

## Sequence and Genetic Organization of Adenovirus Type 35 Early Region 3

PHYLLIS R. FLOMENBERG,<sup>1\*</sup> MEI CHEN,<sup>2</sup> AND MARSHALL S. HORWITZ<sup>2,3,4</sup>

*Departments of Cell Biology,<sup>2</sup> Medicine,<sup>1</sup> Microbiology-Immunology,<sup>3</sup> and Pediatrics,<sup>4</sup> Albert Einstein College of Medicine, Bronx, New York 10461*

Received 18 April 1988/Accepted 19 July 1988

The early transcription region 3 (E3) of group B adenovirus type 35 (Ad35), a serotype isolated primarily from patients with acquired immunodeficiency syndrome and other immunodeficiency disorders, has been partially sequenced. We had previously identified an Ad35 29-kilodalton (kDa) early glycoprotein which, analogous to group C Ad2 E3-19K, associated with major histocompatibility complex class I antigens in the endoplasmic reticulum of infected cells. The open reading frame (ORF) of the Ad35 29-kDa protein has now been identified within a 2-kilobase-pair cloned Ad35 E3 fragment. The predicted amino acid sequence was very similar to that of group B Ad3 E3-19K. In contrast, homology between the Ad35 and Ad2 glycoproteins was limited to five cysteines in identical positions and a 20-amino-acid region proximal to the transmembrane domain. In addition, 20.3- and 20.6-kDa ORFs have been identified downstream from the ORF for the Ad35 glycoprotein. Analogous 20-kDa ORFs are present in the Ad3 E3 region but are not present in Ad2 and Ad5. In contrast, the region analogous to an Ad2 11.6-kDa ORF, which is 9 kDa in size in Ad3, was absent from the expected position within the Ad35 E3 region. Because the E3 region is likely to play an important role in the interaction between virus and host, analysis of the function of the Ad35 E3 proteins should further our understanding of adenovirus pathogenesis.

Adenovirus type 35 (Ad35) is a group B adenovirus that has been isolated primarily from patients with acquired immunodeficiency syndrome (AIDS) or other immunodeficiencies (8, 28, 30). We have shown that over 10% of patients with AIDS or AIDS-related complex surveyed at our institution had urine cultures positive for Ad35 (19). In contrast, this serotype was isolated rarely from normal individuals with febrile illnesses, and specific Ad35 antibody was present only at a very low titer in pooled gamma globulin (12). We have therefore proposed that Ad35 may be unique from other adenoviruses in its interaction with the immune system. Currently, there is no well-established animal model in which to test this hypothesis.

We have focused on the characterization of the Ad35 early transcription region 3 (E3) because this region is likely to play an important role in the interaction between the virus and the host. It has been demonstrated that the group C Ad2 and Ad5 E3 regions code for a glycoprotein, E3-19K, which associates with major histocompatibility complex (MHC) class I antigens (22). In addition, the E3 region is not essential for viral replication *in vitro* (20) yet appears to have been conserved among all adenovirus serotypes studied so far (6, 11, 16, 17, 26, 29). The Ad2 glycoprotein has been localized to the endoplasmic reticulum (ER), where it interferes with the transport of class I antigens to the cell surface (1, 3). Inhibition of the expression of class I antigens, which are needed for cytotoxic T-lymphocyte recognition of virus-infected cells (9), may enable adenovirus-infected cells to evade immune recognition and facilitate latency. Besides E3-19K, the Ad2 E3 region also contains eight other open reading frames (ORFs) which could code for proteins important for infection *in vivo*. Currently, only two other E3 proteins, coded by the 11.6-kilodalton (kDa) and 14.7-kDa ORFs, have been shown to be synthesized in adenovirus-infected cells (31, 33). (Another 11.6-kDa ORF is present in

the Ad2 E3 region but is predicted to be noncoding.) Recently, the product of the 14.7-kDa ORF has been shown to protect infected cells against lysis by tumor necrosis factor (14).

We have previously identified and characterized a group B Ad35 29-kDa early glycoprotein which associated with MHC class I antigens in the ER, analogous to the Ad2 E3-19K (11). In contrast to the Ad2 glycoprotein, the Ad35 glycoprotein, E29, was synthesized in much smaller amounts, was more extensively glycosylated, and did not cross-react with polyclonal antibody against the Ad2 glycoprotein.

We now report the identification of the Ad35 E29 ORF within the Ad35 E3 region. A 2-kilobase (kb) Ad35 E3 fragment from map coordinates 79.5 to 83 was cloned and sequenced. Ad35 genomic DNA was extracted from purified virions grown in HeLa cells infected with Ad35 Holden (30), as previously described (18). The Ad35 *Bam*HI A fragment, located between 60.5 to 85 map units (m.u.) (32), was cloned into pUC12 by standard recombinant DNA techniques. The right-hand end of the fragment, containing the E3 region, was subcloned by making use of one *Eco*RI restriction site at 79.5 m.u. and another within the multicloning site of the plasmid, adjacent to the *Bam*HI site at 60.5 m.u. The construct was cut with *Eco*RI and religated, and the 60.5-to-79.5-m.u. *Eco*RI fragment was removed. A pUC12 plasmid containing the remaining Ad35 fragment between the 79.5-m.u. *Eco*RI site and the 85-m.u. *Bam*HI site was obtained. The Ad35 E3 fragment was cloned into both M13mp18 and M13mp19, between the *Eco*RI and *Bam*HI sites in the multicloning regions of these vectors, and sequenced in both directions. A series of overlapping deletions within the insert were subcloned into both M13 vectors, through the use of the IBI Cyclone protocol, based on the method of Dale et al. (7). Sequencing reactions were performed by using a modification of the Sanger dideoxy method (27). The sequence of the entire 2-kb fragment was determined for both strands.

Translation of the L strand of the Ad35 1,978-base-pair E3

\* Corresponding author.

DNA sequence in all three reading frames identified four ORFs which have a capacity to code for proteins larger than 6 kDa. The predicted proteins had molecular masses of 18.5, 6.4, 20.3, and 20.6 kDa. The 2-kb Ad35 E3 nucleotide sequence exhibited a high degree of homology to a region within the group B Ad3 E3 sequence published previously (29). The three largest Ad35 ORFs, 18.5, 20.3, and 20.6 kDa, are analogous to Ad3 ORFs previously identified. The nucleotide sequence of the Ad35 E3 fragment and the predicted amino acid sequences of the four ORFs, together with the alignment of the Ad3 sequences, are shown in Fig. 1. Figure 2 shows a schematic representation and alignment of the Ad35 and Ad3 DNA sequences and ORFs. These data represent the first nucleotide sequence established for the unique group B Ad35.

Comparison of the Ad35 18.5-kDa ORF with the Ad3 E3-19K ORF revealed approximately 80% identity at both the DNA and protein levels. Analysis of the predicted amino acid sequence of the Ad35 ORF revealed that there are four canonical glycosylation sites (Asn-X-Thr/Ser), an observation which is consistent with our studies of the carbohydrate residues on the Ad35 E29 glycoprotein. The Ad35 glycoprotein has been shown by sodium dodecyl sulfate-polyacrylamide gel electrophoresis to migrate as a 29-kDa band because of the presence of four N-linked high-mannose oligosaccharides per molecule. The Ad3 glycoprotein has not been characterized, but its ORF also contains four putative glycosylation sites. In contrast, Ad2 E3-19K contains two high-mannose oligosaccharides per molecule and migrates as a 25-kDa protein (21). Hydrophobicity plots generated for both Ad35 and Ad3 glycoproteins reveal that both proteins have a typical NH<sub>2</sub>-terminal signal sequence, a large hydrophobic transmembrane domain, and a short hydrophilic cytoplasmic tail, similar to Ad2 E3-19K (Fig. 3).

Both the Ad35 18.5-kDa and the Ad3 19.2-kDa ORFs exhibited much more limited homology to the Ad2 E3-19K ORF at the DNA (50%) and protein (30%) levels. This observation was consistent with our inability to demonstrate immunoreactivity between the Ad35 E29 glycoprotein and a polyclonal antibody to Ad2 E3-19K. However, comparison of the Ad35, Ad3, Ad2, and Ad5 MHC class I antigen-binding proteins has identified one prominent region of homology among all four proteins (Fig. 3 and 4). A conserved stretch of 20 amino acids is located adjacent to the transmembrane domain, within the luminal globular domain of the protein. In this region, 15 amino acids are identical and 3 amino acids are related (L-I, F-Y, A-T), whereas only 2 amino acids are not conserved among all four proteins. Since group B and C glycoproteins all share the ability to bind to MHC class I antigens, one hypothesis is that the conserved domain represents a common binding site. This possibility is currently being addressed by site-directed mutagenesis experiments. In addition, there are five Cys residues which are absolutely conserved, located at identical positions within each protein. It is likely that these Cys residues are involved in maintaining a structure that is conserved by critical disulfide bonds.

Further comparison of the four class I antigen-binding proteins has revealed that the last three COOH-terminal amino acids, Lys-Met-Pro, are absolutely conserved. The cytoplasmic-tail amino acids have been shown to be important for localization of the Ad2 glycoprotein to the ER. Deletion of the last eight amino acids is sufficient to allow Ad2 E3-19K to move out of the ER (25). Our current data suggest that three residues may be sufficient for retention of the glycoproteins in the ER or that the three amino acids are

an important conserved domain of the octapeptide ER retention signal.

In addition, we have identified Ad35 E3 ORFs of 20.3 and 20.6 kDa downstream from the 18.5-kDa ORF. Both of these ORFs are strongly predicted by TESTCODE analysis to be coding (10). Hydrophobicity plots of the predicted amino acid sequences indicate that both ORFs have features typical of transmembrane proteins (data not shown). Both proteins coded by the 20.3- and 20.6-kDa ORFs were predicted to be heavily glycosylated because the sequences contain four and five putative glycosylation sites, respectively. Despite these similarities to the E29 protein, neither ORF exhibited any homology to the 18.5-kDa ORF at either the DNA or protein level. The 20.3- and 20.6-kDa ORFs may be distantly related to each other, but the alignment at either the protein (27%) or DNA (41%) level was not statistically significant.

Both the Ad35 E3 20.3- and 20.6-kDa ORFs are analogous to two ORFs in the group B Ad3 E3 unit. Specifically, the Ad35 20.3-kDa ORF showed 79% identity with the Ad3 20.1-kDa ORF at the DNA level and a 76% alignment at the protein level. The Ad35 20.6-kDa ORF was 70% homologous to the Ad3 20.5-kDa ORF at the DNA level, and the predicted proteins were 60% identical. Neither group C Ad2 nor Ad5 had an analogous 20-kDa ORF within its E3 region, which was consequently 1 kb smaller. The function of the unique group B E3 proteins remains unclear. In addition, a 6.4-kDa Ad35 ORF was present in another reading frame within the same DNA sequence as the 18.5-kDa ORF. It had no analog within the E3 units of the other serotypes sequenced, and TESTCODE analysis yielded no prediction in terms of probability of coding.

Unexpectedly, the analog of the Ad2 11.6-kDa and the Ad3 9-kDa ORFs was absent from its anticipated location in the Ad35 E3 region. Instead, directly downstream from the 20.6-kDa ORF, a DNA sequence closely related to the Ad3 10.2-kDa ORF was present (Fig. 1 and 2). The Ad35 E3 sequence data extend approximately two-thirds into this ORF. Since both complementary Ad35 DNA strands were sequenced at least once and the deletion observed corresponds specifically to a single complete ORF, the absence of the E3 ORF is unlikely to represent a sequencing or cloning artifact.

It is intriguing that the equivalent of the Ad2 11.6-kDa ORF was absent from the Ad35 E3 region. The Ad2 protein has previously been shown to be expressed in infected cells by analysis with antibody to synthetic peptides (33). An analogous 9-kDa ORF is present in Ad3, and a related 7.7-kDa ORF has been identified in group B Ad7 (J. Engler, personal communication). Perhaps the absence of an analogous ORF in Ad35 is related to the unique epidemiology of Ad35 infections. Ad35 has been isolated primarily from immunosuppressed hosts such as patients with AIDS, bone marrow transplant recipients, and children with severe combined immunodeficiency disease. Ad35 infection usually causes asymptomatic viremia but occasionally has resulted in fatal pneumonias. Perhaps the Ad2 11.6-kDa ORF codes for a virulence factor which is absent from Ad35. On the other hand, it may not be an essential protein in the adenovirus life cycle *in vivo*. Less likely, the ORF may be present in an aberrant location within the Ad35 E3 region and therefore was not included in the sequence data presented here.

In addition, the Ad35 E3 sequence does not contain the polyadenylation [poly(A)] consensus signal that would be predicted by comparison with the Ad2 E3 region. The Ad2 and Ad5 E3 units have been shown to encode multiple overlapping mRNAs; the mRNAs are separated into two







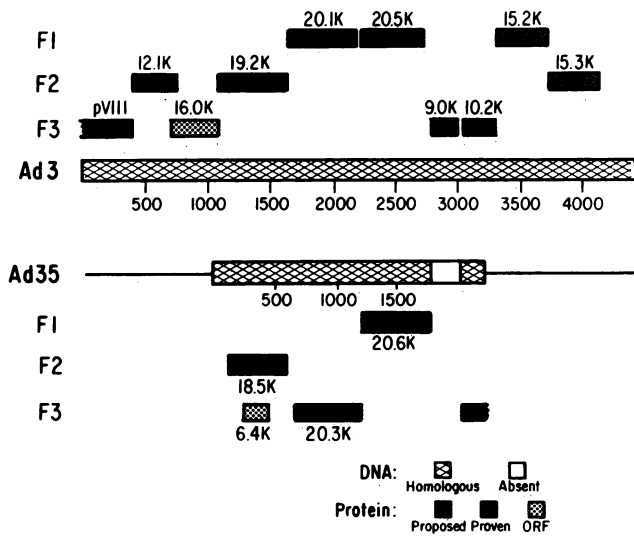


FIG. 2. Comparison of the E3 regions of Ad3 and Ad35. Alignment of the Ad35 sequence within the 4-kb Ad3 *Xho*I-*Bal*I fragment (from 76.5 to 89.2 m.u.) (29) is illustrated schematically. The gap in the Ad35 E3 region indicates the absence of approximately 250 nucleotides which code for a 9-kDa ORF in the Ad3 sequence. Boxes represent ORFs larger than 6 kDa. F1, F2, and F3 are the three reading frames; K, kilodaltons.

groups, E3A and E3B, on the basis of the use of two different poly(A) sites (4, 5). The major Ad2 E3 mRNA transcribed, which codes for E3-19K, uses the first poly(A) site downstream from the 11.6-kDa ORF; the poly(A) site is preceded by the hexanucleotide signal sequence ATTAAG. The consensus nucleotide sequence AATAAG is present at the second poly(A) site, downstream from the Ad2 5'-terminal ORF. In the Ad3 E3 region, the sequences AATAAG and

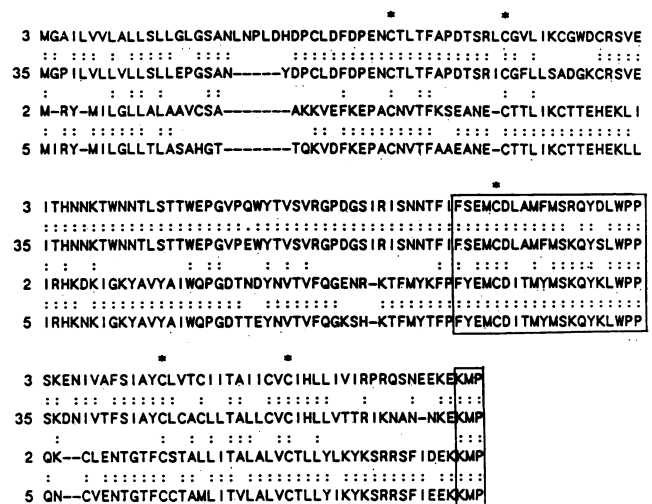


FIG. 4. Alignment between the Ad35, Ad3, Ad2, and Ad5 MHC class I antigen-binding proteins. Identical residues are indicated by a colon. Symbols: □, conserved domains; \*, Cys residues which are absolutely conserved among the four proteins.

ATTAAG are present close together within the 20.1-kDa ORF (at nucleotides 1996 and 2008), and an atypical sequence ATGAAA, is present at the end of the 9-kDa ORF (nucleotide 3016); one of these sequences may function as a poly(A) site for the Ad3 19.2-kDa ORF. In contrast, none of the above sequences is present in an analogous location in the Ad35 E3 sequence. It is possible that the absence of a poly(A) site adjacent to the Ad35 18.5-kDa ORF could result in less efficient transcription. This hypothesis is consistent with our data that show that the Ad35 E29 glycoprotein is produced in significantly smaller amounts compared with Ad2 E3-19K. For instance, studies of Ad2 E3 deletion

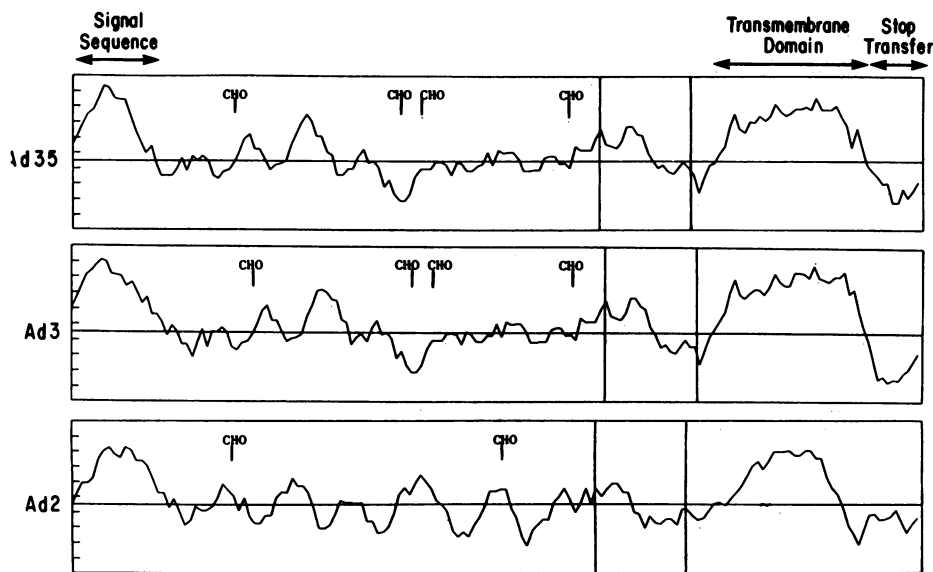


FIG. 3. Comparison of hydrophobicity plots for the MHC class I antigen-binding glycoproteins of Ad35, Ad3, and Ad2. Plots were generated by the computer program ANALYSEP (Roger Staden, Medical Council Center, Cambridge, England) with the hydrophobicity values of Kyte and Doolittle (23). A window span of seven was used. The y axis represents the hydrophobicity index from +50 (top) to -50 (bottom). Hydrophilic C-terminal stop transfer regions and domains which are typical for hydrophobic N-terminal signal sequences and transmembrane regions are indicated. CHO, Putative glycosylation sites. A 20-amino-acid domain that is conserved among all three glycoproteins is indicated within the vertical lines adjacent to the transmembrane domains on each plot.

mutants have demonstrated that the elimination of the first poly(A) site or other defined sites downstream from the 19-kDa ORF can decrease levels of Ad2 E3-19K mRNA (2). Alternatively, a novel sequence which can function as a poly(A) site for the 18.5-kDa mRNA may be present in the Ad35 E3 region.

The functions of the E3 proteins in adenovirus pathogenesis remain unclear. The MHC class I antigen-binding glycoprotein may effect virulence or facilitate latency, but direct evidence is lacking. Although recent studies of infection of the cotton rat with Ad5 mutants have indicated that deletion of E3 enhances virulence, similar experiments with hamsters have yielded contradictory results (13, 24). We are currently investigating an alternative model, using a mouse adenovirus in its natural host (15). Further studies will also provide important information in regard to the safety of using an adenovirus E3 deletion mutant as a live-virus vaccine vector in humans, as has been proposed elsewhere (24).

We thank E. Guetta for assistance in cloning the Ad35 *Bam*HI A fragment, R. Korn and A. Weinreb for technical advice, G. DeFilippi for typing the manuscript, and J. Engler for critically reading the manuscript.

This work was supported by Public Health Service grants AI21105, AI20671, and CA13330 from the National Institutes of Health.

#### LITERATURE CITED

- Andersson, M., S. Paabo, T. Nilsson, and P. A. Peterson. 1985. Impaired intracellular transport of class I MHC antigens as a possible means for adenoviruses to evade immune surveillance. *Cell* **43**:215-222.
- Bhat, B. M., and W. S. M. Wold. 1987. A small deletion distant from a splice or polyadenylation site dramatically alters pre-mRNA processing in region E3 of adenovirus. *J. Virol.* **61**:3938-3945.
- Burgert, H., and S. Kvist. 1985. An adenovirus type 2 glycoprotein blocks cell surface expression of human histocompatibility class I antigens. *Cell* **41**:987-997.
- Chow, L. T., T. R. Broker, and J. B. Lewis. 1979. Complex splicing patterns of RNAs from the early regions of adenovirus 2. *J. Mol. Biol.* **134**:265-303.
- Cladaras, C., B. Bhat, and W. S. M. Wold. 1985. Mapping the 5' ends, 3' ends, and splice sites of mRNAs from the early E3 transcription unit of adenovirus 5. *Virology* **140**:44-54.
- Cladaras, C., and W. S. M. Wold. 1985. DNA sequence of the early E3 transcription unit of adenovirus 5. *Virology* **140**:28-43.
- Dale, R. M. K., B. A. McClure, and J. P. Houchins. 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. *Plasmid* **13**:31-40.
- deJong, P. J., G. Valderrama, I. Spigland, and M. S. Horwitz. 1983. Adenovirus isolates from the urines of patients with the acquired immunodeficiency syndrome. *Lancet* **ii**:1293-1295.
- Doherty, P. C., and R. M. Zinkernagel. 1975. H-2 compatibility is required for T-cell mediated lysis of target cells infected with lymphocytic choriomeningitis virus. *J. Exp. Med.* **141**:502-507.
- Fickett, J. W. 1982. Recognition of protein coding regions in DNA sequence. *Nucleic Acids Res.* **10**:5303-5318.
- Flomenberg, P. R., M. Chen, and M. S. Horwitz. 1987. Characterization of a major histocompatibility complex class I antigen-binding glycoprotein from adenovirus type 35, a type associated with immunocompromised hosts. *J. Virol.* **61**:3665-3671.
- Flomenberg, P. R., M. Chen, G. Munk, and M. S. Horwitz. 1987. Molecular epidemiology of adenovirus type 35 infections in immunocompromised hosts. *J. Infect. Dis.* **155**:1127-1134.
- Ginsberg, H. S., J. Valdesus, R. Horswood, R. M. Chanock, and G. Prince. 1987. Adenovirus gene products affecting pathogenesis, p. 322-326. *In* R. Chanock, R. A. Lerner, F. Brown, and H. Ginsberg (ed.), *Vaccines* 87. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Gooding, L. R., L. W. Elmore, A. E. Tollefson, H. A. Brody, and W. S. M. Wold. 1988. A 14,700 MW protein from the E3 region of adenovirus inhibits cytolysis by tumor necrosis factor. *Cell* **53**:341-346.
- Hartley, J. W., and W. P. Rowe. 1960. A new mouse virus apparently related to the adenovirus group. *Virology* **11**:645-647.
- Herisse, J., G. Courtois, and F. Galibert. 1980. Nucleotide sequence of the EcoRI D fragment of adenovirus 2 genome. *Nucleic Acids Res.* **8**:2173-2192.
- Herisse, J., and F. Galibert. 1981. Nucleotide sequence of EcoRI E fragment of adenovirus 2 genome. *Nucleic Acids Res.* **9**:1229-1240.
- Horwitz, M. S. 1971. Intermediates in the synthesis of type 2 adenovirus deoxyribonucleic acid. *J. Virol.* **8**:675-683.
- Horwitz, M. S., G. Valderrama, R. Korn, and I. Spigland. 1984. Adenovirus isolates from the urines of AIDS patients: characterization of group B recombinants, p. 187-207. *In* M. S. Gottlieb and J. E. Groopman (ed.), *Acquired immune deficiency syndrome*. Alan R. Liss, Inc., New York.
- Kelly, T. J., Jr., and A. M. Lewis, Jr. 1973. Use of nondefective adenovirus-simian virus 40 hybrids for mapping the simian virus 40 genome. *J. Virol.* **12**:643-652.
- Kornfeld, R., and W. S. M. Wold. 1981. Structures of the oligosaccharides of the glycoprotein coded by early region E3 of adenovirus 2. *J. Virol.* **40**:440-449.
- Kvist, S., L. Ostberg, H. Persson, and L. Philipson. 1978. Molecular association between transplantation antigens and a cell surface antigen in an adenovirus-transformed cell line. *Proc. Natl. Acad. Sci. USA* **75**:5674-5678.
- Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**:105-132.
- Morin, J. E., M. D. Lubeck, J. E. Barton, A. J. Conley, A. R. Davis, and P. P. Hung. 1987. Recombinant adenovirus induces antibody response to hepatitis B virus surface antigen in hamsters. *Proc. Natl. Acad. Sci. USA* **84**:4626-4630.
- Paabo, S., B. M. Bhat, W. S. M. Wold, and P. A. Peterson. 1987. A short sequence in the COOH-terminus makes an adenovirus membrane glycoprotein a resident of the endoplasmic reticulum. *Cell* **50**:311-317.
- Paabo, S. T., T. Nilsson, and P. A. Peterson. 1986. Adenoviruses of subgenera B, C, D, and E modulate cell-surface expression of major histocompatibility complex class I antigens. *Proc. Natl. Acad. Sci. USA* **83**:9665-9669.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
- Shields, A. F., R. C. Hackman, K. H. Fife, L. Corey, and J. D. Meyers. 1985. Adenovirus infections in patients undergoing bone-marrow transplantation. *N. Engl. J. Med.* **312**:529-533.
- Signas, C., G. Akusjarvi, and U. Pettersson. 1986. Region E3 of human adenoviruses: differences between the oncogenic adenovirus-3 and the nononcogenic adenovirus-2. *Gene* **50**:173-184.
- Stalder, H., J. C. Hierholzer, and M. N. Oxman. 1977. New human histocompatibility complex class I antigen (p. 35) causing fatal disseminated infection in a renal transplant recipient. *J. Clin. Microbiol.* **6**:257-265.
- Tollefson, A. E., and W. S. M. Wold. 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.
- Valderrama-Leon, G., P. Flomenberg, and M. S. Horwitz. 1985. Restriction endonuclease mapping of adenovirus 35, a type isolated from immunocompromised hosts. *J. Virol.* **56**:647-650.
- Wold, W. S. M., C. Cladaras, S. C. Magie, and N. Yacoub. 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.