Clustering of rare peptide segments in the HCV immunome

Angela Stufano,¹ Giovanni Capone,¹ Barbara Pesetti,² Lorenzo Polimeno² and Darja Kanduc^{1,*}

¹Department of Biochemistry and Molecular Biology; and ²Department of Emergency and Organ Transplantation; Section of Gastroenterology; University of Bari; Bari, Italy

Key words: HCV, B-cell epitopes, low-similarity peptides, nonself HCV peptides, human proteome, anti-HCV vaccine

Our previous research and a comprehensive meta-analysis of data from the literature on epitope mapping has revealed that the B cell epitope repertoire is allocated to rare peptide motifs, i.e., antigenic peptide sequences endowed with a low level of similarity to the host proteome. From a clinical point of view, low-similarity peptides able to evoke an immune response appear to be of special interest for the rational design of vaccines for poorly treatable diseases such as hepatitis-C virus (HCV) infection. Indeed, low similarity peptides would guarantee the highest specificity and lowest cross-reactivity, i.e., effectiveness without adverse side-effects. In this study, aimed at gaining further information for the development of effective anti-HCV peptide-based vaccines, the HCV epitopes recognized by human antibodies and currently catalogued in the Immune Epitope Data Base (IEDB) were examined for pentamer sequence similarities to the human proteome. We report that the analyzed HCV determinants are characterized by the presence of fragment absent from (or scarcely represented in) human proteins. These data confirm the low-similarity hypothesis, according to which a low-similarity to the host proteome defines the nonself character of microbial antigens and modulates peptide immunogenicity. Moreover, this study indicates a concrete and safe immunotherapeutic approach which might be used in a universal anti-HCV vaccine.

Introduction

HCV is a leading cause of chronic liver disease, end-stage cirrhosis and liver cancer.¹ Despite relevant progress in the understanding of the pathogenesis of HCV infection,^{2,3} the development and implementation of strategies for HCV detection and therapy are lagging behind. Prevention of HCV infection in chimpanzees by hyper-immune serum has been reported.⁴ This datum indicates that a neutralizing humoral immune response occurs during HCV infection and suggests that amplification of the humoral end products, namely anti-HCV immunoglobulins, might allow effective neutralization of the virus. However, notwithstanding the number of immune-based therapies that are under development,⁵ specific and effective vaccines for the treatment of chronic HCV infection are still lacking.

Our labs are working on the hypothesis that rare peptide motifs, i.e., sequences rarely found in the host proteins, have the potential to elicit an immune response. We have already validated the relationship between sequence rarity and peptide immunoreactivity in a number of experimental models: breast cancer and the associated EC Her-2/Neu oncoprotein;⁶ *Pemphigus vulgaris* and the desmoglein-3 autoantigen;⁷ melanoma and melan-a/MART-1,⁸ high-molecular-weight melanoma-associated-antigen,^{9,10} and tyrosinase;^{11,12} prostate cancer and PSA;¹³ HPV16 E7,¹⁴ HIV-1,¹⁵ and Influenza A¹⁶ proteins. In addition, browsing the scientific literature on epitope mapping has revealed that most peptide

epitopes adhere to the low-similarity rule;¹⁷⁻²⁰ thus, widely supporting our hypothesis.

Here, we extend our analysis to the HCV immunome in order to define immunoreactive HCV peptide sequences not shared with the human proteome,^{21,22} which may be used for effective and specific anti-HCV immunotherapeutic approaches. Indeed, the definition of a peptidome platform unique to HCV and able to evoke a humoral immune response might allow effective anti-HCV vaccines without concomitant harmful cross-reaction with host molecules. We have analyzed HCV-derived B-cell epitope sequences already experimentally validated as epitopes and currently catalogued in the Immune Epitope Data Base (IEDB; www.immuneepitope.org).²³ We report here that almost all of the investigated HCV peptide sequences are fragments absent from (or scarcely represented in) human proteins, thereby confirming our previous reports on sequence uniqueness as a molecular signature of HCV peptide immunogenicity²¹ and opening the way to new anti-HCV strategies.

Results

HCV-derived B-cell linear epitopes were analysed for similarities to the human proteome with the aim of identifying shared amino acid groupings.^{19,24} According to the literature, peptides consisting of 5 to 6 amino acids represent sufficient minimal antigenic determinants^{25,26} and, therefore, a putative determinant

*Correspondence to: Darja Kanduc; Email: d.kanduc@biologia.uniba.it; dkanduc@gmail.com

Submitted: 01/06/10; Revised: 02/01/10; Accepted: 02/03/10

Previously published online: www.landesbioscience.com/journals/selfnonself/article/11391

is reasonably represented by a sequence of five amino acids. Therefore, we analyzed epitope similarity versus the human proteome by dividing the epitopes into pentamers sequentially overlapping by four residues, and determined the number of times each viral pentamer occurs in the human proteome. The similarity level of a pentapeptide is zero when the pentapeptide is absent from the proteome under analysis, and incrementally higher as the pentapeptide is repeatedly represented in the proteome investigated. Previous experimental data have shown that the similarity level of a pentapeptide can range from no matches to hundreds of matches. A pentapeptide fragment that has five or fewer perfect matches to the host proteome is considered a low-similarity sequence.⁶⁻¹⁶

Table 1 illustrates the data obtained following the similarity analysis of the human proteome versus the HCV peptides previously validated as epitopes by a number of different immunoassays in a number of different research laboratories worldwide. The collection of peptide fragments described in **Table 1** forms a specific HCV immuno-peptidome. Although the epitopic peptides belong to different HCV proteins and encompass different peptide lengths (from 7 to 30 amino acids, such as the MMNWSPT and MSTNPKPQKKNKRNTRRPQDVKFPGGGQI epitopes) (**Table 1**), nevertheless, the viral peptides form a peptide set characterized by the common property of being targeted by the immune humoral response. As an additional commonality, the sequence similarity analysis shows that the viral epitopic sequences have low-similarity pentameric fragments as a recurrent feature. All of the analysed HCV epitopes have pentapeptides scarcely represented in (or absent from) the human proteome. Sequences containing pentapeptide fragment(s) with low similarity to the human proteome constantly appear to characterize the determinant region of the viral antigens. A paradigmatic example is given by the first eight epitopes listed in **Table 1**. The sequence alignment of the eight epitopes clearly indicates that the low-similarity NTNRR pentapeptide is a minimal determinant region recognized by a number of different human antibody preparations.²⁷⁻³³ More generally, the analysis in **Table 1** reveals that low-similarity segments characterize the antigen portions involved in the immune recognition, independent of the epitope length and viral antigen derivation.^{21,27-47}

A graphic visualization of this concept is presented in **Figure 1**, which illustrates the location of linear epitopes along the envelope glycoprotein E2 from HCV, genotype 1, (UniProt accession number: P26664, aa 384–746), as a function of the antigen similarity to the human proteome. In brief, using the computational procedure described above, the entire primary sequence of HCV E2 was divided into 5-mer motifs that were probed against the human proteome utilizing the exact peptide matching program at the PIR website (<http://pir.georgetown.edu/pirwww>).⁴⁹ The HCV E2 epitopic sequences listed in **Table 1** were then allocated along the HCV E2-versus-human similarity profile.

Figure 1 reveals that the 5-mer sharing between the HCV E2 protein and the human proteome has a wave-like behaviour, with viral peptide motifs scarcely represented in the human proteome alternating with areas of intensively repeated peptides. There are 49 pentamers unique to HCV E2; they account for the 13.6%

of the total HCV E2 pentamers. The remaining 310 pentapeptides are widespread throughout the human proteome, occurring a total of 4113 times. That is, the 4113 occurrences include multiple HCV E2 matches to human proteins. Moreover, and importantly, **Figure 1** documents that the HCV E2 epitopes recognized by human sera are clustered within viral regions characterized by a low level of similarity to the human proteins and representing unique signatures of the viral E2 glycoprotein. That is, the epitope clustering practically assigns an immunological information to the rare peptide segments of the viral E2 primary sequence.

Discussion

HCV causes chronic infections that can be associated with development of liver diseases ranging from mild liver inflammation to hepatocellular carcinomas. Currently, the therapeutic agents used to treat HCV infections are limited in effectiveness, poorly tolerated, toxic, and, in general, can lead to development of resistance. Therefore, immunotherapeutic approaches are being actively pursued in order to complement or replace the current anti-HCV treatments.

Focusing on proteomic approaches and applications to anti-HCV peptide-based vaccines, we have examined the HCV amino acid sequences already experimentally validated as epitopes. We found that fragment(s) not present (or scarcely represented) in human proteins are the common feature of the HCV epitopes targeted by the human humoral immune response.

The data appear to be of special importance for the following reasons. First, they offer a principle for designing effective anti-HCV vaccines. In fact, low-similarity peptides present in the viral antigens may bypass immune escape mechanisms because of their being practically unknown to the immune system,⁵⁰ thus breaking the immunotolerance to HCV antigens which presumably is a major obstacle to the physiological immune recognition and destruction of the virus.

Moreover, vaccine preparations based on no-similarity segments from HCV antigens are expected to specifically target the HCV antigen, without harmful cross-reactions with other proteins. HCV is composed of 11 proteins and 3,011 amino acids, for a total of 3,007 pentapeptides sequentially overlapping by four residues. The total viral overlap with the human proteome (including multiple occurrences) amounts to 45,545 occurrences with 19,605 human proteins involved in the overlap.⁵⁰ Consequently, a full virus vaccine that contains all peptide determinants would produce a high risk of possible cross-reactivity with human proteins and a sequela of inflammatory autoimmune reactions, thereby aggravating the HCV-induced inflammation status. Additionally, vaccines based on single antigens, i.e., an E2-based vaccine, would imply a high risk of cross-reactivity as illustrated by the HCV E2-versus-human proteome overlap illustrated in **Figure 1**. Therefore, from a clinical point of view, low-similarity peptide approaches might pave the way to the concrete possibility of safe anti-HCV immunotherapies.

Finally, the data reported here are promising for vaccine formulations that are highly flexible, since peptide consensus sequences can be modulated for activity against a broad range

Table 1. HCV-derived B-cell epitopes: sequence similarity analysis to the human proteome at the pentapeptide level

HCV ID ¹	Antigen ²	Epitope ID ³	Epitopic Sequence ³⁻⁵	Matches ⁶	Refs.
31646	p21	42688	MSTNP kqkknkr NTNRR p QDVKF pgggqi	0, 1, 2	27
31646	p21	42687	MSTNP kqkknkr NTNRR p QDVKF pggg	0, 1, 2	28
11104	p21	42684	MSTNP kqkknkr NTNRR	0, 1	29
11108	p21	49116	pqrktkr NTNRR	1	30
11104	p21	51204	qkknkr NTNRR pqdv	1	31
31646	p21	33747	ktkr NTNRR p QDVKF	1, 2	32
11104	p21	44588	nkr NTNRR p QDVKF pggg	1, 2	29
11108	p21	55071	r NTNRR p QDVKF	1, 2	33
11108	p21	65414	tnrrp QDVKF pgggq	2	33
11108	p21	55680	rrp QDVKF pg	2	34
11108	p21	48984	p QDVKF pgggqi	2	33
11103	p21	50525	QDVKF pggg QIVGGV	2, 4	32
31646	p21	50525	QDVKF pggg QIVGGV	2, 4	35
11108	p21	10636	DVKFP gggqi	4	34
11104	p21	10638	DVKFP ggg QIVGG vylp	4, 4	29
11108	p21	47662	pggg QIVGG vYLLPRr	4, 5	33
11108	p21	51148	QIVGG vylp	4	34
31646	p21	51150	QIVGG vYLLPRrgpr	4, 5	32
11108	p21	23223	gvYLLPRrgp	5	34
11103	p21	55581	rrgprlgvra TRKTS	5	32
31646	p21	55581	rrgprlgvra TRKTS	5	35
11108	p21	54570	rlgvra TRKTSe	5	33
11108	p21	66044	TRKTS ersqp	5	34
31646	p21	66046	TRKTS ersqprgr	5	32
11104	p21	33858	ktersqprg RRQPI pka	4	29
11108	p21	55930	rsqprg RRQPI P	4	33
31646	p21	49193	prg RRQPI pkarrpe	4	32
11108	p21	55698	RRQPI pkarrpe	4	33
11104	p21	79575	ipkarrpeg RTWAQ pgy	1	36
11108	p21	29936	karrpeg RTWAQ	1	33
31646	p21	29936	karrpeg RTWAQ	1	37
31646	p21	29937	karrpeg RTWAQ pgy	1	32
11103	p21	47289	peg RTWAQ	1	38
11104	p21	47289	peg RTWAQ	1	36
11108	p21	22223	g RTWAQ pGYPWP	1, 0	33
11108	p21	4023	aqpGYPWPly	0	34
11108	p21	4024	aqpGYPWPlygneg	0	32
31646	p21	47792	pGYPWPlygneg	0	33
11108	p21	75431	ypWPLYGneg	1	34
11108	p21	40772	lyg NEGCG wa	1	34
31646	p21	74061	ygnegcg WAGWL isp	0	32
11108	p21	6365	cg WAGWL isprg	0	33

¹HCV IEDB ID: 31646: HCV1a; 11104: HCV1; 11108: HCV1a, isolate H; 11103: HCV; 63746: HCV1a, isolate H77; 10000457: HCV1a, Chiron Corp. ²Viral antigen: p21, Core protein p21; p19, Core protein p19; E1, Envelope glycoprotein E1; E2, Envelope glycoprotein E2; NS2-3, Protease NS2-3; NS3, Serine protease/ NTPase/helicase NS3; NS4A, Non-structural protein 4A; NS4B, Non-structural protein 4B; NS5A, Non-structural protein 5A; RdRP, RNA-directed RNA polymerase. ³The epitopic peptides were immunoassayed by ELISA using sera from HCV-infected patients. Epitope ID, sequence and experimental assay description from IEDB (www.immuneepitope.org) (23). ⁴Low similarity pentapeptides are given in capital. ⁵The sequence alignment indicates the epitopic commonality of low-similarity fragment(s). ⁶Similarity level is measured as number of exact matches to the host proteome, i.e., as the number of times the low-similarity pentapeptide occurs in the entire set of human proteins [see Methods and refs. 6–16].

Table 1. HCV-derived B-cell epitopes: sequence similarity analysis to the human proteome at the pentapeptide level

11108	p21	12134	egcg WAGWLI	0	34
31646	p21	72755	wllsprgs RPSWG pt	4	32
11108	p21	37814	llsprgs RPSWG	4	33
11104	p21	53988	rgsrps WGPTD prrrsrn	2	29
11108	p21	60770	srps WGPTD prrr	2	33
31646	p21	62453	s WGPTD prrrsrnlq	2	32
11108	p21	39825	ltcgfad LMGYI p	2	33
11108	p21	19491	gfad LMGYI p	2	34
11103	p21	38031	LMGYI plvga	2	39
11108	p21	38031	LMGYI plvga	2	34
11108	p21	38496	IPGCSF siflla	3	33
11104	E1	71427	VTNDC pnss IVYEA ada	1, 1	29
31646	E1	61123	ss IVYEA a DAILH t	0, 1	37
31646	E1	73614	yeaa DAILH tp	1	37
31646	E1	82	aa DAILH tpgcvpc	1	37
11108	E1	5029	atlc SALYV g	4	34
11108	E1	56888	salyv GDLCG	3	34
11108	E1	68638	v GDLCG svfl	3	34
63746	E1	65846	t QGCNC siyp	0	34
31646	E1	58765	siypghitg HRMAW	0	37
31646	E1	47673	pghitg HRMAW dmmm	0	37
11108	E1	28909	itg HRMAW dm	0	34
11104	E1	28912	itg HRMAW d MMMNW sptt	0	29
31646	E1	63848	tg HRMAW d MMMNW	0, 0	37
11103	E1	24694	HRMAW dmmm	0	40
11104	E1	24694	HRMAW dmmm	0	21
11104	E1	54860	RMAW dmm	0	22
11108	E1	54863	RMAW d MMMNW	0, 0	34
11104	E1	42159	MMNWS pt	0	22
11108	E1	75567	YQVRN ssglyhv	1	33
11108	E1	47054	pc VREGN asrcw	0	33
11108	E1	1439	aga HWGVL ag	1	34
10000457	E2	22271	gsaghtvsgfvslap GAKQN	2	41
31646	E2	33043	kqniqlint NGSWH	1	37
11104	E2	46459	nvqlint NGSWH lnstl	1	29
31646	E2	51575	qlvnt NGSWH in	1	42
11108	E2	59375	s LNTGW laglfyqhkfns	0	33
11104	E2	36025	LFYHH kfns	0	21
11104	E2	55382	RPYCW HYPP	0	21
11104	E2	50070	PYCW Hyppk PCGIV pa	0, 0	30
31646	E2	50070	PYCW Hyppk PCGIV pa	0, 0	28
11108	E2	57978	sga PTYSW Gand	0, 0	33
31646	E2	67404	tynwgand TDVFI	5	37

¹HCV IEDB ID: 31646: HCV1a; 11104: HCV1; 11108: HCV1a, isolate H; 11103: HCV; 63746: HCV1a, isolate H77; 10000457: HCV1a, Chiron Corp. ²Viral antigen: p21, Core protein p21; p19, Core protein p19; E1, Envelope glycoprotein E1; E2, Envelope glycoprotein E2; NS2-3, Protease NS2-3; NS3, Serine protease/ NTPase/helicase NS3; NS4A, Non-structural protein 4A; NS4B, Non-structural protein 4B; NS5A, Non-structural protein 5A; RdRP, RNA-directed RNA polymerase. ³The epitopic peptides were immunoassayed by ELISA using sera from HCV-infected patients. Epitope ID, sequence and experimental assay description from IEDB (www.immuneepitope.org) (23). ⁴Low similarity pentapeptides are given in capital. ⁵The sequence alignment indicates the epitopic commonality of low-similarity fragment(s). ⁶Similarity level is measured as number of exact matches to the host proteome, i.e., as the number of times the low-similarity pentapeptide occurs in the entire set of human proteins [see Methods and refs. 6–16].

Table 1. HCV-derived B-cell epitopes: sequence similarity analysis to the human proteome at the pentapeptide level

11108	E2	62447	swgand TDVFI	5	33
11104	E2	60120	spIII TTQWQVPCSFT	0, 0, 0	29
11104	E2	52739	qvl PCSFT tlpalstgli	0	29
63746	E2	2899	als TGLIH	5	43
63746	E2	39697	ls TGLIH	5	43
11108	E2	2902	alstgli HLHQ N	2	33
63746	E2	61621	stg LIHLH	3	43
63746	E2	63882	tg LIHLH q	3	43
11108	E2	62065	svcg PVYCFT	0, 0	34
11108	E2	68808	VGTTDR sgapty	3, 3	33
11104	E2	46532	NWFGCTW mn	0	21
31646	E2	72869	MNSTgFTKVC gappc	2, 2	28
31646	p7	58999	slag THGLV sflvf	5	37
11104	p7	72171	VYTFY gmwp	0	21
31646	NS2-3	35304	ldtevaas CGGVV I	5	37
31646	NS2-3	33259	kryis WCLWW lqyf	0, 0	37
31646	NS2-3	56259	rvea QLHVV ippln	0, 0	37
31646	NS2-3	5380	avill MCAVH ptlv	0	37
31646	NS2-3	16031	fgpl WILQA sllkv	0	37
31646	NS2-3	70612	VQMVI ikgaltgt	1	37
31646	NS2-3	36241	lgpa DGMVS kgwrl	1	37
31646	NS3	4080	aqtflatci NGVCW	1	37
31646	NS3	22806	g TRTIAS pkgpviq	4, 4	37
31646	NS3	65474	tnvdq DLVGW papq	0	37
31646	NS3	68302	ven ETTMRS pvft	0	37
31646	NS3	25795	ieevalstt GEIPF	3	37
31646	NS3	12927	elaaklv ALGIN av	5	37
31646	NS3	33066	kqsgenlpy LVAYQ a	1	37
11104	NS3	72326	WDQM Wkcli	0	21
11104	NS3	37987	llyrlgavqneitlthpv TKYIM tc	0	44
31646	NS3-NS4A	41012	MACMS adlevvtst	0	37
11104	NS4A	22244	grvvlsqgpaii PDREVI	1	29
31646	NS4A	20654	g KPAII pdre	2	37
31646	NS4A	20653	g KPAII pdr	2	45
31646	NS4A	32648	KPAII pdre	2	45
31646	NS4A	46825	paii PDREV	1	45
31646	NS4A	1951	aii PDREVI	1	45
31646	NS4A	26619	ii PDREVI ly	1	45
11104	NS4A	26622	ii PDREVI lyrefdemec	1	29
31646	NS4A	27812	i PDREVI lyr	1	45
31646	NS4A	27813	i PDREVI lyrefdem	1	37
31646	NS4A	47189	PDREVI lyre	1	45
31646	NS4A	9966	drev LYREF	1	45

¹HCV IEDB ID: 31646: HCV1a; 11104: HCV1; 11108: HCV1a, isolate H; 11103: HCV; 63746: HCV1a, isolate H77; 10000457: HCV1a, Chiron Corp. ²Viral antigen: p21, Core protein p21; p19, Core protein p19; E1, Envelope glycoprotein E1; E2, Envelope glycoprotein E2; NS2-3, Protease NS2-3; NS3, Serine protease/ NTPase/helicase NS3; NS4A, Non-structural protein 4A; NS4B, Non-structural protein 4B; NS5A, Non-structural protein 5A; RdRP, RNA-directed RNA polymerase. ³The epitopic peptides were immunoassayed by ELISA using sera from HCV-infected patients. Epitope ID, sequence and experimental assay description from IEDB (www.immuneepitope.org) (23). ⁴Low similarity pentapeptides are given in capital. ⁵The sequence alignment indicates the epitopic commonality of low-similarity fragment(s). ⁶Similarity level is measured as number of exact matches to the host proteome, i.e., as the number of times the low-similarity pentapeptide occurs in the entire set of human proteins [see Methods and refs. 6–16].

Table 1. HCV-derived B-cell epitopes: sequence similarity analysis to the human proteome at the pentapeptide level

31646	NS4A	53666	revLYREFd	1	45
31646	NS4A	14822	evLYREFde	1	45
31646	NS4A	69900	vLYREFdem	1	45
11104	NS4A/4B	11973	EFDEM ^e ECSQHlpyieqg	1, 1	29
31646	NS4A/4B	11265	ECSQHlpyi	1	45
31646	NS4A/4B	7061	csqhLPYIE	1	45
31646	NS4B	60449	sqhLPYIEq	1	45
31646	NS4B	51014	qhLPYIEgg	1	45
31646	NS4B	24283	hLPYIEqgm	1	45
11104	NS4B	24284	hlp ^y iEQGMMl ^a eqfk	0	29
31646	NS4B	38777	lp ^y iEQGMM	0	45
31646	NS4B	50102	pyiEQGMMI	0	45
31646	NS4B	74201	YIEQGMql ^a eqfkq	2	37
11104	NS4B	34747	IAEQFKQKalgllqtasr	3	29
31646	NS4B	16437	FKQKAlgllqtasr	4	37
11104	NS4B	21019	gllqtasrqaEVIAPavq	1	29
31646	NS4B	60774	srqaEVIAP	1	45
31646	NS4B	55390	rqaEVIAPa	1	45
31646	NS4B	14721	EVIAPaVQTNWqkl	1, 0	37
11108	NS4B	36450	liafasr ^g nhvsPTHYV	1	46
11108	NS4B	44175	nhvsPTHYVpesdaaar	1	46
31646	NS4B	60285	sPTHYVpesdaaar	1	37
11104	NS4B	72236	WAKHMwnfi	0	21
31646	NS4B	27557	IMSGEv ^p stedl ^v n	3	37
31646	NS4B	55646	rrl ^H QWISsecttp	0	37
31646	NS5A	47027	PCKPLreevsfrv	5	37
31646	NS5A	53494	reevsfrvgLHEY ^P	1	37
11108	NS5A	46856	pal ^P VWAR	1	46
11108	NS5A	2830	alp ^V WAR ^P	1	46
11108	NS5A	38764	lpv ^W AR ^P D	0	46
11108	NS5A	50016	pv ^W AR ^P Dy	0	46
11108	NS5A	71921	v ^W AR ^P Dyn	0	46
11108	NS5A	72259	WAR ^P Dynp	0	46
11108	NS5A	15438	FDPLV ^a ee	3	46
31646	NS5A	15439	FDPLV ^a eedereis	3	37
31646	NS5A	11543	ee ^D EREIsv ^p aeil	4	37
11108	NS5A	45392	np ^L VET ^W	2	46
31646	NS5A	48521	p ^L VET ^W kkpdy ^e pp	2	37
11108	NS5A	40323	L ^V ET ^W kkp	2	46
11108	NS5A	68391	vet ^W KK ^P D	2	46
11108	NS5A	14599	et ^W KK ^P Dy	2	46
11108	NS5A	67270	t ^W KK ^P Dy ^e	2	46
11108	NS5A	72684	WKK ^P Dy ^e p	2	46

¹HCV IEDB ID: 31646: HCV1a; 11104: HCV1; 11108: HCV1a, isolate H; 11103: HCV; 63746: HCV1a, isolate H77; 10000457: HCV1a, Chiron Corp. ²Viral antigen: p21, Core protein p21; p19, Core protein p19; E1, Envelope glycoprotein E1; E2, Envelope glycoprotein E2; NS2-3, Protease NS2-3; NS3, Serine protease/ NTPase/helicase NS3; NS4A, Non-structural protein 4A; NS4B, Non-structural protein 4B; NS5A, Non-structural protein 5A; RdRP, RNA-directed RNA polymerase. ³The epitopic peptides were immunoassayed by ELISA using sera from HCV-infected patients. Epitope ID, sequence and experimental assay description from IEDB (www.immuneepitope.org) (23). ⁴Low similarity pentapeptides are given in capital. ⁵The sequence alignment indicates the epitopic commonality of low-similarity fragment(s). ⁶Similarity level is measured as number of exact matches to the host proteome, i.e., as the number of times the low-similarity pentapeptide occurs in the entire set of human proteins [see Methods and refs. 6–16].

Table 1. HCV-derived B-cell epitopes: sequence similarity analysis to the human proteome at the pentapeptide level

11108	NS5A	31657	KKPDYEP p	5	46
11108	NS5A	32683	KPDYEP pv	5	46
11108	NS5A	47230	PDYEP pvv	5	46
31646	NS5A	32684	kpdyep VVHGC plppp	0	37
11108	NS5A	73741	yep PVVHG	3	46
11108	NS5A	13728	epp VVHGC	0	46
31646	NS5A	71686	VVHGC plpppkpp	0	37
31646	NS5A	46900	papsg CPPDS daes	5	37
31646	NS5A	47200	pdsdaes YSSMP pl	1	37
31646	NS5A	10812	d WICEV lsdfktwl	0	37
31646	NS5A	29839	k AKLMPQ lpgipfv	2, 2	37
31646	NS5A	8408	d GIMHT rc HCGAE i	0, 0	37
11104	NS5A	63096	tcn MWSGT	0	21
31646	NS5A	12571	eirqvg DFHYV tgm	1	37
31646	NS5A	62933	ta NHDSP daeliea	1	37
11108	NS5A	55135	RPDYN pll	2	46
31646	NS5A	70426	vppprk KRTVV lte	3	37
31646	NS5A	2380	alae IATRSF gsss	5	37
31646	NS5A	66219	tsgitg DNTTT sse	5	37
31646	NS5A	61996	stvseana EDVVC	4	37
31646	RdRP	53276	r CHARK avahinsv	1	37
31646	RdRP	6375	CHARK avathins	1	47
31646	RdRP	21953	gqrvefl VQAWK sk	1	37
31646	RdRP	16963	fl VQAWK skktp	1	47
31646	RdRP	47865	piaaagql LSGWF	1	37
31646	RdRP	58308	SGWFT agysggdiy	0	37
31646	RdRP	65515	tpcaae QKLPI na	2	37
31646	RdRP	52735	qv LDSHY qdvakev	1	37
31646	RdRP	33366	k SKFGY sakdvrl	0	37
31646	RdRP	15356	fc VQPEK ggrkpar	3	37
31646	RdRP	10787	dvvtkl PLAVM g	4	47
11104	RdRP	73595	y DTRCF dst	0	21
31646	RdRP	40980	l YVGGPI tnsrgen	5	37
31646	RdRP	6061	CCDLD pqarvaiks	3	37
31646	RdRP	58293	sgvltts CGNTL tc	1	37
31646	RdRP	74272	YIKAR aacraag	4	47
31646	RdRP	25497	i CESAG vqedaasl	1	37
31646	RdRP	67625	vahdgag KRVYY lt	2	37
31646	RdRP	54633	rllarg GRAAI cgk	2	37
31646	RdRP	62326	svsharp RWFWF cl	0	37

¹HCV IEDB ID: 31646: HCV1a; 11104: HCV1; 11108: HCV1a, isolate H; 11103: HCV; 63746: HCV1a, isolate H77; 10000457: HCV1a, Chiron Corp. ²Viral antigen: p21, Core protein p21; p19, Core protein p19; E1, Envelope glycoprotein E1; E2, Envelope glycoprotein E2; NS2-3, Protease NS2-3; NS3, Serine protease/ NTPase/helicase NS3; NS4A, Non-structural protein 4A; NS4B, Non-structural protein 4B; NS5A, Non-structural protein 5A; RdRP, RNA-directed RNA polymerase. ³The epitopic peptides were immunoassayed by ELISA using sera from HCV-infected patients. Epitope ID, sequence and experimental assay description from IEDB (www.immuneepitope.org) (23). ⁴Low similarity pentapeptides are given in capital. ⁵The sequence alignment indicates the epitopic commonality of low-similarity fragment(s). ⁶Similarity level is measured as number of exact matches to the host proteome, i.e., as the number of times the low-similarity pentapeptide occurs in the entire set of human proteins [see Methods and refs. 6–16].

Figure 1. Similarity profile of HCV E2 protein versus the human proteome at the pentapeptide level, and location of the viral epitopes. The columns indicate the number of HCV E2 pentapeptide occurrences in the human proteome. The arrows indicate the location of immune epitopic sequences recognized by human sera. Arrow numbering refers to the epitope IEDB ID: (1) 22271; (2) 33043; (3) 46459; (4) 51575; (5) 59375; (6) 36025; (7) 55382; (8) 50070; (9) 50070; (10) 62065; (11) 68808; (12) 57978; (13) 67404; (14) 62447; (15) 46532; (16) 72869; (17) ref. 48; (18) ref. 43; (19) 60120; (20) 52739; (21) 2899; (22) 39697; (23) 2902; (24) 61621; (25) 63882. Amino acid sequence details are reported in Table 1. HCV E2 protein UniProt accession number: P26664; amino acid position: 384–746.

Do not distribute.

of HCV species, and quickly modified for fighting newly emerging subtypes. Indeed, HCV can mutate rapidly, thus contributing to the high sequence divergence of multiple viral isolates in the world. HCV exhibits complex genetic variability that can be classified into four hierarchical strata: genotypes, sub-genotypes, isolates and quasispecies.⁵¹ Moreover, infected patients may suffer from infection with multiple HCV genotypes or experience changing patterns of HCV sequence variability over time.⁵² Thus, the amino acid sequence variability of various HCV types might alter peptide immunoreactivity.²¹ In this regard, HCV conserved sequences endowed with a low-similarity score to the human proteome might represent the basic platform for a universal vaccine program. The advantages would be paramount in terms of costs and global protection from HCV infection.

Methods

B-cell linear epitopes from HCV, genotype 1, and catalogued in the IEDB resource (www.immuneepitope.org)²³ were analyzed for

similarity to the human proteins. In the immunological context, sequence similarity search signifies to identify identical amino acid groupings that are common sequences.^{19,24} Procedurally, the HCV epitopes were dissected into pentamer motifs (i.e., minimal sufficient determinants) offset by one residue: i.e., MSTNP, STNPK, TNPKP, NPKPQ, etc. Then, each 5-mer was used as a probe to scan the entire human proteome searching for exact identical matches using the PIR Perfect Peptide Match program (<http://pir.georgetown.edu/pirwww/>).⁴⁹ The resulting number of matches provides the similarity level to the human proteome of each viral pentamer.

Acknowledgements

This study was supported by the Ministry of University, Italy.

Authors' contribution: A.S., G.C., and B.P. performed the computational analyses, with a major contribution of A.S. L.P. supervised the computational analysis. D.K. proposed the original idea, interpreted the data and wrote the paper. All authors revised and approved the paper.

References

- Mühlberger N, Schwarzer R, Lettmeier B, Sroczynski G, Zeuzem S, Siebert U. HCV-related burden of disease in Europe: a systematic assessment of incidence, prevalence, morbidity and mortality. *BMC Public Health* 2009; 9:34.
- Schmitz U, Tan SL. NS5A: from obscurity to new target for HCV therapy. *Recent Pat Antiinfect Drug Discov* 2008; 3:77-92.
- Lukavsky PJ. Structure and function of HCV IRES domains. *Virus Res* 2009; 139:166-71.
- Farci P, Shimoda A, Wong D, Cabezon T, De Gioannis D, Strazzer A, et al. 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* 1996; 93:15394-9.
- Chang DY, Shin EC. Immune-based therapy for chronic hepatitis C. *J Leukoc Biol* 2009; 86:33-9.
- Mittelman A, Lucchese A, Sinha AA, Kanduc D. Monoclonal and polyclonal humoral immune response to EC HER-2/NEU peptides with low similarity to the host's proteome. *Int J Cancer* 2002; 98:741-7.
- Lucchese A, Mittelman A, Lin MS, Kanduc D, Sinha AA. Epitope definition by proteomic similarity analysis: identification of the linear determinant of the anti-Dsg3 MAb 5H10. *J Transl Med* 2004; 2:43.

8. Tiwari R, Geliebter J, Lucchese A, Mittelman A, Kanduc D. Computational peptide dissection of Melan-a/MART-1 oncoprotein antigenicity. *Peptides* 2004; 25:1865-71.
9. Mittelman A, Tiwari R, Lucchese G, Willers J, Dummer R, Kanduc D. Identification of monoclonal anti-HMW-MAA antibody linear peptide epitope by proteomic database mining. *J Invest Dermatol* 2004; 123:670-5.
10. Dummer R, Mittelman A, Fanizzi FP, Lucchese G, Willers J, Kanduc D. Non-self-discrimination as a driving concept in the identification of an immunodominant HMW-MAA epitopic peptide sequence by autoantibodies from melanoma cancer patients. *Int J Cancer* 2004; 111:720-6.
11. Lucchese A, Willers J, Mittelman A, Kanduc D, Dummer R. Proteomic scan for tyrosinase peptide antigenic pattern in vitiligo and melanoma: role of sequence similarity and HLA-DR1 affinity. *J Immunol* 2005; 175:7009-20.
12. Willers J, Lucchese A, Mittelman A, Dummer R, Kanduc D. Definition of anti-tyrosinase MAb T311 linear determinant by proteome-based similarity analysis. *Exp Dermatol* 2005; 14:543-50.
13. Stufano A, Kanduc D. Proteome-based epitopic peptide scanning along PSA. *Exp Mol Pathol* 2009; 86:36-40.
14. Kanduc D, Lucchese A, Mittelman A. Individuation of monoclonal anti-HPV16 E7 antibody linear peptide epitope by computational biology. *Peptides* 2001; 22:1981-5.
15. Lucchese A, Serpico R, Crincoli V, Shoenfeld Y, Kanduc D. Sequence uniqueness as a molecular signature of HIV-1-derived B-cell epitopes. *Int J Immunopathol Pharmacol* 2009; 22:639-46.
16. Lucchese G, Stufano A, Kanduc D. Proteome-guided search for influenza A B-cell epitopes. *FEMS Immunol Med Microbiol* 2009; 57:88-92.
17. Kanduc D, Lucchese A, Mittelman A. Non-redundant peptidomes from DAPs: towards "the vaccine"? *Autoimmun Rev* 2007; 6:290-4.
18. Kanduc D. Correlating low-similarity peptide sequences and allergenic epitopes. *Curr Pharm Des* 2008; 14:289-95.
19. Kanduc D. Immunogenicity in peptide-immunotherapy: from self/nonself to similar/dissimilar sequences. In *Multichain Immune Recognition Receptor Signaling: From Spatiotemporal Organization to Human Disease* A. Sigalov, Ed., Landes Bioscience, Austin TX, USA 2008; 198-207.
20. Kanduc D. Self-nonself peptides in the design of vaccines. *Curr Pharm Des* 2009; 15:3283-9.
21. Kanduc D, Tessitore L, Lucchese G, Kusalik A, Farber E, Marincola FM. Sequence uniqueness and sequence variability as modulating factors of human anti-HCV humoral immune response. *Cancer Immunol Immunother* 2008; 57:1215-23.
22. Polimeno L, Mittelman A, Gennero L, Ponzetto A, Lucchese G, Stufano A, et al. Sub-epitopic dissection of HCV E1315-328HRMAWDMMMNWSPT sequence by similarity analysis. *Amino Acids* 2008; 34:479-84.
23. Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, et al. The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol* 2005; 3:91.
24. May AC. Percent sequence identity; the need to be explicit. *Structure* 2004; 12:737-8.
25. Oldstone MB. Molecular mimicry and immune-mediated diseases. *FASEB J* 1998; 12:1255-65.
26. Lucchese G, Stufano A, Trost B, Kusalik A, Kanduc D. Peptidology: short amino acid modules in cell biology and immunology. *Amino Acids* 2007; 33:703-7.
27. Sato A, Ida N, Ishikawa M, Tanahashi K, Nakamura H, Shio Y, et al. A sensitive serodiagnosis of hepatitis C virus (HCV) infection with two non-fused peptides: comparison of antibody responses detected with a newly developed assay and a commercial second-generation test. *Microbiol Immunol* 1993; 37:295-304.
28. Zhang ZX, Sönnernborg A, Sällberg M. Antigenic structure of the hepatitis C virus envelope 2 protein. *Clin Exp Immunol* 1994; 98:382-7.
29. Sällberg M, Rudén U, Wahren B, Magnus LO. Antigenic regions within the hepatitis C virus envelope 1 and non-structural proteins: identification of an IgG3-restricted recognition site with the envelope 1 protein. *Clin Exp Immunol* 1993; 91:489-94.
30. Carlos MP, Yamamura Y, Vu Q, Conzen K, Anderson DE, Torres JV. Humoral immunity to immunodominant epitopes of hepatitis C virus in individuals infected with genotypes 1a or 1b. *Clin Immunol* 2004; 111:22-7.
31. Pirisi M, Fabris C, Toniutto P, Vitulli D, Soardo G, Falletti E, et al. Reactivity to B cell epitopes within hepatitis C virus core protein and hepatocellular carcinoma. *Cancer Res* 1995; 55:111-4.
32. León P, López JA, Elola C, Lee SR, Calmann M, Echevarría JM. Use of overlapping synthetic peptides to characterize samples from blood donors with indeterminate results to hepatitis C virus core antigen. *Vox Sang* 1998; 75:32-6.
33. Wang YF, Brotman B, Andrus L, Prince AM. Immune response to epitopes of hepatitis C virus (HCV) structural proteins in HCV-infected humans and chimpanzees. *J Infect Dis* 1996; 173:808-21.
34. Ching WM, Wychowski C, Beach MJ, Wang H, Davies CL, Carl M, et al. Interaction of immune sera with synthetic peptides corresponding to the structural protein region of hepatitis C virus. *Proc Natl Acad Sci USA* 1992; 89:3190-4.
35. Tobler LH, Busch MP, Wilber J, Dinello R, Quan S, Polito A, et al. Evaluation of indeterminate c22-3 reactivity in volunteer blood donors. *Transfusion* 1994; 34:130-4.
36. Machida A, Ohnuma H, Tsuda F, Munekata E, Tanaka T, Akahane Y, et al. Two distinct subtypes of hepatitis C virus defined by antibodies directed to the putative core protein. *Hepatology* 1992; 16:886-91.
37. Rodríguez-López M, Riezu-Boj JJ, Ruiz M, Berasain C, Civeira MP, Prieto J, et al. Immunogenicity of variable regions of hepatitis C virus proteins: selection and modification of peptide epitopes to assess hepatitis C virus genotypes by ELISA. *J Gen Virol* 1999; 80:727-38.
38. Jolivet-Reynaud C, Dalbon P, Viola F, Yvon S, Paranhos-Baccala G, Piga N, et al. HCV core immunodominant region analysis using mouse monoclonal antibodies and human sera: characterization of major epitopes useful for antigen detection. *J Med Virol* 1998; 56:300-9.
39. Akatsuka T, Donets M, Scaglione L, Ching WM, Shih JW, Di Bisceglie AM, et al. B-cell epitopes on the hepatitis C virus nucleocapsid protein determined by human monospecific antibodies. *Hepatology* 1993; 18:503-10.
40. Siemoneit K, Cardoso Mda S, Koerner K, Wölpl A, Kubanek B. Human monoclonal antibodies for the immunological characterization of a highly conserved protein domain of the hepatitis C virus glycoprotein E1. *Clin Exp Immunol* 1995; 101:278-83.
41. Shirai M, Arichi T, Chen M, Masaki T, Nishioka M, Ikeda K, et al. T cell recognition of hypervariable region-1 from hepatitis C virus envelope protein with multiple class II MHC molecules in mice and humans: preferential help for induction of antibodies to the hypervariable region. *J Immunol* 1999; 162:568-76.
42. Tarr AW, Owsianka AM, Jayaraj D, Brown RJ, Hickling TP, Irving WL, et al. Determination of the human antibody response to the epitope defined by the hepatitis C virus-neutralizing monoclonal antibody AP33. *J Gen Virol* 2007; 88:2991-3001.
43. Olenina LV, Nikolaeva LI, Sobolev BN, Blokhina NP, Archakov AI, Kolesanova EF. Mapping and characterization of B cell linear epitopes in the conservative regions of hepatitis C virus envelope glycoproteins. *J Viral Hepat* 2002; 9:174-82.
44. Wienhues U, Ihlenfeldt HG, Seidel C, Schmitt U, Kraas W, Jung G. Characterization of a linear epitope in the nonstructural region 4 of hepatitis C virus with reactivity to seroconversion antibodies. *Virology* 1998; 245:281-8.
45. Simmonds P, Rose KA, Graham S, Chan SW, McOmish F, Dow BC, et al. Mapping of serotype-specific, immunodominant epitopes in the NS4 region of hepatitis C virus (HCV): use of type-specific peptides to serologically differentiate infections with HCV types 1, 2 and 3. *J Clin Microbiol* 1993; 31:1493-503.
46. Rosa C, Osborne S, Garetto F, Griva S, Rivella A, Calabresi G, et al. Epitope mapping of the NS4 and NS5 gene products of hepatitis C virus and the use of a chimeric NS4-NS5 synthetic peptide for serodiagnosis. *J Virol Methods* 1995; 55:219-32.
47. Quiroga JA, Martin J, Pernas M, Pardo M, Herrero M, Castillo I, et al. Evidence of subtype-specific antibodies to antigenic epitopes in the NS5 region of hepatitis C virus in the circulation of patients with chronic hepatitis C. *Clin Diagn Lab Immunol* 1994; 1:545-51.
48. Petit MA, Jolivet-Reynaud C, Peronnet E, Michal Y, Trépo C. Mapping of a conformational epitope shared between E1 and E2 on the serum-derived human hepatitis C virus envelope. *J Biol Chem* 2003; 278:44385-92.
49. Wu CH, Yeh LS, Huang H, Arminski L, Castro-Alvares J, Chen Y, et al. The protein information resource. *Nucleic Acids Res* 2003; 31:345-7.
50. Kusalik A, Bickis M, Lewis C, Li Y, Lucchese G, Marincola FM, et al. Widespread and ample peptide overlapping between HCV and *Homo sapiens* proteomes. *Peptides* 2007; 28:1260-7.
51. Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* 2005; 42:962-73.
52. Pol S, Thiers V, Noursbaum JB, Legendre C, Berthelot P, Kreis H, et al. The changing relative prevalence of hepatitis C virus genotypes: evidence in endodialyzed patients and kidney recipients. *Gastroenterology* 1995; 108:581-3.