type Culture Collection). Late viral RNA was prepared from Ad7-infected A549 cells grown in F12 medium at 24 hours postinfection by using guanidinium isothiocynate (Maniatis et al., 1982). The 5' probe for S1 mapping was isolated by SacI and BstNI digestion of a EcoRI-BamHI fragment cloned from pJB740 into pSP65 (Promega) and was labelled at its 5' end with T4 polynucleotide kinase and [g-^{32}P]rATP. The probe for S1 analysis of the 3' end of fiber mRNA was isolated from DNA from an SV40 recombinant virus that contains the Ad7 fiber gene; this fiber-gene-containing fragment (from plasmid pQP7F; J.S.Hong and J.A.Engler, unpublished) was cloned into the BamHI cleavage site of Bluescript-SK(+) (Stratagene Cloning Systems); a BstNI-PstI fragment from within the fiber gene (nucleotide 3061 to a PstI site located in an adjacent polylinker sequence located a few bp beyond the BamHI site at nucleotide 3566 in Fig. 1) was labelled by fill-in of the 5' overhang of the BstNI cleavage site with [a-^{32}P]dATP and the Klenow fragment. The 5'- or 3'-end labelled DNA fragments were heat-denatured for 15 minutes at 80°C and hybridized to 50 mg of total cytoplasmic RNA for 3 hrs at 50°C. S1 nuclease treatment of the hybrids formed was performed as described by Maniatis et al. (1982), using 10 units of S1 nuclease per mg of RNA for each S1 nuclease reaction. The S1 nuclease-resistant material was analyzed by electrophoresis on a 5% polyacrylamide denaturing gel containing 7 M urea and TBE buffer.

RESULTS

The organization and protein coding capacity of the 3' half of early region 3 in Ad7. The nucleotide sequence of the Ad7 BamHI D fragment is shown in Fig. 1. A comparison of the ORFs encoded in the Ad7 sequence with the corresponding sequences from Ad3 is shown in Fig 2. The organization in the region compared is highly homologous with regard to both amino acids and nucleotides. The size of the E3 regions compared are 2200 nucleotides in Ad3 and 2176 nucleotides in Ad7; most of the observed difference in length is due to a series of deletions and insertions between Ad3 and Ad7, particularly within the 7.7 KDa (Ad7)/9.0 KDa (Ad3) ORF and in the interval between the end of the 15.3 KDa ORF and the beginning of the fiber gene (see below). Based on the homology with Ad3, approximately 50% of the sequence of the entire E3 region of Ad7 has been determined, since the BamHI site at g.c. 83.1 of Ad7 (Tibbetts, 1977) lies within the carboxy terminal portion of the 20.1 KDa gene identified in the Ad3 sequence (Signäs et al., 1986). The Ad7 region E3 sequence that resides within the BamHI D fragment encodes five open reading frames that encode putative proteins with molecular weights greater than 9 KDa. The only AATAAA polyadenylation signal observed in the Ad7 E3 region presented here is found between nucleotides 1987 and 1992 (Fig. 1).

The reading frame for 20.6 KDa is the first complete ORF found in the Ad7 BamHI D fragment. The 20.6 KDa shares high nucleotide (97.9%) and amino acid (96.8%) homology with the Ad3 20.5 KDa (see Table 1), but no equivalent ORF has been observed in Ad2; this ORF is also found in Ad35 (Flomenberg et al., 1988). The TEST CODE algorithm (Fickett, 1982), a mathematical method that can be used to predict whether an ORF is protein coding and non-coding, suggests that this ORF of Ad7 probably encodes a polypeptide. Signäs et al. (1986) have shown by S1 analysis that

Figure 1: The nucleotide sequence of the BamHI D fragment of Ad7 (Gomen) DNA. Restriction cleavage sites within the sequence are underlined with the name of the enzyme written above. Potential splice acceptors and the initial ATG codons within ORFs that correspond to ORFs found in Ad3 are also underlined; the positions of two AATAAA sequences important for selection of polyadenylation sites (at nucleotides 1987 and 3159) are also underlined. The predicted amino acid sequence for these ORFs is written below the nucleotide sequence and identified by the calculated molecular weight of the predicted polypeptide. The positions of three potential N-linked glycosylation sites found in the predicted amino acid sequence of Ad7 fiber (encoded starting at nucleotides 2418, at 2751, and at 3027) are also underlined. The Ad7 ORF that shares homology with an unidentified ORF in early region 4 of Ad2 (41 out of 72 amino acids) is encoded in the opposite DNA strand from Ad7 nucleotides 3409 to 3191 (see Results).


BamHI + + + + + 60
GGATCCACTACTAAATTTTAACATTTAATTTTTTATACAGATGATTTCCACTACAATTTT
20.6 Kd: M I S T T I F
+ + + + + 120
TATCATTACTAGCCTTGCGGCTGTAACTTATGGCCGTTCACACCTAACTGTACCTGTTGG
I I T S L A A V T Y G R S H L T V P V G
+ + + + + 180
CTCAACATGTACACTACAAGGACCCCAACAAGGCTATGTCACTTGGTGGAGAATATATGA
S T C T L Q G P Q Q G Y V T W W R I Y D
+ + BclI+ + + 240
TAATGGAGGGTTCGCTAGACCATGTGATCAGCCTGGTACAAAATTTTCATGCAACGGAAG
N G G F A R P C D Q P G T K F S C N G R
+ + + + + 300
AGACTTAACCATAATTAACATAACATCAAATGAGCAAGGCTTCTATTATGGAACCAACTA
D L T I I N I T S N E Q G F Y Y G T N Y
+ + + + + 360
TAAAGATAGTTTAGATTACAACATTATTGTAGTGCCAGCCACCACTTCTGCTCCCCGCAA
K D S L D Y N I I V V P A T T S A P R K
+ + + + + 420
AACCACTTTCTCTAGCAGCAGTGCCAAAGCAAGCACAATTCCTAAAACAGCTTCTGCTAT
T T F S S S S A K A S T I P K T A S A M
HindIII + + + + 480
GTTAAAGCTTCAAAAAATCGCTTTAAGTAATTCCACAGCCGCTCCCAAAACAATTCCTAA
L K L Q K I A L S N S T A A P K T I P K
+ + + + + 540
ATCAACAATTGGCATCATTACTGCCGTGGTAGTGGGATTAATTATTATATTTTTGTGCAT
S T I G I I T A V V V G L I I I F L C I
+ + + + + 600
AATGTACTATGCCTGCTGCTATAGAAAACATGAACAAAAAGGTGATGCATTACTAAATTT
M Y Y A C C Y R K H E Q K G D A L L N F
+ + + + + 660
TGACATTTAATTTTTTATAGAATTATGATATTGTTTCAATCAAATACCACTAACACTATC
D I * 7.7 Kd: M I L F Q S N T T N T I
+ DraI+ + + + 720
AATGTGCAGACTACTTTAAATCATGACATGGAAAACCACACTACCTCCTATGCATACATA
N V Q T T L N H D M E N H T T S Y A Y I
+ + + + + EcoRI
AACATTCAGCCTAAATACGCTATGCATCTAAAAATCACCATACTAATTGTAATTGGAATT
N I Q P K Y A M H L K I T I L I V I G I

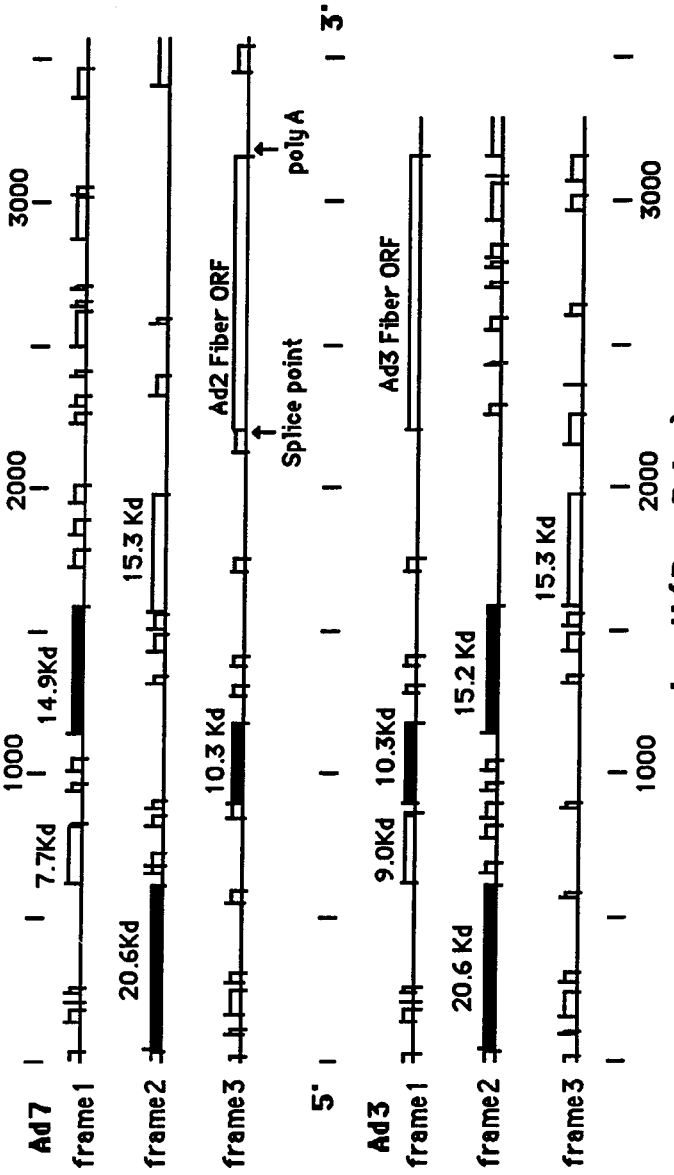
```
          +         +         +         +         +        840
CTTATACTATCTGTTATTCTTTATTTTCTATTCTCCTATGATTAGTCGTCCCCATATAGC
L  I  L  S  V  I  L  Y  F  L  F  S  Y  D  *

          +      BglII      +         +         +        900
TCTGAATGAAATCTAAGATCTTTTTTTTTTCTCTTACAGTATGGTGAACACCAATCATGA
                                                 10.3 Kd:  M

          +         +         +         +         +        960
TCCCTAGAAATTTCTTCTTCACCATACTCATCTGTGCTTTCAATGTCTGTGCTACTTTCA
I  P  R  N  F  F  F  T  I  L  I  C  A  F  N  V  C  A  T  F

          +         +         +         +         +       1020
CAGCAGTAGCCACTGCAAGCCCAGACTGTATAGGACCATTTGCTTCCTATGCACTTTTTG
T  A  V  A  T  A  S  P  D  C  I  G  P  F  A  S  Y  A  L  F

          +         +         +         +         +       1080
CCTTCGTTACTTGCATCTGCGTGTGTAGCATAGTCTGCCTGGTTATTAATTTTTTCCAAC
A  F  V  T  C  I  C  V  C  S  I  V  C  L  V  I  N  F  F  Q

          +         +         +         +         +       1140
TGGTAGACTGGATCTTTGTACGAATTGCCTACCTACGTCACCATCCCGAATACCGCAATC
L  V  D  W  I  F  V  R  I  A  Y  L  R  H  H  P  E  Y  R  N

          +         +      DraI       +         +       1200
AAAATGTTGCGGCACTTCTTAGGCTTATTTAAAACCATGCAGGCTATGCTACCAGTCATT
Q  N  V  A  A  L  L  R  L  I  *
                             14.9 Kd:  M  Q  A  M  L  P  V  I

          +         +         +         +         +       1260
TTAATTCTGCTACTACCCTGCATTGCCCTAGCTTCCACCGCCACTCGCGCTACACCTGAA
 L  I  L  L  L  P  C  I  A  L  A  S  T  A  T  R  A  T  P  E

          +         +      NcoI       +         +       1320
CAACTTAGAAAATGCAAATTTCAACAACCATGGTCATTTCTTGATTGCTACCATGAAAAA
 Q  L  R  K  C  K  F  Q  Q  P  W  S  F  L  D  C  Y  H  E  K

          +         +         +         +         +       1380
TCTGATTTCCCCACATACTGGATAGTGATTGTTGGAATAATTAACATACTTTCATGTACA
 S  D  F  P  T  Y  W  I  V  I  V  G  I  I  N  I  L  S  C  T

          +         +         +         + EcoRI   +       1440
TTTTTCTCAATCACAATATACCCCACATTTAATTTTGGGTGGAATTCTCCCAATGCACTG
 F  F  S  I  T  I  Y  P  T  F  N  F  G  W  N  S  P  N  A  L.

BstEII    +         +         +         +         +       1500
GGTTACCCACAAGAACCAGATGAACATATTCCACTACAACACATACAACAACCACTAGCA
 G  Y  P  Q  E  P  D  E  H  I  P  L  Q  H  I  Q  Q  P  L  A

          +         +         +         +         +       1560
CTGGTAGAGTATGAAAATGAGCCACAACCTTCACTGCCTCCTGCCATTAGTTACTTCAAC
 L  V  E  Y  E  N  E  P  Q  P  S  L  P  P  A  I  S  Y  F  N
```

```
          +         +         +         +         +      1620
CTAACCGGCGGAGATGACTGAAATACTCACCACCTCCAATTCCGCCGAGGATCTGCTTGA
 L  T  G  G  D  D  *
    15.3 Kd:  M  T  E  I  L  T  T  S  N  S  A  E  D  L  L  D

          +         +         +         +         +      1680
TATGGACGGCCGCGTCTCAGAACAGCGACTCGCCCAACTACGCATCCGCCAGCAGCAGGA
  M  D  G  R  V  S  E  Q  R  L  A  Q  L  R  I  R  Q  Q  Q  E

          +    SacI +         +         +         +      1740
ACGCGTGACCAAAGAGCTCAGAGATGTCATCCAAATTCACCAATGCAAAAAAGGCATATT
  R  V  T  K  E  L  R  D  V  I  Q  I  H  Q  C  K  K  G  I  F

          +         +         +         +         +      1800
TTGCTTGGTAAAACAAGCCAAAATATCCTACGAGATCACCGCTACTGACCATCGCCTCTC
  C  L  V  K  Q  A  K  I  S  Y  E  I  T  A  T  D  H  R  L  S

          +         +         +         +         +      1860
TTACGAACTTGGCCCCCAACGACAAAAATTTACATGCATGGTGGGAATCAACCCTATAGT
  Y  E  L  G  P  Q  R  Q  K  F  T  C  M  V  G  I  N  P  I  V

          +         +         +         +         +      1920
TATCACTCAGCAAAGTGGAGATACTAAGGGGTGCATTCACTGCTCTTGTGATTCCATCGA
  I  T  Q  Q  S  G  D  T  K  G  C  I  H  C  S  C  D  S  I  E

          +         +         +         +         +      1980
GTGCACCTACACCCTGCTAAAGACCCTATGCGGCTTAAGAGACCTGCTACCCATGAATTA
  C  T  Y  T  L  L  K  T  L  C  G  L  R  D  L  L  P  M  N  *

          +         +         +         +         +      2040
AAAATTAATAAAAAATCACTTACTTGAAATCAGCAATAAGGTCTCTGTTGAAATTTTCTC

          +         +         +         +         +      2100
CCAGCAGCACCTCACTTCCCTCTTCCCAACTCTGGTATTCTAAACCCCGTTCAGCGGCAT

          +    DraI +         +         +         +      2160
ACTTTCTCCATACTTTAAAGGGGATGTCAAATTTTAACTCCTCTCCTGTACCCACAATCT

          +         +         +         +         +      2220
TCATGTCTTTCTTCCAGATGACCAAGAGAGTCCGGCTCAGTGATTCCTTCAACCCTGTCT
           fiber:  M  T  K  R  V  R  L  S  D  S  F  N  P  V

          +         +         +         +         +      2280
ACCCCTATGAAGATGAAAGCACCTCCCAACACCCCTTTATAAACCCAGGGTTTATTTCCC
Y  P  Y  E  D  E  S  T  S  Q  H  P  F  I  N  P  G  F  I  S

          +         +         +        DraI       +      2340
CAAATGGCTTTACACAAAGCCCAGACGGAGTTCTTACTTTAAAATGTTTAACCCCACTAA
P  N  G  F  T  Q  S  P  D  G  V  L  T  L  K  C  L  T  P  L

          +         +         +         +         +      2400
CAACCACAGGCGGGTCTCTACAGTTAAAAGTGGGAGGGGGTCTTACAATAGATGACACCG
T  T  T  G  G  S  L  Q  L  K  V  G  G  G  L  T  I  D  D  T
```

```
          +         +         +         +         +      2460
ACGGTTTTTTGAAAGAAAACATAAGTGCCGCCACACCACTCGTTAAGACTGGTCACTCTA
D  G  F  L  K  E  N  I  S  A  A  T  P  L  V  K  T  G  H  S

          +         +         +         +         +      2520
TAGGTTTGTCGCTAGGACCCGGATTAGGAACAAATGAAAACAAACTTTGTGCCAAATTGG
I  G  L  S  L  G  P  G  L  G  T  N  E  N  K  L  C  A  K  L

          +         +         +         +         +      2580
GAGAAGGACTTACATTCAATTCCAACAACATTTGCATTGATGACAATATTAACACCCTAT
G  E  G  L  T  F  N  S  N  N  I  C  I  D  D  N  I  N  T  L

          HpaI      +         +         +         +      2640
GGACAGGAGTTAACCCCACCACAGCCAACTGTCAAATAATGGCCTCCAGTGAATCTAATG
W  T  G  V  N  P  T  T  A  N  C  Q  I  M  A  S  S  E  S  N

          +         +         +      StuI+        +      2700
ATTGCAAATTAATTCTAACACTAGTTAAAACTGGAGGCCTTGTCACTGCATTTGTTTATG
D  C  K  L  I  L  T  L  V  K  T  G  G  L  V  T  A  F  V  Y

          +         +         +         +         +      2760
TTATAGGAGTATCTAACGATTTTAATATGCTAACTACACATAAAAATATAAATTTCACTG
V  I  G  V  S  N  D  F  N  M  L  T  T  H  K  N  I  N  F  T

PstI      +         +         +         +         +      2820
CAGAGCTGTTTTTTGATTCTACTGGTAATTTACTAACTAGCCTTTCATCCCTAAAAACTC
A  E  L  F  F  D  S  T  G  N  L  L  T  S  L  S  S  L  K  T

          +         +         +         +         +      2880
CACTTAATCATAAATCAGGGCAAAACATGGCTACTGGTGCCCTTACTAATGCTAAAGGTT
P  L  N  H  K  S  G  Q  N  M  A  T  G  A  L  T  N  A  K  G

          +         +         +         +         +      2940
TCATGCCCAGCACAACTGCCTATCCTTTCAATGTTAATTCCAGAGAAAAAGAAAACTACA
F  M  P  S  T  T  A  Y  P  F  N  V  N  S  R  E  K  E  N  Y

          +         +      BclI         +         +      3000
TTTACGGAACTTGTTACTACACAGCTAGTGATCACACTGCTTTTCCCATTGACATATCTG
I  Y  G  T  C  Y  Y  T  A  S  D  H  T  A  F  P  I  D  I  S

          +         +         +         +         +      3060
TCATGCTTAACCAAAGAGCATTAAATAATGAGACATCATATTGTATTCGTGTAACTTGGT
V  M  L  N  Q  R  A  L  N  N  E  T  S  Y  C  I  R  V  T  W

          +         +         +         +         +      3120
CCTGGAACACAGGAGTTGCCCCAGAAGTGCAAACCTCTGCTACTACCCTAGTCACCTCTC
S  W  N  T  G  V  A  P  E  V  Q  T  S  A  T  T  L  V  T  S

          +         +         +         +         +      3280
CATTTACCTTTTACTACATTAGAGAAGACGACTGACAAAATAAAGTTTAACTTGTTTATT
P  F  T  F  Y  Y  I  R  E  D  D  *
```

```
DraI    +           +           +           +           +     3240
TAAAATCAATTCATAAAATTCGAGTAGTTATTTTGCCTCCCCCTTCCCATTTAACAGAAT
          * L I R T T I K G G G E W K V S Y

        +           +      DraI         +           +     3300
ACACCAATCTCTCCCCACGCACAGCTTTAAACATTTGGATACCATTACAGATAGACATAG
  V L R E G R V A K F M Q I G N C I S M T

        +           +           +           +           +     3360
TTTTAGATTCCACATTCCAAACAGTTTCAAAGCGAGCCAATCTGGGGTCAGTGATACATA
  K S E V N W V T E F R A L R P D T I C L

        +           + DraI      +           +           +     3420
AAAATGCATCGGGATAGTCTTTTAAAGCGCTTTCACAGTCCAACTGCTGCGGATGGACTC
  F A D P Y D K L A S E C D L Q Q  :unknown

        +           +           +           +           +     3480
CGGGGTCTGGATCACGGTCATCTGGAAGAAGAACGATGGGAATCATAATCCGAAAACGGA
E4 ORF (encoded in reverse strand)

        +           +           +           +           +     3540
ATCGGGCGATTGTGTCTCATCAAACCCACAAGCAGCCGCTGTCXTGCGTCGCTCCGTGCG

        +           + 3566
ACTGCTGCTTATAGGATCGGGATCC
                    BamHI
```

Figure 2: Comparison of the protein coding capacity of the Ad7 BamHI D fragment with the corresponding region of Ad3. Potential ORFs in each of the three reading frames encoded in the DNA are shown as rectangles above the solid line; the position of the first ATG in these frames is indicated with a vertical line above the rectangle, while the termination codon is shown as a vertical line extending below the rectangle. Open reading frames that correspond to ORFs reported for Ad2 (Hérissé et al., 1980; Hérissé and Galibert, 1981), Ad3 (Signäs et al., 1985, 1986), or Ad5 (Cladaras and Wold, 1985) are shown as filled rectangles with the calculated molecular weight of the encoded polypeptide written above the ORF; filled rectangles show Ad7 ORFs that share greater than 95% amino acid homology with Ad3, while stipled rectangles show Ad7 ORFs that are between 60% and 84% homologous with the corresponding Ad3 ORF. The positions of the 3' acceptor splice point and of the poly(A) addition site for Ad7 fiber mRNA are shown with vertical arrows.


Ad7
frame1
7.7Kd
14.9Kd
frame2
20.6Kd
15.3 Kd
frame3
10.3 Kd
Ad2 Fiber ORF
Splice point
poly A
5'
3'
Ad3
frame1
9.0Kd
10.3Kd
Ad3 Fiber ORF
frame2
20.6 Kd
15.2 Kd
frame3
15.3 Kd
1000
2000
3000
Length(Base Pairs)

Legend to Table I: Nucleotide and predicted amino acid sequence homology for ORFs in early region 3 and the fiber gene. In part A, the predicted molecular weights of the products encoded by the corresponding ORFs from Ad2, Ad3, and Ad7 are shown. In parts B, C, and D, the number of identical nucleotides or amino acids observed is written next to the total number of nucleotides and amino acids in the ORF; the calculated homology for each ORF is shown in parentheses.

Table I: Nucleotide and predicted amino acid sequence homology for ORFs in early region 3 and fiber gene.

A. ANALOGOUS ORFS OBSERVED IN AD2, AD3 AND AD7

AD2	AD3	AD7
---	20.5K	20.6K
11.6K	9.0K	7.7K
10.4K	10.3K	10.3K
14.5K	15.2K	14.9K
14.7K	15.3K	15.3K

B. NUCLEOTIDE HOMOLOGY OBSERVED

ORF	Ad7/Ad3	Ad7/Ad2
20.6	324/559 (98%)	no corresponding gene
7.7	204/209 (98%)	---
10.3	272/275 (99%)	149/275 (54%)
14.9	340/346 (99%)	162/346 (47%)
15.3	332/400 (83%)	288/400 (58%)

C. AMINO ACID HOMOLOGY OBSERVED

ORF	Ad7/Ad3	Ad7/Ad2
20.6	183/189 (97%)	no corresponding gene
7.7	43/66 (65%)	10/66 (16%) - scattered throughout gene
10.3	91/91 (100%)	45/91 (49%)
14.9	131/134 (99%)	46/131 (35%)
15.3	118/135 (87%)	65/135 (48%)

D. HOMOLOGIES OBSERVED WITHIN THE FIBER GENE

	Nucleotide	Amino Acid
Ad7/Ad2	413/962 (43%)	64/325 (20%)
Ad7/Ad3	615/962 (64%)	191/320 (60%)

an mRNA that may encode this polypeptide can be found in Ad3-infected cells.

A short ORF for a 7.7 KDa polypeptide shares 98% nucleotide homology with the ORF for a 9 KDa protein of Ad3. The Ad7 and Ad3 predicted polypeptides have diverged in an unusual way: there is a 51 bp (17 amino acids) insert near the beginning of this ORF in Ad7, and a 54 bp (18 amino acids) insert near the end of this ORF in Ad3. Between these two insertion points, there is relatively high amino acid homology (88%). Signäs et al. (1986) reported that the Ad3 polypeptide shared 27% amino acid homology to an 11.6 KDa polypeptide found at a corresponding location in the E3 region of Ad2 and that the highest homology was found in the C-terminal portion of the sequences. This region of homology between the Ad3 and the Ad2 polypeptides is encoded in the segment of the Ad3 ORF that is deleted in Ad7; there is not substantial amino acid homology between the Ad7 7.7 KDa and the Ad2 11.6 KDa polypeptide (10 out of 66 amino acids compared). Wold et al. (1984) have shown that the Ad2 11.6 KDa (and corresponding Ad5 10.5 KDa) polypeptide can be immune-precipitated from Ad2- and Ad5-infected cells. Whether proteins encoded by the corresponding Ad3 and Ad7 ORFs are also found in infected cells awaits further experiments; another group B adenovirus, Ad35, does not contain this ORF (Flomenberg et al., 1988). An ATTAAA polyadenylation signal for termination of the E3A family of mRNAs is found at the end of this ORF in Ad2 and Ad5, but is not observed in this region of either the Ad3 (Signäs et al., 1986) or Ad7 sequences.

Three additional E3 ORFs can also be found. A 10.3 KDa ORF of Ad7 is identical in amino acid sequence to the 10.3 KDa ORF of Ad3, and shares lower amino acid homology (49%) to the 10.4 KDa predicted polypeptide for Ad2. The TESTCODE algorithm predicts that this ORF in Ad7 may encode

a polypeptide. Genetic evidence for the existence of the 10.4 KDa polypeptide in Ad5 has recently been obtained (C. Carlson, A.E. Tollefson and W.S.M. Wold, personal communication).

A 14.9 KDa ORF overlaps the end of the 10.3 KDa ORF. This ORF shares high nucleotide and amino acid homology (99% and 98%, respectively) to the Ad3 ORF. This ORF also shares 31% amino acid homology with the predicted polypeptide sequence of the Ad2 14.5 KDa ORF. The initial ATG in this ORF (AAAATGT, starting at nucleotide 1144) is probably not a strong site for initiation of translation, based on the rules suggested by Kozak (1986). Both the second internal ATG (ACCATGC, encoded five nucleotides beyond the end of the 10.3 KDa ORF at nucleotide 1177), and the third ATG (GCTATGC, at nucleotide 1183) are in a stronger context for initiation of translation and may be the actual start point(s) for translation. The amino acid homology observed between this ORF and the Ad2 14.5 KDa ORF starts nearest to the third ATG in the Ad7 ORF.

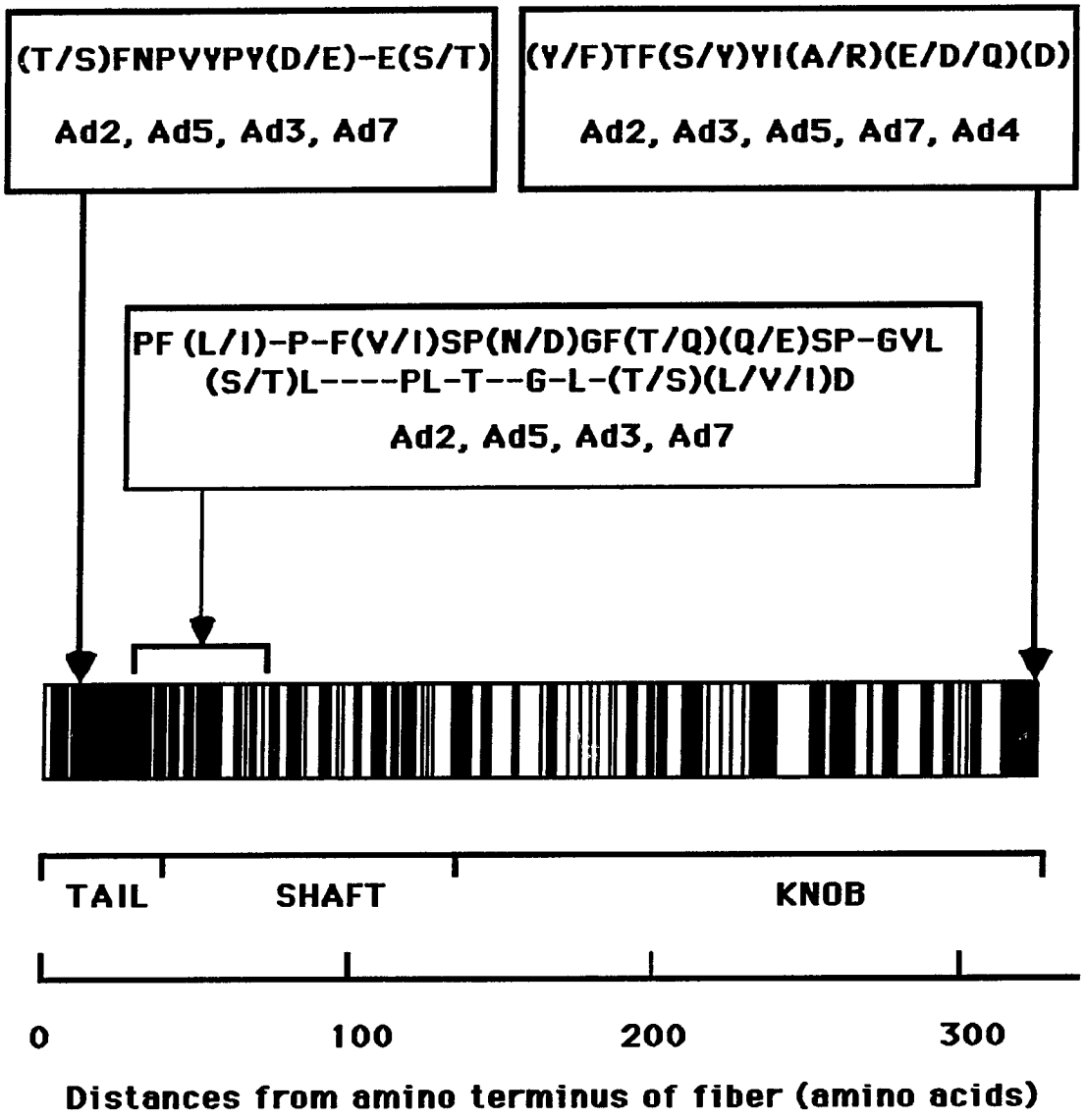
The end of the 14.9 KDa ORF overlaps with a 15.3 KDa ORF in early region 3. The predicted amino acid sequence for the Ad7 15.3 KDa ORF is also homologous to that of the Ad3 15.3 KDa ORF (84%) and of the Ad2 and Ad5 14.7 KDa ORF (45%). The TESTCODE algorithm makes no predictions about the protein coding potential of this ORF in Ad7, but the corresponding 14.7 KDa polypeptide in Ad5 can be immunoprecipitated from extracts of cells infected with Ad2 or Ad5 (Tollefson and Wold, 1988; Wang et al., 1988), suggesting that the Ad7 ORF might also be expressed. Gooding et al. (1988) have shown that the product of the Ad5 14.7 KDa ORF protects C3HA cells infected by Ad5 virus from the lytic effects of tumor necrosis factor, suggesting that this ORF may encode a protein that protects infected cells from lysis by one or more antiviral factors.

Organization and homology of the fiber gene region. The fiber gene of Ad7 is encoded starting at nucleotide 2178 and ending at nucleotide 3155 (Fig. 1). The nucleotide sequence homology within fiber gene between Ad7 and Ad2 is 43% (413/962 nucleotides compared) and the corresponding amino acid sequence homology is 20% (64/325 amino acids compared); the homology between the Ad7 and Ad5 fiber genes is similar. The degree of DNA homology observed within the fiber gene is lower than in many other regions in the genome; where comparison is possible, the nucleotide sequences of Ad5 and Ad7 in other protein-coding regions of the genome show between 52% (in the 289 amino acid protein from early region 1A; van Ormondt et al., 1980) and 80% DNA sequence homology (for the preterminal protein and the Ad Pol genes; Engler et al., 1983). This result matches well with the fact that the fiber protein contains type-specific antigenic sites.

Fig. 3 shows a comparison of predicted amino acid sequences of fiber polypeptide of Ad3 and Ad7. The percent homology within the predicted amino acid sequences of Ad3 and Ad7 fiber is 60%, lower than that observed between the predicted amino acid sequences for the Ad2 and the Ad5 fiber (69% homology; Chroboczek and Jacrot, 1987). Although the nonhomologous regions are scattered over the entire protein-coding region, the amino terminal portion of the predicted polypeptide (that may encode the region that binds to penton base) is generally more conserved than other parts of the protein, probably due to the fact that this part of the fiber protein plays a role in joining of fiber protein to penton base (Devaux et al., 1987).

S1 nuclease mapping of the exon that encodes Ad7 fiber. The acceptor splice site for the fiber gene in Ad2 and Ad3 has been shown to be located directly in front of the ATG initiation codon; since the DNA

Figure 3: The structural organization of Ad7 fiber and homology to Ad3 fiber. The black regions represent homologous amino acid sequences between the Ad7 and Ad3 proteins (ignoring deletions and insertions). Written above the figure are three rectangles which show the sequences of conserved amino acids found in the fiber polypeptides of Ad2, Ad3 (Signäs et al., 1985), Ad4 (J.S. Hong and J.A. Engler, unpublished), Ad5 (Cladaras and Wold, 1985; Chroboczek and Jacrot, 1987), and Ad7. Positions in the consensus sequence where more than one similar amino acid can be present are shown within parentheses; positions where dissimilar amino acids may be present are shown with a dash (-). The arrows beneath the rectangles point to the relative locations of these conserved amino acid sequences in the fiber protein.


(T/S)FNPVYPY(D/E)-E(S/T)
Ad2, Ad5, Ad3, Ad7
(Y/F)TF(S/Y)YI(A/R)(E/D/Q)(D)
Ad2, Ad3, Ad5, Ad7, Ad4
PF (L/I)-P-F(V/I)SP(N/D)GF(T/Q)(Q/E)SP-GVL
(S/T)L----PL-T--G-L-(T/S)(L/V/I)D
Ad2, Ad5, Ad3, Ad7
TAIL
SHAFT
KNOB
0
100
200
300
Distances from amino terminus of fiber (amino acids)

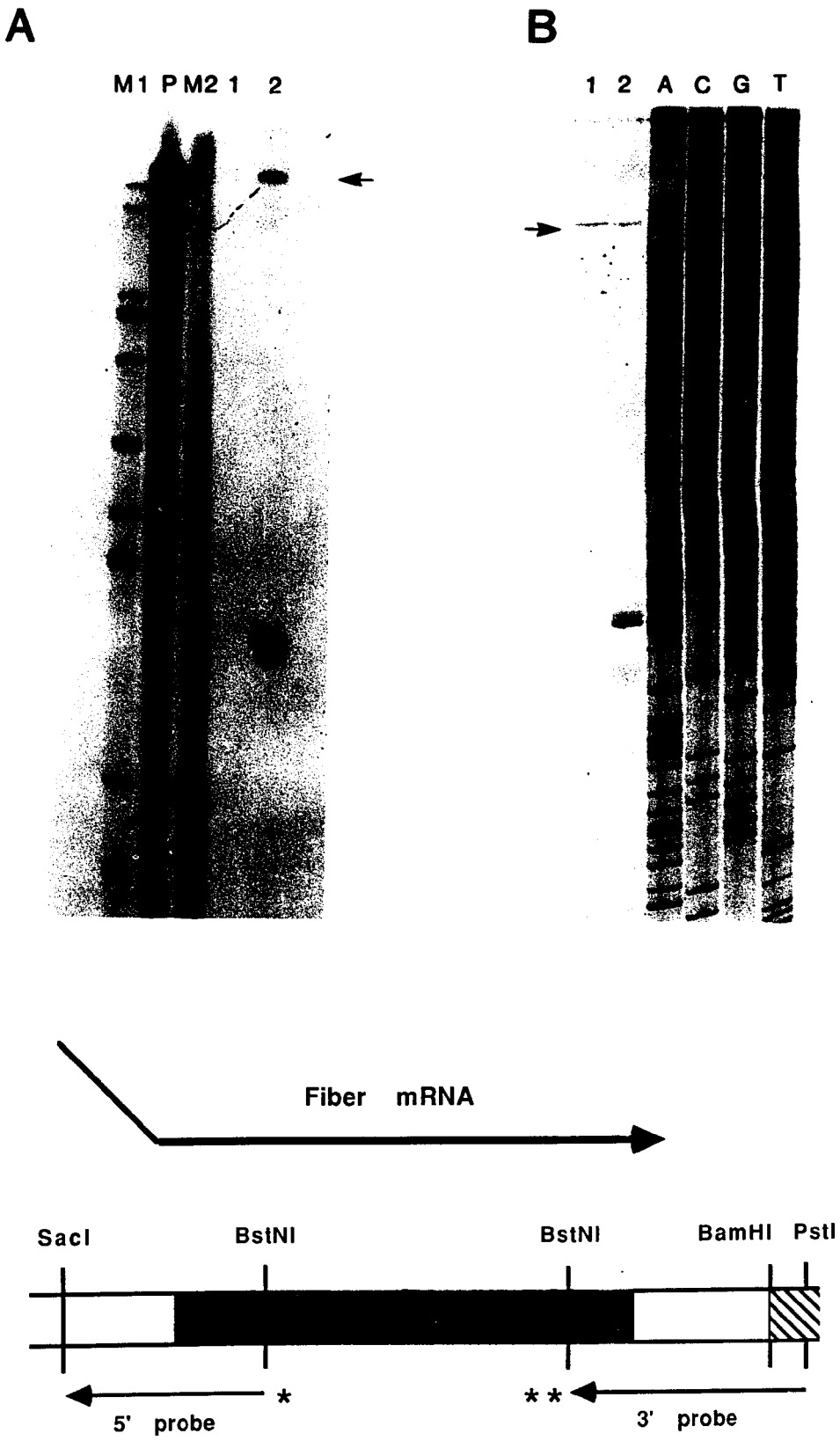
sequence in this region of Ad7 is identical to that of Ad3, S1 nuclease mapping was used to determine whether the same splice points were utilized during processing of the main exon encoding the Ad7 fiber mRNA. The results of an experiment to map the 5' and 3' boundaries of the exon that encodes fiber are shown in Fig. 4 (Panels A and B, respectively). To map the 5' splice site (where the tripartite leader is attached to the body of the Ad7 fiber mRNA), late RNAs from Ad7 infected A549 cells were hybridized to a SacI-BstNI fragment (nucleotides 1694 to 2265) labeled at the 5' end of BstNI restriction site, treated with S1 nuclease and separated through 5% denaturing polyacrylamide gel. The observed S1 resistant band is 90 nucleotides long, indicating that the 5' end of the fiber mRNA to the AG dinucleotide located immediately upstream of the AUG initiation codon is at nucleotide 2177; this same 5' acceptor site is also used in Ad3 fiber mRNA (Signäs et al., 1985).

To map the poly(A) addition site for the Ad7 fiber mRNA, late viral RNA was hybridized to 3' end labelled DNA fragment located between nucleotide 3062 (a BstNI cleavage site) and a PstI site in polylinker region of plasmid (Fig. 4). The DNA probe was made by filling in the BstNI site with [a-^{32}P]dATP. This mapping experiment shows that the poly(A) addition site of the Ad7 fiber mRNA is located at nucleotide 3388 or 3389 (Fig. 1) An AATAAA sequence is located between nucleotides 3159-3164, 23 or 24 bases upstream from this poly(A) addition site.

Amino acid sequence homology in an unidentified E4 ORF between Ad2 and Ad7. A short ORF for an unidentified protein is encoded in the complementary (leftwardly transcribed) strand of the sequence shown in Fig.1 between nucleotides 3499 and 3191.

This ORF terminates prior to the presumed polyadenylation signal for early region 4 of Ad7 (at nucleotide 3180 in the complementary strand).

Figure 4: S1 nuclease mapping of the 5' and 3' ends of the exon that encodes Ad7 fiber mRNA. The sizes of the S1 nuclease-resistant products of RNA:DNA hybrids (prepared as described in Materials and Methods) were measured by electrophoresis on a 5% urea/polyacrylamide sequencing gel. The position covered by each probe is shown below the panels. The filled box shows the approximate position of the exon, as determined by the S1 mapping experiments reported here. The single and double asterisks show the position of the ^{32}P-label incorporated into the probes used. Panel A: Mapping the 5' end of the fiber exon. Lane M1: DNA size markers are MspI-cut pSP65 (Promega) which gives fragments whose lengths are (from top) 512, 489, 258, 242, 228, 190, 147, 123, 110, 105, 67, 55. Lane P: undigested 5' probe (572 bp). Lane M2: an end labeled 94 bp DNA fragment. Lane 1: RNA from uninfected A549 cells hybridized to the 5' probe and treated with S1 nuclease. Lane 2: RNA from Ad7-infected A549 hybridized to probe and treated with S1 nuclease. Panel B: Mapping the 3' end and polyadenylation site of the fiber exon. The arrow indicates the length of full length probe (510 nucleotides). Lane 1: RNA from uninfected A549 cells hybridized to the 3' probe and treated with S1 nuclease. Lane 2: RNA from Ad7-infected A549 hybridized to probe and treated with S1 nuclease. Lanes 3-6 are a dideoxynucleotide DNA sequencing ladder using a specific oligonucleotide primer, whose 5' end is 3 nucleotides upstream of the BstNI site used for labelling; the position of the S1 resistant material in lane 2 corresponds to the sequence shown 3 nucleotides higher on the ladder.


A
M1 P M2 1 2
B
1 2 A C G T
Fiber mRNA
SacI
BstNI
BstNI
BamHI
PstI
*
* *
5' probe
3' probe

When this ORF is compared to a similar unidentified ORF encoded in the E4 region of Ad2, it shares 57% amino acid homology over the carboxy-terminal 72 amino acids of the reading frame (encoded between nucleotides 3409 and 3193 in Ad7 and between 33,132 and 32,914 of Ad2 (numbered according to Roberts et al., 1986). At nucleotide 3409, a glutamine is encoded; this codon could also serve as an acceptor splice site for splicing to upstream exons in early region 4, to connect partial ORFs from upstream exon onto these 72 amino acids. Tigges and Raskas (1984) have suggested that there may be as many as 16 different mRNAs with splice acceptors in this region; however, none of the ten mRNAs whose splice points are known in this region (either by S1 nuclease mapping (Tigges and Raskas, 1984) or by cDNA cloning (Freyer et al., 1984)) utilize this potential splice acceptor.

DISCUSSION

The Ad2 and Ad12 fiber are both glycoproteins (Ishbashi and Maizel, 1974; Brüggemann et al., 1985), although the Ad7 and Ad3 fiber proteins are probably also glycosylated. However, little is known about the precise composition or position of attachment of oligosaccharides to the fiber protein. Ishbashi and Maizel (1974) demonstrated that the fiber polypeptide of Ad2 contained two residues of GlcNAc linked to the polypeptide chain by an alkali-sensitive O-glycosidic bond. Although there is no evidence for N-linked glycosylation of fiber in any serotype, there are three possible sites for N-linked glycosylation (Asn-X-Ser/Thr) found in the sequence of Ad7 fiber polypeptide (encoded starting at nucleotides 2418, at 2751, and at 3027 in Fig. 1); none of these potential sites are found at the comparable positions in the predicted amino acid sequences of the Ad2 or Ad3 fiber polypeptides.

Green et al. (1983) have suggested that a portion of the amino acid sequence of Ad2 fiber polypeptide contains a 15-residue motif which is repeated 22 times. These repeating units contain two alternating b-bends and b-strands and could form an amphipathic sheet-like structure which could be stabilized by dimer formation to give the rod of the fiber. To elucidate the possible secondary structure of Ad7 fiber, the results obtained from five structure prediction programs (Nagano, 1973; Burgess et al., 1974; Chou and Fasman, 1974; Lim, 1974; and Garnier et al., 1978) were compared. The consensus of these five programs suggested that 6 repeats of a 15 amino acid motif of alternating b-bends and b-strands could be a feature of this region of Ad7. These predictions are similar to those for the corresponding region of Ad3 (Signäs et al., 1985) and further strengthen the idea that the number of these repeating motifs helps to determine the length of the rod segment and total length of the fiber polypeptide observed in the electron microscope.

There are three short regions of partial homology between Ad2, Ad3, Ad5 and Ad7. A consensus sequence for these conserved amino acids is shown in Fig. 3. Devaux et al. (1987) have shown that the N terminal 16 amino acids of Ad2 fiber are required for binding to penton base: a proteolytically-cleaved 60 KDa form of fiber missing these N-terminal amino acids is unable to bind to penton base. The N-terminal homologous segment shown in Fig.3 contains the site at which the 60 KDa form is cleaved ([T/S]-F-N-P-V-Y-P/cleavage site/Y-[D/E]) and may be important for fiber binding to penton base. The partially conserved sequence from amino acids 27 to 73 contains the segment of the polypeptide predicted to represent the transition from the penton-binding tail domain to the shaft structure; conservation of these sequences may be important to formation of the base of the shaft. The homologous amino acid segment near the

carboxy-terminus of fiber probably resides in the fiber knob end and might interact with the cellular receptor for adenovirus binding. Inspection of the preliminary nucleotide sequence of Ad4 fiber also shows the carboxy-terminus of an ORF that also contains this C-terminal sequence (J.S. Hong and J.A. Engler, unpublished observation). Other possible functions for these conserved sequences in processing or transport of fiber are the focus of further studies in this laboratory.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Institutes of Health (GM-38234). Oligonucleotide synthesis of short primers for DNA sequencing was supported by NCI grant CA-13148.

REFERENCES

Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA 80, 3963-3965.

Birnbaum, H. C., and Doly, J. C. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7, 1513-1523.

Brüggemann, U., Klenk, H.-D., and Doerfler, W. (1985). Increased infectivity of extracellular adenovirus type 12. J. Virol. 55, 117-125.

Burgess, A. W., Ponnuswamy, P. K., and Scheraga, H. A. (1974). Analysis of conformations of amino acid residues and prediction of backbone topography in proteins. Israel J. Chem. 12, 239-286.

Chou, P. Y., and Fasman, G. D. (1974). Prediction of protein conformation. Biochemistry 13, 222-245.

Chow, L. T., Broker, T. R., and Lewis, J. B. (1979). Complex splicing patterns of RNAs from the early regions of adenovirus. J. Mol. Biol. 134, 265-303.

Chow, L. T., Roberts, J. M., Lewis, J. B., and Broker, T. R. (1977). A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. Cell 11, 819-836.

Chroboczek, J., and Jacrot, B. (1987). The sequence of adenovirus fiber: Similarities and differences between serotypes 2 and 5. Virol. 161, 549-554.

Cladaras, C., Bhat, B., and Wold, W. S. M. (1985). Mapping the 5' ends, 3' ends, and splice sites of mRNAs from the early E3 transcription unit of adenovirus 5. Virol. 140, 44-54.

Cladaras, C., and Wold, W. S. M. (1985). DNA sequence of the early E3 transcription unit of Ad5. Virol. 140, 28-43.

Devaux, C., Caillet-Boudin, M.-L., Jacrot, B. and Boulanger, P. (1987). Crystallization, enzymatic cleavage, and the polarity of the adenovirus type 2 fiber. Virol. 161, 121-128.

Devaux, C., Zulauf, M., Boulanger, P., and Jacrot, B. (1982). Molecular weight of adenovirus serotype 2 capsomers: A new characterization. J. Mol. Biol. 156, 927-937.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12, 387-395.

Dorsett, P. H., and Ginsberg, H. S. (1975). Characterization of the type 5 adenovirus fiber protein. J. Virol. 15, 208-216.

Engler, J. A., Hoppe, M. S., and van Bree, M. P. (1983). The nucleotide sequence of the genes encoded in early region 2b of human adenovirus type 7. Gene 21, 145-159.

Engler, J. A., and Kilpatrick, B. A. (1981). Cloning and characterization of Class B adenoviruses Ad3 and Ad7 DNA. Gene 13, 125-132.

Fickett, J. W. (1982). Recognition of protein coding regions in DNA sequences. Nucl. Acids Res. 10, 5303-5318.

Flomenberg, P. R., Chen, M., and Horwitz, M. S. (1988). Characterization of the adenovirus type 35 early region 3. J. Virol. 62, in press.

Freyer, G. A., Katoh, Y., and Roberts, R. J. (1984). Characterization of the major mRNAs from adenovirus 2 early region 4 by cDNA cloning and sequencing. Nucl.Acids Res. 12, 3503-3519.

Garnier, J., Osguthorpe, D. J., and Robeson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120, 97-120.

Gooding, L. R., Elmore, L. W., Tollefson, A. E., Brady, H. A., and Wold, W. S. M. (1988). A 14,700 MW protein from the E3 region of adenovirus inhibits cytolysis by tumor necrosis factor. Cell 53, 341-346.

Green, N. M., Wrigley, N. G., Russell, W. C., Martin, S. R., and McLachlan, A. D. (1983). Evidence for a repeating cross-ß sheet structure in the adenovirus fibre. EMBO J. 2, 1357-1365.

Hérissé, J., Courtois, G., and Galibert, F. (1980). Nucleotide sequence of the EcoRI D fragment of adenovirus 2 genome. Nucl. Acids Res. 8, 2173-2192.

Hérissé, J., and Galibert, F. (1981). Nucleotide sequence of the EcoRI E fragment of adenovirus genome. Nucl. Acids Res. 9, 1229-1240.

Ishibashi, M., and Maizel, J. V., Jr. (1974). The polypeptides of adenovirus: VI. Early and late glycopolypeptides. Virol. 58, 345-361.

Kilpatrick, B. A., Gelinas, R. E., Broker, T. R., and Chow, L. T. (1979). Comparison of late mRNA splicing among class B and class C adenoviruses. J. Virol. 30, 899-912.

Kitchingman, G. R., and Westphal, H. (1980). The structure of adenovirus 2 early nuclear and cytoplasmic RNAs. J. Mol. Biol. 137, 23-48.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44, 283-292.

Lim, V. I. (1974). Algorithms for prediction of alpha-helical and beta-structural regions in globular proteins. J. Mol. Biol. 88, 873-894.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Nagano, K. (1973). Logical analysis of the mechanism of protein folding: I. Predictions of helices, loops and beta-structures from primary structure. J. Mol. Biol. 75, 401-420.

Norrby, E. (1969). The structural and functional diversity of adenovirus capsid components. J. Gen. Virol. 5, 221-236.

Pettersson, U. (1984). Structural and nonstructural adenovirus proteins. In "The Adenoviruses" (H.S. Ginsberg, Ed.), pp. 205-269. Plenum, New York.

Roberts, R. J., Akusjärvi, G., Aleström, P., Gelinas, R.E., Gingeras, T.R., Sciaky, D., and Pettersson, U. (1986). A consensus sequence for the adenovirus 2 genome. In "Adenovirus DNA: The viral genome and its expression" (W. Doerfler, Ed.), pp. 1-51. Martinus Nijhoff Publishing, Boston.

Sanger, F., Coulson, A. R., Barrell, B. ., Smith, A. J. H., and Roe, B. A. (1980). Cloning in single-strand bacteriophage as an aid to rapid DNA sequencing. J. Mol.Biol. 143, 161-178.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Signäs, C., Akusjärvi, G., and Pettersson, U. (1985). Adenovirus 3 fiber polypeptide gene: implications for the structure of the fiber protein. J. Virol. 53, 672-678.

Signäs, C., Akusjärvi, G., and Pettersson, U. (1986). Region E3 of human adenovirus: Differences between the oncogenic adenovirus-3 and the non-oncogenic adenovirus-2. Gene 50, 173-184.

Staden, R. (1977). Sequence handling by computer. Nucl. Acids Res. 4, 4037-4051.

Staden, R. (1978). Further procedures for sequence analysis by computer. Nucl. Acids Res. 5, 1010-1015.

Tibbetts, C. (1977). Physical organization of subgroup B human adenovirus genomes. J. Virol. 24, 564-579.

Tigges, M. ., and Raskas, H. J. (1984). Splice junctions in adenovirus 2 early region 4 mRNAs: multiple splice sites produce 18 to 24 RNAs. J. Virol. 50, 106-117.

Tollefson, A. E., and Wold, W. S. M. (1988). Identification and gene mapping of a 14,700-molecular-weight protein encoded by region E3 of group C adenoviruses. J. Virol. 62, 33-39.

van Ormondt, H., Maat, J., and Dijkema, R. (1980). Comparison of nucleotide sequences of the early E1a regions for subgroups A, B and C of human adenoviruses. Gene 12, 63-76.

Vieira, J., and Messing, J. (1987). Production of single-stranded plasmid DNA. In "Methods in Enzymology: Recombinant DNA, Part D " (R. Wu and L. Grossman, Eds.), pp. 3-11. Academic Press, New York.

Wang, E. W., Scott, M. O., and Ricciardi, R. P. (1988). An adenovirus mRNA which encodes a 14,700-M_r protein that maps to the last open reading frame of region E3 is expressed during infection. J. Virol. 62, 1456-1459.

Wold, W. S. M., Cladaras, C., Magie, S. C., and Yacoub, N. (1984). Mapping a new gene that encodes an 11,600-molecular-weight protein in the E3 transcription unit of adenovirus 2. J. Virol. 52, 307-313.

NUCLEAR TRANSPORT OF AN ADENOVIRUS FIBER MUTANT THAT LACKS O-LINKED N-ACETYLGLUCOSAMINE

by

JEONG SHIN HONG, KAREN GHEESLING MULLIS, AND JEFFREY A. ENGLER

AUTHORSHIP

This was a collaborative work. The studies of glycosylation are the work of Karen Mullis.

ABBREVIATIONS

Ad2, Ad5, Ad7, adenovirus serotypes 2, 5 and 7 (respectively); DMEM, Dulbecco's modified Eagle's medium; ELISA, enzyme-linked immunosorbent assay; Gal, galactose; O-GlcNAc, O-linked N-acetylglucosamine; SDS-PAGE, sodium dodecylsulfate-polyacrylamide gel electrophoresis.

ABSTRACT

O-linked N-acetylglucosamine (O-GlcNAc) is a posttranslational modification attached to many cytoplasmic and nuclear proteins. Roles that have been suggested for this modification include regulatory functions, participation in the assembly of multimeric complexes, and participation in targeting of proteins to the nucleus. Adenovirus types 2 and 5 fibers are nuclear proteins containing O-GlcNAc; they are thought to form trimeric structures prior to assembly into virions in the nucleus. A temperature-sensitive mutant of adenovirus type 5 fiber (H5ts142) does not contain O-GlcNAc at 39.5°C and was used to test whether O-GlcNAc was a necessary signal for proper targeting of nuclear proteins or participated in the formation of fiber multimers. Using monoclonal antibodies specific for fiber monomers or fiber trimers, adenovirus fiber from wild type or from H5ts142 was localized to the nucleus by indirect immunofluorescence at both 32°C and 39.5°C. Mutant fibers lacking O-GlcNAc can be properly transported to the nucleus, but appeared to have a defect in formation of trimers and their subsequent assembly into virions.

INTRODUCTION

In contrast to the well-studied glycosylation events that occur within the lumen of the endoplasmic reticulum and Golgi (14, 15), a novel posttranslational modification, O-linked N-acetylglucosamine, has been found to be enriched on many proteins localized to the cytoplasmic and nucleoplasmic compartments of the cell. Although the biological role of O-GlcNAc on cytoplasmic and nuclear proteins is not yet known, a number of possible roles for this modification have been suggested: (1) regulatory functions (16), for example by reversibly blocking sites of action of other types of modification, such as phosphorylation; (2) participation in the formation or stabilization of multiprotein or multimeric complexes (16, 17); (3) participation in protein transport through nuclear pores and into the nucleus (4, 9, 28, 32, 36); (4) transcriptional activation, as in the case of factor SP1 (20); and (5) the targeting of proteins to specific intracellular compartments such as the nucleus (32), similar to the role played by phosphorylated mannose residues on the targeting of specific N-linked glycoproteins to the lysosomes (reviewed in 23).

The suggestion that O-linked GlcNAc is part of the signal for nuclear targeting has been controversial. Although there are many examples of nuclear proteins that contain this modification, there are also some O-GlcNAc-modified proteins that are found exclusively in the cytoplasm (14, 15). Short peptide sequences required for correct targeting of a number of nuclear proteins have been identified (3, 8, 10, 21, 25, 33, 37); these sequences generally contain stretches of basic residues. The sequence motif proposed for attachment of O-GlcNAc to proteins (11) is not contained within known nuclear targeting signals. Fusion of nuclear targeting signals to normally cytoplasmic proteins results in proper localization of the hybrid polypeptide to the nucleus (3, 8, 10, 25, 30),

although it cannot yet be excluded that these test proteins are not themselves substrates for attachment of O-GlcNAc or other saccharide modifications when expressed in or injected into animal cells. It is also possible that some (but not all) nuclear proteins require the action of more than one signal (specific amino acid sequence; carbohydrate modification, etc.) for proper transport or that multiple signals act together to increase the efficiency of the transport process (for example, see 37).

Adenovirus is a DNA virus whose capsids are assembled in the cell nucleus. The adenovirus capsid consists of three major proteins: hexon, penton base, and fiber (120K, 70K, and 62K in Ad2, respectively). These components are synthesized in the cytoplasm and are then transported to the nucleus for assembly into virus particles (18). Fibers from Ad2 and Ad5 (both in subgroup C) have been shown to contain O-linked GlcNAc by a number of criteria: (1) fiber can be labelled with [^{14}C]-glucosamine in vivo (1, 19; Mullis, K.G. and Engler, J.A., unpublished observation); (2) fiber contains terminal GlcNAc residues that can be galactosylated in vitro (Mullis et al., submitted). On the other hand, fibers from Ad3, 4, 7, and 9 do not contain this carbohydrate modification (1; Mullis et al., submitted). A temperature-sensitive mutant of Ad5 fiber (Ad5 H5ts142) has been reported that lacks this GlcNAc modification at the nonpermissive temperature, based on labeling studies with ^{3}H-glucosamine (2). At the non-permissive temperature, this mutant produces empty capsids rather than infectious virus particles; in addition, fibers from H5ts142 are unable to form trimers required for assembly, as judged by sedimentation in sucrose gradients (2).

In this study, adenovirus fiber from the H5ts142 mutant virus has been used to test the possible role of GlcNAc modification on the transport of this protein to the nucleus and on the formation of trimers; the idea that

two or more separate signals (a peptide sequence within the protein and a carbohydrate modification (O-GlcNAc)) might be required for correct targeting of this protein could be tested directly, since both types of signals could be found on fiber from H5ts142 at the permissive but not at the nonpermissive temperature. The results of this study demonstrate that both glycosylated and nonglycosylated Ad5 fiber can be transported efficiently to the nucleus, suggesting that the presence of O-linked GlcNAc is not obligatory for correct targeting of fiber. However, as shown with a monoclonal antibody specific for trimers, fibers lacking O-GlcNAc appear to be defective in formation of trimers at the nonpermissive temperature, confirming the earlier results of Chee-Sheung and Ginsberg (2); this defect in trimer assembly is at least partially reversible upon incubation at the permissive temperature.

MATERIALS AND METHODS

Cells and Viruses. Ad2 and Ad5 viruses were grown on plates of HeLa cells maintained in DMEM (Gibco), 5% fetal bovine serum (Hyclone Laboratories), gentamycin (50μg/ml, Sigma Chemical, St. Louis, MO), and nystatin (250 u/ml). H5ts142 was a gift from Dr. H. Ginsberg and was grown as described above, but at 32°C or 39°C. Ad7 virus was grown on plates of A549 cells maintained in F-12 medium (Cellgro) with 10% fetal bovine serum and gentamycin.

Polyclonal and Monoclonal Antibodies. Ad2 virus inactivated by UV irradiation was used as antigen in Balb/c mice. Tissue culture supernatants from cloned hybridoma cells were first screened by ELISA assay, using CsCl-purified adenovirions; ELISA-positive supernatants were further screened by immunoblot of adenovirus proteins, by immunoprecipitation, and by indirect immunofluorescence. High titer ascites fluid from Balb/c mice injected with antibody-secreting hybridoma

cells was used for all experiments. Polyclonal rabbit anti-Ad5 antibody was purchased from ATCC (VR-1081 AS/Rab). Monoclonal antibody specific for GlcNAc-containing proteins (RL2; 34) was a gift of Dr. L. Gerace (Scripps Clinic, La Jolla, CA).

Immunoprecipitation of virus samples was carried out using polyclonal antibody directed against whole virus (from ATCC) or using monoclonal antibodies 4D2-5, 2A6-36, and 9C6-41, according to procedures described in Harlow and Lane (13).

Galactosylation. Autogalactosylation reactions were carried out as described previously (35), using 60 mM UDP-Gal (Boehringer Mannheim, Indianapolis, IN) and 9 munits 4b-galactosyltransferase (4.7 units/mg; Sigma Chemical) for 30 minutes at 37°C. Viral proteins were galactosylated in a reaction mixture containing 60 mM $MnCl_2$ and 200 mM sodium cacodylate, pH 6.8. The labeling reaction was initiated by addition of 0.4 mCi UDP-(U-^{14}C)-Gal (257 mCi/mmol; Amersham), followed by 9 munits of autogalactosylated bovine galactosyltransferase (containing 60 mM cold UDP-Gal carried over from the prior pregalactosylation of the enzyme); the reaction was then carried out for 30 minutes at 37°C. The labeled sample was immunoprecipitated using monoclonal antibodies directed against fiber (as described below). A fraction of this precipitate was analyzed by SDS-PAGE on a 10% gel. The gel was stained (in 2% Coomassie blue, 20% methanol, 10% acetic acid), destained (in 20% methanol, 10% acetic acid), impregnated with ENHANCE (New England Nuclear, Boston, MA), and visualized by autoradiography.

Temperature-shift experiment. A549 cells were infected with H5ts142 at either 32°C (permissive temperature) or 39.5°C (nonpermissive temperature). At 42 hrs postinfection, cycloheximide (30 mg/ml; Sigma Chemical) was added and cells were incubated at the original temperature

for 1-2 hours. The infected cells were then placed at the final temperature (as detailed in legend of each figure) and the infection was allowed to proceed for additional 24 hours.

Indirect immunofluorescence microscopy. A549 cells were plated on a glass coverslip in 35mm plate. At 60 hr postinfection, cells were fixed with 3% paraformaldehyde in PBS (phosphate buffered saline) either at 32°C (for cells grown at the permissive temperature) or 39.5°C (for cells grown at the nonpermissive temperature). TBS-TX (Tris buffered saline-1% Triton X-100) was used to permeabilize the cell membrane, prior to incubation with anti-fiber antibody (4D2-5) or anti-trimer antibody (2A6-36) at room temperature for 1 hour. Goat anti-mouse Ig conjugated to fluorescein isothiocyanate (Jackson Immunoresearch, West Grove PA) was used as secondary antibody. The cells were photographed on a Nikon Optiphot microscope equipped for fluorescent illumination at a magnification of 20x or 40x using Kodak Gold 400 film.

Immunoblot assay. Protein samples were prepared by sonication and Freon extraction of infected cells, as described by Maizel et al. (26). Fifteen μg of each protein sample was loaded in sample buffer (24) and then separated by SDS-PAGE on a 10% gel; the proteins in the gel were then blotted onto nitrocellulose and incubated in Blotto (5% nonfat dry milk in PBS) to block nonspecific antibody binding. Either anti-fiber (4D2-5) or anti-trimer (2A6-36) antibody was used as primary antibody (1:1000 dilution in Blotto). Biotin-conjugated goat anti-mouse Ig antibody (Fisher Biotech) and alkaline phosphatase conjugated to streptavidin (Fisher Biotech or Jackson ImmunoResearch) were used to probe the specific binding of monoclonal antibodies to fiber; the color reaction for alkaline phosphatase was developed with NBT and BCIP (Bethesda Research Labs).

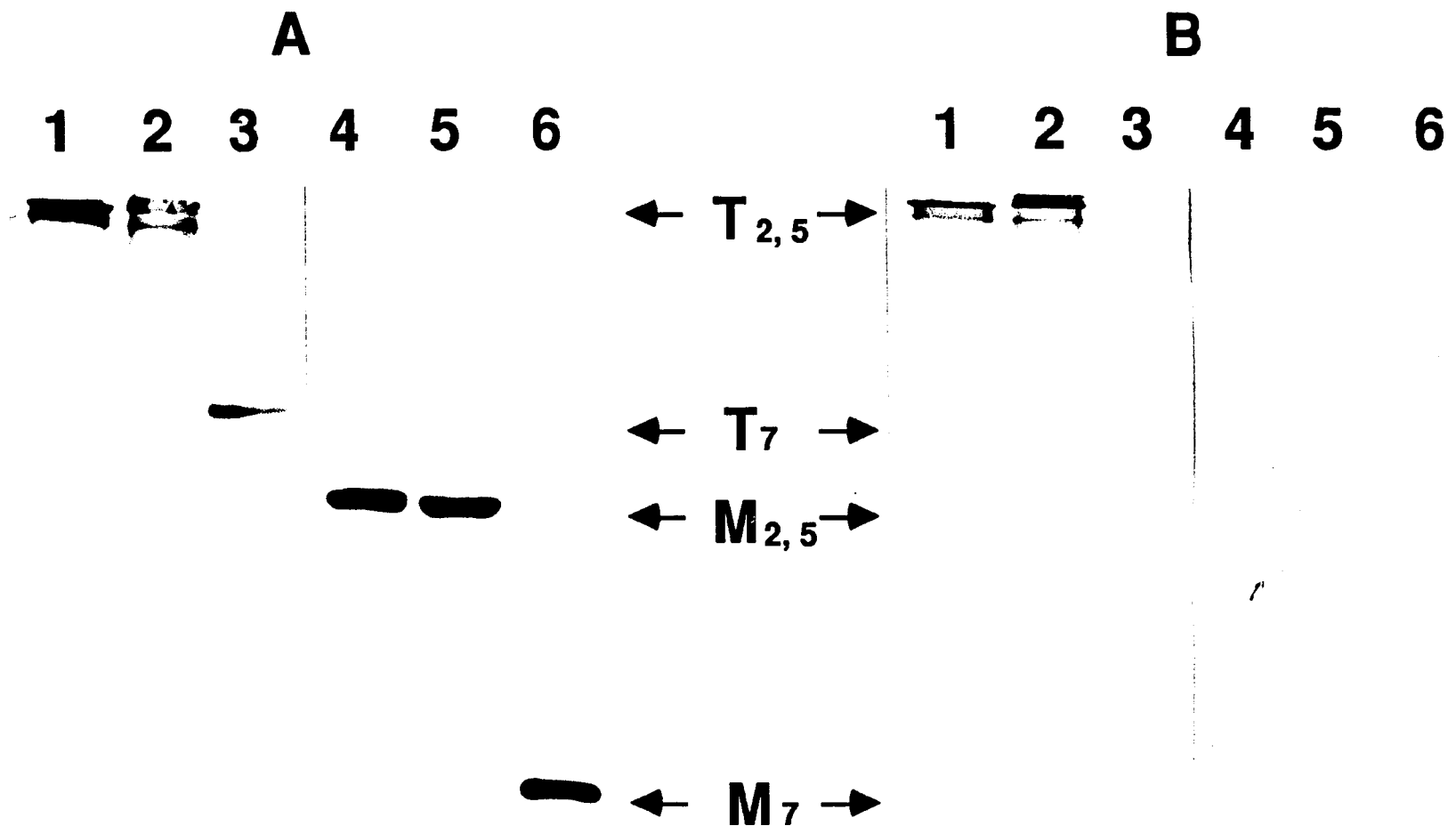
RESULTS

Specificity of monoclonal antibodies. Monoclonal antibodies that recognize adenovirus capsid proteins were prepared by standard techniques, using UV-irradiated Ad2 virus as an antigen. After initial screening of the hybridoma cell supernatants by ELISA assay, antibody specificity was determined by immunoblot, by immune precipitation, and by indirect immunofluorescence of infected cells. The results of the immunoblot analysis are shown in Figure 1; these immunoblots were done with both boiled and unboiled lysates from adenovirus-infected cells. Antibody 4D2-5 recognizes the monomer (boiled samples) and trimer (unboiled) forms of fiber from serotypes Ad2, Ad5, and Ad7 (Figure 1, panel A). However, antibody 2A6-36 recognizes only trimer form of fiber from Ad2 and Ad5 (Figure 1, panel B); in unboiled samples, the fiber remained as an intact trimer and could be detected with this antibody, even after SDS-PAGE. Antibody 9C6-41 recognizes penton base from Ad2 and Ad5 and appears to be group-specific (data not shown). Both antibodies recognized Ad2 and Ad5 fiber in the nuclei of the infected cells by indirect immunofluorescence (as shown for Ad5 in Figure 3, panels A and B). Antibody 4D2-5 also recognized Ad7 fiber by immunofluorescence in the nucleus (data not shown).

Presence of GlcNAc on the wild type and mutant fiber polypeptides.

An *in vitro* galactosylation reaction can identify terminal GlcNAc residues using radioactive UDP-galactose and bovine milk 4b-galactosyl-transferase (35). It has been shown that Ad2 and Ad5 (but not Ad7) fibers can be labelled using galactosyltransferase and UDP-(U-^{14}C)-galactose (Mullis et al., submitted). Ad5 H5*ts*142 is a temperature-sensitive mutant of Ad5 whose fiber polypeptides cannot be labelled *in vivo* with ^{3}H-glucosamine at the nonpermissive temperature (2); at the nonpermissive

Figure 1: Characterization of monoclonal antibodies against fiber. Antibodies 4D2-5 (Panel A) and 2A6-36 (Panel B) were used as primary antibodies on identical immunoblots. Lanes 1, 2, and 3 are unboiled lysates from Ad-infected cells; lanes 4 through 6 are the same as lanes 1 through 3, except that the samples were boiled prior to loading on the 10% SDS-PAGE gel. $T_{2,5}$ and $M_{2,5}$ mark the position expected for trimers and monomers of Ad2 or Ad5 (respectively); T_7 and M_7 show the position expected for trimers and monomers of Ad7 (respectively). Lanes 1 and 4: Ad2-infected cells; lanes 2 and 5: Ad5-infected cells; lanes 3 and 6: Ad7-infected cells. Uninfected cell lysates gave no apparent bands on these immunoblots with these monoclonal antibodies (data not shown).


A
B
1 2 3 4 5 6
1 2 3 4 5 6
T 2, 5
T 7
M 2, 5
M 7

temperature, this mutant cannot make fiber trimers (as determined by sedimentation in a sucrose gradient) and forms empty virus particles (2). The in vitro galactosyltransferase reaction was used to investigate the differences in glycosylation of fiber on wild-type Ad5 and on H5ts142 grown at both the permissive (32°C) and nonpermissive temperatures (39.5°C; Figure 2A). Both H5ts142 and Ad5 fibers grown at permissive temperature could be galactosylated with UDP-[^{14}C]-galactose in vitro, although the amount of ^{14}C-galactose added to the mutant fiber was lower than the wild type level. The lower amount of galactose label observed on H5ts142 fiber could be due to:

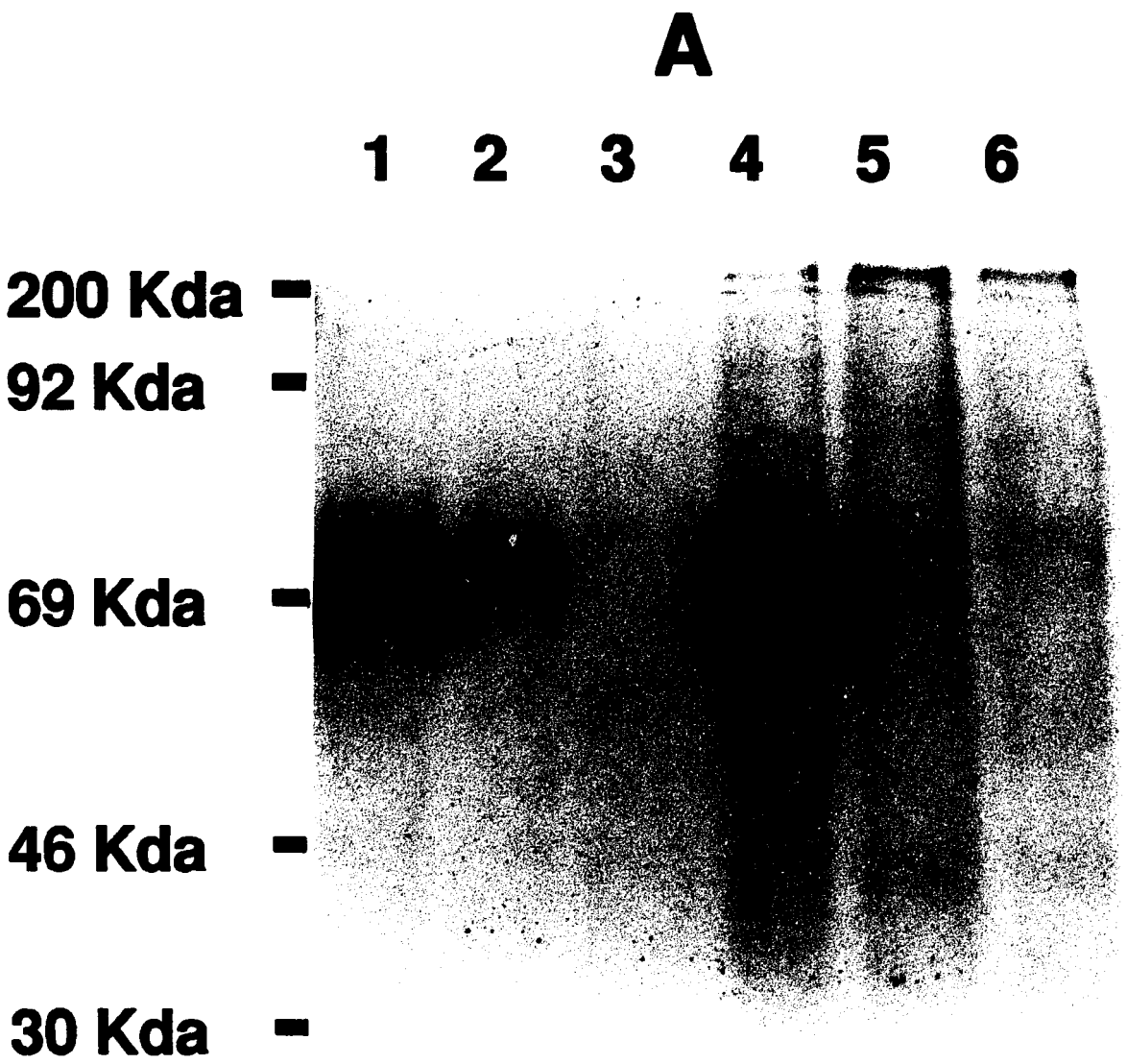
1) lower amounts of fiber made by the mutant at 32°C, although identical immunoblots developed with the 4D2-5 monoclonal antibody show nearly identical amounts of wild-type and mutant fiber at 32°C (data not shown),

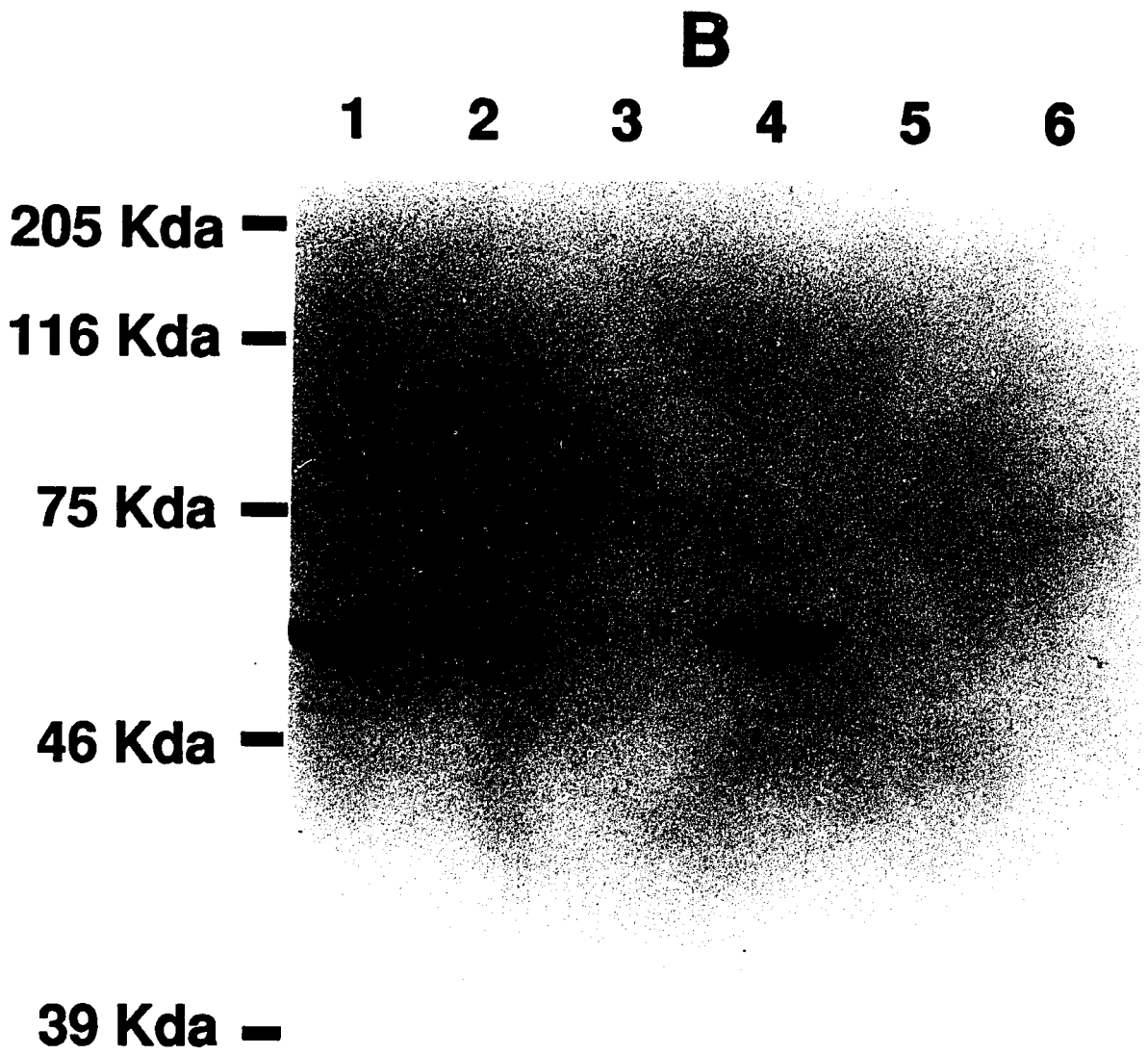
2) decreased accessibility of the GlcNAc acceptor to galactosyltransferase, or 3) the presence of fewer GlcNAc acceptors on the H5ts142 fiber.

Wild type Ad5 fiber grown at 39.5°C could also be galactosylated with radioactive galactose in this reaction; on the other hand, H5ts142 fiber grown at this temperature could not be measureably galactosylated.

Immunoblots developed with an antibody specific for GlcNAc-containing proteins (RL2; 34) confirmed the presence of GlcNAc on wild type fiber at 32°C and at 39.5°C but only on H5ts142 fibers at 32°C (Figure 2; Panel B). These results confirm the previous conclusion of Chee-Sheung and Ginsberg (2), based on the lack of detectable labeling of H5ts142 fiber by ^{3}H-glucosamine in vivo. As was discussed previously, there appears to be a decreased reactivity of the RL2 antibody with H5ts142 fiber compared to wild type at 32°C, although identical blots developed with the 4D2-5 antibody show similar amounts of mutant and wild-type fiber protein (data not shown).

Figure 2: Characterization of galactosylation of Ad5 and of H5ts142 fiber. Panel A: Characterization by labelling of GlcNAc acceptors with 4b-bovine galactosyltransferase. Samples were galactosylated with ^{14}C-UDP-galactose as described in Materials and Methods, prior to immune precipitation with monoclonal antibody 4D2-5. Size markers are shown on the left of the figure. Lanes 1 through 3 are lysates from infected HeLa cells at 32°C; lanes 4 through 6 are lysates from infected HeLa cells at 39.5°C. Lanes 1 and 4 are samples from wild type Ad5 infected cells, lanes 2 and 5 are samples from H5ts142-infected cells; lanes 3 and 6 are from uninfected HeLa cells.
Panel B: Immunoblot developed with monoclonal antibody RL2. Size markers are shown at the left of the figure. The order of lanes is identical to that in panel A.


A
1 2 3 4 5 6
200 Kda
92 Kda
69 Kda
46 Kda
30 Kda


B
1
2
3
4
5
6
205 Kda
116 Kda
75 Kda
46 Kda
39 Kda

Transport of H5ts142 fiber to the nucleus. Indirect immunofluorescence was performed on cells infected with wild-type Ad5 or with H5ts142 at either 32°C or 39.5°C (Figure 3). Wild-type fiber could be detected in the nucleus by both the anti-fiber (4D2-5) and the anti-trimer (2A6-36) monoclonal antibodies at each temperature. Fiber from H5ts142 was detected in the nucleus at both 32°C and 39.5°C with anti-fiber antibody but only at 32°C with the anti-trimer antibody. This observation suggests that the presence of O-linked GlcNAc is not required to insure correct transport of Ad5 fiber to the cell nucleus. In all cases, fiber-specific immunofluorescence in the cytoplasm was very low. Indirect immunofluorescence analysis of Ad7-infected cells with anti-fiber antibody showed that Ad7 fiber (which also lacks attached O-linked GlcNAc) was also localized in the nucleus (data not shown).

To investigate whether the mutant fiber synthesized at the nonpermissive temperature could form fiber trimers if shifted down to the permissive temperature, immunoblot and immunofluorescence experiments were done. At 42 hrs postinfection, cycloheximide (30 μg/ml) was added to the media and the cells were incubated at the original temperature for 1 hour before the culture was shifted to the final temperature; the infection was allowed to continue for an additional 24 hours after the temperature shift. Cell lysates were made for immunoblot as described in Materials and Methods. Samples were either boiled (Figure 4, Panel B) or not boiled (Figure 4; Panel A) before SDS-PAGE on a 10% gel. Fibers from both wild type Ad5 and H5ts142 fibers could be detected as trimers (with both 4D2-5 and 2A6-36 Ab) following this temperature shift from 32°C to 39.5°C. None of the fiber appears as monomers, despite the shift to 39.5°C; all of the trimer could be converted to monomers by boiling prior to SDS-PAGE. When tested by immunoblot for

Figure 3: Nuclear localization of Ad5 wild type and of H5ts142 fiber at 32° and 39.5°C. Indirect immunofluorescence of virus-infected A549 cells was determined at 70 hours (32°C) or 41 hours (39.5°C) post infection, using monoclonal antibodies 4D2-5 (Panel A) and 2A6-36 (Panel B).

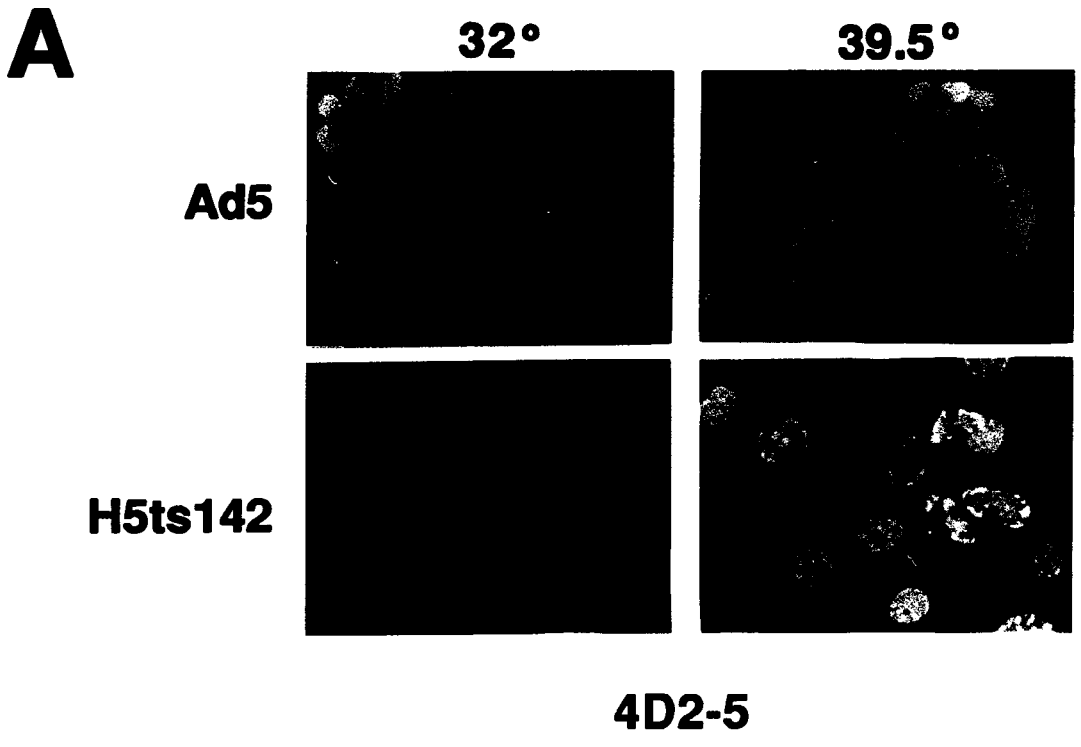
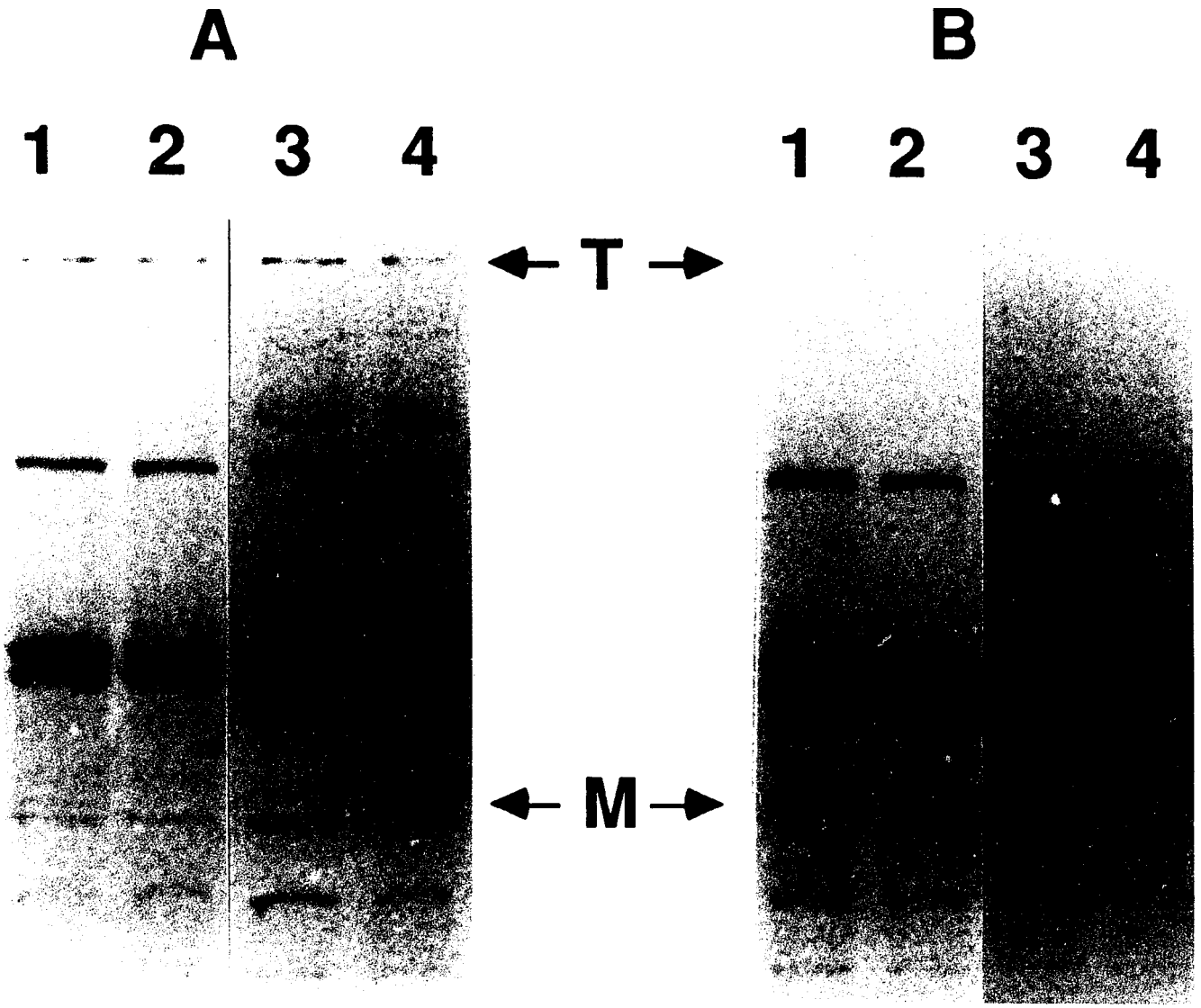
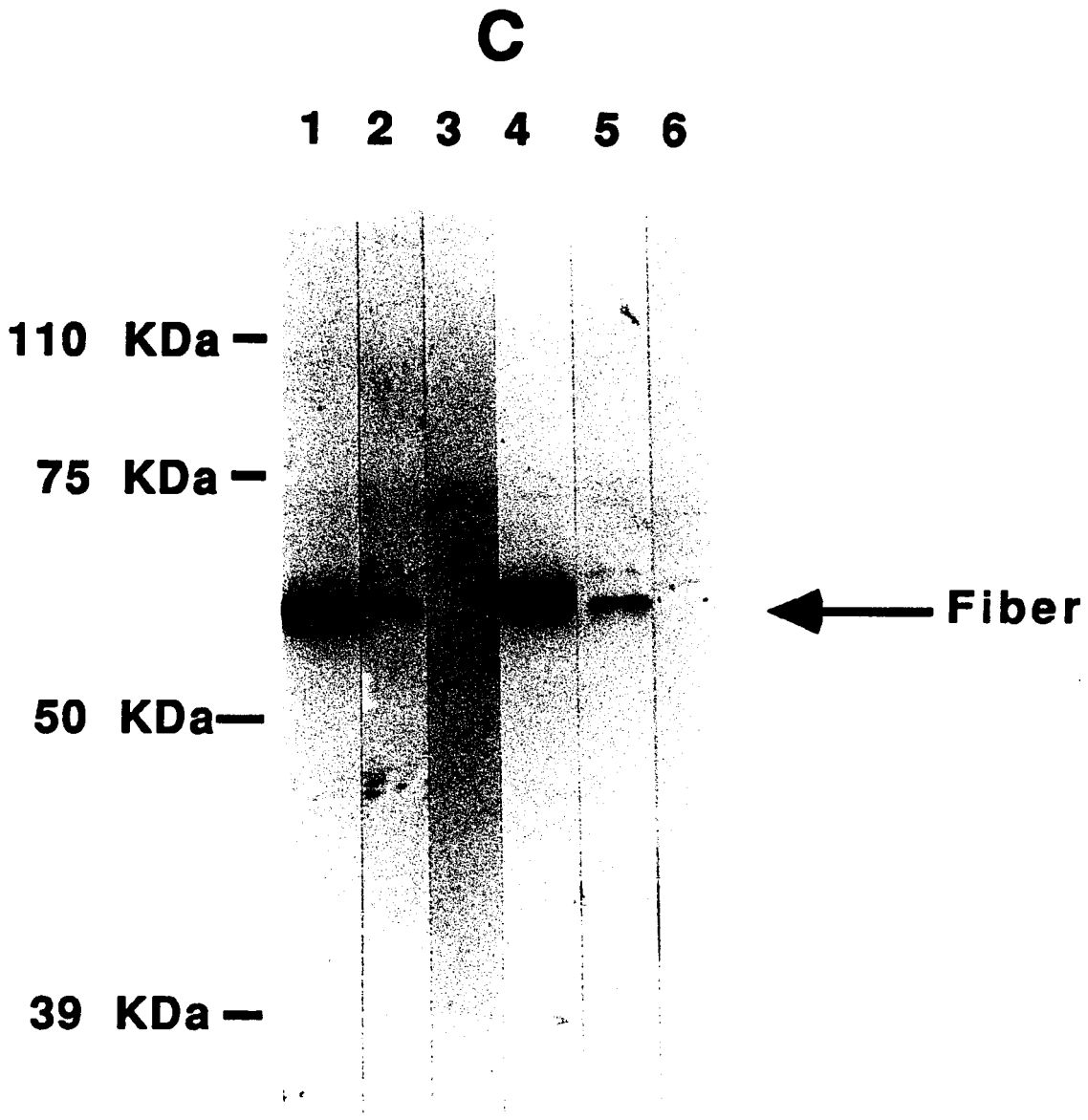

A
32°
39.5°
Ad5
H5ts142
4D2-5
B
32°
39.5°
Ad5
H5ts142
2A6-36

Figure 4: Immunoblot on H5ts142 infected cell lysates after the addition of cycloheximide, followed by temperature shift. Samples were either unboiled (panel A) or boiled (panels B and C) prior to loading on a 10% SDS-PAGE. T and M or fiber show the position expected for trimers or monomers, respectively. In panels A and B, lanes 1 and 2 were developed with antibody 4D2-5 (specific for monomers and trimers) and lanes 3 and 4 were developed with 2A6-36 (specific only for trimers). Lanes 1 and 3 were held at 32°C after cycloheximide addition, while lanes 2 and 4 were shifted from 32°C to 39.5°C 1 hour after cycloheximide addition. In Panel C, all lanes were developed with antibody RL2 (specific for O-linked GlcNAc); lanes 1 through 3 were held at 32°C after cycloheximide addition, while lanes 4 through 6 were shifted from 32° to 39.5°C 1 hour after cycloheximide addition. Lanes 1 and 4 are samples from wild type Ad5 infected cells; lanes 2 and 5 are from Ad5 H5ts142 infected cells; lanes 3 and 6 are from uninfected HeLa cells.


A
B
1 2 3 4
1 2 3 4
T
M


C
1 2 3 4 5 6
110 KDa
75 KDa
50 KDa
39 KDa
Fiber

the presence of GlcNAc, both wild type and H5ts142 fibers could be stained with RL2, showing that they contained O-linked GlcNAc (Figure 4, Panel C). Indirect immunofluorescence confirmed the results of the immunoblot analysis (Figure 5): trimers synthesized at 32°C persist even after shift to 39.5°C. Taken together, these results suggest that the fiber trimer must be stable once it is formed.

Figure 5 also shows the results of a temperature shift-down experiment: although fibers formed at 39.5°C were primarily monomers (detectable by monoclonal antibody 4D2-5, but not by 2A6-36), trimers could be detected upon shift-down to 32°C, as shown by the immunofluorescent staining of H5ts142-infected cells with each of the monoclonal antibodies. Immunoblot analysis of these infected cells also shows that trimers can be formed upon shift-down to the permissive temperature (Figure 6, panels A and B). This process must be somewhat inefficient, since only a small (but detectable) amount of trimers are formed at the permissive temperature within the 16 hours between addition of cycloheximide and the preparation of samples for analysis. Interestingly, the fibers show no addition of GlcNAc, as shown by the lack of staining of H5ts142 fiber on an immunoblot developed with RL2 (Figure 6, panel C).

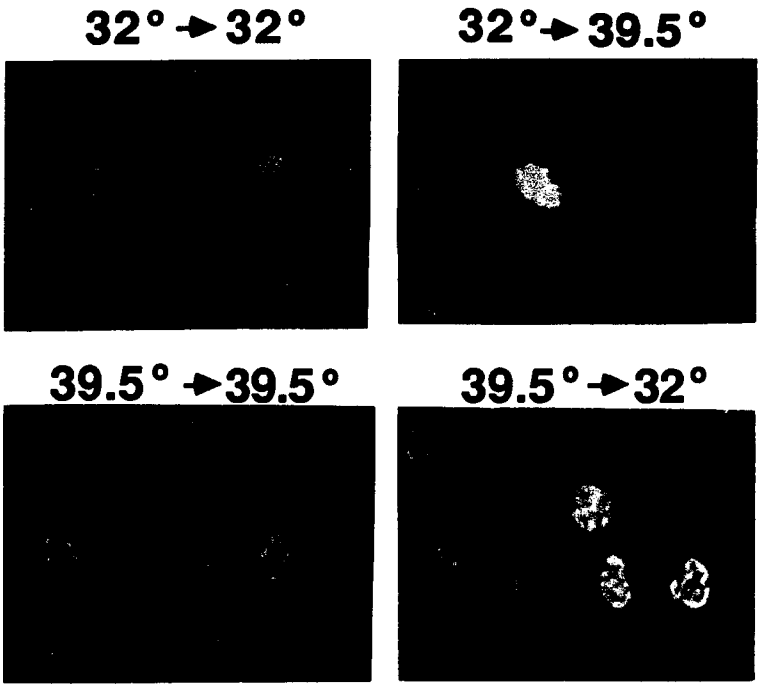
DISCUSSION

The data presented here demonstrate that O-linked GlcNAc is not required for the correct transport of adenovirus fiber within the cell, since the glycosylation defect of Ad5 H5ts142 fiber does not affect its eventual entry in the nucleus. Other arguments can also be made to support this conclusion:

1. Since fibers from other adenovirus serotypes (such as Ad7) lack O-linked GlcNAc (1), it seems unlikely that this modification is an essential

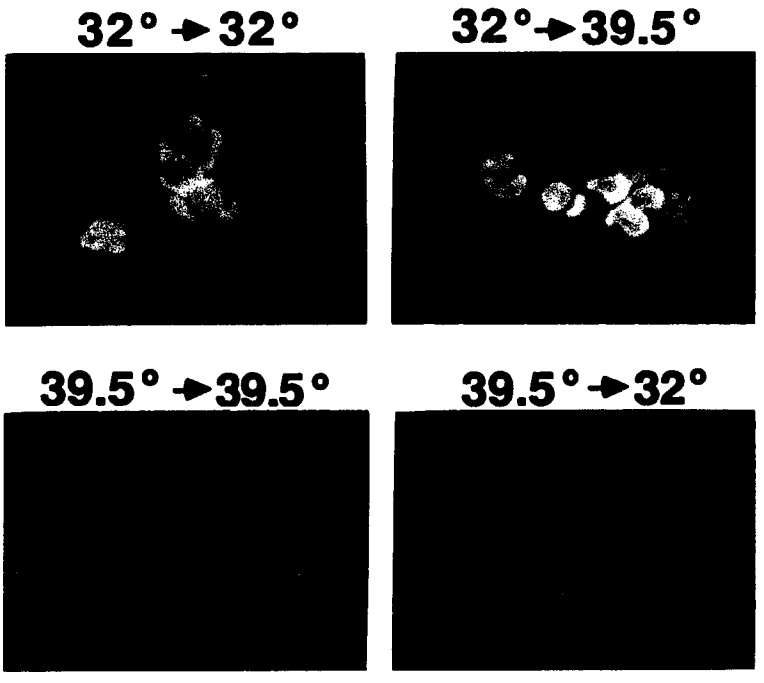
Figure 5: Nuclear localization of H5ts142 fiber after addition of cycloheximide and temperature shift. Indirect immunofluorescence of virus-infected A549 cells was determined at 48 hours post infection, using monoclonal antibodies 4D2-5 (anti-fiber; Panel A) and 2A6-36 (anti-trimer; Panel B) as described in Materials and Methods.

A



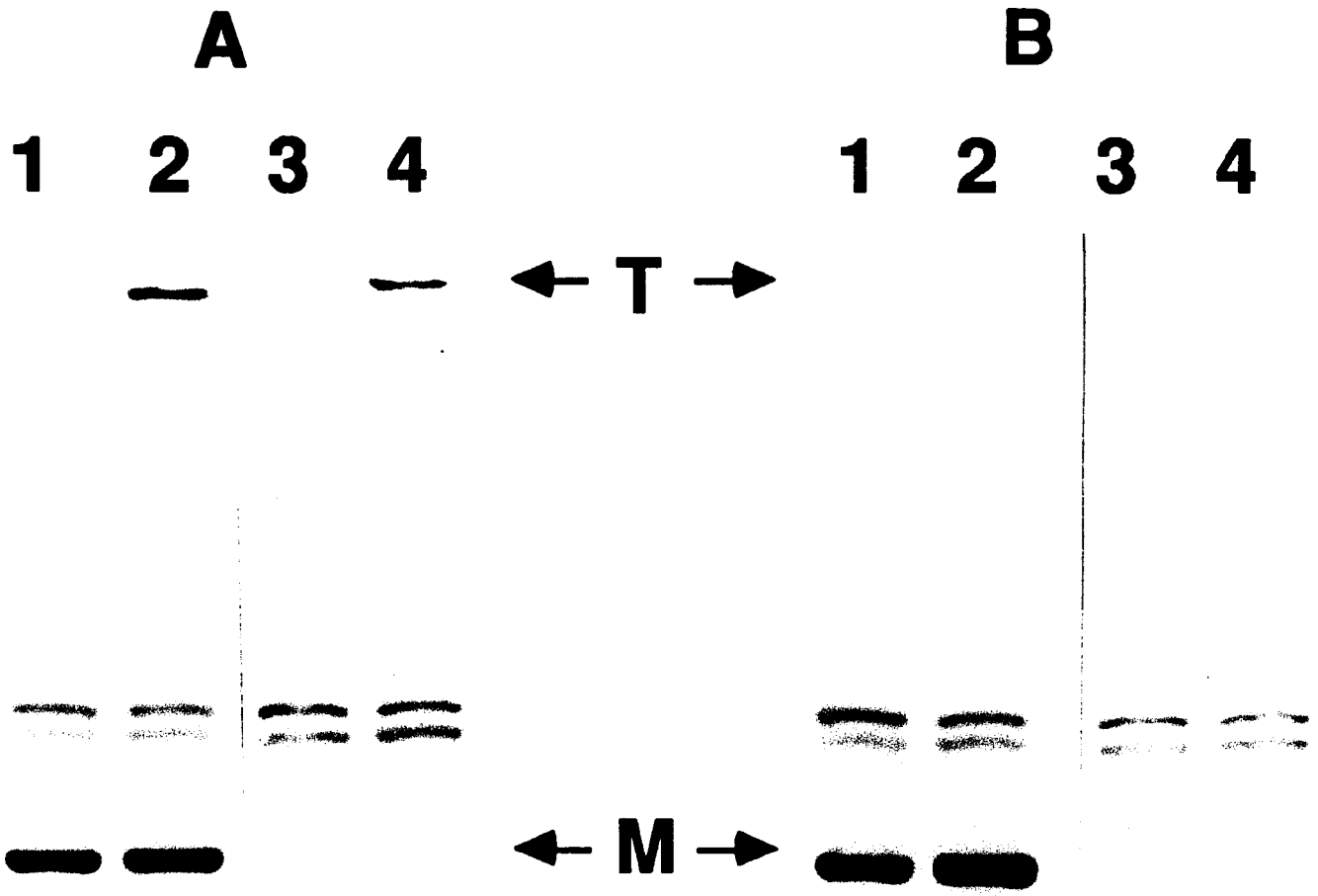
4D2-5

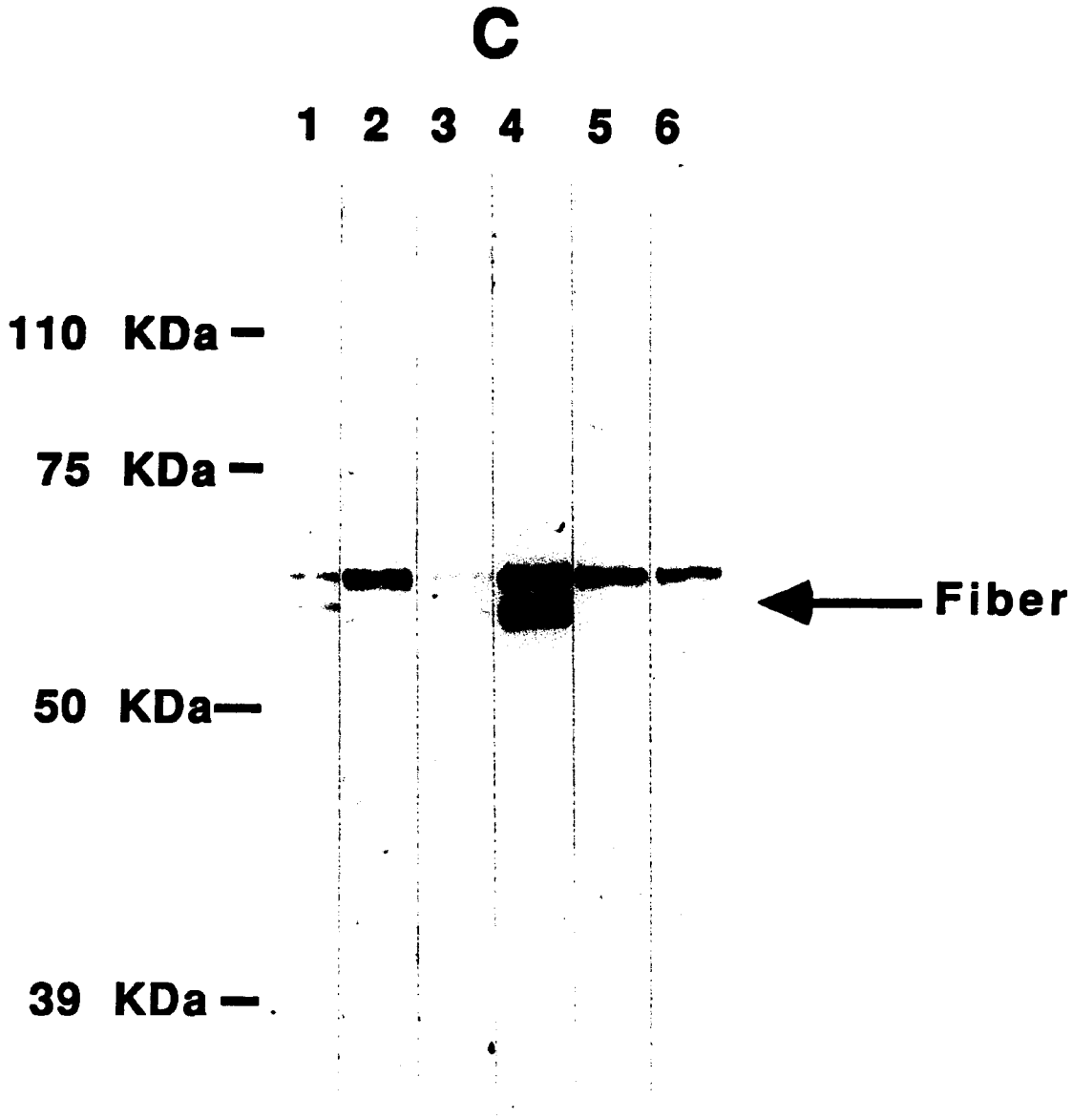
B



2A6-36

Figure 6: Immunoblot on H5ts142-infected cell lysates after the addition of cycloheximide, followed by temperature shift. Samples were either unboiled (panel A) or boiled (panels B and C) prior to loading on a 10% SDS-PAGE. T and M or fiber show the position expected for trimers or monomers, respectively. In panels A and B, lanes 1 and 2 were developed with antibody 4D2-5 (specific for monomers and trimers) and lanes 3 and 4 were developed with 2A6-36 (specific only for trimers). Lanes 1 and 3 were held at 39.5°C after cycloheximide addition, while lanes 2 and 4 were shifted from 39.5°C to 32°C one hour after cycloheximide addition. In Panel C, all lanes were developed with antibody RL2 (specific for O-linked GlcNAc); lanes 1 through 3 were held at 39.5°C after cycloheximide addition, while lanes 4 through 6 were shifted from 39.5° to 32°C one hour after cycloheximide addition. Lanes 1 and 4 are samples from wild type Ad5 infected cells; lanes 2 and 5 are from Ad5 H5ts142 infected cells; lanes 3 and 6 are from uninfected HeLa cells.


A
B
1 2 3 4
1 2 3 4
T
M


C
1 2 3 4 5 6
110 KDa
75 KDa
50 KDa
39 KDa
Fiber

component for the function and targeting of fibers from other serotypes of adenovirus.

2. Some cellular proteins contain O-linked GlcNAc and remain in the cytoplasm or associated with the cytoplasmic surface of other organelles (14, 15).

If fiber is typical of other nuclear proteins, then it seems unlikely that GlcNAc will be a required part of the signal for transport of proteins destined for the nucleus. The precise sequences required for targeting of the adenovirus fiber protein to the nucleus will be determined in future studies.

Fiber shares an epitope with several of the proteins of nuclear pore complex that contain O-linked GlcNAc, since it is recognized by the monoclonal antibody RL2 on immunoblots (Figure 2, panel B); the mutant fiber grown at 39.5°C lacks this epitope. A sequence motif has been proposed for the attachment site for O-linked GlcNAc (11). Since mutant H5ts142 fibers made at 32°C contain GlcNAc, this sequence motif for GlcNAc addition must be present on the molecule at 32°C. Since this sequence must also be present on H5ts142 fiber at 39.5°C (where GlcNAc addition is blocked), there must also be structural determinants required to correctly present the sequence motif to putative GlcNAc transferases. Since H5ts142 fibers grown at 39.5°C are found in the nucleus and yet are not glycosylated when shifted down to 32°C, either an enzyme capable of GlcNAc addition to fiber is not available in the nucleus or the molecule is not able to reform the structure to correctly present the sequence motif to the glycosyltransferase after the shift back to the permissive temperature.

Little is known about the role of monomeric O-linked GlcNAc on glycoproteins. In the case of fibers from Ad5 (and presumably Ad2), O-linked GlcNAc may help in the stabilization of trimers, as suggested by the

indirect immunofluorescence experiments on H5ts142 at 39.5°C: when shifted to the nonpermissive temperature, H5ts142 trimers formed at 32°C and containing GlcNAc did not appear to dissociate. However, when mutant fibers synthesized at 39.5°C are shifted to 32°C, they still lack GlcNAc but can form trimers, suggesting that GlcNAc is required for stabilization and not for assembly of trimers. This observation supports the possibility that O-linked GlcNAc may play a role in the stabilization of multimeric complexes (5, 6, 12, 16, 17).

ACKNOWLEDGEMENTS

We thank Drs. Richard Marchase and Casey Morrow for their advice and their critique of the manuscript. We thank Dr. Harold Ginsberg for the gift of Ad5 H5ts142 virus and Dr. Larry Gerace for the gift of the RL2 antibody. This work was supported by NIH grant GM38234 (to J.A.E.). One of us (K.G.M.) was supported by training grant (T32 AI07150 from the National Institutes of Health). Preparation of monoclonal antibodies was done by the Hybridoma Core Facility of the Multipurpose Arthritis Center (grant NIH 5P60 AR20614).

REFERENCES

1. Caillet-Boudin, M.-L., G. Strecker and J.-C. Michalski. 1989. O-linked GlcNac in serotype-2 adeonvirus fibre. Eur.J.Biochem. 184:205-211.

2. Chee-Sheung, C.C. and H.R. Ginsberg. 1982. Characterization of a temperature-sensitive fiber mutant of type 5 adenovirus and effect of the mutation on virion assembly. J. Virol. 42:932-950.

3. Chelsky, D., R. Ralph and G. Jonak. 1989. Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol. and Cell Biol. 9:2487-2492.

4. Dabauvalle, M.-C., B. Schulz, U. Scheer, and R. Peters. 1988. Inhibition of nuclear accumulation of karyophilic proteins in living cells by micro-injection of the lectin wheat germ agglutinin. Exp. Cell Res. 174:291-296.

5. Davis, L.I. and G. Blobel. 1986. Identification and characterization of a nuclear pore complex. Cell. 45:699-709.

6. Davis, L.I. and G. Blobel. 1987. Nuclear pore complex contains a family of glycoproteins that includes p62: Glycosylation through a previously unidentified pathway. Proc. Natl. Acad. Sci. USA. 84:7552-7556.

7. Dingwall, C. and R.A. Laskey. 1986. Protein import into the cell nucleus. Ann. Rev. Cell Biol. 2:367-90.

8. Dingwall, C., J. Robbins, S.M. Dilworth, B. Roberts, and W.D. Richardson. 1988. The nucleoplasmin nuclear location sequence is larger and more complex that that of SV-40 large T antigen. J. Cell Biol. 107:841-849.

9. Finlay, D.R., D.D. Newmeyer, T.M. Price, and D.J. Forbes. 1987. Inhibition of in vitro nuclear transport by a lectin that binds to nuclear pores. J. Cell Biol. 104:189-200.

10. Hall, M.N., L.M. Hereford, and I. Herskowitz. 1984. Targeting of E. coli ß-galactosidase to the nucleus in yeast. Cell 36:1057-1065.

11. Haltiwanger, R.S., G.D. Holt and G.W. Hart. 1990. Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetyl-glucosaminyltransferase. J. Biol. Chem. 265:2563-2568.

12. Hanover, J.A., C.K. Cohen, M.C. Willingham and M.K. Park. 1987. O-linked N-acetylglucosamine is attached to proteins of the nuclear pore. J. Biol. Chem. 262:9887-9894.

13. Harlow, E. and D. Lane. "Antibodies: a laboratory manual." 1988 Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

14. Hart, G.W., R.S. Haltiwanger, G.D. Holt, and W.G. Kelly. 1989. Glycosylation in the nucleus and cytoplasm. Ann. Rev. Biochem. 58:841-874.

15. Hart, G.W., G.D. Holt and R.S. Haltiwanger. 1988. Nuclear and cytoplasmic glycosylation. Trends Biochem. Sci. 13:380-384.

16. Holt, G.D., R.S. Haltiwanger, R. Torres and G.W. Hart. 1987. Erythrocytes contain cytoplasmic glycoproteins. J.Biol.Chem. 262:14847-14850.

17. Holt, G.D., C.M. Snow, A. Senior, R.S. Haltiwanger, L. Gerace and G.W. Hart. 1987. Nuclear pore complex glycoproteins contain cytoplasmically disposed O-linked N-acetylglucosamine residues. J.Cell Biol. 104:1157-1164.

18. Horwitz, M.S., M.D. Scharff and J.V. Maizel. 1969. Synthesis and assembly of adenovirus 2. Virology. 39:682-694.

19. Ishibashi, M. and J.V. Maizel. 1974. The polypeptides of adenovirus VI. Early and late glycoproteins. Virology. 58:345-361.

20. Jackson, S.P. and R. Tjian. 1988. O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. Cell. 55:125-133.

21. Kalderon, D., W.D. Richardson, A.F. Markham, and A.E. Smith. 1984. Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature. 311:33-38.

22. Kelly, W.G. and G.W. Hart. 1989. Glycosylation of chromosomal proteins: Localization of O-linked N-acetylglucosamine in drosophila chromatin. Cell. 57:243-251.

23. Kornfeld, S. 1987. Trafficking of lysosomal enzymes. FASEB J. 1:462-468.

24. Laemmli, E.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227:680-685.

25. Lyons, R.H., B.Q. Ferguson, and M. Rosenberg. 1987. Pentapeptide nuclear localization signal in adenovirus E1a. Mol. Cell. Biol. 7:2451-2456.

26. Maizel, J.V., D.O. White and M.D. Scharff. 1968. The polypeptides of adenovirus. I. Evidence for multiple protein components in the virion and a comparison of types 2,7A, and 12. Virology. 36:115-125.

27. Newmeyer, D.D., D.R. Finlay and D.J. Forbes. 1986. In vitro transport of a flourescent nuclear protein and exclusion of non-nuclear proteins. J. Cell Biol. 103:2091-2102.

28. Newmeyer, D.D. and D.J. Forbes. 1988. Nuclear import can be separated into distinct steps in vitro: Nuclear pore binding and translocation. Cell. 52:641-653.

29. Newmeyer, D.D., J.M. Lucocq, T.G. Burlin and E.M. De Robertis. 1986. Assembly in vitro of nuclei active in nuclear protein transport: ATP is required for nucleoplasm accumulation. EMBO J. 5:501-510.

30. Roberts, B.L., W.D. Richardson, and A.E. Smith. 1987. The effect of protein context on nuclear location signal function. Cell 50:465-475.

31. Schindler, M. 1984. The carbohydrate moieties of nuclear glycoproteins are predominantly N-acetylglucosamine. J. Cell Biol. 99:99a.

32. Schindler, M., M. Hogan, R. Miller, and D. DeGaetano. 1987. A nuclear specific glycoprotein representative of a unique pattern of glycosylation. J. Biol. Chem. 262:1254-1260.

33. Silber, P.A., L.P. Keegan, and M. Ptashne. 1984. Amino terminus of the yeast GAL 4 gene product is sufficient for nuclear localization. Proc. Natl. Acad. Sci. USA 81:5951-5955.

34. Snow, C.M., A. Senior and L. Gerace. 1987. Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. J. Cell Biol.104:143-156.

35. Torres, C.-R. and G.W. Hart. 1984. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. J. Biol. Chem. 259:3308-3317.

36. Yoneda, Y., N. Imamoto-Sonobe, M. Yamaizumi, and T. Uchida. 1987. Reversible inhibition of protein import into the nucleus by wheat germ agglutinin injected into cultured cells. Exp. Cell Res. 173:586-595.

37. Zhao, L.-J., and R. Padmanabhan. 1988. Nuclear transport of adenovirus DNA polymerase is facilitated by interaction with preterminal protein. Cell. 55:1005-1015.

THE AMINO TERMINUS OF THE ADENOVIRUS FIBER PROTEIN ENCODES THE NUCLEAR LOCALIZATION SIGNAL

by

JEONG SHIN HONG AND JEFFREY A. ENGLER

ABBREVIATIONS

Ad2, Ad5, Ad7, adenovirus serotypes 2, 5 and 7 (respectively); BCIP, 5-bromo-4-chloro-3-indolyl-phosphate; BUdR, 5-bromodeoxyuridine; DMEM, Dulbecco's modified Eagle's medium; gpt, the E. coli gene which codes for the enzyme xanthine-guanine phosphoribosyl-transferase; mAb, monoclonal antibody; MPA, mycophenolic acid; NBT, 4-nitro blue tetrazolium chloride; X-gal, 5-bromo-4-chloro-3-indolyl-b-D-galacto-pyranoside.

ABSTRACT

Using a recombinant vaccinia virus vector, the fiber protein from adenovirus serotype 2 has been expressed in human cells; the protein expressed was correctly assembled into trimers, glycosylated and transported to the nucleus. Deletion of amino acids 2-5 (KRAR) resulted in accumulation of fiber in the cytoplasm; fusion of the sequence TKRVRL, found at the beginning of Ad7 fiber, to the N-terminus of this mutant restored correct targeting. Changing the charge of amino acids 91 and 92 within another potential targeting sequence (LKKTK to LEETK) had little effect on nuclear targeting. Deletions at the C-terminus of Ad2 fiber had no apparent effect on targeting. When fused to the N-terminus of ß-galactosidase and expressed in recombinant vaccinia virus, neither MKRARP nor MTKRVRL (from Ad2 and Ad7 fibers, respectively), were sufficient for efficient transport of the hybrid protein to the nucleus; on the other hand, fusion of MKRARPSEDTF (from Ad2 fiber) or of MKRPRP (a known targeting sequence from the C-terminus of Ad2 E1A proteins) to ß-galactosidase were localized to the nucleus. These results suggest that sequences at the N-terminus of Ad2 and Ad7 fiber are required for correct nuclear targeting.

INTRODUCTION

Adenovirus capsid proteins are synthesized in the cytoplasm and transported to the nucleus for assembly into the virus particles (for a review, see Tooze, Pettersson). The three major capsid proteins (hexon, penton base, and fiber) are synthesized in large excess late in infection. Fiber plays a crucial role in adenovirus infection by attaching the virus to a specific receptor of the cell surface. Green et al.(14) proposed that Ad2 and Ad5 fibers consists of three domains: an N-terminal tail that interacts with penton base (7), a shaft composed of 22 repeats of a 15 amino acid segment that forms ß-sheet and ß-bends, and a knob at the C-terminus that contains the type-specific antigen and is responsible for binding to the cell surface receptor. A similar model has been proposed for Ad3 and Ad7 fiber (15, 32). An analysis of the virion composition by Van Oostrum and Burnett(35) showed that the fiber of Ad2 is most likely a trimer when found on the virion. Devaux et al. (7) proposed a triple-helical model for the fiber in which the shaft of the fiber is stabilized by both inter-chain hydrogen bonds and hydrophobic interactions.

The uptake of proteins by the nucleus is known to be extremely selective. The entry of proteins into the nucleus is mediated by the nuclear pore complexes in the nuclear envelope (10, for a review 8). The transport through the nuclear pore complex is a multistep process and requires ATP (1, 29, 30). It is believed that nuclear proteins contain a sequence that allows their selective accumulation in the nucleus. Amino acid sequences required for nuclear localization are known for many proteins; for example, an internal seven amino acid sequence for SV40 T Ag (16, 17). Although the precise sequence requirements for the nuclear targeting signal is somewhat variable, these sequences seem to be rich in positively charged amino acids, lysine and arginine (3, 8). The presence of this

signal with a primary amino acid sequence is necessary but not sufficient to ensure nuclear entry. It is also known that multiple signals can cooperate to enhance nuclear accumulation (11). Although SV40 prototypic signal sequence (PKKKRKVE) can function independently and is sufficient for nuclear localization (17), its activity is crucially dependent on the protein context within which it is present (31). There are several examples of proteins which require two distant regions in the polypeptide sequence involved in the nuclear translocation (18, 24, 26, 28).

Nuclear transport of several adenovirus proteins has been reported. Cepko and Sharp (2) previously reported that hexon requires an action of a major late protein of adenovirus, 100K, in trimer assembly and nuclear transport. Also the nuclear transport of adenovirus DNA polymerase is facilitated by interaction with pTP (37). On the other hand, adenovirus E1a protein and DBP (DNA binding protein) include nuclear targeting signals in itself. The nuclear localization signal (NLS) of E1A, KRPRP, is located at the C-terminal end of the protein (20). DBP contains a two part NLS: both of those sequences are required for the nuclear targeting of DBP (24). However, the nuclear localization defect of DBP mutants can be complemented by viral infection, perhaps through the interaction of the mutant polypeptide with a virus-encoded or -induced factor (24).

In this study, a recombinant vaccinia virus system has been used to express the Ad2 fiber protein. The fiber produced in this way is indistinguishable from that isolated from adenovirus-infected cells. Mutations in the fiber gene have been used to map sequences important for the correct accumulation of this protein in the nucleus: only the amino-terminal sequence, KRAR, appears to be essential. These mutant fibers are still able to form trimers, suggesting that nuclear transport is not a

prerequisite for trimer formation and that sequences at the amino terminus are not required for trimer formation.

MATERIALS AND METHODS

Enzymes and chemicals. Restriction endonucleases, T4 DNA ligase, T4 DNA polymerase, calf intestine alkaline phosphatase, and Klenow enzyme were purchased from Boehringer Mannheim Biochemicals (Boehringer-Mannheim, Indianapolis, IN); Sequenase^{TM} for DNA sequencing was purchased from United States Biochemicals (Cleveland, Ohio). MPA (10mg/ml; from Sigma Chemical, St Louis, MO) and xanthine (10mg/ml; from Sigma) were dissolved in 0.1 N NaOH, and hypoxanthine (10mg/ml; from Sigma) was dissolved in water and sterile filtered;the solutions were stored frozen. Stock solutions of BUdR (2.5 mg/ml; from Sigma) were made in water, sterile filtered, and kept at 4°C. Stock solutions of X-gal (20mg/ml or 100mg/ml; from Boehringer-Mannheim) were made in dimethylformamide and stored at -20°C.

Cells and viruses. HeLa T4 cells and HEp2 cells were maintained in DMEM (from GIBCO) with 10% bovine calf serum (Hyclone, Logan, Utah). CV1 cells were maintained in DMEM with 10% fetal bovine serum (Hyclone). TK$^-$ cells were maintained in DMEM with 10% bovine calf serum and 25 μg/ml BUdR. Wild-type vaccinia virus was replicated in rabbit kidney cells maintained in DMEM with 10% bovine calf serum and purified as reported previously (21).

Construction of recombinant plasmids. The Ad2 fiber gene was removed from plasmid pUC2Fm (DraI-SmaI DNA fragment containing Ad2 fiber cloned in pUC119) using KpnI or Asp718 restriction enzyme; in this plasmid, the Ad2 fiber gene was inserted into pUC119 with one KpnI cleavage site inserted directly in front of the ATG initiation codon and another located in the polylinker after the end of the fiber gene. The

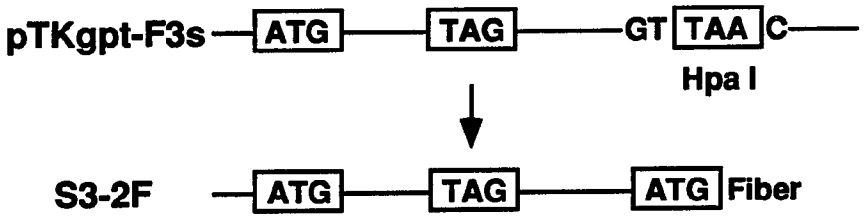
overhangs that resulted from the restriction digests were removed or repaired by T4 DNA polymerase or Klenow enzyme (22). The DNA fragment bearing Ad2 fiber gene was isolated by electroelution from an agarose gel using a DEAE membrane (Schleicher & Schuell, Keene, NH) and was cloned into HpaI site of the vaccinia virus expression vector pTKgpt-F3s (9) by blunt end ligation. In the resulting construct (plasmid S3-2F; Fig. 1), translation stop codons are located immediately upstream of the initation codon for translation of fiber, so that the fiber protein expressed starts at the authentic ATG used in vivo.

Formation and selection of gpt^+ recombinant virus. Recombinant viruses were prepared as described by Falkner and Moss (9) with the following modifications. Confluent monolayers of CV1 cells were infected with 0.1 PFU of vaccinia virus per cell. At 2 hours postinfection, 20 μg of recombinant plasmid was transfected into the cells by calcium-phosphate procedure described previously (4). A pooled virus stock was prepared the next day by resuspending the infected cells in 1 ml of DMEM (without serum), freezing and thawing 3 times, and sonicating for 2 min (4 times at 30 sec each). The virus stocks were aliquoted and kept at -100°C. To select gpt^+ recombinant viruses, CV1 cells preincubated in gpt selection medium (27) for 20-24 hours were infected with diluted virus pool described above. gpt-resistant recombinant viruses were plaque purified from infected TK^- cells in BUdR selection medium; plaques were picked after 2 days and expanded in HEp2 cells.

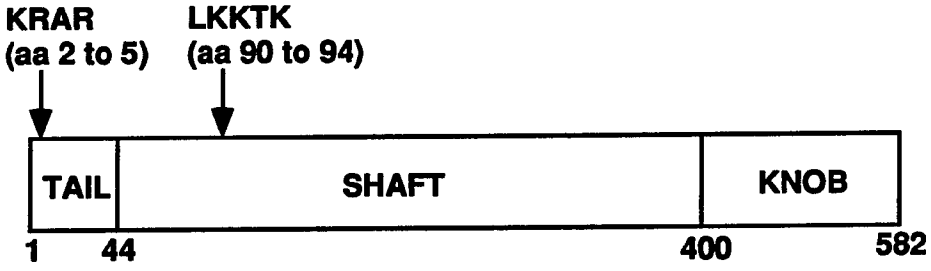
Immunoblot. Cell lysates were made from recombinant vaccinia virus infected CV1 cells by freeze-thawing the cells 3 times, sonicating for 2 min, and extracting with freon. Fifteen μg of each protein sample was loaded in sample buffer (19) and then separated by SDS-PAGE on a 10% gel; the proteins in the gel were then blotted onto nitrocellulose

Figure 1: Panel A: Construction of recombinant plasmid. Vector plasmid pTKgpt-F3s (9) was used for expression of fiber in vaccinia virus recombinants; this vector allows expression of foreign genes under control of the p11 promoter and allows selection of recombinant viruses by selection for either gpt-resistance or TK^- phenotype. The fiber gene was cloned into the unique HpaI cleavage site by blunt end ligation. The ATG boxes show the relative positions of the first (vaccinia specific) and second (fiber specific) initiation codons downstream from the start of transcription (+1); the encoded fiber protein is expressed after reinitiation of translation at the bonafide ATG present in the fiber open reading frame. Panel B: Location of possible nuclear targeting sequences within the protein. Based on the consensus sequences for nuclear targeting (3, 34), two potential signals can be identified within Ad2 and Ad5 fibers: KRAR (amino acids 2 to 5) and LKKTK (amino acids 90-94). Ad3 and Ad7 fiber contain a potential signal at the amino terminus similar to KRAR, but have no sequence comparable to LKKTK (15, 32).

A.


pTKgpt-F3s
ATG
TAG
GT
TAA
C
Hpa I
S3-2F
ATG
TAG
ATG
Fiber

B. Ad2 Fiber


KRAR
(aa 2 to 5)
LKKTK
(aa 90 to 94)
TAIL
SHAFT
KNOB
1
44
400
582

and incubated in Blotto (5% nonfat dry milk in PBS) to block nonspecific antibody binding. Either anti-fiber (4D2-5) or anti-trimer (2A6-36) mAb were used as the primary antibody (1:1000 dilution in Blotto). Biotin-conjugated goat anti-mouse Ig antibody (Fisher Biotech) and alkaline phosphatase conjugated to streptavidin (Fisher Biotech or Jackson ImmunoResearch) were used to probe the specific binding of monoclonal antibodies to fiber; the color reaction for alkaline phosphatase was developed with NBT and BCIP (Bethesda Research Labs).

Immunofluorescence. Cells were plated on a glass coverslip in 35mm plate. At 60 hr postinfection, cells were fixed with 3% paraformaldehyde in PBS (phosphate buffered saline) either at 32°C (for cells grown at the permissive temperature) or 39.5°C (for cells grown at the nonpermissive temperature). TBS-TX (Tris buffered saline-1% Triton X-100) was used to permeabilize the cell membrane, prior to incubation with anti-fiber antibody (4D2-5) or anti-trimer antibody (2A6-36) at room temperature for 1 hour. Goat anti-mouse Ig conjugated to fluorescein isothiocyanate (Jackson Immunoresearch, West Grove PA) was used as secondary antibody. The cells were photographed on a Nikon Optiphot microscope equipped for fluorescent illumination at a magnification of 20x or 40x using Kodak Gold 400 film.

Mutagenesis of Ad2 fiber genes. Single-stranded pUC2Fm DNA was prepared by infecting TG1 cells containing the plasmid with R408 helper phage (Stratagene), as described by Viera and Messing (36). DNA oligomers for mutagenesis and sequencing were synthesized on an Applied Biosystems 380 DNA synthesizer and purified by HPLC. Oligonucleotide-directed mutagenesis was performed as described elsewhere (38). The mutants were screened by colony hybridization with ^{32}P-labeled mutagenic oligomers and confirmed by plasmid sequencing

(23) using Sequenase™. The mutant Ad2 fiber genes were shuttled into the expression vector, as described previously for the construction of the wild-type plasmid.

Construction of ß-galactosidase fusion proteins. Plasmid pMLB1034 (33) contains the lacZ gene lacking a promoter and ATG initiation codon and was the source of ß-galactosidase gene used for these experiments. The plasmid was digested with EcoRI and BamHI and treated with alkaline phosphatase. Five sets of DNA oligomers (two 11-mers, four 26-mers, and two 41-mers; Table 1) were designed to make M-, MKRARP-, MTKRVRL-, MKRPRP-, or MKRARPSEDTF-ß-galactosidase fusion proteins. Each set of purified DNA oligomers were kinased, annealed together, and cloned into the digested pMLB1034 vector. The linker insertion was confirmed by screening for blue colonies on DYT-agar plates (22) containing X-gal, by restriction endonuclease digestion, and by plasmid sequencing. The ß-galactosidase gene with linker was shuttled into the HpaI site of pTKgpt-F3s expression vector by ligating the gene after cutting it out with EcoRI and MscI (or DraI) and repairing the overhangs with Klenow enzyme.

RESULTS

Construction of vaccinia recombinant viruses expressing Ad fiber. Vaccinia virus expression vector pTKgpt-F3s(9) was used to clone and express the Ad2 fiber gene. The vector has multiple cloning sites downstream of the gene encoding the 11K polypeptide and also has the E. coli gpt gene under the control of the vaccinia p7.5 promoter. Coexpression of the gpt gene provides an efficient dominant selection for recombinants in the presence of MPA, an inhibitor of purine metabolism which reversibly blocks the replication of vaccinia virus in normal cell lines (9). This inhibition could be overcome by adding xanthine and hypoxanthine to the

Table 1. Sequences of the DNA oligomers used in this study

A. Mutagenic oligonucleotides for Ad2 fiber gene

oligonucleotide 1575: 28-mer, for KRAR (aa 2 to 5) deletion
5' TGTCT TCAGA CGGAG CCATG GATCC TGC 3'

oligonucleotide 1576: 34-mer, LKKTK (aa 90 to 94) to LEETK substitution
5' TTATG TTTGA CTTTG TTTCT TCAAG TGGCT GAGT 3'

oligonucleotide 1790: 43-mer, KRAR to TKRVRL substitution
5' AGGTG TCTTC AGAGA GCCGG ACTCT CTTGG TCATG GTACC TGC 3'

oligonucleotide 1791: 45-mer, relocation of KRAR to the C-terminus of Ad2 fiber gene
5' TTCAC GATTC TTTAC GGTCT GGCGC GTTTT GGTTC CTGGG CAATG 3'

All of the mutagenic oligonucleotides are annealed to the sense strand of the gene.

B. Oligonucleotide adapters for fusion of test targeting sequences to the ß-galactosidase gene.

M-:
```
5'  AATTCACCATG      3'
3'      GTGGTACCTAG  5'
```

MKRARP-:
```
5'  AATTC ACCAT GAAAC GCGCC AGACC G      3'
3'      G TGGTA CTTTG CGCGG TCTGG CCTAG  5'
```

MTKRVRL-:
```
5'  AATTC ACCAT GACCA AGAGA GTCCG GCTC      3'
3'      G TGGTA CTGGT TCTCT CAGGC CGAGCTAG  5'
```

MKRPRP-:
```
5'  AATTC ACCAT GAAAC GCCCC AGGCC A      3'
  3'    G TGGTA CTTTG CGGGG TCCGG TCTAG  5'
```

MKRARPSEDTF-:
```
5'  AATTC ACCAT GAAAC GCGCC AGACC GTCTG AAGAC ACCTT C      3'
3'      G TGGTA CTTTG CGCGG TCTGG CAGAC TTCTG TGGAA GCTAG  5'
```

All of the linkers are designed to have overhangs corresponding to the overhangs of EcoRI and BamHI on each end.

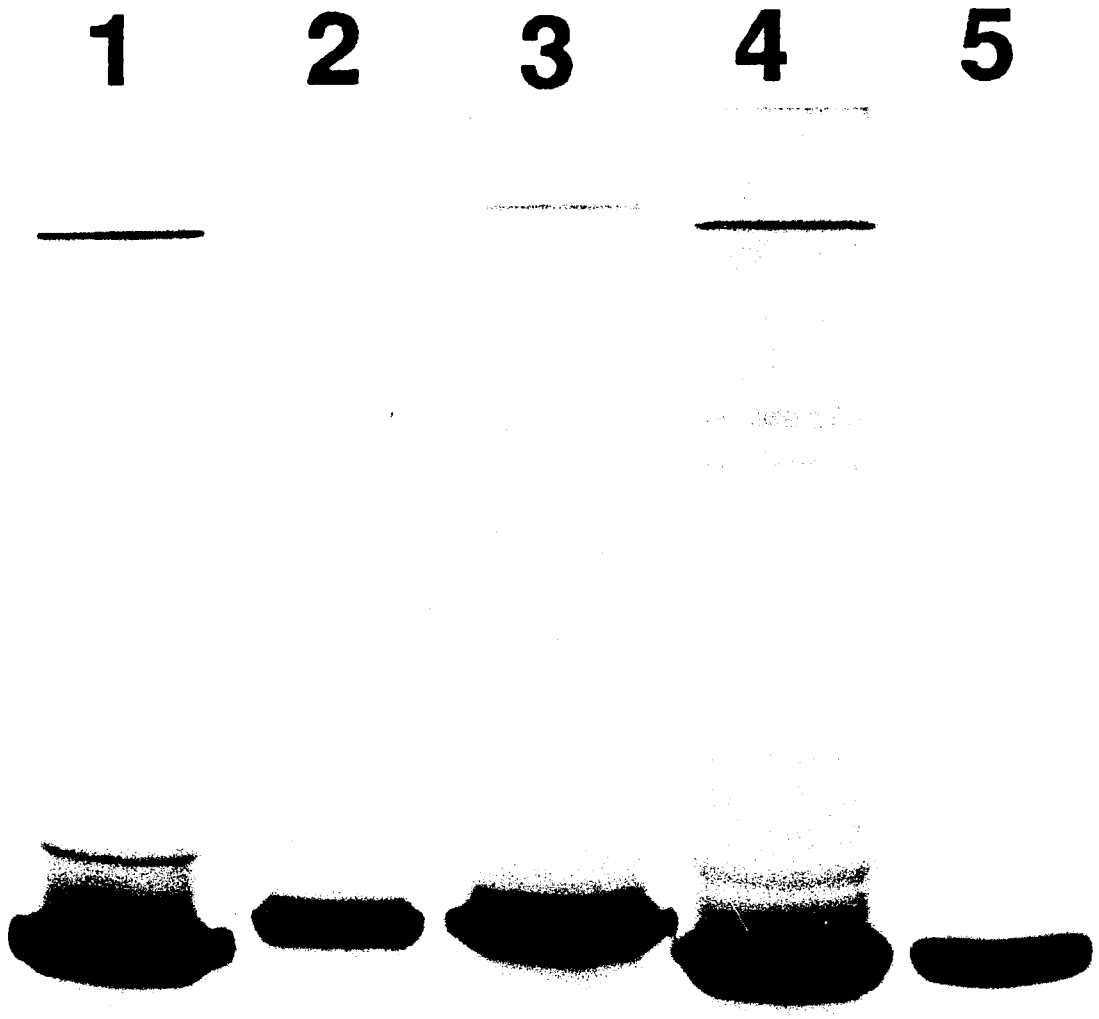
growth medium. In order to avoid adding foreign amino acid sequences onto the amino terminus of the protein, the Ad2 fiber gene was cloned into HpaI site of pTKgpt-F3s, downstream from the first termination codon in the vector (Fig. 1A). The fiber was then expressed by reinitiating protein translation 15 nucleotides downstream at its own ATG codon.

In order to test whether these vaccinia virus recombinants express fiber protein, Western blot analysis of infected cell lysates was performed. As can be seen in Figure 2, fiber protein can be detected in cell lysates from recombinant-infected cells. Other plasmid constructs in which fiber was expressed directly from the vaccinia virus ATG in this vector (with a few extra N-terminal amino acids) expressed 5 to 10 fold lower levels of fiber, when incorporated into recombinant vaccinia virus (data not shown).

Nuclear localization of Ad2 fiber in vaccinia recombinants. In order to test whether fiber protein directs its own transport to the nucleus, indirect immunofluorescence experiments were performed using cells infected with the vaccinia virus recombinants expressing fiber without the presence of other adenovirus proteins. Since vaccinia virus grows entirely in the cytoplasm (for review, see 25), it presumably could not provide compensating factors that might be necessary for fiber transport to the nucleus. As shown in the indirect immunofluorescence experiments in Figure 3 (panels A and B), wild-type fiber expressed in recombinant vaccinia virus was transported to the nucleus, without other adenoviral proteins. This result suggests that there are at least two different systems for nuclear targeting of adenovirus structural proteins: one independent (fiber) and one dependent (hexon) on the action of the 100K protein.

The ability of fiber to form trimers when expressed in recombinant vaccinia virus was also investigated. Both immunofluorescence with 2A6-36 mAb (which recognizes only trimers) and immunoblots of unboiled

Figure 2. Immunoblot on cell lysates infected with recombinant vaccinia virus. Lane 1:S3-2F (wild type Ad2 fiber); lane 2: 2F-ΔKRAR (Ad2 fiber with KRAR deletion at the N-terminus); lane 3: 2F-LEETK (Ad2 fiber with substitution from KK to EE at aa 91 and 92); lane 4: 2F-TKRVRL (Ad2 fiber with substitution of KRARP to TKRVRL from Ad7 fiber); lane 5: 2F-ΔKRR-PKRARP (Ad2 fiber with relocation of KRARP from N-terminus to the C-terminus). The cell lysates were made from recombinant vaccinia virus infected HEp2 cells. The blot was developed with 4D2-5 Ab. All of the recombinant fibers were recognized by this antibody.


1
2
3
4
5

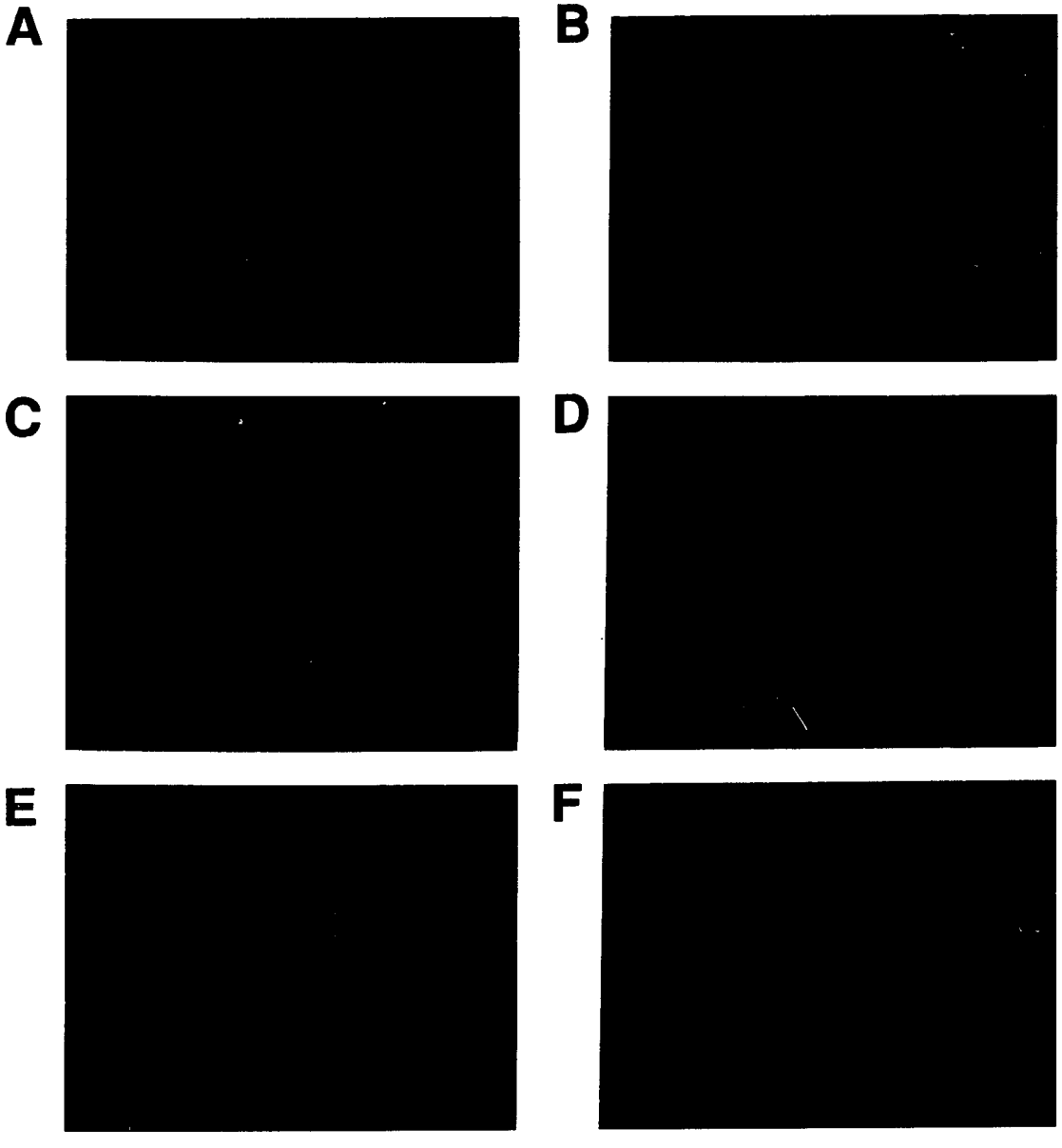
samples from infected cells (data not shown) show that trimerization of the fiber does not require other adenovirus proteins.

Identification of nuclear localization signal for Ad2 fiber. Since all of the known nuclear localization signals (NLS) are rich in basic amino acid residues (3, 8), two potential NLS sequences with stretches of positively charged amino acids were identified within the predicted Ad2 fiber sequence (Fig. 1B): KRAR (aa 2 to aa 5) and LKKTK (aa 90 to aa 94). The KRAR and LKKTK sequences are conserved in fibers from Ad2 and Ad5 (subgroup C); fibers from Ad3 and Ad7 (subgroup B) (15, 32) have a sequence, TKRVRL, analogous in composition and position to KRAR, but lack a homolog to LKKTK. Fiber mutants deleted for the KRAR sequence or containing LEETK (rather than LKKTK) were constructed by site-directed mutagenesis. When expressed in vaccinia recombinants and analyzed by immunoblot, each mutant showed levels of protein expression similar to that observed for wild type fiber (data not shown).

Immunofluorescence of a KRAR deletion mutant (Fig. 3, C and D) shows that deletion of these four amino acids are enough to abort the nuclear transport of the fiber protein and the mutant fiber stays solely in the cytoplasm of the cell. Interestingly, this mutant fiber forms trimers in the cytoplasm, as shown with the 2A6-36 mAb (Fig. 3D). Although it is not clear whether the trimer is formed before or after the transport to the nucleus during normal adenovirus infection, correct translocation is not a prerequisite for trimer formation.

The indirect immunofluorescence of the fiber mutant with substitution of LEETK for LKKTK (2F-LEETK; Fig. 3, E and F) shows correct targeting to the nucleus and trimerization of this mutant fiber. Either this sequence is not essential for nuclear transport or its removal is not sufficient to disturb the targeting of this protein.

Figure 3. Localization of fibers in recombinant virus infected HEp2 cells. A and B are S3-2F infected, C and D are 2F-ΔKRAR infected, and E and F are 2F-LEETK infected. 4D2-5 Ab (anti-fiber) was used for A, C, and E; 2A6-36 Ab (anti-trimer) was used for B, D, and F. Ad2 fiber targets to the nucleus in the absence of other adenovirus protein. Deletion of KRAR from the N-terminus of Ad2 fiber completely aborts nuclear targeting of the protein. Substitution of KK to EE at aa 91 and 92 does not seem to have an effect on the targeting of the protein to the nucleus. All of the three recombinant fibers were able to form trimers in either cytoplasm or nucleus, indicating that nuclear transport of the protein is independent from trimer formation.

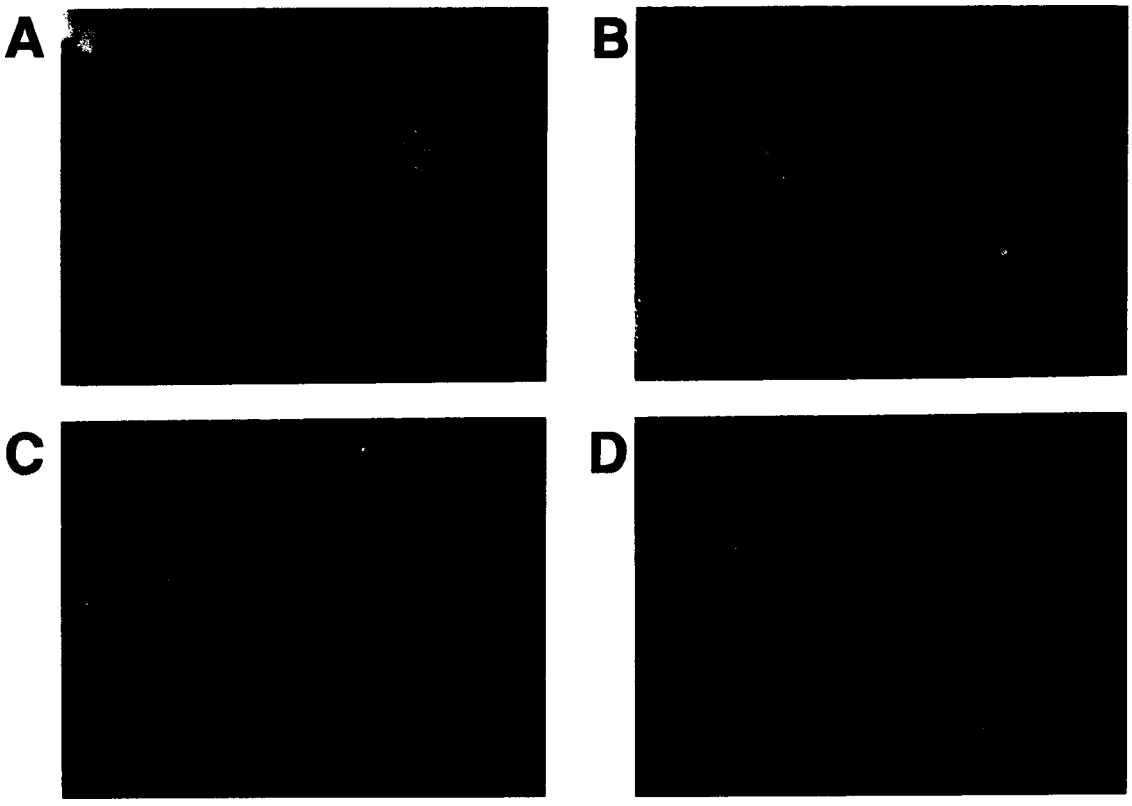

A
B
C
D
E
F

The effect of substitution of TKRVRL for KRAR and relocation of KRARP to the C-terminus of Ad2 fiber. Since the possible nuclear targeting signal KRAR of Ad2 fiber is conserved in fibers from other serotypes, the effect of inserting the corresponding sequence from Ad7 fiber (TKRVRL) into the KRAR deletion mutant was studied. As shown in Figure 4 (A and B), the insertion of the Ad7 sequence restores correct nuclear localization to the ΔKRAR mutant. This hybrid fiber can also form trimers.

In order to test whether the action of the KRAR localization is dependent on the context within the fiber, this signal was relocated to the C-terminus of Ad2 fiber. When this mutant protein was expressed in vaccinia recombinants, it was still located in the cytoplasm of the cells (Fig. 4, C and D). This mutant also retains the ability to form trimers. It has been reported that a nuclear location signal can function at a variety of positions within a protein but that in some locations its activity is masked (31). From these results, we concluded that either KRARP can function only within its proper context at the amino-terminus or KRARP is only a part of the complete signal for nuclear transport of fiber.

The location of ß-galactosidase fusion proteins in a cell. In order to confirm the results of the deletion and replacement experiments, the ability of these sequences to alter the localization of a cytoplasmic protein (ß-galactosidase) was tested. Plasmid pMLB1034 contains the ß-galactosidase gene without a promoter and initiation codon (33); by cloning different adaptors between EcoRI and BamHI restriction sites at the 5' end of the ß-galactosidase gene in this plasmid, the initiation codon and other codons can be fused to ß-galactosidase, which result in proteins with a variety of different N-terminal sequences. The resulting fusion genes were then cloned in the vaccinia virus expression vector, pTKgpt-F3s (9). The

Figure 4. Localization of fibers in 2F-TKRVRL and 2F-ΔKRAR-PKRARP infected HEp2 cells. A and B were infected with 2F-TKRVRL and C and D were infected with 2F-ΔKRAR-PKRARP. 4D2-5 Ab was used for A and C; 2A6-36 Ab was used for B and D. TKRVRL, a corresponding sequence of Ad7 fiber to KRAR in Ad2 fiber, restored the nuclear localization of 2F-ΔKRAR mutant fiber. Relocation of KRARP sequence from N-terminus of Ad2 fiber to the C-terminus fail to target the protein to the nucleus. Both fibers were able to form trimers.


A
B
C
D

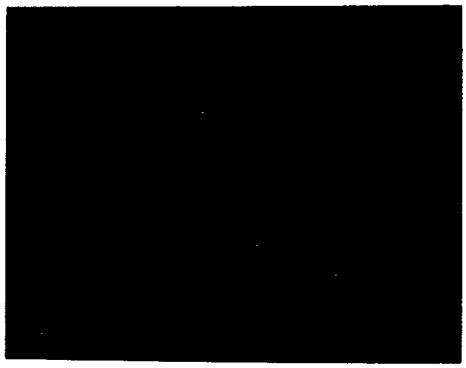
recombinant vaccinia virus infected cells were stained with a mAb for ß-galactosidase and a fluorescent secondary antibody. Wild-type ß-galactosidase containing no fusion at the N-terminus was located in the cytoplasm and the nucleus (Fig. 5A). Both MKRARP-ß-galactosidase and MTKRVRL-ß-galactosidase were also located in both the cytoplasm and the nucleus (Fig. 5, C and D, and 5B, respectively). This result suggested that KRARP and TKRVRL sequences, which appear to be essential for translocation of fiber to the nucleus, are not sufficient to translocate ß-galactosidase into the nucleus. This result was unexpected because the KRPRP sequence from adenovirus E1a protein (which differs from KRARP of Ad2 fiber by only one amino acid) can direct the accumulation of E. coli galactokinase to the nucleus (20). When the KRPRP sequence was fused to the N-terminal end of the ß-galactosidase construct, the protein was found predominantly in the nucleus (Fig. 5, E and F). When 10 amino acids from the N-terminal end of Ad2 fiber (KRARPSEDTF) were fused to ß-galactosidase, the protein was predominanly located in the nucleus (Fig. 5, G and H); the five additional amino acids (SEDTF) probably help to present the sequence in a correct conformation. In some cases (Fig. 5, D, F and H), cycloheximide was added to the infected cells 12 hours prior to fixing for immunofluorescence to determine the relative efficiency of transport: complete transport to the nucleus was observed for the KRPRP and the KRARPSEDTF fusion proteins, but not for the KRARP fusion, suggesting that KRARP is either an inefficient or an incomplete transport signal when coupled to a hybrid protein.

DISCUSSION

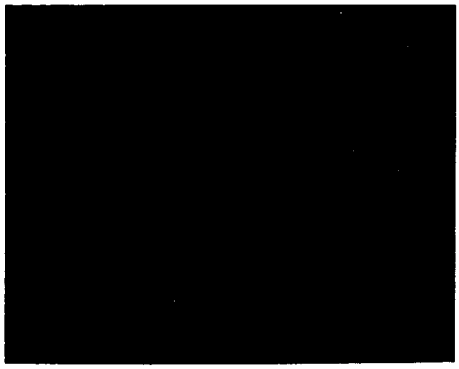
Vaccinia virus expression vector has been used to study the nuclear transport of adfenovirus fiber protein. Since the cells infected with wild type vaccinia virus synthesize viral DNA and assemble progeny virus

Figure 5. Localization of recombinant ß-galactosidase in HeLa T4 cells. A: HeLa T4 cells expressing wild-type ß-galactosidase; it localizes both in the cytoplasm and the nucleus. B: expressing TKRVRL-ß-galactosidase; TKRVRL of Ad7 fiber fail to translocate the recombinant ß-galactosidase into the nucleus. C and D: KRARP-ß-galactosidase (KRARP from Ad2 fiber); E and F: KRPRP-ß-galactosidase (KRPRP from Ad E1a protein); G and H: KRARPSEDTF-ß-galactosidase (KRARPSEDTF from Ad2 fiber). In D, F, and H, cycloheximide (100µg/ml) was added to the infected cells at 48 hours postinfection. Cells were fixed approximately 12 hours after the addition of cycloheximide. Fusion of either KRPRP or KRARPSEDTF to ß-galactosidase localized the protein into the nucleus. In E and G, strong cytoplasmic fluorescence was observed in addition to nuclear fluorescence, indicating that the transport of ß-galactosidase is slower than that of adenovirus fiber. F and H show that both of KRPRP- and KRARPSEDTF-ß-galactosidase located exclusively in the nucleus 12 hours after its synthesis. Fusion of KRARP to ß-galactosidase failed to localize the fusion protein into the nucleus, indicating that nuclear localization of other ß-gal recombinants are by specific targeting signal.

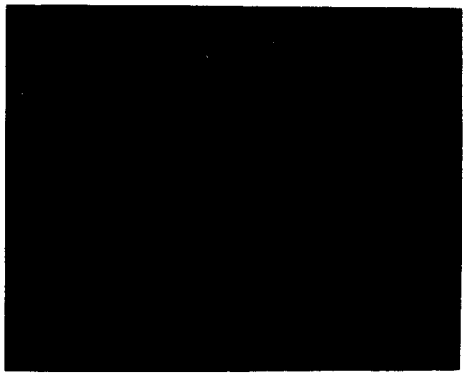
A



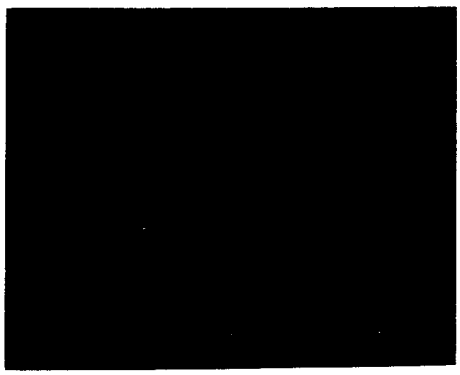
B



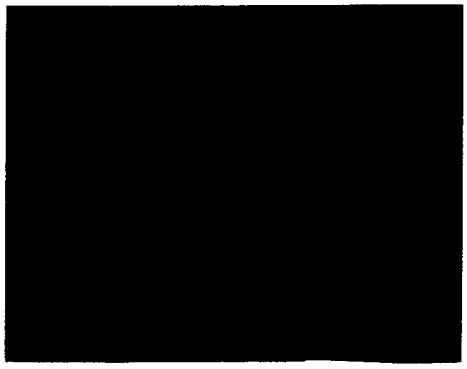
C



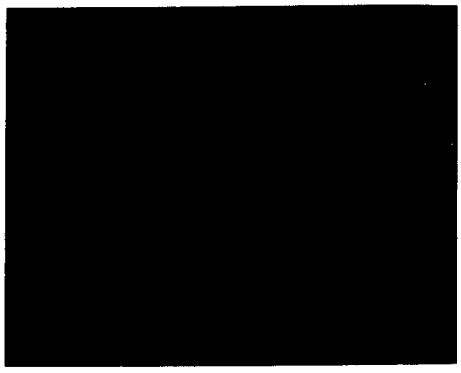
D



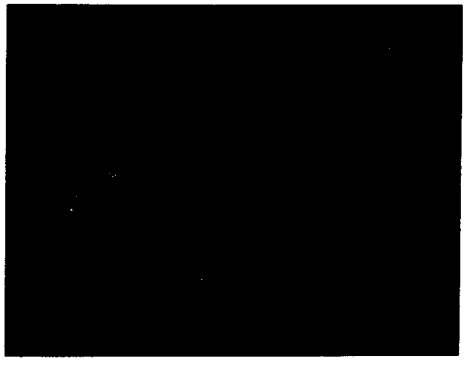
E



F



G



H

particles within the cytoplasm, the nucleus of the infected cell remains relatively intact and it seemed reasonable to use this system to study nuclear translocation of proteins. PB1 and PB2 (polymerase basic protein 1 and 2) of influenza virus have been expressed using vaccinia virus vector and the nuclear translocation of these proteins have been demonstrated (26, 28).

Adenovirus structural proteins are transported to the nucleus to be assembled into virions (12, 13); the mechanism by which these proteins are transported efficiently is not clear. Cepko and Sharp (2) showed that the 100K protein is required for trimerization and translocation of Ad5 hexon to the nucleus. Since Ad2 fiber expressed alone (S3-2F) is transported to the nucleus and trimerized, fiber must contain all of the signals required for its own transport to the nucleus and at least two different systems exist in adenovirus to assemble the structural proteins in the correct conformation and transport the protein to the nucleus.

The data presented in this paper demonstrate that the deletion of 4 amino acids (KRAR) at the N-terminus of Ad2 fiber is sufficient to abolish translocation of the fiber to the nucleus. This defect can be restored by replacing the corresponding sequence, TKRVRL, from the Ad7 fiber onto the N-terminus of the mutant 2F-ΔKRAR. The KRAR sequence is conserved among fibers from several serotypes, such as Ad5, Ad3, and Ad7 (5, 15, 32). Furthermore, these four amino acids of fiber also fit the proposed consensus sequence, K R/K X R/K, predicted for the nuclear localization signal (3). When these four residues were relocated to the C-terminal end of the fiber, it could not function as a nuclear targeting signal anymore. Although these four amino acids are essential for the targeting of the fiber to the nucleus, it was not able to translocate effectively a nonnuclear protein, E. coli ß-galactosidase, into the nucleus. This result

indicates that either other sequences might also be required or that the KRARP signal is much more context dependent than other signals for nuclear translocation of the fiber. When KRPRP- and KRARPSEDTF-sequences were fused onto the N-terminus of ß-galactosidase, both could direct ß-galactosidase to the nucleus. The proline in KRPRP may help to present the signal in a correct conformation and the addition of SEDTF along with KRARP has compensated, by placing KRARP in the proper context, so that both fiber and the fusion protein could be transported into the nucleus.

2F-ΔKRAR fiber is a mutant defective in the nuclear localization, but it forms trimers and could be recognized by 2A6-36 mAb both on immunofluorescence and immunoblot. TKRVRL substitution at the N-terminus of the fiber did not affect on the trimerization of the protein. This result suggests that the N-terminus probably does not have much influence on trimerization of the fiber. Furthermore the N-terminus of the fiber is more sensitive to proteases than the rest of the fiber (7), indicating that the fiber has a less compact structure at the N-terminus. Unlike the translocation of HA in influenza virus which requires trimer formation for translocation through Golgi apparatus (6), these results indicate that the trimerization is not required for translocation of fiber into the nucleus.

ACKNOWLEDGEMENTS

We thank Drs. L. Andrew Ball, Gillian Air, and Dick Compans for their critical comments about the manuscript. This work was supported by Public Health Service grant GM38234 (to J.A.E.). Costs for DNA synthesis were supported in part by grant CA13148 to the UAB Comprehensive Cancer Center.

REFERENCES

1. Akey, C. W. and Goldfarb, D. S. 1989. Protein import through the nuclear pore complex is a multistep process. J. Cell Biol. 109:971-982.

2. Cepko, C. L. and Sharp, P. A. 1983. Analysis of Ad5 hexon and 100K temperature sensitive mutants using conformation-specific monoclonal antibodies. Virol. 129:137-154.

3. Chelsky, D., Ralph, R. and Jonak, G. 1989. Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol. Cell. Biol. 9:2487-2492.

4. Chen, C. and Okayama, H. 1987. High-efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. 7:2745-2752.

5. Chroboczek, J. and Jacrot, B. 1987. The sequence of adenovirus fiber: similarities and differences between serotypes 2 and 5. Virol. 161:549-554.

6. Copeland, C. S., Doms, R. W., Bolzau, E. M., Webster, R. G. and Helenius, A. 1986. Assembly of influenza hemagglutinin trimers and its role on intracellular transport. J. Cell Biol. 103:1179-1191.

7. Devaux, C., Adrian, M., Berthet-Colominas, C., Crusack, S. and Jacrot, B. 1990. Structure of Adenovirus Fibre I. Analysis of Crystals of fibre from Adenovirus serotype 2 and 5 by electron microscopy and X-ray crystallography. J. Mol. Biol. 215:567-588.

8. Dingwall, C. and Laskey, R. A. 1986. Protein import into the cell nucleus. Ann. Rev. Cell Biol. 2:367-390.

9. Falkner, F. G. and Moss, B. 1988. Escherichia coli gpt gene provides dominant selection for vaccinia virus open reading frame expression vectors. J. Virol. 62:1849-1854.

10. Feldherr, C. M., Kallenbach, E. and Schultz, N. 1984. Movement of a karyophilic protein through the nuclear pores of oocytes. J. Cell Biol. 99:2216-2222.

11. Fischer-Fantuzzi, L. and Vesco, C. 1988. Cell-dependent efficiency of reiterated nuclear signals in a mutant simian virus 40 oncoprotein targeted to the nucleus. Mol. Cell. Biol. 8:5495-5503.

12. Flint, S. J. and Broker, T. R. 1980. "The adenoviruses." In DNA tumor viruses. J. Tooze ed., CSH Laboratory, Cold Spring Harbor, New York.

13. Ginsberg, H. S. 1984. The adenoviruses. Plenum Press. New York.

14. Green, N. M., Wrigley, N. G., Russell, W. C., Martin, S. R. and

McLachlan, A. D. 1983. Evidence for a repeating cross-ß sheet structure in the adenovirus fibre. EMBO J. 2:1357-1365.

15. Hong, J. S., Mullis, K. G. and Engler, J. A. 1988. Characterization of the early region 3 and fiber genes of Ad7. Virol. 167:545-553.

16. Kalderon, D., Richardson, W. D., Markham, A. F. and Smith, A. E. 1984. Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature. 311:33-38.

17. Kalderon, D., Roberts, B. L., Richardson, W. D. and Smith, A. E. 1984. A short amino acid sequence able to specify nuclear location. Cell. 39:499-509.

18. Kleinschmidt, J. A. and Seiter, A. 1988. Identification of domains involved in nuclear uptake and histone binding of protein N1 of Xenopus laevis. EMBO J. 7:1605-1614.

19. Laemmli, E. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227:680-685.

20. Lyons, R. H., Ferguson, B. Q. and Rosenberg, M. 1987. Pentapeptide nuclear localization signal in adenovirus E1a. Mol.Cell.Biol. 7:2451-2456.

21. Mackett, M., Smith, G. L. and Moss, B. 1985. "The construction and characterization of vaccinia virus recombinants expressing foreign genes." In DNA cloning: a practical approach. D.M. Glover, D.M. Glovers. IRL Press, Oxford. p191-211.

22. Maniatis, T., Fritsch, E. F. and Sambrook, J. 1982. "Molecular cloning: a laboratory manual." CSH Laboratory. Cold Spring Harbor, New York.

23. Mierendorf, R. C. and Pfeffer, D. 1987. Direct sequencing of denatured plasmid DNA. Meth. in Enzym. 152:556-562.

24. Morin, N., Delsert, C. and Klessig, D. F. 1989. Nuclear localization of adenovirus DNA-binding protein: requirements of two signals and complementation during virus infection. Mol. Cell. Biol. 9:4372-4380.

25. Moss, B. 1990. "Poxviridae and their replication." In Virology. B.N. Fields. Raven Press, Ltd., New York. Chapter 74. p2079-2111.

26. Mukaigawa, J. and Nayak, D. P. 1991. Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2. J. Virol. 65:245-253.

27. Mulligan, R. and Berg, P. 1981. Selection for animal cells, that express the E. coli gene coding for xanthine-guanine phosphoribosyl transferase. Proc. Natl. Acad. Sci. USA. 78:2072-2076.

28. Nath, S. T. and Nayak, D. P. 1990. Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). Mol. Cell Biol. 10:4139-4145.

29. Newmeyer, D. D. and Forbes, D. J. 1988. Nuclear import can be separated into distinct steps in vitro: Nuclear pore binding and translocation. Cell. 52:641-653.

30. Richardson, W. D., Mills, A. D., Dilworth, S. M., Laskey, R. A. and Dingwall, C. 1988. Nuclear protein migration involves two steps: Rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. Cell. 52:655-664.

31. Roberts, B. L., Richardson, W. D. and Smith, A. E. 1987. The effect of protein context on nuclear location signal function. Cell. 50:465-475.

32. Signas, C., Akusjarvi, G. and Petterson, U. 1985. Adenovirus 3 fiber polypeptide gene: Implications for the structure of the fiber protein. J. Virol. 53:672-678.

33. Silhavy, T. J., Berman, M. L. and Enquist, L. W. 1984. "Appendix I. Cloning vectors: pMLB1034." In Experiments with gene fusions. CSH Laboratory, Cold Spring Harbor. p250-252.

34. Smith, A. E., Kalderon, D., Roberts, B. L., Colledge, W. H., Edge, M., Gillett, P., Markham, A., Paucha, E. and Richardson, W. D. 1985. The nuclear location signal. Proc. R. Soc. Lond. B. 226:43-58.

35. van Oostrum, J. and Burnett, R. M. 1985. Molecular composition of the adenovirus type 2 virion. J. Virol. 56, No.2:439-448.

36. Viera, J. and Messing, J. 1987. Production of single-stranded plasmid DNA. Meth. Enzym. 153:3-11.

37. Zhao, L.-J. and Padmanabhan, R. 1988. Nuclear transport of adenovirus DNA polymerase is facilitated by interaction with preterminal protein. Cell. 55:1005-1015.

38. Zoller, M. J. and Smith, M. 1983. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. Meth. Enzym. 100:468-500.

TRIMER FORMATION DEPENDS ON AMINO ACID SEQUENCES IN THE CARBOXYL-TERMINUS OF AD2 AND AD5 FIBER

by

JEONG SHIN HONG AND JEFFREY A. ENGLER

ABBREVIATIONS

aa, amino acid; Ad2 and Ad5, adenovirus serotypes 2 and 5, respectively; BCIP, 5-bromo-4-chloro-3-indolyl-phosphate; kDa, kilodaltons; mAB, monoclonal antibody; NBT, 4-nitro blue tetrazolium chloride; SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis; Nucleotides and amino acids have been numbered starting with the first nucleotide (A) and the first amino acid (M) in the open reading frame that encodes Ad5 fiber.

ABSTRACT

The nucleotide sequence of the fiber gene of H5ts142 (2), a temperature-sensitive fiber defective mutant of Ad5, has been determined. Three nucleotide changes were found within this mutant gene; one of these substitutions resulted in an amino acid change from alanine to valine at amino acid 440. The temperature-sensitive phenotype of H5ts142 fiber results from this single mutation, since the H5ts142 fiber protein shows the same phenotype when expressed independently from other adenovirus proteins using a vaccinia virus vector. When the same mutation was introduced into the Ad2 fiber gene, it did not show the same temperature-sensitive phenotype. Deletion mutants from the C-terminus of Ad2 fiber have also been constructed; indirect immunofluorescence experiments show that these mutant fibers were correctly localized in the nucleus but that they were not able to form a trimer.

INTRODUCTION

Adenovirus is a DNA virus whose capsids are assembled in the cell nucleus. The adenovirus capsid consists of three major proteins: hexon, penton base, and fiber. These components are synthesized in the cytoplasm and are transported to the nucleus for assembly into virus particles. At least two different systems (virus-dependent and virus-

independent) exist for the assembly and the transport of structural proteins during adenovirus infection. A major late protein of adenovirus, 100K, plays a role in the assembly of hexon trimers. The 100K protein plays a major role in hexon assembly, by forming a complex with monomeric hexon molecules that results in assembly and release of hexon trimers. Cepko and Sharp (1) studied a series of temperature sensitive mutants of adenovirus type 5 deficient in hexon assembly and discovered two kinds of mutants: (1) 100K gene mutants, which fail to produce hexon trimers at the nonpermissive temperature (NPT), and (2) transport mutants within the hexon gene, which produce hexon trimers but fail to transport these trimers into the nucleus. A second pathway for assembly can be seen for adenovirus fiber. Trimer formation from Ad2 fiber monomers does not depend on other adenovirus proteins, since trimers form when fiber is expressed in a vaccinia virus expression system (9). Trimerization of fiber is apparently independent from nuclear translocation, since fiber mutants with N-terminal deletions fail to accumulate in the nucleus but form trimers in the cytoplasm (9).

H5ts142 is a temperature-sensitive mutant of Ad5 characterized by Chee-Sheung and Ginsberg (2). It lacks an O-GlcNAc modification when synthesized at the NPT. At this temperature, this mutant produces empty capsids rather than infectious virus particles. Fibers from H5ts142 are unable to form trimers required for assembly at the NPT, as judged by sedimentation in sucrose gradients (2). Fibers made at the NPT are still transported to the nucleus (see the second manuscript in this thesis).

In this study, the fiber gene of H5ts142 has been isolated and its nucleotide sequence was determined. Three nucleotide changes have been found, resulting in only one amino acid substitution. H5ts142 fiber was expressed independently from other adenovirus proteins using a

recombinant vaccinia virus (4); the recombinant H5ts142 fiber did not form trimers at the NPT, suggesting that this single amino acid change was sufficient to disrupt trimer formation. Surprisingly, when the same mutation was introduced into the Ad2 fiber gene, it did not have the same effect on trimerization of Ad2 fiber protein. Deletions from the C-terminus of Ad2 fiber protein were also tested and shown to be unable to form trimers, although these mutants continued to accumulate in the nucleus.

MATERIALS AND METHODS

Enzymes and chemicals. Restriction endonucleases, T4 DNA ligase, T4 DNA polymerase, calf intestine alkaline phosphatase, and Klenow enzyme were purchased from Boehringer Mannheim Biochemicals (Boehringer-Mannheim, Indianapolis, IN); Sequenase^{TM} for DNA sequencing was purchased from United States Biochemicals (Cleveland, Ohio). MPA (10mg/ml; from Sigma Chemical, St Louis, MO) and xanthine (10mg/ml; from Sigma) were dissolved in 0.1 N NaOH, and hypoxanthine (10mg/ml; from Sigma) was dissolved in water and sterile filtered;the solutions were stored frozen. Stock solutions of BUdR (2.5 mg/ml: from Sigma) were made in water, sterile filtered, and kept at 4°C. Stock solutions of X-gal (20mg/ml or 100mg/ml; from Boehringer-Mannheim) were made in dimethylformamide and stored at -20°C.

Cells and Viruses. Ad5 H5ts142 (a gift from Dr. H. Ginsberg) was grown in HeLa cells at 32°C. HeLa cells and HEp2 cells were maintained in DMEM (GIBCO), 10% bovine calf serum (Hyclone Laboratories), and gentamycin (50 μg/ml final concentration, Sigma Chemical, St.Louis, MO). CV1 cells were maintained in DMEM with 10% fetal bovine serum. TK$^-$ cells were maintained in DMEM with 10% bovine calf serum and 25μg/ml 5-Bromodeoxyuridine.

Isolation of adenovirus DNA from H5ts142 and Ad5. Virus DNA for cloning of the H5ts142 fiber gene was isolated from infected cells by a modified Hirt lysate method (8, 10). 500ml of HeLa cells in Spinner culture was infected with H5ts142 at 32°C (moi approx. 50). At 40 hours postinfection, the cells were harvested by centrifugation and resuspended in 5 ml of ice cold TE (10 mM Tris, 1 mM EDTA (pH 7.5). SDS was added to a final concentration of 0.2% and the sample was incubated at 37°C for 15 min with 500 μg of Proteinase K (Boehringer-Mannheim). The volume was adjusted to 9 mls with 10 mM Tris, 1 mM EDTA (pH 7.5) containing 1% SDS and the sample was incubated an additional 5 min at 37°C. NaCl was added to 1M (final concentration) and the sample was incubated at 4°C for approximately 16 hours. The supernatant containing the DNA was recovered by centrifugation and DNA was precipitated by ethanol precipitation. The DNA was cleaned by phenol/chloroform extraction and ethanol precipitation. DNA from Ad5 virus was isolated from CsCl-banded Ad5 virus by standard procedures (14).

Cloning of the fiber gene from adenovirus genome (Ad5 and H5ts142). The H5ts142 DNA from the Hirt extract was digested with HindIII. Since the HindIII-B (approx. 5.9kb) and the HindIII-F (approx. 2.9kb) DNA fragments of Ad5 contain the fiber gene (5), DNA fractions near those sizes were isolated by electroelution onto DEAE membrane (Schleicher & Schuell). The HindIII-digested DNA was cloned into HindIII-digested pUC119 DNA (Vieira and Messing) and screened by colony hybridization (Maniatis) with ^{32}P-labeled oligonucleotides. Plasmid DNA from colonies giving a positive signal was isolated and the entire fiber gene was sequenced by double-strand DNA sequencing using 10 oligonucleotides which anneal throughout the fiber gene (listed in Table I).

The Ad5 DNA was digested with KpnI and the KpnI-D fragment (approx. 4.5kb) containing the fiber gene (5) was cloned into the KpnI site of pUC119.

Cloning of H5ts142 fiber gene. HindIII-F fragment cloned in pUC119 was digested with SmaI and self-ligated, resulting in a construct that contained the 5' half of the fiber gene. This plasmid was then digested with HindIII and the HindIII-B fragment was ligated into this DNA to obtain an intact H5ts142 fiber gene.

Mutagenesis on Ad2 fiber gene. Site-directed mutagenesis was performed to insert KpnI restriction sites in front of the fiber genes from Ad5 and from H5ts142 for cloning into the vaccinia expression vector. The sequences of oligonucleotides used for this mutagenesis are listed in Table 1. Mutagenesis on Ad2 fiber gene was performed to change codon GCT (alanine) to GTT (valine) at amino acid 440 in the Ad2 fiber gene. The sequence of the oligonucleotide used for mutagenesis is 5'ATCTC CAGAT ACAAC CAAAG CAGCT A 3'(26-mer). The mutation (H2ts142) was confirmed by plasmid sequencing using Sequenase™ (13).

Construction of deletion mutants of Ad2 fiber. 138 amino acids were deleted from the C-terminal end of Ad2 fiber (2F-dl444) by cloning the BamHI (in the vector) and BglII (in the fiber gene) fragment of Ad2 fiber gene from pUC2Fm into BamHI site of pTKgpt-F3s (Figure 5B). Since the termination codon from the fiber gene has been deleted, the termination codon in the vector was utilized and this resulted in adding 5 more codons (PLKLT) onto the C-terminal end of the fiber. Another deletion mutant was made by creating a termination codon at the SpeI site of Ad2 fiber gene. It was done by cutting the S3-2F plasmid (9) with SpeI, repairing the overhangs with Klenow enzyme, and religating the ends together. This clone has resulted in Ad2 fiber with 42 amino acids deletion at the C-

terminal end (2F-dl540). The recombinant vaccinia viruses expressing these mutant fiber genes were prepared as described above.

Cloning and expression of the fiber genes. The fiber genes of Ad5, H5ts142, and H2ts142 were cloned into the vaccinia virus expression vector as described before (8). Recombinant vaccinia viruses expressing the fiber polypeptides were prepared as described previously (9).

Immunoblot. Cell lysates were made from recombinant vaccinia virus infected cells by freeze-thawing the cells 3 times, sonicating for 2 min, and extracting with freon. Fifteen μg of each protein sample was loaded in sample buffer (12) and then separated by SDS-PAGE on a 10% gel; the proteins in the gel were then blotted onto nitrocellulose and incubated in Blotto (5% nonfat dry milk in PBS) to block nonspecific antibody binding. Either anti-fiber (4D2-5) or anti-trimer (2A6-36) antibody was used as primary antibody (1:1000 dilution in Blotto). Biotin-conjugated goat anti-mouse Ig antibody (Fisher Biotech) and alkaline phosphatase conjugated to streptavidin (Fisher Biotech or Jackson ImmunoResearch, West Grove, PA) were used to probe the specific binding of monoclonal antibodies to fiber; the color reaction for alkaline phosphatase was developed with NBT and BCIP (Bethesda Research Labs).

Immunofluorescence. HeLa T4 or Hep2 cells were plated on a glass coverslip in 35mm plate. At 48 hr postinfection, cells were fixed with 3% paraformaldehyde in PBS (phosphate buffered saline) either at 32°C (for cells grown at the permissive temperature) or 39.5°C (for cells grown at the nonpermissive temperature). TBS-TX (Tris buffered saline-1% Triton X-100) was used to permeabilize the cell membrane, prior to incubation with anti-fiber antibody (4D2-5) or anti-trimer antibody (2A6-36) at room temperature for 1 hour. Goat anti-mouse Ig conjugated to fluorescein isothiocyanate (Jackson Immunoresearch) was used as

secondary antibody. The cells were photographed on a Nikon Optiphot microscope equipped for fluorescent illumination at a magnification of 20x or 40x using Kodak Gold 400 film.

RESULTS

Mutations in the fiber gene from H5ts142. A modified Hirt extraction procedure was used to obtain the viral genome from H5ts142 ; unlike wild-type adenoviruses, H5ts142 virus did not form distinct virus band during CsCl gradient centrifugation and it was not possible to obtain pure virus preparation by this procedure for extraction of virus DNA. The HeLa cells infected with H5ts142 were lysed at 40 hours postinfection, so that the viral DNA still remains in the nucleus rather than in the virus particles. In the restriction endonuclease cleavage map of Ad5 (5), two HindIII digested fragments share the fiber gene: the HindIII-B fragment (approx. 5.9 kb) which encodes the 5' half of the gene and HindIII-F fragment (approx. 2.9kb) which contains the 3' half of the gene (Fig. 1). Each of these two fragments were cloned from the H5ts142 DNA pool into the HindIII site of pUC119 and screened with oligonucleotides for the correct insert; four isolates of each hybridization-positive clone were picked and sequenced by using synthetic oligonucleotides (Table 1) which anneal to the antisense strand throughout the Ad5 fiber gene. Three nucleotide changes were found in the fiber gene of H5ts142 (Fig. 2), when compared to that of Ad5 (3): G to A at nucleotide 78, C to T at nucleotide 1086, and C to T at nucleotide 1319. Only the nucleotide change at nucleotide 1319 resulted in an amino acid change: from alanine (GCT) to valine (GTT) at aa440 (see Fig. 2). This third change was sequenced on both strands of the DNA in each of the four isolated clones to insure the accuracy of this result.

Expression and characterization of the fiber of H5ts142 and its equivalent mutation in Ad2 fiber. KpnI restriction endonuclease sites

Figure 1. Restriction fragments containing the fiber genes of Ad5 and H5ts142. The boxes indicate the fiber genes. The fiber gene of H5ts142 was obtained by putting two HindIII digested fragments (HindIII-B and HindIII-F) together in pUC119 plasmid. The fiber gene of Ad5 was obtained by cloning KpnI-D fragment into pUC119 plasmid. The arrows indicate the location of KpnI restriction site insertions for cloning of the fiber genes into expression vector, pTKgpt-F3s.

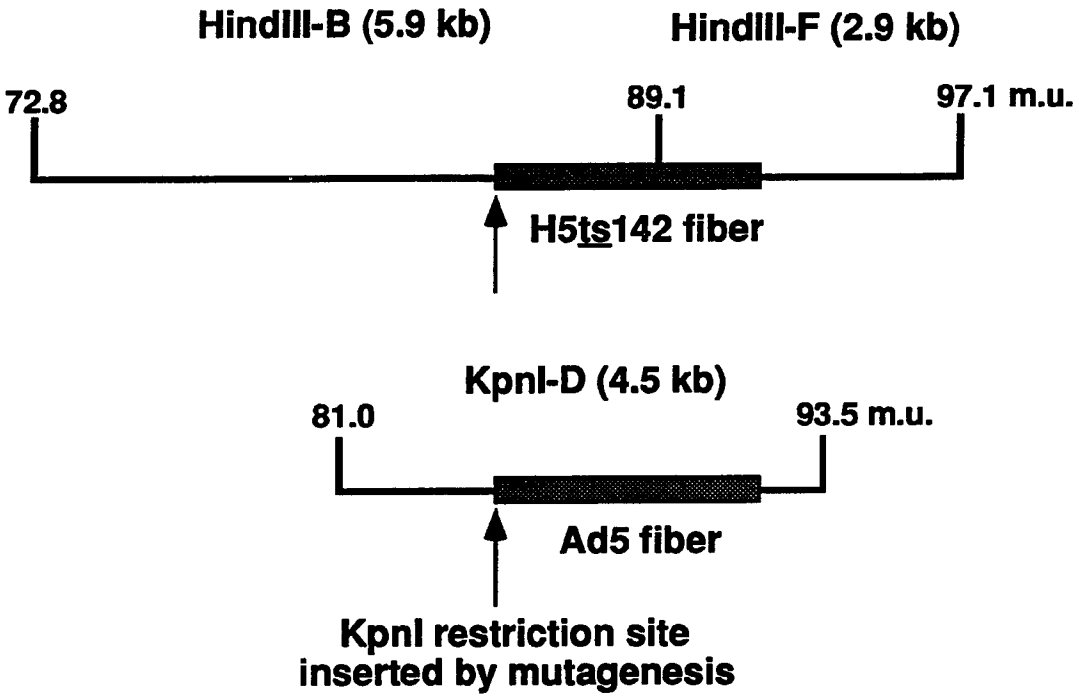

HindIII-B (5.9 kb)
HindIII-F (2.9 kb)
72.8
89.1
97.1 m.u.
H5ts142 fiber
KpnI-D (4.5 kb)
81.0
93.5 m.u.
Ad5 fiber
KpnI restriction site
inserted by mutagenesis

Table 1 List of the oligonucleotides used for sequencing.

A. Oligonucleotides used for sequencing of fiber gene from H5ts142; the oligonucleotides anneal to the antisense strand DNA of fiber gene except 2571 which anneals to the sense strand.

2007; 5' TTCCT CCTGT TCCTG 3'
2008; 5' TCCTC CCTTT GTATC 3'
2009; 5' TGTAA CCACT GTGAG 3'
2010; 5' AAACT TAGCA TTGCC 3'
2011; 5' TTGCA TGTAA CAGAC 3'
2012; 5' TGCTC AAAAC CAACT 3'
2013; 5' GCACC AAACA CAAAT C 3'
2014; 5' GCTAA ACTCA CTTTG 3'
2015; 5' AGATC TTACT GAAGG 3'
2016; 5' AGACA CAACT CCAAG 3'
2571; 5' GCACT CCATT TTCGT 3'

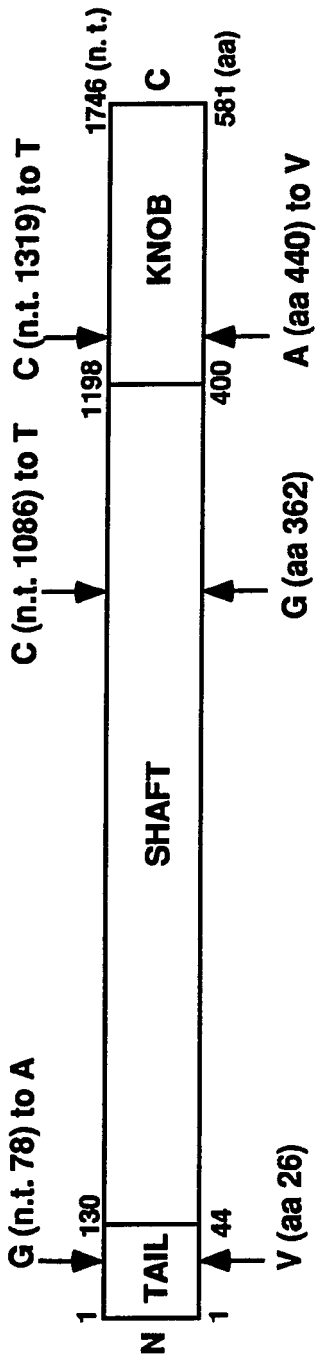
B. Oligonucleotides used for mutagenesis.

2572; A mutagenic oligonucleotide for insertion of KpnI site in front of fiber genes from Ad5 and H5ts142;
5' TTGCG CGCTT CATGG TACCT GCAAC AA 3'

2570; A mutagenic oligonucleotide for alanine to valine change at aa 440 in Ad2 fiber gene;
5' ATCTC CAGAT ACAAC CAAAG CAGCT A 3'

2569; A sequencing primer for 2570 mutation;
5' TGGTC AAATC TAAGG 3'

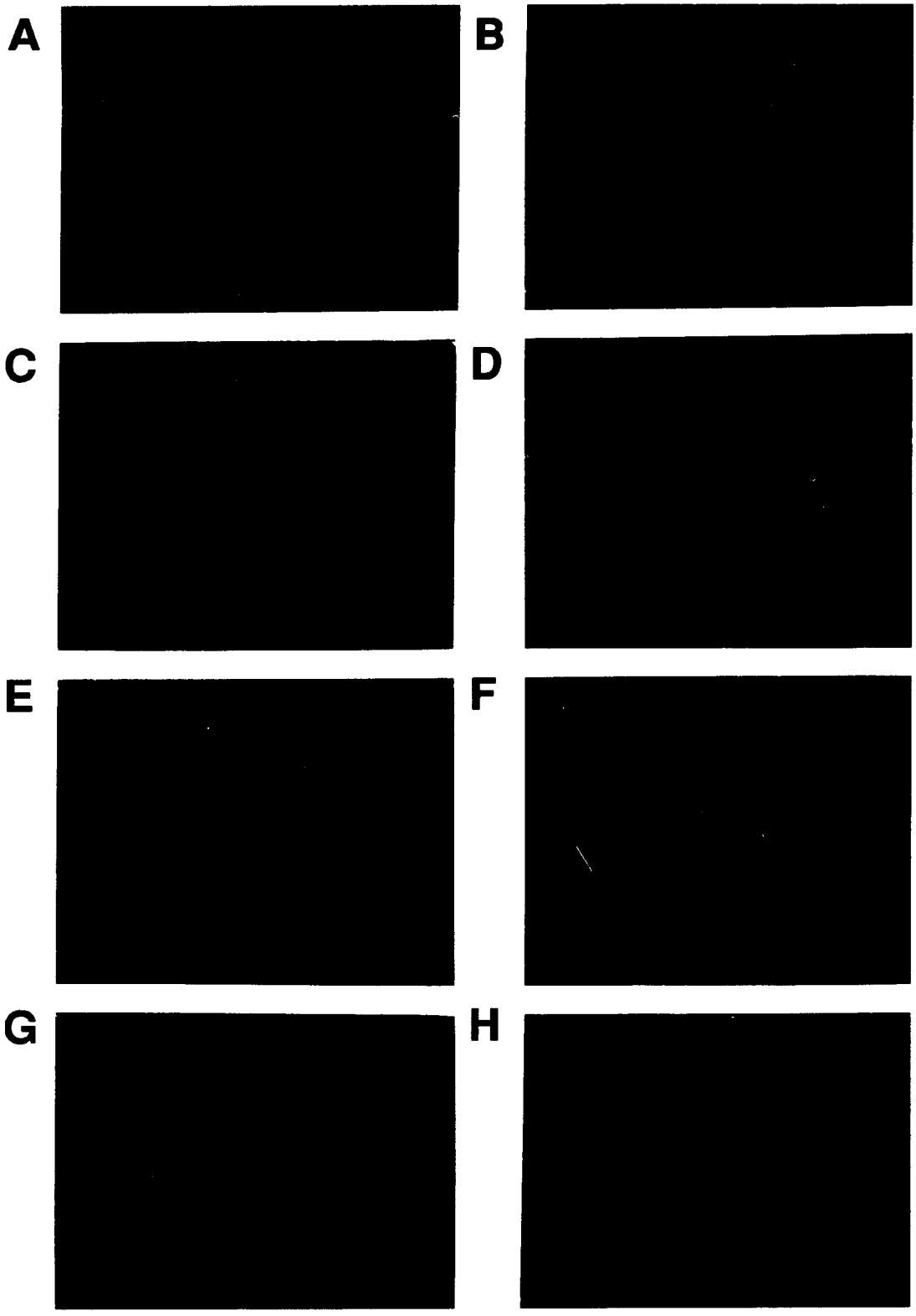
Figure 2. The location of mutations in fiber gene of H5ts142. The numbers on top indicate nucleotide numbers starting from 1 at the A of ATG initiation codon; the numbers at the bottom indicate amino acid numbers starting from 1 at the initiation codon (the first methionine). The secondary structure of the fiber (a tail, a shaft, and a knob) is described by Green et al. (6) Only one of the mutations resulted in an amino acid change: A to V at a.a. 440.


G (n.t. 78) to A
C (n.t. 1086) to T
C (n.t. 1319) to T
1746 (n. t.)
C
581 (aa)
A (aa 440) to V
G (aa 362)
V (aa 26)
1
130
44
1
1198
400
N
TAIL
SHAFT
KNOB

were inserted in front of the fiber genes from both H5ts142 and Ad5 to clone the genes in the vaccinia virus vector pTKgpt-F3s (4). The sequences of the mutagenic oligonucleotides are listed in Table 1. KpnI was chosen because it can be used to provide part of the optimal sequence for translation initiation in eukaryotic cells, (GGT)ACCATG (11). By expressing the H5ts142 fiber independently from other polypeptides found during adenovirus infection, whether the behavior of ts142 fiber was due to the single amino acid change in the fiber itself or to an interaction of the mutant fiber with other adenovirus proteins could be tested. Two monoclonal antibodies were used to detect fiber monomers and trimers: 4D2-5 recognizes both monomer and trimer form of fibers from Ad2 and Ad5, while 2A6-36 recognizes only trimer form of fiber from Ad2 and Ad5 (see the second manuscript in this thesis). Both recombinant fibers of S3-5F (recombinant Ad5 fiber) and S3-5ts142F (recombinant H5ts142 fiber) were translocated into the nucleus at either 32°C (permissive temperature) or 39.5°C (nonpermissive temperature) as shown by immunofluorescence with 4D2-5 mAb (Fig. 3A and 3B). Both S3-5F and S3-5ts142F recombinant fibers form trimers at 32°C as shown by immunofluorescence with 2A6-36 mAb (Fig. 3C and 3D). As expected, 2A6-36 mAb recognized only S3-5F, but not S3-5ts142F at 39.5°C by immunofluorescence (Fig. 3E and 3F). This data correlates well with the results from the adenovirus infected cells using these same antibodies. Therefore, a single amino acid change from alanine to valine at aa 440 in H5ts142 apparently results in temperature-sensitive phenotype of recombinant ts142 in trimerization and glycosylation of the fiber.

An alanine residue is also found at aa 440 in Ad2 fiber (3). When the same alanine to valine change at aa 440 was constructed in Ad2 fiber

Figure 3. Immunofluorescence of recombinant Ad5 wild-type fiber (S3-5F) and of recombinant H5ts142 fiber (S3-5ts142). Immunofluorescence was performed on HeLa T4 cells infected with recombinant vaccinia viruses; A, B, C, and D are the cells infected with S3-5F and E, F, G, and H are the cells infected with S3-5ts142. The antibodies used are 4D2-5 (anti-fiber) for A, C, E, and G; 2A6-36 (anti-trimer) for B, D, F, and H. Infection was proceeded at 32°C for A, B, E, and F and at 39.5°C for C, D, G, and H.


A
B
C
D
E
F
G
H

(S3-2ts142F), the fiber was able to form trimers at both 32°C and 39.5°C as shown by immunofluorescence with 4D2-5 and 2A6-36 mAbs (Fig. 4).

Deletions from the C-terminus of Ad2 fiber do not form trimers.

From previous studies to map the nuclear targeting signal for Ad2 fiber, deletions from the N-terminus of the protein did not affect the ability of fiber to form trimers (see the third manuscript in this thesis). In order to map domains important for trimer formation in fiber, two deletion mutants were constructed and expressed in vaccinia virus recombinants. Mutant 2F-dl444 is missing the last 138 amino acids of fiber and has an additional 5 amino acids, PLKLT, at the C-terminus since the termination codon located downstream from the cloning site in the vector was used (Fig. 5B). A second deletion mutant, 2F-dl540, has 42 amino acids removed from the C-terminus (Fig. 5C).

Indirect immunofluorescence experiments were performed on cells infected with vaccinia recombinants expressing dl444 (virus 2F-dl444) and dl540 (virus 2F-dl540). When mAb 4D2-5 was used to visualize fiber in the cells expressing 2F-dl444 (Fig. 6A) or 2F-dl540 (Fig. 6C), the mutant proteins were found in the nucleus. When mAb 2A6-36 was used to visualize these mutants, neither mutant showed detectable fluorescence (Fig. 6B and 6D). There are two possibilities to explain this result: (1) these mutants are no longer able to form trimers, or (2) these mutants can still form trimers, but these trimers have a different conformation from wild-type fiber and are not recognized by mAb 2A6-36.

In order to discriminate between these two possibilities, the mutants were tested for ability to form trimers on a Western blot with mAb 4D2-5 (Fig. 7). This blot contained both boiled (to show protein expression) and unboiled cell lysates (to detect trimers); electrophoresis of unboiled fiber trimers on an SDS-polyacrylamide gel is not sufficient to disrupt the

Figure 4. Immunofluorescence of the recombinant H2ts142 fiber (S3-2ts142). The fiber of H2ts142 contains a mutation at aa 440 from A to V. Immunofluorescence was performed on HeLa T4 cells infected with S3-2ts142; A and B were visualized with mAb 4D2-5 and C and D were with mAb 2A6-36. Infection was proceeded at 32°C for A and C and 39.5°C for B and D. The fiber of S3-2ts142 still forms trimer even with the same mutation as in S3-5ts142 fiber.

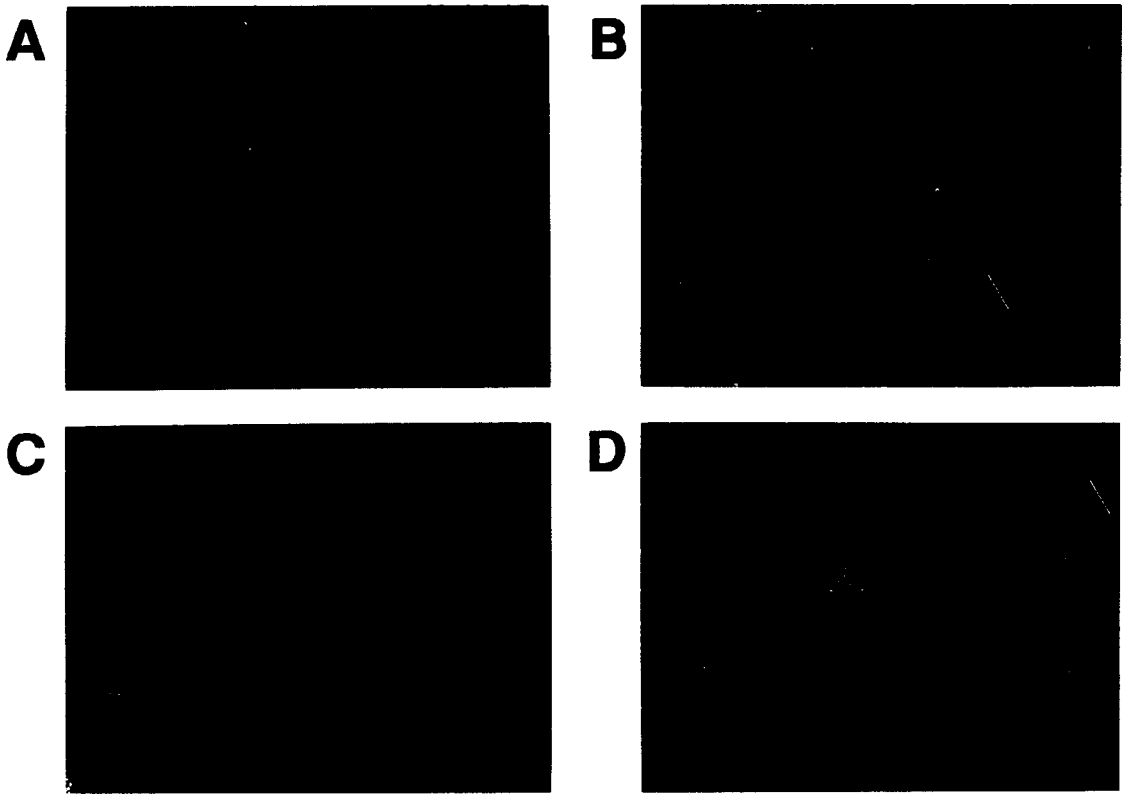
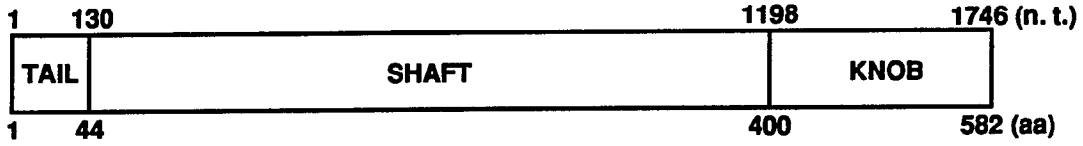

A
B
C
D

Figure 5. Construction of deletion mutants of Ad2 fiber. A (S3-2F) is wild type Ad2 fiber. B (2F-dl444) is Ad2 fiber with 138 amino acid deletion at the C-terminus; it contains 5 foreign amino acids at the C-terminus. C (2F-dl540) is Ad2 fiber with 42 amino acid deletion at the C-terminus. The nucleotide numbers and the amino acid numbers are shown above and below the box, respectively.

A. S3-2F



B. S3-2F-dl444

C. S3-2F-dl540

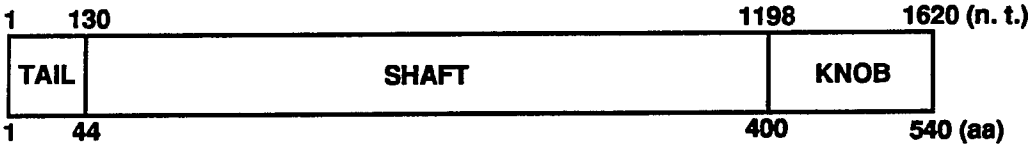


Figure 6. Immunofluorescence on C-terminal deletion mutants of Ad2 fiber. A and B are the HeLa T4 cells infected with 2F-dl444 and C and D are the HeLa T4 cells infected with 2F-dl540. A and C were probed with mAb 4D2-5 (anti-fiber); B and D were probed with mAb 2A6-36 (anti-trimer).

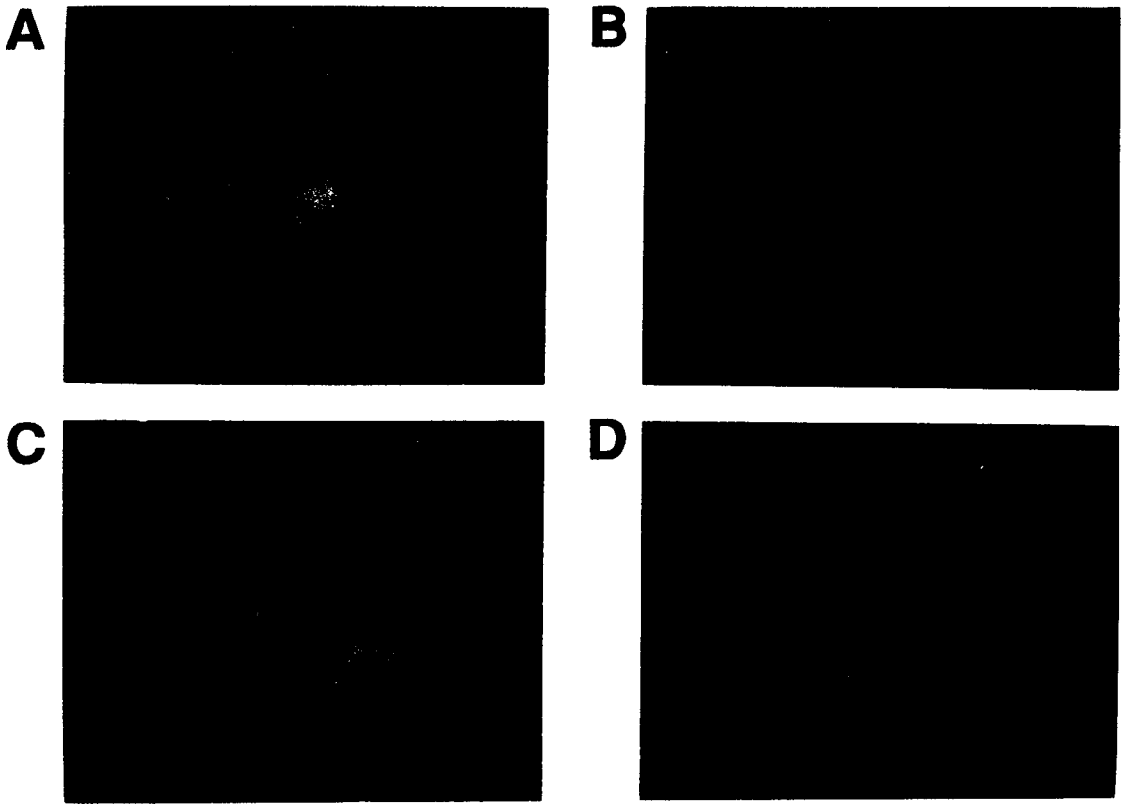
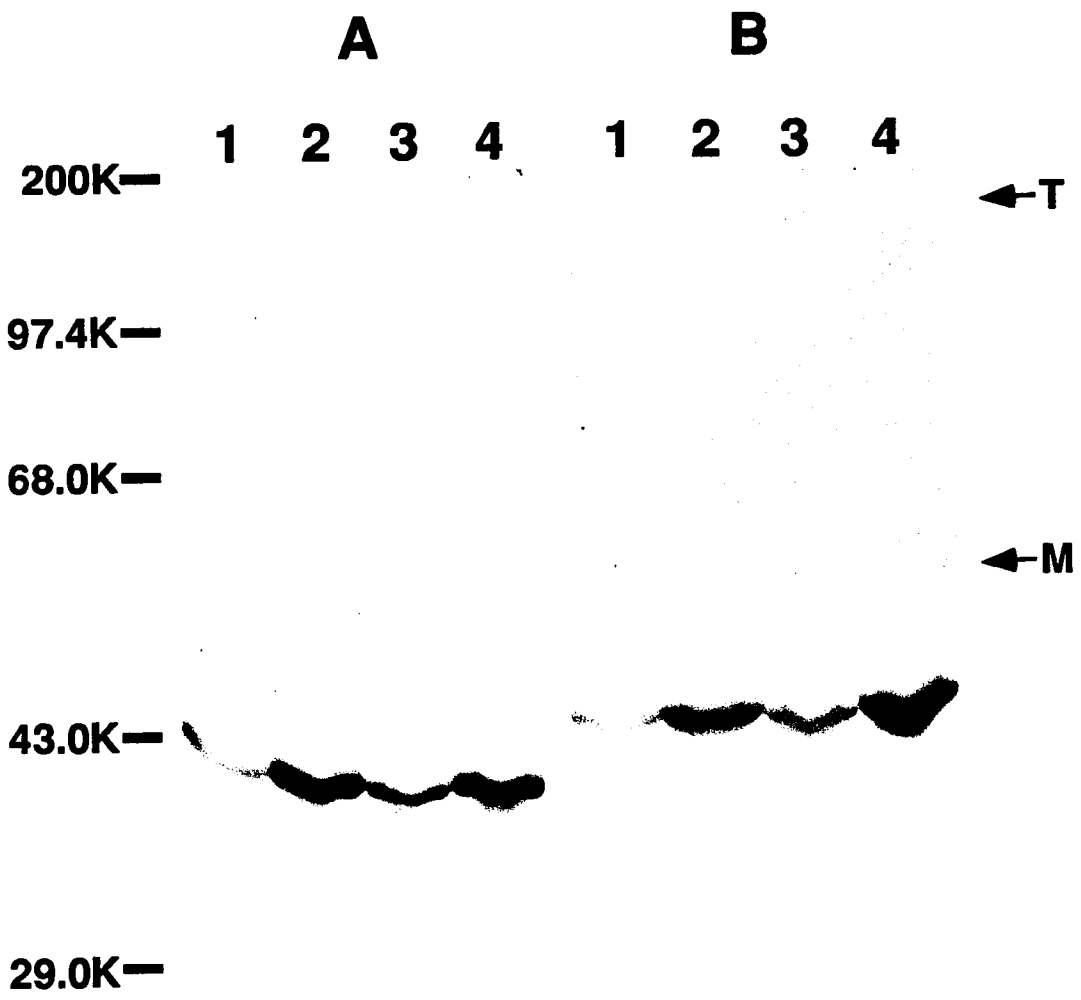

A
B
C
D

Figure 7. Immunoblot on C-terminal deletion mutants infected cell lysates. Samples were either boiled or unboiled to detect the trimer formation of fiber. The blot was developed with the 4D2-5 antibody. Panel A is the HEp2 cell lysates infected with 2F-dl444; panel B is the HEp2 cell lysates infected with 2F-dl540. Lanes 1 and 3: infected at 32°C; lanes 2 and 4: infected at 39.5°C. Lanes 1 and 2 are boiled lysates and lanes 3 and 4 are unboiled lysates. When a duplicate blot was developed with the 2A6-36 antibody, no specific band was detected (data not shown).


A
B
1 2 3 4
1 2 3 4
200K
97.4K
68.0K
43.0K
29.0K
T
M

complex (see the second manuscript in this thesis). Since the fiber band appeared only at monomer size on both denatured and nondenatured samples (which could be recognized only with the 4D2-5 mAb, but not with the 2A6-36 mAb), both deletions completely block the trimer formation of the fiber.

DISCUSSION

This manuscript presents evidence that domains important for trimer formation of Ad2 and Ad5 fiber reside in the C-terminus of the protein. The results that lead to this conclusion can be summarized as follows:

1. The mutation that is responsible for the trimer-deficient phenotype observed in the Ad5 H5ts142 fiber lies at amino acid 440 in the protein.
2. Deletions of 138 and 42 amino acids from the C-terminus of fiber are both unable to form trimers at 32° or at 39°C.
3. In another manuscript (Hong and Engler, submitted), deletions at the N-terminus of fiber or substitutions at amino acid 90-94 were used to map the nuclear localization signal of fiber. Both of these mutants are still able to form trimers; further, mutants blocked in nuclear transport accumulate trimers in the cytoplasm, suggesting that trimer formation and nuclear transport are independent processes.

Taken together, these results suggest that trimer formation is dependent on one or more domains within the C-terminus of fiber. Little is known about the actual pathway for assembly of trimers, although one might speculate on the basis of these experiments that trimerization starts in the C-terminus. This process apparently proceeds independently of other adenovirus proteins (although not necessarily other cellular proteins), since vaccinia recombinants expressing fiber are still capable of forming trimers; if another adenovirus protein were required to assemble

trimers, then one would have to postulate that vaccinia virus itself could also provide a protein with a similar function to the hypothetical adenovirus molecular chaperone. The self-assembly of fiber trimers is apparently different from assembly of hexon trimers which also requires the action of the adenovirus 100K protein (1). These observations are substantially in agreement with similar studies of C-terminal fiber deletions expressed in baculoviruses (P. Boulanger, personal communication).

When the Ad2 H2ts142 fiber was expressed using vaccinia virus, it did not show the temperature sensitive phenotype for trimer formation as had been observed with the Ad5 H5ts142 fiber. This result was quite unexpected, since the alanine at this position is conserved in both Ad2 and Ad5 fiber (3). One potential explanation for this discrepancy is that the structures of the two C-terminal domains is extremely variable and that domains important for trimerization of Ad5 fiber will be completely different from the domains found for Ad2. A second possibility is that the sequences around this alanine are somewhat different between the two serotypes: perhaps nearby amino acid residues in Ad2 fiber (AALAVS) can compensate for the alteration in structure in ways that the Ad5 fiber (SVLAVK) cannot.

A third possibility is that posttranslational modifications of Ad2 fiber may help to stabilize fiber trimers. Previously, it has been shown that both Ad2 and Ad5 fibers contain 3-4 moles and 2 moles of O-GlcNAc (14). Although the specific sites for attachment of these sugars to these fibers are not known, the fiber sequences of Ad2 and Ad5 contain 5 and 2 potential attachment sites (respectively), based on a proposed consensus sequence for attachment of O-GlcNAc (7): Ad2 and Ad5 share 2 potential sites in the N-terminus of the protein, while Ad2 contains additional potential sites in

the C-terminus. It has been proposed elsewhere that the shared GlcNAc sites in the N-terminus of fiber may help to stabilize trimers once formed (14); if this hypothesis is correct, it may also be possible that additional O-GlcNAcs in the C-terminus of (only) Ad2 fiber may stabilize the trimer in ways not available to Ad5 fiber (which lacks apparent C-terminal O-GlcNAc sites). Experiments to test this hypothesis by mapping the sites of O-GlcNAc attachment and by removing each of the predicted attachment sites are currently underway.

ACKNOWLEDGEMENTS

We thank Dr. Harold Ginsberg for the gift of Ad5 H5ts142 virus. This work was supported by NIH grant GM38234 (to J.A.E.). Preparation of monoclonal antibodies was done by the Hybridoma Core Facility of the Multipurpose Arthritis Center (grant NIH 5P60 AR20614).

REFERENCES

1. Cepko, C. L. and Sharp, P. A. 1983. Analysis of Ad5 hexon and 100K temperature sensitive mutants using conformation-specific monoclonal antibodies. Virol. 129:137-154.

2. Chee-Sheung, C. C. and Ginsberg, H. S. 1982. Characterization of a temperature-sensitive fiber mutant of type 5 adenovirus and effect of the mutation on virion assembly. J. Virol. 42:932-950.

3. Chroboczek, J. and Jacrot, B. 1987. The sequence of adenovirus fiber: similarities and differences between serotypes 2 and 5. Virol. 161:549-554.

4. Falkner, F. G. and Moss, B. 1988. Escherichia coli gpt gene provides dominant selection for vaccinia virus open reading frame expression vectors. J. Virol. 62:1849-1854.

5. Flint, S. J. and Broker, T. R. 1980. "The Adenoviruses." In DNA tumor viruses. J. Tooze ed. CSH Laboratory, Cold Spring Harbor, New York.

6. Green, N. M., Wrigley, N. G., Russell, W. C., Martin, S. R. and McLachlan, A. D. 1983. Evidence for a repeating cross-ß sheet structure in the adenovirus fibre. EMBO J. 2:1357-1365.

7. Haltiwanger, R. S., Holt, G. D. and Hart, G. W. 1990. Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. J. Biol. Chem. 265:2563-2568.

8. Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. 26:365-369.

9. Hong, J. S. and Engler, J. A. The amino terminus of the adenovirus fiber protein encodes the nuclear localization signal. Submitted.

10. Horwitz, M. S. 1976. Bidirectional replication of adenovirus type 2 DNA. J. Virol. 18:307-315.

11. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell. 44:283-292.

12. Laemmli, E. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227:680-685.

13. Mierendorf, R. C. and Pfeffer, D. 1987. Direct sequencing of denatured plasmid DNA. Meth. in Enzym. 152:556-562.

14. Mullis, K. G., Haltiwanger, R. S., Hart, G. W., Marchase, R. B. and Engler, J. A. 1990. Relative accessibility of N-acetylglucosamine in trimers of the adenovirus types 2 and 5 fiber proteins. J. Virol. 64:5317-5323.

SUMMARY

The fact that viruses use established cellular pathways in their life cycle makes them an excellent probe with which to examine a variety of intracellular processes. The study of adenovirus fiber as a nuclear trimeric glycoprotein offers a unique opportunity to study the processing and nuclear transport of a nonmembrane multimeric glycoprotein. Knowledge gained from the study of fiber may well provide insights into the interrelationship between protein structure, posttranslational modification, and intracellular protein transport.

Fiber provides several advantages to study translocation and assembly of a multimeric nuclear protein: i) the adenovirus fiber protein is made in large excess in virus infected cells, ii) adenovirus must have a very efficient transport mechanism to translocate the structural proteins into the nucleus to be assembled into virus particles, iii) fiber genes from several adenovirus serotypes have been cloned and subjected to extensive sequence analysis, iv) fiber forms a very stable trimer, perhaps due to a novel form of glycosylation, and v) an expression system for the production of recombinant fiber is available. Vaccinia virus expression system has allowed the investigation of fiber independently from other adenovirus proteins, so that the relationships among the viral proteins could be studied. This expression system also made it possible to express mutants of fiber without constructing adenovirus mutants. Since fiber is an essential protein for virus assembly and infectivity, lethal mutants in this gene cannot be studied by replacement in the adenovirus genome.

Characterization of these mutants provided indispensible information about the requirements for nuclear transport and the relationship between assembly and glycosylation of adenovirus fiber protein.

The first manuscript presented the nucleotide sequence and predicted the amino acid sequences for open reading frames (ORF) encoded in the BamHI-D fragment of Ad7, including a part of the E3 region, the entire L5 region which encodes the fiber gene, and a portion of early region 4. This study made it possible to compare the derived amino acid sequences not only with serotypes in the same subgroup (Ad3, subgroup B) but also with serotypes in different subgroups (Ad2, subgroup C).

The exon encoding the fiber polypeptide (L5) of Ad7 was identified by S1 nuclease mapping. The acceptor splice site for Ad7 fiber gene, the AG dinucleotide, is located immediately upstream of the AUG initiation codon as in Ad3 (35). The 3'-end mapping indicated that the poly(A) addition site is located 23 or 24 bases downstream from the AATAAA sequence of the Ad7 fiber gene. Ad7 fiber shares 64% nucleotide homology and 60% amino acid homology with Ad3 fiber. However, it shares only 43% nucleotide homology and 20% amino acid homology with Ad2 fiber. The degree of homology observed within the fiber gene is lower than in many other regions in the genome; for example, the preterminal protein and the Ad Pol genes of Ad5 and Ad7 share 80% DNA sequence homology. This result matches well with the fact that the fiber protein contains type-specific antigenic sites (12, 13). It was first thought that type-specific antigenic determinants might be identified by locating regions of non-homology. Unfortunately, the nonhomologous regions were more numerous than had been expected and were spread throughout the entire sequence. However, the amino-terminal portion is generally more conserved than other parts of the protein, probably due to the fact that this part of the fiber protein plays a

role in joining the fiber protein to penton base (10). Three short regions of partial homology between Ad2, Ad3, Ad5, and Ad7 are identified: (T/S)FNPVYPY(D/E)-E(S/T), PF(L/I)-P-F(V/I)SP(N/D)GF(T/Q)(Q/E)SP-GVL(S/T)L----PL-T--G-L-LK-G-GL-(T/S)(L/V/I)D, and (Y/F)TF(S/Y)YI(A/R)(E/D/Q)(D). These sequences may be important for fiber binding to penton base, assembly of fiber trimer, or interaction with the cellular receptor for adenovirus binding. Identification of the significance of these conserved regions will be the focus of future studies.

Antibodies specific for the adenovirus fiber were an essential component throughout the course of this study. The detection of intracellular localization and trimer formation was made possible by using two special antibodies: monoclonal antibodies (mAb) 4D2-5 and 2A6-36. 4D2-5 (anti-fiber) Ab recognizes both monomer and trimer forms of fiber from Ad2, Ad5, and Ad7, while 2A6-36 recognizes only trimer form of fiber from Ad2 and Ad5, but not from Ad7. Development of these two special antibodies played a key role in furthering this study.

Ishbashi and Maizel (19) previously reported that Ad2 fiber contained an O-linked amino sugar derivative which could be labeled with radioactive glucosamine. The glycosyl moieties of Ad2 and Ad5 fibers have been further characterized (3, 29) to be monosaccharide O-linked glucosamine residues. Many of the glycoproteins bearing this novel form of glycosylation have been reported (2, 16, 20, 23) and found to be localized predominantly in the nucleus and cytoplasm, instead of the ER and the Golgi.

One of the suggested functions for this modification is participation in targeting of proteins to the nucleus (20, 34). The availability of a temperature-sensitive glycosylation-defective mutant of Ad5, H5ts142, made fiber an excellent tool for testing this hypothesis; the fiber of H5ts142

does not contain O-GlcNAc at the nonpermissive temperature (39.5°C). The data presented in the second manuscript demonstrates that O-linked GlcNAc is not required for the correct targeting of adenovirus fiber protein to the nucleus, since the glycosylation defect of H5ts142 does not affect its eventual entry into the nucleus. Other arguments can also be made to support this conclusion: i) Since fibers from other adenovirus serotypes (such as Ad7) lack O-GlcNAc (3, 29), it seems unlikely that this modification is an essential component for the function and targeting of fibers from other serotypes of adenovirus; ii) Some cellular proteins contain O-linked GlcNAc and remain in the cytoplasm or associated with the cytoplasmic surface of other organelles (17). If fiber is typical of other nuclear proteins, then it seems unlikely that GlcNAc will be part of the signal for transport of proteins destined for the nucleus.

Little is known about the role of monomeric O-linked GlcNAc on glycoproteins. In the case of fibers from Ad5 (and presumably Ad2), O-linked GlcNAc may help in the stabilization of trimers; when H5ts142 fibers synthesized at the permissive temperature are shifted to the nonpermissive temperature, fiber trimers of H5ts142 formed at the permissive temperature which contained GlcNAc did not appear to dissociate. However, when mutant fibers synthesized at 39.5°C are shifted to 32°C, they still lack GlcNAc but can form trimers, suggesting that GlcNAc is required for stabilization and not for assembly of trimers. This observation supports the possibility that O-linked GlcNAc may play a role in the stabilization of multimeric complexes (29). Further characterization of fibers from H5ts142 will aid in identifying the role of O-GlcNAc modification on proteins.

The next step in defining the requirements for the nuclear transport of the fiber was the identification of the essential peptide sequence for

nuclear localization in fiber. Sequences required for nuclear localization have been identified for a number of proteins (Table 2 in introduction, see references 5, 11, 21, 25, 28, 30, 33, and 36); the first of them was PKKKRKV of SV40 large T antigen. Many of these studies involved deletion or alteration of the nuclear proteins. In some studies, these sequences were fused to a cytosolic polypeptide in order to determine the sequence sufficient to redirect the cytoplasmic protein into the nucleus (8, 11, 15, 22, 24, 33). The same approach was used to identify the peptide sequence required for nuclear localization of Ad2 fiber. By using a vaccinia virus expression vector (27), fiber protein could be studied independently from adenovirus infection. Posttranslational modification such as glycosylation occurs faithfully and the nucleus remains intact throughout the vaccinia virus infection (26). When Ad2 fiber is expressed alone (S3-2F), the protein is transported to the nucleus and assembled into trimers. Therefore, fiber contains all of the signals required for its own transport to the nucleus and for its assembly. Cepko and Sharp (4) showed that 100K protein is required for trimerization and translocation of Ad5 hexon to the nucleus. This data suggests that there are at least two different systems that exist in adenovirus to assemble the structural proteins in the correct conformation and transport the protein to the nucleus.

A mutant of Ad2 fiber (2F-ΔKRAR) showed that the deletion of four amino acids (KRAR) at the N-terminus of Ad2 fiber is sufficient to abolish translocation of the fiber to the nucleus. This defect can be restored by placing the corresponding sequence from the Ad7 fiber, TKRVRL, in the N-terminus of the mutant 2F-ΔKRAR. The KRAR sequence is highly conserved among fibers from several serotypes, such as Ad5, Ad3, and Ad7 (6, 18, 35). Furthermore, these four amino acids of fiber also fit the proposed consensus sequence, K-R/K-X-R/K, predicted for the nuclear

localization signal (5). When these four residues were relocated to the C-terminal end of fiber, they could no longer function as a nuclear targeting signal. Deletions up to 138 amino acids at the C-terminus of Ad2 fiber have no apparent effect on targeting. When fused to the N-terminus of ß-galactosidase and expressed in a recombinant vaccinia virus, neither MKRARP nor MTKRVRL (from Ad2 and Ad7 fibers, respectively), were sufficient for transport of the hybrid protein to the nucleus. This result indicates that other adjacent or distant sequences might also be required for nuclear translocation of fiber. This observation was unexpected because a very similar sequence (KRPRP, (24)) from adenovirus E1A protein can translocate a cytoplasmic protein, E. coli galactokinase, into the nucleus. On the other hand, fusion of MKRARPSEDTF (from Ad2 fiber) or of MKRPRP (a known targeting sequence from the C-terminus of Ad2 E1A proteins) to ß-galactosidase allowed correct localization to the nucleus. The proline in KRPRP may help to present the signal in a correct conformation and the addition of SEDTF along with KRARP has compensated, by placing KRARP in the proper context so that the protein could be transported into the nucleus.

Another interesting result from these nuclear targeting mutants of adenovirus fiber is that all of the mutants were able to form trimers regardless of the location of fiber in the cell. 2F-ΔKRAR is a nuclear translocation-defective mutant, but it is still able to form trimers in the cytoplasm. Fiber from H5ts142 is transported to the nucleus as a monomer at the nonpermissive temperature (NPT), since it can not form trimers at NPT. It is known that the translocation of HA in influenza virus requires trimer formation for translocation through Golgi apparatus (7). This is not the case for the nuclear transport of fiber; the trimerization and the nuclear translocation of fiber seem to be independent of each other.

Little is known about how the trimerization of fiber occurs and what features of the protein contribute to the extreme stability of the fiber trimer. DNA sequence determination of the fiber gene from H5ts142 was the reasonable first step to explore the answer to this question. The mutation that is responsible for the trimer-deficient phenotype observed in the H5ts142 fiber lies near the C-terminus of fiber at amino acid 440 in the protein. This change from alanine to valine at aa 440 is located in the first part of the knob based on the secondary structure of fiber predicted by Green et al. (14) Two deletion mutants, 2F-dl444 and 2F-dl540 (138 and 42 amino acid deletions from the C-terminus of fiber, respectively), are both unable to form trimers; this defect in trimerization is not temperature-dependent. These results suggest that trimer formation is dependent on one or more domains within the C-terminus of fiber. Although more work needs to be done to find out the actual pathway for assembly of trimers, it can be speculated on the basis of these experiments that trimerization of fiber starts in the C-terminus. This process apparently proceeds independently of other adenovirus proteins (although not necessarily other cellular proteins), since vaccinia recombinants expressing fiber are still capable of forming trimers. This self-assembly of fiber is obviously different from the assembly of hexon trimers which requires the action of the adenovirus 100K protein (4); thus adenovirus contains at least two different pathways for assembly and transport of its structural proteins.

To understand more about the assembly of fiber, the same change in H5ts142 fiber was introduced into Ad2 fiber (2F-2ts142). The alanine residue at aa 440 is conserved in both Ad2 and Ad5 fiber. Surprisingly, the mutant Ad2 fiber, 2F-2ts142, did not show the temperature-sensitive phenotype for trimer formation as had been observed with the H5ts142 fiber. One potential explanation for this discrepancy is that the structures

of the two C-terminal domains is extremely variable and that domains important for trimerization of Ad5 fiber will be completely different from the domains found for Ad2. A second possibility is that the sequences around this alanine are somewhat different between the two serotypes: perhaps nearby amino acid residues in Ad2 fiber (AALAVAS) can compensate for the alteration in structure in ways that the Ad5 fiber (SVLAVK) cannot. The third possibility is that posttranslational modifications of Ad2 fiber may help to stabilize fiber trimers. Previously, it has been shown that Ad2 fiber contains 3-4 moles and Ad5 fiber contains 2 moles of O-linked N-acetylglucosamine residues per mole of fiber (29). Although the specific sites for attachment of these sugars are not known, Ad2 and Ad5 fibers share 2 potential sites in the N-terminus of fiber, while Ad2 contains additional potential sites in the C-terminus. It has been proposed that the shared GlcNAc sites in the N-terminus of fiber may help to stabilize trimers once they are formed (29). It is also possible that additional O-GlcNAc at the C-terminus of (only) Ad2 fiber may stabilize the trimer in ways not available to Ad5 fiber (which lacks apparent C-terminal O-GlcNAc sites).

Taken together, the signal for nuclear transport of fiber involves a peptide sequence, KRARPSEDTF, and it probably takes the same pathway through the nuclear pore complex as do other karyophilic proteins (1, 11, 31, 32). Little is known about how the transport occurs through the nuclear pore complex in general. The availability of an expression system and the analysis of various mutants can help to address the questions for this transport process. H5ts142 offered us an unique opportunity to learn that O-GlcNAc addition is not part of the signal for nuclear transport. This mutant also clarified the relationship between posttranslational modification and the formation and stabilization of multimeric complexes.

Mapping the sites of GlcNAc addition and generating mutants at these sites will provide valuable information to examine this relationship further.

Deletion of the C-terminal end of the fiber polypeptide resulted in a defect in trimer formation. This correlates well with the fact that H5ts142 contains a mutation only at the C-terminal end. Thus the C-terminal end of the protein must play a vital role in the trimerization of fiber, probably by forming a "hook" type of configuration and allowing three monomers to lock into each other. If this hypothesis is correct, the lower temperature (32°C) will allow the fiber from H5ts142 to form the right configuration; on the other hand, the higher temperature (39.5°C) will disturb the "hook" configuration and the fiber will no longer be able to form trimers. It has been suggested that the shaft part of fiber plays a role in the stabilization of fiber by forming inter-chain hydrogen bonds as well as hydrophobic interactions (9). Defining the minimum C-terminal sequence requirement for trimer formation as well as studying mutations at the glycosylation addition sites will provide a greater understanding of the mechanism of fiber assembly. The knowledge obtained from the study of fiber will be invaluable in the further elucidation of similar cellular processes in a wide variety of organisms.

REFERENCES FOR SUMMARY

1. Akey, C. W. and Goldfarb, D. S. 1989. Protein import through the nuclear pore complex is a multistep process. J. Cell Biol. 109:971-982.

2. Benko, D. M., Haltiwanger, R. S., Hart, G. W. and Gibson, W. 1988. Virion basic phosphoprotein from human cytomegalovirus contains O-linked N-acetylglucosamine. PNAS. 85:2573-2577.

3. Caillet-Boudin, M.-L., Strecker, G. and Michalski, J.-C. 1989. O-linked GlcNAc in serotype-2 adenovirus fibre. Eur.J.Biochem. 184:205-211.

4. Cepko, C. L. and Sharp, P. A. 1983. Analysis of Ad5 hexon and 100K temperature sensitive mutants using conformation-specific monoclonal antibodies. Virol. 129:137-154.

5. Chelsky, D., Ralph, R. and Jonak, G. 1989. Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol. Cell. Biol. 9:2487-2492.

6. Chroboczek, J. and Jacrot, B. 1987. The sequence of adenovirus fiber: similarities and differences between serotypes 2 and 5. Virol. 161:549-554.

7. Copeland, C. S., Doms, R. W., Bolzau, E. M., Webster, R. G. and Helenius, A. 1986. Assembly of influenza hemagglutinin trimers and its role on intracellular transport. J. Cell Biol. 103:1179-1191.

8. Davey, J., Dimmock, N. J. and Colman, A. 1985. Identification of the sequence responsible for the nuclear accumulation of the influenza virus nucleoprotein in Xenopus oocytes. Cell. 40:667-675.

9. Devaux, C., Adrian, M., Berthet-Colominas, C., Crusack, S. and Jacrot, B. 1990. Structure of Adenovirus Fibre I. Analysis of Crystals of fibre from Adenovirus serotype 2 and 5 by electron microscopy and X-ray crystallography. J. Mol. Biol. 215:567-588.

10. Devaux, C., Caillet-Boudin, M.-L., Jacrot, B. and Boulanger, P. 1987. Crystallization, enzymatic cleavage, and the polarity of the adenovirus type 2 fiber. Virol. 161:121-128.

11. Dingwall, C. and Laskey, R. A. 1986. Protein import into the cell nucleus. Ann. Rev. Cell Biol. 2:367-390.

12. Flint, S. J. and Broker, T. R. 1980. "The adenoviruses." In DNA tumor viruses. J. Tooze ed., CSH Laboratory, Cold Spring Harbor, New York.

13. Ginsberg, H. S. 1984. The adenoviruses. Plenum Press. New York.

14. Green, N. M., Wrigley, N. G., Russell, W. C., Martin, S. R. and McLachlan, A. D. 1983. Evidence for a repeating cross-b sheet structure in the adenovirus fibre. EMBO J. 2:1357-1365.

15. Hall, M. N., Hereford, L. and Herskowitz, I. 1984. Targeting of E.coli ß-galactosidase to the nucleus in yeast. Cell. 36:1057-1065.

16. Holt, G. D., Haltiwanger, R. S., Torres, C.-R. and Hart, G. W. 1987. Erythrocyte contains cytoplasmic glycoproteins. J. Biol. Chem. 262:14847-14850.

17. Holt, G. D. and Hart, G. W. 1986. The subcellular distribution of terminal N-Acetylglucosamine moieties. J. Biol. Chem. 261:8049-8057.

18. Hong, J. S., Mullis, K. G. and Engler, J. A. 1988. Characterization of the early region 3 and fiber genes of Ad7. Virol. 167:545-553.

19. Ishibashi, M. and Maizel, J. V. 1974. The polypeptides of adenovirus VI. Early and late glycoproteins. Virol. 58:345-361.

20. Jackson, S. P. and Tjian, R. 1988. O-Glycosylation of eukaryotic transcription factors:Implications for mechanisms of transcriptional regulation. Cell. 55:125-133.

21. Kalderon, D., Richardson, W. D., Markham, A. F. and Smith, A. E. 1984. Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature. 311:33-38.

22. Kalderon, D., Roberts, B. L., Richardson, W. D. and Smith, A. E. 1984. A short amino acid sequence able to specify nuclear location. Cell. 39:499-509.

23. Kelly, W. G. and Hart, G. W. 1989. Glycosylation of chromosomal proteins: Localization of O-linked N-acetylglucosamine in Drosophila chromatin. Cell. 57:243-251.

24. Lyons, R. H., Ferguson, B. Q. and Rosenberg, M. 1987. Pentapeptide nuclear localization signal in adenovirus E1a. Mol.Cell.Biol. 7:2451-2456.

25. Morin, N., Delsert, C. and Klessig, D. F. 1989. Nuclear localization of adenovirus DNA-binding protein: requirements of two signals and complementation during virus infection. Mol. Cell. Biol. 9:4372-4380.

26. Moss, B. 1990. "Poxviridae and their replication." In Virology. B.N. Fields ed. Raven Press, Ltd., New York. Chapter 74. p2079-2111.

27. Moss, B. and Flexner, C. 1987. Vaccinia virus expression vectors. Ann. Rev. Immunol. 5:305-324.

28. Mukaigawa, J. and Nayak, D. P. 1991. Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2. J. Virol. 65:245-253.

29. Mullis, K. G., Haltiwanger, R. S., Hart, G. W., Marchase, R. B. and Engler, J. A. 1990. Relative accessibility of N-acetylglucosamine in trimers of the adenovirus types 2 and 5 fiber proteins. J. Virol. 64:5317-5323.

30. Nath, S. T. and Nayak, D. P. 1990. Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). Mol. Cell Biol. 10:4139-4145.

31. Newmeyer, D. D. and Forbes, D. J. 1988. Nuclear import can be separated into distinct steps in vitro: Nuclear pore binding and translocation. Cell. 52:641-653.

32. Richardson, W. D., Mills, A. D., Dilworth, S. M., Laskey, R. A. and Dingwall, C. 1988. Nuclear protein migration involves two steps: Rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. Cell. 52:655-664.

33. Richardson, W. D., Roberts, B. L. and Smith, A. E. 1986. Nuclear location signals in polyoma virus large-T. Cell. 44:77-85.

34. Schindler, M., Hogan, M., Miller, R. and DeGaetano, D. 1987. A nuclear specific glycoprotein representative of a unique pattern of glycosylation. J. Biol. Chem. 262:1254-1260.

35. Signas, C., Akusjarvi, G. and Petterson, U. 1985. Adenovirus 3 fiber polypeptide gene: Implications for the structure of the fiber protein. J. Virol. 53:672-678.

36. Smith, A. E., Kalderon, D., Roberts, B. L., Colledge, W. H., Edge, M., Gillett, P., Markham, A., Paucha, E. and Richardson, W. D. 1985. The nuclear location signal. Proc. R. Soc. Lond. B. 226:43-58.

APPENDIX A

UNPUBLISHED WORK ON THE SEQUENCE COMPARISON OF ADENOVIRUS BAM HI D FRAGMENT

Nucleotide as well as amino acid sequences from regions encoding E3, fiber and E4 were compared to Ad2, Ad3, Ad4, and Ad5. The results of these comparisons are shown in Figures 1, 2, 3, 4, and 5. The nucleotide homology, predicted amino acid homology, and conserved regions were derived from this data and are presented in the manuscript, "Characterization of the early region 3 and fiber genes of Ad7."

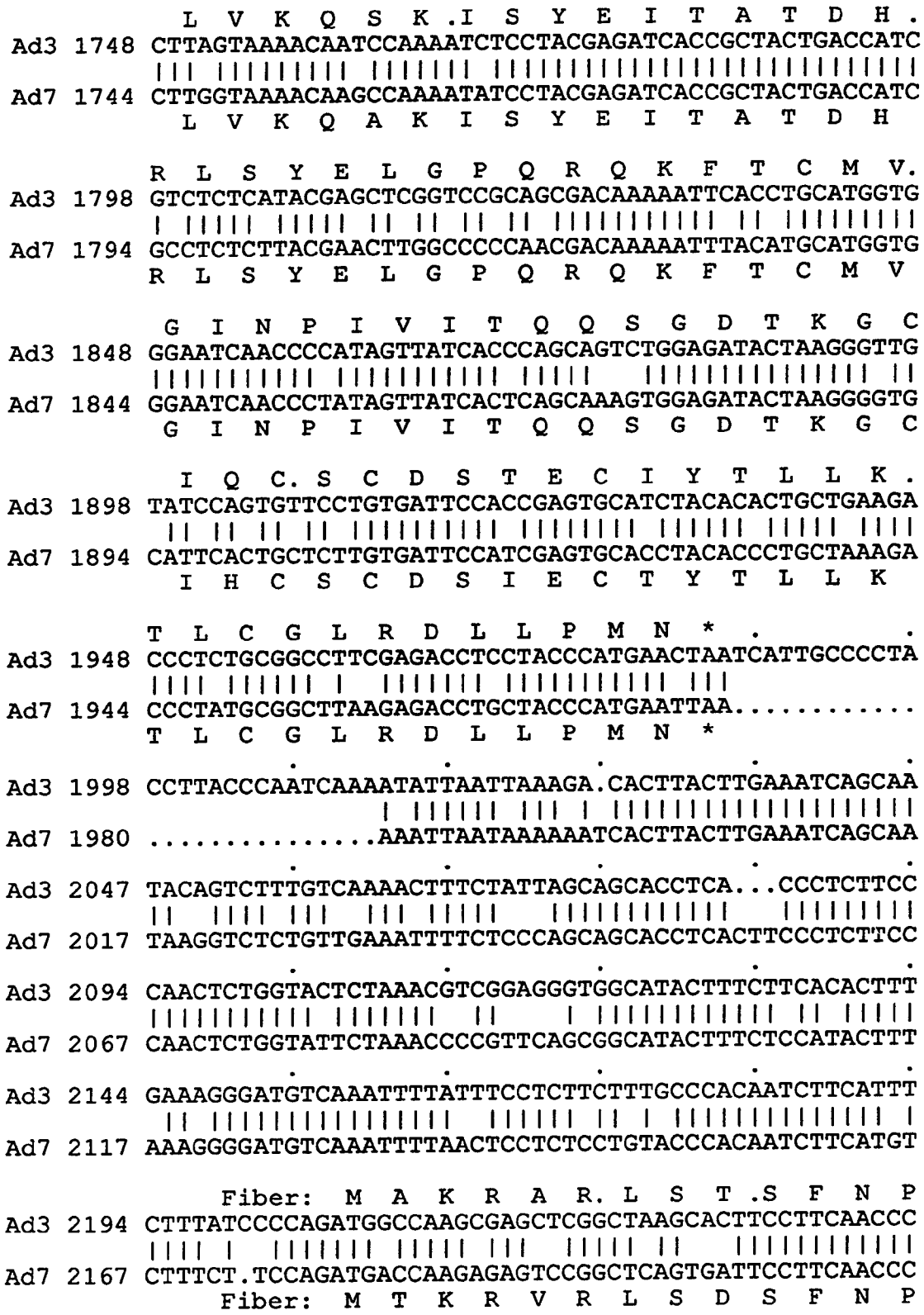
Figure 1: Detailed comparison of Ad3 and Ad7 nucleotide and predicted amino acid sequences in early region 3. The nucleotide sequences are aligned using the algoritm of Needleman and Wunsch (J. Mol. Biol. 48, 443-453, 1970). The predicted amino acids are represented by one-letter codes, and the nucleotide number for each sequence is listed at the beginning of each line. Matches between the serotypes are designated by a perpendicular line between the aligned sequences. The predicted amino acid sequences from Ad3 are taken from Signas et al. (Gene, 50, 173-184,1986). The predicted amino acid sequences for Ad7 were taken from Hong et al. (Virology, 167, 545-553, 1988).

```
20.1 Kd:   D  P  L. L  N  F .N  I  *  . 20.5 Kd:. M  I  S .
Ad3    1 GGATCCACTACTAAATTTTAACATTTAATTTTTTATACAGATGATTTCCA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Ad7    1 GGATCCACTACTAAATTTTAACATTTAATTTTTTATACAGATGATTTCCA
20.1 Kd:   D  P  L  L  N  F  N  I  *    20.6 Kd:  M  I  S

           T  T  I  F  I  I  T  S  L  A  A  V  T  Y  G  R  S.
Ad3   51 CTACAATTTTTATCATTACTAGCCTTGCAGCTGTAACTTATGGCCGTTCA
         |||||||||||||||||||||||||| ||||||||||||||||||||
Ad7   51 CTACAATTTTTATCATTACTAGCCTTGCGGCTGTAACTTATGGCCGTTCA
           T  T  I  F  I  I  T  S  L  A  A  V  T  Y  G  R  S

            H  L  T  V  P  V  G  S  T  C  T  L  Q  G  P  Q  E
Ad3  101 CACCTAACTGTACCTGTTGGCTCAACATGTACACTACAAGGACCCCAAGA
         |||||||||||||||||||||||||||||||||||||||||||||||| |
Ad7  101 CACCTAACTGTACCTGTTGGCTCAACATGTACACTACAAGGACCCCAACA
            H  L  T  V  P  V  G  S  T  C  T  L  Q  G  P  Q  Q

              G  Y  V  T  W  W  R  I  Y  D  N  G  G  F  A  R.
Ad3  151 AGGCTATGTCACTTGGTGGAGAATATATGATAATGGAGGGTTCGCTAGAC
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Ad7  151 AGGCTATGTCACTTGGTGGAGAATATATGATAATGGAGGGTTCGCTAGAC
              G  Y  V  T  W  W  R  I  Y  D  N  G  G  F  A  R

           P  C  D  Q  P  G  T  K  F  S  C  N  G  R  D  L  T.
Ad3  201 CATGTGATCAGCCTGGTACAAAATTTTCATGCAACGGAAGAGACTTGACC
         |||||||||||||||||||||||||||||||||||||||||||| |||
Ad7  201 CATGTGATCAGCCTGGTACAAAATTTTCATGCAACGGAAGAGACTTAACC
           P  C  D  Q  P  G  T  K  F  S  C  N  G  R  D  L  T

            I  I  N  I  T  S  N  E  Q  G  F  Y  Y  G  T  N  Y
Ad3  251 ATTATTAACATAACATCAAATGAGCAAGGCTTCTATTATGGAACCAACTA
         ||  ||||||||||||||||||||||||||||||||||||||||||||||
Ad7  251 ATAATTAACATAACATCAAATGAGCAAGGCTTCTATTATGGAACCAACTA
           I  I  N  I  T  S  N  E  Q  G  F  Y  Y  G  T  N  Y

              K  N  S  L  D  Y  N  I  I  V  V  P  A  T  T  S.
Ad3  301 TAAAAATAGTTTAGATTACAACATTATTGTAGTGCCAGCCACCACTTCTG
         ||||  ||||||||||||||||||||||||||||||||||||||||||||
Ad7  301 TAAAGATAGTTTAGATTACAACATTATTGTAGTGCCAGCCACCACTTCTG
              K  D  S  L  D  Y  N  I  I  V  V  P  A  T  T  S

           A  P  R  K  S  T  F  S  S  S  S  A  K  A  S  T  I.
Ad3  351 CTCCCCGCAAATCCACTTTCTCTAGCAGCAGTGCCAAAGCAAGCACAATT
         ||||||||||| ||||||||||||||||||||||||||||||||||||||
Ad7  351 CTCCCCGCAAAACCACTTTCTCTAGCAGCAGTGCCAAAGCAAGCACAATT
           A  P  R  K  T  T  F  S  S  S  S  A  K  A  S  T  I

            P  K  T  A  S  A  M  L  K  L  P  K  I  A  L  S  N
Ad3  401 CCTAAAACAGCTTCTGCTATGTTAAAGCTTCCAAAAATCGCTTTAAGTAA
         |||||||||||||||||||||||||||| ||||||||||||||||||
Ad7  401 CCTAAAACAGCTTCTGCTATGTTAAAGCTTCAAAAAATCGCTTTAAGTAA
            P  K  T  A  S  A  M  L  K  L  Q  K  I  A  L  S  N
```

```
            S  T  A  A  P  N  T  I  P  K  S  T  I  G  I  I .
Ad3   451 TTCCACAGCCGCTCCCAATACAATTCCTAAATCAACAATTGGCATCATTA
          |||||||||||||||||| ||||||||||||||||||||||||||||||
Ad7   451 TTCCACAGCCGCTCCCAAAACAATTCCTAAATCAACAATTGGCATCATTA
            S  T  A  A  P  K  T  I  P  K  S  T  I  G  I  I

          T  A  V  V  V  G  L  M  I  I  F  L  C  I  M  Y  Y.
Ad3   501 CTGCCGTGGTAGTGGGATTAATGATTATATTTTTGTGTATAATGTACTAC
          |||||||||||||||||||| |||||||||||||| ||||||||||||
Ad7   501 CTGCCGTGGTAGTGGGATTAATTATTATATTTTTGTGCATAATGTACTAT
          T  A  V  V  V  G  L  I  I  I  F  L  C  I  M  Y  Y

           A  C  C  Y  R  K  H  E  Q  K  G  D  A  L  L  N  F
Ad3   551 GCCTGCTGCTATAGAAAACATGAACAAAAAGGTGATGCATTACTAAATTT
          ||||||||||||||||||||||||||||||||||||||||||||||||
Ad7   551 GCCTGCTGCTATAGAAAACATGAACAAAAAGGTGATGCATTACTAAATTT
           A  C  C  Y  R  K  H  E  Q  K  G  D  A  L  L  N  F

            D  I  *        9.0 Kd:  M  I  L  F  Q  S  N  T  .
Ad3   601 TGATATTTAATTTTTTATAGAATTATGATATTGTTTCAATCAAATACC..
          |||  ||||||||||||||||||||||||||||||||||||||||
Ad7   601 TGACATTTAATTTTTTATAGAATTATGATATTGTTTCAATCAAATACCAC
            D  I  *        7.7 Kd:  M  I  L  F  Q  S  N  T  T
                   .        .        .        .        .
Ad3   649 ..........................................A
                                                    |
Ad7   651 TAACACTATCAATGTGCAGACTACTTTAAATCATGACATGGAAAACCACA
            N  T  I  N  V  Q  T  T  L  N  H  D  M  E  N  H

          T  T  S  Y  A  Y  T  N  I  Q  P  K  Y  A  M  Q  L.
Ad3   650 CTACCTCCTATGCATACACAAACATTCAGCCTAAATACGCTATGCAACTA
          |||||||||||||||||| |||||||||||||||||||||||| |||
Ad7   701 CTACCTCCTATGCATACATAAACATTCAGCCTAAATACGCTATGCATCTA
          T  T  S  Y  A  Y  I  N  I  Q  P  K  Y  A  M  H  L
                   .        .        .        .        .

           K  I  T  I  L  I  V  I  G  I  L  I  L  S  V  I  .
Ad3   700 GAAATCACAATACTAATTGTAATTGGAATTCTTATACTATCTGTTATTC
          |||||||  ||||||||||||||||||||||||||||||||||||||
Ad7   751 AAAATCACCATACTAATTGTAATTGGAATTCTTATACTATCTGTTATTC
           K  I  T  I  L  I  V  I  G  I  L  I  L  S  V  I

          L  Y  F  I  F  C  R. Q  I  P .N  V  H  R  N  S  K.
Ad3   749 TTTATTTTATATTCTGCCGTCAAATACCCAATGTTCATAGAAATTCTAAA
          ||||||||
Ad7   800 TTTATTTT.....................................

          L  Y  F

           R  R  P .I  Y  S  P  M  I  S. R  P  H .M  A  L  N
Ad3   799 AGACGTCCCATCTATTCTCCTATGATTAGTCGTCCCCATATGGCTCTGAA
                    |||||||||||||||||||||||||||| ||||||||
Ad7   807 ...........CTATTCTCCTATGATTAGTCGTCCCCATATAGCTCTGAA
                     L  F  S  Y  D  *
```

```
            E  I  *.          .          .          .          .
Ad3   849 TGAAATCTAAGATC.TTTTTTTTTCTCTTACAGTATGGTGAACATCAATC
          |||||||||||||| |||||||||||||||||||||||||||| |||||
Ad7   847 TGAAATCTAAGATCTTTTTTTTTTTCTCTTACAGTATGGTGAACACCAATC

10.2 Kd:  M  I  P  R  N  F  F  F  T  I  L  I  C  A  F  N  V
Ad3   898 ATGATCCCTAGAAATTTCTTCTTCACCATACTCATCTGTGCTTTTAATGT
          |||||||||||||||||||||||||||||||||||||||||||| |||||
Ad7   897 ATGATCCCTAGAAATTTCTTCTTCACCATACTCATCTGTGCTTTCAATGT
10.3 Kd:  M  I  P  R  N  F  F  F  T  I  L  I  C  A  F  N  V

            C  A  T  F  T  A  V  A  T  A  S  P  D  C  I  G.
Ad3   948 CTGTGCTACTTTCACAGCAGTAGCCACTGCAAGCCCAGACTGTATAGGAC
          |||||||||||||||||||||||||||||||||||||||||||||||||
Ad7   947 CTGTGCTACTTTCACAGCAGTAGCCACTGCAAGCCCAGACTGTATAGGAC
            C  A  T  F  T  A  V  A  T  A  S  P  D  C  I  G

          P  F  A  S  Y  A  L  F  A  F  V  T  C  I  C  V  C.
Ad3   998 CATTTGCTTCCTATGCACTTTTTGCCTTCGTTACTTGCATCTGCGTGTGT
          |||||||||||||||||||||||||||||||||||||||||||||||||
Ad7   997 CATTTGCTTCCTATGCACTTTTTGCCTTCGTTACTTGCATCTGCGTGTGT
          P  F  A  S  Y  A  L  F  A  F  V  T  C  I  C  V  C

           S  I  V  C  L  V  I  N  F  F  Q  L  V  D  W  I  F
Ad3  1048 AGCATAGTCTGCCTGGTTATTAATTTTTTCCAACTGGTAGACTGGATCTT
          |||||||||||||||||||||||||||||||||||||||||||||||||
Ad7  1047 AGCATAGTCTGCCTGGTTATTAATTTTTTCCAACTGGTAGACTGGATCTT
           S  I  V  C  L  V  I  N  F  F  Q  L  V  D  W  I  F

            V  R  I  A  Y  L  R  H  H  P  E  Y  R  N  Q  N.
Ad3  1098 TGTGCGAATTGCCTACCTACGTCACCATCCCGAATACCGCAATCAAAATG
          |||  |||||||||||||||||||||||||||||||||||||||||||||
Ad7  1097 TGTACGAATTGCCTACCTACGTCACCATCCCGAATACCGCAATCAAAATG
            V  R  I  A  Y  L  R  H  H  P  E  Y  R  N  Q  N

                              15.2 Kd:  M  Q  A  M  L  P  V
          V  A  A  L  L  R  L  I  *     .          .          .
Ad3  1148 TTGCGGCACTTCTTAGGCTTATTTAAAACCATGCAGGCTATGCTACCAGT
          |||||||||||||||||||||||||||||||||||||||||||||||||
Ad7  1147 TTGCGGCACTTCTTAGGCTTATTTAAAACCATGCAGGCTATGCTACCAGT
          V  A  A  L  L  R  L  I  *
                              14.9 Kd:  M  Q  A  M  L  P  V

            I  L  I  L  L  L  P  C  I  P  L  A  S  T  A  T.
Ad3  1198 CATTTTAATTTTGCTACTACCCTGCATTCCCCTAGCTTCCACCGCCACTC
          |||||||||| |||||||||||||||| ||||||||||||||||||||
Ad7  1197 CATTTTAATTCTGCTACTACCCTGCATTGCCCTAGCTTCCACCGCCACTC
            I  L  I  L  L  L  P  C  I  A  L  A  S  T  A  T

          R  A  T  P  E  Q  L  R  K  C  K  F  Q  Q  P  W  S.
Ad3  1248 GCGCTACACCTGAACAACTTAGAAAATGCAAATTTCAACAACCATGGTCA
          |||||||||||||||||||||||||||||||||||||||||||||||||
Ad7  1247 GCGCTACACCTGAACAACTTAGAAAATGCAAATTTCAACAACCATGGTCA
          R  A  T  P  E  Q  L  R  K  C  K  F  Q  Q  P  W  S
```

```
            F  L  D  C  Y  H  E  K  S  D  F  P  T  Y  W  I  V
Ad3 1298 TTTCTTGATTGCTACCATGAAAAATCTGATTTTCCCACATACTGGATAGT
         ||||||||||||||||||||||||||||||| |||||||||||||||||
Ad7 1297 TTTCTTGATTGCTACCATGAAAAATCTGATTTCCCCACATACTGGATAGT
            F  L  D  C  Y  H  E  K  S  D  F  P  T  Y  W  I  V

              I  V  G  I  I  N  I  L  S  C  T  F  F  S  I  T .
Ad3 1348 GATTGTTGGAATAATTAACATACTTTCATGTACCTTTTTCTCAATCACAA
         ||||||||||||||||||||||||||||||| |||||||||||||||||
Ad7 1347 GATTGTTGGAATAATTAACATACTTTCATGTACATTTTTCTCAATCACAA
              I  V  G  I  I  N  I  L  S  C  T  F  F  S  I  T

            I  Y  P  T  F  N  F  G  W  N  S  P  N  A  L  G  Y.
Ad3 1398 TATACCCCACATTTAATTTTGGGTGGAATTCTCCCAATGCACTGGGTTAC
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Ad7 1397 TATACCCCACATTTAATTTTGGGTGGAATTCTCCCAATGCACTGGGTTAC
            I  Y  P  T  F  N  F  G  W  N  S  P  N  A  L  G  Y

            P  Q  E  P  D  E  H  I  P  L  Q  H  I  Q  Q  P  L.
Ad3 1448 CCACAAGAACCAGATGAACATATTCCACTACAACACATACAACAACCACT
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Ad7 1447 CCACAAGAACCAGATGAACATATTCCACTACAACACATACAACAACCACT
            P  Q  E  P  D  E  H  I  P  L  Q  H  I  Q  Q  P  L

                  .           .           .           .           .
Ad3 1498 AGCACTGGTACAGTATGAAAATGAGCCACAACCTTCACTGCCCCCTGCCA
         |||||||||| |||||||||||||||||||||||||||| |||||||
Ad7 1497 AGCACTGGTAGAGTATGAAAATGAGCCACAACCTTCACTGCCTCCTGCCA
            A  L  V  E  Y  E  N  E  P  Q  P  S  L  P  P  A

                                  15.3 Kd:  M  T  D  P  I  A  T  S
            I  S  Y  F  N  L  T  G  G  D  D  *     .           .
Ad3 1548 TTAGTTACTTCAACCTAACCGGCGGAGATGACTGACCCAATCGCCACATC
         |||||||||||||||||||||||||||||||||   |||   |||  |
Ad7 1547 TTAGTTACTTCAACCTAACCGGCGGAGATGACTGA...AATACTCACCAC
            I  S  Y  F  N  L  T  G  G  D  D  *
                                  15.3 Kd:  M  T  E     I  L  T  T

              S  T  A. A  K  E .L  L  D  M  D  G  R. A  S  E .
Ad3 1598 ATCCACCGCTGCCAAGGAGCTGCTGGACATGGACGGACGTGCCTCAGAAC
          ||||   | ||| |||| ||||| || |||||||| || | ||||||||
Ad7 1594 CTCCAATTCCGCCGAGGATCTGCTTGATATGGACGGCCGCGTCTCAGAAC
            S  N  S  A  E  D  L  L  D  M  D  G  R  V  S  E

            Q  R  L  I  Q  L  R. I  R  Q .Q  Q  E  R  A  V  K.
Ad3 1648 AGCGACTCATCCAACTACGCATTCGTCAGCAGCAGGAACGAGCAGTAAAA
         ||||||||  |||||||||||  || |||||||||||||  |     |||
Ad7 1644 AGCGACTCGCCCAACTACGCATCCGCCAGCAGCAGGAACGCGTGACCAAA
         Q  R  L  A  Q  L  R  I  R  Q  Q  Q  E  R  V  T  K

             E  L  R .D  A  I  G  I  H  Q. C  K  K  G  I  F  C
Ad3 1698 GAGCTAAGGGATGCCATTGGGATTCACCAGTGCAAAAAAGGCATATTCTG
         |||||  ||  ||||  |||    |||||||| |||||||||||||||  ||
Ad7 1694 GAGCTCAGAGATGTCATCCAAATTCACCAATGCAAAAAAGGCATATTTTG
             E  L  R  D  V  I  Q  I  H  Q  C  K  K  G  I  F  C
```


L V K Q S K.I S Y E I T A T D H.
Ad3 1748 CTTAGTAAAACAATCCAAAATCTCCTACGAGATCACCGCTACTGACCATC
Ad7 1744 CTTGGTAAAACAAGCCAAAATATCCTACGAGATCACCGCTACTGACCATC
L V K Q A K I S Y E I T A T D H
R L S Y E L G P Q R Q K F T C M V.
Ad3 1798 GTCTCTCATACGAGCTCGGTCCGCAGCGACAAAAATTCACCTGCATGGTG
Ad7 1794 GCCTCTCTTACGAACTTGGCCCCCAACGACAAAAATTTACATGCATGGTG
R L S Y E L G P Q R Q K F T C M V
G I N P I V I T Q Q S G D T K G C
Ad3 1848 GGAATCAACCCCATAGTTATCACCCAGCAGTCTGGAGATACTAAGGGTTG
Ad7 1844 GGAATCAACCCTATAGTTATCACTCAGCAAAGTGGAGATACTAAGGGGTG
G I N P I V I T Q Q S G D T K G C
I Q C.S C D S T E C I Y T L L K.
Ad3 1898 TATCCAGTGTTCCTGTGATTCCACCGAGTGCATCTACACACTGCTGAAGA
Ad7 1894 CATTCACTGCTCTTGTGATTCCATCGAGTGCACCTACACCCTGCTAAAGA
I H C S C D S I E C T Y T L L K
T L C G L R D L L P M N * .
Ad3 1948 CCCTCTGCGGCCTTCGAGACCTCCTACCCATGAACTAATCATTGCCCCTA
Ad7 1944 CCCTATGCGGCTTAAGAGACCTGCTACCCATGAATTAA.....
T L C G L R D L L P M N *
Ad3 1998 CCTTACCCAATCAAAATATTAATTAAAGA.CACTTACTTGAAATCAGCAA
Ad7 1980 .....AAATTAATAAAAAATCACTTACTTGAAATCAGCAA
Ad3 2047 TACAGTCTTTGTCAAAACTTTCTATTAGCAGCACCTCA...CCCTCTTCC
Ad7 2017 TAAGGTCTCTGTTGAAATTTTCTCCCAGCAGCACCTCACTTCCCTCTTCC
Ad3 2094 CAACTCTGGTACTCTAAACGTCGGAGGGTGGCATACTTTCTTCACACTTT
Ad7 2067 CAACTCTGGTATTCTAAACCCCGTTCAGCGGCATACTTTCTCCATACTTT
Ad3 2144 GAAAGGGATGTCAAATTTTATTTCCTCTTCTTTGCCCACAATCTTCATTT
Ad7 2117 AAAGGGGATGTCAAATTTTAACTCCTCTCCTGTACCCACAATCTTCATGT
Fiber: M A K R A R.L S T.S F N P
Ad3 2194 CTTTATCCCCAGATGGCCAAGCGAGCTCGGCTAAGCACTTCCTTCAACCC
Ad7 2167 CTTTCT.TCCAGATGACCAAGAGAGTCCGGCTCAGTGATTCCTTCAACCC
Fiber: M T K R V R L S D S F N P

```
            V  Y  P. Y  E  D .E  S  S  S  Q  H  P. F  I  N .P
Ad3 2244 GGTGTACCCTTATGAAGATGAAAGCAGCTCACAACACCCATTTATAAATCC
          || ||||| |||||||||||||||| ||| |||||||| |||||||| ||
Ad7 2216 TGTCTACCCCTATGAAGATGAAAGCACCTCCCAACACCCCTTTATAAACCC
2266
            V  Y  P  Y  E  D  E  S  T  S  Q  H  P  F  I  N  P

            G  F  I. S  P  D .G  F  T  Q  S  P  N. G  V  L .
Ad3 2295 TGGTTTCATTTCCCCTGACGGGTTCACACAAAGTCCAAACGGGGTTTTAA
          || || |||||||| | || || |||||||| ||| |||| ||| | |
Ad7 2267 AGGGTTTATTTCCCCAAATGGCTTTACACAAAGCCCAGACGGAGTTCTTA
            G  F  I  S  P  N  G  F  T  Q  S  P  D  G  V  L

          S  L  K  C  V  N  P. L  T  T .A  S  G  S  L  Q  L.
Ad3 2345 GTCTTAAATGTGTTAATCCACTTACCACTGCAAGCGGCTCCCTCCAACTT
          | | |||||| | |  ||||| || ||  || |||| || || ||  |
Ad7 2317 CTTTAAAATGTTTAACCCCACTAACAACCACAGGCGGGTCTCTACAGTTA
          T  L  K  C  L  T  P  L  T  T  T  G  G  S  L  Q  L

           K  V  G .S  G  L  T  V  D  T. T  D  G .S  L  E  E
Ad3 2395 AAAGTGGGAAGTGGTCTTACAGTAGACACTACTGATGGATCCTTAGAAGA
          ||||||||| | ||||||||| ||||    || || || |  || ||||
Ad7 2367 AAAGTGGGAGGGGGTCTTACAATAGATGACACCGACGGTTTTTTGAAAGA
           K  V  G  G  G  L  T  I  D  D  T  D  G  F  L  K  E

            N  I  K. V  N  T .P  L  T  K  S  N  H. S  I  N..
Ad3 2445 AAACATCAAAGTTAACACCCCCCTAACAAAGTCAAACCATTCTATAAATT
          |||||| |  |    ||| || ||   ||| |   || ||||||  ||
Ad7 2417 AAACATAAGTGCCGCCACACCACTCGTTAAGACTGGTCACTCTATAGGTT
            N  I  S  A  A  T  P  L  V  K  T  G  H  S  I  G

          L  P  I  G  N  G  L. Q  I  E  Q  N  K  L. C  S  K .
Ad3 2495 TACCAATAGGAAACGGTTTGCAAATAGAACAAAACAAACTTTGCAGTAAA
          |  |  |||||  ||| ||   || | |  ||||||||||||    |||
Ad7 2467 TGTCGCTAGGACCCGGATTAGGAACAAATGAAAACAAACTTTGTGCCAAA
          L  S  L  G  P  G  L  G  T  N  E  N  K  L  C  A  K

           L  G  N .G  L  T  F  D  S  S. N     S .I  A  K  K
Ad3 2545 CTCGGAAATGGTCTTACATTTGACTCTTCCAA...TTCTATTGCACTGAA
          | ||| | || |||||||| | ||  |||  ||  ||||     ||
Ad7 2517 TTGGGAGAAGGACTTACATTCAATTCCAACAACATTTGCATTGATGACAA
           L  G  E  G  L  T  F  N  S  N  N  I  C  I  D  D  N

            N  N  T. L  W  T .G  P  K  .  P  E  A. N  C    .
Ad3 2592 AAATAACACTTTATGGACAGGTCCAAAAC...CAGAAGCCAACTGC...A
          | |||||| |||||||||    || |   |  |||||||||    |
Ad7 2567 TATTAACACCCTATGGACAGGAGTTAACCCCACCACAGCCAACTGTCAAA
            I  N  T  L  W  T  G  V  N  P  T  T  A  N  C  Q

          I  I  E  Y  G  K  Q. N  P  D .S  K  L  T  L  I  L.
Ad3 2636 TAATTGAATACGGGAAACAAAACCCAGATAGCAAACTAACTTTAATCCTT
          |||| |  | | |  ||   ||    ||| ||||| ||| | |||  ||
Ad7 2617 TAATGGCCTCCAGTGAATCTAAT...GATTGCAAATTAATTCTAACACTA
          I  M  A  S  S  E  S  N     D  C  K  L  I  L  T  L
```

```
          V  K  N .G  G  I  V  N  G  Y. V  T  L .M  G  A  S
Ad3 2686 GTAAAAAATGGAGGAATTGTTAATGGATATGTAACGCTAATGGGAGCCTC
         || |||| ||||||  |||| | || || |||    | || ||||  ||
Ad7 2664 GTTAAAACTGGAGGCCTTGTCACTGCATTTGTTTATGTTATAGGAGTATC
          V  K  T  G  G  L  V  T  A  F  V  Y  V  I  G  V  S

           D  Y  V. N  T  L .F  K  N  K  N  V  S. I  N  V .
Ad3 2736 AGACTACGTTAACACCTTATTTAAAAACAAAAATGTCTCCATTAATGTAG
           || |  |||| |   ||  || | | |||||| |     | | || ||
Ad7 2714 TAACGATTTTAATATGCTAACTACACATAAAAATATAAATTTCACTGCAG
           N  D  F  N  M  L  T  T  H  K  N  I  N  F  T  A

         E  L  Y  F  D  A  T. G  H  I .L  P  D  S  S  S  L.
Ad3 2786 AACTATACTTTGATGCCACTGGTCATATATTACCAGACTCATCTTCTCTT
         | || |  |||||| | |||||| || || || |   |   || || ||
Ad7 2764 AGCTGTTTTTTGATTCTACTGGTAATTTACTAACTAGCCTTTCATCCCTA
         E  L  F  F  D  S  T  G  N  L  L  T  S  L  S  S  L

          K  T  D .L  E  L  K         . Y  K  Q .T  A  D  F
Ad3 2836 AAAACAGATCTAGAACTAAAAT.........ACAAGCAAACCGCTGACTT
         |||||    ||  | |  ||||         ||| |   || | || | |
Ad7 2814 AAAACTCCACTTAATCATAAATCAGGGCAAAACATGGCTACTGGTGCCCT
          K  T  P  L  N  H  K  S  G  Q  N  M  A  T  G  A  L

             S  A. R  G  F .M  P  S  T  T  A  Y. P  F  V .
Ad3 2877 T...AGTGCAAGAGGTTTTATGCCAAGTACTACAGCGTATCCATTTGTCC
         |   | ||| | |||||| ||||| || || || || |||||   | ||
Ad7 2864 TACTAATGCTAAAGGTTTCATGCCCAGCACAACTGCCTATCC...TTTCA
           T  N  A  K  G  F  M  P  S  T  T  A  Y  P     F

         L  P  N  A  G  T  H. N  E  N .Y  I  F  G  Q  C  Y.
Ad3 2924 TTCCTAATGCGGGAACACATAATGAAAATTATATTTTTGGTCAATGCTAC
          |  |||| |    | | | || ||||| || |||| ||    || |||
Ad7 2911 ATGTTAATTC...CAGAGAAAAAGAAAACTACATTTACGGAACTTGTTAC
         N  V  N  S     R  E  K  E  N  Y  I  Y  G  T  C  Y

          Y  K  A .S  D  G  A  L  F  P. L  E  V .T  V  M  L
Ad3 2974 TACAAAGCAAGCGATGGTGCCCTTTTTCCGTTGGAAGTTACTGTTATGCT
         |||| ||| || |||    |  ||||||  | ||  |  |||| |||||
Ad7 2958 TACACAGCTAGTGATCACACTGCTTTTCCCATTGACATATCTGTCATGCT
          Y  T  A  S  D  H  T  A  F  P  I  D  I  S  V  M  L

           N  K  R. L  P  D .S  R  T  S  Y  V  M. T  F  L .
Ad3 3024 TAATAAACGCCTGCCAGATAGTCGCACATCCTATGTTATGACTTTTTTAT
         |||  || |      | ||| |  ||||| |||  |||   | |    |
Ad7 3008 TAACCAAAGAGCATTAAATAATGAGACATCATATTGTATTCGTGTAACTT
           N  Q  R  A  L  N  N  E  T  S  Y  C  I  R  V  T

         W  S  L  N  A  G  L. A  P  E .      T  T  Q  A  T.
Ad3 3074 GGTCCTTGAATGCTGGTCTAGCTCCAGAA......ACTACTCAGGCAACC
         |||||| |||  | ||  | || ||||||      ||  ||    | |||
Ad7 3058 GGTCCTGGAACACAGGAGTTGCCCCAGAAGTGCAAACCTCTGCTACTACC
         W  S  W  N  T  G  V  A  P  E  V  Q  T  S  A  T  T
```

```
          L  I  T .S  P  F  T  F  S  Y. I  R  E .D  D  *  .
Ad3 3118 CTCATAACCTCCCCATTTACCTTTTCCTATATTAGAGAAGATGACTGACA
         ||  | ||||| |||||||||||||| ||| |||||||||||| ||||||||
Ad7 3108 CTAGTCACCTCTCCATTTACCTTTTACTACATTAGAGAAGACGACTGACA
          L  V  T  S  P  F  T  F  Y  Y  I  R  E  D  D  *
                .           .           .           .           .
Ad3 3168 ACAAAAATAAAGTTCAACATTTTTTATTGAAATTCCTTTTACAGTATTCG
         |    ||||||||| ||| || |||||| ||| ||  || | |  |||||
Ad7 3158 A....AATAAAGTTTAAC.TTGTTTATTTAAAATCAATTCATAAAATTCG
                .           .           .           .           .
Ad3 3218 AGTAGTTATTTTGCCTCCCCCTTCCCATTTAACAGAATACACCAATCTCT
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Ad7 3203 AGTAGTTATTTTGCCTCCCCCTTCCCATTTAACAGAATACACCAATCTCT
                .
Ad3 3268 CCCCACGCACAGCTTTAAA 3286
         ||||||||||||||||||
Ad7 3253 CCCCACGCACAGCTTTAAA 3271
```

Figure 2: Comparison of the predicted amino acid sequences for Ad2, Ad3, Ad5, and Ad7 fiber genes. The amino acid numbers for Ad2 are shown above the Ad2 sequence, and the amino acid numbers corresponding to Ad7 are shown directly below that sequence. Predicted amino acid sequences are shown in one-letter code. The amino acid sequences are aligned using the algorithm of Needleman and Wunsch (J. Mol. Biol. 48, 443-453, 1970). The predicted amino acid sequences from Ad2 are taken from Roberts et al. (Adenovirus DNA: The viral genome and its expression, (W. Doerfler, ed.), pp1-51, Martinus Nijhoff Publishing, Boston, 1986). The predicted amino acid sequences from Ad3 are taken from Signas et al. (J. Virol., 53, 672-678, 1985). The predicted amino acid sequences for Ad5 are taken from Chroboczek, and Jacrot (Virology, 161, 549-554, 1987). The predicted amino acid sequences for Ad7 were taken from Hong et al. (Virol, 167, 545-553, 1988).

COMPARISON OF FIBER PROTEIN SEQUENCES FROM SEVERAL SEROTYPES

```
                  10          20          30          40
                   +           +           +           +
AD2     .MKRARPSED TFNPVYPYDT EFGPPTVPFL TPPFVSPNGF QESPPGVLSL
AD5     .MKRARPSED TFNPVYPYDT ETGPPTVPFL TPPFVSPNGF QESPPGVLSL
AD3     MAKRARLS.T SFNPVYPYED E.SSSQHPFI NPGFISPDGF TQSPNGVLSL
AD7     MTKRVRLS.D SFNPVYPYED E.STSQHPFI NPGFISPNGF TQSPDGVLTL
                   +           +           +           +
                   10          20          30          40   :

       50          60          70          80          90
        +           +           +           +           +
AD2     RVSEPLDTSH GMLALKMGSG LTLDKAGNLT SQNVTTVTQP LKKTKSNISL
AD5     RLSEPLVTSN GMLALKMGNG LSLDEAGNLT SQNVTTVSPP LKKTKSNINL
AD3     KCVNPLTTAS GSLQLKVGSG LTVDTTDGS. .......... .....LEENI
AD7     KCLTPLTTTG GSLQLKVGGG LTIDDTDGF. .......... .....LKENI
         +           +           +                       +
        50          60          70                      80

      100         110         120         130         140
        +           +           +           +           +
AD2     DTSAPLTITS GALTVATTAP LIVTSGALSV QSQAPLTVQD SKLSIATKGP
AD5     EISAPLTVTS EALTVAAAAP LMVAGNTLTM QSQAPLTVHD SKLSIATQGP
AD3     KVNTPLTKSN .......... .......... .......... ..........
AD7     SAATPLVKTG .......... .......... .......... ..........
              +
             90

      150         160         170         180         190
        +           +           +           +           +
AD2     ITVSDGKLAL QTSAPLSGSD SDTLTVTASP PLTTATGSLG INMEDPIYVN
AD5     LTVSEGKLAL QTSGPLTTTD SSTLTITASP PLTTATGSLG IDLKEPIYTQ
AD3     .......... .......... .......... .......... ..........
AD7     .......... .......... .......... .......... ..........

      200         210         220         230         240
        +           +           +           +           +
AD2     NGLIGIKISG PLQVAQNSDT LTVVTGPGVT VEQNSLRTKV AGAIGYDSSN
AD5     NGKLGLKYGA PLHVTDDLNT LTVATGPGVT INNTSLQTKV TGALGFDSQG
AD3     .......... .......... .......... .......... ..........
AD7     .......... .......... .......... .......... ..........

      250         260           270         280         290
        +           +             +           +           +
AD2     NMEIKTGGGM RI..NNNLLI LDVDYPFDAQ TKLRLKLGQG PLYINASHNL
AD5     NMQLNVAGGL RIDSQNRRLI LDVSYPFDAQ NQLNLRLGQG PLFINSAHNL
AD3     .......... .......... .......... .......... ..........
AD7     .......... .......... .......... .......... ..........
```

Fiber amino acid sequence comparison

```
        300         310         320         330         340
          +           +           +           +           +
AD2     DINYNRGLYL FNASNNTKKL EVSIKKSSGL NFDNTAIAIN AGKGLEFDTN
AD5     DINYNKGLYL FRASNNSKKL EVNLSTAKGL MFDATAIAIN AGDGLEF..G
AD3     .......... .......... .......... ....HSINLP IGNGLQIEQ.
AD7     .......... .......... .......... ....HSIGLS LGPGLGTNE.
                                                     +
                                                    100

        350         360         370         380         390
          +           +           +           +           +
AD2     TSESPDINPI KTKIGSGIDY NENGAMITKL GAGLSFDNSG AITIGNKNDD
AD5     SPNAPNTNPL KTKIGHGLEF DSNKAMVPKL GTGLSFDSTG AITVGNKNND
AD3     .......... .......... ..NKLCS.KL GNGLTFDSSN SIALKN....
AD7     .......... .......... ..NKLCA.KL GEGLTFFNSN.NICIDDNI..
                                   +           +           +
                                  110         120         130

        400         410           420         430         440
          +           +             +           +           +
AD2     KLTLWTTPDP SPNCRI...H SD.NDCKFTL VLTKCGSQVL ATV.AALAVS
AD5     KLTLWTTPAP SPNCR...LN AE.KDAKLTL VLTKCGSQIL ATV.SVLAVK
AD3     .NTLWTGPKP EANCIIEYGK QN.PDSKLTL ILVKNGGIVN GYVTLMGASD
AD7      NTLWTGVNP TTANCQIMAS SESNDCKLIL TLVKTGGLVT AFVYV IGVS
               +           +           +           +           +
             140         150         160         170         180

          450         460         470         480         490
            +           +           +           +           +
AD2     GDLSSMTGTV ASVSIFLRFD QNGVLMENSS LKKHYWNFRN GNSTNANPYT
AD5     GSLAPISGTV QSAHLIIRFD ENGVLLNNSF LDPEYWNFRN GDLTEGTAYT
AD3     YVNTLFKNKN VSINVELYFD ATGHILPDSS SLK....TDL ELKYKQTADF
AD7     NDFNMLTTHK NINFTAELFF DSTGNLLTSL SSL...KTPL NHKSGQNMAT
               +           +           +                   +
             190         200         210                 220

                  500         510         520               530
                    +           +           +               +
AD2     N....AVGFM PNLLAYPKTQ SQTAKNNIVS QVYLHGDK.. ...TKPMILT
AD5     N....AVGFM PNLSAYPKSH GKTAKSNIVS QVYLNGDK.. ...TKPVTLT
AD3     S....ARGRM PSTTAYPFVL PNAGTHNENY IFGQCYYKAS DGALFPLEVT
AD7     GALTNAKGFM PSTTAYPFNV NSRE..KENY IYGTCYYTAS DHTAFPIDIS
          +           +           +           +           +
        230         240         250         260         270

            540         550         560                 570
              +           +           +                 +
AD2     ITLNGTSEST ETSEVSTYSM SFTWSWESGK Y.....TTET FATNSYTFSY
AD5     ITLNGTQE.T GDTTPSAYSM SFSWDWSGHN Y.....INEI FATSSYTFSY
AD3     VMLNKRLPDS RTSYVMTFL. ...WSLNAGL APE..TTQAT LITSPFTFSY
AD7     VMLNQRALNN ETSYCIRV.. ..TWSWNTGV APEVQTSATT LVTSPFTFYY
            +           +           +           +           +
          280         290         300         310         320
```

Fiber amino acid comparison

```
          581
            +
AD2      IAQE*
AD5      IAQE*
AD3      IREDD*
AD7      IREDD*
             +
           326
```

Figure 3: Conserved amino acid segments at N-terminus and C-terminus of the predicted amino acid sequences for Ad2, Ad3, Ad4, Ad5, and Ad7. Amino acid sequences were aligned using the algorithm of Needleman and Wunsch (J. Mol. Biol. 48, 443-453, 1970). The predicted amino acid sequences are represented by one-letter codes, with the consensus sequences written below the aligned sequences. The predicted amino acid sequences from Ad2 are taken from Roberts et al. (Adenovirus DNA: The viral genome and its expression, (W. Doerfler, ed.),pp1-51, Martinus Nijhoff Publishing, Boston, 1986). The predicted amino acid sequences from Ad3 are taken from Signas et al. (J. Virol., 53, 672-678, 1985). The predicted amino acid sequences for Ad4 are from Hong and Engler (unpublished data). The predicted amino acid sequences for Ad7 were taken from Hong et al. (Virology, 167,545-553, 1988)

```
A.  SEQUENCES AT N-TERMINUS
                  10                  20
                   +                   +
Ad2   .MKRARPSED..T..FNPVYPY..D..TE..T..GPP
Ad5   .MKRARPSED..T..FNPVYPY..D..TE..T..GPP
Ad3   MAKRARLS.T..S..FNPVYPY..E..DE..S..SSQ
Ad7   MTKRVRLS.D..S..FNPVYPY..E..DE..S..TSQ
CONCENSUS:      ..........(T/S)FNPVYPY(D/E).E(S/T)

B.  SEQUENCES FROM AMINO ACIDS 27-73
            30                  40                  50
             +                   +                   +
Ad2   PF..L..TPPF..V..SP..N..GF..Q....E..SPPGVL..S..LRVSEPL
Ad5   PF..L..TPPF..V..SP..N..GF..Q....E..SPPGVL..S..LRLSEPL
Ad3   PF..I..NPGF..I..SP..D..GF..T....Q..SPNGVL..S..LKCVNPL
Ad7   PF..I..NPGF..I..SP..N..GF..T....Q..SPDGVL..T..LKCLTPL
CONCENSUS:
      PF(L/I).P.F(V/I)SP(N/D)GF(T/Q)(Q/E)SP.GVL(S/T)L....PL

          60          70          73
           +           +           +
Ad2   DTSHGMLALKMGSGL..T.....L...D
Ad5   VTSNGMLALKMGNGL..S.....L...D
Ad3   TTASGSLQLKVGSGL..T.....V...D
Ad7   TTTGGSLQLKVGGGL..T.....I...D
CONCENSUS:
      .T..G.L.LK.G.GL(T/S)(L/V/I)D

C.  C-TERMINAL SEQUENCES

Ad2   WESGKYTTETFATNS..Y..TF..S..YI..A....Q.....E*
Ad5   WSGHNYINEIFATSS..Y..TF..S..YI..A....Q.....E*
Ad3   GLAPETTQATLITSP..F..TF..S..YI..R....E.....D....D*
Ad7   APEVQTSATTLVTSP..F..TF..Y..YI..R....E.....D....D*
Ad4   WTNG?YIGATFGANS..Y..TF..S..YI..A....Q.....Q*
CONCENSUS:
      ..............(Y/F)TF(S/Y)YI(A/R)(Q/E)(E/D/Q)(D)
```

Figure 4: Detailed comparison of Ad2, Ad3, and Ad7 nucleotide and predicted amino acid sequences in comparable ORFs in early region 3. ORFs designations are taken from Hong et al. (Virology, 167, 545-553,1988). The predicted amino acid sequences are represented by one-letter codes, and the amino acid sequences are aligned using the algorithm of Needleman and Wunsch (J. Mol. Biol. 48, 443-453, 1970). Matches between the serotypes are designated by a line between the aligned sequences. The predicted amino acid sequences from Ad2 are taken from Robers et al. (Adenovirus DNA: The viral genome and its expression, (W. Doerfler, ed.), pp1-51, Martinus Nijhoff Publishing, Boston, 1986). The predicted amino acid sequences from Ad3 are taken from Signas et al. (Gene, 50, 173-184, 1986). The predicted amino acid sequences for Ad7 were taken from Hong et al. (Virology, 167, 545-553, 1988).

ORF 1: 20.5 Kd (Ad3)/20.6Kd (Ad7):

```
                10         20         30         40         50
                 +          +          +          +          +
Ad3   MISTTIFIIT SLAAVTYGRS HLTVPVGSTC TLQGPQFGYV TWWRIYDNGG
      |||||||||| |||||||||| |||||||||| |||||| ||| ||||||||||
Ad7   MISTTIFIIT SLAAVTYGRS HLTVPVGSTC TLQGPQQGYV TWWRIYDNGG

                60         70         80         90         100
                 +          +          +          +          +
Ad3   FARPCDQPGT KFSCNGRDLT IINITSNEQG FYYGTNYKNS LDYNIIVVPA
      |||||||||| |||||||||| |||||||||| |||||||| | ||||||||||
Ad5   FARPCDQPGT KFSCNGRDLT IINITSNEQG FYYGTNYKDS LDYNIIVVPA

               110        120        130        140        150
                 +          +          +          +          +
Ad3   TTSAPRKSTF SSSSAKASTI PKTASAMLKL PKIALSNSTA APNTIPKSTI
      ||||||| || |||||||||| |||||||||| |||||||||  || |||||||
Ad5   TTSAPRKTTF SSSSAKASTI PKTASAMLKL QKIALSNSTA APKTIPKSTI

               160        170        180        189
                 +          +          +
Ad3   GIITAVVVGL IIIFFCIMYY ACCYRKHEQK GDALLNFDI*
      |||||||||| |||| ||||| |||||||||| |||||||||
Ad5   GIITAVVVGL IIIFLCIMYY ACCYRKHEQK GDALLNFDI*
```

183/189 amino acids match (97% homology)

ORF 2: 9.0Kd (Ad3)/7.7Kd (Ad7):
Mismatches between Ad3 and Ad7 are shown in bold face.

```
Ad3   MILFQSNT.. .......... .....TTSYA YTNIQPKYAM QLEITILIVI
      ||||||||              .      ||||| | ||||||||  ||||||||
Ad7   MILFQSNTTN TINVQTTLNH DMENHTTSYA YINIQPKYAM HLKITILIVI

Ad3   GILILSVILY FIFCRQIPNV HRNSKRSMSR PHMALNEI*
      |||||||||| | |
Ad7   GILILSVILY FLFSYD*
```

Matches excluding deletions/insertions: 43/49 amino acids (88% homology)
Matches including insertion in Ad7: 43/66 (65% homology)
No apparent amino acid homology to 11.6Kd (Ad2)

ORF 3: 10.4Kd (Ad2)/10.3Kd (Ad3)/10.3Kd (Ad7):
Mismatches between Ad2 and (Ad3 and Ad7) are shown in bold face.

```
Ad2   MIPRVLILLT LVALFCACST LAAVAHIEVD CIPPFTVYLL YGFVTLILIC
      ||||       |   |  | |    |||   | |  ||  | |   ||| |  |
Ad3   MIPRNFFFTI LICAFNVCAT FTAVATASPD CEGPFASYAL FAFVTCICVC
      ||||       |   |  | |    |||   | |  ||  | |   ||| |  |
Ad7   MIPRNFFFTI LICAFNVCAT FTAVATASPD CEGPFASYAL FAFVTCICVC
```

```
Ad2  SLVTVVIAFI QFIDWVCVRI AYLRHHPQYR DRTIADLLRI L*
     | |  |  |  |  ||  ||| ||||||| ||     | |||
Ad3  SIVCLVENFF QLVDWIFVRI AYLRHHPEYR NQNVAALLRL I*
     | |  |  |  |  ||  ||| ||||||| ||     | |||
Ad7  SIVCLVENFF QLVDWIFVRI AYLRHHPEYR NQNVAALLRL I*
```

Ad3 and Ad7 share 100% amino acid homology in this ORF.
Ad2 shares 45/91 amino acids (49% homology) with Ad3 and Ad7 in this ORF.

ORF 4: 14.5Kd (Ad2)/15.2Kd (Ad3)/14.9Kd (Ad7):

```
Ad2  ---MKRSVIF VLLIFCALPV LCSQTSAPPK RHISCRFTQI WNIPSCYNKQ
        |   ||   ||  |         | | |      | | |  |    ||
Ad3  MQAMLP-VIL ILLLPC-IAL ASTATRATPE QLRKCKFQQP WSFLDCYHEK
     |||||| ||| |||||| ||| ||||||||||| |||||||||| ||||||||||
Ad7  MQAMLP-VIL ILLLPC-IAL ASTATRATPE QLRKCKFQQP WSFLDCYHEK

Ad2  SDLSEAWLYA IISVMVFCST IFALAIYPYL DIGWNAIDAM NHPTFPVPAV
     ||    |           |  |   |||    | |   |    |  |
Ad3  SDFPTYWIVI VGIINILSCT FFSITIYPTF NFGWNSPNAL GYPQEPD.EH
     |||||||||| |||||||||| |||||||||| |||||||||| ||||||| ||
Ad7  SDFPTYWIVI VGIINILSCT FFSITIYPTF NFGWNSPNAL GYPQEPD.EH

Ad2  IPLQQV.IAP IN...QPRPP SPTPTEISYF NLTGGDD*
     ||||               |     |  |||| |||||||
Ad3  IPLQHIQQPL ALVQYENEPQ PSLPPAISYF NLTGGDD*
     |||||||||| ||| |||||| |||||||||| |||||||
Ad7  IPLQHIQQPL ALVEYENEPQ PSLPPAISYF NLTGGDD*
```

Ad3 shares 133/134 amino acids (99% homology) with Ad7 in this ORF.
Ad2 shares 46/131 amino acids (35% homology) with Ad7 in this ORF. (First 3 amino acids at N-terminal end were ignored in this comparison.)

ORF 5: 14.7Kd (Ad2)/15.3Kd (Ad3)/15.3Kd (Ad7)
A. homology between Ad3 and Ad7
Mismatches between Ad3 and Ad7 are shown in bold face.

```
Ad3  MTDPIATSST AAKELLDMDG RASEQRLIQL RIRQQQERAV KELRDAIGIH
     ||  | | |   |  |||||| | ||||| || ||||||||   ||||| | ||
Ad7  MTE.ILTTSN SAEDLLDMDG RVSEQRLAQL RIRQQQERVT KELRDVIQIH

Ad3  QCKKGIFCLV KQSKISYEIT ATDHRLSYEL GPQRQKFTCM VGINPIVITQ
     |||||||||| || ||||||| |||||||||| |||||||||| ||||||||||
Ad7  QCKKGIFCLV KQAKISYEIT ATDHRLSYEL GPQRQKFTCM VGINPIVITQ

Ad3  QSGDTKGCIQ CSCDSTECIY TLLKTLCGLR DLLPMN*
     |||||||||  ||||| || | |||||||||| ||||||
Ad7  QSGDTKGCIH CSCDSIECTY TLLKTLCGLR DLLPMN*
```

Ad3 shares 118/135 amino acids (87% homology) with Ad7 in this ORF.

B. Homology between Ad2 and Ad7
Mismatches between Ad2 and Ad7 are shown in bold face.

```
Ad2   MTE.SLD... ....LELDGI NTEQRLLERR KAASERERLK QEVEDMVNLH
      |||             |  ||    ||||   |       ||    |  |   ||
Ad7   MTEILTTSNS AEDLLDMDGR VSEQRLAQLR I.RQQQERVT KELRDVIQIH

Ad2   QCKRGIFCVV KQAKLTYEKT TTGNRLSYKL PTQRQKLVLM VGEKPITVTQ
      ||| |||| | ||||  || |  |  |||| |   ||||   | ||  ||  ||
Ad7   QCKKGIFCLV KQAKISYEIT ATDHRLSYEL GPQRQKFTCM VGINPIVITQ

Ad2   HSAETEGCLH FPYQGPEDLC TLIKTMCGIR DLIPFN*
       |  | || |       |    || || || | || | |
Ad7   QSGDTKGCIH CSCDSIECTY TLLKTLCGLR DLLPMN*
```

Ad2 shares 65/135 amino acids (48% homology) with Ad7 in this ORF.

Figure 5: Comparison of the early region 4 sequences from Ad2 and Ad7. The amino acid sequences were aligned using the algorithm of Needleman and Wunsch (J. Mol. Biol. 48, 443-453, 1970). The amino acids are represented by a one-letter code, and matches are designated by a line between the aligned sequences. Predicted amino acid sequences for Ad2 are taken from Roberts et al. (Adenovirus DNA: The viral genome and its expression, (W. Doerfler, ed.), pp1-51, Martinus Nijhoff Publishing, Boston, 1986), and represented sequences are encoded starting at nucleotide 33132 of the r-strand. The predicted amino acid sequences for Ad7 were taken from Hong et al. (Virology, 167, 545-553, 1988), and represented sequences are encoded in the r-strand starting at nucleotide 3409.

```
     -20
Ad2   *AWTSPSPPVEQPQVGQQPVA
     -29
Ad7   *DTIARFSFRIMIPIVLLPDDSDPDPGVDP

     1
Ad2  QQLDSDMNLS ELPGEFINIT DERLARQETV WNITPKNMSV THDMMLFKAS
     ||||    |    |  |  || | |||| ||| ||   | ||        |||
Ad7  QQLDCESALK DYPDAFLCIT DPRLARFETV WNVESKTMSI CNGIQMFKAV

                              72
Ad2  RGERTVYSVC WEGGGRLNTR VL*
     |||| |||| ||||| || |
Ad7  RGERLVYSVK WEGGGKITTR IL*
```

Ad2 shares 41/72 amino acids (57% homology) with Ad7 in the region of this ORF starting at amino acid.1 and ending at 72.

APPENDIX B

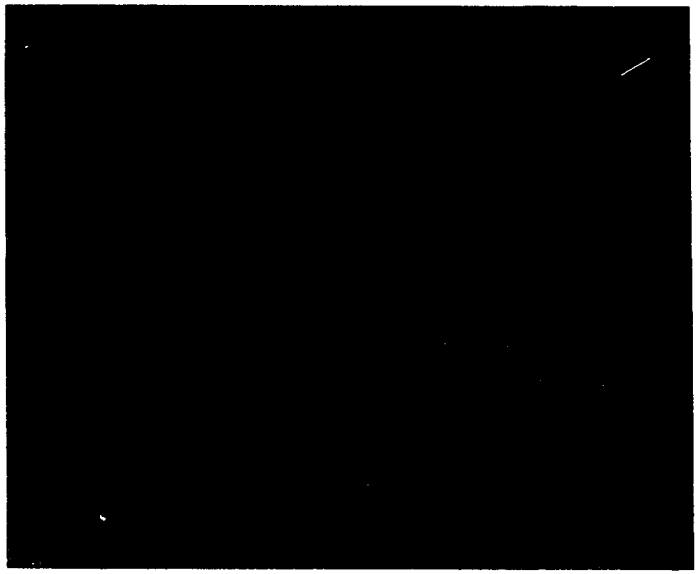
UNPUBLISHED WORK RELATING TO THE STUDY OF THE ROLE OF O-LINKED N-ACETYLGLUCOSAMINE IN NUCLEAR TRANSPORT

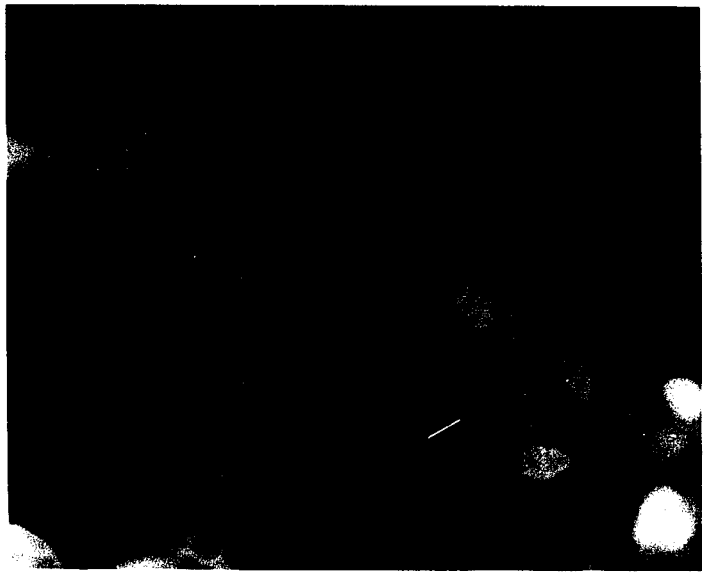
The following work was mentioned as data not shown in the manuscript "Nuclear transport of an adenovirus fiber mutant that lacks O-linked N-acetylglucosamine." Figure 1 shows the transport of wild type Ad7 fiber using a monoclonal antibody directed against fiber monomers and trimers. This serotype lacks glycosylation, and further supports our assertion that the GlcNAc is not necessary for nuclear transport of fiber protein. Figure 2 shows that fiber is present in cell extracts from cells infected with Ad5 or H5ts142 at 32°C or 39.5°C. This result is consistent with our finding that H5ts142 is made at 39.5°C but does not contain O-linked GlcNAc residues.

Figure 1: Nuclear localization of Ad7 wild type fiber. Indirect immunofluorescence of virus-infected A549 cells was determined at 48 hours post infection, using monoclonal antibody 4D2-5, and compared to uninfected cells.

Uninfected
A549 cells

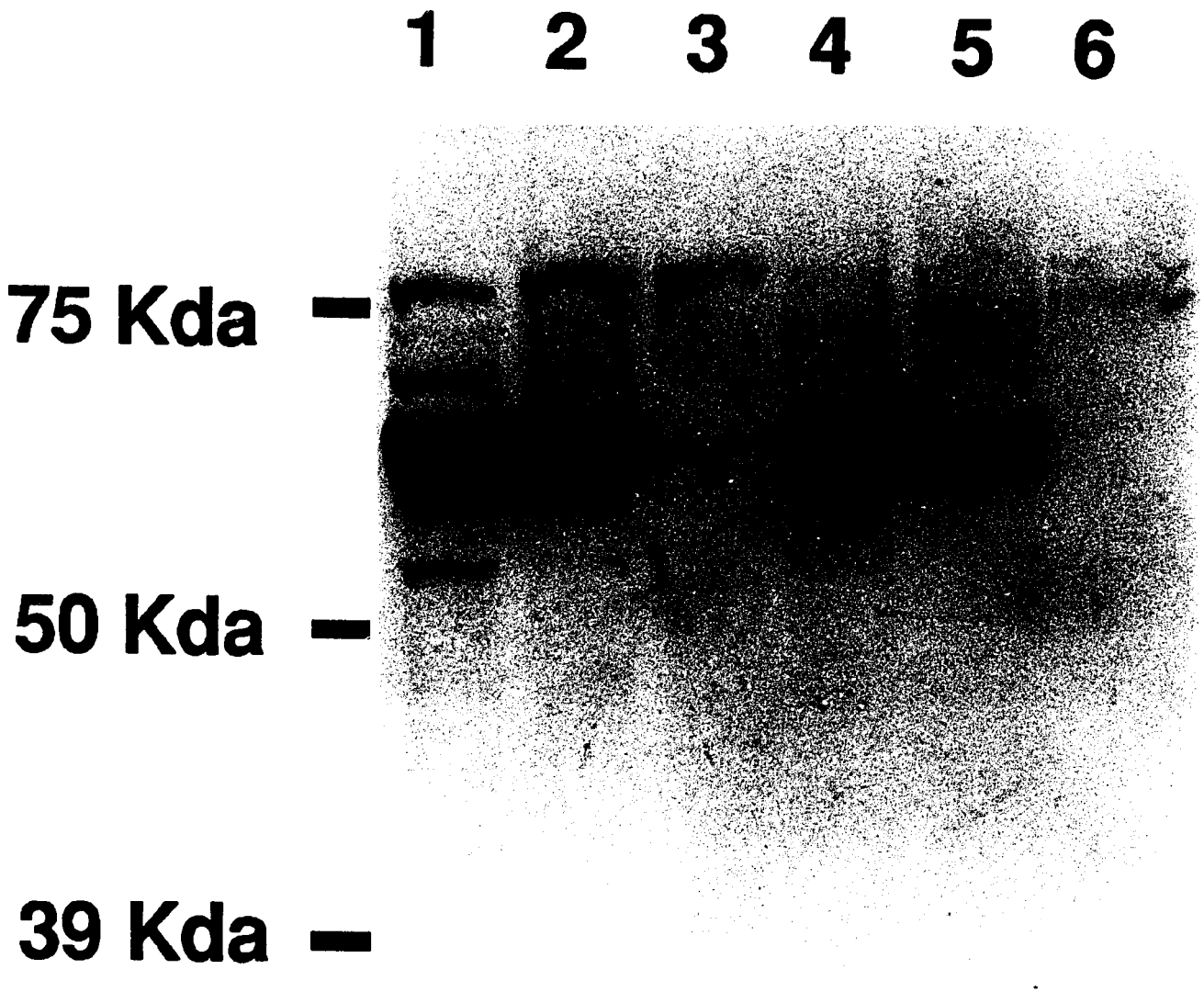
Ad7





Anti Fiber

Figure 2: Western blot of H5ts142, Ad5, and uninfected cell lysates. Characterization of Ad5 and of H5ts142 fiber from cell lysates. Panel A: Western blot developed with monoclonal antibody 4D25. Size markers are shown at the left of the figure. Lanes 1 through 3 are lysates from infected HeLa cells at 32°C; Lanes 4 through 6 are lysates from infected HeLa cells at 39.5°C. Lanes 1 and 4 are samples from wild type Ad5 infected cells; Lanes 2 and 5 are samples from H5ts142-infected cells; Lanes 3 and 6 are from uninfected HeLa cells.


1 2 3 4 5 6
75 Kda
50 Kda
39 Kda

APPENDIX C

NUCLEOTIDE SEQUENCE OF THE GENES ENCODED IN EARLY REGION 2B OF HUMAN ADENOVIRUS TYPE 12

by

LIMING SHU, JEONG SHIN HONG, YING-FEI WEI,
AND JEFFREY A. ENGLER

Gene, 46:187-195, 1986.

AUTHORSHIP

This work was a collaborative effort. A part of the sequencing work was accomplished by Dr. Jeffrey A. Engler, Liming Shu, and Ying-fei Wei. In addition, the sequence analysis was performed by Dr. Jeffrey A. Engler.

ABBREVIATIONS

aa, amino acid(s); Ad2, Ad7, Ad12, adenovirus serotypes 2, 7, and 12 (respectively); Ad pol, adenovirus DNA polymerase; bp, base pair(s); nt, nucleotide(s); ORF, open reading frame; pollk, Klenow (large) fragment of E.coli DNA polymerase I; pTP, adenovirus preterminal protein; URF, unidentified ORF.

ABSTRACT

The nucleotide (nt) sequence of a cloned DNA segment containing the early 2b region of the class A adenovirus Ad12 has been determined. When compared to the corresponding region of Ad2 or Ad7, there is a high degree of nt and predicted amino acid (aa) sequence homology within the r-strand regions that encode the preterminal protein and the viral DNA polymerase. A gene coding region comparable to the M_r 13600 gene product found in Ad2 can be identified; this hypothetical gene product shares 30% aa homology with its Ad2 counterpart and has a very similar hydropathy profile.

INTRODUCTION

At late times during adenovirus infection of permissive human cells, many different mRNAs are synthesized. Most of these molecules arise from processing of rightward RNA transcripts that originate at the major late promoter at genome coordinate 16.5 (genome coordinate is the distance from the left end of the genome, measured in %). Three noncontiguous leader segments (encoded at genome coordinates 16.6, 19.6, and 26.6) are found at the 5' ends of these transcripts and are spliced to alternative

main-body coding regions. At intermediate times in infection, an extra leader segment ('i') is found between the second and third leader segment; in Ad2, this extra segment has been shown to encode a stable perinuclear polypeptide of unknown function (Virtanen et al., 1982; Lewis and Anderson, 1983; Lewis et al., 1985; Symington et al., 1986). An M_r 13,500 polypeptide has also been mapped to this region of Ad2, based on in vitro translation of mRNA selected by hybridization to M13 clones containing Ad2 DNA from between the first and second leader segments (Lewis and Anderson, 1983; Lewis et al., 1985). The VA RNA genes, important for the efficient translation of adenovirus mRNA, have also been mapped on the same strand, near genome coordinate 30 (Thimmappaya et al., 1982).

In addition to mRNA transcripts from the late promoter, the main body regions of three low abundance l-strand mRNAs have been mapped to this region (Stillman et al., 1981). These early region 2b mRNAs are transcribed from the early region 2 promoter at genome coordicate 75 and share leader segments at coordinates 75, 68, and 39 before splicing to one of three possible main-body splice points at coordinates 30, 26, or 23. Two polypeptides required for adenovirus DNA replication are encoded in this region: the 80-kDa pTP and the 140-kDa Ad pol (summarized in Sussenbach and Van der Vliet, 1983; Kelly, 1984).

In previous papers, the nt sequences of the E2b region of class C (Alestrom et al., 1982; Gingeras et al., 1982; Dekker and Van Ormondt, 1984) and class B adenoviruses (Engler et al., 1983) have been reported. In this paper, the nt sequence of the E2b region of a class A adenovirus Ad12 is presented. An extensive analysis and comparison of genetic information contained in this region shows that the DNA replication polypeptides of the different serotypes exhibit a high degree of aa homology, as expected for proteins playing an essential role in the viral life cycle. A possible 'i' leader

polypeptide is also present, but is not as highly conserved in aa sequence as the DNA replication polypeptides. In addition, there appears to be only one VA RNA gene in this region of Ad12.

MATERIALS AND METHODS

(a) Preparation of DNA

Plasmids Ad12RID, Ad12HINF, and Ad12BAMG contain the Ad12 EcoRI D, HindIII F, and BamHI G fragments of Ad12 (Vogel et al., 1981). The EcoRI B fragment of Ad12 was subcloned after digestion by EcoRI + HindIII to give plasmid Ad12VA7 which contains Ad12 DNA from genome coordinates 28.0-31.9. Each of these Ad12 fragments was cloned into M13-mWB2348 in both orientations (Barnes et al., 1983). Plasmid DNA was prepared for chemical sequencing by the alkaline lysis procedure described by Birnbaum and Doly (1979). Dideoxynucleotide sequencing was performed by the standard methods (Sanger et al., 1977; 1980; Biggin et al., 1983), using bacteriophage M13 as a vector for preparing template.

(b) Enzymes and buffers

Restriction endonucleases were purchased from New England Biolabs, from Bethesda Research Laboratories, or from Boehringer Mannheim Biochemicals. Polynucleotide kinase and the PolIk were purchased from Pharmacia PL Biochemicals. E buffer and TBE buffer used for gel electrophoresis are described in Maniatis et al. (1982).

(c) Nucleotide sequence analysis

Some preliminary sequencing results were obtained by the chemical degradation method, using 5'- or 3'-terminally labelled DNA fragments (Maxam and Gilbert, 1980). The remainder of the sequence was determined by dideoxynucleotide sequencing (Sanger et al., 1977; 1980). M13 phage containing either sonicated Ad12 DNA fragments (obtained from plasmid DNA) or directed deletions (Barnes et al., 1983) were used as

a source of single-stranded template DNA. Oligodeoxynucleotide primers were synthesized to fill gaps in the DNA sequence left after analysis of templates prepared by other methods. Each strand of the sequence was determined at least twice; approx. 64% of the sequence was determined by both chemical and dideoxynucleotide methods. The sequences obtained were recorded and analyzed by computer programs developed by Staden (1977; 1978) and by Devereux et al. (1984). Computer-generated alignments of sequences have been furnished to the referees and are available upon request.

RESULTS AND DISCUSSION

(a) The nucleotide and amino acid sequences

Fig.1 shows the nt sequence of the l-strand of the E2b region DNA, as determined from the Ad12 EcoRI D, EcoRI B, BamHI G, and the HindIII F fragments. A number of important regulatory elements, such as the major late promoter, the three late leader segments found at the 5' ends of most adenovirus late mRNAs, and the VA genes are found on this strand. The pTP and Ad pol genes are encoded on the complementary (r) strand. The region shows the same high degree of nt and predicted aa sequence homology to Ad2 as was shown previously for Ad7; when Ad2, Ad7, and Ad12 are compared over the region that encodes the pTP and the Ad pol genes, the nt homology is 63% (pTP) and 62% (Ad pol) and the corresponding aa sequence homology is 75% and 73%, respectively (not shown). The nt sequences that encode the first, second, and third leader segments of late adenovirus mRNAs can also be identified by analogy to Ad2, because of the high homology of this region of the genome (Fig. 1). By analogy to the size and location of the VA RNA genes of Ad2 and Ad7, Ad12 apparently has only one VA gene (VAI), probably located between nt 10259 and 10410 (Fig.1). This result confirms an earlier observation of Föhring et

Figure 1. The nucleotide sequence of the l-strand of Ad12 in the E2b region. The sequence is written from 5' (at nucleotide 4830) to 3'. The nucleotide numbers refer to the distance from the left end of Ad12; although the sequence for the IVa2 gene of the Ad12 is not complete, the distance between the end of Ad12 left end sequence (which includes the C-terminus of Ad12 IVa2 gene) and the sequence presented here was estimated, based on the high homology that exists in the IVa2 gene of Ad2, Ad5, Ad7, and Ad12 (where it is known). For this calculation, it has been assumed that no deletions or insertions have occured in the IVa2 gene of Ad12 relative to Ad2, Ad5, of Ad7. A portion of the sequence presented here has been reported previously (nt 5297-5665; Engler and Van Bree, 1982). Major restriction sites and predicted locatiohs for major splice sites, protein coding regions, and VA RNA gene are underlined.

```
      4840          50          60          70          80          90        4900          10
AAGGAAAGTA ACTCTTGTAA AGATTGCAAA TTTTTAAGTG GCTTTAGCCC ATCCGCCATA GGCATGTGGT CCAGGGTTTG

        20          30          40          50          60          70          80          90
CTTCAGCAGT TGCAAGCGAT CCCATAGCTC AGTTATATTT TCTATGCCAT CTCGATCCAG CAAACTTCCT CGTTGCGGGG
                                              End of Ad pol reading frame (opposite strand)

      5000          10          20          30          40          50          60          70
GTTTGGCTGG CTGTTGCTGT AAGGAACGAG GCGGTGAGCA TCCAAATGGA CGAGGGTCAT GTCCTTCCAG GGACGTAATG

        80          90        5100          10          20          30          40          50
TGCGCGTCAG GGTTGTTTCG GTCACGGTGA ATGGATGCGC TCCTGGTTGA GCGCTGGCCA GTGTGCGCTT TAAACTGAGG
                                                          Bal I

        60          70          80          90        5200          10          20          30
CGGCTGGTGC TGAAGCGCGT GTCTTCTCCC TGTGCTTCGG CAAGGTAGCA TTTTAACATA AGATCATAAG ACAAAGCCTC
                                Acceptor splice site for IVa2? (opposite strand)

        40          50          60          70          80          90        5300          10
TGTAGCGTGG CCTTTAGCCC GTATTTTTCC TTTGGAGGTG CTCCCGCAGT GAGGACACTG AAGGCATTTA AGGGCGTACA

        20          30          40          50          60          70          80          90
GTTTTGGAGC CAAAAAAACA GATTCTGGAG AATAAGCATC TGCGCCACAA TAACTACAAA CAGTTTCACA TTCAACTGAC

      5400          10          20          30          40          50          60          70
                              Donor splice site for IVa2 mRNA? (opposite strand)
CAGGTCAGCT CGGGACATGA TGGATCAAAA AGAAGTTTCC CTCCGTACTT TTTGATGCGT TTCTTACCTT GCGTCTCCAT
                                                          Start IVa2 gene? (opposite strand)

        80          90        5500          10          20          30          40          50
AAGGCGGCGT CCTTTCTCTG TGACAAAAAG ACTGTCAGTG TCTCCGTATA CAGATTTAAG GGGTCTATCC TTGAGTGGTA

        60          70          80          90        5600          10          20          30
TTCCGCGGTC CTCCTCGTAC AGGAATTCTG ACCACTCTGA CACAAAAGCT CTAGTCCAAG CAAGTACAAA GGAAGCCACA
                          EcoR I

        40          50          60          70          80          90        5700          10
TGGGAAGGGT ACCGATCGTT GTTAATTAAA GGGTTAGAAC TTTCTAAGGT GTGTAAACAC ATGTCTCCTT CTTCAGCGTC
       Kpn I

        20          30          40          50          60          70          80          90
CATGAATGTG ATTGGTTTGT AGGTGTAAGT CACGTGTTCA CAATTTTCTG GTGGTGGGCT ATAAAAAGGG GCGGGTCCTT
                                                  "TATA" box for major late promoter?

      5800          10          20          30          40          50          60          70
GGTCTTCATC GCTTTCTTCT GCTTCGCTGT TTACGAGCGC CAACTGGTTG GGTGAGTACA CGCGCTCAAA GGCAGGCATT
            Start of first leader?                    Donor splice site for first leader?

        80          90        5900          10          20          30          40          50
ACCTCTGTAC TCAACGTGTC AGTTTCTATA AACGATGAGG ATTTGATGTT TAATCGCCCC GCTGCAATTT CTTTCATTAG

        60          70          80          90        6000          10          20          30
GCTTTCTTCC ATTTGATCAG AAAAAACTAT TTTTTTGTTA TCTAGTTTGG TAGCAAAAGA TCCGTACAAG GCATTGGAAA
              Bcl I

        40          50          60          70          80          90        6100          10
GCAGCTTGGC TATAGATCTT AGGGTTTGAT TTTTGTCCCT ATCGGCCCGT TCTTTTGCGG CAATATTGAG TTGCACATAT
              Bgl II

        20          30          40          50          60          70          80          90
TCGCGTGCCA GGCATTTCCA GGTGGGGAAA ATGGTGGTGC GCTCGTCAGA TAGCAAGCGT AAGCGCCACC CGCGATTATG

      6200          10          20          30          40          50          60          70
CAGTGTAACC AGATCTACGC TGGTAACTAC TTCACCGCGC AAGCTTTCAT TGGTCCAGGC TAAACGACCG CCTTTTCTAG
            Bgl II                            Hind III                                Xba I
```

```
        80          90        6300          10          20          30          40          50
AACAAAAAGG AGGAAGAACA TCCAACTGAT TTTCATCTGG GGGGTCGGCA TCTATAGTAA AAATGCCAGG ACAAAGATTT

        60          70          80          90        6400          10          20          30
TTGTCAAAAT AATCAATTTT GCAAGTGTAA TTTTCCAGCG CCACCTGCCA TTGCCGCACG GCGAATGCCC GCTCATAGGG

        40          50          60          70          80          90        6500          10
GTTAAGGGGA GGACCCCAAG GCATGGGGTG TGTGAGGGCC GATGCATACA TGCCGCAAAT ATCATATACA TATATGGGCT

        20          30          40          50          60          70          80          90
CTTTTAGTAC TCCTATGTAA GTAGGATAGC ACCTGCCGCC ACGAATGCTG GCGCGAACGT AGTCATATAG CTCATGTGAA

      6600          10          20          30          40          50          60          70
GGCGCCAGGA TGTTGGGCCC AAGATGTGTG CGCTGTGGTT TTCGGCGCG GTACAAAATT TGTCTGAAAA TTGCATGAGA

        80          90        6700          10          20          30          40          50
GTTAGAGGAA ATGGTAGGAC GCTGAAACAC ATTAAAATGT GCCGCGTCAA GACCCACTGC GTCAGTAACA AACTGGGCGT

        60          70          80          90        6800          10          20          30
ATGACGTACG CAGTTTTTCT ACCAATGAGG CAGTCACAAG TACATCCAGG GCACAATAGT TTAATGTTTC CCCGATAAGA

        40          50          60          70          80          90        6900          10
TTGTAATTTT TTTCTCCTTT TTTTTTCCAT AGTTCTTGAT TTAGGAGGTA TTCCTCCTTA TCCTTCCAGT ACTCCTCCAG
                                  Acceptor splice site for second leader?

        20          30          40          50          60          70          80          90
GGGAAACCCG TTTGCATCTG CACGGTAAGA ACCAAGCATA TAAAACTGAT TTACCGCCTT GTACGGACAA CATCCTTTTT
                          Donor splice site for second leader?

      7000          10          20          30          40          50          60          70
CTACAGGCAG GGCATACGCT TGTGCAGCCT TTCTTAAAGA TGTATGAGTA AGAGCAAAGG TATCTCTGAC CATTACTTTT

        80          90        7100          10          20          30          40          50
AAATACTGGT ATTTAAAATC TTGGTCGTCA CACCCTCCGT GTTCCCACAG TAGGAAGTTA GTTCGCTTTT TGTAGTGGGG

        60          70          80          90        7200          10          20          30
ATTGGGAAGG GCAAAAGTAA TATCATTAAA TAATATTTTG CCAGCTCTTG GAATAAAATT TCTAGAAATT TTAAAGGGTC
                                                                  Xba I

        40          50          60          70          80          90        7300          10
CAGGGACGTC CAAGCGGTTA TTGATTACCT GAGCGGCAAG AACAATTTCA TCAAATCCAT TAATATTGTG TCCTACTATA

        20          30          40          50          60          70          80          90
TACAACTCTA CAAATCTTGG CTCACCCTTA ATTGCAGGGG CTCTTTTAAG ATCTTCGTAG GAAAGATCTT CAAGCGCGAC
                                              Bgl II              Bgl II

      7400          10          20          30          40          50          60          70
TAGTCCGTTT TCTTCTTGAG CCCATTGAGA CAAGTGTGGA TTTTTTTGTA AAAAAGTCAT CCAAAGATCA GTAGCTAAGG

        80          90        7500          10          20          30          40          50
AGGTTTGTAA GCGGTTTCTA TAGGTACGAA ACTGTTGACC GACCTTCATT TTTTCTGGGG TTAAGCAGTA GAAAGTAGTA

        60          70          80          90        7600          10          20          30
GAGTCTTTTT CCCATTGGTC CCATCCAAGT TCTAATGCCA AGTTGTAAGC ATGTTTGACA AGATTGTCAT CCCCAGACAG

        40          50          60          70          80          90        7700          10
TTTCATCACC AGCATAAATG GGAGAAGTTG CTTTCCAAAT GCCCCCATCC AGGTGTAGGT TTCTACATCA TAGGTAATAA
                                                  Acceptor splice site for "i" leader?

        20          30          40          50          60          70          80          90
AAAGGCGCTC AGTGCGAGGA TGCGAACCGA TTGGGAAAAA GTGGATCTCC TGCCACCAGT TGGAAGAATG GCTGTTGATG
                    Start of "i" leader gene product?

      7800          10          20          30          40          50          60          70
TGAGTAAAGT AGAAATCTCG TCGGCGGACA GAGCATTCAT GCTGATGTTT GTAAAAGCGT GCGCAGTGTT CGCATCGTTG
```

```
        80          90        7900          10          20          30          40          50
CACGGGCTGT ATCTGTTGAA TGAGGTGTAC CTGGCGGCCT CGCACCAGAA AGCAGATGGG AAAATCAATA CCACTTGGCA

        60          70          80          90        8000          10          20          30
GCTGCCGTTC GTCCTCTTCC TCTTCTGCTG CATTGCCACT ACCGTTTGGA TCCTCGAAAG CGAGAACGGA GAGGGTGACG
Pvu II                                                 BamH I

        40          50          60          70          80          90        8100          10
GTGCCCCTCG AGCTGCATGT CCAGCTTTCA GCACGAGAGG GGCGGAAACG GGAAATCAGG GCGTACAGCC TGGAGCTGTC
      Xho I

        20          30          40          50          60          70          80          90
CATGGTATCA GTCAGAGAGA AAAGCATGTC CGCGGGGACA GCGCGCAAGT TGACTTCGCA CAGGCGGGTA AGAGCAGGCT
Nco I

      8200          10          20          30          40          50          60          70
        Donor splice site for "i" leader?
GGAGGTGCAG GTAATACTTA ATTTCTAGAG GCGTGCCGTT GGCAGAGTCT ATTGCGTGAA GTATTCCGTG AACCCGGGAA
                        Xba I                                                       Sma I

        80          90        8300          10          20          30          40          50
                                                End of pTP reading frame (opposite strand)
CCAACCACGG TTCCACGGTG CACTTTTCCA ATGCGCCTGC TTAAAATCGG CGGCGCGGAC GAGCTCCCGG AGGAAGCGGC
                                                                    Sac I

        60          70          80          90        8400          10          20          30
GGTTCGGGTC CTGCGGGAAG CGGGGGAAGC GGTATGTCGG CCTGACGCTC TGGCAGGGGA AGGTGTTGAG CCCGAAGTTG

        40          50          60          70          80          90        8500          10
ACTGGCATGG GCGACTACCC GGCGATTGAT ATCTTGAATC TGTCGGCGTT GTGTAAACAC TACCGGCCCT GTTGTTTTGA
                              EcoR V

        20          30          40          50          60          70          80          90
  Acceptor splice site for Ad pol mRNA? (opposite strand)
ACCTGAAAGA AAGTTCAACA GAATCAATCT CAGTGTCATT TACTGCAGCC TGTCTTAAAA TCTCCTGAAC GTCGCCTGAG
                                                Pst I

      8600          10          20          30          40          50          60          70
TTATCTTGGT AGGCAATTTC TGCCATTAAT TGATCAATTT CTTCCTCCTG GAGGTCTCCA TGTCCCGCAC GTTCAATAGT
                                  Bcl I

        80          90        8700          10          20          30          40          50
GGCTGCAAGG TCATTAGATA TCCGACTCAT AAGCTGTGAA AATGCGTTTA GTCCAATTTC GTTCCAGACT CGGCTGTATA
                EcoR V

        60          70          80          90        8800          10          20          30
CTACCCCTCC TTCGCTGTCC CGAGCGCGCA TAACCACTTG CGCCAAGTTG AGTTCCACGA GCCGTGCGAA CACGCCGTAG

        40          50          60          70          80          90        8900          10
TTGCGCAAGC GCTGAAACAG GTAGTTTAAG GTGGTGGCAA CGTGTTCTGA GACGAAGAAA TACAGAATCC ACCGACGAAG

        20          30          40          50          60          70          80          90
CGTCACGTCG TTGATGTCAC CTAAGGCTTC AAGACGTTCC ATGGCTTCGT AAAAGTCTAC TGCAAAATTG AAAAACTGGG
                                              Nco I

      9000          10          20          30          40          50          60          70
AGTTGCGAGC TGCCACCGTC AATTCTTCTT CCAACAGACG AATAAGCTCG GCCACCGTCT CGCGCACTTC TTGCTGAAAT

        80          90        9100          10          20          30          40          50
GCGCCCGGAA CTATTTCTTG TTCTTCCTCT TCTACCTCCA TTATTTCTTC CTCGACCACA GGTGGTGGGG GTTGTCTTCT

        60          70          80          90        9200          10          20          30
TCGACGCCGG CGAACGGGCA GCCTGTCTAC AAATCTTTCA ATCATTTCGC CGCGACGGCG GCGCATAGTT TCGGTTACTG

        40          50          60          70          80          90        9300          10
CTCGACCGTT TTCACGTGGT CGTAACTCAA AAACTCCACC TCTAAGTTCT GTTTCATGTA AAATGGGAAA TGAGGCGTTG
```

```
        20          30          40          50          60          70          80          90
                                                                 End of "i" leader gene product?
CGAGGGGCGT TAGGTAGGGA TACAGCGCTG ATTATGCATT TTATTATTTG CTGCGTAGGA ACTCCGCGCA AGGAGCTAAA
                              Acceptor splice site for third leader?

      9400          10          20          30          40          50          60          70
CGTCTGCATA TCCACCGGGT CGGAGAACCT TTCAAGAAAG GCATCTAGCC AGTCACAGTC ACAAGGTAGG CTAAGTTTTG
                                                          Donor splice site for third leader?

        80          90        9500          10          20          30          40          50
TTTCTTCTAA AGTACCAGGA AGCTGAGCAA TGCTACTAAT AATGTAATTG AAGTAAGCTG TTTTAAGCCC ACGAATGGTT

        60          70          80          90        9600          10          20          30
TTAAGAAGCA CCACATCTTT GGGTCCGGCT TGTTGAATTC GCAGGCGGTC TGCCATTCCC CACACGTCAC TTTGACATCG
                                    EcoR I

        40          50          60          70          80          90        9700          10
TCCAAGATCT TTGTAGTAGT CTTGCATTAA CCTTTCCACC TCTACCTCGC GGTTTCCGCG ATCAGCCATG TGCGTGCTTC
    Bgl II

        20          30          40          50          60          70          80          90
CGTAGCCTTG CAGCGGTTGT AATAAAGCTA AATCTGCCAC TACCCGTTCC GCAAGCACTG CCTGTTGAAT TTGGGTAAGG

      9800          10          20          30          40          50          60          70
GTGGTTGCAA AGTCATCCAC ATCTACAAAG CGGTGATAAG CTCCTGCATT AATGGTGTAG CTGCAGTTTG TCATTACTGA
                                                                    Pst I

        80          90        9900          10          20          30          40          50
CCAATTAACA GTTTGCGTGC CTGGCTGTAC AGTTTCTGTG TATCGCAAGC GTGAGTAAGC CCGAGAGTCA AAAACATAGT

        60          70          80          90       10000          10          20          30
CATTGCAGGT GCGCACTAGG TATTGATAGC CCACAAGGAA ATGAGGAGGA GGTTCGCGAT ACAACGGCCA GCCAAGCGTA

        40          50          60          70          80          90       10100          10
GCCGCAGCAC CTGGAGCGAG ATCTTCCAAC ATGAGGCGGT GGTATTCATA TATGTATCTG GACATCCATG TGATGCCGGC
                  Bgl II

        20          30          40          50          60          70          80          90
AGCGGTAGTT GTTGCTCGCA TAAATTCGCG GGCTCGGTTC CAAATATTGC GCAGGGGTAA AAAGCGTTCA ATAGTTGCCA

      10200          10          20          30          40          50          60          70
                                                                              Start of VAI?
CGCTTTGACC GGTCAGGCGT GCGCAGTCTT GAATGCTCTG GACATGGAAA AAATGAAAGT CGGTAAGCGA CTCCCTTCCG
                                    Acceptor splice site for pTP mRNA? (opposite strand)

        80          90       10300          10          20          30          40          50
TGGTTTGGTG GAAAAGTCAC AAGGGTACCA TAGCGAGGAA CCCCGGTTCG AAACCGGCAG GATCCGCTAT GAGCACAAGT
                          Kpn I                                          BamH I

        60          70          80          90       10400          10          20          30
                                                                 Acceptor splice site for 52, 55K mRNA?
GAGGCGCTTG CGCGTTGAAC CCGGCCAAGG ACCCCCAGAC ACGGAGAGGA GTCTTTTTTT TATTTATTTT TTCTTAGATG
                                                             End of VAI?

        40          50          60       10470
CATCCTGTCC TGCGACAAAT GCGACCTCAG CCCAGGGCAA
```

al., (1979) who detected only one VA RNA species in Ad12. The predicted RNA can be folded so that it has substantial secondary structure, as has been shown for the VA RNAs from Ad2 (Akusjärvi et al., 1980).

(b) Coding

Fig.2 shows the protein-coding potential of both strands of the DNA sequence. As has been shown for Ad2 and Ad7, the IVa2 gene of Ad12 apparently overlaps the coding region for Ad pol in a different reading frame; in this overlap interval, the Ad pol has been conserved (when compared to Ad2 and/or Ad7) at the expense of IVa2. URFs identified in Ad2 and/or Ad7 (in particular, the URF that encodes the M_r 13,500 polypeptide in Ad2, near genome coordinate 17) are not apparent in the sequence of Ad12.

Fig.3a shows a comparison of the protein-coding capacity for serotypes Ad2, Ad7, and Ad12 in the region that has been shown to encode the 'i' leader polypeptide (Virtanen et al., 1982). In Ad2, this polypeptide is found in the perinuclear region and accumulates starting about 8h after infection (Lewis et al., 1985; Symington et al., 1986); the function of this gene product is not known. The presence of this polypeptide in Ad7 or Ad12 infections has not yet been demonstrated. Although the nt sequence of this region is less highly conserved when compared to Ad2 and Ad7 (not shown), the region that encodes this 'i' leader polypeptide is always retained. When the predicted amino acids are compared by the algorithm of Needleman and Wunsch (1970), there is about 30% aa sequence homology (not shown). Further, there is a remarkable correspondence in the plots of the hydropathy index (Kyte and Doolittle, 1982) for the predicted aa sequence in each serotype, suggesting that the corresponding parts of the sequences of the polypeptides are hydrophilic or hydrophobic.

Figure 2. The protein-coding capacity of the E2b region of Ad12. The position of stop codons in each potential reading frame is shown as vertical lines descending from the horizontal. The ORFs predicted to encode the preterminal protein (pTP), the virus polymerase (Ad pol), the IVa2 gene, and the potential reading frame analogous to the 'i' leader of Ad2 are all shown as rectangles; regions of high homology in the predicted aa sequences of Ad2, Ad7, and Ad12 are shown as filled rectangles.

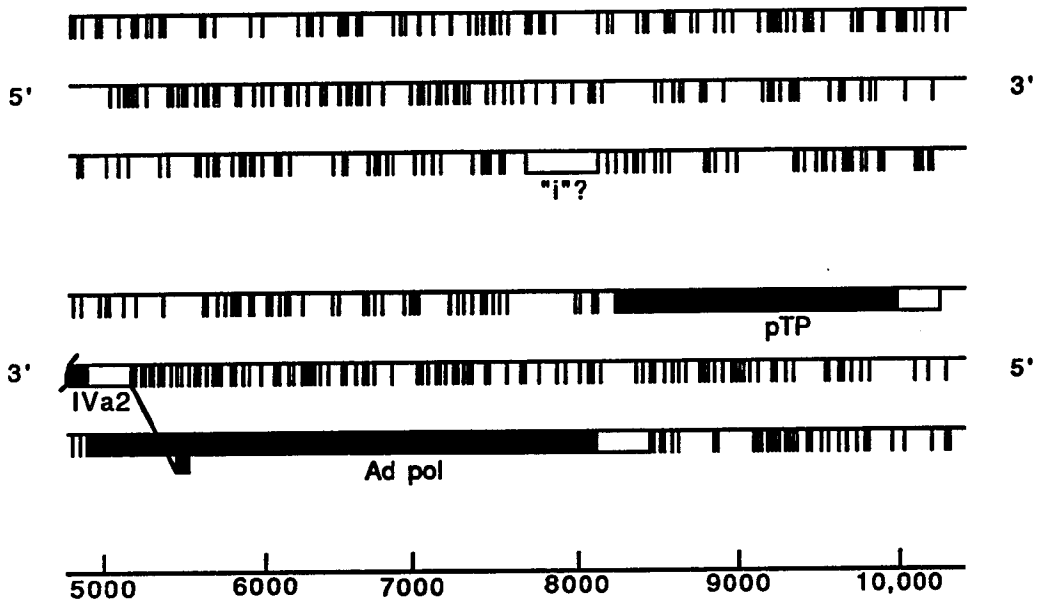
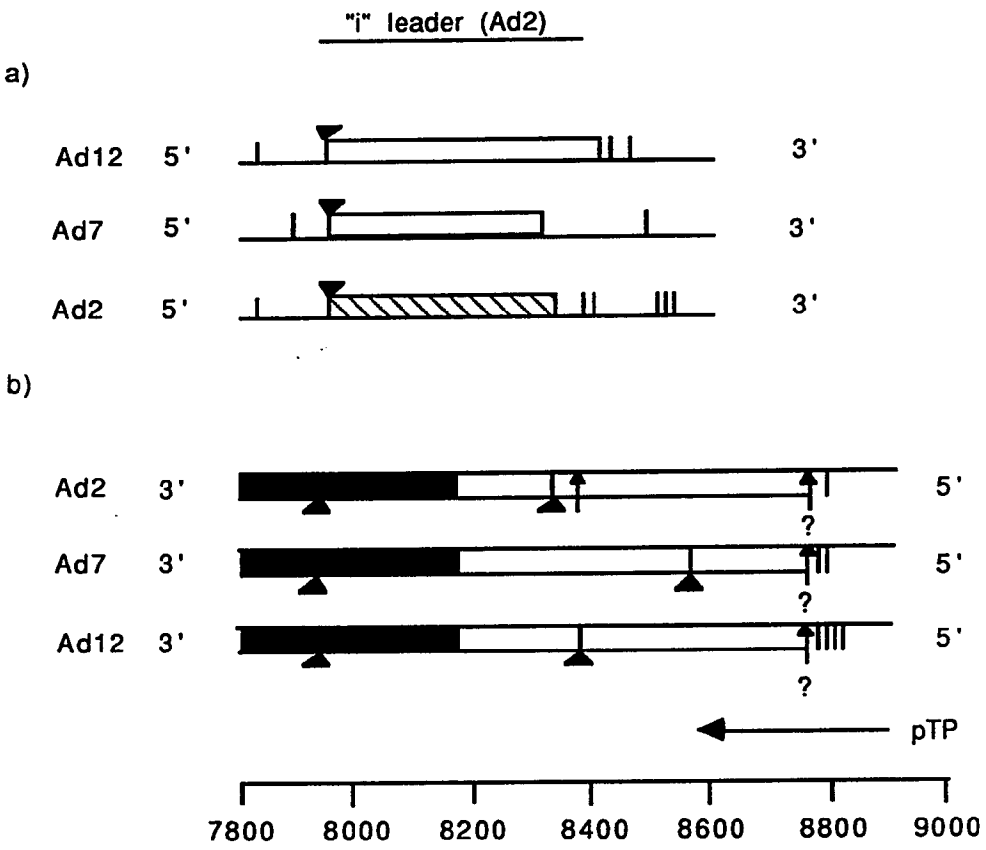

5'
3'
"i"?
pTP
3'
5'
IVa2
Ad pol
5000
6000
7000
8000
9000
10,000
Distance from left end of Ad12 (bp)

Figure 3. Comparison of the protein-coding potential of Ad2 (class C), Ad7 (class B), and Ad12 (class A) in the region that encodes the M_r 13600 protein (a) and the Ad Pol (b) protein products of Ad2. Vertical lines show the positions of termination codons in the same frame. Vertical lines with filled triangles show the positions of the initial ATG codons in the frame. The highly conserved region of the Ad pol gene sequences is shown as a filled rectangle; the position of the first ATG in the highly conserved portion of the gene is shown with a filled triangle. The position of the acceptor splice site for Ad pol mapped by Dekker and Van Ormondt (1984) is shown as a vertical arrow; the positions of the putative splice sites that might also be used (see RESULTS AND DISCUSSION, section b and Fig. 4) are shown as vertical arrows with a question mark. The C-terminal end of the pTP gene is shown with a horizontal arrow.


"i" leader (Ad2)
a)
Ad12 5' 3'
Ad7 5' 3'
Ad2 5' 3'
b)
Ad2 3' 5'
Ad7 3' 5'
Ad12 3' 5'
?
?
?
pTP
7800
8000
8200
8400
8600
8800
9000
Distance from Ad2 left end (base pairs)

Analysis of the protein-coding potential of the r-strand of Fig.3b and comparison of the nt sequences in this same region show that although the sequence of DNA of the three adenovirus serotypes is less highly conserved between Ad12 nucleotide 7934 to 8311, no termination codons have occurred in this interval. Further, the number of nt found in this less conserved region in each serotype is always a multiple of three, suggesting that a protein might be encoded in this region. The acceptor splice point for the Ad pol has been mapped to this region by electron microscopy (in Ad2; Stillman et al., 1981) or by S1 nuclease mapping (at nt 8418 in Ad5; Dekker and Van Ormondt, 1984). There is no comparable consensus splice site in Ad7 and Ad12 near the position mapped with S1 in Ad5 (Fig.4a). However, a consensus splice acceptor sequence can be found at nt 8513 in the Ad12 sequence (nt 8775 in Ad2 and 8625 in Ad7; Fig.4b). This possible splice acceptor has several attractive features: (1) it is highly conserved at the nt level in all three serotypes; (2) after RNA splicing, both this putative acceptor as well as the probable pTP acceptor could connect the Ad pol or the pTP reading frames to the same reading frame encoded in the leader segment at coordinate 39; (3) it would explain why the nt sequences do not contain termination codons and are always a multiple of 3 nt long. It is not clear whether or not this putative splice point is the site mapped by electon microscopy to genome coordinate 23 or 26 (Stillman et al., 1981). Use of this splice site to encode Ad pol would mean that the end of the pTP gene and the beginning of the Ad pol gene would overlap in different reading frames, a situation similar to that predicted for IVa2 (Engler and Van Bree, 1982; Gingeras et al., 1982; Alestrom et al., 1982; this work).

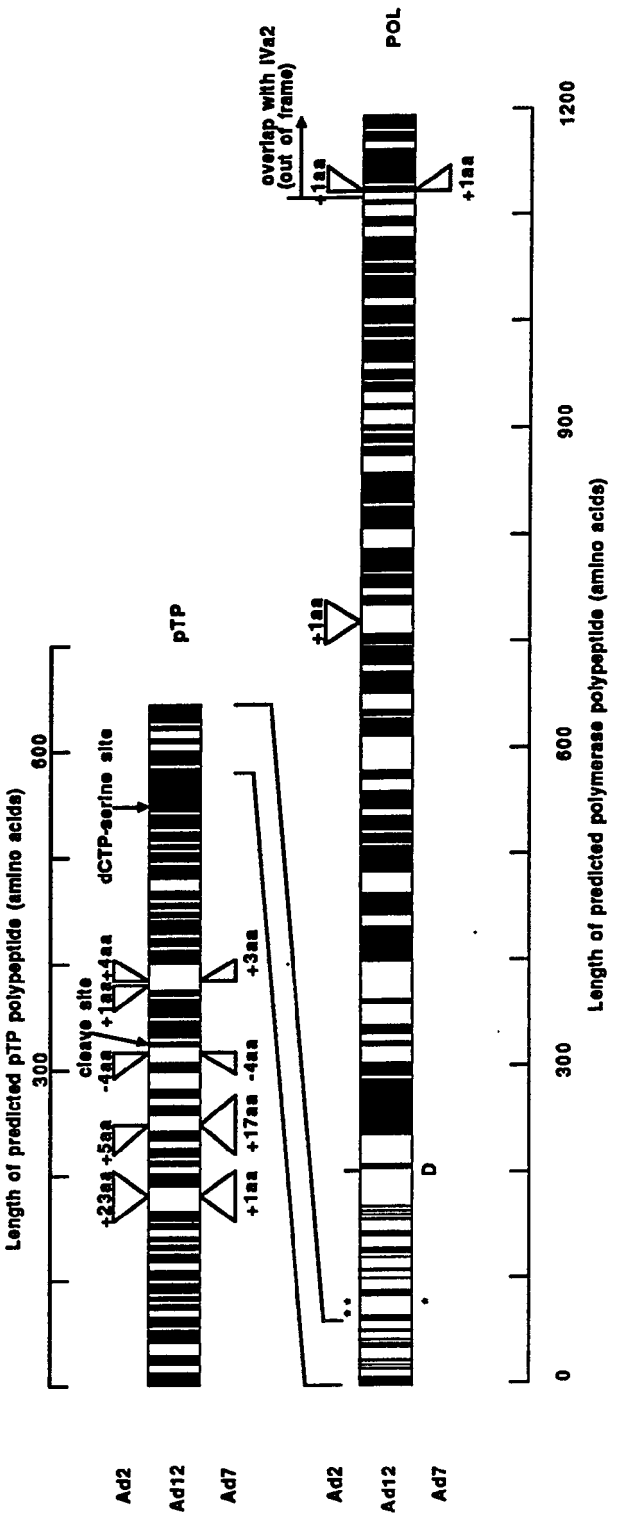
(c) Comparison of the predicted pTP and Ad pol polypeptides

Fig.5 shows a comparison of the predicted aa sequences of the pTP and Ad pol polypeptides in Ad2, Ad7, and Ad12; in this figure, the predicted

Figure 4. Comparison of predicted acceptor splice sites for the Ad Pol mRNA. (a) The acceptor splice site for Ad pol mapped in Ad5 by Dekker and Van Ormondt (1984) compared to the corresponding sequences from Ad7 and Ad12. The three DNA sequences in this less highly conserved region were first aligned using the algorithm of Needleman and Wunsch (1970), prior to the comparison shown. (b) Comparisons of the putative acceptor splice site for Ad pol mRNA, identified by its high homology in Ad2, Ad7, and Ad12. Use of this splice point for Ad pol mRNA would result in overlap of the coding frames for pTP and Ad pol. (c) Comparisons of the putative acceptor splice points for pTP mRNA, identified by the abrupt appearence of high nt homology near genome coordinate 29 in Ad2, Ad7, and Ad12.

```
                                                8414
a)  AD5:  CTAGCCCGCG CCCTGACCCG GCTATGCGAG GTA AAC CTG CAG
                                                V   N   L   Q
                                                8256
    AD7:  CTCCCTCACG CCCTCGAAAA GATCATGCAA GTC AAT CTC CTG
                                                V   N   L   L
                                                8165
   AD12:  CTCCAGCCTG CTCTTACCCG CCTGTGCGAA GTC AAC TTG CGC
                                                V   N   L   R

                                                8775
b)  AD2:  CGAAATTGAT TCTGTCGAAC TCTCTTTCAG GTT CAA GCT CAC
                                                V   Q   A   H
                                                8625
    AD7:  CGAGATTGAT TCTGTTGAAC TCTCTTTCAG GTT CAA GCT CAC
                                                V   Q   A   H
                                                8513
   AD12:  CGAAATTGAT TCTGTTGAAC TTTCTTTCAG GTT CAA AAC AAC
                                                V   Q   N   N

                                               10580
c)  AD2:  GCTTACAGGC TCTCCTTTTG CACGGTCTAG AGC GTC AAC GAC
                                                S   V   N   D
                                               10391
    AD7:  CATCGCTTTC CGTTTTCTCC GTGTCTACAG AGC GTC AAT GAC
                                                S   V   N   D
                                               10228
   AD12: CTTACCGACT TTCATTTTTT CCATGTCCAG AGC ATT CAA GAC
                                                S   I   Q   D
```

Figure 5. Comparison of the predicted pTP and adenovirus DNA polymerase aa sequences of Ad2, Ad7, and Ad12. The optimal alignment between the three nt sequences was determined using the algorithm of Needleman and Wunsch (1970). Positions in the aligned sequences where three or more consecutive aa were identical in all three serotypes are shown in black. Positions where deletions or insertions in the aa sequences have occurred in Ad2 or Ad7 relative to Ad12 are shown as triangles above or below the comparison (respectively); the size (in aa) of the deletion (negative numbers) or insertion (positive numbers) is also shown. The possible overlap between the C-terminus of pTP and the beginning of Ad pol is shown by the lines connecting the two regions. In the case of Ad pol, in the first 250 aa, there is only low homology; in this region, any position where the predicted aa matches in all three serotypes has been shown in black. The asterisks show the positions of deletions or insertions in the Ad pol ORF of Ad2 or Ad7 relative to Ad12; these changes occur within 45 nt after the end of the pTP coding region. In addition, the position marked I represents a 2-aa insertion in Ad2 relative to Ad12, while the position marked D represents a 3-aa deletion in Ad7 relative to Ad12; these changes occur immediately before the beginning of the highly homologous region in Ad pol.


Length of predicted pTP polypeptide (amino acids)
300
600
pTP
dCTP-serine site
cleave site
+23aa +5aa
-4aa
+1aa+4aa
+1aa +17aa -4aa +3aa
Ad2
Ad12
Ad7
overlap with IVa2 (out of frame)
+1aa
+1aa
+1aa
POL
D
Ad2
Ad12
Ad7
0
300
600
900
1200
Length of predicted polymerase polypeptide (amino acids)

amino acid sequence of Ad pol have been determined starting at the hypothetical splice acceptor at nt 8513 (in Ad12). A low level of aa homology between the three adenovirus serotypes in the less conserved nt sequence region was detected, using the matching algorithm of Needleman and Wunsch (1970). Interestingly, deletions and insertions in each protein have occurred in regions that are not highly conserved in aa sequence, suggesting that these regions have little effect on the function of the polypeptides. The Ad pol has fewer insertions and deletions than does pTP, although most of the insertions and deletions in pTP lie in the N-terminal section of the protein which is removed by cleavage with the adenovirus protease.

ACKNOWLEDGEMENTS

We thank W. Doerfler for providing us with the plasmids and the l phage carrying the Ad12 DNA that was used for sequencing. Several of the initial sequencing experiments were carried out in the laboratory of Thomas R. Broker and Louis T. Chow at Cold Spring Harbor Laboratory. Work at UAB was supported by NIH grant AI-20408 and CA-13148-12. The core facility for oligodeoxynucleotide synthesis provided oligodeoxynucleotide primers for sequencing and is supported by the UAB Comprehensive Cancer Center (NCI grant CA-13148-15).

REFERENCES

Akusjärvi, G., Mathews, M.B., Andersson, P., Vennstrom, B. and Pettersson, U.: Structure of genes for virus-associated RNAI and RNAII of adenovirus type 2. Proc. Natl. Acad. Sci. USA 77 (1980) 2424-2428.

Alestrom, P., Akusjarvi, G., Pettersson, M. and Pettersson, U.: DNA sequence analysis of the region encoding the terminal protein and the hypothetical *N* -gene product of adenovirus type 2. J. Biol. Chem. 257 (1982) 13492-13498.

Barnes, W.M., Bevan, M. and Son, P.H.: Kilo-sequencing: creation of an ordered nest of asymmetric deletions across a large target sequence carried on phage M13. Methods Enzymol. 101 (1983) 98-123.

Biggin, M.D., Gibson, T.J. and Hong, G.F.: Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA 80 (1983) 3963-3965.

Birnbaum, H.C. and Doly, J.C.: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7 (1979) 1513-1523.

Dekker, B.M.M. and Van Ormondt, H.: The nucleotide sequence of fragment *Hind* III-C of human adenovirus type 5 DNA (map positions 17.1-31.7). Gene 27 (1984) 115-120.

Devereux, J., Haeberli, P. and Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12 (1984) 387-395.

Engler, J.A., Hoppe, M.S. and Van Bree, M.P.: The nucleotide sequence of the genes encoded in early region 2b of human adenovirus type 7. Gene 21 (1983) 145-159.

Engler, J.A. and Van Bree, M.P.: The nucleotide sequence of the gene encoding protein IVa2 in human adenovirus type 7. Gene 10 (1982) 71-80.

Föhring, B., Geis, A., Koomey, M. and Raska, K.: Adenovirus type 12 VA RNA, I. Synthesis in productive infection and gene mapping. Virology 95 (1979) 295-302.

Gingeras, T.R., Sciaky, D., Gelinas, R.E., Bing-Dong, J., Yen, C.E., Kelly, M.M., Bullock, P.A., Parsons, B.L., O'Neill, K.E. and Roberts, R.J.: Nucleotide sequences from the adenovirus genome. J. Biol. Chem. 257 (1982) 13475-13491.

Kelly Jr., T.J.: Adenovirus DNA replication. In Ginsberg, H.S. (Ed.), The Adenovirus. Plenum, New York, 1984, pp. 271-308.

Kyte, J. and Doolittle, R.F.: A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157 (1982) 105-132.

Lewis, J.B. and Anderson, C.W.: Proteins encoded near the adenovirus late messenger RNA leader segments. Virology 127 (1983) 112-123.

Lewis, J.B., Fahnestock, M.L., Hardy, M.M. and Anderson, C.W.: Presence in infected cells of nonvirion proteins encoded by adenovirus messenger RNAs of the late transcription regions L0 and L1. Virology 143 (1985) 452-466.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, pp. 454.

Maxam, A.M. and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65 (1980) 499-560.

Needleman, S.B. and Wunsch, D.D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. (1970) 443-453.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A.: Cloning in single-strand bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143 (1980) 161-178.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463-5467.

Staden, R.: Sequence handling by computer. Nucl. Acids Res. 4, (1977) 4037-4051.

Staden, R.: Further procedures for sequence analysis by computer. Nucl. Acids Res. 5 (1978) 1010-1015.

Stillman, B.W., Lewis, J.B., Chow, L.T., Mathews, M.B. and Smart, J.E.: Identification of the gene and mRNA for the adenovirus terminal protein precursor. Cell 23 (1981) 497-508.

Sussenbach, J.S. and Van der Vliet, P.C.: The mechanism of adenovirus DNA replication and the characterization of replication proteins. Curr. Topics Microbiol. Immunol. 109 (1983) 53-73.

Symington, J.S., Lucher, L.A., Brackman, K.H., Virtanen, A., Pettersson, U. and Green, M.: Biosynthesis of adenovirus type 2 i-leader protein. J. Virol. 57 (1986) 848-856.

Thimmappaya, B., Weinberger, C., Schneider, R.J. and Shenk, T.: Adenovirus VAI RNA is required for efficient translation of viral mRNAs at later times after infection. Cell 31 (1982) 543-551.

Virtanen, A., Aleström, P., Persson, H., Katze, M.G. and Pettersson, U.: An adenovirus agnogene. Nucl. Acids Res. 10 (1982) 2539-2547.

Vogel, S., Brötz, M., Kruczek, I., Neumann, R., Eick, D., Winterhoff, U. and Doerfler, W.: Cloned fragments of human adenovirus type-12 DNA. Gene 15 (1981) 273-278.

GRADUATE SCHOOL
UNIVERSITY OF ALABAMA AT BIRMINGHAM
DISSERTATION APPROVAL FORM

Name of Candidate Jeong Shin Hong

Major Subject Biochemistry

Title of Dissertation Assembly and Transport of Adenovirus Fiber

Dissertation Committee:

Jeffrey A. Engler, Chairman

Director of Graduate Program Jeffrey A. Engler

Dean, UAB Graduate School

Date February 15, 1991

PS-1428