

Positively selected amino acid sites in the entire coding region of hepatitis C virus subtype 1b

Yoshiyuki Suzuki, Takashi Gojobori*

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, 1111 Yata, Mishima-shi, Shizuoka-ken 411-8540, Japan

Received 6 December 2000; received in revised form 10 April 2001; accepted 26 July 2001

Received by G. Bernardi

Abstract

To predict the amino acid sites important for the clearance of hepatitis C virus (HCV) subtype 1b *in vivo*, positively selected amino acid sites were detected by analyzing the sequence data collected from the international DNA databank. The rate of nonsynonymous substitutions per nonsynonymous site was compared with that of synonymous substitutions per synonymous site for each codon site in the entire coding region. As a result, 13 out of 3010 amino acid sites were found to be positively selected. Among the 13 positively selected amino acid sites, eight were located in the structural proteins and five were in the nonstructural proteins. Moreover, eight were located in B-cell epitopes and two were in T-cell epitopes. These observations suggest that both the antibody and the cytotoxic T lymphocyte are involved in the clearance of HCV subtype 1b *in vivo*. These positively selected amino acid sites represent candidate vaccination targets for HCV subtype 1b. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Hepatitis C virus; Synonymous substitution; Nonsynonymous substitution; Positive selection; Epitope; Vaccine

1. Introduction

Hepatitis C virus (HCV), the sole member of the genus *Hepacivirus* in the family *Flaviviridae* (van Regenmortel et al., 2000), is an enveloped, non-segmented, single-stranded, and positive-sense RNA virus (Choo et al., 1989). The genome of HCV is approximately 9.5 kilobases long, encoding a polyprotein of approximately 3000 amino acids (Kato et al., 1990). The polyprotein is co- and post-translationally cleaved into core protein (C), envelope glycoprotein 1 (E1), E2, p7, nonstructural protein 2 (NS2), NS3, NS4A, NS4B, NS5A, and NS5B in order from its N-terminus by the cellular signalase and the viral proteinases (Grakoui et al., 1993).

The genomic sequences from different HCV isolates are highly divergent (Kato et al., 1989). According to the phylogenetic analysis, HCV has been classified into six clades, in which various numbers of subtypes are included (Robertson

et al., 1998). In the polyprotein of HCV, a region including the N-terminal 27–31 amino acid sites in E2 is known to be the most variable and is called hypervariable region 1 (HVR1; Hijikata et al., 1991; Weiner et al., 1991). HVR1 has been proposed as a major target of the immune response (Weiner et al., 1992; Farci et al., 1996; Zibert et al., 1997a).

More than 50% of humans infected with HCV establish chronic hepatitis, which may progress to cirrhosis and hepatocellular carcinoma (Alter et al., 1992). Interferon alpha and ribavirin are used for treatment of HCV infection. However, they are not highly effective, especially for HCV subtype 1b (Davis et al., 1998; McHutchison et al., 1998). Therefore, development of effective therapies and vaccines against HCV is an urgent subject worldwide.

For developing effective vaccines against HCV, it is important to identify epitopes involved in the clearance of HCV *in vivo*, because they may be the candidate vaccination targets. Although it may be difficult to identify those epitopes experimentally because of the lack of an efficient *in vitro* cell culture system and an *in vivo* animal model system for proliferation of HCV (Blight et al., 2000), they may be identified as the positively selected amino acid sites. This is because the amino acid mutations at those epitopes may provide selective advantage to the mutants by allowing them to escape from the immune response (Endo et al., 1996; Fitch et al., 1997; Nielsen and Yang, 1998; Suzuki

Abbreviations: C, core protein; DDBJ, DNA databank of Japan; E, envelope glycoprotein; HCV, hepatitis C virus; HVR, hypervariable region; NS, nonstructural protein; *P*, probability; c_s , number of synonymous substitutions per codon; c_n , number of nonsynonymous substitutions per codon; d_s , number of synonymous substitutions per synonymous site; d_n , number of nonsynonymous substitutions per nonsynonymous site; s_s , number of synonymous sites per codon; s_n , number of nonsynonymous sites per codon

* Corresponding author. Tel.: +81-559-81-6847; fax: +81-559-81-6848.

E-mail address: tgojobor@genes.nig.ac.jp (T. Gojobori).

and Gojobori, 1999; Yamaguchi-Kabata and Gojobori, 2000).

Positive selection operating at the amino acid sequence level can be detected by comparing the rate of nonsynonymous substitution per nonsynonymous site (r_N) with that of synonymous substitution per synonymous site (r_S). A higher value of r_N than r_S is an indicator of positive selection, whereas the inverse is an indicator of negative selection (Hughes and Nei, 1988, 1989). By comparing r_N and r_S at single codon sites, it is possible to detect positive selection at single amino acid sites (Fitch et al., 1997; Nielsen and Yang, 1998; Suzuki and Gojobori, 1999; Yamaguchi-Kabata and Gojobori, 2000).

In this study, we identified positively selected amino acid sites in the entire coding region of HCV subtype 1b by using the method of Suzuki and Gojobori (1999) to predict the amino acid sites important for the clearance of HCV subtype 1b *in vivo*.

2. Materials and methods

2.1. Sequence data

In this study, we focused on HCV subtype 1b because of its clinical importance and a large number of sequence data deposited in the international DNA databank. Note that a relatively large number of sequences are required for detecting positively selected amino acid sites by using the method of Suzuki and Gojobori (1999). Actually, the numbers of sequences for another subtype were too small to detect positively selected amino acid sites in the entire coding region.

For detecting positively selected amino acid sites in the entire coding region of HCV subtype 1b, the simplest way is to analyze only sequence data which contain the entire coding region. However, the number of such sequences was relatively small. Thus, we divided the entire coding region of HCV subtype 1b into eight regions encoding C, E1, E2, NS2, NS3, NS4, NS5A, and NS5B and analyzed each region separately. p7 and E2 were combined as E2, and NS4A and NS4B were combined as NS4, because p7 (63 amino acid sites) and NS4A (54 amino acid sites) were too short to analyze positively selected amino acid sites. The numbers of amino acid sites for C, E1, E2, NS2, NS3, NS4, NS5A, and NS5B were 191, 192, 426, 217, 631, 315, 447, and 591, respectively.

A total of 7262 entries which included the successive terms 'hepatitis C virus' in their organism names were collected from the international DNA databank (DDBJ release 40). These entries included all clades and subtypes of HCV. To collect nucleotide sequences for each coding region of HCV subtype 1b, we defined HCV-JS (Accession number: D85516; Tanaka et al., 1995) as a reference sequence for HCV subtype 1b and made 7261 pairwise alignments, each of which consisted of HCV-JS and one of the other sequences, by using the computer program

CLUSTAL W (Thompson et al., 1994). We then extracted sequences which did not contain any gaps within each coding region in the pairwise alignment. The subtype 1b sequences were identified by reconstructing a phylogenetic tree for each coding region. The numbers of nucleotide sequences collected for C, E1, E2, NS2, NS3, NS4, NS5A, and NS5B were 129, 135, 59, 74, 68, 70, 69, and 61, respectively. These sequences originated largely from unrelated patients and only a few were from the same patients. In this paper, the amino acid positions are numbered according to HCV-JS.

2.2. Data analysis

A multiple alignment was made for each coding region by using CLUSTAL W. The positively selected amino acid sites were identified by using the method of Suzuki and Gojobori (1999). Briefly, a phylogenetic tree was reconstructed by the neighbor-joining method (Saitou and Nei, 1987) using the number of synonymous substitutions (Nei and Gojobori, 1986). The ancestral sequence was inferred at each node in the phylogenetic tree by using the maximum parsimony method (Hartigan, 1973). Then, the average numbers of synonymous (s_S) and nonsynonymous (s_N) sites and the total numbers of synonymous (c_S) and nonsynonymous (c_N) substitutions throughout the phylogenetic tree were estimated for each codon site. A probability (P) of obtaining the observed or more biased numbers of synonymous and nonsynonymous substitutions was computed for each codon site, assuming a binomial distribution. In the computation, $s_S/(s_S + s_N)$ and $s_N/(s_S + s_N)$ were used as the probabilities of the occurrences of synonymous and nonsynonymous substitutions, respectively. The significance level was set at 5%. A significantly larger value of c_N than c_S was considered as an indicator of positive selection, whereas the inverse was considered as an indicator of negative selection. The number of synonymous substitutions per synonymous site (d_S) and that of nonsynonymous substitutions per nonsynonymous site (d_N) were estimated by c_S/s_S and c_N/s_N , respectively.

3. Results and discussion

The results for identifying positively selected amino acid sites in the entire coding region of HCV subtype 1b are summarized in Fig. 1. d_S exceeded d_N at most of the amino acid sites (2426/3010, 80.60%), and negative selection was detected at more than half (1560/3010, 51.83%) of all amino acid sites. The HCV polyprotein contains many B-cell and T-cell epitopes. For example, nearly the entire coding region of E1 and E2, and NS3 have been reported as B-cell and T-cell epitopes, respectively (Zibert et al., 1997a, 1999; Tabatabai et al., 1999). However, the amino acid sites involved in the clearance of HCV subtype 1b *in vivo* may be limited, because most of the sites in the entire

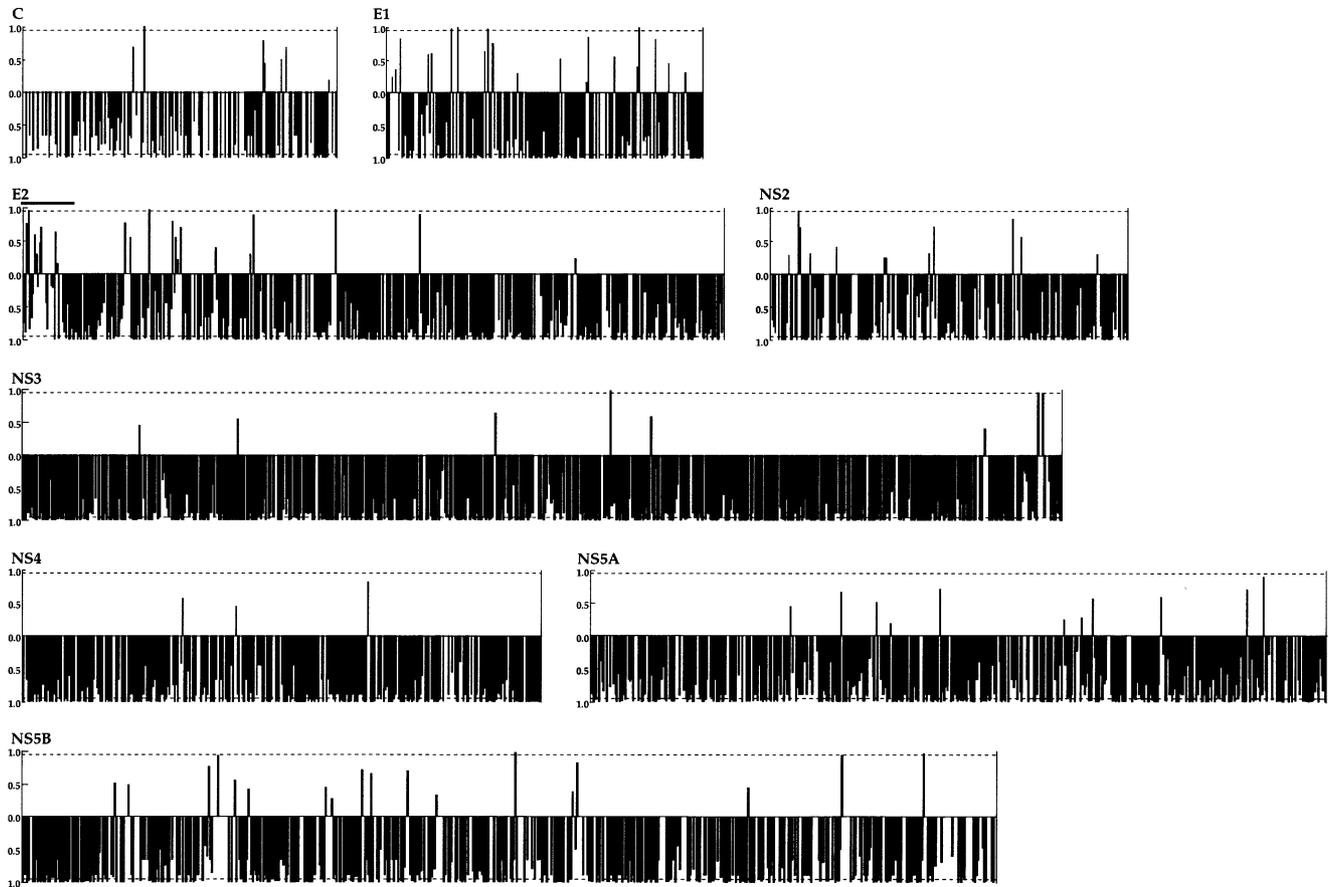


Fig. 1. Distribution of the value of $(1 - P)$ in the entire coding region of HCV subtype 1b. The entire coding region is divided into eight coding regions (C, E1, E2, NS2, NS3, NS4, NS5A, and NS5B). The abscissa indicates the amino acid positions and the ordinate indicates the value of $(1 - P)$ for each amino acid site. When d_N is larger than d_S , the value is indicated above the abscissa, whereas in the opposite situation, the value is indicated below the abscissa. Dotted lines indicate the 5% significance level. A filled rectangle in E2 indicates the position of HVR1.

coding region, including E1, E2, and NS3, are negatively selected (Fig. 1).

Indeed, d_N exceeded d_S at only 265 amino acid sites (8.80%), and positive selection was detected at only 13 amino acid sites (0.43%) (Fig. 1). Among the 13 positively selected amino acid sites, eight were located in the structural proteins (C, E1, and E2) and five were in the nonstructural proteins (NS2, NS3, NS4, NS5A, and NS5B). Since the structural proteins occupy only 26.9% (809/3010) of the HCV polyprotein, positively selected amino acid sites seemed to be located more densely in the structural proteins than in the nonstructural proteins ($P = 0.009$).

Table 1 summarizes the functions of positively selected amino acid sites for HCV subtype 1b. Most of the sites are located in B-cell (Zibert et al., 1997a, 1999; Jolivet-Reynaud et al., 1998; Pereboeva et al., 1998, 2000; Nakano et al., 1999) and T-cell (Tabatabai et al., 1999; Wang and Eckels, 1999) epitopes, suggesting that both antibodies and cytotoxic T lymphocytes (CTLs) are involved in the clearance of HCV subtype 1b *in vivo*. However, a larger number of positively selected amino acid sites were located in B-cell epitopes than in T-cell epitopes. This is probably because

Table 1
Functions of positively selected amino acid sites in the entire coding region of HCV subtype 1b

Position ^a	Protein	Function	References
75	C	B-cell epitope	Pereboeva et al., 1998
231	E1	B-cell epitope	Zibert et al., 1999
235	E1	B-cell epitope	Zibert et al., 1999
253	E1	B-cell epitope	Zibert et al., 1999
345	E1	NA ^b	NA
387	E2	B-cell epitope	Zibert et al., 1997a
461	E2	B-cell epitope	Nakano et al., 1999; Zibert et al., 1999
574	E2	B-cell epitope	Nakano et al., 1999; Zibert et al., 1999
827	NS2	NA	NA
1384	NS3	T-cell epitope	Tabatabai et al., 1999; Wang and Eckels, 1999
1644	NS3	B-cell epitope	Pereboeva et al., 2000
2719	NS5B	T-cell epitope	Tabatabai et al., 1999
2968	NS5B	NA	NA

^a The amino acid position is numbered according to HCV-JS.

^b NA, not available.

the recognition of T-cell epitopes is restricted by the haplotype of human leukocyte antigen, whereas that of B-cell epitopes is not. The sequence data used in this study were collected from the international DNA databank, which included HCV subtype 1b sequences largely from unrelated patients. Therefore, the positive selection may be more efficiently detected in B-cell epitopes than in T-cell epitopes in this study. This is consistent with the above observation that positively selected amino acid sites are more densely located in the structural proteins than in the nonstructural proteins, because B-cell epitopes are mainly located in structural proteins. Since most positively selected amino acid sites were located in B-cell and T-cell epitopes, amino acid positions 345, 827, 2719, and 2968 may also be parts of B-cell and T-cell epitopes.

The positively selected amino acid sites in B-cell and T-cell epitopes may be the vaccination targets against HCV subtype 1b, because these sites should be highly immunogenic and involved in the clearance of HCV subtype 1b *in vivo*. However, it should be noted that these sites are often highly variable, so that HCV may escape from the immune response by producing antigenic mutants (Nowak et al., 1991; Weiner et al., 1992; Farci et al., 2000). This may be facilitated by a high evolutionary rate of HCV (Ina et al., 1994; Smith et al., 1997; Allain et al., 2000; Suzuki et al., 2000) and shifting immunodominance of the immune response (Nowak et al., 1995). Nevertheless, it has been reported that if the immune response is sufficiently strong in an acute infection, HCV is unable to escape from the immune response even if it is directed against highly variable epitopes (Missale et al., 1996; Zibert et al., 1997a,b; Farci et al., 2000). Moreover, the composite vaccines containing different amino acid residues, particularly predicted future amino acid residues, at positively selected amino acid sites may be useful for preventing proliferation of escape mutants (Bush et al., 1999). Since both antibodies and CTLs seem to be involved in the clearance of HCV subtype 1b, it may be more effective to use both B-cell and T-cell epitopes as the vaccination targets.

In conclusion, we identified 13 positively selected amino acid sites in the entire coding region of HCV subtype 1b. These sites may be candidate vaccination targets against HCV subtype 1b.

Acknowledgements

The authors thank Dr Allison Wyndham at the National Institute of Genetics, Japan, for providing valuable comments on an earlier version of the manuscript. We are grateful to two anonymous reviewers for suggestions to improve the manuscript. This work was supported, in part, by grants from the Ministry of Education, Culture, Sports, Science, and Technology, Japan. Y.S. is supported by the JSPS Research Fellowships for Young Scientists.

References

- Allain, J.-P., Dong, Y., Vandamme, A.-M., Moulton, V., Salemi, M., 2000. Evolutionary rate and genetic drift of hepatitis C virus are not correlated with the host immune response: studies of infected donor-recipient clusters. *J. Virol.* 74, 2541–2549.
- Alter, M.J., Margolis, H.S., Krawczynski, K., Judson, F.N., Mares, A., Alexander, W.J., Hu, P.Y., Miller, J.K., Gerber, M.A., Sampliner, R.E., Meeks, E.L., Beach, M.J., 1992. The natural history of community-acquired hepatitis C in the United States. *N. Engl. J. Med.* 327, 1899–1905.
- Blight, K.J., Kolykhalov, A.A., Rice, C.M., 2000. Efficient initiation of HCV RNA replication in cell culture. *Science* 290, 1972–1974.
- Bush, R.M., Bender, C.A., Subbarao, K., Cox, N.J., 1999. Predicting the evolution of human influenza A. *Science* 286, 1921–1925.
- Choo, Q.-L., Kuo, G., Ralston, R., Weiner, A.J., Overby, L.R., Bradley, D.W., Houghton, M., 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B hepatitis genome. *Science* 244, 359–362.
- Davis, G.L., Esteban-Mur, R., Rustgi, V., Hoefs, J., Gordon, S.C., Trepo, C., Shiffman, M.L., Zeuzem, S., Craxi, A., Ling, M.-H., Albrecht, J., 1998. Interferon alfa-2b alone or in combination with ribavirin for the treatment of relapse of chronic hepatitis C. *N. Engl. J. Med.* 339, 1493–1499.
- Endo, T., Ikeo, K., Gojobori, T., 1996. Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* 13, 685–690.
- Farci, P., Shimoda, A., Wong, D., Cabezon, T., De Gioannis, D., Strazzer, A., Shimizu, Y., Shapiro, M., Alter, H.J., Purcell, R.H., 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* 93, 15394–15399.
- Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J.C., Strazzer, A., Chien, D.Y., Munoz, S.J., Balestrieri, A., Purcell, R.H., Alter, H.J., 2000. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* 288, 339–344.
- Fitch, W.M., Bush, R.M., Bender, C.A., Cox, N.J., 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl. Acad. Sci. USA* 94, 7712–7718.
- Grakoui, A., Wychowski, C., Lin, C., Feinstone, S.M., Rice, C.M., 1993. Expression and identification of hepatitis C virus polyprotein cleavage products. *J. Virol.* 67, 1385–1395.
- Hartigan, J.A., 1973. Minimum mutation fits to a given tree. *Biometrics* 29, 53–65.
- Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Shimotohno, K., 1991. Hypervariable regions in the putative glycoprotein of hepatitis C virus. *Biochem. Biophys. Res. Commun.* 175, 220–228.
- Hughes, A.L., Nei, M., 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335, 167–170.
- Hughes, A.L., Nei, M., 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* 86, 958–962.
- Ina, Y., Mizokami, M., Ohba, K., Gojobori, T., 1994. Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. *J. Mol. Evol.* 38, 50–56.
- Jolivet-Reynaud, C., Dalbon, P., Viola, F., Yvon, S., Paranhos-Baccala, G., Piga, N., Bridon, L., Trabaud, M.A., Battail, N., Sibai, G., Jolivet, M., 1998. HCV core immunodominant region analysis using mouse monoclonal antibodies and human sera: characterization of major epitopes useful for antigen detection. *J. Med. Virol.* 56, 300–309.
- Kato, N., Ohkoshi, S., Shimotohno, K., 1989. Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variations from the original isolate in the U.S.A. *Proc. Jpn. Acad.* 65, 219–223.
- Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T., Shimotohno, K., 1990. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc. Natl. Acad. Sci. USA* 87, 9524–9528.

- McHutchison, J.G., Gordon, S.C., Schiff, E.R., Shiffman, M.L., Lee, W.M., Rustgi, V.K., Goodman, Z.D., Ling, M.-H., Cort, S., Albrecht, J.K., 1998. Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. *N. Engl. J. Med.* 339, 1485–1492.
- Missale, G., Bertoni, R., Lamonaca, V., Valli, A., Massari, M., Mori, C., Rumi, M.G., Houghton, M., Fiaccadori, F., Ferrari, C., 1996. Different clinical behaviors of acute hepatitis C virus infection are associated with different vigor of the anti-viral cell-mediated immune response. *J. Clin. Invest.* 98, 706–714.
- Nakano, I., Fukuda, Y., Katano, Y., Hayakawa, T., 1999. Conformational epitopes detected by cross-reactive antibodies to envelope 2 glycoprotein of the hepatitis C virus. *J. Infect. Dis.* 180, 1328–1333.
- Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. *Mol. Biol. Evol.* 3, 418–426.
- Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Nowak, M.A., Anderson, R.M., McLean, A.R., Wolfs, T.F., Goudsmit, J., May, R.M., 1991. Antigenic diversity thresholds and the development of AIDS. *Science* 254, 963–969.
- Nowak, M.A., May, R.M., Phillips, R.E., Rowland-Jones, S., Laloo, D.G., McAdam, S., Klenerman, P., Koppe, B., Sigmund, K., Bangham, C.R.M., McMichael, A.J., 1995. Antigenic oscillations and shifting immunodominance in HIV-1 infections. *Nature* 375, 606–611.
- Pereboeva, L.A., Pereboev, A.V., Morris, G.E., 1998. Identification of antigenic sites on three hepatitis C virus proteins using phage-displayed peptide libraries. *J. Med. Virol.* 56, 105–111.
- Pereboeva, L.A., Pereboev, A.V., Wang, L.F., Morris, G.E., 2000. Hepatitis C epitopes from phage-displayed cDNA libraries and improved diagnosis with a chimeric antigen. *J. Med. Virol.* 60, 144–151.
- Robertson, B., Myers, G., Howard, C., Brettin, T., Bukh, J., Gaschen, B., Gojobori, T., Maertens, G., Mizokami, M., Nainan, O., Netesov, S., Nishioka, K., Shin-i, T., Simmonds, P., Smith, D., Stuyver, L., Weiner, A., 1998. Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. *Arch. Virol.* 143, 2493–2503.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Smith, D.B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P.L., Simmonds, P., 1997. The origin of hepatitis C virus genotypes. *J. Gen. Virol.* 78, 321–328.
- Suzuki, Y., Gojobori, T., 1999. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* 16, 1315–1328.
- Suzuki, Y., Yamaguchi-Kabata, Y., Gojobori, T., 2000. Nucleotide substitution rates of HIV-1. *AIDS Rev.* 2, 39–47.
- Tabatabai, N.M., Bian, T.-H., Rice, C.M., Yoshizawa, K., Gill, J., Eckels, D.D., 1999. Functionally distinct T-cell epitopes within the hepatitis C virus non-structural 3 protein. *Hum. Immunol.* 60, 105–115.
- Tanaka, T., Kato, N., Cho, M.-J., Shimotohno, K., 1995. A novel sequence found at the 3' terminus of hepatitis C virus genome. *Biochem. Biophys. Res. Commun.* 215, 744–749.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B., 2000. *Virus Taxonomy*. Academic Press, San Diego, CA.
- Wang, H., Eckels, D.D., 1999. Mutations in immunodominant T cell epitopes derived from the nonstructural 3 protein of hepatitis C virus have the potential for generating escape variants that may have important consequences for T cell recognition. *J. Immunol.* 162, 4177–4183.
- Weiner, A.J., Brauer, M.J., Rosenblatt, J., Richman, K.H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q.-L., Houghton, M., Han, J.H., 1991. Variable and hypervariable domains are found in the regions of HCV corresponding to the Flavivirus envelope and NS1 proteins and the Pestivirus envelope glycoproteins. *Virology* 180, 842–848.
- Weiner, A.J., Geysen, H.M., Christopherson, C., Hall, J.E., Mason, T.J., Saracco, G., Bonino, F., Crawford, K., Marion, C.D., Crawford, K.A., Brunetto, M., Barr, P.J., Miyamura, T., McHutchinson, J., Houghton, M., 1992. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc. Natl. Acad. Sci. USA* 89, 3468–3472.
- Yamaguchi-Kabata, Y., Gojobori, T., 2000. Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. *J. Virol.* 74, 4335–4350.
- Zibert, A., Kraas, W., Meisel, H., Jung, G., Roggendorf, M., 1997a. Epitope mapping of antibodies directed against hypervariable region 1 in acute self-limiting and chronic infections due to hepatitis C virus. *J. Virol.* 71, 4123–4127.
- Zibert, A., Meisel, H., Kraas, W., Schulz, A., Jung, G., Roggendorf, M., 1997b. Early antibody response against hypervariable region 1 is associated with acute self-limiting infections of hepatitis C virus. *Hepatology* 25, 1245–1249.
- Zibert, A., Kraas, W., Ross, R.S., Meisel, H., Lechner, S., Jung, G., Roggendorf, M., 1999. Immunodominant B-cell domains of hepatitis C virus envelope proteins E1 and E2 identified during early and late time points of infection. *J. Hepatol.* 30, 177–184.