Amino Acids (2008) 34: 479–484
DOI 10.1007/s00726-007-0539-7
Printed in The Netherlands

Amino Acids

Sub-epitopic dissection of HCV $E1_{315–328}$HRMAWDMMMMNWSPT sequence by similarity analysis

L. Polimeno[1], **A. Mittelman**[2], **L. Gennero**[3], **A. Ponzetto**[3], **G. Lucchese**[4], **A. Stufano**[4], **A. Kusalik**[5], and **D. Kanduc**[4]

[1] Department of Emergency and Organ Transplantation, Gastroenterology Section, University of Bari, Bari, Italy
[2] Division of Oncology/Hematology, Department of Medicine, and Department of Microbiology and Immunology, New York Medical College, Valhalla, USA
[3] Department of Internal Medicine, University of Turin, Turin, Italy
[4] Department of Biochemistry and Molecular Biology "Ernesto Quagliariello", University of Bari, Bari, Italy
[5] Department of Computer Science, University of Saskatchewan, Saskatoon, Canada

Received January 15, 2007
Accepted February 23, 2007
Published online April 26, 2007;

Summary. Our labs are focused on identifying amino acid sequences having the ability to react specifically with the functional binding site of a complementary antibody. Our epitopic definition is based on the analysis of the similarity level of antigenic amino acid sequences to the host proteome. Here, the similarity profile to the human proteome of an HCV E1 immunodominant epitope, i.e. the HCV $E1_{315–328}$HRMAW DMMMNWSPT sequence, led to i) characterizing the immunoreactive HCV E1 315–328 region as a sequence endowed with a low level of similarity to human proteins; ii) defining 2 contiguous immunodominant linear determinants respectively located at the NH_2 and COOH terminus of the conserved viral antigenic sequence. This study supports the hypothesis that low sequence similarity to the host's proteome modulates the pool of epitopic amino acid sequences in a viral antigen, and appears of potential value in defining immunogenic viral peptide sequences to be used in immunotherapeutic approaches for HCV treatment.

Keywords: Epitope mapping – Similarity level – Amino acid sequence uniqueness – Computational biology – Anti-HCV peptide vaccines

Introduction

We are defining peptide-based vaccine constructs using the sequence similarity hypothesis, according to which amino acid sequences characterized by a low level of similarity (or no similarity at all) to the host's proteome have the potential to evoke immunogenic responses in the host (Natale et al., 2000; Lucchese et al., 2005). The final aim consists in using immunogenic peptides from tumor-associated-antigens to evoke specific anti-tumor antibodies or, on the opposite, using immunogenic peptides from auto-antigens to block circulating harmful autoantibodies in autoimmune diseases. Preliminary clinical data appear positive and encourage pursuing such a research line (Angelini et al., 2006).

Given this premise, HCV polyprotein appears an interesting experimental model since it seems that this RNA virus escapes immunosurveillance by being highly mutable. Indeed, the rate of HCV nucleotide change is approximately 10^{-3} substitutions per site per year and generates numerous different HCV types, subtypes and quasispecies (Forns and Bukh, 1999). So, even if immunogenic HCV protein sequences exist, they might change (and possibly their immunogenicity fall) because of subsequent amino acid variations. This leads to two interwoven consequences: from one side the difficulty of designing an effective vaccine due to rapidly changing viral targets; from the other side the continuous and unflagging increase of the number in infected individuals in the world. HCV peculiarly is able to establish chronic infection and progressive incumbent hepatocellular carcinoma risk in a large proportion of those exposed (Tanaka et al., 2006).

In this context, the present study has utilized HCV E1 protein to study the role of sequence uniqueness in humoral anti-HCV immune responses since this envelope glycoprotein: 1) represents a main target of host immune reaction by being located on the virion surface; 2) has been already shown to host immunoreactive portions

(Olenina et al., 2002), and, finally, 3) one third of E1 sequence is highly conserved (Sobolev et al., 2000), so offering the possibility of evaluating the role of sequence uniqueness in immunogenicity. Due to the extreme heterogeneity of the HCV genotypes and the reported large number of complete and partial sequences (Forns and Bukh, 1999), we examined HCV genotype 1a, this subtype being also more common in Europe with a tendency to increase in Italy (Bortolotti et al., 2005).

Materials and methods

Sequence similarity analysis of E1 sequence (aa191-382, HCV genotype 1a, accession no. P26664) to the human proteome was conducted using overlapping viral pentamers as probes, and the peptide match program from Protein Information Resource (pir.georgetown.edu/pirwww) (Wu et al., 2002) as previously described (Natale et al., 2000; Lucchese et al., 2005). By using T-Coffee program (Notredame et al., 2000), peptide sequence alignment was conducted on HCV genotype 1a, isolate 1 (P26664); 1a, isolate H (P27958); 1b, isolate Taiwan (P29846); 1b, isolate BK (P26663); 1b, isolate Con1 (Q9WMX2); 1b, isolate Japanese (P26662); 1b, isolate HC-JT (Q00269); 2a, isolate JFH-1 (Q99IB8); 2a, isolate HC-J6 (P26660); 2b, isolate HC-J8 (P26661); 2b, isolate JPUT971017 (Q9DHD6). The program Drawhca (Gaboriaud et al., 1987) was utilized in hydrophobic cluster analysis.

The HCV synthetic peptides used in dot-blot immunoassays were: $E1_{315-328}$HRMAWDMMMNWSPT; $E1_{315-320}$HRMAWD; $E1_{316-321}$RMAWDM; $E1_{317-322}$MAWDMM; $E1_{318-323}$AWDMMM; $E1_{319-324}$WDMMMN; $E1_{320-325}$DMMMNW; $E1_{321-326}$MMMNWS; $E1_{322-327}$MMNWSP; $E1_{323-328}$MNWSPT; E2-$p7_{741-750}$SQAEAALENL; $NS3_{1160-1173}$LKGSSGGPLLCPAG; $NS3_{1628-1637}$GAVQNEITLT; $NS5A_{2182-2191}$AEAAGRRLAR. Control peptides were: $Tyrosinase_{4-18}$AVLYCLLWSFQTSAG; $Tyrosinase_{233-247}$IPYWDWRDAEKCDIC. Peptides were synthesized by Fmoc (N-(9-fluorenyl) methoxycarbonyl) solid phase peptide synthesis (Primm srl, Milan, Italy).

Serum samples from HCV-infected patients were obtained from the Dept. of Internal Medicine, University of Turin, Division of Gastroenterology, Hospital San Giovanni Battista, Turin, Italy. HCV studies had been approved by the Hospital Institutional Review Board. Sera were treated with RNase/DNase for 2 h at 37 °C, partially purified by precipitation with 40% saturated $(NH_4)_2SO_4$ (×2), and dissolved in phosphate-buffered saline (PBS).

Nitrocellulose membrane (0.1 μm, Schleicher&Schuell, Milan, Italy) was pretreated with 1% glutaraldehyde. Peptides (10 μg) were spotted on the activated membranes and immunoreacted with partially purified polyclonal antibodies (pAbs) from HCV-infected patients as described (Lucchese et al., 2005).

Results

Searching for sequence uniqueness along the HCV E1 protein

To understand the role of sequence similarity in determining HCV E1 immunogenicity in humans, HCV E1 protein was analyzed for similarity to human proteins.

To this aim, the E1 primary sequence (amino acids 191–382) was dissected into pentapeptides that were used as probes to scan the human proteome, since a pentapeptide unit is a sufficient minimal antigenic determinant in a protein (Lucchese et al., 2006, and refs. therein). Consecutive pentamers were offset by one residue, i.e. overlapped by four residues: YQVRN, QVRNS, VRNST, RNSTG, NSTGL, etc. The 5-mer sequences were analyzed for similarity utilizing the PIR non-redundant reference human protein database (Wu et al., 2002). The HCV E1 similarity profile is reported in Fig. 1. As already found in other experimental models (Natale et al., 2000; Lucchese et al., 2005), the sequence-sequence profiling has an alternating behaviour. Many viral pentamers appear repeatedly in human proteome and, at the same time, there are HCV E1 fragments with a low number of pentamers in common with the human proteome. In this regard, the sequence HCV $E1_{315-328}$HRMAWDMMMN WSPT (indicated by an arrow in Fig. 1) is relevant to the similarity analysis since most of its pentamers had no matches at all to the human proteome (see Fig. 1, lower insert). According to the hypothesis pursued in our labs, the HCV $E1_{315-328}$HRMAWDMMMNWSPT sequence appeared suitable to test the relationship between sequence uniqueness and immunogenicity in immunoassays with sera from HCV-infected patients.

Consequently, the $E1_{315-328}$HRMAWDMMMNWSPT peptide was synthesised for use as an antigen. HCV $NS3_{1160-1173}$LKGSSGGPLLCPAG peptide was used as a control, being characterized by a high level of similarity to the human proteome (total number of 5-mer matches >400) and a degree of sequence conservation comparable to that of the low-similarity $HCV_{315-328}$HRMAWDMMMNWSPT sequence under analysis (see Table 1).

As additional controls, we used HCV and tyrosinase peptides at hand in the lab and having high/medium similarity to human proteome: HCV E2-$p7_{741-750}$SQAEAA LENL (matches = 289), HCV $NS3_{1628-1637}$GAVQNEITLT (matches = 66) and HCV $NS5A_{2182-2191}$AEAAGRR LAR (matches = 263), $tyrosinase_{4-18}$AVLYCLLWSFQT SAG (matches = 59), and $tyrosinase_{233-247}$IPYWDWRD AEKCDIC (matches = 36). This last peptide appeared interesting as an external control by hosting a zero similarity fragment, i.e. $tyrosinase_{233-241}$IPYWDWRDA (Lucchese et al., 2005).

Immune response against the low-similarity HCV $E1_{315-328}$HRMAWDMMMNWSPT peptide

As a second step, the low-similarity HCV $E1_{315-328}$HRM AWDMMMNWSPT peptide, the conserved high-similarity HCV $NS3_{1160-1173}$LKGSSGGPLLCPAG peptide

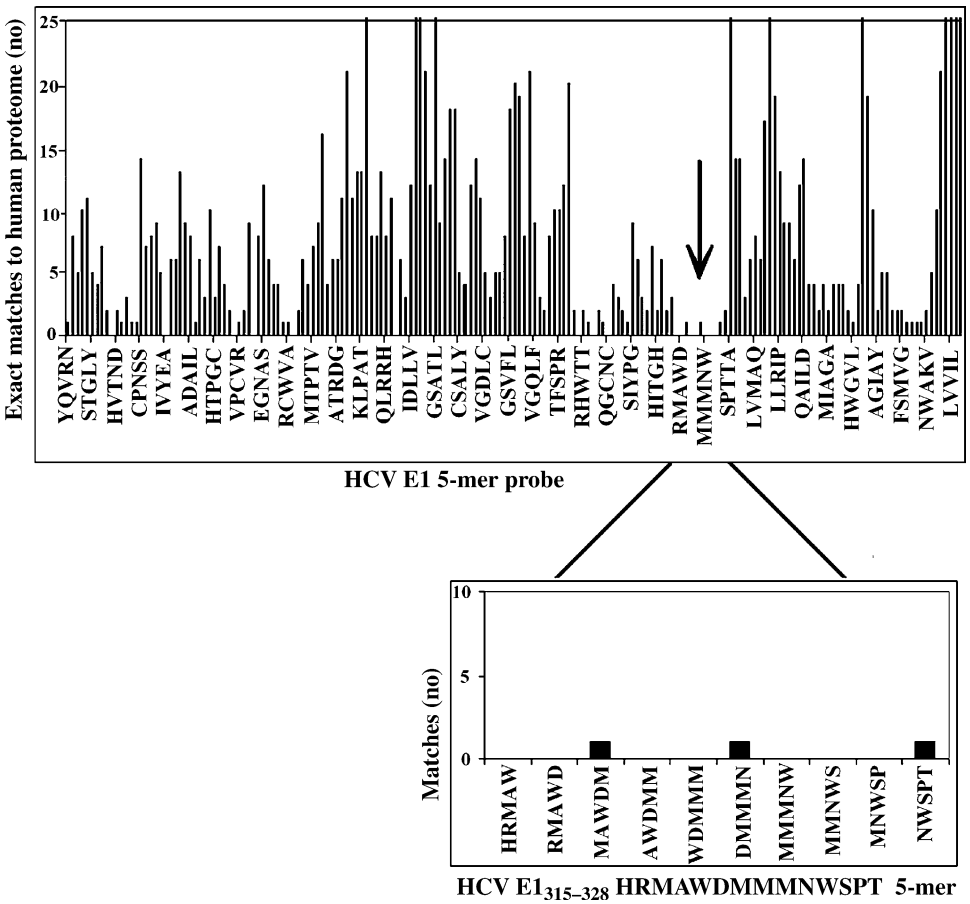

Fig. 1. Similarity profile of HCV E1 protein to human proteome by using 5-mer motifs as probes. Lower insert: similarity profile of the $E1_{315–328}$HRMAWDMMMNWSPT sequence

Table 1. Sequence conservation of the low-similarity $HCV_{315–328}$HRMAWDMMMNWSPT peptide and the high-similarity $HCV_{1160–1173}$LKGSSGGPLLCPAG peptide control among different HCV isolate genotypes

HCV genotypes	Peptide sequence
1a	HRMAWDMMMNWSPT LKGSSGGPLLCPAG
1a (H)	HRMAWNMMMNWSPT LKGSSGGPLLCPTG
2a (JFH-1)	HRMAWDMMMNWSPT LKGSSGGPVLCPRG
2a (HC-J6)	HRMAWDMMMNWSPT LKGSSGGPVLCPRG
1b (Taiwan)	HRMAWDMMMNWSPT LKGSSGGPLLCPSG
1b (BK)	HRMAWDMMMNWSPT LKGSSGGPLLCPFG
1b (Con1)	HRMAWDMMMNWSPT LKGSSGGPLLCPSG
1b (Japanese)	HRMAWDMMMNWSPT LKGSSGGPLLCPSG
1b (HC-JT)	HRMAWDMMMNWSPT LKGSSGGPLLCPSG
2b (HC-J8)	HRMAWDMMLSWSPT LKGSSGGPVLCSRG
2b (JPUT971017)	QRMAWDMMLNWSPT LKGSSGGPVLCPRG

Changed amino acids are given underlined

control, plus the above described additional HCV and tyrosinase peptide controls were analyzed for immunoreactivity using sera from 11 HCV-infected and 4 healthy subjects. Representative peptide dot-blot immunoassay analyses are shown in Fig. 2. Table 2 details the recognition pattern of the 15 sera monitored in the study.

Dot-blot immunoassays indicate that the unshared $E1_{315–328}$HRMAWDMMMNWSPT peptide is immunoreactive with almost all of the HCV-infected sera (81%, 9 out of 11 sera). No signal was detected using the 4 HCV peptides endowed with medium to high level of similarity to the human proteins (similarity is given in Fig. 2 as match number in parentheses). Moreover, whereas the $tyrosinase_{4–18}$AVLYCLLWSFQTSAG peptide control did not react with any of the sera, the second external control represented by $tyrosinase_{233–247}$IPYWDWRDAEKCDIC showed intense immunoreactivity with all of the sera used in this study, both the healthy and HCV-infected ones. While the HCV $E1_{315–328}$HRMAWDMMMNWSPT-induced signal was weak, the intensity of the $tyrosinase_{233–247}$IPYWDWRDAEKCDIC signal was comparatively much higher (Fig. 2 and Table 2). As recently reported (Lucchese et al., 2005), the $tyrosinase_{233–247}$IPYWDWRDAEKCDIC sequence immunoreacts with vitiligo, melanoma and healthy sera, apparently deriving

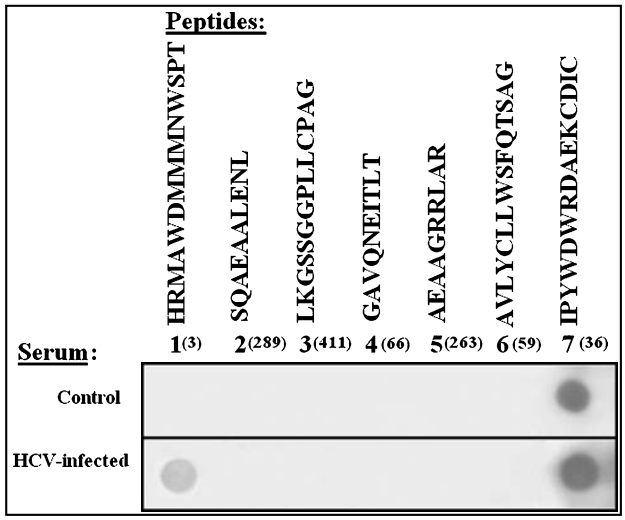

Fig. 2. Immunoreactivity of the low-similarity HCV $E1_{315–328}$HRMAWDMMMNWSPT peptide with pAbs from human HCV-infected serum. Dot immunoassay was performed using pAbs from normal healthy subject as a control. HCV peptide: 1) $E1_{315–328}$HRMAWDMMMNWSPT; 2) E2-$p7_{741–750}$SQAEAALENL; 3) $NS3_{1160–1173}$LKGSSGGPLLCPAG; 4) $NS3_{1628–1637}$GAVQNEITLT; 5) $NS5A_{2182–2191}$AEAAGRRLAR. Tyrosinase peptides: 6) $Tyrosinase_{4–18}$AVLYCLLWSFQTSAG; 7) $Tyrosinase_{233–247}$IPYWDWRDAEKCDIC. In parentheses, peptide similarity level given as number of perfect 5-mer matches shared between the viral peptide sequence and the human proteome

its immunogenicity from a low-similarity portion in the NH_2 terminus.

Identifying antibody binding sites on the $E1_{315–328}$HRMAWDMMMNWSPT peptide sequence

Successively, we characterized the linear epitope by identifying residues on the $E1_{315–328}$HRMAWDMMMNWSPT peptide sequence that are important for binding human pAbs. Indeed, as shown in Fig. 3, the $E1_{315–328}$HRMAW DMMMNWSPT peptide is characterized by hydrophobic clustering that makes this viral amino acid sequence susceptible to three-dimensional conformer structures. Therefore the possibility exists that the peptide under study might interact with the human pAbs through conformational epitope(s). Given our starting hypothesis, according to which linear viral pentamers not represented in the host proteome should be the potential epitopic targets of the immune response, it was important to verify this point.

To this aim, fine epitope mapping was performed using nine consecutive 6-mer peptides containing five overlapping residues and covering the $E1_{315–328}$HRMAWDMM MNWSPT sequence. Figure 4 illustrates the representative immunoreactivity pattern of the nine 6-mers when assay-

Table 2. Reactivity of the low-similarity HCV $E1_{315–328}$HRMAWDMMMNWSPT peptide with sera from HCV-infected patients

Serum	HCV peptide					Tyrosinase peptide	
	1 (3)	2 (289)	3 (411)	4 (66)	5 (263)	6 (59)	7 (36)
Healthy							
1	−	−	−	−	−	−	+++
2	±	−	−	−	−	−	+++
3	−	−	−	±	+	−	++
4	±	−	−	−	+	−	+++
HCV infected							
5	++	−	−	−	−	−	++
6	++	−	−	−	−	−	+++
7	++	−	−	−	−	−	+++
8	++	−	−	+	−	−	+++
9	+++	−	−	−	+	−	+++
10	+++	−	−	−	−	−	+++
11	++	−	−	−	−	±	+
12	+++	−	−	−	−	−	+++
13	+++	−	−	−	−	−	+++
14	±	±	−	−	−	−	+++
15	+	+	−	+	−	−	++

The synthetic peptides correspond to the amino acid sequences: 1) HCV $E1_{315–328}$HRMAWDMMMNWSPT; 2) HCV E2-$p7_{741–750}$SQAEAALENL; 3) HCV $NS3_{1160–1173}$LKGSSGGPLLCPAG; 4) HCV $NS3_{1628–1637}$GAVQNEITLT; 5) HCV $NS5A_{2182–2191}$AEAAGRRLAR; 6) $Tyrosinase_{4–18}$AVLYCLLWSFQTSAG; 7) $Tyrosinase_{233–247}$IPYWDWRDAEKCDIC. In parentheses: peptide similarity level as number of perfect 5-mer matches shared between the viral peptide sequence and the human proteome. Signal intensity shown with increased numbers of plus signs

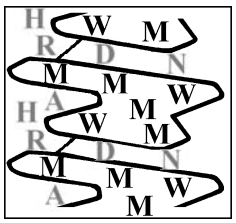

Fig. 3. Two-dimensional surface-like map of the potential hydrophobic domain structures of the HCV $E1_{315-328}$HRMAWDMMMNWSPT peptide epitope analysed in this study

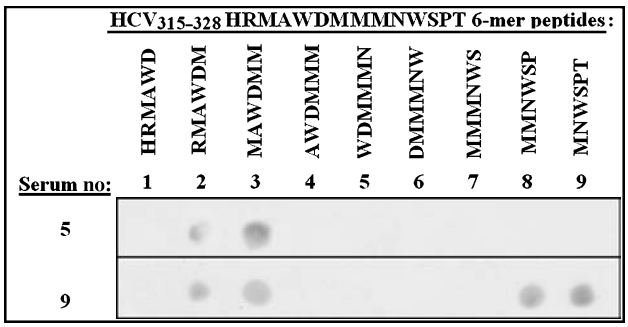

Fig. 4. Identification of epitopic fragments on the low-similarity HCV $E1_{315-328}$HRMAWDMMMNWSPT peptide sequence. Epitope scanning by dot-blot immunoassay was carried out using pAbs from HCV-infected sera (see Table 2)

ed by dot-blot experiments using HCV-infected sera. It can be seen that the main epitope sequence on $E1_{315-328}$HR MAWDMMMNWSPT is between amino acids 316 and 322 (RMAWDMM) located at the N-terminal of the $E1_{315-328}$HRMAWDMMMNWSPT peptide. All of the HCV-infected sera that were immunoreactive towards the $E1_{315-328}$HRMAWDMMMNWSPT peptide (that is from no. 5 to 13, see Table 2) recognized the aa 316–322 sequence (9 out of 9; 100%). Additionally, Fig. 4 shows that HCV-infected serum 9 (and serum 6, not shown) recognized a second domain between aa 322 and 328 (MMNWSPT), thus indicating a possible phenomenon of intramolecular epitope spreading along the $E1_{315-328}$HR MAWDMMMNWSPT peptide sequence.

Discussion

In this study we examined the role of sequence similarity in determining HCV E1 immunogenicity in human subjects. To this aim, we have 1) undertaken a molecular dissection of the conserved HCV E1 protein sequence searching for peptide stretches scarcely represented in the human proteome, and 2) measured the immunoreactivity of the selected low-similarity HCV peptides in immunoassays with sera from HCV-infected patients. We report that $E1_{316-322}$RMAWDMM peptide forms a common sequence signature of the humoral immune response from the HCV-infected cohort examined here. The epitope is located in the low-similarity $E1_{315-328}$HRMAWDM MMNWSPT peptide sequence, and is flanked by a contiguous second non-immunodominant linear $E1_{322-328}$MMN WSPT determinant, possibly arising by intramolecular epitope spreading.

As a first general note, we observe that the procedure described here results in the identification of an immunodominant viral epitopic sequence with only a minimum number of synthetic peptides.

As regards the specific anti-HCV response, this study indicates the presence of effective autoantibodies in HCV-infected patients (see Fig. 2 and Table 2). Therefore, even if HCV infection (and maintenance of the chronic infection) appears associated with an immunosuppressed status of the host (Nagai et al., 2005), our results appear to suggest that the inability of the human humoral response to induce HCV eradication might be due to its low extent. In fact, the data of Fig. 2 clearly illustrate the weakness of anti-HCV $E1_{315-328}$HRMAWDMMMNW SPT when compared to the more intense response to the $tyrosinase_{233-247}$IPYWDWRDAEKCDIC peptide. Consequently, using specific immunogenic peptide fragments to elicit more powerful neutralizing antibodies or construct heterologous antibody preparations might represent a proper therapeutical solution for augmenting the potency of anti-HCV reaction.

In this context, the exact identification of the molecular amino acid sequence targeted by the immune response is important. Indeed, once that conserved viral epitopes are identified in vitro, they might be used to selectively produce neutralizing antibodies to be used in vivo for therapeutic purposes. In this context, it is pertinent to recall that previous approaches (such as peptide scanning through phage display technology) to identify viral HCV epitopes have been unsuccessful (Allander et al., 2000), possibly because of the difficulty in separating the antiviral response against conformational and glycosylated regions (Hadlock et al., 2000) from that against linear peptide sequences.

As a final consideration, the existence of two closely contiguous, still distinct and well separated sites of immunoreactivity on a single low-similarity stretch (Fig. 4) suggests a mechanism of epitope spreading that could be analysed and used in the serologic monitoring of viral infection progression. In conclusion, the present study offers preliminary data that may contribute to a better understanding of the mechanisms leading to HCV persis-

tence as well to a molecular definition of peptide-based therapeutical protocols in HCV infection.

Acknowledgement

Support by the Canadian Fund for Innovation grant (AK); MIUR 60% (DK) is gratefully acknowledged.

References

Allander T, Drakenberg K, Beyene A, Rosa D, Abrignani S, Houghton M, Widell A, Grillner L, Persson MA (2000) Recombinant human monoclonal antibodies against different conformational epitopes of the E2 envelope glycoprotein of hepatitis C virus that inhibit its interaction with CD81. J Gen Virol 81: 2451–2459

Angelini G, Bonamonte D, Lucchese A, Favia G, Serpico R, Mittelman A, Simone S, Sinha AA, Kanduc D (2006) Preliminary data on *Pemphigus vulgaris* treatment by a proteomics-defined peptide: a case report. J Transl Med 4: 43

Bortolotti F, Resti M, Marcellini M, Giacchino R, Verucchi G, Nebbia G, Zancan L, Marazzi MG, Barbera C, Maccabruni A, Zuin G, Maggiore G, Balli F, Vajro P, Lepore L, Molesini M, Guido M, Bartolacci S, Noventa F (2005) Hepatitis C virus (HCV) genotypes in 373 Italian children with HCV infection: changing distribution and correlation with clinical features and outcome. Gut 54: 852–857

Bottger V, Lane EB (1994) A monoclonal antibody epitope on keratin 8 identified using a phage peptide library. J Mol Biol 235: 61–67

Forns X, Bukh J (1999) The molecular biology of hepatitis C virus. Genotypes and quasispecies. Clin Liver Dis 3: 693–716

Gaboriaud C, Bissery V, Benchetrit T, Mornon JP (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett 224: 149–155

Hadlock KG, Lanford RE, Perkins S, Rowe J, Yang Q, Levy S, Pileri P, Abrignani S, Foung SK (2000) Human monoclonal antibodies that inhibit binding of hepatitis C virus E2 protein to CD81 and recognize conserved conformational epitopes. J Virol 74: 10407–10416

Lucchese A, Willers J, Mittelman A, Kanduc D, Dummer R (2005) Proteomic scan for tyrosinase peptide antigenic pattern in vitiligo and melanoma. Role of sequence similarity and HLA-DR1 affinity. J Immunol 175: 7009–7020

Lucchese G, Stufano A, Trost B, Kusalik A, Kanduc D (2007) Peptidology: short amino acid modules in cell biology and immunology. Amino Acids (in press)

Nagai H, Matsumaru K, Shiozawa K, Momiyama K, Wakui N, Shinohara M, Watanabe M, Ishii K, Nonaka H, Hasegawa A, Teramoto T, Yamamuro W, Sumino Y, Miki K (2005) Disappearance of HCV after cessation of immunosuppression in a patient with ulcerative colitis and renal transplantation. J Gastroenterol 40: 848–853

Natale C, Giannini T, Lucchese A, Kanduc D (2000) Computer-assisted analysis of molecular mimicry between HPV16 E7 oncoprotein and human protein sequences. Immunol Cell Biol 78: 580–585

Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for multiple sequence alignments. J Mol Biol 302: 205–217

Olenina LV, Nikolaeva LI, Sobolev BN, Blokhina NP, Archakov AI, Kolesanova EF (2002) Mapping and characterization of B cell linear epitopes in the conservative regions of hepatitis C virus envelope glycoproteins. J Viral Hepat 9: 174–182

Tanaka Y, Kurbanov F, Mano S, Orito E, Vargas V, Esteban JI, Yuen MF, Lai CL, Kramvis A, Kew MC, Smuts HE, Netesov SV, Alter HJ, Mizokami M (2006) Molecular tracing of the global hepatitis C virus epidemic predicts regional patterns of hepatocellular carcinoma mortality. Gastroenterology 130: 703–714

Wu CH, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu ZZ, Ledley RS, Lewis KC, Mewes HW, Orcutt BC, Suzek BE, Tsugita A, Vinayaka CR, Yeh LS, Zhang J, Barker WC (2002) The protein information resource: an integrated public resource of functional annotation of proteins. Nucleic Acids Res 30: 35–37

Authors' address: D. Kanduc, Department of Biochemistry and Molecular Biology, University of Bari, Via Orabona 4, Bari 70125, Italy, Fax: +39 080 544 3317, E-mail: d.kanduc@biologia.uniba.it