Hydropathic characteristics of adenovirus hexons

Q.-G. Li, K. Lindman, and G. Wadell

Department of Virology, Umeå University, Umeå, Sweden

Accepted February 10, 1997

Summary. The complete nucleotide sequence and the predicted amino acid sequence of the adenovirus type 7 hexon gene were determined. The hydrophathy of the hexon proteins from human adenovirus types 2, 3, 4, 5, 7, 12, 16, 40, 41, and 48, bovine adenovirus type 3, murine adenovirus type 1, and avian adenovirus types 1 and 10 was analysed. The presence of purines and pyrimidines in the second position of the codons was correlated to hydrophilicity and hydrophobicity, respectively. Comparison of the hydrophilicity plots of eight hexons showed seven hypervariable regions to be distributed on the surface. A large portion of the hypervariable regions manifests hydrophilicity. The strength of the surface charge accumulated on the hydrophilic and hydrophobic regions correlated to the tissue tropism of the different adenovirus types. Analysis of codon usage for adenovirus hexons showed that among synonymous codons those with cytidine in the third position were preferably used to a great extent. Analysis of the nucleotide and amino acid sequence pair distances and the phylogenetic tree of 14 hexon proteins showed members of subgenera B, D and E to be closely related, especially Ad4 and Ad16, and subgenus A to be closely related to subgenus F.

Introduction

The family *Adenoviridae* comprises two genera *Mastadenovirus* and *Aviadenovirus*. The genus *Mastadenovirus* consists of 101 species that have been isolated from nine different host species [24]. The most extensively studied group is that of the human adenoviruses. So far, 49 human adenovirus (Ad) serotypes have been identified [27] and divided into six different subgenera, A to F, which differ from each other in various characteristics such as tropism. Adenovirus type 7 (Ad7) is the serotype most frequently associated with severe diseases [17].

Adenoviruses are non-enveloped icosahedral viruses. The virion contains at least eleven different structural polypeptides. The hexon is the most abundantly

produced protein. Each virion contains 240 hexons which form the facets of the icosahedron. The nine central hexon capsomers in each facet are cemented together by polypeptide IX to form a group-of-nine [7]. The hexon is a homotrimer consisting of three identical polypeptide chains. The monomer is an unusually large structural protein with a M_r ranging from 102–109 k (Table 2).

As several entire or partial hexon sequences of different serotypes have been published (references cited in Table 2), the hexons of several different serotypes have been compared at the nucleotide and predicted amino acid sequence levels [9, 14, 31, 36].

The biochemical and immunological properties and even the three-dimensional structure of adenoviruses have been extensively studied. The trimeric hexon molecule has a pseudo-hexagonal base with a large central cavity and a triangular top. The base contains two pedestal domains, P1 and P2. The top contains three long loops L1, L2 and L4 [3]. However, the hydrophobic character of the adenovirus hexons and the codon usage has not been analysed. The hydrophobicity of a protein is very important for predicting putative antigenic regions of the protein. The antigenic determinants can be deduced by searching the amino acid sequence for the areas of greatest local hydrophilicity. Generally, the highest peak of hydrophilicity correctly predicts an antigenic site [11].

Codon usage bias for codons with thymidine (T) or cytidine (C) at second positions in mitochondrial DNA has been shown to be correlated with hydrophobicity [21]. Having analysed nine adenovirus hexon genes, we found the presence of purine and pyrimidine at the second codon position to be correlated to hydrophilicity and hydrophobicity, respectively [19]. This finding was recently confirmed by analysis of further data obtained from GenBank. Here, we report the hydrophobicity analysis of 14 adenovirus hexon sequences predicted from a newly determined Ad7 hexon DNA sequence and thirteen published hexon sequences of Ad2, Ad3, Ad4, Ad5, Ad12, Ad16, Ad40, Ad41, Ad48, Bav3, Mav1, Fav1 and Fav10. There is strong correlation between hydrophilicity and the codons with purine in the second position (CPUSs), and between hydrophobicity and the codons with pyrimidine in the second position (CPYSs) in adenovirus hexons. Comparison of the surface charge on these 14 hexons suggests the strength of the surface charge accumulated in hydrophilic regions or hydrophobic regions to be correlated with the tissue tropism of the different adenovirus types.

Materials and methods

Virus and cells

The Ad7 prototype strain Gomen, originally from the American Type Culture Collection (ATCC), was obtained from Dr. G. von Zeipel, Stockholm County Council Central Microbiological Laboratory, Stockholm, Sweden [34]. This strain was propagated in A549 cells.

Cloning of hexon DNA fragments

The viral DNA of Ad7 strain Gomen was prepared as described [18]. The purified DNA was used as the template for PCR to amplify five fragments which covered the entire hexon gene. The PCR products were cloned into pT7 Blue vector. Ligation, transformation and rapid screening were performed using the pT7 Blue T-vector Kit protocols (Novagen, Madison, WI, USA). The recombinant pT7-Blue plasmid DNA was prepared for sequencing.

As the PCR method occasionally gave rise to error when the complementary DNA strand was synthesised using Taq polymerase, the sequence so obtained was confirmed with the following procedure: The Ad7 hexon DNA Bam HI restriction fragments A and E, which contain the hexon gene, were cloned into plasmid vector pBR322 and multiplied in *E. coli* strain HB101 using conventional methods described previously [25]. These recombinant plasmid DNAs were used as the sequencing templates.

Sequencing

The nucleotide sequences were determined for both strands with the dideoxynucleotide chain determination method, following procedure C of the AutoRead Sequencing Kit (Pharmacia LKB Biotech, Sweden) and the manufacturer's instructions for the ALF DNA Sequencer (Pharmacia).

Amino acid sequence analysis

All the sequence data were analysed with the LaserGene software (DNASTar Inc., Madison, WI) programs, Editseq, MegAlign and Protean.

Every hexon DNA sequence was translated to protein sequence by using program EditSeq-Translation. The codon usage data could be obtained when a protein file was opened by EditSeq. The hydrophilicity plot (Kyte-Doolittle method) could be shown by program Protean. All the eight hydrophilicity plots could be moved together by the program Microsoft PowerPoint (Fig. 1). The isoelectric point and the net charge of a protein could be obtained by program Protean-Composition (Table 2).

A tabular data of hydrophilicity and hydrophobicity for each amino acid could be obtained by program Protean-TabularData. The accumulated charge of all hydrophilic regions of a protein could be obtained by deleting manually all the hydrophobic regions according to the tabular data. An easier way was that, first, to delete the hydrophobic regions within DNA sequence; second, to translate this DNA sequence to an amino acid sequence; then, to obtain the accumulated charge of hydrophilic regions by program Protean-Composition. The accumulated charge of hydrophobic regions could be obtained in the same way (Table 2).

The data of codon usage of accumulated hydrophilic and hydrophobic regions from 14 different serotypes were analysed by the computer program Microsoft Excel (Tables 1 and 5).

Results*Ad7 hexon DNA sequence*

The complete hexon gene DNA sequence of Ad7 prototype strain, Gomen, was determined. It consists of 2 814 nucleotides. This DNA sequence, together with

Fig. 1. Comparison of the hydrophilicity plots (Kyte-Doolittle method) of the hexons of 8 adenovirus serotypes. The boxed areas correspond to the hypervariable regions in Fig. 3. The figures around the boxes denote the amino acid numbers at the start and the end of each hypervariable region

Table 1. Codon usage in 14 adenovirus hexons

Amino acid	Codon	Codon Number		Frequency (%)		Amino acid	Codon	Codon Number		Frequency (%)	
		Hydrophilic	Hydrophobic	Hydrophilic	Hydrophobic			Hydrophilic	Hydrophobic	Hydrophilic	Hydrophobic
Lys(K)	AAA	194	37	4,0	2,3	Thr (T)	ACA	132	39	3,1	1,6
	AAG	186	42	3,9	2,6		ACC	305	126	7,2	5,2
Asn(N)	AAC	537	165	11,2	10,1	Ile (I)	ACG	81	58	1,9	2,4
	AAU	248	90	5,2	5,5		ACU	170	47	4,0	2,0
Arg (R)	AGA	121	17	2,5	1,0		AUA	59	39	1,4	1,6
	AGG	60	16	1,3	1,0		AUC	88	96	2,1	3,9
	CGA	39	9	0,8	0,6		AUU	133	87	3,1	3,6
	CGC	166	45	3,5	2,8	Met(M)	AUG	260	173	6,1	7,1
Ser (S)	CGG	66	13	1,4	0,8	Pro (P)	CCA	141	32	3,3	1,3
	CGU	65	13	1,4	0,8		CCC	288	74	6,8	3,0
	AGC	114	76	2,4	4,7		CCG	87	23	2,1	0,9
	AGU	70	34	1,5	2,1		CCU	148	31	3,5	1,3
Gln(Q)	CAA	172	42	3,6	2,6	Leu (L)	CUA	46	39	1,1	1,7
	CAG	257	79	5,4	4,4		CUC	82	93	1,9	3,8
His (H)	CAC	134	27	2,8	1,7		CUG	198	167	4,7	6,6
	CAU	46	9	1,0	0,6		CUU	80	92	1,9	3,8
Glu (E)	GAA	214	59	4,5	3,6	Ala (A)	UUA	28	17	0,7	0,7
	GAG	187	31	3,9	1,9		UUG	70	72	1,7	3,0
Asp (D)	GAC	401	143	8,4	8,8		GCA	82	37	1,9	1,5
	GAU	231	64	4,8	3,9		GCC	241	157	5,8	6,4
Gly (G)	GGA	145	62	3,0	3,8	Val (V)	GCG	73	52	1,7	2,1
	GGC	217	120	4,5	7,4		GCU	173	97	4,1	4,0
	GGG	94	60	2,0	3,7		GUA	74	50	1,8	2,1
	GGU	108	67	2,3	4,1		GUC	79	68	1,9	2,8
Tyr (Y)	UAC	389	181	8,1	11,1	Ser (S)	GUG	215	145	5,1	5,9
	UAU	131	61	2,7	3,7		GUU	100	72	2,4	3,0
Cys (C)	UGC	38	20	0,8	1,2		UCA	65	22	1,5	0,9
	UGU	11	7	0,2	0,4		UCC	130	98	3,1	4,0
Trp (W)	UGG	153	38	3,2	2,3	Phe (F)	UCG	81	24	1,9	1,0
							UCU	113	44	2,7	1,8
							UUC	208	137	4,9	5,6
							UUU	198	130	4,7	5,3
Total		4794	1627	100,0	100,0			4228	2438	100,0	100,0

two short flanking regions at each end, has been entered in the European Molecular Biology Laboratory (EMBL, accession number: Z48571). The 5' flanking region has a splice acceptor site. The 3' flanking region ends just before the start codon of proteinase. The sequence of the predicted protein, consisting of 937 amino acids, was obtained with the LaserGene software program EditSeq.

Table 2. Predicted biophysical characteristics^a of hexon proteins of 14 adenovirus serotypes

Sub-genus	Sero-type	Length (aa)	Molecular mass	Iso-electric point	Charge at pH 7		C + D content (%)	GeneBank accession number	Reference
					entire hexon	H-phobic region			
A	Ad12	919	103 050	5.80	-11.54	-3.85	47.48	X73487	[29]
B	Ad3	943	106 192	5.23	-18.66	-15.13	47.70	X76549	[23]
	Ad7	937	105 720	5.31	-16.83	-12.30	47.51	Z48571	This paper
C	Ad16	940	105 511	5.30	-16.71	-5.19	48.94	X74662	[23]
	Ad2	968	109 163	4.96	-26.73	-20.97	49.67	J01917	[2]
	Ad5	952	108 017	5.15	-22.57	-17.81	49.07	M73260	[14]
D	Ad48	947	106 851	5.65	-14.36	-7.83	55.30	U20821	[9]
E	Ad4	936	105 273	5.30	-17.55	-8.16	55.06	X84646	[23]
F	Ad40	923	103 988	6.00	-8.71	-4.99	50.78	L19443	[31]
	Ad41	925	103 985	5.73	-10.91	-7.18	50.63	X51783	[30]
BAV	Bav3	911	103 061	6.44	-6.42	+3.99	54.59	X53990	[12]
MAV	Mav1	909	102 337	5.74	-10.90	-0.26	47.78	M81889	[36]
FAV	Fav1	943	106 835	5.27	-15.88	-4.02	56.09	Z67970	[1]
	Fav10	936	105 462	5.66	-10.21	-4.51	58.23	U26221	[28]

^a All characteristics were elicited with the program: Protein Composition, DNASTar

*CPUS and CPYS are correlated to hydrophilicity
and hydrophobicity, respectively*

The hydropathy data of hexon proteins from human adenovirus types 2, 3, 4, 5, 7, 12, 16, 40, 41, and 48, bovine adenovirus type 3, murine adenovirus type 1, and avian adenovirus types 1 and 10 were derived using the prediction method of Kyte-Doolittle in the LaserGene computer program Protean. Analysis of the codon usage for the codons corresponding to hydrophilic and hydrophobic regions showed the presence of CPUSs and CPYSs to be strongly correlated with hydrophilicity and hydrophobicity, respectively. Statistical analysis using the chi square test showed a highly significant difference ($\chi^2 = 192.79$, $P < 0.001$) for the total numbers of CPUSs and CPYSs in hydrophilic and hydrophobic regions (Table 1).

Alignment of the 14 complete sequences of adenovirus hexons

The multiple sequence alignments of the 14 complete nucleotide sequences and the amino acid sequences of adenovirus hexons were determined by the program MegAlign (data not shown). The nucleotide and amino acid sequence pair distances and the phylogenetic tree of 14 hexon proteins showed serotypes of subgenera B, D and E to be closely related (Table 3 and Fig. 2). They manifest 82.0–93.9% amino acid sequence homology. Ad4 (subgenus E) is very similar to Ad16 (subgenus B), the amino acid sequence homology between these two serotypes reaching 93.4%. This is in agreement with the cross neutralisation seen between anti-hexon sera specific for Ad4 and Ad16 [22]. The alignment data also showed Ad12 (subgenus A) to be closely related to Ad40 and Ad41 (subgenus F), the level of sequence homology being 81.5–87.1%. In contrast, the sequence similarity of E3a proteins between Ad12 and Ad41 was very low, only 30% in E3 RL1 and 34% in E3 RL2 [37].

The alignment of the 14 complete sequences of adenovirus hexons showed nucleotide sequence homology to be 77–94% within a subgenus of human adenoviruses, 65–77% between members of different subgenera with the exception between Ad4 and Ad16 (86.2%); 64–94% between adenoviruses from the same host species and 35–59% between adenoviruses from different host species; and 35–43% between the two adenovirus genera *Mastadenovirus* and *Aviadenovirus*.

*Comparison of hydropathy plots demonstrates the existence
of seven hypervariable regions*

The alignment of the 14 complete amino acid sequences of adenovirus hexons revealed the existence of seven hypervariable regions (Fig. 3). Hydrophilicity plots derived from the hexon sequences of eight serotypes representing eight subgenera of *Mastadenovirus* were compared (Fig. 1). Seven hypervariable regions were demonstrated (in boxes, corresponding to the boxes in Fig. 3). The first four hypervariable regions, A1 to A4, covered most of the external loop 1 (L1) of the hexon. They were separated by three short relatively conserved

Table 3. Nucleotide and amino acid sequence homology^a among 14 adenovirus hexons

Subgenus	Serotype	Ad12	Ad3	Ad7	Ad16	Ad2	Ad5	Ad48	Ad4	Ad40	Ad41	Bav3	Mav1	Fav1	Fav10
A	Ad12		64,3	63,6	63,7	65,0	66,1	64,7	63,6	69,1	68,3	57,9	54,0	41,3	35,0
B	Ad3	77,4		94,0	76,9	64,9	66,3	73,0	72,9	65,2	65,5	51,5	53,4	39,4	34,7
	Ad7	78,3	93,9		78,5	67,3	67,8	73,2	73,9	67,0	67,2	56,9	53,4	38,7	38,9
	Ad16	77,9	84,0	85,8		66,6	66,5	72,7	86,2	67,9	67,1	57,0	53,7	38,6	39,8
C	Ad2	77,9	76,0	78,1	77,0		80,7	69,4	70,0	68,1	66,9	52,2	54,4	38,7	40,3
	Ad5	77,0	75,9	76,7	76,5	86,7		68,4	70,0	69,0	68,0	52,4	53,6	38,5	36,5
D	Ad48	79,4	82,0	83,2	82,4	77,3	77,0		77,3	68,8	68,0	58,7	53,7	40,3	42,6
E	Ad4	78,2	82,4	83,5	93,4	77,0	76,7	83,3		68,5	67,7	58,6	54,2	41,8	42,8
F	Ad40	82,3	77,5	79,3	79,2	78,2	78,3	80,5	78,8		78,9	57,9	56,8	40,7	40,6
	Ad41	81,5	77,1	78,4	79,0	78,2	77,5	78,8	78,9	87,1		52,7	54,7	40,5	39,4
BAV	Bad3	66,6	62,8	63,7	65,3	62,9	63,4	63,9	64,8	65,5	66,7		50,4	38,0	41,7
MAV	Mav1	63,7	62,7	62,4	63,9	64,1	62,7	63,9	63,5	64,8	64,8	60,5		37,3	37,3
FAV	Fav1	44,9	44,2	44,9	44,9	44,2	43,8	44,4	44,8	45,3	45,8	39,7	42,9		69,9
	Fav10	42,4	41,7	42,0	42,4	42,6	37,2	42,3	42,5	43,0	43,0	42,4	41,9	73,7	

^aValues in the upper right denote the percent similarity of the nucleotide sequences; each pair of serotypes and those in the lower left denote the percent similarity of the amino acid sequences of each pair of serotypes

Fig. 2. Phylogenetic tree of 14 adenovirus serotypes. This tree was obtained by using program MegAlign-Phylogenetic-tree. The scale indicates the distance between sequences. Ancestral residues are determined by the Hartigan-Fitch-Sankoff method (see Laser-Gene User's Guide provided by the manufacturer)

sequences which may stabilise the outer shell structure of the hexon. The L2 contains two hypervariable regions, B1 and B2. All of the hypervariable regions in L1 and L2 contained longer or shorter deletions/insertions. The amino acid sequences of these regions were serotype specific)

The hypervariable region D of *Mastadenovirus* consisting of 13–14 amino acids was located in L4. Amino acid sequence homology showed the region D to be subgenus specific. Sequence homology was 93–100% for pairs of serotypes within a subgenus, and less than 79% between serotypes belonging to different subgenera.

Interestingly, the major portion of the hypervariable regions manifest hydrophilicity. According to the ribbon representation of the Ad2 hexon subunit [3], the major portion of the hydrophilic regions in hypervariable areas was located at the surface of the hexon molecule.

Codons with cytidine in the third position are highly preferred

Analysis of codon usage for the 14 serotypes showed that among the synonymous codons those with cytidine in the third position (NNC) were highly preferred (Table 1). Among CPUSs the bias for NNC codons was strong. However, although the codon preference for NNC among CPYSs was generally manifest, this was not the case for the NNC codons for the hydrophobic amino acids Ile, Leu and Val.

Hexon surface charge

The isoelectric point and the surface charge for hexons of the 14 serotypes were deduced with the program LaserGene-Protean-Composition (Table 2). At pH 7,

A1

AD12 VLDGRGSPFKPYSGTAYNSLAPKGAAPNSQW-----SDNAKLNITFAQ-APY--LSDTITAADGIKVGTD---TAQAGAAVYANK 177
AD3 VLDGRGSPFKPYSGTAYNSLAPKGAAPNTSQWIVT--TN-----GDNVATTTTNTFGI-ASM---KGDNITKEGLQIGKDITTEGEKPIYADK 187
AD7 VLDGRGSPFKPYSGTAYNSLAPKGAAPNTSQWIVT--AG-----BERAVTTTNTFGI-ASM---KGDNITKEGLEIGKDITA--DNKPIYADK 184
AD16 VLDGRGSPFKPYSGTAYNSLAPKGAAPNTSQW-----KSDSKMHTFGV-AAMPVGTGKKIEADGLPIG--IDSTSGTDTVIYADK 181
AD2 VLDGRGTFPKPYSGTAYNALAPKGAAPNSCEWEQTEDSGRAVADEEEDEDEEEDEEQNARDQATKKTHVYQA-APLSE---TITKSGLQIGSD--NAETQAKFVYADP 209
AD5 VLDGRGTFPKPYSGTAYNALAPKGAAPNSCEWEQTEDSGRAVADEEEDEDEEEDEEQNARDQATKKTHVYQA-APLSE---TITKSGLQIGSD--NAETQAKFVYADP 209
AD48 VLDGRGSPFKPYSGTAYNSLAPKGAAPNSQWEEKKNGG-----GSDANQMOTHTFGV-AAMGGI--EITAKGLQIGIDATKEEDNGKRIYADK 189
AD4 VLDGRGSPFKPYSGTAYNSLAPKGAAPNTSQW-----KSDSKMHTFGV-AAMPVGTGKKIEADGLPIG--IDSTSGTDTVIYADK 181
AD40 VLDGRGSPFKPYSGTAYNSLAPKGAAPNSQW-----TNQNKTNSTFGV-APY--IGQKITNQ-GVQVGLD---SNNR--DVFADK 174
AD41 VLDGRGSPFKPYSGTAYNSLAPKGAAPNTSQW-----KDNKIKVRGQ-APF--IGTNINKDNGIQIGTD---TTNQ--PIYADK 175
BAV3 TLDRGASGFKPYSGTAYNSFAPKSAAPNTQFRQANNG-----HPAQITIAQ-ASYVATIGG--ANNLQMGVD---ERQLFVYANT 178
MAV1 RLDRGSPFKPYSGTAYNSLAPRAAPNLMFKTGADSKTMTIAQAN-----IVGSTIDQDGLKINNVALNPAGEPAV--- 178
FAV1 VLDGRGSPFKPYSGTAYNSLAPRAAPNLMFKTGADSKTMTIAQAN-----IVGSTIDQDGLKINNVALNPAGEPAV--- 178
FAV10 ILDRGSPFKPYSGTAYNSLAPKESMEN--NWESETAPGQNVTA-----SQLSNVYTN--TNTSKATTAAQVTKISGVFPNPNQPGVGNPL--RQVENDNTG--- 200

A2

AD12 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNEHGGQSK-----DD-----NIELKFFDSANN-----AAN-TAQVVFYTEDVNLEM 261
AD3 TYQPEPQVGESEWNTS-D-GTNEKFGRALKPATNMKPCYGSFVKTYNIGKQAKNRKVTNTEGGVETE-EPDIDMEFFD-----GRDAVAGALAPEVILVTENVNLET 288
AD7 TYQPEPQVGESEWNTS-D-GTNEKFGRALKPATNMKPCYGSFARPTNIGKQAKNRKVPTEGDEVETE-EPDIDMEFFD-----GREAA-ADAFSPFVILVTENVNLET 284
AD16 TFQPEPQVGNASWVDA-N-GTEEKYGGRALKDTTKMKPCYGSFAKPTNKEGGQAN---LKDSETAATTP-NYDIDLAFPD-----NKN-IAANYDDIVMVTENVDLQT 278
AD2 SYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
AD5 TFQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
AD48 TFQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
AD4 TFQPEPQVGNDSWVDT-N-GAEKYGGRALKDTTKMKPCYGSFAKPTNKEGGQAN---LKDSETAATTP-NYDIDLAFPD-----NKN-IAANYDDIVMVTENVDLQT 278
AD40 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
AD41 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
BAV3 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
MAV1 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
FAV1 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
FAV10 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307

A3

AD12 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNEHGGQSK-----DD-----NIELKFFDSANN-----AAN-TAQVVFYTEDVNLEM 261
AD3 TYQPEPQVGESEWNTS-D-GTNEKFGRALKPATNMKPCYGSFVKTYNIGKQAKNRKVTNTEGGVETE-EPDIDMEFFD-----GRDAVAGALAPEVILVTENVNLET 288
AD7 TYQPEPQVGESEWNTS-D-GTNEKFGRALKPATNMKPCYGSFARPTNIGKQAKNRKVPTEGDEVETE-EPDIDMEFFD-----GREAA-ADAFSPFVILVTENVNLET 284
AD16 TFQPEPQVGNASWVDA-N-GTEEKYGGRALKDTTKMKPCYGSFAKPTNKEGGQAN---LKDSETAATTP-NYDIDLAFPD-----NKN-IAANYDDIVMVTENVDLQT 278
AD2 SYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
AD5 TFQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
AD48 TFQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
AD4 TFQPEPQVGNDSWVDT-N-GAEKYGGRALKDTTKMKPCYGSFAKPTNKEGGQAN---LKDSETAATTP-NYDIDLAFPD-----NKN-IAANYDDIVMVTENVDLQT 278
AD40 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
AD41 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
BAV3 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
MAV1 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
FAV1 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307
FAV10 TYQPEPQVGESEWNTS-I-ENV-KAGGRALKQTTAMQPCYGSYARPTNPFGGQSVLPVDEKGVPLP-----KVDLQFFSNTTSLNDRQG--NATPKPVVLYSEDVNMET 307

A4

AD12 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
AD3 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
AD7 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
AD16 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
AD2 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
AD5 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
AD48 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
AD4 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
AD40 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
AD41 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
BAV3 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
MAV1 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
FAV1 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371
FAV10 PDTHLVFKPFTVINGTIAESLLGQQAAPNRRNYIAFRDNFIGLMYYNSTGNMGVLGAQASQLNAVVDLQDRNTELSYQLMLDALGDRTRYFSLWNQAVDSYDPDVRVIE 371

B1

AD12 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD3 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD7 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD16 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD2 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD5 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD48 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD4 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD40 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD41 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
BAV3 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
MAV1 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
FAV1 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
FAV10 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472

B2

AD12 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD3 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD7 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD16 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD2 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD5 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD48 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD4 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD40 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
AD41 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
BAV3 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
MAV1 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
FAV1 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472
FAV10 HGVEDELPNYCFPLSAGVEIKNYKGIKPDNGGGG-----WTADNT-VSEANHIGIGNIAAMEINLANLWRSFLYSNVALYLPDKLYTPGNIKLPDNK--NTYEYMN 472

D

AD12 NFQPMRQVVDITTEYKNYKVT-----EFQH 790
AD3 NFQPMRQVVDITTEYKNYKVT-----EFQH 813
AD7 NFQPMRQVVDITTEYKNYKVT-----EFQH 808
AD16 NFQPMRQVVDITTEYKNYKVT-----EFQH 811
AD2 NFQPMRQVVDITTEYKNYKVT-----EFQH 839
AD5 NFQPMRQVVDITTEYKNYKVT-----EFQH 823
AD48 NFQPMRQVVDITTEYKNYKVT-----EFQH 818
AD4 NFQPMRQVVDITTEYKNYKVT-----EFQH 807
AD40 NFQPMRQVVDITTEYKNYKVT-----EFQH 794
AD41 NFQPMRQVVDITTEYKNYKVT-----EFQH 796
BAV3 NFQPMRQVVDITTEYKNYKVT-----EFQH 782
MAV1 NFQPMRQVVDITTEYKNYKVT-----EFQH 779
FAV1 NFQPMRQVVDITTEYKNYKVT-----EFQH 805
FAV10 NFQPMRQVVDITTEYKNYKVT-----EFQH 799

the charges of the hexon proteins from different subgenera varied widely. The hexons of Ad2 and Ad5 belonging to subgenus C carried the highest negative charge, -26.73 and -22.57 , respectively. These charges were about 2.5 times greater than those of hexons of subgenus F, Bav3, Mav1 and Fav10.

More interestingly, the strength of the accumulated charge of hydrophilic or hydrophobic regions correlated with tissue tropism. The prototypes of Ad3, Ad7, Ad2 and Ad5 were isolated from respiratory tract specimens. The prototype of Ad3 was isolated from nasal washing, Ad7 from throat washing, and both Ad2 and Ad5 from adenoid tissue cultures [6]. These adenoviruses were frequently isolated from patients with respiratory diseases [26]. The hexons of these four serotypes were characterised by strong accumulated charges of -12.3 to -20.97 in hydrophilic regions, but manifestly weaker charges of only -3.62 to -5.85 in hydrophobic regions. The prototypes of Ad16 and Bav3 were isolated from conjunctival scrapings and conjunctiva, respectively [6], the hydrophobic regions of their hexons manifest charges of -11.62 and -10.51 . Ad4 can be isolated both from the eye and the respiratory tract [16]. The negative charges of the hydrophilic and the hydrophobic regions of the Ad4 hexon were similar, -8.16 and -9.48 , respectively. The prototypes of Ad12, Ad40, Ad41 and Fav10 were isolated from faeces or rectal swabs [6, 8, 32]. All four of these enteric adenoviruses were characterised by lower negative charges of -3.81 to -7.78 in both hydrophilic and hydrophobic regions of their hexons. Mav1 isolated from spleen and Fav1 isolated from allantois [6] manifested individually unique patterns of surface charge.

Discussion

The hydropathic type and value of an amino acid is dependent on its codon usage and its position in a protein structure.

Scales for evaluating the hydropathic characteristics of amino acids have been developed by many different research groups. However, the hydrophilicity and hydrophobicity values obtained differ substantially [10, 11, 15]. The most frequently used scales, introduced by Kyte and Doolittle were used in the present study. As ranked on hydropathy scales [15], the 20 common amino acids can be divided into three different classes. The first is the hydrophobic class including Ile, Val, Leu, Phe and Cys. Of these five amino acids, only Cys is encoded by CPUSs, the other four being encoded by CPYSs. The second is the hydrophilic class consisting of His, Glu, Gln, Asp, Asn, Lys and Arg, all of

◀
Fig. 3. Seven hypervariable regions of 14 adenovirus hexons. A1 to A4 are localised in loop 1, B1 and B2 are localised in loop 2, and D is localised in loop 4. The numeral at the right designates the amino acid residue in each hexon

which are encoded by CPUSs. The third, intermediate class comprises Met, Ala, Gly, Thr, Trp, Ser, Tyr and Pro. Met and Ala are encoded by CPYSs and manifest weak hydrophobicity. Thr and Pro (encoded by CPYSs), and Gly, Trp and Tyr (encoded by CPUSs) manifest weak hydrophilicity. Ser, the unique amino acid encoded by both CPUSs and CPYSs, manifests weak hydrophilicity. Statistical analysis showed the hydropathy of 19 amino acid (i.e., excluding Ser) to differ significantly between CPUSs and CPYSs (exact method of chi square test, $P < 0.025$). Of the 63 Ser residues in the Ad7 hexon molecule, encoded by 18 CPUSs and 45 CPYSs, 23 occur in hydrophobic regions and 40 in hydrophilic regions, though the difference in distribution is non-significant (chi square test, $\chi^2 = 0.109$, $P > 0.05$).

In general, a A + G-rich region (A, adenosine; G, guanine) in a nucleotide sequence contains more CPUS codons, and C + T-rich region contains more CPYS codons. Therefore, the part of a protein, which is encoded by A + G-rich region usually manifests hydrophilicity and a C + T-rich region encoded peptide usually shows hydrophobic characteristics.

The hydropathy value of an amino acid in a protein chain is dependent on the protein conformation. The residue side-chains can protrude at the interior or exterior portion of a protein chain. The hydropathy value of an amino acid was determined by averaging over a window which contains several consecutive residues surrounding the amino acid in question [15]. Therefore, the strength and even the type of the hydropathy of an individual amino acid may vary according to its location within a peptide chain. In particular, the hydropathy of the amino acids in the intermediate class was found to vary more frequently and in a more pronounced way (Table 1).

The major portion of hydrophobic sequences in a protein will be found in the interior of the native structure, and the major portion of hydrophilic sequences will be found on the exterior [15]. Branden and Tooze [5] found that the hydropathy plots (Kyte-Doolittle method) agree with the crystal structure data in a polypeptide of *R. Sphaeroides*. In this study, we found the major portion of the hypervariable areas to manifest hydrophilicity. The major portion of the hydrophilic regions in hypervariable areas is located at the surface of the crystal structure of the hexon.

The NNC codons in adenovirus hexons manifested high usage. The result is compatible with that derived from analysis of 28 different genes of Ad2 (except the Phe) [33]. Preference for the synonymous codon for Phe, UUC was greater than UUU in 14 hexon genes (Table 1). However, the reverse result was true of 28 different Ad2 genes, UUC accounting for 1.44% and UUU was 2.07%. Phe with its aromatic side chain is highly hydrophobic, nonpolar and chemically stable. The variability of the Phe residue is one of the lowest during divergence of homologous proteins [10]. The patterns of codon usage for Phe in hexon genes were shown to be subgenus specific (Table 4). The codon usage for Phe in the hexon genes of the species of each subgenus is highly similar within subgenera B, C and F. The codon preferably used for Phe in hexon genes of members of subgenera A, C and F is UUU, whereas, UUC is preferably used in

Table 4. Comparison of codon usage for Phe in adenovirus hexon proteins

Codon	Ad12	Ad3	Ad7	Ad16	Ad2	Ad5	Ad48	Ad4	Ad40	Ad41	Bav3	Mav1	Fav1	Fav10
UUC	12	30	28	30	20	19	38	36	21	17	18	5	31	40
UUU	37	17	20	19	25	28	13	11	29	30	29	46	15	9

hexon genes of the species of subgenera B, D and E. These findings corroborate the relatedness of the VA RNA₁ genes [13], and also the relatedness of the hexon genes of human adenoviruses [4].

We found the NNC codons in adenovirus hexon genes to be highly preferred. Analysis of the data obtained from GenBank [33] suggested that the codon usage bias for each virus genus was conserved. A high usage of NNC codons was found in adenoviruses type 2 (*Mastadenovirus*), herpes simplex virus types 1 and 2 (*Simplexvirus*), Epstein-Barr virus (*Lymphocryptovirus*) and human cytomegalovirus (*Cytomegalovirus*); a high usage of NNA codons was found in varicella-zoster virus (*Varicellovirus*), influenzavirus types A and B (*Influenzavirus A,B*), human immunodeficiency virus (*Lentivirus*), human parainfluenza virus and Sendai virus (*Paramyxovirus*), Newcastle disease virus (*Rubulavirus*), human polyomavirus and simian virus 40 (*Polyomavirus*), Human papilloma virus (*Papillomavirus*), and vesicular stomatitis virus (*Vesiculovirus*); a high usage of NNU codons was found in mouse hepatitis virus (*Coronavirus*), reovirus (*Orthoreovirus*), cucumber mosaic virus (*Cucumovirus*), bacteriophage T4 (“T4-like Phages”); and a high-usage of NNA + NNU codons was found in vaccinia virus (*Orthopoxvirus*).

The C + G content of different organisms is used as one of the criteria in taxonomy. Among various eubacteria the C + G content of the genome DNA ranges from 25% to 75% [20]. In *Adenoviridae* the DNA C + G content varies from 48–61% for mastadenoviruses and 54–55% for aviadenoviruses [24]. In the six different subgenera of human adenoviruses the C + G content of genome DNA ranges from 48% to 58% [35]. The C + G content of all 14 serotypes reveals the existence of two groups: higher C + G content (54.59–58.23%) group including Ad48, Ad4, Bav3, Fav1 and Fav10, and lower C + G content (47.48–50.78%) group including the remaining nine serotypes (Table 2). The C + G content of human adenovirus hexons was consistent with the level of C + G content of whole adenovirus genome DNA with the exception of subgenus C, Ad2 and Ad5, which has higher C + G content (58%) in genome DNA.

There are 15 amino acids encoded by synonymous codons which contain NNC (Table 5). Of these 15 amino acids, nine manifested NNC high usage, and three (Asp, Ile and Pro) showed at same level of NNC usage, whereas only three (Cys, Leu and Val) showed low NNC usage. Cys is a rare amino

Table 5. Codon usage in the Ad7 hexon

Amino acid	Codon	Codon Number		Frequency (%)		Amino acid	Codon	Codon Number		Frequency (%)	
		Hydrophilic	Hydrophobic	Hydrophilic	Hydrophobic			Hydrophilic	Hydrophobic	Hydrophilic	Hydrophobic
Lys (K)	AAA	18	3	5,1	2,8	Thr (T)	ACA	12	4	3,9	2,3
	AAG	14	3	4,0	2,8		ACC	20	9	6,5	5,2
Asn (N)	AAC	33	8	9,4	7,5	Ile (I)	ACG	1	2	0,3	1,2
	AAU	23	9	6,6	8,5		ACU	18	3	5,8	1,7
Arg (R)	AGA	17	1	4,8	0,9		AUA	4	3	1,3	1,7
	AGG	5	2	1,4	1,9		AUC	4	8	1,3	4,7
	CGA	2	0	0,6	0,0	Met(M)	AUU	13	7	4,2	4,1
	CGC	10	2	2,8	1,9		AUG	22	12	7,1	7,0
Ser (S)	CGG	1	2	0,3	1,9	Pro (P)	CCA	19	4	6,2	2,3
	CGU	3	0	0,9	0,0		CCC	13	3	4,2	1,7
	AGC	6	3	1,7	2,8	Leu (L)	CCG	2	1	0,6	0,6
	AGU	6	3	1,7	2,8		CCU	13	4	4,2	2,3
Gln (Q)	CAA	12	3	3,4	2,8		CUA	0	3	0,0	1,7
	CAG	13	4	3,7	3,8		CUC	4	1	1,3	0,6
His (H)	CAC	9	2	2,6	1,9	Ala (A)	CUG	13	14	4,2	8,1
	CAU	2	1	0,6	0,9		CUU	10	13	3,2	7,6
Glu (E)	GAA	20	5	5,7	4,7		UUA	3	0	1,0	0,0
	GAG	14	1	4,0	0,9		UUG	3	3	1,0	1,7
Asp (D)	GAC	24	6	6,8	5,7	Val (V)	GCA	9	4	2,9	2,3
	GAU	26	6	7,4	5,7		GCC	15	9	4,9	5,2
Gly (G)	GGA	15	4	4,3	3,8		GCG	4	0	1,3	0,0
	GGC	14	9	4,0	8,5		GCU	16	6	5,2	3,5
	GGG	9	3	2,6	2,8	Ser (S)	GUA	6	3	1,9	1,7
Tyr (Y)	GGU	6	3	1,7	2,8		GUC	4	7	1,3	4,1
	UAC	26	14	7,4	13,2		GUG	14	6	4,5	3,5
	UAU	10	5	2,8	4,7		GUU	10	6	3,2	3,5
Cys (C)	UGC	1	0	0,3	0,0	Phe (F)	UCA	6	3	1,9	1,7
	UGU	2	1	0,6	0,9		UCC	5	9	1,6	5,2
Trp (w)	UGG	10	3	2,8	2,8		UCG	3	1	1,0	0,6
							UCU	14	4	4,5	2,3
							UUC	14	14	4,5	8,1
							UUU	14	6	4,5	3,5
Total		351	106	100,0	100,0			308	172	100,0	100,0

acid and only four cysteines in Ad7 hexon. Among the synonymous codons for Val, seven GUCs appeared in the hydrophobic region. Therefore, the conclusion is that the NNC codons in Ad7 hexon were frequently used although Ad7 is one of the serotypes which contain lower level of C + G content in hexon DNAs.

Acknowledgement

This work was supported by the Swedish Medical Research Council (grant No. 16V-11191 and 16 X-05688) and the Fund for Advanced Medical Research at the University Hospital of Northern Sweden.

References

1. Akopian TA, Karpov VA, Doronin KK, Naroditsky BS (1996) Fav1 (CELO) genes encoding for major core, hexon-associated and hexon proteins. GenBank Database, Accession no. Z67970
2. Akusjavi G, Aleström P, Pettersson M, Lager M, Jörnvall H, Pettersson U (1984) The gene for the adenovirus 2 hexon polypeptide. *J Biol Chem* 259: 13 976–13 979
3. Athappilly FK, Murali R, Rux JJ, Cai Z, Burnett RM (1994) The refined crystal structure of hexon, the major coat protein of adenovirus type 2, at 2.9 Å resolution. *J Mol Biol* 242: 430–455
4. Bailey A, Mautner V (1994) Phylogenetic relationships among adenovirus serotypes. *Virology* 205: 438–452
5. Branden C, Tooze J (1991) Introduction to protein structure. Garland, New York London
6. Buck C, Paulino G (1990) Adenoviruses. In: American Type Culture Collection (ATCC) Catalogue of Animal & Antisera, Chlamydiae & Rickettsiae, 6th ed. American Type Culture Collection, Rockville, pp 1–17
7. Burnett RM (1985) The structure of the adenovirus capsid II. The packing symmetry of hexon and its implications for viral architecture. *J Mol Biol* 185: 125–143
8. Calnek BW, Cowen BS (1974) Adenoviruses of chickens: serologic groups. *Avian Dis* 19: 91–103
9. Crawford-Miksza L, Schnurr DP (1996) Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. *J Virol* 70: 1 836–1 844
10. Creighton TE (1993) Reconstructing evolution from contemporary sequences (chapter 3.3) and The properties of liquid water and the characteristics of noncovalent interactions in this solvent (chapter 4.2). In: Creighton TE Protein, structure and molecular properties 2nd ed. W. H. Freeman, New York, pp 114–127, 148–156
11. Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA* 78: 3 824–3 828
12. Hu S-L, Hays WW, Potts DE (1984) Sequence homology between bovine and human adenoviruses. *J Virol* 49: 604–608
13. Kidd AH, Garwicz D, Öberg M (1995) Human and simian adenoviruses: phylogenetic inferences from analysis of VA RNA genes. *Virology* 207: 32–45
14. Kinloch R, Mackay N, Mautner V (1984) Adenovirus hexon: Sequence comparison of subgroup C serotypes 2 and 5. *J Biol Chem* 259: 6 431–6 436
15. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105–132
16. Levandowski RA, Rubenis M (1981) Nosocomial conjunctivitis caused by adenovirus type 4. *J Infect Dis* 143: 28–31
17. Li Q-G, Wadell G (1986) Analysis of 15 different genome types of adenovirus type 7 isolated on five continents. *J Virol* 60: 331–335
18. Li Q-G, Hambræus J, Wadell G (1991) Genetic relationship between thirteen genome types of adenovirus 11, 34 and 35 with different tropisms. *Intervirology* 32: 338–350
19. Li Q-G, Lindman K, Wadell G (1995) Hydropathic character analysis of nine adenovirus hexon sequences. In: Abstracts of Papers Presented at The Imperial Cancer Research Found Tumour Virus Meeting on Papovaviruses, Papillomavirus and Adenoviruses, Cambridge
20. Moto A, Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci USA* 84: 166–169

21. Naylor GJP, Collins TM, Brown WM (1995) Hydrophobicity and phylogeny. *Nature* 373: 565–566
22. Norrby E, Wadell G (1969) Immunological relationships between hexons of certain human adenoviruses. *J Virol* 4: 663–670
23. Pring-Åkerblom P, Trussenaar FEJ, Adrian T (1995) Sequence characterisation and comparison of human adenovirus subgenus B and E hexons. *Virology* 212: 232–236
24. Russell WC, Adrian T, Bartha A, Fujinaga K, Ginsberg HS, Hierholzer JC, de Jong JC, Li Q-G, Mautner V, Nasz I, Wadell G (1995) Adenoviridae. In: Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA (eds) *Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses*. Springer, Wien, New York, pp 128–133 (*Arch Virol [Suppl]* 10)
25. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning*, 2nd ed. Cold Spring Harbor Laboratory Press, New York, pp 1.53–1.91
26. Schmitz H, Wigand R, Heinrich W (1983) World-wide epidemiology of human adenovirus infections. *Am J Epidemiol* 117: 455–466
27. Schnurr D, Dondero ME (1993) Two new candidate adenovirus serotypes. *Intervirology* 36: 79–83
28. Sheppard M, McCoy RJ, Werner W (1995) Genomic mapping and sequence analysis of the fowl adenovirus serotype 10 hexon gene. *J Gen Virol* 76: 2 595–2 600
29. Sprengel J, Schmitz B, Heuss-Neitzel D, Zock C, Doerfler W (1994) Nucleotide sequence of human adenovirus type 12 DNA: comparative functional analysis. *J Virol* 68: 379–389
30. Toogood CIA, Hay RT (1988) DNA sequence of the adenovirus type 41 hexon gene and predicted structure of the protein. *J Gen Virol* 69: 2 291–2 301
31. Toogood CIA, Murali R, Burnett RM, Hay RT (1989) The adenovirus type 40 hexon: sequence, predicted structure and relationship to other adenovirus hexons. *J Gen Virol* 70: 3 203–3 214
32. Taylor PJ, Calnek W (1962) Isolation and classification of avian enteric cytopathogenic agents. *Avian Dis* 6: 51–58
33. Wada K-N, Wada Y, Doi H, Ishibashi F, Gojobori T, Ikemura T (1991) Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res* 19: 1 981–1 986
34. Wadell G, Varsanyi TM (1978) Demonstration of three different subtypes of adenovirus type 7 by DNA restriction site mapping. *Infect Immun* 21: 238–246
35. Wadell G (1994) Adenoviruses. In: Webster RG, Granoff A (eds) *Encyclopedia of virology*, vol 1. Academic Press, Harcourt Brace & Co, London, pp 1–7
36. Weber JM, Cai F, Murali R, Burnett RM (1994) Sequence and structural analysis of murine adenovirus type 1 hexon. *J Gen Virol* 75: 141–147
37. Yeh H-Y, Pieniazek N, Pieniazek D, Luftig RB (1996) Genetic organization, size, and complete sequence of eight region 3 genes of human adenovirus type 41. *J Virol* 70: 2 658–2 663

Authors' address: Dr. G. Wadell, Department of Virology, Umeå University, S-901 85 Umeå, Sweden.

Received October 25, 1996