

## Molecular Epidemiology of Hepatitis C Virus Genotype 4 Isolates in Egypt and Analysis of the Variability of Envelope Proteins E1 and E2 in Patients with Chronic Hepatitis

D. Genovese,<sup>1†</sup> S. Dettori,<sup>1†</sup> C. Argentini,<sup>1</sup> U. Villano,<sup>1</sup> P. Chionne,<sup>1</sup> M. Angelico,<sup>2</sup>  
and M. Rapicetta<sup>1\*</sup>

*Viral Hepatitis Unit, Department of Infectious, Parasitic and Immune-Mediated Disease, Istituto Superiore di Sanità,<sup>1</sup> and Gastroenterology Unit, Department of Public Health, “Tor Vergata” University,<sup>2</sup> Rome, Italy*

Received 30 July 2004/Returned for modification 24 September 2004/Accepted 1 November 2004

**We analyzed hepatitis C virus (HCV) genotype 4 isolates circulating in the Alexandria District (Egypt) in terms of genetic divergence and the presence of different subtypes. Hypervariable region 1 (HVR1) and the NH2 region of the E2 protein were characterized, and the heterogeneity of subtype 4a isolates was evaluated by analyzing epitope frequencies, immunoproteasome prediction, and possible glycosylation patterns. The heterogeneity of the nucleotide sequences was greater than that found in previous studies, which reported only subtype 4a. Subtype 4a was most common (78% of cases), yet four new subtypes were found, with subtype 4m representing 11% of the cases and the other three subtypes representing another 11%. Substantial heterogeneity was also found when the intrasubtype 4a sequences were analyzed. Differences in the probability of glycosylation and in the positions of the different sites were also observed. The analysis of the predicted cytotoxic-T-lymphocyte epitopes showed differences in both the potential proteasome cleavage and the prediction score. The Egyptian isolates in our study also showed high variability in terms of the HVR1 neutralization epitope. Five of these isolates showed amino acid substitutions never previously observed (a total of six positions). Four of these residues (in four different isolates) were in positions involved in anchoring to the E2 glycoprotein core and in maintaining the HVR1 conformation. The results of this study indicate that HCV genotype 4 in Egypt is extremely variable, not only in terms of sequence, but also in terms of functional and immunological determinants. These data should be taken into account in planning the development of vaccine trials in Egypt.**

Hepatitis C virus (HCV) is the major etiological agent of chronic hepatitis and liver disease worldwide, and it is a major cause of morbidity and mortality. In 80% of cases, infection with HCV results in chronic hepatitis, possibly leading to cirrhosis and hepatocellular carcinoma (11, 28).

HCV is characterized by a high degree of nucleotide sequence variability. Overall, the heterogeneity of the viral genome ranges from 30 to 35% between different genotypes, although it varies with the specific region of the genome, and it has been used to define three types of regions: highly conserved regions (e.g., the 5' untranslated region), variable regions (e.g., envelope 1 [E1] and nonstructural 5b [NS5b]), and hypervariable regions (HVR) (e.g., HVR1 and HVR2 in E2) (48). Furthermore, because of errors of the RNA-dependent RNA polymerase, HCV has a high rate of mutation during replication, and it exists in the bloodstreams of infected persons as complex distributions of mutants known as viral quasispecies, which fluctuate during the course of the disease, mainly as a result of immune pressure (47, 51, 54). These coexisting mutant genomes always have a consensus, or master, sequence. In general, despite the potentially high mutation rate and variability of RNA viruses, changes in the consensus sequence of a viral population occur only if the population equilibrium is

altered by a selection mechanism (16), which is generally produced by the host immune response and its influence on the distributions of the different viral variants. Variants may be related to differences in transmissibility, immunogenicity, or pathogenicity (12, 13, 24, 34, 35), which can be important in developing prophylactic and therapeutic vaccines.

HCV is one of the most suitable virus models for studying genomic variability. The classification into six major genotypes has been accepted (45), and there is further intrasubtype and intrasubject heterogeneity (31). The development of an effective HCV vaccine to protect humans from HCV infection and chronic liver disease is a public health priority, yet differences in antigenic epitopes in genotypes, subtypes, and quasispecies could make cross-protection unlikely.

The E2 glycoprotein of the HCV envelope, one of the possible targets for the development of an effective vaccine, encodes as many as 11 N-linked glycosylation sites, many or all of which may be utilized during the posttranslational processing of nascent E1-E2 complexes (14, 27, 32, 52). Multiple N-linked glycans, in addition to assisting in the folding of antigenically complex proteins, may have other functions, such as masking proteins from reactivity with virus-specific antibodies, facilitating escape from neutralization by antibodies or the complement, and interfering with antigen processing. The antigenicity of viral glycoproteins can in some cases be enhanced by eliminating N-linked glycosylation sites (19, 36). Another essential step in the immune response to HCV infection is the recognition of epitopes by cytotoxic T lymphocytes (CTLs). For viral proteins to be recognized by CTLs, they must be cleaved into

\* Corresponding author. Mailing address: Viral Hepatitis Unit, Department of Infectious, Parasitic and Immune-Mediated Disease, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy. Phone: (39) 06 4990 3233. Fax: (39) 06 4990 2662. E-mail: rapicett@iss.it.

† D. Genovese and S. Dettori contributed equally to this work.

short peptides and translocated into the endoplasmic reticulum or loading onto HLA class I molecules (1, 39, 41). The HLA-peptide binding depends on amino acids known as anchor residues (41). The complex of a peptide and a class I molecule is then presented on the cell surface, which allows it to be recognized by epitope-specific CTLs (39). Each step of CTL epitope generation and recognition is potentially constrained by sequence specificity. Mutations both within and proximal to epitopes can influence their immunogenic potential (7, 8, 33, 55). Mutations that influence cleavage during epitope processing can also be critical (4, 9, 21). Cleavage sites generated by the immunoproteasome are sensitive to the surrounding amino acid sequences (5, 37), although no simple cleavage signal is apparent. In light of these considerations, an effective vaccine has to be multivalent and include those genotypes that are common in a given geographical region.

Egypt could be considered a candidate country for performing trials of prophylactic and therapeutic vaccines, because it has one of the world's highest HCV prevalence rates (2, 48). In Egypt, subtype 4a is predominant (15), and new putative subtypes have been described (43).

In the present study, we analyzed HCV genotype 4 isolates circulating in the Alexandria District, with particular regard to genetic divergence and the presence of different subtypes. The analysis of epitope frequencies, immunoproteasome prediction, and possible glycosylation patterns were used to characterize the HVR1 and NH2 regions of the E2 protein with the aim of studying the heterogeneity of subtype 4a isolates.

#### MATERIALS AND METHODS

**Study population.** The studied isolates were extracted from sera collected during a previous survey of 135 patients with chronic hepatitis who presented to the Alexandria University Hospital (Egypt) between August 1993 and January 1995 and who came from different areas of the Alexandria District (2). As described, the anti-HCV prevalence was 67%, as determined by HCV version 3.0 enzyme-linked immunosorbent assay (Ortho-Clinical Diagnostics, Inc., Raritan, N.J.) and confirmed using the RIBA HCV 3.0 Strip Immunoblot assay (Chiron Corp., Emeryville, Calif.), and 74 of the 84 HCV RNA-positive patients were infected with genotype 4. In the present study, serum samples were analyzed for only 36 of the 74 patients with genotype 4 (~50%), since sufficient quantities of serum were not available for the other patients. There were no epidemiological or clinical differences between the patients included in the analysis and those who were excluded. Sera were stored at  $-80^{\circ}\text{C}$ .

**HCV RNA amplification of NS5b, E1, and E2 genome regions.** HCV RNA was extracted from 100  $\mu\text{l}$  of serum using the guanidinium isothiocyanate method (44). The regions NS5b and E1 were amplified as previously described (3). The E2 region was amplified by means of a reverse transcription-nested PCR, using the forward primer GGF1 (nucleotides [nt] 1171 to 1193 of the HCV genotype 4 prototype Ed43; accession number Y11604), 5' CACTGGACYACBCARGA NTGYAA 3' (where Y is C or T; B is G, T, or C; R is G or A; and N is A, C, G, or T); as reverse primers, we used an equimolar mixture of the primers GGR1 (nt 1946 to 1967), 5' TTGGTGAACCCDGTTCYRTTCA 3' (where D is G, A, or T; R is G or A; and Y is C or T), and GGR2 (nt 1942 to 1963), 5' TGAAC CCGTTCYRTTCA 3' (where D is G, A, or T; R is G or A; and Y is C or T). The 50- $\mu\text{l}$  reaction mixture consisted of 10  $\mu\text{l}$  of RNA, 5  $\mu\text{l}$  of 10 $\times$  buffer (150 mM Tris-HCl, pH 8.9, 625 mM KCl, 27.5 mM MgCl<sub>2</sub>, and 0.01% gelatin), 0.5 mM (each) primer, 200 mM (each) dideoxynucleotide, 5 U of avian myeloblastosis virus reverse transcriptase, and 1.25 U of recombinant *Taq* polymerase. The thermal profile was 43 $^{\circ}\text{C}$  for 20 min, 94 $^{\circ}\text{C}$  for 8 s, and 40 cycles at 94 $^{\circ}\text{C}$  for 40 s, 53 $^{\circ}\text{C}$  for 20 s, and 72 $^{\circ}\text{C}$  for 1 min, with a final extension at 72 $^{\circ}\text{C}$  for 7 min. The second amplification step in 50  $\mu\text{l}$  of reaction mixture consisted of 5  $\mu\text{l}$  of 10 $\times$  buffer (100 mM Tris-HCl, pH 8.3, 500 mM KCl, 15 mM MgCl<sub>2</sub>, and 0.01% gelatin), 0.5 mM (each) a mixture of forward primers GGF2a (nt 1231 to 1253), 5' ATGGCNTGGGAYATGATGATGAA 3' (where N is A, C, G, or T and Y is C or T), and GGF2b (nt 1231 to 1253), 5' ATGGCNTGGGAYATGATGC WGAA 3' (where N is A, C, G, or T; Y is C or T; and W is A or T), and a mixture

of 0.5 mM (each) reverse primers GGR3a (nt 1873 to 1895), 5' AGGAAGAC ATCNGTNTCRRTTCTC 3' (where N is A, C, G, or T and R is G or A), and GGR3b (nt 1873 to 1895), 5' AGGAAGACATCNGTNTCRRTTTTC 3' (where N is A, C, G, or T and R is G or A), 200 mM (each) deoxynucleotide, 1.25 U of recombinant *Taq* polymerase, and 5  $\mu\text{l}$  of the first PCR amplification product. The profile was 95 $^{\circ}\text{C}$  for 9 min and 35 cycles of 94 $^{\circ}\text{C}$  for 40 s, 53.5 $^{\circ}\text{C}$  for 10 s, and 72 $^{\circ}\text{C}$  for 1 min, with a final extension at 72 $^{\circ}\text{C}$  for 7 min. All PCR products were analyzed using 2% agarose gel electrophoresis with ethidium bromide staining.

**Sequencing and analysis of nucleotide sequences.** To remove excess primers, all positive samples were purified on a 100,000-kDa Centricon column (Millipore, Billerica, Mass.); they were then sequenced using the BigDye Terminator kit (Applied Biosystems, Foster City, Calif.) following the manufacturer's instructions and using the same forward and reverse primers adopted for PCR. All samples were sequenced on an ABI 373 automatic sequencer. Genotype 4 sequences were retrieved from GenBank. The sequences were aligned with the consensus sequences using the program ClustalX implemented in the Bioedit package (23) and were then manually adjusted. The phylogenetic analysis was performed using MEGA2 (26). The neighbor-joining method was used with 1,000 bootstrap replications.

The analysis was carried out with an HCV genotype 1 prototype as the outspecies. However, to enlarge the results, the outspecies was deleted from the trees in the figures, without changes in the topology of the representations.

The following sequences were included in the HCV E1 analysis: 1196-5 (D43678), 1196-8 (D43680), 1196-10 (D43681), B14 (L39282), B203 (L39284), CAM600 (L29587), CAM736-42 (L38323), CAR4/901 (L36440), CAR4/1205 (L36439), DK13 (L16656), ED43 (Y11604), FR12 (L38332), G22A (L22592), G22B (L22595), G27(L29597), G358 (L29606), GB116 (L29601), GB215 (L29604), GB438 (L29610), GB549 (L29620), GB809 (L29624), HEMA51 (D45193), NL40 (L39303), NL52 (L39310), NL81 (L39314), NL85 (L39315), NL88 (L39316), Z1 (L16677), Z4 (L16652), Z6 (L16678), and Z7 (L16653).

The following sequences were included in the HCV NS5b analysis: B14 (L44597), CAM600 (L29590), CAR1/501 (L36438), CAR4/1205 (L36437), EG13 (L23469), EG19 (L23470), EG81 (L78841), FR9 (L38376), FR12 (L38370), FR14 (L48493), G22 (L29596), GB116 (L29602), GB215 (L29605), GB358 (L29607), GB438 (L29611), GB48 (L29614), GB809 (L29626), and NL81 (L44603).

**Computer-assisted analysis of N glycosylation sites and epitope prediction.** N glycosylation sites were predicted using the on-line prediction server NetNGlyc version 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>), which predicts N glycosylation sites in proteins by using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequons (22). The prediction of T-cell epitopes was performed following the method of Yusim et al. (57) with minor modifications. In particular, the absence of a fully reconstructed database for experimentally defined HCV epitopes was overcome by means of T-cell epitope prediction using the on-line software SYFPEITHI (version 1.0), a database of major histocompatibility complex ligands and peptide motifs (<http://www.syfpeithi.de/>). SYFPEITHI is a database comprising >4,000 peptide sequences known to bind class I and class II major histocompatibility complex molecules compiled from published reports only (41, 42).

**Nucleotide sequence accession numbers.** All of the E1 sequences characterized in the present study have been submitted to GenBank under the indicated accession numbers: Egy4 (AY786475), Egy12 (AY787785), Egy13 (AY786476), Egy15 (AJ002758), Egy27 (AY786477), Egy28 (AY786478), Egy31 (AY786479), Egy33 (AY786480), Egy34 (AJ002755), Egy36 (AJ002753), Egy37 (AY786481), Egy39 (AY786482), Egy40 (AY786483), Egy42 (AY786484), Egy44 (AJ002757), Egy47 (AJ002759), Egy53 (AY786485), Egy58 (AJ002751), Egy68 (AJ002761), Egy75 (AJ002752), Egy76 (AY786486), Egy83 (AJ002754), Egy84 (AJ002762), Egy86 (AY786487), Egy87 (AY786488), Egy98 (AY786489), Egy104 (AY786490), Egy111 (AY786491), Egy127 (AY786492), Egy130 (AY786493), Egy133 (AY786494), Egy157 (AY786495), Egy162 (AY786496), Egy171 (AY786497), Egy193 (AJ002756), Egy198 (AY786498), and Egy205 (AJ002760). All of the NS5b sequences characterized in the present study have been submitted to GenBank under the indicated accession numbers: Egy13 (AJ002825), Egy15 (AJ002832), Egy28 (AJ002818), Egy29 (AJ002822), Egy31 (AJ002824), Egy34 (AJ002819), Egy37 (AJ002827), Egy39 (AJ002823), Egy44 (AJ002831), Egy47 (AJ002833), Egy49 (AJ002817), Egy58 (AJ002820), Egy68 (AJ002826), Egy75 (AJ002828), Egy84 (AJ002821), Egy193 (AJ002830), and Egy205 (AJ002816).

#### RESULTS

Partial E1 coding region sequences were aligned with type 4 prototypes derived from Maertens and Stuyver (reference 31 and data not shown) and with GenBank sequences. The phy-

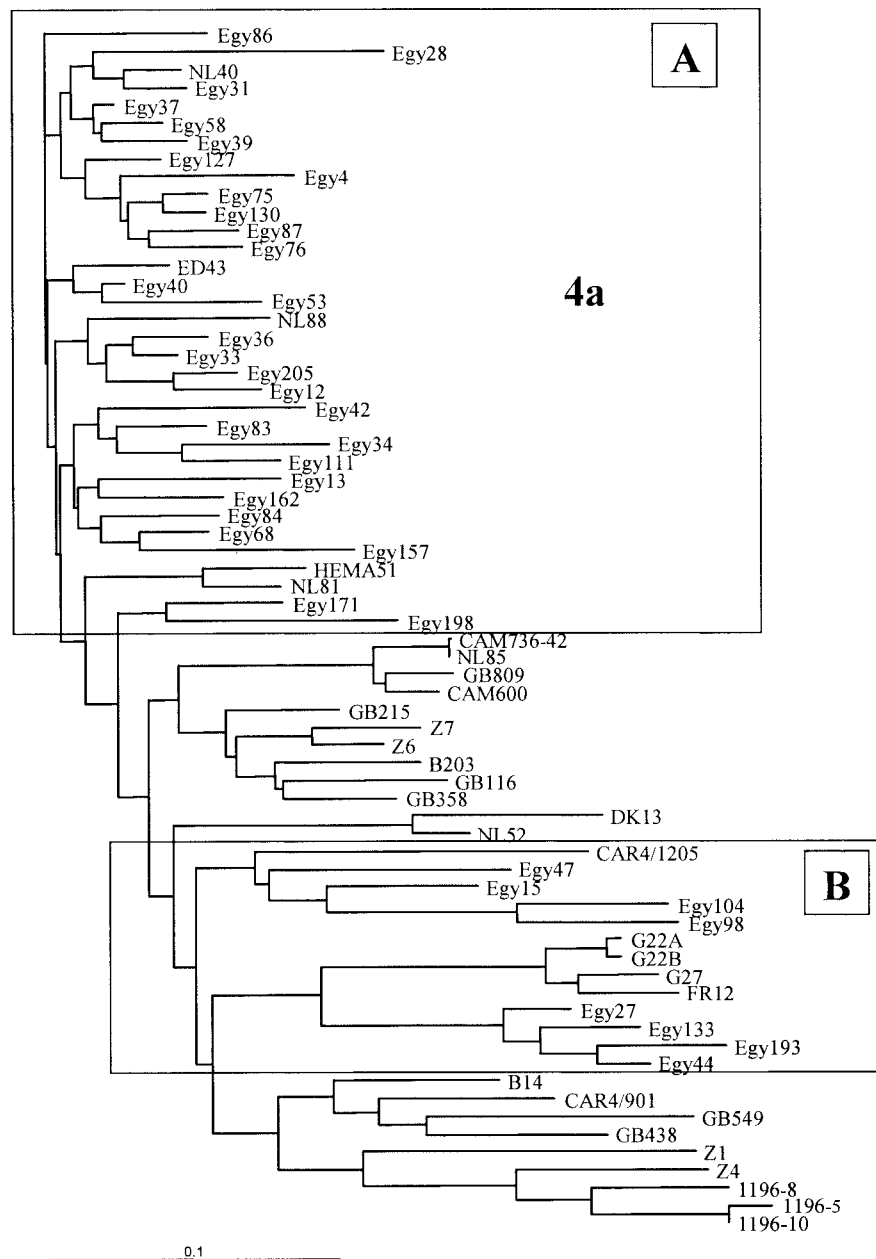


FIG. 1. Phylogenetic tree of the E1 region of genotype 4 isolates. (A) Subtype 4a isolates. (B) All other isolates. The scale bar shows substitutions per nucleotide site.

logenic tree obtained by performing neighbor-joining analysis of the alignment of sequences is shown in Fig. 1. The subtypes were named following the nomenclature of Maertens and Stuyver, with minor modifications. In particular, Z4 and GB809-4 were assigned to subtype 4l, and G22 and related isolates were assigned to subtype 4f. Based on the topology of the tree, we initially classified 28 of the isolates as subtype 4a (Fig. 1A). The molecular analysis of subtype 4a isolates showed a high level of heterogeneity. Compared to the sequences of 5 prototypes obtained from GenBank, 26 of the 29 isolates showed a genetic distance of 10 to 19%; 2 isolates, namely, Egy33 and Egy37, showed distances of <10%; and 1, Egy34, showed a distance of >19% (data not shown).

Eight isolates (Egy15, Egy27, Egy44, Egy47, Egy98, Egy104, Egy133, and Egy193) clustered in a defined branch of the tree (Fig. 1B). To better classify these isolates, we performed amplification of the NS5b region and genetic characterization of the four isolates for which a sufficient amount of serum was available (Egy15, Egy44, Egy47, and Egy193) and for nine randomly chosen isolates of subtype 4a. The results of the phylogenetic analysis, shown in Fig. 2, indicated the presence of three new subtypes (4m, 4n, and 4p).

Comparison of the genetic distances between the unclassified isolates and the subtype 4 prototypes in the E1 and NS5b regions confirmed the results of the phylogenetic analysis (data not shown) and indicated the presence of a fourth new subtype

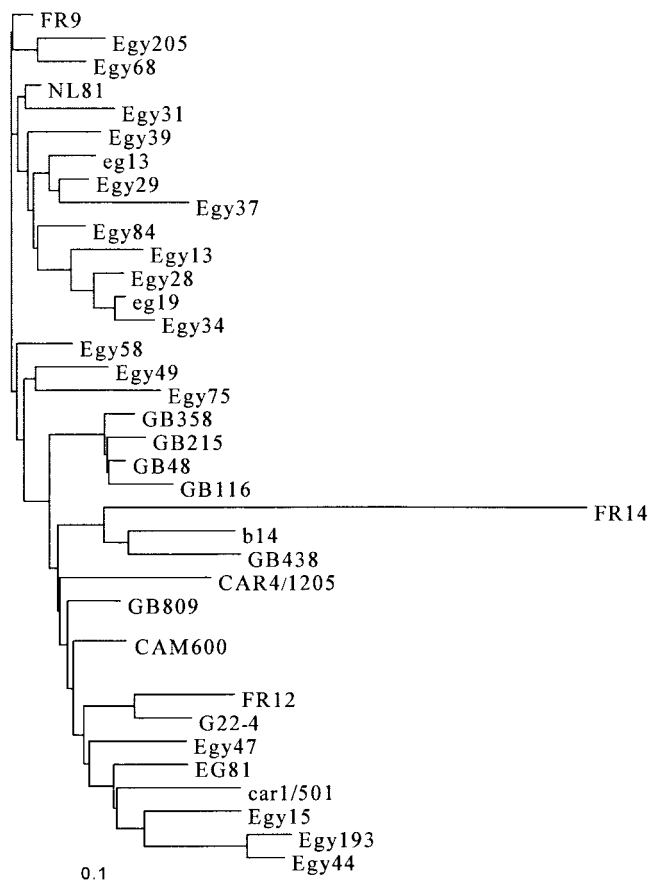


FIG. 2. Phylogenetic tree of the NS5b region of genotype 4 isolates. The scale bar shows substitutions per nucleotide site.

(4o) (Table 1). Four isolates (Egy27, Egy44, Egy133, and Egy193) were assigned to subtype 4m, with the intragroup genetic distances ranging from 7.44 to 13.47% (mean, 10.92%). Egy15 and Egy47 were classified as 4p and 4n, respectively. Egy15 (4p) was highly related to subtypes 4c and 4e, whereas Egy47 (4n) was equally related to subtypes 4c, 4e, and 4f. Two isolates (Egy98 and Egy104) were assigned to subtype 4o, with a genetic distance of 9.91%.

The distribution of the genotype 4 isolates was as follows: subtype 4a (*n* = 29), 78%; subtype 4m (*n* = 4), 11%; subtype 4o (*n* = 2), 5.5%; subtype 4n (*n* = 1), 2.7%; and subtype 4p (*n* = 1), 2.7%.

TABLE 1. Genetic distances in the E1 region of newly classified Egyptian isolates

Isolate	Distance <sup>a</sup>								Subtype
	Egy27	Egy44	Egy133	Egy193	Egy15	Egy47	Egy98	Egy104	
Egy27		0.1347	0.1141	0.1290	0.2492	0.3379	0.2674	0.2234	4m
Egy44			0.1195	0.0836	0.2604	0.3426	0.2990	0.2799	4m
Egy133				0.0740	0.2140	0.3390	0.2908	0.2538	4m
Egy193					0.2308	0.3340	0.2504	0.2269	4m
Egy15						0.3597	0.2473	0.2169	4p
Egy47							0.3098	0.3316	4n
Egy98								0.0991	4o
Egy104									4o

<sup>a</sup> The values range between 0 (0%) and 1 (100%) substitutions per nucleotide site.

TABLE 2. Probability of glycosylation in E2 ectodomain

aa position:	HVR1					HVR2	No. of sites
	34	40	47	65	93		
Isolate	Probability at position <sup>a</sup> :						
	34	40	47	65	93		
Ed43	++	-	++	-	+		3
Egy13	++	-	+	-	+		3
Egy28	++	-	++	-	+		3
Egy31	++	-	++	-	+		3
Egy37	++	-	++	+	+		3
Egy49	++	-	++	+ <sup>b</sup>	+		4 <sup>b</sup>
Egy68	++	-	++	+	+		3
Egy76	++	-	++	-	+		3
Egy87	++	-	++	+	-		3
Egy133	++	-	+	-	+		3
Egy157	++	-	++	-	-		2
Egy162	++	-	++	-	+		3
Egy198	+++	+	-	+	-		3
Egy205	++	-	++	-	+		3

<sup>a</sup> Glycosylation probability is shown by +++ (probability > 70%), ++ (probability between 60 and 70%), + (probability between 50 and 60%), and - (not present). In Egy198, N47 is replaced with V. aa, amino acid.

<sup>b</sup> Egy49 is represented by two main variants. One variant did not present N glycosylation at position 65, and thus only three sites can be predicted.

The portion of the E2 ectodomain that contains HVR1 in subtype 4a isolates was further studied. The variations in the predicted glycosylation sites are shown in Table 2. Whereas other studies have reported five glycosylation sites in this region, in the present study, all of the isolates showed three different glycosylation sites, with the exception of Egy157, which showed only two sites. Differences in the probability of glycosylation and in the positions of the different sites were also observed. The isolate Egy49 is present in two main variants, one of which had an additional glycosylation site in position 65.

The presence of predicted epitopes in HVR1 and the E2 ectodomain is shown in Table 3. The two regions showed similar features, with the probability of epitopes per amino acid ranging from 0.00 (in HVR1 of Egy28, Egy31, and Egy162) to 0.13 (in HVR1 of Egy68 and Egy157 and in the E2 ectodomain of Egy198). The number of predicted epitopes ranged from 6 (Egy31) to 16 (Egy68 and Egy198).

Recently, a combination of computer-assisted analyses has been shown to predict epitopes actually utilized in vivo (20). In our study, these analyses were applied to the E2 ectodomain



TABLE 3. Number of T-cell epitopes predicted in HVR1 and E2 ectodomain

Isolate	No. of T-cell epitopes <sup>a</sup>	
	HVR1 (aa 1-30)	E2 (aa 31-133)
Ed43	1 (0.03)	9 (0.08)
Egy13	3 (0.10)	9 (0.08)
Egy28	0 (0.00)	9 (0.08)
Egy31	0 (0.00)	6 (0.06)
Egy37	2 (0.07)	9 (0.08)
Egy49	3 (0.10)	9 (0.08)
Egy68	4 (0.13)	12 (0.12)
Egy76	3 (0.10)	9 (0.08)
Egy87	2 (0.07)	7 (0.07)
Egy157	4 (0.13)	10 (0.07)
Egy162	0 (0.00)	8 (0.07)
Egy198	3 (0.10)	13 (0.13)
Egy205	3 (0.10)	9 (0.08)

<sup>a</sup> aa, amino acids. The value in parentheses indicates the probability of being involved in an epitope per amino acid site.

(Table 4). We observed five classes of epitopes. One class (HLA-A0201) represents an epitope that is not utilized because it lacks a proteasome cleavage site; amino acid mutations (e.g., L55F in Egy13) are responsible for decreases in the score that indicates affinity of binding. In a second class of epitopes (HLA-B2705), again the epitope is not utilized because it has no cleavage site, yet some of the mutations can abolish the sequence recognized as the epitope (i.e., Egy28, Egy76, and Egy87). In a third class of epitopes (HLA-A3 and HLA-B5101), there is a mutation that can abolish the proteasome cleavage (HLA-A3) or the predicted epitope (HLA-B5101). Finally, in a fourth class of epitopes (HLA-B5151), a single amino acid position (no. 111) determines the affinity of binding.

The variability in the HVR1 neutralization epitope is shown in Fig. 3. The Egyptian isolates in our study showed high variability (two isolates of subtype 2m and two isolates of subtype 2o were added to the analysis). Five of the isolates showed amino acids in positions in which they had never been observed (a total of six positions). Four of these residues (in isolates Egy27, Egy68, Egy162, and Egy205) were in positions involved in anchoring to the E2 glycoprotein core.

DISCUSSION

In Africa, circulating HCV is extremely heterogeneous (31). Genotype 5 is endemic in southern Africa. Genotype 1 circulates in north and central Africa, and provisional new subtypes have been characterized in Morocco, Cameroon, and Nigeria. Genotype 2 has been found in north, western, and central Africa. New subtypes have been identified in Guinea Conakry (46), The Gambia, and Nigeria (31). In Surinam and the French Antilles, subtypes of genotype 2 found in certain patients are supposed to have originated in Africa (31). Finally, genotype 4 is widespread in East Africa and Egypt (48).

Recently, there has been increased interest in the variability of HCV, especially in light of the fact that new vaccines based on the structural proteins of HCV are being developed. The two HCV N glycosylated envelope glycoproteins, E1 and E2, are believed to be type 1 transmembrane proteins with N-terminal ectodomains and C-terminal hydrophobic anchors. Chimpanzees vaccinated with purified recombinant E1-E2 het-

TABLE 4. Epitope prediction analysis of E2 ectodomain<sup>a</sup>

Isolate	HLA-A0201			HLA-B2705			HLA-B5151			HLA-A3			HLA-B5101			
	49 SLNTGFLASL 58	C	Pr	71 GRLASCRSL 79	C	Pr	107 YAPRPCGI 114	C	Pr	54 FLASLFYVHK 63	C	Pr	87 GPLGVANI 94	C	Pr	Score
Ed43		-	N	E C K	-	N		+	P	T	+	N		+	P	31
Egy13	F	-	N	E R K	-	N		+	P	F	+	N		+	P	25
Egy28		-	N	E R K	-	N		+	P	T	+	N		+	P	31
Egy31		-	N	E E	-	N	Q	+	P	A	+	N	24	+	P	31
Egy37		-	N	E K	-	N		+	P	T	+	N		-	N	31
Egy49a		-	N	E G K	-	N		+	P	A	+	N	24	-	N	31
Egy49b		-	N	E G K	-	N		+	P	T	+	N		-	N	31
Egy68	Q	-	N	Q R KR	-	N	Q	+	P	N	+	N	25	-	N	28
Egy76		-	N	E S K	-	N		+	P	FQ	+	N		+	P	31
Egy87		-	N	D S K	-	N		+	P	T	+	N		+	P	31
Egy157	ISG	-	N	E RR	-	N	P	+	P	ISG AF	+	N		+	P	28
Egy162		-	N	E Q	-	N	Q	+	P	G TR	+	N		-	N	31
Egy198	G	-	N	E Q	-	N		+	P	G T	+	N	24	-	N	30
Egy205		-	N	E K	-	N		+	P		+	N	25	+	P	31

<sup>a</sup> Each isolate was analyzed for the presence of alpha helix, proteasome cleavage, and prediction. The table shows the epitopes common to all isolates. The sequence and the position are shown. The presence or absence of proteasome cleavage (C) is indicated by + or -, respectively. The probability of epitope existence is summarized in column Pr (N, not probable; P, probable). Only scores of >23 are considered.

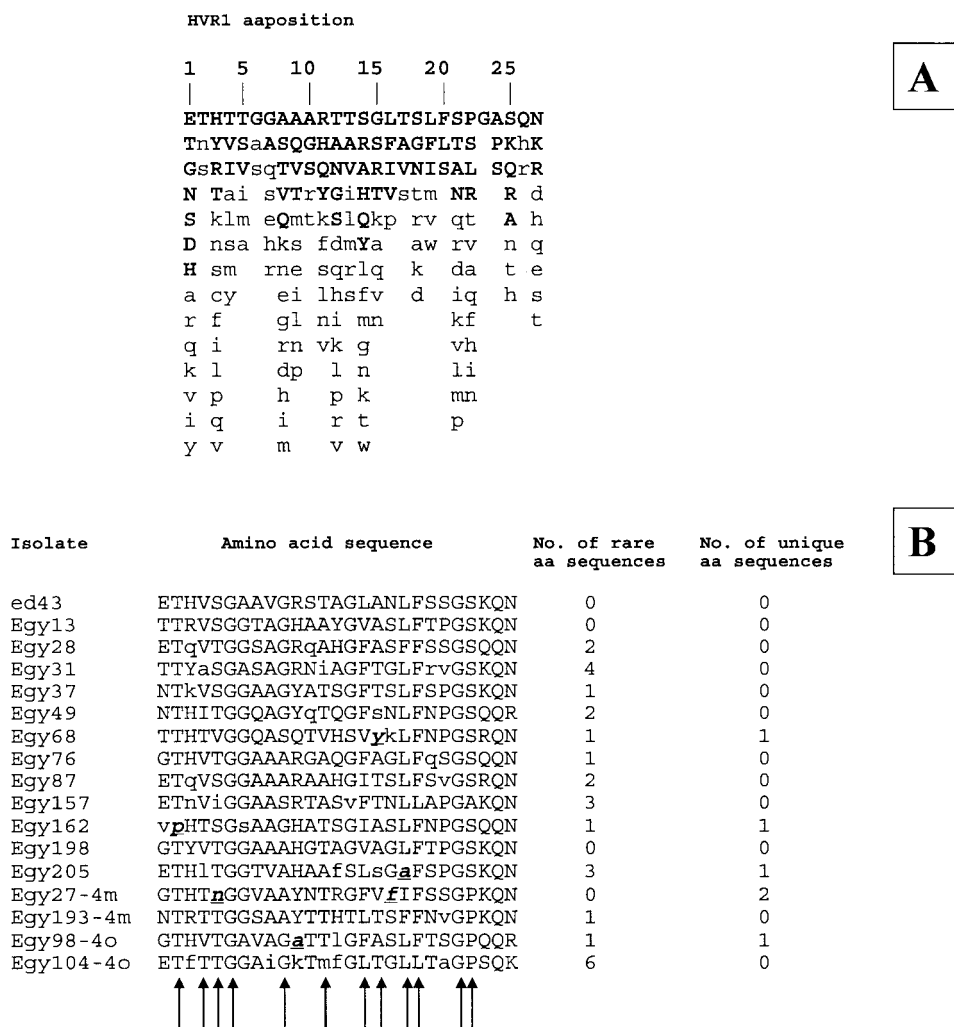


FIG. 3. Variability in the HVR1 neutralization epitope. (A) Amino acid (aa) variations among 1,382 isolates of different genotypes and subtypes extracted from DataBank (40). The first line represents the consensus sequence. (B) Amino acids are listed in decreasing order of observed frequency, from top to bottom. Uppercase letters represent amino acids with a >10% presence; bold lowercase letters represent amino acids found only in the Egyptian isolates of this study. The arrows indicate the residues essential for the anchoring of the E2 glycoprotein core and in maintaining the HVR1 conformation.

erodimeric proteins have been shown to be protected against challenge with homologous virus (10). However, given that quasispecies circulate in the bloodstreams of infected individuals, it remains to be determined whether the anti-E2 response elicited by one recombinant protein is effective against heterologous viral inocula. One of the targets of this immune response is the 27-amino-acid-long N-terminal segment of the E2 glycoprotein, known as HVR1, which is the most variable region of the HCV polyprotein and which contains the only neutralization determinant that has been identified. Anti-HVR1 antibodies specific for one variant display only a limited ability to neutralize different viral variants (17, 18). The efficacy of a vaccine based on structural proteins could be impaired if escape mutants were present (6, 25, 30, 38, 53, 58).

We analyzed the circulation and distribution of subtypes of clade 4 in Egypt and the variability of selected regions of the structural proteins, which could be important for the efficacy of an anti-HCV vaccine. The study population had been recruited in a previous survey in a hospital in the city of Alexandria,

although these individuals came from all over the Alexandria District, including rural villages (15).

The observed heterogeneity was much greater than expected, since previous studies mainly reported subtype 4a (for a review, see reference 49). In our study, although subtype 4a was found to be prevalent (78% of the cases), four new subtypes were found, with subtype 4m representing 11% of the cases and the other three subtypes representing another 11%. Substantial heterogeneity was also found when the intrasubtype 4a sequences were analyzed. The previously characterized prototypes (i.e., ED43, HEMA51, NL81, and NL40) showed distances ranging from 5 to 11.5% (mean, 9.4%), whereas, compared to the Egyptian isolates in our study, the mean genetic distance for these prototypes was 10.5%, with a range from 3.9 to 19.18%. Egy34 was the least related isolate, with a distance ranging from 13.5 to 19.2%.

Although Egypt is eligible as a model site for vaccine trials because the prevalence of HCV infection is high (10 to 14%) (2), the extreme heterogeneity found at all levels in this HCV

clade must be taken into consideration. Factors that contribute to the development of an effective vaccine and that are influenced by viral heterogeneity include the proper folding of the HCV structural proteins used, the recognition of epitopes by CTLs, and the eliciting of a neutralizing immune response.

A structural model based on homology modeling with the flavivirus tick-borne encephalitis virus has predicted the positions of the 11 potential N-linked glycosylation sites in the E2 glycoprotein (56). Mutagenesis studies suggest that these sites are used during the posttranslational processing of nascent E1-E2 complexes (5, 19, 50). Our analysis of the first five N-linked glycosylation sites revealed that our isolates were quite different from the model, in that in most cases there were only three glycosylated sites. Moreover, among our isolates, there were differences in the pattern of glycosylation (Table 4). However, these data do not take into account differences in conformational epitopes, which could be influenced by the interaction of E2 with E1.

Conformational epitopes are crucial in cell-mediated immunity, and there is growing awareness of the importance of the HCV-specific CTL response in viral clearance. This response may contribute to the long-term control of HCV (29), and it is critical to the development of an effective vaccine. The pattern of variation of the predicted CTL epitopes in the E2 protein ectodomain (which contains, in HVR1, the only known neutralization epitope) in the subtype 4a isolates showed no differences between HVR1 and the rest of the E2 ectodomain, expressed as the probability per position of being included in the CTL epitope. As shown in Table 4, the variability could drastically alter epitope prediction. In human immunodeficiency virus type 1 (57), escape mutations that inhibit processing or HLA binding have been observed to occur more often in variable regions, conferring a generic form of CTL escape. If a cleavage site is lost or a peptide cannot bind to any class I molecule, then the accumulation of such mutations decreases the immunogenic potentials of the regions where they are located, especially in regions that readily tolerate change. At the population level, this means that the virus may have evolved toward a state more refractive to the CTL response. Our data support the possible influence of variability in the escape of the virus from cellular recognition. The fact that HVR1 contains the only known neutralization epitope may also be important for vaccine development. Nucleotide sequences encoding HVR1 fragments are highly variable (up to 80%), apparently because of antibody selection of immune escape variants. However, the examination of aligned amino acid sequences has suggested that there is restricted amino acid usage in certain positions, probably because HVR1 could interact with negatively charged molecules at the cell surface, playing a role in host cell recognition and attachment, as well as in the cellular compartmentalization of the virus. In our study, five of the isolates showed amino acids in six novel positions. Four of these residues (in four different isolates) were in positions involved in anchoring to the E2 glycoprotein core and in maintaining the HVR1 conformation.

The results of this study indicate that HCV genotype 4 in Egypt is extremely variable, not only in terms of sequence, but also in terms of functional and immunological determinants. These data should be taken into account in planning vaccine trials in Egypt. Moreover, given that a cellular system of infec-

tion for HCV does not exist, *in vitro* neutralization studies cannot be performed, and thus, it is not possible to determine whether genotypes can be recognized as serotypes. In fact, cross-neutralization could clarify whether specific genotypes and subtypes represent an actual obstacle to effective mass vaccination.

#### ACKNOWLEDGMENTS

We thank Wael Refai (Medical Research Institute, Alexandria, Egypt) for his invaluable clinical support and the collection of serum samples, Mark Kanieff for his helpful review of the manuscript, and Romina Tomasetto for secretarial and editorial assistance.

This work was performed as part of the Viral Hepatitis National Projects of the Istituto Superiore di Sanità (D. leg.vo no. 502) and was also funded by the project of the Istituto Superiore di Sanità Etiological, Diagnostic, and Structural Characterization of Hepatitis Viruses (no. C3NE).

#### REFERENCES

- Abele, R., and R. Tampe. 1999. Function of the transport complex TAP in cellular immune recognition. *Biochim. Biophys. Acta* **1461**:405-419.
- Angelico, M., E. Renganathan, C. Gandin, M. Fathy, M. C. Profili, W. Refai, A. De Santis, A. Nagi, G. Amin, L. Capocaccia, F. Callea, M. Rapicetta, G. Badr, and G. Rocchi. 1997. Chronic liver disease in the Alexandria governorate, Egypt: contribution of schistosomiasis and hepatitis virus infections. *J. Hepatol.* **26**:236-243.
- Argentini, C., S. Dettori, U. Villano, V. Guadagnino, D. Infantolino, P. Dentico, R. C. Coppola, and M. Rapicetta. 2000. Molecular characterisation of HCV genotype 4 isolates circulating in Italy. *J. Med. Virol.* **62**:84-90.
- Beekman, N. J., P. A. van Veelen, T. van Hall, A. Neisig, A. Sijts, M. Camps, P. M. Kloetzel, J. J. Neefjes, C. J. Melief, and F. Ossendorp. 2000. Abrogation of CTL epitope processing by single amino acid substitution flanking the C-terminal proteasome cleavage site. *J. Immunol.* **164**:1898-1905.
- Bochtler, M., L. Ditzel, M. Groll, C. Hartmann, and R. Huber. 1999. The proteasome. *Annu. Rev. Biophys. Biomol. Struct.* **28**:295-317.
- Breitburd, F., and P. Coursaget. 1999. Human papillomavirus vaccines. *Semin. Cancer Biol.* **9**:431-444.
- Buseyne, F., and Y. Riviere. 2001. The flexibility of the TCR allows recognition of a large set of naturally occurring epitope variants by HIV-specific cytotoxic T lymphocytes. *Int. Immunol.* **13**:941-950.
- Cao, H., P. Kanki, J. L. Sankale, A. Dieng-Sarr, G. P. Mazzara, S. A. Kalams, B. Korber, S. Mboup, and B. D. Walker. 1997. Cytotoxic T-lymphocyte cross-reactivity among different human immunodeficiency virus type 1 clades: implications for vaccine development. *J. Virol.* **71**:8615-8623.
- Chassin, D., M. Andrieu, W. Cohen, B. Culmann-Penciolelli, M. Ostankovitch, D. Hanau, and J. G. Guillet. 1999. Dendritic cells transfected with the *nef* genes of HIV-1 primary isolates specifically activate cytotoxic T lymphocytes from seropositive subjects. *Eur. J. Immunol.* **29**:196-202.
- Choo, Q.-L., G. Kuo, R. Ralston, A. Weiner, D. Chien, G. Van Nest, J. Han, K. Berger, K. Thudium, C. Kuo, J. Kansopon, J. McFarland, A. Tabrizi, K. Ching, B. Moss, L. B. Cummins, M. Houghton, and E. Muchmore. 1994. Vaccination of chimpanzees against infection by the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **91**:1294-1298.
- Cohen, J. 1999. The scientific challenge of hepatitis C. *Science* **285**:26-30.
- Domingo, E. 1996. Biological significance of viral quasispecies. *Viral Hepatitis Rev.* **2**:247-261.
- Domingo, E., L. Menéndez-Arias, M. E. Quinones-Mateu, A. Holguin, M. Gutierrez-Rivas, M. A. Martinez, J. Quer, I. S. Novelli, and J. J. Holland. 1997. Viral quasispecies and the problem of vaccine-escape and drug-resistant mutants. *Prog. Drug. Res.* **48**:99-128.
- Dubuisson, J., H. H. Hsu, R. C. Cheung, H. B. Greenberg, D. G. Russel, and C. M. Rice. 1994. Formation and intracellular localization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia and Sindbis viruses. *J. Virol.* **68**:6147-6160.
- Dushenko, G., H. Schmilovitz-Weiss, D. Brown, F. McOmish, P. L. Yap, S. Sherlock, N. McIntyre, and P. Simmonds. 1994. Hepatitis C virus genotypes: an investigation of type-specific differences in geographic origin and disease. *Hepatology* **19**:13-18.
- Elena, S. F., R. Miralles, J. M. Cuevas, P. E. Turner, and A. Moya. 2000. The two faces of mutation: extinction and adaptation in RNA viruses. *IUBMB Life* **49**:5-9.
- Farci, P., H. J. Alter, D. C. Wong, R. H. Miller, S. Govindarajan, R. Engle, M. Shapiro, and R. H. Purcell. 1994. Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated *in vitro* neutralization. *Proc. Natl. Acad. Sci. USA* **16**:7792-7796.
- Farci, P., A. Shimoda, D. Wong, T. Cabezon, D. De Gioannis, A. Strazera, Y. Shimizu, M. Shapiro, H. J. Alter, and R. H. Purcell. 1996. Prevention of

- hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *Proc. Natl. Acad. Sci. USA* **96**:15394–15399.
19. Fournillier, A., C. Wychowski, D. Boucreux, T. F. Baumert, J. C. Meunier, D. Jacobs, S. Muguet, E. Depla, and G. Inchaupse. 2001. Induction of hepatitis C virus E1 envelope protein-specific immune response can be enhanced by mutation of N-glycosylation sites. *J. Virol.* **75**:12088–12097.
  20. Frahm, N., B. T. Korber, C. M. Adams, J. J. Szinger, R. Draenert, M. M. Addo, M. E. Feeney, K. Yusim, K. Sango, N. V. Brown, D. SenGupta, A. Piechocka-Trocha, T. Simonis, F. M. Marincola, A. G. Wurcel, D. R. Stone, C. J. Russell, P. Adolf, D. Cohen, T. Roach, A. StJohn, A. Khatri, K. Davis, J. Mullins, P. J. R. Goulder, B. D. Walker, and C. Brander. 2004. Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities. *J. Virol.* **78**:2187–2200.
  21. Goulder, P., D. Price, M. Nowak, S. Rowland-Jones, R. Phillips, and A. McMichael. 1997. Co-evolution of human immunodeficiency virus and cytotoxic T-lymphocyte responses. *Immunol. Rev.* **159**:17–29.
  22. Gupta, R., E. Jung, and S. Brunak. Unpublished data.
  23. Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**:95–98.
  24. Holland, J. J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol. 1982. Rapid evolution of RNA genomes. *Science* **215**:1577–1585.
  25. Janssens, W., A. Buvé, and J. N. Nkengasong. 1997. The puzzle of HIV-1 subtypes in Africa. *AIDS* **11**:705–712.
  26. Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**:1244–1245.
  27. Lanford, R. E., L. Notvall, D. Chavez, R. White, G. Frenzel, C. Simonsen, and J. Kim. 1993. Analysis of hepatitis C virus capsid, E1, and E2/NS1 proteins expressed in insect cells. *Virology* **197**:225–235.
  28. Lauer, G. M., and B. D. Walker. 2001. Hepatitis C virus infection. *N. Engl. J. Med.* **345**:41–52.
  29. Lechner, F., D. K. Wong, P. R. Dunbar, R. Chapman, R. T. Chung, P. Dohrenwend, G. Robbins, R. Phillips, P. Klenerman, and B. D. Walker. 2000. Analysis of successful immune responses in persons infected with hepatitis C virus. *J. Exp. Med.* **191**:1499–1512.
  30. Lok, A. S. 2000. Hepatitis B infection: pathogenesis and management. *J. Hepatol.* **32**:89–97.
  31. Maertens, G., and L. Stuyver. 1997. Genotypes and genetic variation of hepatitis C virus, p. 183–233. *In* T. J. Harrison and A. J. Zuckerman (ed.), *The molecular medicine of viral hepatitis*. John Wiley & Sons, Ltd., Chichester, United Kingdom.
  32. Matsuura, Y., T. Suzuki, R. Suzuki, M. Sato, H. Aizaki, I. Saito, and T. Miyamura. 1994. Processing of E1 and E2 glycoproteins of hepatitis C virus expressed in mammalian and insect cells. *Virology* **205**:141–150.
  33. McAdam, S., P. Klenerman, L. Tussey, S. Rowland-Jones, D. Laloo, R. Phillips, A. Edwards, P. Giangrande, A. L. Brown, and F. Gotch. 1995. Immunogenic HIV variant peptides that bind to HLA-B8 can fail to stimulate cytotoxic T lymphocyte responses. *J. Immunol.* **155**:2729–2736.
  34. McKnight, A., R. A. Weiss, C. Shotton, Y. Takeuchi, H. Hoshino, and P. R. Clapham. 1995. Change in tropism upon immune escape by human immunodeficiency virus. *J. Virol.* **69**:3167–3170.
  35. McMichael, A. J., and R. E. Phillips. 1997. Escape of human immunodeficiency virus from immune control. *Annu. Rev. Immunol.* **15**:271–296.
  36. Mori, K., Y. Yasutomi, S. Ohgimoto, T. Nakasone, S. Takamura, T. Shioda, and Y. Nagai. 2001. Quintuple deglycosylation mutant of SIVmac239 in rhesus macaques: robust primary replication, tightly contained chronic infection, and elicitation of potent immunity against the parental wild-type strain. *J. Virol.* **75**:4023–4028.
  37. Niedermann, G., E. Geier, M. Lucchiari-Hartz, N. Hitziger, A. Ramsperger, and K. Eichmann. 1999. The specificity of proteasomes: impact on MHC class I processing and presentation of antigens. *Immunol. Rev.* **172**:29–48.
  38. Osmanov, S., W. L. Heyward, and J. Esparza. 1996. HIV-1 genetic variability: implications for the development of HIV vaccines. *Antibiot. Chemother.* **48**:30–38.
  39. Pamer, E., and P. Cresswell. 1998. Mechanisms of MHC class I-restricted antigen processing. *Annu. Rev. Immunol.* **16**:323–358.
  40. Penin, F., C. Combet, G. Germanidis, P. O. Frainais, G. Deléage, and J. M. Pawlotsky. 2001. Conservation of the conformation and positive charges of hepatitis C virus envelope glycoprotein hypervariable region 1 points to a role in cell attachment. *J. Virol.* **75**:5703–5710.
  41. Rammensee, H. G., J. Bachmann, N. N. Emmerich, O. A. Bachor, and S. Stevanovic. 1999. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**:213–219.
  42. Rammensee, H. G., J. Bachmann, and S. Stevanovich. 1997. MHC ligands and peptide motifs. Landes Bioscience, Georgetown, Tex.
  43. Rapicetta, M., C. Argentini, S. Dettori, E. Spada, G. Pellizzer, and C. Gandini. 1998. Molecular heterogeneity and new subtypes of HCV genotype 4. *Res. Virol.* **149**:293–297.
  44. Rapicetta, M., A. F. Attili, A. Mele, A. De Santis, P. Chionne, K. Cristiano, E. Spada, E. Giuliani, L. Carli, F. Goffredo, and L. Capocaccia. 1992. Prevalence of hepatitis C virus antibodies and hepatitis C virus-RNA in an urban population. *J. Med. Virol.* **37**:87–92.
  45. Robertson, B., G. Myers, C. Howard, T. Brettin, J. Bukh, B. Gaschen, T. Gojobori, G. Maertens, M. Mizokami, O. Nainan, S. Netesov, K. Nishioka, T. Shin, P. Simmonds, D. Smith, L. Stuyver, A. Weiner, et al. 1998. Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. *Arch. Virol.* **143**:2493–2503.
  46. Ruggieri, A., C. Argentini, F. Kouruma, P. Chionne, E. D'Ugo, E. Spada, S. Dettori, S. Sabbatani, and M. Rapicetta. 1996. Heterogeneity of hepatitis C virus genotype 2 variants in West Central Africa (Guinea Conakry). *J. Gen. Virol.* **77**:2073–2076.
  47. Shimizu, Y. K., M. Hijikata, A. Iwamoto, H. J. Alter, R. H. Purcell, and H. Yoshikura. 1994. Neutralizing antibodies against hepatitis C virus and the emergence of neutralization escape mutant viruses. *J. Virol.* **65**:1494–1500.
  48. Simmonds, P. 1995. Variability of hepatitis C virus. *Hepatology* **21**:570–583.
  49. Simmonds, P. 2004. Genetic diversity and evolution of hepatitis C virus—15 years on. *J. Gen. Virol.* **85**:3173–3188.
  50. Slater-Handshy, T., D. A. Droll, X. Fan, A. M. Di Bisceglie, and T. J. Chambers. 2004. HCV E2 glycoprotein: mutagenesis of N-linked glycosylation sites and its effect on E2 expression and processing. *Virology* **319**:36–48.
  51. Sobolev, B. N., V. V. Poroikov, L. V. Olenina, E. F. Kolesanova, and A. L. Archakov. 2000. Comparative analysis of amino acid sequences from envelope proteins isolated from different hepatitis C virus variants: possible role of conservative and variable regions. *J. Viral Hepatol.* **7**:368–374.
  52. Spaete, R. R., D. Alexander, M. E. Rugroden, Q. L. Choo, K. Berger, K. Crawford, C. Kuo, S. Leng, C. Lee, R. Ralston, K. Thudium, J. W. Tung, G. Kuo, and M. Houghton. 1992. Characterization of the hepatitis C virus E2/NS1 gene product expressed in mammalian cells. *Virology* **188**:819–830.
  53. Sullender, W. M. 2000. Respiratory syncytial virus genetic and antigenetic diversity. *Clin. Microbiol. Rev.* **13**:1–15.
  54. Van Doorn, L. J., I. Capriles, G. Maertens, R. DeLeys, K. Murray, T. Kos, H. Schellekens, and W. Quint. 1995. Sequence evolution of the hypervariable region in the putative envelope region E2/NS1 of hepatitis C virus is correlated with specific humoral immune response. *J. Virol.* **69**:773–778.
  55. Wilson, C. C., R. C. Brown, B. T. Korber, B. M. Wilkes, D. J. Ruhl, D. Sakamoto, K. Kunstman, K. Luzuriaga, I. C. Hanson, S. M. Widmayer, A. Wiznia, S. Clapp, A. J. Ammann, R. A. Koup, S. M. Wolinsky, and B. D. Walker. 1999. Frequent detection of escape from cytotoxic T-lymphocyte recognition in perinatal human immunodeficiency virus (HIV) type 1 transmission: the Ariel project for the prevention of transmission of HIV from mother to infant. *J. Virol.* **73**:3975–3985.
  56. Yagnik, A. T., A. Lahm, A. Meola, R. M. Roccasecca, B. B. Ercole, A. Nicosia, and A. Tramontano. 2000. A model for the hepatitis C virus envelope glycoprotein E2. *Proteins* **40**:355–366.
  57. Yusim, K., C. Kesmir, B. Gashen, M. M. Addo, M. Altfeld, S. Brunak, A. Chigaev, V. Detours, and B. T. Korber. 2002. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J. Virol.* **76**:8757–8768.
  58. Zeuzem, S. 2000. Hepatitis C virus: kinetics and quasispecies evolution during antiviral therapy. *Forum* **10**:32–42.