JOURNAL OF VIROLOGY, Oct. 1994, p. 6778–6781
0022-538X/94/$04.00+0
Copyright © 1994, American Society for Microbiology
Vol. 68, No. 10

The Transactivator Gene of Human T-Cell Leukemia Virus Type I Is More Variable within and between Healthy Carriers than Patients with Tropical Spastic Paraparesis

STEFAN NIEWIESK,[1]* SUSAN DAENKE,[1] CLAIRE E. PARKER,[1] GRAHAM TAYLOR,[2] JONATHAN WEBER,[2] SIMON NIGHTINGALE,[3] AND CHARLES R. M. BANGHAM[1,4]

Institute of Molecular Medicine,[1] and Department of Virology, Public Health Laboratory,[4] John Radcliffe Hospital, Headington, Oxford OX3 9DU, St. Mary's Hospital Medical School, London W2 1NY,[2] and The Midland Centre for Neurosurgery and Neurology, West Midlands, Birmingham B67 7JX,[3] United Kingdom

Received 2 February 1994/Accepted 18 July 1994

Human T-cell leukemia virus type I (HTLV-I) causes T-cell leukemia and tropical spastic paraparesis (TSP) in a minority of infected people, whereas the majority remain healthy. No association between a particular HTLV-I sequence and disease manifestation has been found in previous studies. We studied here the sequence variability of the gene for the HTLV-I Tax protein, which is the dominant target antigen of the very strong cytotoxic T-lymphocyte response to the virus. In HTLV-I infection, the intraisolate nucleotide variability is much greater than the variability between isolates. The predicted protein sequence of Tax was significantly more variable in the healthy seropositive individuals' proviruses than in those of the patients with TSP. Thus, *tax* sequence heterogeneity, rather than the presence of particular sequences, distinguishes healthy HTLV-I-seropositive individuals from patients with TSP.

Human T-cell leukemia virus type I (HTLV-I)-infected individuals may develop adult T-cell leukemia (ATL) or tropical spastic paraparesis (TSP) (9, 16, 26). The pathogenesis of these diseases is unresolved, and no link between disease progression and viral strain has been found so far. Little sequence difference between HTLV-I isolates has been found, although there are enough nucleotide differences to allow isolates to be grouped according to geographical regions (5, 6, 11, 19, 24). In contrast to the low level of variability of HTLV-I between isolates, recent studies have shown some intrastrain variability, indicating a quasispecies structure of HTLV-I (1, 4, 5).

We have investigated the nature of the HTLV-I quasispecies by sequencing 20 clones of the proviral *tax* gene from each of four TSP patients and four healthy seropositive individuals. TD, HD, HH, and HM have been described elsewhere (20–22). Each subject is denoted by a two-letter code: for the first letter, T denotes a patient with TSP and H denotes a healthy HTLV-I carrier. The subjects TD, TB, TG, and TI were 63-, 56-, 56-, and 57-year-old females from the West Indies, who developed TSP 12, 27, 21, and 3 years earlier, respectively. HD was a 69-year-old male, and HH was a 22-year-old female from the West Indies. HM and HN, respectively, were 45- and 36-year-old females. HN was diagnosed HTLV-I seropositive 2 years previously. The duration of infection for all the patients is not known, although it is presumed from their personal histories to have been many years. Genomic DNA was extracted from 5×10^6 to 10×10^6 peripheral blood lymphocytes by the proteinase K technique (15). One-microgram quantities of DNA were amplified by PCR using buffer containing 50 mM KCl, 10 mM Tris (pH 8.3), 1 mM $MgCl_2$, 0.2 mM deoxynucleoside triphosphates, 2.5 U of *Taq* polymerase (Perkin-Elmer Cetus), and 1 μM primers. A nested set of primers was used: HT1 (5′-GATAGCAAACCG TCAAGCACAG-3′; positions 7158 to 7179) and HT2B (5′-TTGAGCCATATGCGTGCCATGA-3′; 8608 to 8587) for the primary amplification and BG II (5′-TACAGATCTC-CATGGCCCACTTCCCAGG-3′; 7309 to 7336) and TSRII (5′-CTGAGAATTCAGAGCCTTAGTCT-3′; 8396 to 9019) for the secondary. The amplified DNA products were inserted into M13mp18. DNA was purified and sequenced as described previously (15), with T7 polymerase and a DNA sequencing kit (United States Biochemical) with the M13 forward primer, TS1 (5′-GCTCTACAGTTCCTTATC-3′; 7518 to 7535), TS2 (5′-GAACCTGTACACCCTCTG-3′; 7743 to 7758), TS3 (5′-CCTTTTCCAGCCTGTTAG-3′, 7967 to 7984), or TS4 (5′-CTCATACAGTACTCTTCCT-3′; 8205 to 8223).

Consensus sequences differ little between isolates. In order to obtain a reference sequence for the *tax* gene, a grand consensus was generated by comparing sequences from four healthy seropositive individuals (this study) and the patients with TSP (four subjects from this study and H5 [27], Xav, Bou, Akr, and Jap3 [11]). The grand consensus differed from the first reported ATK-1 sequence (25) at four nucleotide positions (nucleotide 7919 C→T, 7981 C→T, 8230 A→G, and 8366 A→C [Table 1]; nucleotides are numbered according to the HTLV-I [ATK-1] sequence in the EMBL database). The consensus sequences from our eight subjects differed from the grand consensus in only 1 to 6 nucleotides (0.1 to 0.6%) (Table 1). There was no association between a particular consensus sequence and the presence or absence of disease. The data obtained in this study confirm the strong sequence similarity between different HTLV-I isolates, as reported elsewhere (2, 3, 6, 11, 14, 19, 24).

The *tax* gene is highly variable within isolates. The low degree of variability between consensus sequences from different subjects contrasts markedly with the variability within a

* Corresponding author. Present address: Institut für Virologie und Immunbiologie, Versbacher Str. 7, 97078 Würzburg, Germany. Phone: 49/931/201-5954. Fax: 49/931/201-3934. Electronic mail address: Stefan Niewiesk@vax.rz.uni-wuerzburg.d400.de.

TABLE 1. Comparison of isolate sequences with the grand consensus of Tax[a]

Sequence	Nucleotide change at position:																							
	7359	7599	7644	7658	7709	7755	7769	7833	7877	7880	7917	7919	7981	8000	8013	8112	8190	8230	8249	8294	8335	8336	8348	8366
ATK-1												C	C					G						A
Grand consensus	G	A	A	C		C	C	A	T	C	C	T	T	A	A	G	C	A	G	G	G	C	C	C
HD											T													
HH	A							G							C						A	G		
HM		G	G	T		A		G	C											A	G			
HN							T												G					
TB								G							C		T				A	G		
TD			G					G							C						A	G		
TG																				A			T	
TI					T									G		A								
Amino acid change	G→R	I→V	M→V			L→E		I→V			L→F	*	*		N→H	G→S		*			G→E	G→E		*

[a] The grand consensus represents the average sequence of the subjects' sequences from this study, H5 (27), Xav, Bou, Akr, and Jap3 (24). The subjects' own consensus sequences were compared with the grand consensus. *, deviation from the grand consensus by the first described ATK-1 sequence (25).

patient. This is unlike human immunodeficiency virus, where interisolate variation exceeds intraisolate variation (28). The 20 clones sequenced per patient were compared with the patient's own consensus sequence. Within the viral swarm, between 16 and 100 mutations were found, that is, 0.8 to 4.8 mutations per clone (Table 2). The percentage of variable nucleotide positions within the viral swarm (1.5 to 9.4%) was higher than that of positions varying between isolate consensus sequences (0.1 to 0.6%). In each isolate, we found between 12 and 17 distinct *tax* sequences (Table 2, with the exception of HH). Individual clones differed in 1 to 8 positions from the patient's own consensus. This variability within an isolate confirmed the quasispecies nature of HTLV-I (1, 4, 5). However, one-third to one-quarter of the clones were identical to the consensus sequence (with the exception of HH). Therefore the master sequence (i.e., the most common sequence) was identical to the consensus. The consensus sequence does not seem to change over time (5, 13, 18).

The proportion of nonsynonymous nucleotide changes is higher in healthy HTLV-I carriers. The quasispecies did not seem to differ between healthy seropositive subjects and patients with TSP, except in respect of the percentage of nucleotide mutations that changed the amino acid code (healthy seropositive subjects, 78 to 87%; and patients with TSP, 52 to 64%). By random genetic drift, between 66 and 77% of the mutations would be expected to be nonsynonymous (10). In absolute numbers, the proportion of nonsynonymous to synonymous changes was 157/190 (83%) for healthy seropositive subjects and 66/107 (62%) for patients with TSP. This difference is significant (chi-square test, $P < 0.001$), indicating that the virus population in healthy seropositive subjects is more variable, at the protein level, than that in patients with TSP.

The percentage of nonsynonymous nucleotide mutations gives only an approximate indication of the intensity of selection for amino acid change. Nei and Gojobori (17) developed an algorithm to estimate the rate of nonsynonymous changes per nonsynonymous site (d_N) and the rate of synonymous changes per synonymous site (d_S). The ratio d_N/d_S gives a more accurate measure of the selection intensity for amino acid change, corrected for the proportion of nonsynonymous sites in the sequences examined, and so allows a formal comparison to be made between samples. A d_N/d_S ratio of <0.5 is usually interpreted as showing conservation of amino acid sequence, while a ratio of >0.5 implies selection for amino acid change (12, 17). This algorithm has been widely used to compare gene sequences between animal species but has not been used to analyze intraisolate sequence variation in a virus.

The calculation of the d_N/d_S ratio for the different clones within an isolate revealed a marked difference between the two groups of subjects. Whereas the ratio for the viral swarm within the patients with TSP was relatively low (d_N/d_S between 0.3 and 0.6), the ratio for the healthy seropositive subjects was around or slightly above 1 (Table 3). The clustering of the patients with TSP in a group with a low ratio and the healthy seropositive subjects in a group with a high ratio is significant ($P < 0.03$; two-tailed sign test). The difference between the highest d_N/d_S ratio observed for a patient with TSP and the lowest ratio in a healthy carrier was significant ($P < 0.05$; two-tailed t test). Comparing the respective consensus sequences between isolates, we found again that the d_N/d_S ratio is higher among healthy carriers than among patients with TSP (Table 4). This demonstrates that the smaller protein sequence diversity observed for the patients with TSP is not due to the

TABLE 2. Comparison of HTLV-I quasispecies[a]

Patient	No. of clones	No. of mutations	No. of mutations per clone	No. of different sequences	Difference from grand consensus (%)	Variability within isolate (%)	Coding changes (%)
HD	20	24	1.2	15	0.1	2.3	80
HH	21	100	4.8	21	0.5	9.4	87
HM	22	28	1.3	16	0.6	2.6	79
HN	20	38	1.9	14	0.2	3.6	78
TB	19	36	1.9	17	0.5	3.4	64
TD	22	29	1.3	13	0.5	2.7	52
TG	21	16	0.8	12	0.2	1.5	63
TI	19	31	1.6	14	0.3	2.9	64

[a] Eight subjects were included in this comparison; four healthy seropositive subjects (signified by the prefix H) and four patients with TSP (signified by the prefix T). Up to 22 clones of the *tax* gene per individual were sequenced, and comparisons were made between the viral clones and the subject's own consensus sequence. The only exception is the column showing differences from the grand consensus, where the patient's own consensus sequence was compared with the grand consensus.

proliferation of a small number of HTLV-I sequences in these patients and strengthens the conclusion that the observed difference in the d_N/d_S ratio is real. Interestingly, the d_N/d_S ratio among ATL isolates was very low (Table 4), in agreement with a previous study by Ina and Gojobori (7). This suggests that there are stronger constraints on sequence variation in Tax among patients with ATL than among those with TSP or healthy carriers.

These results indicate that there is stronger selection for amino acid change in Tax, or less constraint on sequence variation, in healthy HTLV-I carriers than in patients with TSP. The question arises: what are the selection forces that act on Tax which might explain this difference? We consider two possibilities here. (i) Although the age at infection with HTLV-I in these subjects is not known, the average age of the patients with TSP is greater (58 years) than that of the healthy carriers (43 years). It is possible that synonymous changes accumulate randomly during long-term infection, and this might lower the d_N/d_S ratio of the patients with TSP. However, the frequency of synonymous changes we observed for patients with TSP was not significantly higher than that for healthy carriers, so we reject this explanation. (ii) Nonsynonymous changes in *tax* may lead to less efficient recognition of the protein by cytotoxic T lymphocytes (CTL), which are present in large numbers in HTLV-I-infected people (8, 21, 22). We have found that naturally occurring amino acid changes abolish the recognition of CTL epitopes in Tax (18), so we conclude that CTL may indeed exert a significant selection force on Tax in vivo. However, we cannot yet explain the difference in selection intensities between patients with TSP and healthy carriers on this basis. One possibility is that the higher copy number of HTLV-I in patients with TSP leads to a different pattern of selection by CTL.

Exclusion of *Taq* polymerase errors. Half of the described nucleotide substitutions were seen in more than one PCR amplification and are therefore highly unlikely to be *Taq* polymerase errors. Random substitutions will produce a d_N/d_S ratio of 1. The consistently low d_N/d_S ratios for genomic DNA of patients with TSP, which indicates selection against amino acid changes, therefore suggests that *Taq* errors contributed little to the observed variation in *tax* sequences. However, with substitutions that were found only once, it is not possible to exclude *Taq* errors.

In summary, we have analyzed HTLV-I quasispecies by examining the sequence variation in the *tax* gene. We found that within-isolate variation in *tax* exceeded the between-isolate variation. Unexpectedly, we found a significantly higher proportion of nonsynonymous nucleotide changes in healthy HTLV-I carriers than in patients with TSP, indicating that the selection for amino acid conservation in Tax is stronger for patients with TSP. We conclude that although no particular sequence of HTLV-I was associated with a distinct disease outcome, the virus populations of patients with TSP and healthy carriers differed significantly.

S.N. was supported by AIDS-Stipendienprogramm (Bundesministerium für Forschung und Technologie). C.R.M.B. and S.D. were supported by Welcome Trust. C.E.P. was supported by an MRC training fellowship.

TABLE 3. Comparison of the d_N/d_S ratios for the *tax* gene[a]

Isolate	d_N	d_S	d_N/d_S
HD	0.0025	0.0028	0.892
HH	0.0091	0.0071	1.281
HM	0.0031	0.0026	1.192
HN	0.0052	0.0042	1.238
TB	0.0034	0.0054	0.629
TD	0.0026	0.0074	0.351
TG	0.0015	0.0049	0.306
TI	0.0028	0.0068	0.411

[a] The rate of synonymous changes per synonymous site (d_S), the rate of the nonsynonymous changes per nonsynonymous site (d_N), and the d_N/d_S ratio were calculated by pairwise comparison of the viral clones from each patient by the method of Nei and Gojobori (17) for the full-length *tax* genes.

TABLE 4. Comparison of the d_N/d_S ratios for the *tax* consensus sequences of healthy seropositive individuals, TSP patients, and patients with ATL[a]

Group (n)	d_N	d_S	d_n/d_S
Patients with ATL (6)	0.0016	0.0210	0.076
Patients with TSP (9)	0.0046	0.0125	0.368
Healthy carriers (4)	0.0033	0.0052	0.635

[a] The rate of synonymous changes per synonymous site, the rate of the nonsynonymous changes per nonsynonymous site, and the d_N/d_S ratio were calculated by pairwise comparison of consensus sequences from patients with ATL and TSP and healthy seropositive subjects by the method of Nei and Gojobori (17). n represents the number of sequences used. In the ATL group, six previously published sequences were included (ATK-1 [25], HS-35 [14], MC-1 [23], and Sie, Jap1, and Jap3 [11]). The TSP group consisted of four subjects from this study and H5 (27), Xav, Bou, Akr, and Jap3 (11); the sequences used for the healthy seropositive subjects were obtained from this study.

REFERENCES

1. **Berneman, Z. N., R. B. Gartenhaus, M. S. Reitz, M. E. Klotman, and R. C. Gallo.** 1992. cDNA sequencing confirms HTLV-I expression in adult T-cell leukemia/lymphoma and different sequence variations in vivo and in vitro. Leukemia **6**(Suppl. 3):67S–71S.
2. **Daenke, S., S. Nightingale, J. K. Cruickshank, and C. R. M. Bangham.** 1990. Sequence variants of human T-lymphotropic virus type I from patients with tropical spastic paraparesis and adult T-cell leukemia do not distinguish neurological from leukemic isolates. J. Virol. **64:**1278–1282.
3. **De, B. K., M. D. Lairmore, K. Griffis, L. J. Williams, F. Villinger, T. C. Quinn, C. Brown, Nzilambi, M. Sugimoto, S. Araki, and T. M. Folks.** 1991. Comparative analysis of nucleotide sequences of the partial envelope gene (5′ domain) among human T lymphotropic virus type I (HTLV-I) isolates. Virology **182:**413–419.
4. **Ehrlich, G. D.** 1992. DNA sequencing analysis of the gene encoding the HTLV-I P21E transmembrane protein reveals inter- and intraisolate genetic heterogeneity. Virology **186:**619–627.
5. **Gessain, A., R. C. Gallo, and G. Franchini.** 1992. Low degree of human T-cell leukemia/lymphoma virus type I genetic drift in vivo as a means of monitoring viral transmission and movement of ancient human populations. J. Virol. **66:**2288–2295.
6. **Gray, G. S., M. White, T. Bartman, and D. Mann.** 1990. Envelope gene sequence of HTLV-I isolate MT-2 and its comparison with other HTLV-I isolates. Virology **177:**391–395.
7. **Ina, Y., and T. Gojobori.** 1990. Molecular evolution of human T-cell leukemia virus. J. Mol. Evol. **31:**493–499.
8. **Jacobson, S., H. Shida, D. E. McFarlin, A. S. Fauci, and S. Koenig.** 1990. Circulating CD8+ cytotoxic T lymphocytes specific for HTLV-I pX in patients with HTLV-I associated neurological disease. Nature (London) **348:**245–248.
9. **Kaplan, J. E., M. Osame, and H. Kubota.** 1990. The risk of development of HTLV-I associated myelopathy/tropical spastic paraparesis among persons infected with HTLV-I. J. Acquired Immune Defic. Syndr. **3:**1096–1101.
10. **Kimura, M.** 1968. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11:**247–269.
11. **Komurian, F., F. Pelloquin, and G. de Thé.** 1991. In vivo genomic variability of human T-cell leukemia virus type I depends more upon geography than upon pathologies. J. Virol. **65:**3770–3778.
12. **Li, W.-H., C.-I. Wu, and C.-C. Luo.** 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2:**150–174.
13. **Major, M., S. Nightingale, and U. Desselberger.** 1993. Complete sequence conservation of the human T cell leukaemia virus type 1 tax gene within a family cluster showing different pathologies. J. Gen. Virol. **74:**2531–2537.
14. **Malik, K. T. A., J. Even, and A. Karpas.** 1988. Molecular cloning and complete nucleotide sequence of an adult T cell leukemia virus type 1 (ATLV/HTLV-I) isolate of Caribbean origin: relationship to other members of the ATLV/HTLV-I subgroup. J. Gen. Virol. **69:**1695–1710.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Murphy, E. L., B. Hanchard, J. P. Figueroa, W. N. Gibbs, W. S. Lofters, M. Campbell, J. J. Goedert, and W. A. Blattner.** 1989. Modelling the risk of adult T-cell leukemia/lymphoma in persons with human T-lymphotropic virus type I. Int. J. Cancer **43:**250–253.
17. **Nei, M., and T. Gojobori.** 1986. Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
18. **Niewiesk, S., and C. Bangham.** Unpublished data.
19. **Paine, E., J. Garcia, T. C. Philpott, G. Shaw, and L. Ratner.** 1991. Limited sequence variation in human T-lymphotropic virus type 1 isolates from North American and African patients. Virology **182:**111–123.
20. **Parker, C.** 1994. Ph.D. thesis. University of Oxford, Oxford.
21. **Parker, C. E., S. Daenke, S. Nightingale, and C. R. M. Bangham.** 1992. Activated, HTLV-1-specific cytotoxic T-lymphocytes are found in healthy seropositives as well as in patients with tropical spastic paraparesis. Virology **188:**628–636.
22. **Parker, E. P., S. Nightingale, G. P. Taylor, J. Weber, and C. R. M. Bangham.** 1994. Circulating anti-Tax cytotoxic T lymphocytes from human T-cell leukemia virus type I-infected people, with and without tropical spastic paraparesis, recognize multiple epitopes simultaneously. J. Virol. **68:**2860–2868.
23. **Ratner, L., S. F. Josephs, B. Starcich, B. Hahn, G. M. Shaw, R. C. Gallo, and F. Wong-Staal.** 1985. Nucleotide sequence analysis of a variant human T-cell leukemia virus (HTLV-Ib) provirus with a deletion in pX-I. J. Virol. **54:**781–790.
24. **Ratner, L., T. Philpott, and D. B. Trowbridge.** 1991. Nucleotide sequence analysis of isolates of human T-lymphotropic virus type 1 of diverse geographical origins. AIDS Res. Hum. Retroviruses **7:**923–941.
25. **Seiki, M., J. I. Inoue, T. Takeda, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80:**3618–3622.
26. **Tokudome, S., O. Tokunaga, Y. Shimamoto, Y. Miyamoto, I. Sumidy, M. Kikuchi, M. Takeshita, T. Ikeda, K. Fujiwara, M. Yoshihara, T. Yanagawa, and M. Nishizumi.** 1989. Incidence of adult T-cell leukemia/lymphoma among human T-lymphotropic virus type I carriers in Saga, Japan. Cancer Res. **49:**226–228.
27. **Tsujimoto, A., T. Teruuchi, J. Imamura, K. Shimotohno, I. Miyoshi, and M. Miwa.** 1988. Nucleotide sequence analysis of a provirus derived from HTLV-I associated myelopathy (HAM). Mol. Biol. Med. **5:**29–42.
28. **Wain-Hobson, S.** 1992. Human immunodeficiency virus type 1 quasispecies in vivo and ex vivo, p. 181–194. *In* J. J. Holland (ed.), Genetic diversity of RNA viruses. Springer-Verlag, Berlin.