Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 6039–6043, October 1984
Cell Biology

Homology of adenoviral E3 glycoprotein with HLA-DR heavy chain

(viral protein/sequence analysis by computer/major histocompatibility complex antigens)

DEVJANI CHATTERJEE AND JACOB V. MAIZEL, JR.

Section of Molecular Structure, Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20205

Communicated by Matthew D. Scharff, June 12, 1984

ABSTRACT **The M_r 19,000 adenovirus glycoproteins coded by the E3 region that is expressed on the cell membrane and presumably binds to HLA class I antigen shows sequence homology to the α-chain of human HLA-DR and to the κ-chain of human IgM. The homology extends to a conserved domain present in all of the major histocompatibility complex antigens, β_2-microglobulins, and immunoglobulins. Predicted β-sheet secondary structures are similar between the M_r 19,000 adenovirus protein and these immune system proteins. Evolutionary and functional implications of the homology are discussed.**

The human adenovirus (adeno), type 2 genome has seven transcription units that are expressed early after infection, the L1, E1a, E1b, E2a, E2b, E3, and E4 regions (1). Of these the E3 region codes for a M_r 19,000 protein (2). This protein is glycosylated (3) and is localized in the plasma membrane (4), where it appears soon after infection (5, 6). Association of this viral glycoprotein with the cytoskeleton has also been reported (7). The COOH terminus of the protein contains a region of 23 amino acids containing mainly hydrophobic residues followed by a hydrophilic tail of about 15 amino acids. This resembles the structural pattern of many membrane-integrated glycoproteins like glycophorin and the HLA antigens, which span the cell membrane and have a hydrophilic COOH terminus on the cytoplasmic side of the cell membrane (8). The adeno glycoprotein has two potential glycosylation sites located on the exterior side of the membrane.

It is known that the successful killing of virus-infected cells by cytotoxic T lymphocytes is dependent on the corecognition by the cytotoxic T lymphocytes of a viral antigen and antigens of the major histocompatibility complex (MHC) on the same cell surface. The formation of a complex between the viral glycoprotein and class I MHC antigen *in vivo* has been demonstrated by immunoprecipitation and affinity chromatography (6). In addition to a physical relationship, other studies (9, 10) have shown a functional relationship between the viral and class I MHC antigens. In the case of adenovirus, Signas *et al.* (6) suggest that the interaction between the two antigens occurs close to the cell membrane.

The MHC controls the expression of two types of genetically polymorphic cell surface antigens, the class I and the class II HLA antigens (11), both of which are glycosylated, membrane-integrated molecules. The class I molecules are associated noncovalently with a small protein called β_2-microglobulin (β_2m), whose exact function is yet undefined (10, 11). The class II HLA antigens, unlike the class I molecules, are only expressed on certain types of cells involved in immune response (12). Both the class I and class II proteins serve as restrictive elements involved in the cell-mediated recognition and killing of infected cells. The HLA antigens, β_2m, and the constant region of the immunoglobulins show amino acid sequence homology in the domain nearest to the cell membrane (13, 14). Though all of these proteins have a common feature of being functionally involved in some way with the immune reaction, the role of the related regions in these proteins is far from clear. The Thy-1 glycoprotein was the first protein outside of this group that showed similarity to the immunoglobulins (15). The Thy-1 protein is found at the cell surface of rodent thymocytes, neuronal cells, and some other cell types. Until now none of these molecules has been found in species more primitive than vertebrates, though a putative Thy-1 analog has been obtained from the squid brain (16).

Since the adeno glycoprotein is expressed on the cell surface and in a complex with the HLA antigen possibly involved in the recognition of virally infected cells, we compared the sequence of this protein with those of the HLA antigens. The adeno protein was found to resemble the HLA class II antigen in domain structure and amino acid sequence. The α-chain of the class II antigen is 100 amino acids longer than the adeno protein. Homology between the proteins is confined to the domain nearest to the cell membrane and to the intramembrane region. However, it is this region that is common among the different HLA proteins and microglobulin and its presence in the adeno glycoprotein suggests that the latter may be involved in the immune process. The viral protein can be deleted and is not essential for the growth of the virus in cell culture, but it is likely to play a role in the overall strategy of this virus in its interaction with its natural, intact host.

METHODS

A VAX 11/750 computer was used for these studies. The graphic matrix method (17) was used to test for the presence of homology. The Goad–Kanehisa program (18) was used to obtain the optimal alignments as well as their statistical significance. The various sequences were randomized 100 times to obtain the significance in terms of standard deviation (SD). All of the sequences were retrieved from the National Biomedical Research Foundation (NBRF) (19). A score of >5 SD indicates evolutionary relatedness of two proteins and scores between 3 and 5 SD support relationship if there are other indications of functional similarity (20).

RESULTS AND DISCUSSION

To test the suitability of the Goad–Kanehisa method for the purpose of these studies we used it to detect potential homologies between two separate groups of proteins. All of the protein sequences were obtained from the NBRF data base. The first group consisted of proteins from diverse sources, different superfamilies, not known to have any functional relation to each other. The proteins were zein precursor, maize (ZIZM3); thaumatin I, *Thaumatococcus daniellii* (QTTC1);

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: β_2m, β_2-microglobulin(s); MHC, major histocompatibility complex; NBRF, National Biomedical Research Foundation.

plastocyanin, shepherd's purse (CUSU); anthopleurin A, sea anemone (NAXA); globin A III, earthworm (GGEWA3); keratin, sheep (KRSHHA); histone H2A, bovine, rat, human, and chicken (HSBO21); *N*-acetylmuramoyl-L-alanine amidase, bacteriophage T7, (MUBPA7). NBRF code numbers are in parentheses.

Predictably, the program did not indicate any relationship between these sequences when the cut-off value for the distance between the best alignment was fixed at −50. [The distance between two sequences can be described as the score obtained, and expressed as a negative value, by the best alignment based on the mutation data matrix (19).]

At the other end of the scale the Goad–Kanehisa program was used to detect homologies between the different members of the HLA class I, class II, and microglobulins, which are known to be homologous and related evolutionarily (Fig. 1). These proteins have very low distance values and consequent high values for significance of relatedness in terms of SD units. On comparing the results of a self alignment with the scores obtained on alignment with other proteins, it is seen that among themselves the class I and the class II proteins have very high SD scores, but, on aligning class I with class II, a 10-fold increase in the distance is obtained with a consequent decrease in the significance of relatedness. It has been shown (11) that the homology between these molecules extends through limited portions or "domains" within these molecules. Similar results are observed with the β_2m.

On comparing the adeno glycoprotein with these proteins it is seen that only a few of the antigens are homologous above the minimal threshold value. These are mainly the class II antigens. The adeno protein is smaller in size and has a comparatively higher distance value on self comparison. On aligning with HLA class II, however, there is a 10-fold increase in the distance value. The alignment of the adeno protein and the HLA class II α-chain stretches over almost the whole of the viral protein and the scores indicate a significant degree of relatedness (Fig. 2). It is interesting to note that the viral protein also demonstrates significant relatedness with the κ-chain V region of IgM (Fig. 2). The latter protein, however, does not show any significant degree of relatedness with the HLA antigens, except the HLA-DR-related DC1, α-chain (Fig. 1). It is a member of the IgM group of immunoglobulins that are expressed on the cell membrane of B cells and probably are responsible for the primary antibody response. The amino acid sequence homology with the viral protein is shown in Fig. 2.

	HLHUB7 HLA-B7 heavy human	HLHU12 HLA-I heavy human	HLMSKB H-2kb heavy mouse	HLHUDA HLA-DR alpha human	HLHUDC HLA-DC1 alpha human	HLHUWB HLA-DR2 beta human	MGHUB2 β2- micrgl. human	MGMSB2 β2- micrgl. mouse	G4HU IgG-4 C-reg human	K1HULY IgM-κ V-I reg human	Q6ADE E-3 prot. Adeno- virus
HLA-B7 heavy human	−1409 271/271 124.20										
HLA-I heavy human	−1257 270/270 92.31	−1823 359/359 138.50									
H-2kb heavy mouse	−1066 271/271 91.81	−1185 329/330 110.50	−1805 348/348 146.90								
HLA-DR alpha human	−133 93/92 14.54	−194 142/134 22.23	−166 153/147 15.25	−1297 254/254 111.00							
HLA-DC1 alpha human	−127 103/103 16.04	−181 143/140 16.97	−174 154/150 17.09	−743 229/229 69.80	−1137 232/232 86.92						
HLA-DR2 beta human	−189 184/184 18.13	−186 182/181 18.79	−183 189/189 18.14	−193 166/171 17.37	−193 103/102 20.03	−1033 198/198 96.82					
β2- micrgl. human	−116 86/87 12.36	−133 104/103 16.66	−130 93/94 13.78	−168 112/109 20.00	−183 103/104 16.52	−168 101/100 18.75	−569 110/110 58.29				
β2- micrgl. mouse	−113 93/92 13.50	−127 99/98 13.10	−114 94/95 10.26	−175 96/95 19.52	−191 100/99 18.99	−165 91/91 19.11	−406 99/99 50.00	−520 99/99 77.39			
IgG-4 C-reg. human	−124 130/130 12.39	−125 132/132 9.88	−113 84/86 12.11	−103 90/87 8.92	−118 109/113 10.49	−138 183/202 11.58	−98 97/101 9.73	−93 88/94 10.15	−1608 327/327 138.50		
IgM-κ V-I reg human	NH	NH	NH	NH	−56 54/55 5.77	NH	NH	NH	NH	−526 108/108 63.10	
E-3 pro Adeno- virus	NH	NH	−56 61/59 5.86	−81 146/141 6.32	−52 141/146 3.33	−54 33/38 3.66	NH	NH	NH	−79 73/76 7.25	−866 159/159 73.50

FIG. 1. Scores generated by the comparison of different HLA antigens, β_2m, immunoglobulins, and adeno E3 glycoprotein by the Kanehisa program. The scores indicate respectively the distance between two sequences, the number of residues involved in the best alignment, and the significance in terms of SD. Comparisons of sequences with themselves, which lie along the main diagonal, serve to define the number of residues in each sequence and the scores for the maximal homology possible. A gap penalty of 8 is used in all cases. NH, no homology observed below a distance of −50. The proteins are HLHUB7, HLA-B7 heavy chain, human; HLHU12, HLA class I, heavy chain precursor, human; HLMSKB, H-2kb heavy chain, mouse; HLHUDA, HLA-DR α-chain precursor, human; HLHUDC, HLA-related DC1 α-chain, human; HLHUWB, HLA-DR2, β-chain, human; MGHUB2, β_2m, human; MGMSB2, β_2m, mouse; G4HU, Ig γ-4 chain C region, human; K1HULY, IgM κ-chain V-I region, human; Q6ADE, adeno E3 glycoprotein.

```
(A). LOCALLY HOMOLOGOUS REGIONS BETWEEN Q6ADE      AND HLHUDA

 DISTANCE   -81,  Q6ADE- 2- 141 ,  HLHUDA- 75- 222
        10         20         30         40         50         60         70
RYMILGLL ALAAVCSAAKKVEFKEPACNVT FKSEANECTTLIKCTTE HE KLIIRHKDKIGKYAVYAIW QPGD  T
:        :::          :       : :       : : :      :  :    :   ::      :   :   :   :
RFASFEAQGALANIAVDKANLEIMTKRSNYTPITNVPPEVTVLTNSPVELREPNVLICFIDKFTPPVVNVTWLRNGKPVT
    80         90        100        110        120        130        140        150

   80         90         100        110        120        130        140
NDYNVTVFQGENRKTFM YKF PFYEMCDITMYMSKQYKLWPPQKCLENTGTF CSTALLITALALVCTL
     :::        :        ::         :          :          :      :  :    :: :
TGVSETVFLPREDHLFRKFHYLPFLPSTE DVYDCRV EHWGLDEPLLKHWEFDAPSPLPETTENVVCAL
   160        170        180        190        200        210        220

(B). LOCALLY HOMOLOGOUS REGIONS BETWEEN Q6ADE      AND K1HULY

 DISTANCE   -79,  Q6ADE- 43- 115 ,  K1HULY- 21-  96
       50         60         70         80         90        100        110
IKC TTEH EKLIIRHKDKIGKYAVYAIWQPGDTNDYNVTVFQGENRKT FMYKFPFYEMCDITMYMSKQYKLWPP
: :              : :        :                    : :    : :          ::  :   ::  :::
ITCQASQNVNAYLNWYQQKPGLAPKLLIYGASTREAGVPSRFSGSGSGTDFTFTISSLQPEDIATYYCQQYNNWPP
       30         40         50         60         70         80         90
```

FIG. 2. Alignment of sequences obtained by the Goad–Kanehisa program. (*A*) Between adeno glycoprotein (Q6ADE) and HLA-DR heavy chain (HLHUDA), human. (*B*) Between Q6ADE and IgM κ-chain (K1HULY), human. The adenoviral sequence is at the top in both of the cases. Residue number and identical amino acids are indicated.

Apart from sequence homology the common domains of the HLA also share secondary structure homology. These regions were shown to have a predominantly β-sheet structure (21). We subjected the adeno protein to a β-sheet prediction program based on the parameters of Chou and Fasman (22). Though the accuracy of these predictions is of the order of 50–60%, the β-sheet structure of the adeno glycoprotein as compared to the homology domain of HLA class II antigen as seen in Fig. 3 shows a striking similarity.

The similarity among the HLA antigens, however, is limited not only to varying degrees of amino acid sequence homology but also to specific arrangement of disulfide bonds that form domains. The similarity between different HLA proteins is mainly limited to the domain nearest to the cell membrane. A close look at the adeno protein sequence next to the cell membrane (residues 45–100) revealed such a domain (Fig. 4).

Certain highly conserved structural features characteristic of these MHC domains (11) are present in the adeno protein: (*i*) the size of the domain between the cysteine residues; (*ii*) matches within the domain of invariant residues: the two cysteines are conserved, and a single tryptophan residue is present and is aligned within the HLA. Other residues like glycine, which may play an important role in secondary structure, as well as aspartic acid, tyrosine, and phenylalanine are found to be aligned when the adeno protein is compared with the HLA and β_2m; (*iii*) there is a homology in the conserved hydrophobicity of certain residues within the domain. If an amino acid is not exactly identical it is replaced by other similar hydrophobic residues. Such substitutions have also been observed among the HLA (23); (*iv*) the concentration of homologous residues around the cysteines. A clustering of identical amino acids, threonine, leucine, and isoleucine, is seen around the NH_2-terminal cysteine. At the other end of the domain a tyrosine residue is invariant. In the region adjoining the cell membrane and outside the disulfide

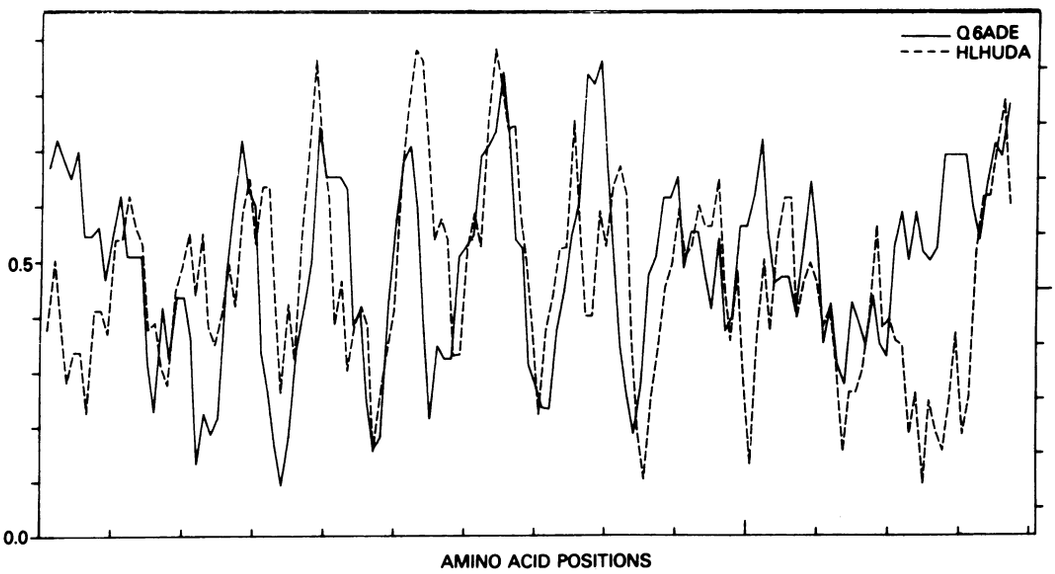

FIG. 3. β-Sheet structure prediction according to Chou and Fasman (22). The window size is 5. Each division in the *x* axis indicates 10 residues. Solid line = Q6ADE adeno glycoprotein, residues 1–141; broken line = HLHUDA, HLA-DR, heavy chain, residues 74–223. The homologous regions as depicted in Fig. 1 are shown here.

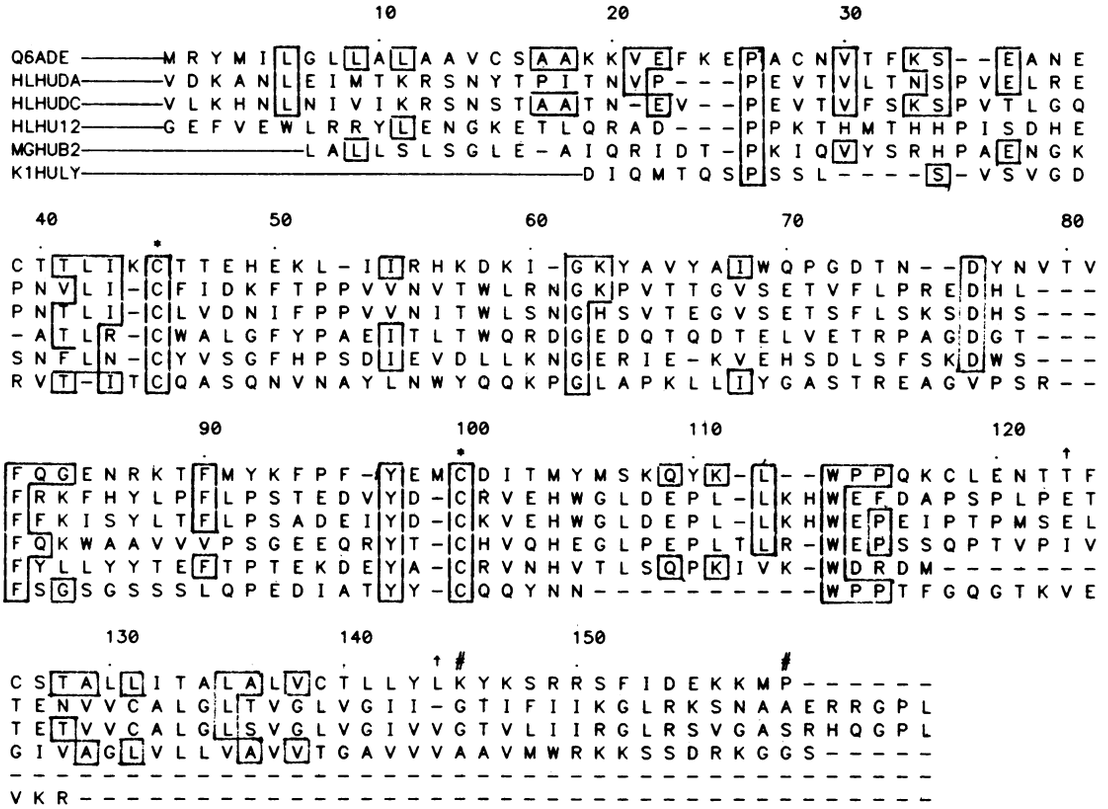

FIG. 4. Alignment of sequences obtained by visual matching. The beginning and the end of domains are indicated by *. The approximate transmembrane region is shown by ↑ and the cytoplasmic region of the adeno protein (8) is indicated by #. Identical residues are boxed. Residue numbers of the adeno protein are indicated. Q6ADE, residues 1–159. Segments of the other proteins are HLHUDA, residues 59–254; HLHUDC, residues 68–232; HLHU12, residues 183–340; MGHUB2, residues 1–110; K1HULY, residues 1–108. For explanation of codes, see the legend to Fig. 1.

domain, the similarity between the adeno protein and the HLA decreases. However, a tryptophan- and proline-containing tripeptide, probably significant in terms of secondary structure, is present in these proteins. The distance of the domain from the hydrophobic region, which may anchor the proteins in the membrane, is also comparable in these proteins; (*v*) the presence of short stretches of hydrophobic residues within the domain that correspond to segments of β-pleated sheets. Orr *et al.* (23) demonstrated four stretches of alternating hydrophobic residues in HLA-B7 and five in β_2m in similar positions within the domain. The presence of such β structures in the viral protein in the homologous region has been shown in Fig. 3.

The alignment of the domains presented in Fig. 4 was obtained by visual matching of the segments. The computer-generated alignment presented in Fig. 2 shows a slightly different alignment in the same region that generates the highest score. This is due to the unique weight assigned to certain amino acids, especially tryptophan, by the mutation data scoring matrix (19) used in these programs. Though the viral protein is related to the class II antigen in a statistically significant manner in an alignment slightly different from the

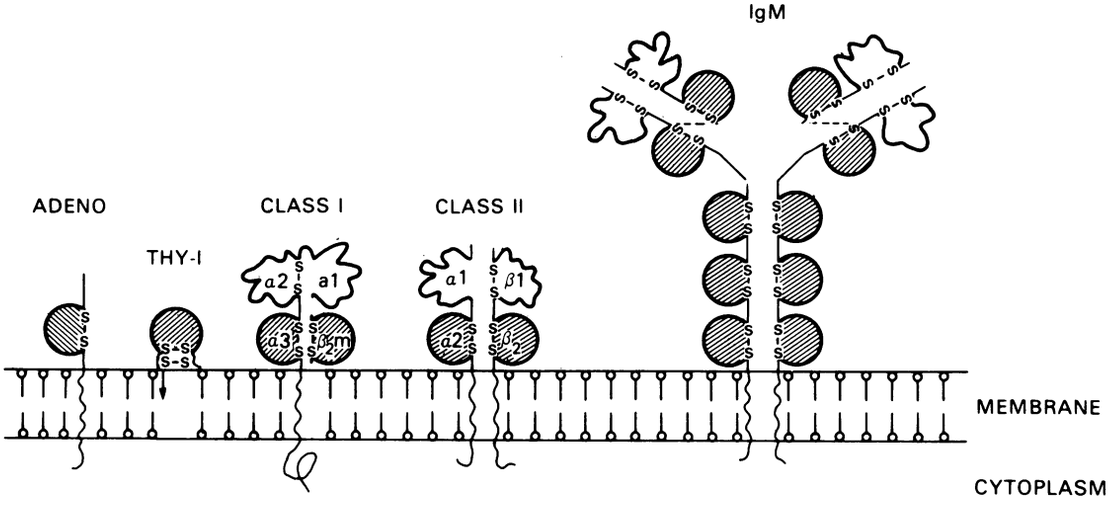

FIG. 5. Schematic representation of the various proteins positioned on the cell membrane.

Table 1. Scores obtained by comparison of adeno glycoprotein with other proteins from NBRF database showing a score of −70 or less

Protein	Adeno E3 glycoprotein Q6ADE		
	Distance	Residues showing homology, no.	Significance in terms of SD
Cytochrome P-450, rat liver (O4RTPB)	−83	136/136	7.15
Acetylcholine receptor protein, δ-chain, electric ray (ACRYD1)	−81	81/79	7.12
Cytochrome *b* mRNA maturase, Baker's yeast mitochondria (MRBY)	−78	142/145	4.04
Probable nucleoprotein, snowshoe hare bunyavirus (VHVUNH)	−76	54/47	5.26
Transposase, *Escherichia coli* (TQECT)	−75	50/50	10.02
Tail tubular protein B, bacteriophage T7 (TLBPB7)	−75	56/54	7.69
Ig κ-chain precursor V-I region, human HK101 (K1HU11)	−74	114/117	8.03
Acetylcholine receptor protein, γ-chain precursor, electric ray (ACRYG1)	−73	123/125	5.96
Proinsulin precursor, anglerfish (IPAF)	−73	43/41	7.93
DNA primase, chains A and B, bacteriophage T7 (YDBFA7)	−73	73/74	9.97
Coat protein VP1, polyoma virus (VVVP13)	−72	80/76	7.82
Coat protein VP1, polyoma virus (VVVP1)	−72	74/69	7.32
α-2u-globulin precursor, rat (UART)	−71	94/93	4.44
Serum albumin precursor, rat (ABRTS)	−71	143/150	5.12
Lactose permease, *E. coli* (GREC)	−70	54/56	5.40
Exodeoxyribonuclease, bacteriophage T7 (NDBPT7)	−70	107/100	6.38

Scores already reported in Fig. 1 are not shown. NBRF code numbers are in parentheses.

alignment presented in Fig. 4, the domain alignment as presented here stresses the biological significance of a conserved structure among the various HLA members, adeno M_r 19,000 glycoprotein, and other proteins belonging to this group.

All of these structural features suggest that the adeno E3 glycoprotein fits well into the schematic representation of the IgM, HLA antigens, and Thy-1 glycoprotein (Fig. 5).

It is difficult at this point to describe precisely the functional relevance of the sequence and structural conformity observed between the viral and the other proteins. The various roles played by the HLA class II antigens are still in the process of being elucidated. Kaufman *et al.* (24) recently suggested that the HLA class II antigens are the primordial MHC molecules. The flow of evolution may have led to the adeno protein to a point similar to that reached by the κ-chain of the IgM immunoglobulins, which are membrane bound and the probable precursor of free antibodies.

Apart from a common origin, the fact that a similarity has been maintained over the period of evolution points to a functional necessity of such structures for cell recognition. The relevance of the conserved structure leads to interesting possibilities. Similarities in the binding of the adeno protein and β_2m to class I antigens are suggested. Recent observations of Kampe *et al.* (25) that smaller amounts of β_2m are immunoprecipitated in infected cells that are producing adeno glycoprotein than in control cells are consistent with this hypothesis. Since the precipitation of both β_2m and the viral protein were observed in conjunction with HLA heavy chains it is possible that the adeno protein competes with β_2m for binding to the HLA antigen. A similar secondary structure of the two proteins strengthens this assumption. The change in conformation as a result of this binding may be recognized by the cytotoxic T lymphocytes, setting into motion the immune response.

In addition to the homologies mentioned above, the adeno glycoprotein has shown varying degrees of relatedness with other proteins (Table 1). Some of these are membrane-related proteins, whereas others are involved in processes involving nucleic acids. The viral protein also shows a significant but low order of homology with the mouse epidermal growth factor (NBRF code, EGMSMG) and also a recently reported (26) yeast cell division control protein. The significance of these homologies is yet to be ascertained.

We thank Minoru Kanehisa for providing us with the VAX version of his programs.

1. Persson, H. & Philipson, L. (1982) *Curr. Top. Microbiol. Immunol.* **97**, 157–203.
2. Walter, G. & Maizel, J. V., Jr. (1974) *Virology* **57**, 402–408.
3. Ishibashi, M. & Maizel, J. V., Jr. (1974) *Virology* **58**, 345–361.
4. Chin, W. & Maizel, J. V., Jr. (1976) *Virology* **71**, 518–530.
5. Storch, T. G. & Maizel, J. V., Jr. (1980) *Virology* **103**, 54–67.
6. Signas, C., Katze, M. G., Persson, H. & Philipson, L. (1982) *Nature (London)* **299**, 175–178.
7. Lenk, R., Storch, T. G. & Maizel, J. V., Jr. (1980) *Virology* **105**, 19–34.
8. Persson, H., Jornvall, H. & Zabielski, J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6349–6353.
9. Geiger, G., Rosenthal, K. L., Klein, J., Zinkernagel, R. M. & Singer, S. J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4603–4607.
10. Kimball, E. S. & Cooligan, J. E. (1983) *Contemp. Top. Mol. Immunol.* **9**, 1–63.
11. Hood, L., Steinmetz, M. & Malissen, B. (1983) *Annu. Rev. Immunol.* **1**, 529–568.
12. Shackelford, W. A., Kaufman, J. F., Korman, A. J. & Strominger, J. L. (1982) *Immunol. Rev.* **66**, 133–187.
13. Korman, A. J., Auffrey, C., Schamboeck, A. & Strominger, J. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6013–6017.
14. Larhammar, D., Gustaffson, K., Claesson, L., Bill, P., Wiman, W., Schenning, L., Sundelin, J., Widmark, E., Peterson, P. A. & Rask, L. (1982) *Cell* **30**, 153–161.
15. Cohen, F. E., Novotny, J., Sternberg, M. J. E., Campbell, D. G. & Williams, A. F. (1981) *Biochem. J.* **195**, 31–35.
16. Williams, A. F. & Gagnon, J. (1982) *Science* **216**, 696–703.
17. Maizel, J. V., Jr., & Lenk, R. P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7665–7669.
18. Goad, W. B. & Kanehisa, M. (1982) *Nucleic Acids Res.* **10**, 247–263.
19. Dayhoff, M. O., ed. (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Georgetown University, Washington, DC), Vol. 5.
20. Barker, W. C. & Dayhoff, M. O. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2836–2839.
21. Krangel, M. S., Orr, H. T. & Strominger, J. L. (1980) *Scand. J. Immunol.* **11**, 561–571.
22. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47**, 251–276.
23. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A. & Strominger, J. L. (1979) *Nature (London)* **282**, 266–270.
24. Kaufman, J. F., Auffrey, C., Korman, A. J., Shackelford, D. & Strominger, J. L. (1984) *Cell* **36**, 1–13.
25. Kampe, O., Bellgrau, D., Hammerling, U., Lind, P., Paabo, S., Severinsson, L. & Peterson, P. A. (1983) *J. Biol. Chem.* **258**, 10594–10598.
26. Lorincz, A. T. & Reed, S. I. (1984) *Nature (London)* **307**, 183–185.