J Mol Evol (1987) 26:148–156

Journal of Molecular Evolution

Molecular Evolutionary Rates of Oncogenes

Takashi Gojobori[1] and Shozo Yokoyama[2]

[1] National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan
[2] Department of Ecology, Ethology and Evolution, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Summary. Using nine sets of viral and cellular oncogenes, the rates of nucleotide substitutions were computed by using Gojobori and Yokoyama's (1985) method. The results obtained confirmed our previous conclusion that the rates of nucleotide substitution for the viral oncogenes are about a million times higher than those for their cellular counterparts. For cellular oncogenes and most viral oncogenes, however, the rate of synonymous substitution is higher than that of nonsynonymous substitution. Moreover, the pattern of nucleotide substitutions for viral oncogenes is more similar to that for functional genes (such as cellular oncongenes) than for pseudogenes. This implies that nucleotide substitutions in viral oncogenes may be functionally constrained. Thus, our observation supports that nucleotide substitutions for the oncogenes in those DNA and RNA genomes are consistent with Kimura's neutral theory of molecular evolution (Kimura 1968, 1983).

Key words: Oncogenes — Retroviruses — Evolutionary rates — Nucleotide substitution — Substitution pattern

Introduction

RNA viruses show high mutation rates. Because of this high mutation rate and high rates of replication, the rates of RNA genome evolution can be more than a millionfold greater than the rates of the DNA genome of their hosts (Holland et al. 1982). The hemagglutinin genes (Air 1981) and neuraminidase gene (Krystal et al. 1983; Martinez et al. 1983) in the influenza A virus seem to show the typical rate of RNA genome evolution (Hayashida et al. 1985; Buonagurio et al. 1986; Saitou and Nei 1986).

For conducting detailed quantitative analyses of evolutionary dynamics of the RNA genomes, oncogenes in RNA tumor viruses (retroviruses) and their host genomes are excellent material. Oncogenes are genes whose activities are responsible for both the initiation and the maintenance of neoplastic transformation in cells of organisms (for a review see Bishop 1983). Since viral oncogenes in retroviruses originated from cellular homologs in DNA genomes (Bishop and Varmus 1982), the evolutionary rates of RNA and DNA genomes can be evaluated directly by comparing cellular and viral oncogenes.

Recently, we developed a method for computing the rates of nucleotide substitution for cellular oncogenes, viral oncogenes, and the retrovirus genome simultaneously (Gojobori and Yokoyama 1985). The method was applied to the nucleotide sequences of a viral oncogene, v-*mos*, in Moloney murine sarcoma virus (Mo-MuSV), cellular oncogenes, c-*mos*, in mouse and human, and the *gag* genes of Mo-MuSV and its helper virus, Moloney murine leukemia virus (Mo-MuLV). We have then shown that the rates of nucleotide substitution for v-*mos* and the *gag* genes are approximately 10^{-3}/site/year, whereas those of c-*mos* in mouse and human are about 10^{-9}/site/year.

At present, out of about 20 known oncogenes, nine sets of viral and cellular oncogenes are available for similar comparisons. In this paper, we apply Gojobori and Yokoyama's method to nine data sets and conduct evolutionary studies of nucleotide substitutions in both viral and cellular oncogenes. The analyses show that the rates of nucleotide substi-

Offprint requests to: T. Gojobori

tution for the viral oncogenes are about a million times higher than those for their cellular counterparts, confirming and extending our previous conclusion. Furthermore, for cellular oncogenes and most viral oncogenes, the rate of synonymous (silent) substitution is higher than that of nonsynonymous (amino acid-altering) substitution. Since these results can be explained by rates of mutation and functional constraints (i.e., negative or purifying selection) against amino acid changes, the evolutionary nucleotide substitutions in those DNA and RNA genomes are compatible with Kimura's neutral (random drift) theory of molecular evolution (Kimura 1968, 1983).

Table 1. Abbreviations for different virus strains and variants used in this study

Abbreviation	Full name
Ab-MuLV	Abelson murine leukemia virus
AMV-BAI/A	Avian myeloblastosis virus strain BAI/A
AMV-E26	Avian myeloblastosis virus strain E26
(AEV-E26)	(Avian erythroblastosis virus strain E26)
FBJ-MuLV	Finkel-Biskis-Jinkins murine leukemia virus
FBJ-MuSV	Finkel-Biskis-Jinkins murine osteosarcoma virus
FBR-MuLV	Finkel-Biskis-Reilly murine leukemia virus
FBR-MuSV	Finkel-Biskis-Reilly murine osteosarcoma virus
LLV	Nondefective lymphoid leukosis virus
MAV	Myeloblastosis-associated virus
MC 29	Avian myelocytomatosis virus
Mo-MuLV	Moloney murine leukemia virus
Mo-MuSV	Moloney murine sarcoma virus
REV-A	Reticuloendotheliosis virus strain A
REV-T	Reticuloendotheliosis virus strain T
rASV1441	Recovered avian sarcoma virus strain 1441
RSV	Rous sarcoma virus
SR-RSV	Schmidt-Ruppin strain of RSV
SR-RSV-A *td* NY108	Schmidt-Ruppin strain of Rous sarcoma virus strain A transformation-defective mutant (*td*) NY108

Nucleotide Sequences Used

In surveying the literature, we have restricted our attention to viral oncogenes and their cellular counterparts whose nucleotide sequence data are both available. Furthermore, we consider only viral oncogenes whose transduction times are known. This is essential, because the divergence time between the viral and cellular oncogenes is needed to compute the rates of nucleotide substitution for the viral oncogenes, as we will see in the next section. If a viral oncogene has been molecularly characterized, then the time of cloning is available. Thus, the divergence time is estimated by the time between the transduction and the cloning of a viral oncogene. Note that this estimation may give the maximum value for some viral oncogenes, especially if they have been frozen in storage for some time. The different viral strains and variants used in this study are listed in Table 1.

When the viral oncogenes have been transduced experimentally, the time of virus isolation can be equated to the transduction time. Experimentally transduced genes include v-*abl*, two v-*fos*, v-*mos*, v-*rel*, and v-*src* (Table 2). Among these, the v-*src* needs special mention. It is known that v-*src* is contained in a transformation-competent Rous sarcoma virus (RSV) in its host, the chicken. Note that the v-*src* in Table 2 differs from the v-*src* in the

Table 2. The estimated divergence period of viral oncogenes since the time of their transduction into retroviral genomes

Viral oncogene	Carrying virus strain	Helper virus	Host species	Time of transduction	Time of cloning	Divergence period (yr)	References for base sequences
v-*abl*	Ab-MuLV	Mo-MuLV	Mouse	1970 [1][a]	1981 [2]	12	[3, 4, 5]
v-*fos* (FBJ)	FBJ-MuSV	FBJ-MuLV	Mouse	1966 [6]	1982 [7]	17	[8]
v-*fos* (FBR)	FBR-MuSV	FBR-MuLV	Mouse	1973 [9, 10, 11]	1984 [12]	12	[13]
v-*mos*	Mo-MuSV	Mo-MuLV	Mouse	1966 [14]	1981 [15, 16]	16	[15, 16, 17]
v-*myb*	AMV-BAI/A	MAV-2(0)	Chicken	1941 [18, 19]	1982 [20]	42	[20, 21]
v-*myb*-E	AMV-E26[b]	MAV	Chicken	1962 [18]	1984 [19]	23	[21, 19]
v-*myc*	MC 29	LLV	Chicken	1964 [22, 23]	1981 [24]	18	[25]
v-*rel*	Rev-T	Rev-A	Turkey	1966 [26]	1981 [27]	16	[28]
v-*src*	rASV1441	SR-RSV-A *td* NY108	Chicken	1977 [29]	1981 [30]	5	[30, 31, 32]

[a] Data source: [1] Abelson and Rabstein (1970), [2] Srinivasan et al. (1981), [3] Reddy et al. (1983), [4] Wang et al. (1984), [5] Ben-Neriah et al. (1986), [6] Finkel et al. (1966), [7] Curran et al. (1982), [8] Van Beveren et al. (1983), [9] Finkel et al. (1973), [10] Finkel et al. (1975), [11] Lee et al. (1979), [12] Curran and Verma (1984), [13] Van Beveren et al. (1984), [14] Moloney (1966), [15] Van Beveren et al. (1981b), [16] Reddy et al. (1981), [17] Van Beveren et al. (1981a), [18] Teich et al. (1982), [19] Nunn et al. (1984), [20] Klempnauer et al. (1982), [21] Rosson and Reddy (1986), [22] Ivanov et al. (1964), [23] Mladenov et al. (1967), [24] Lautenberger et al. (1981), [25] Watson et al. (1983b), [26] Theilen et al. (1966), [27] Chen et al. (1981), [28] Wilhelmsen et al. (1984), [29] Kawai et al. (1977), [30] Takeya et al. (1981), [31] Takeya and Hanafusa (1982), [32] Takeya and Hanafusa (1983)

[b] It is also called AEV-E26, since it causes erythroblastosis as well as myeloblastosis in chicken

Table 3. The sequence comparisons of the cellular oncogenes between the two different species

Cellular oncogenes	Species comparison used in the study	Divergence time (Myr)[a]	References for base sequences[b]
c-*abl*	Human vs mouse	80	[1, 2]
c-*fos*	Human vs mouse	80	[3, 4]
c-*mos*	Human vs mouse	80	[5, 6]
c-*myb*	Mouse vs chicken	270	[7–9]
c-*myc*	Human vs mouse	80	[10, 11]
c-*src*	Chicken vs *Drosophila*	600	[12, 13]

[a] Myr: million years

[b] Data source: [1] Shtivelman et al. (1985), [2] Ben-Neriah et al. (1986), [3] Van Straaten et al. (1983), [4] Van Beveren et al. (1983), [5] Watson et al. (1982), [6] Van Beveren et al. (1981a), [7] Bender and Kuehl (1986), [8] Klempnauer et al. (1982), [9] Rosson and Reddy (1986), [10] Watson et al. (1983a), [11] Stanton et al. (1983), [12] Takeya and Hanafusa (1983), [13] Simon et al. (1985)

RSV. Deletions of part or all of the v-*src* in the RSV lead to transformation-defective (*td*) mutants (Kawai et al. 1977). Transformation-competent viruses have been recovered from host cells after transfection of these *td* mutants. They are called recovered avian sarcoma viruses (rASVs) (Halpern et al. 1979). These rASVs contain v-*src*s that are newly transduced from c-*src* of the chicken genome.

The oncogenes v-*myb*, v-*myb*-E, and v-*myc* are transduced naturally (Table 2). Once these oncogenes are transduced into the viral genome, the host organism will develop an acute neoplastic tumor. Thus, the virus strain becomes extinct as soon as the host organism dies out, unless the virus is isolated. For this reason, even if they are naturally occurring virus strains, we may be able to identify the transduction time as the isolation time of the virus strains.

Among the nine viral oncogenes, v-*abl*, v-*mos*, v-*myc*, v-*rel*, and v-*src* can be compared to the entire cellular counterpart in the host species (see Table 2). However, the viral oncogenes v-*fos* (FBJ) and v-*fos* (FBR) are comparable to the different but overlapping regions of the mouse c-*fos*. Similarly, the oncogenes v-*myb* and v-*myb*-E are comparable to the different regions of the chicken c-*myb*. (See Tables 2 and 3 for details and references to the sequences used.) In these comparisons, we assumed that even if the viral oncogene contains an allelic sequence of the cellular counterpart, it does not affect the measurement of nucleotide changes for the viral oncogene very much. This assumption seems reasonable because of the low level of DNA polymorphism of the functional nuclear genes such as the cellular oncogenes (Nei et al. 1984).

Six out of the seven cellular oncogenes have been characterized in two different species (Table 3). The nucleotide sequences of the cellular oncogenes c-*abl*, c-*fos*, c-*mos*, and c-*myc* have been obtained from human and mouse. The nucleotide sequences of the c-*myb* are obtained from both mouse and chicken, whereas those of the c-*src* are obtained from chicken and *Drosophila*. The oncogene c-*rel* has been characterized only in the host organism, i.e., turkey.

In the present study, only the coding regions were examined, because most viral genes lack intron sequences completely, due to the fact that reverse transcription occurs primarily by using mature mRNA as a template. We also excluded deleted and inserted nucleotides from the sequence comparisons, since we are mainly interested in the evolutionary implications of nucleotide substitutions.

Results and Discussion

Rates of Nucleotide Substitution at Each Codon Position

The number (d) of nucleotide substitutions per site between a viral oncogene sequence and its cellular counterpart was estimated by the formula for the six-parameter method (Gojobori et al. 1982a) based on Kimura's 2FC substitution scheme (Kimura 1981). Since the divergence period (t) of a viral oncogene is at most several decades for the genes studied in this paper (Table 2), the probability of having a nucleotide substitution in the cellular oncogene during such a short period of time should be negligible. Thus, it is reasonable to assume that nucleotide differences between the viral and cellular oncogenes are attributable to nucleotide changes in the viral oncogene and not in the cellular counterpart. Then, the rate (v) of nucleotide substitutions per site per year for the viral oncogenes can be computed by d/t.

With the exception of c-*rel*, the number (D) of nucleotide substitutions for the cellular oncogenes was also estimated by the six-parameter method, using comparable nucleotide sequences for the cellular oncogenes between two appropriate species (see Table 3). The rate (V) of nucleotide substitutions for the cellular oncogenes is given by $V = D/(2T)$, where T represents the divergence time between the two species compared (Table 3).

The rates of nucleotide substitution at each of three codon positions as well as at all positions for the nine viral oncogenes are shown in Table 4. The average rates over the nine viral oncogenes at the first, second, and third codon positions are 0.42×10^{-3}, 0.64×10^{-3}, and 0.79×10^{-3}/site/year, respectively. For all positions, the average rate is 0.62×10^{-3}. Thus, the rates of nucleotide substitution for the viral oncogenes are on the order of 10^{-4} or 10^{-3}. The rates of nucleotide substitution for the cellular oncogenes are given in Table 5. The average rates of nucleotide substitution over the six

Table 4. The rates of nucleotide substitution per year for viral oncogenes

Viral oncogene	Proportion of different nucleotides at codon position				Rates of nucleotide substitution at codon position			
	All	First	Second	Third	All	First	Second	Third
v-*abl*	1/204	0/68	1/68	0/68	0.40×10^{-3} ($\pm 0.41 \times 10^{-3}$)	0.0×10^{-3} ($\pm 0.0 \times 10^{-3}$)	1.02×10^{-3} ($\pm 0.44 \times 10^{-3}$)	0.0×10^{-3} ($\pm 0.0 \times 10^{-3}$)
v-*fos* (FBJ)	11/995	2/332	3/332	6/331	0.66×10^{-3} ($\pm 0.21 \times 10^{-3}$)	0.31×10^{-3} ($\pm 0.33 \times 10^{-3}$)	0.53×10^{-3} ($\pm 0.51 \times 10^{-3}$)	1.17×10^{-3} ($\pm 0.50 \times 10^{-3}$)
v-*fos* (FBR)	9/709	1/237	4/236	4/236	1.05×10^{-3} ($\pm 0.37 \times 10^{-3}$)	0.30×10^{-3} ($\pm 0.52 \times 10^{-3}$)	1.37×10^{-3} ($\pm 0.92 \times 10^{-3}$)	1.48×10^{-3} ($\pm 0.75 \times 10^{-3}$)
v-*mos*	21/1026	7/342	3/342	11/342	1.31×10^{-3} ($\pm 0.29 \times 10^{-3}$)	1.26×10^{-3} ($\pm 0.52 \times 10^{-3}$)	0.56×10^{-3} ($\pm 0.34 \times 10^{-3}$)	2.08×10^{-3} ($\pm 0.06 \times 10^{-3}$)
v-*myb*	9/762	2/254	4/254	3/254	0.29×10^{-3} ($\pm 0.12 \times 10^{-3}$)	0.17×10^{-3} ($\pm 0.18 \times 10^{-3}$)	0.36×10^{-3} ($\pm 0.29 \times 10^{-3}$)	0.29×10^{-3} ($\pm 0.19 \times 10^{-3}$)
v-*myb*-E	1/848	0/283	1/283	0/283	0.06×10^{-3} ($\pm 0.14 \times 10^{-3}$)	0.0×10^{-3} ($\pm 0.0 \times 10^{-3}$)	0.20×10^{-3} ($\pm 0.40 \times 10^{-3}$)	0.0×10^{-3} ($\pm 0.0 \times 10^{-3}$)
v-*myc*	8/1263	2/421	4/421	2/421	0.33×10^{-3} ($\pm 0.17 \times 10^{-3}$)	0.26×10^{-3} ($\pm 0.28 \times 10^{-3}$)	0.55×10^{-3} ($\pm 0.46 \times 10^{-3}$)	0.24×10^{-3} ($\pm 0.29 \times 10^{-3}$)
v-*rel*	21/1422	6/474	9/474	6/474	0.92×10^{-3} ($\pm 0.22 \times 10^{-3}$)	0.79×10^{-3} ($\pm 0.44 \times 10^{-3}$)	1.16×10^{-3} ($\pm 0.44 \times 10^{-3}$)	0.79×10^{-3} ($\pm 0.34 \times 10^{-3}$)
v-*src*	3/1092	1/364	0/364	2/364	0.57×10^{-3} ($\pm 0.05 \times 10^{-3}$)	0.65×10^{-3} ($\pm 1.11 \times 10^{-3}$)	0.0×10^{-3} ($\pm 0.0 \times 10^{-3}$)	1.09×10^{-3} ($\pm 0.90 \times 10^{-3}$)
Average					0.62×10^{-3} ($\pm 0.24 \times 10^{-3}$)	0.42×10^{-3} ($\pm 0.49 \times 10^{-3}$)	0.64×10^{-3} ($\pm 0.66 \times 10^{-3}$)	0.79×10^{-3} ($\pm 0.46 \times 10^{-3}$)

See Gojobori et al. (1982a) for the calculation of standard errors

Table 5. The rates of nucleotides substitution per year for cellular oncogenes

Cellular oncogene	Proportion of different nucleotides at codon position				Rates of nucleotide substitution at codon position			
	All	First	Second	Third	All	First	Second	Third
c-*abl* (human vs mouse)	26/342	3/114	0/114	23/114	0.52×10^{-9} ($\pm 0.11 \times 10^{-9}$)	0.20×10^{-9} ($\pm 0.13 \times 10^{-9}$)	0.0×10^{-9} ($\pm 0.0 \times 10^{-9}$)	1.46×10^{-9} ($\pm 0.33 \times 10^{-9}$)
c-*fos* (human vs mouse)	120/995	25/332	6/332	89/331	0.83×10^{-9} ($\pm 0.79 \times 10^{-9}$)	0.49×10^{-9} ($\pm 0.10 \times 10^{-9}$)	0.11×10^{-9} ($\pm 0.06 \times 10^{-9}$)	2.24×10^{-9} ($\pm 0.28 \times 10^{-9}$)
c-*mos* (human vs mouse)	232/1017	58/339	45/339	129/339	1.71×10^{-9} ($\pm 0.12 \times 10^{-9}$)	1.22×10^{-9} ($\pm 0.17 \times 10^{-9}$)	0.93×10^{-9} ($\pm 0.14 \times 10^{-9}$)	3.61×10^{-9} ($\pm 0.42 \times 10^{-9}$)
c-*myb* (mouse vs chicken)	193/747	51/249	32/249	110/249	0.59×10^{-9} ($\pm 0.04 \times 10^{-9}$)	0.48×10^{-9} ($\pm 0.07 \times 10^{-9}$)	0.27×10^{-9} ($\pm 0.05 \times 10^{-9}$)	1.27×10^{-9} ($\pm 0.16 \times 10^{-9}$)
c-*myc* (human vs mouse)	89/765	15/255	14/256	60/254	0.80×10^{-9} ($\pm 0.09 \times 10^{-9}$)	0.38×10^{-9} ($\pm 0.10 \times 10^{-9}$)	0.37×10^{-9} ($\pm 0.11 \times 10^{-9}$)	2.13×10^{-9} ($\pm 0.38 \times 10^{-9}$)
c-*src* (chicken vs *Drosophila*)	517/1092	162/364	147/364	208/364	0.64×10^{-9} ($\pm 0.04 \times 10^{-9}$)	0.59×10^{-9} ($\pm 0.06 \times 10^{-9}$)	0.51×10^{-9} ($\pm 0.05 \times 10^{-9}$)	1.21×10^{-9} ($\pm 0.31 \times 10^{-9}$)
Average					0.85×10^{-9} ($\pm 0.08 \times 10^{-9}$)	0.55×10^{-9} ($\pm 0.11 \times 10^{-9}$)	0.36×10^{-9} ($\pm 0.08 \times 10^{-9}$)	1.99×10^{-9} ($\pm 0.32 \times 10^{-9}$)

cellular oncogenes at the first, second, and third codon positions are 0.55×10^{-9}, 0.36×10^{-9}, and 1.99×10^{-9}/site/year, respectively, giving an average rate of 0.85×10^{-9} for all positions. These results confirm our previous observation that viral oncogenes evolve at a rate about a million times higher than cellular oncogenes (Gojobori and Yokoyama 1985).

It should be pointed out that for cellular oncogenes, the rate of nucleotide substitution at the third codon position is always highest followed by the first and second codon positions, in that order. This suggests that functional constraints against amino acid changes are operating on the cellular oncogenes, since synonymous substitution can occur more frequently at the third position than at the first two positions (e.g., see Kimura 1983). The viral oncogenes show a somewhat different pattern of nucleotide substitution, though differences may not be significant. The five oncogenes v-*abl*, v-*myb*, v-*myb*-E, v-*myc*, and v-*rel* have the highest rates of nucleotide substitution at the second codon position, whereas the

Table 6. The rates of synonymous and nonsynonymous substitutions for cellular oncogenes[a]

Cellular oncogene	Synonymous or non-synonymous	Proportion of different nucleotides	Rate of nucleotide substitution (per site per yr)
c-*abl* (human vs mouse)			
	SYN	26.0/76.7	$(0.46 \pm 0.10) \times 10^{-9}$
	NON	0.0/266.7	$(0.0 \pm 0.0) \times 10^{-9}$
c-*fos* (human vs mouse)			
	SYN	96.0/242.3	$(3.52 \pm 0.42) \times 10^{-9}$
	NON	24.0/750.7	$(0.20 \pm 0.04) \times 10^{-9}$
c-*mos* (human vs mouse)			
	SYN	129.4/256.7	$(5.23 \pm 0.60) \times 10^{-9}$
	NON	102.6/760.3	$(0.93 \pm 0.09) \times 10^{-9}$
c-*myb* (mouse vs chicken)			
	SYN	113.5/170.3	$(3.05 \pm 0.60) \times 10^{-9}$
	NON	79.5/576.7	$(0.28 \pm 0.03) \times 10^{-9}$
c-*myc* (human vs mouse)			
	SYN	60.5/174.0	$(2.92 \pm 0.42) \times 10^{-9}$
	NON	28.5/585.0	$(0.31 \pm 0.06) \times 10^{-9}$
c-*src* (chicken vs *Drosophila*)			
	SYN	169.5/254.7	$(1.37 \pm 0.22) \times 10^{-9}$
	NON	347.5/837.3	$(0.50 \pm 0.03) \times 10^{-9}$
Average	SYN (total)		$(3.16 \pm 0.50) \times 10^{-9}$
	NON (total)		$(0.37 \pm 0.05) \times 10^{-9}$
	NON (excluding c-*abl*)[b]		$(0.45 \pm 0.06) \times 10^{-9}$

[a] The sequence comparison for c-*rel* is not available

[b] Since the number of nonsynonymous substitutions for c-*abl* is 0, the average was also computed excluding c-*abl*

Table 7. The rates of synonymous and nonsynonymous substitutions for viral oncogenes

Viral oncogene	Synonymous or non-synonymous	Proportion of different nucleotides	Rate of nucleotide substitution (per site per yr)
v-*abl*	SYN	0/47.3	$(0.0 \pm 0.0) \times 10^{-3}$
	NON	1/156.7	$(0.53 \pm 0.53) \times 10^{-3}$
v-*fos* (FBJ)	SYN	6/241.5	$(1.49 \pm 0.61) \times 10^{-3}$
	NON	5/751.5	$(0.39 \pm 0.18) \times 10^{-3}$
v-*fos* (FBR)	SYN	4/172.5	$(1.96 \pm 0.99) \times 10^{-3}$
	NON	5/535.5	$(0.78 \pm 0.35) \times 10^{-3}$
v-*mos*	SYN	11/257.2	$(2.75 \pm 0.84) \times 10^{-3}$
	NON	10/768.8	$(0.82 \pm 0.26) \times 10^{-3}$
v-*myb*	SYN	3/173.8	$(0.42 \pm 0.24) \times 10^{-3}$
	NON	6/588.2	$(0.24 \pm 0.10) \times 10^{-3}$
v-*myb*-E	SYN	0/186.3	$(0.0 \pm 0.0) \times 10^{-3}$
	NON	1/659.7	$(0.07 \pm 0.07) \times 10^{-3}$
v-*myc*	SYN	1/296.8	$(0.19 \pm 0.19) \times 10^{-3}$
	NON	7/966.2	$(0.40 \pm 0.15) \times 10^{-3}$
v-*rel*	SYN	7/311.2	$(1.43 \pm 0.54) \times 10^{-3}$
	NON	14/1107.8	$(0.80 \pm 0.21) \times 10^{-3}$
v-*src*	SYN	2/261.3	$(1.54 \pm 1.09) \times 10^{-3}$
	NON	1/830.7	$(0.24 \pm 0.24) \times 10^{-3}$
Average	SYN (total)		$(1.09 \pm 0.63) \times 10^{-3}$
	SYN (excluding v-*myb*-E and v-*abl*)[a]		$(1.40 \pm 0.72) \times 10^{-3}$
	NON (total)		$(0.48 \pm 0.27) \times 10^{-3}$

[a] Since the numbers of synonymous substitutions for v-*myb*-E and v-*abl* are 0, the average was also computed excluding these two genes

four oncogenes v-*fos* (FBJ), v-*fos* (FBR), v-*mos*, and v-*src* have the highest rates at the third codon position. Furthermore, for the viral oncogenes v-*fos* (FBJ) and v-*fos* (FBR), the rate of nucleotide substitution is higher at the second codon position than at the first codon position. Thus, only v-*mos* and v-*src* have the same substitution pattern as the cellular oncogenes. Since every nucleotide substitution at the second position changes the amino acid, it is possible either that there is no functional constraint against amino acid changes or that positive selection for amino acid changes is operating for many viral oncogenes. The best way to examine this point is to evaluate the rates of synonymous and nonsynonymous substitution directly from the comparisons of the nucleotide sequences.

Rates of Synonymous and Nonsynonymous Substitutions

To examine the degree of functional constraints on viral and cellular oncogenes, we computed the rates of synonymous and nonsynonymous substitutions. The numbers of synonymous and nonsynonymous substitutions were estimated by using both Nei and Gojobori's (NG) method (1986) and Miyata and Yasunaga's (MY) method (1980). The former method gives equal weights for intermediate evolutionary pathways between a given pair of codons compared, whereas the latter method gives unequal weights. The two methods yield similar estimates and, therefore, the results obtained by the NG method for the cellular oncogenes and viral oncogenes are given in Tables 6 and 7.

From Table 6, it is clear that for all of the cellular oncogenes the rate of synonymous substitution is much higher than that of nonsynonymous substitution. For the viral oncogenes, as we saw in the previous section, only two, v-*mos* and v-*src*, have the same pattern as the cellular oncogenes with respect to nucleotide substitution. However, Table 7 shows that most of the viral oncogenes have higher rates of synonymous substitution, with the exception of v-*abl*, v-*myb*-E, and v-*myc*, in which nonsynonymous substitution is higher than synonymous substitution. Thus, most of the viral oncogenes and all cellular oncogenes examined seem to have constraints against amino acid changes. These evolutionary features of oncogenes are consistent with the neutral theory of molecular evolution (Kimura 1968, 1983), which maintains that nonsynonymous substitutions are more functionally constrained than synonymous substitutions.

For v-*abl* and v-*myb*-E, the high rate of nonsynonymous substitution may be easily explained by stochastic errors, since only one nucleotide substitution occurred in these viral oncogenes. For v-*myc*,

however, the rate of nonsynonymous substitution is approximately two times as high as that of synonymous substitution. Although the high rate of nonsynonymous substitution for v-*myc* can still be explained by stochastic errors, it is possible that for v-*myc*, some type of selection for amino acid changes may be operating.

The degree of functional constraints seems to differ between viral and cellular oncogenes. For the viral oncogenes the rate of synonymous substitution is, on the average, only about two or three times as high as that of nonsynonymous substitution, whereas for the cellular oncogenes the former is about 10 times as high as the latter. In general, the stronger the functional constraint is, the larger is the difference in the rates between the synonymous and nonsynonymous substitutions. Thus, functional constraints appear to be much stronger in cellular oncogenes than in viral oncogenes.

Although the biological function of cellular oncogenes is not well understood, it has been speculated that they may play a vital role in regulating cell growth (Bishop 1983). If an amino acid change occurs in the gene product of a cellular oncogene, it may impair a regulatory function for cell growth. If this is the case, the organism could not stand such malfunctioning of the cellular system and the mutation would be quickly eliminated from a population by selection. Thus, it is easily understandable that the cellular oncogenes have stronger constraints against amino acid changes. On the other hand, even defective viruses can survive and replicate themselves to some extent with the aid of a helper virus. Thus, it is expected that a viral oncogene may be somewhat exempt from strong functional constraints. However, most of the viral oncogenes still appear to have some functional constraints, as already shown. This implies that the viral oncogene might have some useful function for the replication and survival of the retrovirus.

Pattern of Nucleotide Substitution

We have shown that the rates of nucleotide substitution in viral oncogenes are roughly a million times as high as those in cellular oncogenes. Since the transduction times of the viral oncogenes studied are very recent, a nucleotide difference between the viral oncogene and the cellular counterpart can be attributed to a substitution in the viral oncogene, as already noted. Thus, we can identify the direction of every single nucleotide substitution, comparing the nucleotide sequence of a viral oncogene with that of its cellular homolog. For example, the v-*rel* sequence has T at the third position of the 22nd codon whereas its cellular counterpart, c-*rel*, has C. In this case, a nucleotide substitution must have occurred from C to T rather than from T to C. Using this logic, we determined the direction of nucleotide substitutions for all nine sets of viral and cellular oncogenes.

Table 8 lists the proportion of nucleotide substitutions from the ith nucleotide to the jth nucleotide (i,j = A, T, C, and G) at all positions for each set of oncogenes. This proportion can be given by dividing the number (n_{ij}) of nucleotide substitutions from the ith to the jth by the number (m_i) of the ith nucleotide contained in the cellular oncogene, i.e., n_{ij}/m_i. For instance, the proportion of substitution A → T at all positions for v-*fos* (FBR) is 2/168 (Table 8), because $n_{AT} = 2$ and $m_A = 168$. The average of these values (P_{ij}) over all viral oncogenes can be computed by dividing the sum of n_{ij} by the sum of m_i for all oncogenes. For example, P_{AT} at all positions is 3/2078 (Table 8). Using the P_{ij} values, we calculated *the relative substitution frequency* using the equation:

$$f_{ij} = \frac{P_{ij}}{\sum_i \sum_{j \neq i} P_{ij}} \times 100\%$$

(Gojobori et al. 1982b). Note that $f_{ij} \times 100$ represents the expected number of nucleotide changes from the ith type to the jth among every 100 substitutions in a random sequence, i.e., in a sequence in which the four kinds of nucleotides are equally frequent. Table 9 shows these f_{ij} values in matrix form for all positions and for the first two codon positions.

In Table 9, we can see two interesting characteristics in the substitution pattern of the viral oncogenes. First, transitions (G → A, C → T, T → C, and A → G) occur much more frequently than transversions (all other nucleotide changes). Indeed, the transitions totaled 74.3% at all positions; cumulative f_{ij} values for transitions are given in brackets in the far right column of each matrix (Table 9). Second, the substitution G → A, which is a transition type, is the most frequent of all. In fact, the frequency of G → A is more than two times higher than that of the second most frequent change, T → C. It is known that for functional genes in the nuclear genome, the change G → A occurs much more frequently than C → T, whereas for pseudogenes these changes occur almost equally (Gojobori et al. 1982b; Li et al. 1984). Thus, the pattern of nucleotide substitutions for viral oncogenes is more similar to that for functional genes than for pseudogenes. This again implies that nucleotide substitutions in viral oncogenes may be functionally constrained.

Functional Constraints and Substitution Patterns

We suggested that functional constraint against amino acid changes may be the most plausible expla-

Table 8. The proportions of nucleotide substitutions at all codon positions and the first two codon positions (in parentheses) for viral oncogenes

Viral oncogene	Cellular oncogene	A → T	A → C	A → G	T → A	T → C	T → G	C → A	C → T
v-*abl*	c-*abl*	0/47 (0/41)	0/47 (0/41)	0/47 (0/41)	0/44 (0/28)	0/44 (0/28)	0/44 (0/28)	0/54 (0/27)	0/54 (0/27)
v-*fos* (FBJ)	c-*fos*	0/218 (0/168)	0/218 (0/168)	0/218 (0/168)	0/199 (0/140)	1/199 (1/140)	0/199 (0/140)	0/306 (0/190)	1/306 (1/190)
v-*fos* (FBR)	c-*fos*	2/168 (1/127)	0/168 (0/127)	1/168 (1/127)	1/140 (1/96)	0/140 (0/96)	0/140 (0/96)	1/211 (1/134)	0/211 (0/134)
v-*mos*	c-*mos*	0/221 (0/169)	1/221 (1/169)	4/221 (3/169)	1/213 (0/154)	1/213 (0/154)	1/213 (1/154)	0/307 (0/188)	3/307 (0/188)
v-*myb*	c-*myb*	0/209 (0/157)	1/209 (1/157)	1/209 (0/157)	0/153 (0/94)	2/153 (2/94)	0/153 (0/94)	0/229 (0/148)	1/229 (0/148)
v-*myb*-E	c-*myb*	0/259 (0/191)	0/259 (0/191)	0/259 (0/191)	0/153 (0/91)	0/153 (0/91)	0/153 (0/91)	0/220 (0/148)	0/220 (0/148)
v-*myc*	c-*myc*	0/285 (0/237)	3/285 (2/237)	1/285 (0/237)	0/178 (0/148)	0/178 (0/148)	0/178 (0/148)	0/443 (0/244)	2/443 (2/244)
v-*rel*	c-*rel*	1/426 (1/305)	3/426 (3/305)	5/426 (4/305)	0/313 (0/197)	5/313 (3/197)	0/313 (0/197)	0/353 (0/223)	3/353 (1/223)
v-*src*	c-*src*	0/245 (0/212)	0/245 (0/212)	0/245 (0/212)	1/178 (0/143)	1/178 (0/143)	0/178 (0/143)	0/352 (0/189)	1/352 (1/189)
Total		3/2078 (2/1607)	8/2078 (7/1607)	12/2078 (5/1607)	3/1571 (1/1091)	10/1571 (6/1091)	1/1571 (1/1091)	1/2475 (1/1491)	11/2475 (5/1491)

Table 9. The pattern of nucleotide substitutions for the nine viral oncogenes[a]

Original nucleotide	Substituted nucleotide A	T	C	G	[74.3]	[71.6]
All codon positions						
A	—	3.5	9.5	14.2	27.2	
T	4.7	—	15.6	1.6	21.9	
C	1.0	10.9	—	2.0	13.9	
G	33.6	3.4	0.0	—	37.0	
	39.3	17.8	25.1	17.8		
Sum of first and second codon positions						
A	—	3.4	12.2	13.9		29.5
T	2.6	—	15.4	2.6		20.6
C	1.9	9.4	—	3.7		15.0
G	32.9	2.0	0.0	—		34.9
	37.4	14.8	27.6	20.2		

[a] To see the trend in the change of base content, the sum of the f_{ij} values in each column is presented in the row below the matrix. These sums represent the numbers of substitutions that result in A, T, C, and G, respectively, among every 100 substitutions in a random sequence (Gojobori et al. 1982b). The relative frequency that one type of base is replaced by any of the others is also given by the sum of the f_{ij} values in each row and is presented in the column under the brackets

nation for the substitution patterns in viral oncogenes. To see if this is the case, we computed the f_{ij} values at the first two codon positions (Table 9) and examined the relationship between nucleotide exchangeability and amino acid exchangeability, because the majority of nucleotide changes at the first two codon positions cause amino acid changes. In particular, we examined whether the frequency of substitution correlates with the chemical dissimilarity between the amino acids interchanged. We did this because it is known that for the genes, such as functional nuclear genes, with constraints against amino acid changes, there is a negative correlation between these two quantities: the larger the dissimilarity between the interchanged amino acids, the less frequently the substitution occurs.

For nucleotide exchangeability in the viral oncogenes, we computed the sum of f_{ij} and f_{ji} for a given pair of nucleotides i and j, following Gojobori et al.'s (1982b) procedure. These values ($f_{ij} + f_{ji}$) at the first two codon positions are given as 24.7, 46.8, 3.7,14.0, 4.6, and 6.0 for the changes between C and T, G and A, C and G, C and A, T and G, and T and A, respectively. For the amino acid exchangeability, a dissimilarity index between a given pair of nucleotides has been obtained from the chemical dissimilarity between the interchanged amino acids (see Gojobori et al. 1982b for details). The dissimilarity indexes between C and T, G and A, C and G, C and A, T and G, and T and A are 120.5, 76.5, 84.5, 89.0, 153.0, and 142.5, respectively. The correlation coefficient between the $f_{ij} + f_{ji}$ values and the dissimilarity indexes is -0.52. Although this value is not significantly different from 0, this again suggests that constraints against amino acid changes certainly exist in the viral oncogenes, but are not as strong as those in the functional nuclear genes.

Table 8. Continued

C → G	G → A	G → T	G → C	Total
0/54 (0/27)	1/59 (1/40)	0/59 (0/40)	0/59 (0/40)	1/204 (1/136)
0/306 (0/190)	9/272 (3/166)	0/272 (0/166)	0/272 (0/166)	11/995 (5/664)
1/211 (1/134)	3/190 (0/116)	0/190 (0/166)	0/190 (0/116)	9/709 (5/473)
1/307 (1/188)	8/285 (4/173)	1/285 (0/173)	0/285 (0/173)	21/1026 (10/684)
0/229 (0/148)	3/171 (3/109)	1/171 (0/109)	0/171 (0/109)	9/762 (6/508)
0/220 (0/148)	0/216 (0/136)	1/216 (1/136)	0/216 (0/136)	1/848 (1/566)
0/443 (0/244)	2/357 (2/213)	0/357 (0/213)	0/357 (0/213)	8/1263 (6/842)
0/353 (0/223)	4/330 (3/223)	0/330 (0/223)	0/330 (0/223)	21/1422 (15/948)
0/352 (0/189)	0/317 (0/184)	0/317 (0/184)	0/317 (0/184)	3/1092 (1/728)
2/2475 (2/1491)	30/2197 (16/1360)	3/2197 (1/1360)	0/2197 (0/1360)	84/8321 (50/5549)

Recently, Graur (1985) studied the pattern of nucleotide substitution and the extent of purifying selection in retroviruses, using several different strains of Rous sarcoma virus. In his study, the direction of nucleotide substitution was estimated from an inferred ancestral sequence in the reconstructed phylogenetic tree. For the virus sequences used in his study, the divergence time is not known. If the divergence time is not short, many multiple substitutions may have occurred in the viral genes before the time of virus isolation. For this reason, the inferred ancestral sequence and the estimated direction of substitution may not be reliable. Thus, our approach appears to have an advantage in estimating the direction of nucleotide substitution. Nevertheless, our observation supports his conclusion that purifying selection in retroviral genomes is weak but the pattern of purifying selection follows the rules obtained for nuclear genes.

Acknowledgments. We thank Drs. J.F. Crow and R. Yokoyama for their helpful comments. We are also grateful to Dr. M. Kimura for his continuous encouragement. This research was supported by NIH grant GM-28672 and NIMH grant MH-31302 to S.Y. and a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan to T.G.

References

Abelson HT, Rabstein LS (1970) Lymphosarcoma virus induced thymic independent disease in mice. Cancer Res 30: 2213–2222

Air GM (1981) Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A virus. Proc Natl Acad Sci USA 78:7639–7643

Bender TP, Kuehl WM (1986) Murine myb protooncogene mRNA: cDNA sequence and evidence for 5′ heterogeneity. Proc Natl Acad Sci USA 83:3204–3208

Ben-Neriah Y, Bernards A, Paskind M, Daley GQ, Baltimore D (1986) Alternative 5′ exons in c-*abl* mRNA. Cell 44:577–586

Bishop JM (1983) Cellular oncogenes and retroviruses. Annu Rev Biochem 52:301–354

Bishop JM, Varmus HE (1982) Functions and origins of retroviral transforming genes. In: Weiss R, Teich N, Varmus H, Loffin J (eds) RNA tumor viruses, ed 2. Cold Spring Harbor Laboratory, Cold Spring Harbor NY, pp 999–1108

Buonagurio DA, Nakada S, Parvin JD, Krystal M, Palese P, Fitch WM (1986) Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. Science 232:980–982

Chen ISY, Mak TW, O'Rear JJ, Temin HM (1981) Characterization of reticuloendotheliosis virus strain T DNA and isolation of a novel variant of reticuloendotheliosis virus strain T by molecular cloning. J Virol 40:800–811

Curran T, Verma IM (1984) The FBR murine osteosarcoma virus. I. Molecular analysis and characterization of a 75,000-Da *gag-fos* fusion product. Virology 135:218–228

Curran T, Peters G, Van Benveren C, Teich NM, Verma IM (1982) FBJ murine osteosarcoma virus: identification and molecular cloning of biologically active proviral DNA. J Virol 44:674–682

Finkel MP, Biskis BO, Jinkins PB (1966) Virus induction of osteosarcomas in mice. Science 151:698–701

Finkel MP, Reilly CA Jr, Biskis BO, Greco IL (1973) Bone tumor virus. In: Colston Papers, Proceedings 24th Symposium of the Colston Research Society, vol 24. Butter-Worths, London, pp 353–366

Finkel MP, Reilly CA Jr, Biskis BO (1975) Viral etiology of bone cancer. Front Radiat Ther Oncol 10:28–39

Gojobori T, Yokoyama S (1985) Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. Proc Natl Acad Sci USA 82:4198–4201

Gojobori T, Ishii K, Nei M (1982a) Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. J Mol Evol 18:414–423

Gojobori T, Li WH, Graur D (1982b) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18:360–369

Graur D (1985) Pattern of nucleotide substitution and the extent of purifying selection in retroviruses. J Mol Evol 21: 221–231

Halpern CC, Hayward WS, Hanafusa H (1979) Characterization of some isolates of newly recovered avian sarcoma virus. J Virol 29:91–101

Hayashida H, Toh H, Kikuno R, Miyata T (1985) Evolution of influenza virus genes. Mol Biol Evol 2:289–303

Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S (1982) Rapid evolution of RNA genomes. Science 215: 1577–1585

Ivanov X, Mladenov Z, Nedyalkov S, Todorov TG, Yakimov M (1964) Experimental investigations into avian leucoses. V. Transmission, haematology and morphology of avian myelocytomatosis. Bull Inst Path Comp Animaux 10:5–38

Kawai S, Duesberg PH, Hanafusa H (1977) Transformation-defective mutants of Rous sarcoma virus with *src* gene deletions of varying length. J Virol 24:910–914

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626

Kimura M (1981) Estimation of evolutionary distances between

homologous nucleotide sequences. Proc Natl Acad Sci USA 78:454–458

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Klempnauer KH, Gonda TJ, Bishop JM (1982) Nucleotide sequence of the retroviral leukemia gene v-*myb* and its cellular progenitor c-*myb*: the architecture of a transduced oncogene. Cell 31:453–463

Krystal M, Buonagurio D, Young JF, Palese P (1983) Sequential mutations in the NS genes of influenza virus field strains. J Virol 45:547–554

Lautenberger JA, Schulz RA, Garon CF, Tsichlis PN, Papas TS (1981) Molecular cloning of avian myeloctyomatosis virus (MC29) transforming sequences. Proc Natl Acad Sci USA 78: 1518–1522

Lee GK, Chan EW, Reilley CA Jr, Pahnke VA, Rockus G, Finkel MP (1979) In vitro properties of FBR murine osteosarcoma virus (40650). Proc Soc Exp Biol Med 162:214–220

Li WH, Wu CI, Luo CC (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21:58–71

Martinez C, del Rio L, Portela A, Domingo E, Ortin J (1983) Evolution of the influenza virus neuraminidase gene during drift of the N2 subtype. Virology 130:539–545

Miyata T, Yasunaga T (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16:23–26

Mladenov Z, Heine U, Beard D, Beard JW (1967) Strain MC29 avian leukosis virus. Myelocytoma, endothelioma, and renal growths: pathomorphological and ultrastructural aspects. J Natl Cancer Inst 38:251–285

Moloney JB (1966) A virus-induced rhabdomyosarcoma of mice. Natl Cancer Inst Monogr 22:139–142

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nei M, Tajima F, Gojobori T (1984) Classification and measurement of DNA polymorphism. In: Chakravarti A (ed) Human population genetics; the Pittsburgh symposium. Van Nostrand Reinhold, New York, pp 307–330

Nunn M, Weiher H, Bullock P, Duesberg P (1984) Avian erythroblastosis virus E26: nucleotide sequence of the tripartite *onc* gene and of the LTR, and analysis of the cellular prototype of the viral *ets* sequence. Virology 139:330–339

Reddy EP, Smith MJ, Aaronson SA (1981) Complete nucleotide sequence and organization of the Moloney murine sarcoma virus genome. Science 214:445–450

Reddy EP, Smith MJ, Srinivasan A (1983) Nucleotide sequence of Abelson murine leukemia virus genome: structural similarity of its transforming gene product to other *onc* gene products with tyrosine-specific kinase activity. Proc Natl Acad Sci USA 80:3623–3627

Rosson D, Reddy EP (1986) Nucleotide sequence of chicken c-*myb* complementary DNA and implications for *myb* oncogene activation. Nature 319:604–606

Saitou N, Nei M (1986) Polymorphism and evolution of influenza A virus genes. Mol Biol Evol 3:57–74

Shtivelman E, Lifshiz B, Gale RP, Canaani E (1985) Fused transcript of *abl* and *bcr* genes in chronic myelogenous leukemia. Nature 315:550–554

Simon MA, Drees B, Kornberg T, Bishop JM (1985) The nucleotide sequence and the tissue-specific expression of *Drosophila* c-*src*. Cell 42:831–840

Srinivasan A, Reddy EP, Aaronson SA (1981) Abelson murine leukemia virus: molecular cloning of infectious integrated proviral DNA. Proc Natl Acad Sci USA 78:2077–2081

Stanton LW, Watt R, Marcu KB (1983) Translocation, breakage and truncated transcripts of c-*myc* oncogene in murine plasmacytomas. Nature 303:401–406

Takeya T, Hanafusa H (1982) DNA sequence of the viral and cellular *src* gene of chickens II. Comparison of the *src* genes of two strains of avian sarcoma virus and of the cellular homolog. J Virol 44:12–18

Takeya T, Hanafusa H (1983) Structure and sequence of the cellular gene homologous to the RSV *src* gene and the mechanism for generating the transforming virus. Cell 32:881–890

Takeya T, Hanafusa H, Junghana RP, Ju G, Skalka AM (1981) Comparison between the viral transforming gene (*src*) of recovered avian sarcoma virus and its cellular homolog. Mol Cell Biol 1:1024–1037

Teich N, Wyke J, Mak T, Bernstein A, Hardy W (1982) Pathogenesis of retrovirus-induced disease. In: Weiss R, Teich N, Varmus H, Loffin J (eds) RNA tumor viruses, ed 2. Cold Spring Harbor Laboratory, Cold Spring Harbor NY, pp 785–998

Theilen GH, Zeigel RF, Twiehaus MJ (1966) Biological studies with RE virus (strain T) that induces reticuloendotheliosis in turkeys, chickens, and Japanese quail. J Natl Cancer Inst 37: 731–743

Van Beveren C, Galleshaw JA, Jonas V, Berns AJM, Doolittle R, Donoghue DJ, Verma IM (1981a) Nucleotide sequence and formation of the transforming gene of a mouse sarcoma virus. Nature 289:258–262

Van Beveren C, van Straaten F, Galleshaw JA, Verma IM (1981b) Nucleotide sequence of the genome of a murine sarcoma virus. Cell 27:97–108

Van Beveren C, van Straaten F, Curran T, Muller R, Verma IM (1983) Analysis of FBJ-MuSV provirus and c-*fos* (mouse) gene reveals that viral and cellular *fos* gene products have different carboxy termini. Cell 32:1241-1255

Van Beveren C, Enami S, Curran T, Verma IM (1984) FBR murine osteosarcoma virus. Virology 135:229–243

Van Straaten F, Muller R, Curran T, Van Beveren CV, Verma IM (1983) Complete nucleotide sequence of a human c-*onc* gene: deduced amino acid sequence of the human c-*fos* protein. Proc Natl Acad Sci USA 80:3183–3187

Wang JYJ, Ledley F, Goff S, Lee R, Groner Y, Baltimore D (1984) The mouse c-*abl* locus: molecular cloning and characterization. Cell 36:349–356

Watson R, Oskarsson M, Woude GFV (1982) Human DNA sequence homologous to the transforming gene (*mos*) of Moloney murine sarcoma virus. Proc Natl Acad Sci USA 79: 4078–4082

Watson DK, Psallidopoulos MC, Samuel KP, Dalla-Favera R, Papas TS (1983a) Nucleotide sequence analysis of human c-*myc* locus, chicken homologue, and myelocytomatosis virus MC29 transforming gene reveals a highly conserved gene product. Proc Natl Acad Sci USA 80:3642–3645

Watson DK, Reddy EP, Duesberg PH, Papas TS (1983b) Nucleotide sequence analysis of the chicken c-*myc* gene reveals homologous and unique coding regions by comparison with the transforming gene of avian myelocytomatosis virus MC29, Δ*gag-myc*. Proc Natl Acad Sci USA 80:2146–2150

Wilhelmsen KC, Eggleton K, Temin HM (1984) Nucleic acid sequence of the oncogene v-*rel* in reticuloendotheliosis virus strain T and its cellular homolog, the proto-oncogene c-*rel*. J Virol 52:172–182

Received August 26, 1986/Revised March 4, 1987