Journal of General Virology (1997), **78**, 659–664. Printed in Great Britain

JC virus regulatory region rearrangements and genotypes in progressive multifocal leukoencephalopathy: two independent aspects of virus variation

Hansjürgen T. Agostini,[1] Caroline F. Ryschkewitsch,[1] Elyse J. Singer[2] and Gerald L. Stoner[1]

[1] Laboratory of Experimental Neuropathology, National Institute of Neurological Disorders and Stroke, Bldg. 36, Room 4A-29, National Institutes of Health, 36 Convent Drive, Bethesda, Maryland 20892-4126, USA

[2] Neurology Service, VAMC West Los Angeles, Los Angeles, CA 90073, USA

JC virus (JCV) causes the central demyelinating disease progressive multifocal leukoencephalopathy (PML). JCV strains excreted in the urine are distinguishable from those in PML tissue by the configuration of their regulatory region to the right of *ori*: the archetypal regulatory region, 267 nucleotides long, is rearranged in PML tissue by deletion and duplication. Within the coding region JCV shows variations as a result of virus evolution. Four major genotypes are distinguishable of which Type 1 is based in Europe and Type 2 in Asia. Here, the regulatory region rearrangements and the viral genotypes of 29 JCV strains from PML brain were determined. Rearrangement patterns and genotypes were not associated. In general, deletions occurred before duplications, but exceptions to this rule exist. Each configuration of the 29 rearranged regulatory regions was unique and could be derived directly from the non-rearranged, archetypal form.

The human polyomavirus JC (JCV) has a genome of about 5·1 kb in length. The genome codes for six major proteins: the capsid proteins VP1–3, agnoprotein and the regulatory proteins large and small T antigen (Frisque *et al.*, 1984). Replication is controlled by the viral regulatory region which can exist in two major configurations. In general, urinary strains have a non-rearranged (archetypal) regulatory region of 267 bp (Yogo *et al.*, 1990; Agostini *et al.*, 1996*a*), whereas strains isolated from brain tissue of patients with progressive multifocal leukoencephalopathy (PML) show rearrangements as a result of deletions and/or duplications from the archetypal form (Ault & Stoner, 1993; Iida *et al.*, 1993). Within the regulatory region it is the promoter/enhancer containing area between the origin of replication (*ori*) and the start of the agnoprotein gene which is altered in PML strains.

Author for correspondence: Gerald L. Stoner.
Fax +1 301 402 1030. e-mail stoner@nih.gov

There are three classifications of these regulatory region rearrangements in JCV. The first of these, antedating the idea of rearrangements from an archetypal structure, distinguished regulatory regions with a single copy of the TATA box or the duplication of this element as observed in the prototype isolate, JCV(Mad-1) (Walker & Frisque, 1986). The other two classifications delineate parts of the archetypal regulatory region which tend to be deleted or duplicated in PML brain. In one of these, three rearranged domains were identified (Iida *et al.*, 1993) and in the other six (Ault & Stoner, 1993).

In contrast to the large number of regulatory region rearrangements in PML strains, the four recognized genotypes of JCV are a result of virus evolution (Ault & Stoner, 1992; Iida *et al.*, 1993; Agostini *et al.*, 1995). These genotypes are distinguishable by nucleotide changes in the coding region and by typing sites in the regulatory region to the left of *ori*. JCV Types 1, 2 and 3 most likely originated in Europe, Asia and Africa, respectively. JCV Type 4 strains are derived from an initial recombination between Type 1 and Type 3 within the VP1 gene (Agostini *et al.*, 1996*a*, *b*). The JCV Type 4 genome is closely related to Type 1 with the exception of a short fragment within the VP1 gene, where it is replaced by a Type 3-like sequence. JCV spreads widely in the general population by an unknown route. Primary infections have not been identified, but seroconversion occurs predominantly in late childhood (Walker & Padgett, 1983). After primary infection the virus persists in the kidneys. Viral DNA can be detected in single urine samples of more than 40% of individuals above 30 years of age. In the United States about two-thirds of the urinary strains are JCV Type 1, whereas JCV Types 2 and 4 are each found in 15–20% of the individuals excreting virus. The few examples of Type 3 strains characterized in the United States were detected in African Americans (Agostini *et al.*, 1996*a*).

In connection with immunosuppression JCV can replicate in oligodendrocytes and astrocytes leading to the fatal demyelinating disease PML. There are indications that infection with JCV Type 2 is associated with a higher risk of developing PML than is infection with Type 1 (Stoner *et al.*, 1996). Once considered rare, PML now contributes to the

0001-4384

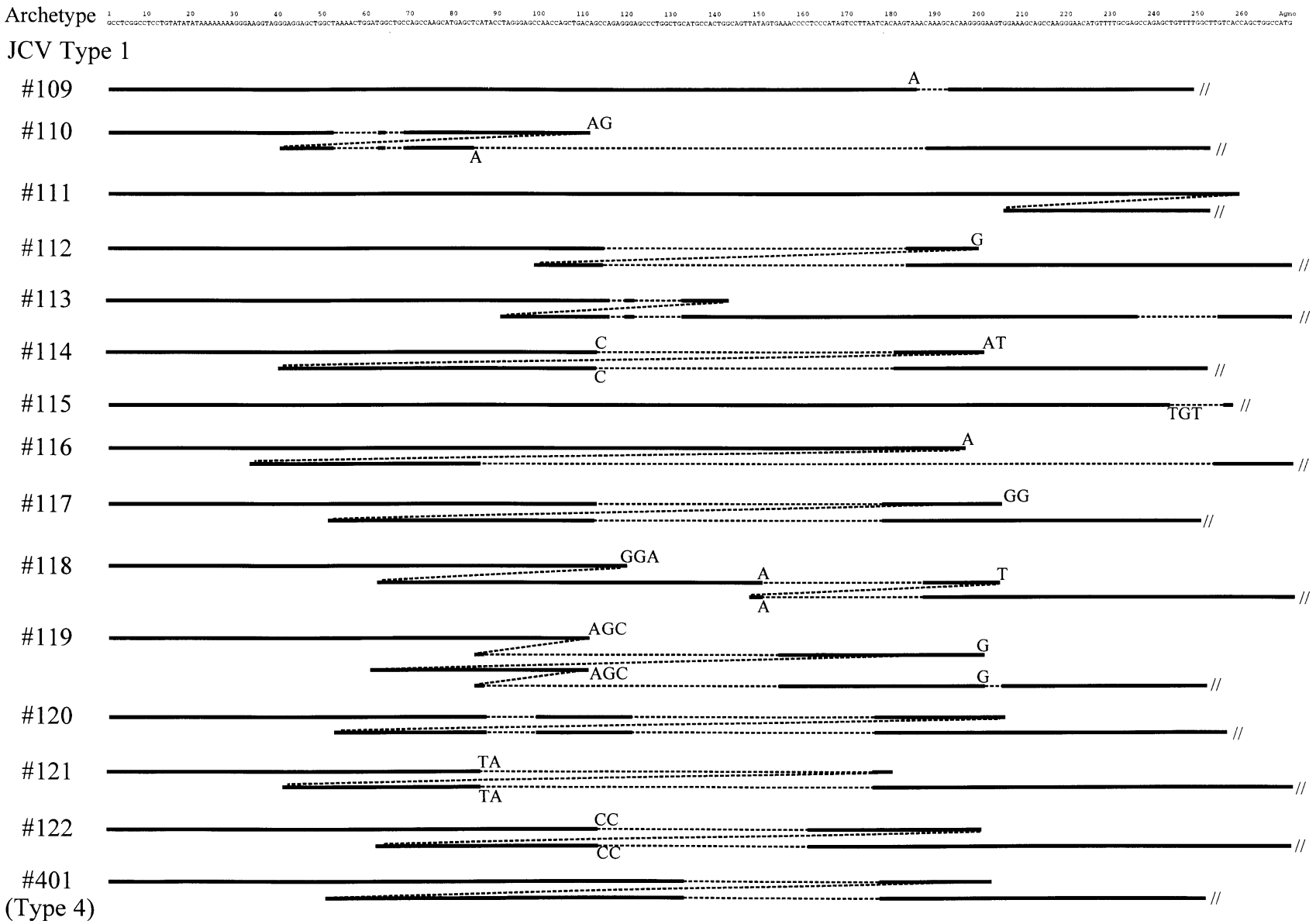

Fig. 1. For legend see facing page.

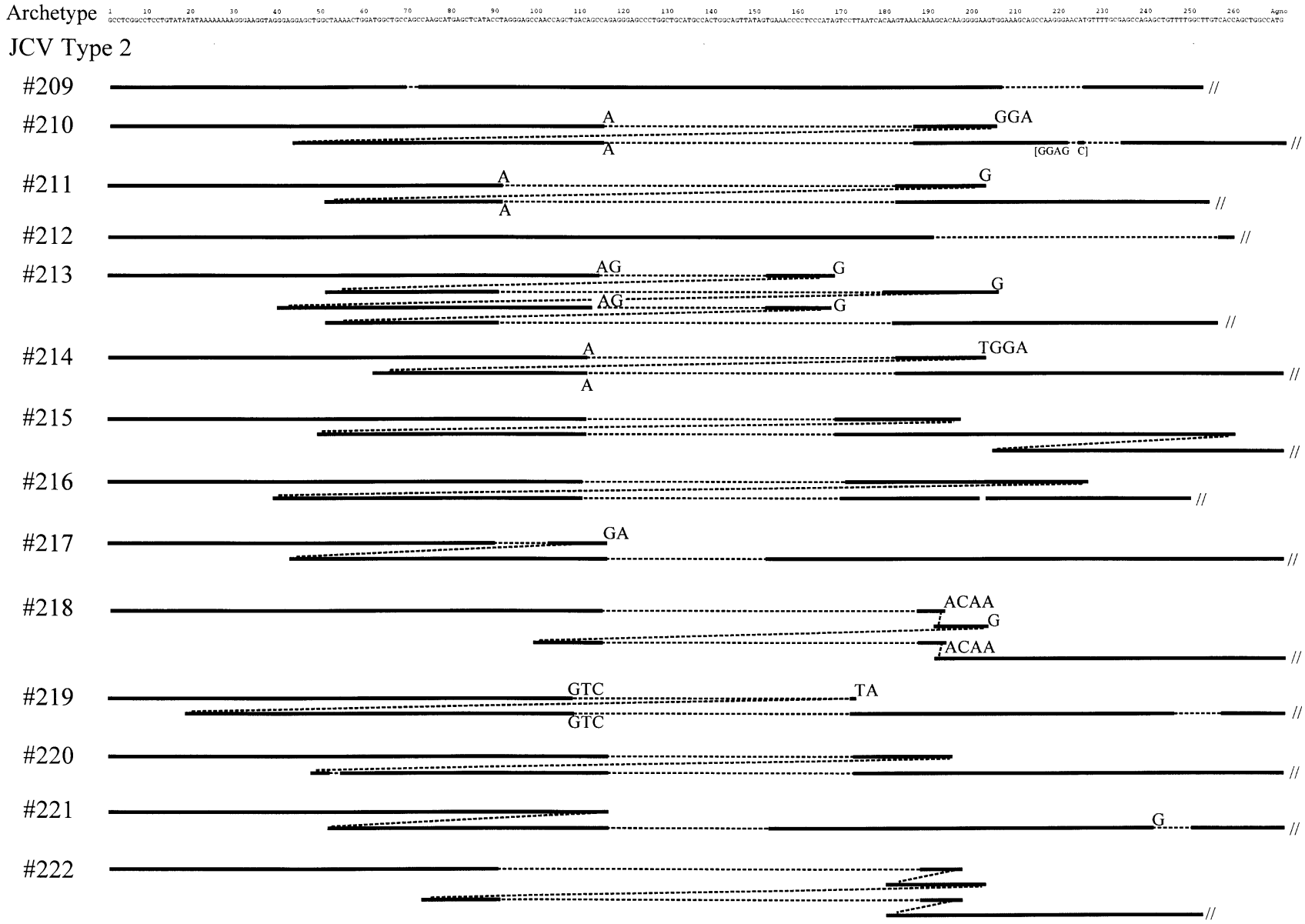

Fig. 1. Regulatory region rearrangements in 29 JCV strains from PML patients. The top line represents the unrearranged, archetypal regulatory region to the right of *ori*. Dotted lines indicate deletions. Nucleotides listed at the beginning of deletions or duplications could be located on either end. Ault & Stoner (1993) defined three general patterns of rearrangements. Typical examples of the so-called 'long duplicate' are the rearrangements in strains #114, #117, #211 and #214. Another pattern, the 'short triplicate' with triple or quadruple copies of the regions flanking the central deletion, was found in four strains (#119, #213, #218, #222). The central region was retained ('D-retaining') in strains #109, #111, #115, #116, #209 and #212. For details of the sequences see Table 1.

terminal illness in about 5% of AIDS patients (Berger & Concha, 1995). The goal of this study was to characterize a large number of PML strains to determine whether particular regulatory region rearrangements in JCV are associated with a specific genotype (Type 1-Type 4) as defined in the coding region.

JCV strains from a total of 29 individual brain or cerebrospinal fluid (CSF) samples from 20 AIDS and nine non-AIDS patients with PML were analysed by PCR and direct cycle sequencing of the PCR products (Agostini *et al*., 1996*a*). Before amplification, brain tissue was digested with proteinase K. CSF was lysed by incubation with NP40 (Stoner & Ryschkewitsch, 1995). A method to PCR amplify the complete viral genome in a single PCR reaction with 5′-overlapping primers (BAM-1 and -2) at the unique *Bam*HI site was used in seven samples (Agostini & Stoner, 1995) (BAM-1: GGGATCCTGTGTTTTCATCATCACTGGC 3′, 4306–4333 [JCV(Mad-1) numbering], BAM-2: AGGATCCCAACACT-CTACCCCACC 3′, 4313–4290). Regulatory region primers were JRR-25 and -28 (JRR-25: CATGGATTCCTCCCTAT-TCAGCA, 4981–5003; JRR-28: TCACAGAAGCCTTAC-GTGACAGC, 313–291) and JRR-1 and -8 (JRR-1: CTTCT-GAGTAAGCTTGGAGGCGG, 5103–5125; JRR-8: GGCG-AAGAACCATGGCCAGCTGG, 289–267). The latter do not include typing sites to the left of *ori*. Primers JLP-1 and -4 amplify a fragment of the VP1 gene [JLP-1: CTCATGTGG-GAGGCTGT(G,T)ACCT, 1769–1790; JLP-4: ATGAAAGC-TGGTGCCCTGCACT, 1897–1876]. Hot start in short range PCRs was performed by adding Ultma DNA polymerase (Perkin Elmer Cetus) at 94 °C followed by 45 cycles with 1 min steps for denaturation (94 °C), annealing (59–63 °C) and elongation (72 °C). PCR products were cleaned by preparative agarose gel electrophoresis and used as templates for bidirectional, direct cycle sequencing with ^{33}P-labelled primers (SequiTherm cycle sequencing, Epicentre Technologies).

Genotyping of JCV strains was based on the partial VP1 sequence amplified by primers JLP-1 and -4. This fragment includes four typing sites which distinguish the four genotypes of JCV (Agostini *et al*., 1996*a*). With 48% each (14 strains), JCV Type 1 and Type 2 were equally distributed among the 29 PML strains as shown in Fig. 1. A single example of a Type 4 strain (#401) was found. The coding region genotype was confirmed by the typing sites to the left of *ori* in the 19 samples amplified by primers JRR-25 and -28.

The viral regulatory region to the right of *ori* was rearranged in all strains examined. None of the rearrangements was identical (Fig. 1). The descriptive method used in Table 1 is precise and universal and reveals areas with essential elements of the regulatory region to the right of *ori*. With the exception of five strains, all changes were a result of both deletions and duplications. In most cases duplication of a particular region occurred after the deletion resulting in identical deletions (e.g. #120, #401, #210). However, single deletions in duplicated areas as shown for strains #116, #216, #217 and #220 indicate that exceptions to this rule exist. Four strains showed only deletions with no duplications (#109, #115, #209, #212). Of these #109 was most closely related to archetype, having only a single 5 bp deletion. The only strain with a duplication but no deletion was #111 (Fig. 1, Table 1). In strain #219 the TATA box was duplicated. The most common pattern of rearrangement, found in 13 strains, was the deletion of the central part of the archetypal regulatory region (basepairs 115–180) associated with a duplication of the flanking upstream and downstream regions (Fig. 1).

Each deletion and each duplication create two breakpoints. Fig. 2 illustrates the cumulative number of breakpoints per site in all strains. The number of breakpoints is increased in the zone of transition between some putative rearrangement domains (Ault & Stoner, 1993; Iida *et al*., 1993). These include the regions between (2)C and D and at the beginning of (1)B. Similar increases were noted between (2)E and F. Although the borders of these elements could not be precisely defined when taking all 29 PML strains into account, they provide a useful tool to characterize patterns of rearrangements (Fig. 1). The profile of Type 1 and Type 2 strains was not distinguishable by this analysis.

The archetypal regulatory region from the right of *ori* to the start of the agnoprotein gene is 267 nucleotides in length. The extremes among the 29 PML strains were #212 (202 bp) and #213 (422 bp). The average length for all strains was 285 ± 47 nucleotides with a median of 287. There were no significant differences between the groups of Type 1 and Type 2 strains. This wide disparity emphasizes that very different structural changes in the regulatory region can generate viable variants. These forms are restricted to an individual CNS and are not transmitted in the population. In this context it is notable that duplications of the TATA box as found in the prototype PML strain, JCV(Mad-1) (Frisque *et al*., 1984), are an unusual event.

The average GC content for the 267 bp archetypal regulatory region is 53%. When analysing the relative GC content in successive frames of 10 bp, values varied from 0% in the TATA box to 80% around position 130. Although some of the PML strains were extensively rearranged (Fig. 1), the average GC content remained at $52{\cdot}5\% \pm 1{\cdot}1\%$ with a median of 53%. The lowest GC content (50%) was found in strain #222; the highest (54%) was in strains #110, #118, #210 and #401. The length of the rearranged regulatory region and the nucleotide composition were not correlated. There was also no obvious relation between nucleotide composition and areas with increased numbers of breakpoints.

This study of 29 JCV strains from PML tissue found no association between regulatory region rearrangements and JCV genotypes. This confirms previous observations which were based on 11 different cases (Ault & Stoner, 1993). In both studies the distribution of JCV Type 1 and Type 2 in PML was balanced. However, in urine the proportion of Type 2 strains in the general population was below 20%. In contrast, the

Table 1. JC virus regulatory region configurations

Strain	Regulatory region based on archetypal numbering*
Mad-1	[1–35]G†[60–115]G[183–198][12–35]G[60–115]G[183– …]
#109	[1–185]A[192– …]
#110‡	[1–51][61,62][68–110]AG[40–51][61,62][68–85]A[187– …]
#111	[1–258][205– …]
#112	[1–114][183–199]G[98–114][183– …]
#113§	[1–115][119–121][132–142][90–115][119–121][132–236][255– …]
#114	[1–112]C[180–200]AT[40–112]C[180– …]
#115	[1–242]TGT[256– …]
#116	[1–195]A[34–86][253– …]
#117	[1–112][177–204]GG[51–112][177– …]
#118	[1–118]GGA[62–149]A[186–203]T[147–149]A[186– …]
#119	[1–110]AGC[84,85][153–199]G[50–110]AGC[84,85][153–199]G[204– …]
#120	[1–86][98–119][175–204][52–86][98–119][175– …]
#121	[1–86]TA[175–179][41–86]TA[175– …]
#122	[1–112]CC[160–199][62–112]CC[160– …]
#401	[1–131][176–201][50–131][176– …]
#209	[1–68][72–205][225– …]
#210	[1–114]A[185–204]GGA[43–114]A[185–216][GGAGC][233– …]
#211	[1–91]A[182–202]G[51–91]A[182– …]
#212	[1–190][256– …]
#213	[1–114]AG[152–167]G[51–90][179–205]G[40–114]AG[152–167]G[51–90][181– …]
#214	[1–110]A[182–203]TGGA[62–110]A[182– …]
#215	[1–110][168–196][49–110][168–259][205– …]
#216	[1–109][170–225][39–109][169–200][202– …]
#217	[1–89][102–115‖]GA[43–115][152– …]
#218	[1–113][186–192]ACAA[190–202]G[98–113][186–192]ACAA[190– …]
#219¶	[1–107]GTC[171,172]TA[19–107]GTC[171–245][256– …]
#220	[1–115][172–194][48–52][54–115][172– …]
#221	[1–115][51–115][152–240]G[251– …]
#222	[1–90][189–198][179–201][73–90][189–198][179– …]

* GenBank accession number M35834 (Yogo *et al.*, 1990) with 'A' replacing 'G' at position 217.
† 'G' could be either position 36 or 59.
‡ 'A' replaces 'G' at position 108.
§ 'A' replaces 'G' at position 108, 'T' replaces 'A' at position 257.
‖ 'C' replaces 'A' at position 115 in repeat.
¶ 'T' replaces 'A' at position 219.

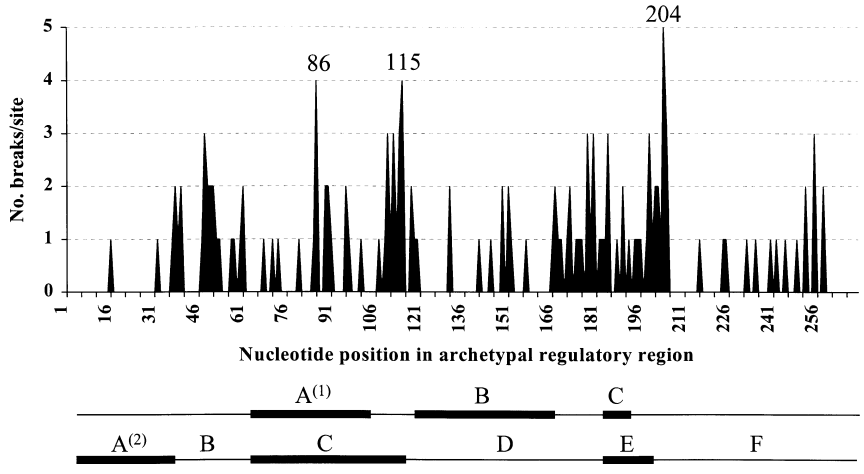

Fig. 2. Breakpoint profile of the archetypal regulatory region. Breakpoints were accounted for only once per sequence at the beginning and end of deletions and duplications. In cases where a nucleotide could be located at either end of a rearranged element, the first possible breakpoint was counted. Rearrangement domains were previously described (1) by Iida *et al.* (1993) and (2) by Ault & Stoner (1993) as indicated.

proportion of Type 1 strains was about 65 %. This difference in the genotype profile is statistically significant and provides evidence for a genotype-specific behaviour (Stoner *et al.*, 1996). As there is no association between genotypes and regulatory region rearrangements, the latter cannot be the only genomic factor determining virus pathogenicity.

When combining all deleted areas of these rearranged regulatory regions, the following portions of the archetypal regulatory region were present in all strains: [1–51][61,62][72–85][256– …]. These include the putative TATA box with the Tst-1 binding site (Krebs *et al.*, 1995). The first NF-1 site (archetypal position 62–74) was interrupted by a deletion in only two strains (#110, #209) (Table 1). Both sequences restore a pseudo-NF-1 site (5′ TGG … CCA 3′) within the borders of the deletion. This emphasizes the important role of the first NF-1 binding site for virus function.

In the prototype JCV(Mad-1) (Table 1) the Sp-1 site is interrupted by a unique 23 bp deletion, and the edges of this deletion create a duplicated pentamer (5′ AGGGAAGGGA 3′). This sequence was found to be involved not only in virus replication but also in activation of the early viral promotor and repression of the late promotor [for review of transcriptional regulation in JCV see Raj & Khalili (1995)]. Since none of the 29 PML strains differed from the archetypal sequence in this Sp-1 site, it would appear that the existence of more than one copy of the AGGGA pentamer is not essential for virus function. An AP-1 binding site is another potentially important *cis*-regulatory element (5′ TGANNTCA 3′, archetypal position 79–86). It was present in all PML strains.

In conclusion, rearrangements of the JCV regulatory region in PML strains are unique. Deletions in the first one-third of the archetypal regulatory region are uncommon, but deletions there can result in restoration of essential elements in the flanking sequence. Duplicated or deleted regions vary widely in length. However, they follow certain patterns with the 'long duplicate' being the most common. Regulatory region rearrangements are not associated with a particular viral genotypes as defined by the coding region.

This work was supported in part by the Deutsche Forschungsgemeinschaft, Bonn (grant Ag 19/1-1). The support and encouragement of Dr Henry deF. Webster is gratefully acknowledged. We thank the National Neurological Research Specimen Bank, VAMC, Los Angeles, CA for CNS specimens.

References

Agostini, H. T. & Stoner, G. L. (1995). Amplification of the complete polyomavirus JC genome from brain, cerebrospinal fluid and urine using pre-PCR restriction enzyme digestion. *Journal of Neurovirology* **1**, 316–320.

Agostini, H. T., Brubaker, G. R., Shao, J., Levin, A., Ryschkewitsch, C. F., Blattner, W. A. & Stoner, G. L. (1995). BK virus and a new type of JC virus excreted by HIV-1 positive patients in rural Tanzania. *Archives of Virology* **140**, 1919–1934.

Agostini, H. T., Ryschkewitsch, C. F. & Stoner, G. L. (1996*a*). Genotype profile of human polyomavirus JC excreted in urine of immunocompetent individuals. *Journal of Clinical Microbiology* **34**, 159–164.

Agostini, H. T., Ryschkewitsch, C. F., Singer, E. J. & Stoner, G. L. (1996*b*). Co-infection with two JC virus genotypes in brain, cerebrospinal fluid or urinary tract detected by direct cycle sequencing of PCR products. *Journal of Neurovirology* **2**, 259–267.

Ault, G. S. & Stoner, G. L. (1992). Two major types of JC virus defined in progressive multifocal leukoencephalopathy brain by early and late coding region DNA sequences. *Journal of General Virology* **73**, 2669–2678.

Ault, G. S. & Stoner, G. L. (1993). Human polyomavirus JC promoter/enhancer rearrangement patterns from progressive multifocal leukoencephalopathy brain are unique derivatives of a single archetypal structure. *Journal of General Virology* **74**, 1499–1507.

Berger, J. R. & Concha, M. (1995). Progressive multifocal leukoencephalopathy: the evolution of a disease once considered rare. *Journal of Neurovirology* **1**, 5–18.

Frisque, R. J., Bream, G. L. & Cannella, M. T. (1984). Human polyomavirus JC virus genome. *Journal of Virology* **51**, 458–469.

Iida, T., Kitamura, T., Guo, J., Taguchi, F., Aso, Y., Nagashima, K. & Yogo, Y. (1993). Origin of JC polyomavirus variants associated with progressive multifocal leukoencephalopathy. *Proceedings of the National Academy of Sciences, USA* **90**, 5062–5065.

Krebs, C. J., McAvoy, M. T. & Kumar, G. (1995). The JC virus minimal core promoter is glial cell specific in vivo. *Journal of Virology* **69**, 2434–2442.

Raj, G. V. & Khalili, K. (1995). Transcriptional regulation – lessons from the human neurotropic polyomavirus, JCV. *Virology* **213**, 283–291.

Stoner, G. L. & Ryschkewitsch, C. F. (1995). Capsid protein VP1 deletions in JC virus from two AIDS patients with progressive multifocal leukoencephalopathy. *Journal of Neurovirology* **1**, 189–194.

Stoner, G. L., Agostini, H. T. & Ryschkewitsch, C. F. (1996). A comparison of JC virus genotype profiles in PML brains and control urines. *Neurology* **46**, A316.

Walker, D. L. & Frisque, R. J. (1986). The biology and molecular biology of JC virus. In *The Papovaviridae*, vol. 1, *The Polyomaviruses*, pp. 327–377. Edited by N. P. Salzman. New York: Plenum Press.

Walker, D. L. & Padgett, B. L. (1983). The epidemiology of human polyomaviruses. *Progress in Clinical and Biological Research* **105**, 99–106.

Yogo, Y., Kitamura, T., Sugimoto, C., Ueki, T., Aso, Y., Hara, K. & Taguchi, F. (1990). Isolation of a possible archetypal JC virus DNA sequence from nonimmunocompromised individuals. *Journal of Virology* **64**, 3139–3143.

Received 3 September 1996; Accepted 28 October 1996