Olga Kalinina,<sup>1,2</sup> Helene Norder,<sup>2</sup>\* Sergey Mukomolov,<sup>1</sup> and Lars O. Magnius<sup>2</sup>

St. Petersburg Pasteur Institute, St. Petersburg 197101, Russia,<sup>1</sup> and Department of Virology, Swedish Institute for Infectious Disease Control, 171 82 Solna, Sweden<sup>2</sup>

Received 31 July 2001/Accepted 21 December 2001

Hepatitis C virus (HCV) evolution is thought to proceed by mutations within the six genotypes. Here, we report on a viable spontaneous HCV recombinant and we show that recombination may play a role in the evolution of this virus. Previously, 149 HCV strains from St. Petersburg had been subtyped by limited sequencing within the NS5B region. In the present study, the core regions of 41 of these strains were sequenced to investigate the concordance of HCV genotyping for these two genomic regions. Two phylogenetically related HCV strains were found to belong to different subtypes, 2k and 1b, according to sequence analysis of the 5' untranslated region (5'UTR)-core and the NS5B regions, respectively. By sequencing of the E2-p7-NS2 region, the crossover point was mapped within the NS2 region, probably between positions 3175 and 3176 (according to the numbering system for strain pi6CF). Sequencing of the 5'UTR-core regions of four other HCV strains. phylogenetically related to the above-mentioned two strains (based on analysis within the NS5B region), revealed that these four strains were also recombinants. Since a nonrecombinant 2k strain was found in St. Petersburg, the recombination may have taken place there around a decade ago. Since the frequency of this recombinant is now high enough to allow the detection of the recombinant in a fraction of the city's population, it seems to be actively spreading there. The reported recombinant is tentatively designated RF1 2k/1b, in agreement with the nomenclature used for HIV recombinants. Recombination between HCV genotypes must now be considered in the classification, laboratory diagnosis, and treatment of HCV infection.

Hepatitis C virus (HCV) is the major agent of parenterally transmitted non-A, non-B hepatitis worldwide (1). HCV is an enveloped virus classified in the family *Flaviviridae*. It has a single-stranded, positive-sense, nonsegmented RNA genome of approximately 9,500 nucleotides (nt) with a single, long open reading frame encoding a polyprotein of  $\sim$ 3,030 amino acids with the gene order C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B. The structural proteins are C (core) and E1 and E2 (envelope glycoproteins). The function of the p7 gene product is presently unknown. The NS2 through NS5 regions encode putative nonstructural proteins (2).

HCV is characterized by a high degree of genetic heterogeneity. Its variability is, however, distributed unequally within the genome, and some regions, such as NS3 and NS5B, are considerably conserved. HCV is classified into six genotypes and an increasing number of subtypes based on degree of genomic sequence divergence (26, 27). Evolution of HCV is believed to be mediated by point, insertion, or deletion mutations. Another mechanism for viral evolution is through recombination. There has so far been little evidence of recombination between HCV strains of different genotypes (38), and it has been suggested that these events are rare in vivo and that the resultant recombinants are usually not viable (26, 28, 34).

Relatively short subgenomic regions within the 5' untranslated region (5'-UTR), core (C), E1, or NS5, which have been shown to be conserved within a given HCV genotype, are used for the classification of HCV strains (26, 27). Presently, identification of HCV genotypes is based on single subgenomic regions and does not take into consideration recombination. Sequence analysis is not presently employed for routine determination of HCV genotypes but is used as a reference for the evaluation of other genotyping procedures. Most methods for direct HCV genotyping include amplification of subgenomic fragments, such as the 5'UTR, core, E1, and NS4, by PCR with type-specific primers or by restriction fragment length polymorphism analysis of PCR products (15, 17, 29, 30). Indirect HCV genotyping may be achieved by demonstration of typespecific antibodies by enzyme-linked immunosorbent assay (3, 25).

In a previous study, we subtyped 149 HCV strains from St. Petersburg by limited sequencing within the NS5B region (10). In order of decreasing frequency, subtypes 3a, 1b, 1a, 2a, 2c, and 4a were found. In the present study, that work was extended to include also sequencing of the core regions of some of these HCV strains to investigate the concordance of HCV genotype determinations among the different regions of the genome.

### MATERIALS AND METHODS

Serum samples. Forty-one HCV strains from St. Petersburg were arbitrarily selected from the previously studied 149 strains for the characterization of the 5'UTR-core region. Fifteen strains derived from injection drug users, and five derived from hemodialysis patients of different hospitals. Two of the investigated strains showed discordant results of genotyping within the 5'UTR-core and the NS5B regions. Therefore, the 5'UTR-core regions of four additional strains were sequenced, since these strains formed a cluster with the two previously mentioned strains in a dendrogram based on the NS5B region (10). In addition, the NS5B and the 5'UTR-core regions of one subtype 2k strain, from a patient with chronic hepatitis C, sampled among 17 other patients in St. Petersburg during January 2001, were sequenced.

<sup>\*</sup> Corresponding author. Mailing address: Department of Virology, Swedish Institute for Infectious Disease Control, 171 82 Solna, Sweden. Phone: 46 8 457 26 39. Fax: 46 8 337272. E-mail: Helene.Norder @smi.ki.se.



FIG. 1. Strategy used for amplification and sequencing of overlapping PCR fragments. The nucleotide positions are numbered according to the numbering system for the subtype 2a strain pj6CF. Locations of sequencing primers are shown above the PCR products. Arrows indicate primer polarity.

**RNA extraction and cDNA synthesis and amplification.** HCV RNA was extracted from 100- $\mu$ l serum samples by the guanidinium isothiocyanate-phenolchloroform method, performed mainly by the procedure of Garson et al. (J. A. Garson, C. Ring, P. Tuke, and R. S. Tedder, letter, Lancet **336**:878–879, 1990). After ethanol precipitation at  $-70^{\circ}$ C overnight, total RNA was dissolved in 20  $\mu$ l of distilled water. cDNA synthesis was performed with 100 U of reverse transcriptase (Superscript II; GIBCO-BRL, Gaithersburg, Md.) and 0.1 U of random hexamer primer (Boehringer Mannheim) in a total volume of 20  $\mu$ l at 42°C for 2 h. Amplification within the NS5B and 5'UTR-core regions by PCR was performed as previously described (13). Primer pairs 105/106 (14) and univ-1/186 (16, 29) were used for nesting of the NS5B and 5'UTR-core fragments, respectively (Fig. 1).

Amplification of long PCR fragments. The core-to-NS2 and E2-to-NS3 regions (positions 342 to 3485 and 2089 to 3741, respectively, according to the numbering system for reference strain pj6CF) of six HCV strains (N747, N796, N611, N674, N687, and N589) were amplified using the Expand Long Template PCR System (Roche Diagnostics GmbH). Six-microliter volumes of cDNA were subjected to first-round PCR with Expand PCR buffer, 2.5 U of Expand Long Template enzyme mixture (thermostable Taq DNA polymerase and a proofreading polymerase), 0.5 mM each nucleotide, and 50 nM outer primer pairs 140 (16)/NS3-2AS at position 3485 for the core-to-NS2 region and 2089S/3741as for the E2-to-NS3 region. The thermal profile of the first PCR included an initial denaturation at 93°C for 2 min; 10 cycles of denaturation at 93°C for 30 s, annealing at 56°C for 30 s, and elongation at 68°C for 3 min; and 30 cycles with denaturation and annealing as in the first 10 cycles but elongation at 68°C for 3 min, with a 20-s increase for each cycle. A final elongation was performed at 68°C for 10 min. Six microliters of the first PCR product was nested under the conditions that were employed for the first-round PCR, except that 50 nM inner primer pairs 256 (16)/NS2-2as at position 3450 for the core-to-NS2 region and E2-2S (at position 2500)/3741as for the E2-to-NS3 region and an annealing temperature of 57°C were used.

**DNA sequencing and sequence analysis.** All PCR products were purified by using GFX PCR DNA and Gel Band purification kits (Amersham Pharmacia Biotech). The sequencing reactions were carried out with BigDye Terminator Cycle Sequencing Ready Reaction kits (PE Biosystems; version 2.0) and 4 pmol of sequencing primer. PCR products amplified within the NS5B region were sequenced with hep-105; those amplified within the 5'UTR-core region were sequenced with primers univ-1 (29) and 186 (16). The core-to-NS2 and E2-to-NS3 regions were sequenced with primers 140, 256, and COREs1 (at position 800); E2-as1 (at position 2000); 2089S, E2-2S, and p7-2as (at position 2600); NS2-s1 (at position 3250); NS2-2as; NS3-2AS; and 3741as (Fig. 1). An ABI PRISM 3100 genetic analyzer (Applied Biosystems) was used for electrophoresis and data collection. The sequences obtained were edited, assembled, and analyzed as described previously (10). Similarity and bootscanning analyses were performed with the SimPlot program (version 1.2.2 for Windows 95/NT; avail-

able from the author, S. C. Ray, online [http://www.welch.jhu.edu/~sray /download])and a bootscanning program package using the PHYLIP package for phylogenetic analysis (22).

Nucleotide sequence accession numbers. The sequences reported herein have been deposited in GenBank under accession no. AY070167 through AY070215).

# RESULTS

**Phylogenetic analysis within the NS5B and 5'UTR-core regions.** A dendrogram based on the NS5B regions of the 46 St. Petersburg HCV strains and another 55 strains from GenBank was obtained by the neighbor-joining method (Fig. 2A). Subtypes 1b, 2a, and 3a were represented by 23, 2, and 17 St. Petersburg strains, respectively, while there was one strain each of subtypes 1a, 2c, 2k, and 4a.

The 5'UTR-core regions of all 46 HCV strains could be amplified. Six strains, subtyped as 1b in the NS5B region, were related to subtype 2k in the 5'UTR-core region (Fig. 2B). These six strains formed a distinct cluster with one Moldova and one St. Petersburg subtype 2k strain (Fig. 2B). However, in the dendrogram based on the NS5B region, the same six strains formed a distinct cluster within subtype 1b (Fig. 2A). For the other 40 St. Petersburg strains there was no discrepancy between genotyping results based on the NS5B and the 5'UTRcore regions.

**Identification of the recombination point.** Since the same six strains from different patients formed a distinct cluster on comparison of the two different regions but in branches belonging to different genotypes, there was unequivocal evidence for the presence of a recombinant. To identify the crossover point, five recombinant strains (747, 796, 611, 674, and 687) and one nonrecombinant strain (589) were chosen for sequencing from the E2 region downstream since the NS2 region was considered a likely spot for recombination. A large fragment covering the core to the NS3 region (positions 342 to 3485) of each of these strains was amplified and subjected to direct sequencing (Fig. 1). To identify the recombination point, SimPlot and bootscan analyses were performed on the ob-



FIG. 2. Phylogenetic analysis of 102 HCV strains, performed on 280 nt within the NS5B region (A) and on 548 nt in the 5'UTR-core region (B) by the neighbor-joining method with GBV-B (accession no. NC 001655) as the outgroup. Sequences of 55 HCV isolates from GenBank are defined by the name of the relevant strain or clone (Table 1). Recombinant strains are indicated by the letter R and are shown in bold. Bootstrap values are given on the branches as percentages from 300 replicates. USA, United States of America.



FIG. 2—Continued.

Strain or clone	Origin	Genotype	EMBL accession no.
HC-J1	United States	1a	D10749
HCV-H	United States	1a	M67463
HCV-A	Australia	1b	AJ000009
HCV-ad78	Germany	1b	A I132996
HD-1	Germany	1b	13132550 1145476
HC C2	China	10 1b	D10034
	China	10	D10934
	Cililia	10	L02650
HCV-L2	Korea	10	001214
Hc 16362	Korea	16	016362
089019	Taiwan	16	089019
HCJTa	Japan	16	D11168
HCJTb	Japan	1b	D11355
HC4/83	Japan	1b	D13558
MD 1-1	Japan	1b	AF165045
MD 1-2	Japan	1b	AF165046
MD 2-1	Japan	1b	AF165047
MD 2-2	Japan	1b	AF165048
MD 3-1	Japan	1b	AF165049
MD 4-1	Japan	1b	AF165051
MD 5-1	Ianan	1b	AF165053
MD 8-2	Janan	1b	AF165060
133	Janan	1b	D14484
K1-R1	Japan	10 1b	D50480
K1 D2	Japan	10 1b	D50480
V1 D2	Japan	10 1b	D50481
KI-KJ V1 S1	Japan	10	D50482
K1-51 K1 52	Japan	10	D30465
K1-52 K1-52	Japan	10	D30463
K1-83	Japan	10	D50484
Hpvncvn	Japan	10	D63857
HCV-JS	Japan	lb	D85516
hpcpp	Japan	16	D30613
hcv-n	Japan	16	AF139594
RB	—	1b	AJ238799
S62220	—	1b	S62220
hc-j4	—	1b	AF054247
Ph77cv-j6s	Chimera	1a/2a	AF177037
pjkCF	United States	2a	AF177036
HC-J6	Japan	2a	D00944
Md2b-1	_	2b	AF238486
BEBE1	Japan	2c	D50409
HCJK025	Jakarta	2e	D49761, D49746
HCJK109	Jakarta	2e	D49772, D49755
HCIK081	Jakarta	2f	D49769, D49754
HCIK139	Thailand	2f	D49777 D49757
VAT-96	Moldova	24	AB031663
CB	Australia	30	AE0/6866
Hpccapa	Ianan	30	D17763
	Japan	20	D28017
	Japan	Ja 10a	D20917
JKU47 ED42	Jakana Middle Deet	108	D49704
EU43	Nildule East	4a 5 -	¥ 11004 ¥12104
EUH1480	South Africa	Sa	¥ 13184
VINZ35	Vietnam	60	D21317, D84263
HCJK046	Jakarta	6e	D49763, D63822
VN004	Vietnam	6h	D21315, D84265
VN405	Vietnam	6k	D21318, D84264

TABLE 1. Sequences of HCV strains retrieved from GenBank and used in the phylogenetic analysis

tained sequences, which included the core-NS2 and parts of the NS3 and NS5B regions (Fig. 3). The whole 5'UTR-to-p7 sequence of the recombinant strains was similar to that of the subtype 2k genome. The corresponding sequence of the nonrecombinant strain (589) was, as expected, similar to that of the subtype 1b genome. While the 5' part of the NS2 sequence of the recombinant strains was similar to that of subtype 2k, the 3' part of this region was similar to that of subtype 1b. Comparison of these sequences with those of two strains from GenBank, k1-s2 (subtype 1b) and Vat-96 (subtype 2k), allowed mapping of the crossover point within the NS2 region, upstream of a stretch of 12 nt encoding the conserved region Tyr-Asp/Asn-His-Leu encompassing the His-952 residue of the cysteine protease (Fig. 4). Nucleotide substitutions within this stretch of 12 bases were all synonymous and did not give definite information regarding the crossover point (Fig. 4A), although the sequences were consistent with the same single recombination site for the five recombinant strains being com-



Window: 120 bp, Step: 10 bp, GapStrip: On, Reps: 1000, Max Likelihood, T/t: 2.0, NEIGHBOR

FIG. 3. (A) Similarity plots for two recombinant strains and reference strains based on seven genomic regions: 5'UTR to NS3 (a) and NS5B (b). The query sequence is the 2k/1b recombinant strain 796. An arrow indicates the crossover point at position 3175/3176. (B) Bootscan for 2k/1b recombinant strain 796 based on 5' UTR-to-NS3 region. The parental genotypes are represented by the genotype 1b strain 589, from St. Petersburg, and the genotype 2k strain Vat-96. The genotype 3a strain CB was used as the outgroup. The bootstrap values are based on 1,000 replicates. Nucleotide positions are numbered as described in the legend to Fig. 1.

Δ

~					
30,50	3060	3070	3080	30,90	
AAGCACGC	CATACTTCGT	CAGAGCGCAA	CGCTGTTGA	GGATATGTGCTG	VAT-96 2k
[G]				· · · · · · · · · [] · [] ·	796-StP 2k/1b
		[c]	$\mathbf{T}$	· · · · · · · · · · · · · · · · · · ·	687-StP 2k/1b
· · · · · · · · · ·				<u>T</u>	611-StP 2k/1b
	· 南· · · · · · · · ·			· · · · · · · · · · · · · · · · · · ·	674-StP 2k/1b
C.AGT C.A.G.T	. G	A C . C T	GC. CA.TC	TGCCATGT	589-SEP 15 K1-S2 1b
C . A A . I				<u></u>	
2100	3110	3120	3130	3140	
					VAT-96 2k
CIGIAAGA	A .	JIGGCAAGIA			796-StP 2k/1b
	A .		. <u>A</u>		747-StP 2k/1b
. C G	A .			G G	687-StP 2k/1b
· · · · · · · ·		$\cdots$	ΔΑ		674-StP 2k/1b
TA.GC.G.	A.AG.TG.C.	. G T C . T	r	GCCT.CA.G.AG	589-StP 1b
TGGC.G	A.AG.G.T.	. д т т	ГА	GCTT.CA.G.AG	K1-S2 1b
		1		:	
		. т	vr <sup>asp/</sup> His Le	u	
		↓ <u>⊢</u>	<u>2 - asn</u>		
3150	3160	3170	3180	3190	
CTTGGTAA	ATGGACTGGC	ACATACATCT	ATGATCACCT	C T C C C C A A T G T C	VAT-96 2k 796-StP 2k/1b
· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · ·		C T	. A . T C C G	747-StP 2k/1b
<u>C</u> . G	C	с т .	с т[	<u>T</u> A.TCCA	687-StP 2k/1b
. <i></i>		C	<u>A</u> . C T	. A . T C C G	611-StP 2k/1b
		C		TAT C.CG	589-StP 1b
G . C C G C	. C T G T		C T	T A . T	K1-S2 1b
3200	3210	3220	3230	3240	
AGGTTGGG	CTGCTAGTGG	GCTACGCGAC	CTCGCGGTTG	CTGTAGAACCCA	VAT-96 2k
G . A C	. CCACGTG	C	[G] .	. A T G G	796-StP 2k/1b
G . A C	. CCACGCG . ,	C A	<u>.</u>	. A T G G	747-StP 2k/1b
G . A C	. CCACGCG	C A	T G .	. A T G T G	687-StP 2k/1b
G.AC	CCAC.CG.		T	AGG	611-StP 2k/1D
G . A	CCA.G.G.	Č.G.A	T	$\overrightarrow{A}$ $\overrightarrow{T}$ $\overrightarrow{G}$ $\overrightarrow{T}$ $\overrightarrow{G}$	589-StP 1b
G . A C	. CCACGCG	T G	T G .	<u> </u>	K1-S2 1b
3250	3260	3270	3280	3290	
TTGTTTC	AGCCCAATGG	AGAAAAAGT	CATTGTGTGG	GGAGCGGAGACG	VAT-96 2k
. <u>C</u> C	TCTGAC	C C G A .	CACC		796-StP 2K/1D
c	TCTGAC	C C G A .	<u>C</u> ACC		687-StP 2k/1b
. C c	TCTGAC	C C G A .	A C C	C	611-StP 2k/1b
. c c	TCTGAC	C C G A'.	CACC	$\cdot \cdot  \mathbf{G} \cdot \cdot \cdot \cdot \cdot \cdot  \mathbf{T} \cdot \cdot  \mathbf{C} $	674-StP 2k/1b
	TCTGAC	C T G A .			K1-S2 1b
<u> </u>	<u> </u>				•
3300	3310	3320	3330	3340	
GCACCGTC	TGGTGACATC	TTGCACGGC	<u>, , , , , , , , , , , , , , , , , , , </u>	CGCCCGCTTGGG	- VAT-96 2k
G	G	A.T.TG.CT	C		796-StP 2k/1b
G	G	A . T . T G T T	. <u>.</u>   c	A A G	747-StP 2k/1b
G	G	A . T . T G C T	. A C	A A G	687-StP 2k/1b
G	G	A T TG CT			674-StP 2k/1b
		ALCTCG.	. A C		589-StP 1b
	C	A.CTTGT.	G . c	A A G	K1-S2 1b

FIG. 4. Alignment of the nucleotide sequences within the NS2 region (A) and the amino acid sequences of the E2-p7 and p7-NS2 junctions (B) of reference strains k1-s2, 589, and Vat-96 and recombinant strains 796, 747, 687, 611, and 674. Arrows show the crossover point. Vertical broken lines show putative cleavage sites for the HCV polyprotein according to De Francesco (2). Nucleotide positions are numbered as for Fig. 1.

pared, in the codon for Val/Ile upstream of the aforementioned four conserved codons.

## DISCUSSION

To rule out the possibility of reverse transcription-PCR artifacts, large fragments from the same five recombinant strains and one nonrecombinant strain, covering the middle of the E2 region to the beginning of the NS3 region (positions 2089 to 3741), were also sequenced (Fig. 1). The obtained sequences were almost identical and verified the same recombination site. Phylogenetic analyses based on two different genomic regions, 5'UTR-core and NS5B, suggested the existence of HCV recombinant strains, since 2 of 41 randomly investigated strains exhibited different subtypes, 2k and 1b, based on the 5'UTRcore and NS5B regions, respectively. Sequencing of the core regions of another four St. Petersburg HCV strains that, based

B	E2	p7			
723 733	743	753	763	773	
LFLLLADARVCACLWMLLI	LLGQAEAAL	EKLVILHAAS	ASSHGMLCFI	IFFIAAWYIKGR	Vat-96 2k
$\cdot$		· · · · · · · · · · · · · · · · · · ·	N . I I	H	796-StP 2k/1h
			FN II.	н	687-StP 2k/11
		· · · · · · · · · · · · · · · · · · ·	N I	H	611-StP 2k/11
· · · · · · · · · · · · · · · · · · ·	. I A	. N V . N	/.GAI.S.I	v	589-StP 1b
	I A	. N V . N V	7. GA F. S. I	VC	K1-S2 1b
	57	NC2			
	<b>▲ Р</b> /				
783 793	803	813	823	833	-
VTPLVTYSYLGMWSFSLLI	LLALPQQAY	ALDTTEQGQIC	GLVLLVVISVE	TLSPAYKILLCR	Vat-96 2k
$\mathbf{V}$ $\mathbf{A}$	 	· · · [3] · · · · · L		· · · · · · ·	747-StP 2k/1
V . A L		<u>N</u> L	· · · · · · · · · · · · [		687-StP 2k/1
A V A A L			· · [A] · · · [1] · · · ·	· · · · · · · · · · · · · · · · · · ·	611-StP 2K/1
LV.GAA.ALY.V.PLL	P R	. M. REMAASC	GAVFAGLMLI	H V A .	589-StP 1b
	P R	. M. REMAASC	GAVFLGLALI	<u> </u>	K1-S2 1b
843 853	863	873	883	893	
SLWWLSYLLVRAEALIQD	WVPPWQARG	GRDGIIWAAT	LECPGVLFDI7	NWLLAILGPGYL	Vat-96 2k
. I	<mark>L</mark> . <u>.</u>	A	. <u>                                    </u>	КАН.	796-StP 2k/1
	L	V		K	687-StP 2k/1
A	. 🗉   L	A	L I V	Кн.	611-StP 2k/1
LIO.FITHL.V	L . V . A L N V		ч	. к	674-StP 2k/1 589-StP 1b
LIO.FITCL.V	. I LNV	A L L T C	ALH.ELI.S.	КІ.VLМV	к1-52 1Ъ
903 913	923	933	943	953	
K A		SGGKYVQMML.	LTLGKWTGTYI	LYDHLSPMSGWAA	yat-96 2k 796-StP 2k/1
. [K]		<u>I</u>	R	T . L R D H	747-StP 2k/1
	 . L		R	н нн	687-StP 2k/1 611-StP 2k/1
. к <u></u>	· ·	_, <u>.</u> .I			674-StP 2k/1
QAGI.RVG.I	A.ML.KV	A.H.AFI	KAALV	/н. Ин.	589-StP 1b
			<u> </u>		_ KI*82 10
963 973	983	993	1003	1013	
SGLRDLAVAVEPIVFSPMI	EKKVIVWGA	ETAACGDILH	GLPVSARLGQI	VLLGPADEYTS	Vat-96 2k
	. T . I . T T . I . T			IFSLEG	796-StP 2k/1 747-StP 2k/1
A	. T . I . T	I I I I I I I I I I I I I I I I I		I	687-StP 2k/1
$\mathbf{T}$ $\mathbf{V}$ $\mathbf{D}$ .	. T . I . T	D I L		ILSLEG	611-StP 2k/1
$\begin{array}{c} \mathbf{A} \\ \mathbf{G} \\ \mathbf{G} \\ . \\ . \\ \mathbf{V} \\ \mathbf{I} \\ \mathbf{I} \\ \mathbf{V} \\ \mathbf{I} \\ $	. T . I . T	D I S		I SFEG	589-StP 2k/1
<u>A</u>	<u>. T . I . T</u>	D	A R . K	IFSLEG	K1-S2 1b

FIG. 4-Continued.

on examination of the NS5B region, were closely related to two recombinant strains indicated that they were all recombinants. Previously discrepant results from genotyping in the same two regions have been reported for two Honduran HCV strains obtained from polytransfused patients, but it could not be convincingly proven that this finding was not due to mixed infections (38). Since concordant results from genotyping HCV in different regions of the genome have been repeatedly reported in the past, it has been commonly believed that recombination events are selected against or that they generally do not generate viable strains (19, 26, 34). Resistance to superinfection at the cellular level and infection-induced immunity are other possible explanations for the rarity of natural recombinants of HCV.

Phylogenetic analyses of some other viruses, such as human immunodeficiency virus and hepatitis B virus, have revealed different evolutionary histories for different genomic regions, indicating that these organisms have undergone recombination in the past (8, 14, 21). Further sequence analysis of the corresponding genomes has allowed mapping of the crossover junctions and confirmed the findings from limited sequencing of different genomic regions.

Our observations from phylogenetic analyses of six strains, within two different regions, support the theory that a spontaneous recombination of HCV genomes had occurred. The identification of the recombination crossover point provided the final proof of the existence of a natural recombinant HCV strain. By direct sequencing of a large PCR fragment covering the core-NS2 region, it was determined that the crossover point was located downstream the NS2 region, possibly between nt 3175 and 3176 (according to the numbering system for the pj6CF strain), within the codon encoding Val/IIe (residue 949). His-952 is the first motif residue associated with the enzymatic activity of the NS2-NS3 cysteine protease (6). Also, the conserved stretch corresponding to codons 950 to 953 downstream might facilitate recombination events.

An infectious HCV chimera, comprising the complete open reading frame of a subtype 1b strain and the 5' and 3' termini of a subtype 1a strain, has been constructed from two related HCV subtypes (36). However, intergenotypic chimeric cDNA clones constructed from infectious clones of subtypes 1a and 2a were not infectious (37). A plausible explanation for this outcome is that the chosen junction, just before or after the p7 region, was not optimal. Our mapping of the junction for a spontaneous recombinant may promote the successful construction of intertypic infectious HCV chimeras. This does not necessarily imply that the identified site always allows the construction of viable HCV hybrids or is the only site that allows them, considering that the mechanism of recombination for HCV is still unknown.

Recombination plays a significant role in the evolution of RNA viruses by creating genetic variation, by reducing mutational load, and by creating viruses with new properties (35). Natural recombinants occur among nonsegmented plus-strand RNA viruses, such as picornaviruses (4, 7) and alphaviruses (12). More importantly, recombinants have recently been reported to occur in other members of the Flaviviridae family, such as dengue virus (9, 32), and in pestiviruses, such as bovine viral diarrhea virus (20). The recombination breakpoints for nonsegmented plus-strand RNA viruses, such as polioviruses and other picornaviruses, as well as members of the family Flaviviridae, are often located in the part of the genome encoding the first nonstructural protein but are sometimes in genes encoding structural proteins (usually envelope proteins) (4, 7, 32). Analysis of the crossover junction reported here suggests that there exists an analogous site for recombination in HCV.

Apparently, a clone resulting from recombination between the widely distributed HCV subtype 1b and the recently described subtype 2k is circulating in St. Petersburg. There remains the question of when and where this recombination occurred. Gene transfer from one genotype to another requires physical cohabitation of both genotypes in the same cell, which requires a mixed infection in a patient. To be further spread, the recombinant not only must be viable but also must have been selected for. Superinfection of patients carrying another HCV strain or with mixed infections has been repeatedly reported (11, 24, 33). However, the frequency of such infections is not well known. The occurrence of a novel intergenotypic recombination will depend on the presence of different genotypes in a population and the viability of the hybrids generated. It is not surprising that one parental strain was subtype 1b, since this has been the predominant subtype in St. Petersburg in the past (10). It was, however, not anticipated that the other parental strain would belong to the most divergent genotype 2 and, moreover, to subtype 2k, which was described only recently as a single strain from Moldova (23). While six recombinant strains involving subtype 2k were found in St. Petersburg, only one nonrecombinant 2k strain was encountered. Since this subtype rarely occurs in the population, there is a low probability of occurrence of novel and independent recombination events involving this subtype. Considering the relative frequency of subtype 2k, the estimated rate of mixed infection involving this subtype should thus be low.

Unless selected for, it is difficult to explain why one of the parental viral strains had the rare 2k subtype. Phylogenetic analysis showed that the subtype 2k strain from St. Petersburg was not more similar to the recombinant than to the Vat-96 strain from Moldova. However, the subtype 1b parental strain was similar to strains circulating in St. Petersburg. Although a nonrecombinant 2k strain was found in St. Petersburg, the small numbers of HCV strains that have been subtyped after collection from other parts of Russia do not allow any definite conclusions on the frequency of the 2k subtype in other parts of that country. However, in the early 1990s, Moldovian and other workers started to travel back and forth between Moldova and St. Petersburg, so it is possible that the 2k strain was introduced to St. Petersburg from Moldova. Therefore, the recombination event may have taken place in St. Petersburg around a decade ago, since some of the patients harboring recombinants were known to be positive for anti-HCV already in the mid-1990s. The finding of a 2k/1b recombinant strain in 5% of HCV-infected patients of St. Petersburg shows that this particular recombinant HCV strain not only is viable but is spreading in that city.

In conclusion, our observations support the concept that natural recombinants of HCV do exist and that hybrid strains may be not only viable but selected for. Although the recombinant is not necessarily more virulent, it will contribute to HCV evolution. Thus, recombination may play a role in creating genetic diversity in HCV. This has previously unanticipated implications in terms of understanding the biology as well as the control of HCV infections, with regard to both laboratory diagnosis and treatment. The genotypes of HCV are considered predictors of the outcome of interferon treatment (39), and interferon resistance has been attributed to sequence variability of the NS5A protein (amino acid residues 2209 to 2248) and of the E2 region (5, 18, 31). Our data indicate that analysis of more than one subgenomic region is necessary to avoid missing recombinant strains, and they show that genotyping within the core region may not always be a valid means of predicting the response to interferon.

In the current system of classification, HCV strains are divided into genotypes, subtypes, and quasispecies. Recombinant strains have not yet been considered in this classification system. In analogy with the nomenclature for human immunodeficiency virus, we suggest that an HCV recombinant strain be designated as a "recombinant form" (RF). RF strains with the same number are progeny resulting from the same recombination event and thus share an identical mosaic structure. According to this nomenclature, we suggest the designation RF1\_2k/1b for the herein-described recombinant strains from St. Petersburg. This means the first recombinant resulting from an event involving the 5' end of subtype 2k and the 3' end of subtype 1b which has the same recombination point within the NS2 region.

#### ACKNOWLEDGMENTS

This work was supported by grant G-126 from Östeuropa Kommitéen and by European Union program INCO Copernicus (contract IC15-CT98-0304).

#### REFERENCES

Choo, Q. L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, R. Medina-Selby, P. J. Barr, et al. 1991. Genetic organi-

zation and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA  $88{:}2451{-}2455.$ 

- De Francesco, F. R. 1999. Molecular virology of the hepatitis C virus. J. Hepatol. 31(Suppl. 1):47–53.
- Dixit, V., S. Quan, P. Martin, D. Larson, M. Brezina, R. DiNello, K. Sra, J. Y. N. Lau, D. Chien, J. Kolberg, A. Tagger, G. Davis, A. Polito, and G. Gitnick. 1995. Evaluation of a novel serotyping system for hepatitis C virus: strong correlation with standard genotyping methodologies. J. Clin. Microbiol. 33:2978–2983.
- Furione, M., S. Guillot, D. Otelea, J. Balanant, A. Candrea, and R. Crainic. 1993. Polioviruses with natural recombinant genomes isolated from vaccineassociated paralytic poliomyelitis. Virology 196:199–208.
- Gale, M. J., M. J. Korth, N. M. Tang, S. L. Tan, D. A. Hopkins, T. E. Dever, S. J. Polyak, D. R. Gretch, and M. G. Katze. 1997. Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. Virology 230:217–227.
- Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice. 1993. A second hepatitis C virus-encoded proteinase. Proc. Natl. Acad. Sci. USA 90:10583–10587.
- Guillot, S., V. Caro, N. Cuervo, E. Korotkova, M. Combiescu, A. Persu, A. Aubert-Combiescu, F. Delpeyroux, and R. Crainic. 2000. Natural genetic exchanges between vaccine and wild poliovirus strains in humans. J. Virol. 74:8434–8443.
- Hannoun, C., H. Norder, and M. Lindh. 2000. An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam. J. Gen. Virol. 81:2267–2272.
- Holmes, E. C., M. Worobey, and A. Rambaut. 1999. Phylogenetic evidence for recombination in dengue virus. Mol. Biol. E vol. 16:405–409.
- Kalinina, O., H. Norder, T. Vetrov, K. Zhdanov, M. Barzunova, V. Plotnikova, S. Mukomolov, and L. O. Magnius. 2001. Shift in subtype of HCV from 1b to 3a in St. Petersburg mediated by increase in drug addiction. J. Med. Virol. 65:517–524.
- Kao, J. H., P. J. Chen, M. Y. Lai, and D. S. Chen. 1993. Superinfection of heterologous hepatitis C virus in a patient with chronic type C hepatitis. Gastroenterogy 105:583–587.
- Lai, M. M. C. 1992. RNA recombination in animal and plant viruses. Microbiol. Rev. 56:61–79.
- Norder, H., Å. Bergström, I. Uhnoo, J. Aldén, L. Weiss, J. Czajkowski, and L. Magnius. 1998. Confirmation of nosocomial transmission of hepatitis C virus by phylogenetic analysis of the NS5-B region. J. Clin. Microbiol. 36: 3066–3069.
- Morozov, V., M. Pisareva, and M. Groudinin. 2000. Homologous recombination between different genotypes of hepatitis B virus. Gene 260:55–65.
- Ohno, O., M. Mizokami, R.-R. Wu, M. G. Saleh, K.-I. Ohba, E. Orito, M. Mukaide, R. Williams, and J. Y. N. Lau. 1997. New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. J. Clin. Microbiol. 35:201–207.
- Okamoto, H., Y. Sugiyama, S. Okada, K. Kurai, Y. Akahane, Y. Sugai, T. Tanaka, K. Sato, F. Tsuda, and Y. Miyakawa. 1992. Typing hepatitis C virus by polymerase chain reaction with type-specific primers: application to clinical surveys and tracing infectious sources. J. Gen. Virol. 73:673–679.
- Okamoto, H., H. Tokita, M. Sakamoto, M. Horikita, M. Kojima, H. Iizuka, and S. Mishiro. 1993. Characterization of the genomic sequence of type V (or 3a) hepatitis C virus isolates and PCR primers for specific detection. J. Gen. Virol. 74:2385–2390.
- Polyak, S. J., K. S. A. Khabar, D. M. Paschal, H. J. Ezelle, G. Duverlie, G. N. Barber, D. E. Levy, N. Mukaida, and D. R. Gretch. 2001. Hepatitis C virus nonstructural 5A protein induces interleukin-8, leading to partial inhibition of the interferon-induced antiviral response. J. Virol. 75:6095–6106.
- Prescott, L. E., A. Berger, J. M. Pawlotsky, P. Conjeevaram, I. Pike, and P. Simmonds. 1997. Sequence analysis of hepatitis C virus variants producing discrepant results with two different genotyping assays. J. Med. Virol. 53: 237–244.

- Ridpath, J. F., and J. D. Neill. 2000. Detection and characterization of genetic recombination in cytopathic type 2 bovine viral diarrhea viruses. J. Virol. 74:8771–8774.
- Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn. 1995. Recombination in HIV-1. Nature 374:124–126.
- Salminen, M. O., and W. Cobb. 1998. Bootscanning package for Unix/Linux version 1. National Public Health Institute, Helsinki, Finland.
- Samokhvalov, E. I., M. Hijikata, R. I. Gylka, D. K. Lvov, and S. Mishiro. 2000. Full-genome nucleotide sequence of a hepatitis C virus variant (isolate name VAT96) representing a new subtype within the genotype 2 (arbitrarily 2k). Virus Genes 20:183–187.
- Sheng, L., M. Willems, K. Peerlinck, J. Vermylrn, and S. H. Yap. 1995. Hepatitis C virus genotypes in Belgian hemophiliacs. J. Med. Virol. 45:211– 214.
- 25. Simmonds, P., K. A. Rose, S. Graham, S.-W. Chan, F. McOmish, B. C. Dow, E. A. C. Follett, P. L. Yap, and H. Marsden. 1993. Mapping of serotypespecific, immunodominant epitopes in the NS-4 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. **31**:1493–1503.
- Simmonds, P., D. B. Smith, F. McOmish, P. L. Yap, J. Kolberg, M. S. Urdea, and E. C. Holmes. 1994. Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. J. Gen. Virol. 75:1053–1061.
- Simmonds, P. 1999. Viral heterogeneity of the hepatitis C virus. J. Hepatol. 31(Suppl. 1):54–60.
- Smith, D. B., and P. Simmonds. 1997. Molecular epidemiology of hepatitis C virus. J. Gastroenterol. Hepatol. 12:522–527.
- Stuyver, L., R. Rossau, A. Wyseur, M. Duhamel, B. Vanderborght, H. Van Heuverswyn, and G. Maertens. 1993. Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. J. Gen. Virol. 74:1093–1102.
- Stuyver, L., A. Wyseur, W. van Arnhem, F. Lunel, P. Laurent-Puig, J. M. Pawlotsky, B. Kleter, L. Bassit, J. Nkengasong, L. J. van Doorn, et al. 1995. Hepatitis C virus genotyping by means of 5'-UR/core line probe assays and molecular analysis of untypeable samples. Virus Res. 38:137–157.
- Taylor, D. R., S. T. Shi, P. R. Romano, G. N. Barder, and M. M. C. Lai. 1999. Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. Science 285:107–110.
- 32. Tolou, H. J., P. Couissinier-Paris, J. P. Durand, V. Mercier, J. J. de Pina, P. de Micco, F. Billoir, R. N. Charrel, and X. de Lamballerie. 2001. Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. J. Gen. Virol. 82:1283–1290.
- Tuveri, R., C. Rothschild, S. Pol, D. Reijasse, T. Persico, C. Gazengel, C. Brechot, and V. Thiers. 1997. Hepatitis C virus genotypes in French haemophiliacs: kinetics and reappraisal of mixed infections. J. Med. Virol. 51:36– 41.
- Viazov, S., A. Widell, and E. Nordenfelt. 2000. Mixed infection with two types of hepatitis C virus is probably a rare event. Infection 28:21–25.
- Worobey, M., and E. C. Holmes. 1999. Evolutionary aspects of recombination in RNA viruses. J. Gen. Virol. 80:2535–2543.
- 36. Yanagi, M., M. St. Claire, M. Shapiro, S. U. Emerson, R. H. Purcell, and J. Bukh. 1998. Transcripts of a chimeric cDNA clone of hepatitis C virus genotype 1b are infectious in vivo. Virology 244:161–172.
- Yanagi, M., R. H. Purcell, S. U. Emerson, and J. Bukh. 1999. Hepatitis C virus: an infectious molecular clone of a second major genotype (2a) and lack of viability of intertypic 1a and 2a chimeras. Virology 262:250–263.
- Yun, Z., C. Lara, B. Johansson, D. R. Lorenzana, and A. Sönnerborg. 1996. Discrepancy of hepatitis C virus genotypes as determined by phylogenetic analysis of partial NS5 and core sequences. J. Med. Virol. 49:155–160.
- Zein, N. N. 2000. Clinical significance of hepatitis C virus genotypes. Clin. Microbiol. Rev. 13:223–235.