Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus

N. Kitamura*, A. Kitamura, K. Toyoshima‡, Y. Hirayama & M. Yoshida†

Cancer Institute, Toshima-ku, Tokyo 170, Japan and ‡ Institute for Medical Sciences, University of Tokyo, Tokyo 108, Japan

From the complete nucleotide sequence of the genome of the avian sarcoma virus Y73, we have predicted amino acid sequence of p90$^{gag\text{-}yes}$, the product of the transforming gene. Contrary to previous evidence from molecular hybridization studies, p90$^{gag\text{-}yes}$ was found to have much homology with the transforming gene product p60^{src} of Rous sarcoma virus, suggesting that the cellular counterparts of the two (c-yes and c-src) originated from a common prototype sequence.

A GROUP of retroviruses, the acute leukaemia/sarcoma viruses, contains the transforming gene (v-*onc*), which is responsible for initiation and maintenance of cellular transformation[1]. More than 13 genes have been identified as v-*onc* and shown to be acquired from a normal cellular sequence as an integral part of their genome. Three distinct classes of transforming gene capable of inducing fibrosarcomas in chicken were defined among avian sarcoma viruses (ASV), that is, *src* in Rous sarcoma virus (RSV), *yes*[2,3] in Y73 and Esh sarcoma virus, and *fps*[4-6] in Fujinami and PRCII strains of ASV. These three transforming genes have no homology in nucleic acid hybridization[2-5], and are derived from the respective sequences in normal cells by recombination with viral sequences[2,7,8]. Furthermore, the cellular counterparts homologous with these transforming genes have been conserved during evolution[9] (M.Y., unpublished data) and expressed in normal cells[10], suggesting some that these genes perform some crucial role in normal cells.

In the light of these observations it is surprising that the transforming gene products of all these classes have protein kinase activity phosphorylating tyrosine on target proteins[11-14], and that this enzyme activity seems to be essential for cellular transformation[15,16]. Therefore, we have investigated the structure of these genes and the gene products in an attempt to understand not only the structural basis of the functions, but also the mechanism of biogenesis of the transforming genes.

The nucleotide sequence of *src* in Schmidt–Ruppin strain RSV (SR-RSV) has been established, and the function of the gene product (p60^{src}) and biogenesis of the *src* have been discussed on the basis of the nucleotide and amino acid sequences[17]. We have studied a recent isolate of ASV, Y73[18], which is defective in its replication and associated with leukosis virus (YAV) of subgroup A. The genome of Y73 is 26S RNA and contains *yes* as a transforming gene in its middle portion[19], which encodes polyprotein with a molecular weight of 90,000 (p90$^{gag\text{-}yes}$) consisting of a *gag*-gene derived part and Y73-specific portion. *In vitro* translation of 26S RNA[4] showed directly that p90$^{gag\text{-}yes}$ is coded by the Y73-specific RNA sequence. The association of protein kinase activity with p90$^{gag\text{-}yes}$ suggests that p90$^{gag\text{-}yes}$ is a transforming gene product.

To investigate the structure of p90$^{gag\text{-}yes}$, we determined the nucleotide sequence of the whole genome DNA. The sequence contains an uninterrupted reading frame that can code for p90$^{gag\text{-}yes}$ and the predicted amino acid sequence of p90$^{gag\text{-}yes}$ is strikingly similar to that of p60^{src}.

Sequencing strategy

For determination of the nucleotide sequence, we used the molecular cloning technique to amplify Y73 genome DNA. Closed circular DNA specific to Y73 was isolated from Y73-infected chick embryo cells by extraction of DNA by Hirt's procedure[20] followed by acid phenol extraction[21] and agarose gel electrophoresis. The purified DNA was linearized with restriction endonuclease *Sst*I and inserted into λ phage Charon 16A[22] at site of *Sst*I. Several clones containing the Y73-specific sequence were isolated and one of them, λY73-11A, was selected for sequencing, because it was shown by transfection assay to contain the complete sequence of Y73 RNA. Cotransfection of this cloned DNA together with cloned helper proviral DNA induced cellular transformation of chick cells and viral replication, the progeny virus was indistinguishable from the original Y73 virus. Details of the cloning the transfection will be reported elsewhere (manuscript in preparation).

Figure 1 shows a restriction map of clone λY73-11A insert containing two long terminal repeats (LTRs). The DNA fragments generated by digestion with *Hpa*II, *Hin*fI, and *Sau*3A were used for sequencing. The fragments were phosphorylated at the 5′ end with polynucleotide kinase and [γ-^{32}P]ATP, separated by either strand separation or further digestion with various restriction enzymes and sequenced by the method of Maxam and Gilbert[23]. Most of the sequences were confirmed by sequencing the complementary strand, or the fragments generated by digestions with different enzymes.

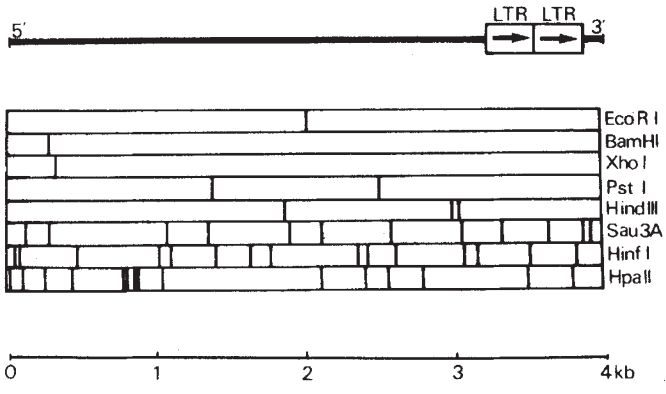

Fig. 1 Restriction map of Y73 genome cloned in λY73-11A. The circular DNA was linearized and inserted into λ Charon 16A vector at *Sst*I site and the cloned DNA contained two direct repeats of LTR.

* Present address: Institute for Immunology, Kyoto University, Kyoto, Japan.

† To whom reprint requests should be addressed.

0028-0836/82/200205—04$01.00

© 1982 Macmillan Journals Ltd

Fig. 2 Total nucleotide sequence of Y73 viral genome and the predicted amino acid sequence of $p90^{gag\text{-}yes}$. The complete sequence of the cloned DNA in λY73-11A was shown in the arrange of RNA genome from 5′ to 3′ direction. U5 and U3 indicate the unique sequence to 5′ and 3′ terminal regions. Thick solid arrows with 'rec' shows the predicted recombination sites between c-*yes* and YAV sequences, and 'spc' represents sequence which is not contained in RSV. The horizontal thick arrows indicate the initiation and termination sites for $p90^{gag\text{-}yes}$. The small arrow with p19 indicates a sequence homologous to the carboxy-terminus of p19 of Pr-RSV. The closed bar between nucleotides 2,477 and 2,479 shows the putative phosphoacceptor tyrosine contained in the same nonapeptide described by Neil *et al.*[26] and the small open circles between 3,174 and 3,176 show the position expected as termination codon for gp37, if present.

Nucleotide sequence

Previously, fingerprint analysis of Y73 26S RNA and YAV RNA[19] demonstrated that 26S RNA contained a specific sequence that represents the transforming gene of Y73. This specific sequence was mapped in the middle of the genome with sequences common to YAV RNA at both ends. The transforming gene product $p90^{gag\text{-}yes}$ encoded by the Y73-specific sequence was coprecipitated with antibody against p19, a *gag*-gene product, suggesting that $p90^{gag\text{-}yes}$ contained at least p19 in the amino-terminal portion[12].

The entire nucleotide sequence of the Y73 genome, shown in Fig. 2, consists of 3,718 nucleotides, confirming the previous data described above. The genome was significantly shorter than 4.8 kilobases, the value estimated previously by agarose gel electrophoresis of the genomic RNA.

The noncoding sequence at the 5′-terminus, including U5 and a primer tRNA binding site, was found to be homologous with that of the SR-RSV genome[31,32] (93% homology). The open reading frame, starting at position 380, contained the entire sequence coding for p19, which is located at the amino-

©1982 Nature Publishing Group

terminus of the *gag*-precursor, Pr76, indicating that AUG at position 380 is the initiation codon for p90$^{gag\text{-}yes}$. The proposed amino acid sequence of p19 was identical with that of p19 of Prague strain of RSV (D. Schwartz *et al.*, personal communication), except for the 2 amino acid insertions (positions 169 and 170) and 2 amino acid substitutions (positions 64 and 161). After the coding sequence for p19, 130 nucleotides were homologous with those of RSV and then there was a specific sequence. The sequence coding for p27, which follows after p19 in Pr76, was not present. Therefore, recombination between the viral sequence and the *yes*-related cellular sequence (c-*yes*) seems to have taken place in the sequence between those coding for p19 and p27. A pyrimidine-rich sequence was present after the putative recombination site, but, the significance of this sequence is not known. From the putative recombination site, the Y73-specific sequence extended to position 2,795. These specific 1,755 nucleotides are the only candidates for a coding sequence of the specific portion of p90$^{gag\text{-}yes}$. After this specific region, there was a sequence homologous to that coding for gp37 of SR-RSV[17], which is one of the *env*-gene products. Thus the recombination site at the 3′-side seemed to be within the sequence coding for gp37. Part of the sequence homologous to that coding for gp37 in SR-RSV was utilized for the carboxy-terminal peptide of p90$^{gag\text{-}yes}$ using a different frame from that for gp37. The UGA codon in the gp37 sequence was the predicted termination codon for p90$^{gag\text{-}yes}$.

Another 161 base sequence around the termination codon for gp37 is not homologous to that in PR-RSV (Schwartz *et al.*, personal communication), however, it seems too small to code any protein required for transformation. The origin of this specific sequence is not clear, but possibly the unique sequence was contained in, and originated from, the helper virus (YAV) genome. The common sequence at the 3′-end was homologous with that of SR-RSV[17,31] (87% homology) except for a 14 nucleotide deletion within the U3 region near the 5′ end. This deletion did not affect expression of the Y73 genome because the cloned DNA had transforming activity in transfection.

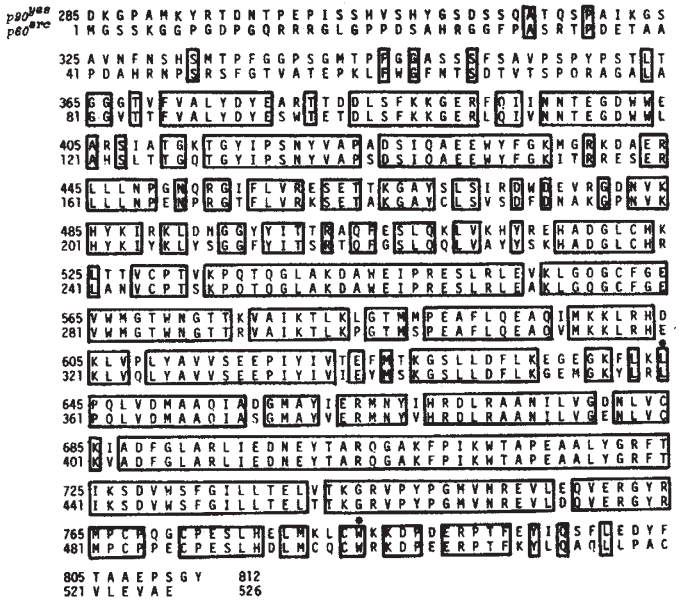

Fig. 3 Comparison of the proposed amino acid sequences for the specific region of p90$^{gag\text{-}yes}$ and p60^{src} of Pr-RSV (N.K. and M.Y., manuscript in preparation). The two amino acid sequences are aligned to give the maximum homology. The numbers in the left are the amino acid residues from the initiation methionine and area of identity are depicted by boxes. The nucleotide sequences between two marks (*) are compared in Fig. 4. A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

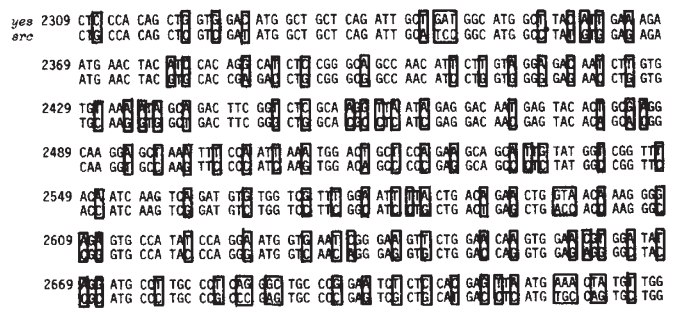

Fig. 4 Comparison between nucleotide sequences for the regions of *yes* and *src* of Pr-RSV coding for the highly homologous amino acid sequences. The nucleotide sequences corresponding to the amino acid sequences marked (*) in Fig. 3 are compared between *yes* and *src*, in which 93% homology was observed in amino acid sequences. Non-identical nucleotides are depicted by boxes.

Amino acid sequence of p90$^{gag\text{-}yes}$

Figure 2 also shows the predicted amino acid sequence of the gene product p90$^{gag\text{-}yes}$ of Y73. P90$^{gag\text{-}yes}$ consists of 812 amino acids and has a molecular weight of 88,697, which is in good agreement with the value of 90,000 estimated by SDS polyacrylamide gel electrophoresis[12]. As mentioned above, p19 was located at the amino acid terminus of the protein followed by a sequence unrelated to the RSV gene product.

Comparison of the amino acid sequence of part of p90$^{gag\text{-}yes}$ and p60^{src} is shown in Fig. 3. The amino acid sequence of p60^{src} was predicted by the nucleotide sequence of *src* in Pr-RSV, which was determined in our laboratory analysing molecularly cloned *Eco*RI fragment of the integrated proviral DNA (N.K. and M.Y., in preparation). Our sequence of *src* in Pr-RSV only differed in 12 nucleotides from that determined by Schwartz *et al.* (personal communication) but differed considerably from that of Czernilofsky *et al.* for SR-RSV, affecting a part of the reading frame for p60^{src}. Seventy-three out of 80 amino acids of the amino-terminus and 10 of the carboxy-terminus of p60^{src} were totally different from the corresponding regions of p90$^{gag\text{-}yes}$ when arranged to get maximum homology, but the other 436 amino acids of p60^{src} were largely homologous (82%) with those of p90$^{gag\text{-}yes}$. In particular 133 amino acids between positions 644 and 766 of p90$^{gag\text{-}yes}$ showed 93% homology with the corresponding region of p60^{src}.

Discussion

In earlier experiments[2,12], the Y73 specific cDNA sequence did not hybridize with *src* RNA. From this finding it was concluded that the transforming gene *yes* is distinct from *src* of RSV. However, our results show a large extent of homology in the amino acid sequences of the transforming gene products, p90$^{gag\text{-}yes}$ and p60^{src}. This discrepancy is due to the diversity of the genetic codes for each amino acid. Of 356 identical amino acids between positions 365 and 800, 221 utilized different codons from those for p60^{src} mainly owing to differences in the third base of the codons (Fig. 4). At the level of nucleotides, 31% of the 1,308 bases of the region are different in *src* and *yes*, and the longest homologous stretch is only 17 nucleotides long. These mismatches can explain the previous finding that cDNA of *yes* did not hybridize with *src* RNA under stringent conditions.

The highly homologous sequences of amino acids in p90$^{gag\text{-}yes}$ and p60^{src} are consistent with the fact that both gene products have protein kinase activity phosphorylating tyrosine[12,24,25]. Furthermore, these homologies strongly suggest that the two transforming genes were derived from a common prototype sequence. This was previously proposed by Neil *et al.*[26] from the finding of similar tryptic peptides containing phosphoaccep-

©1982 Nature Publishing Group

tor tyrosine in p60^{src} and p90$^{gag\text{-}yes}$. Our nucleotide sequence data clearly show homology over a wider range covering most of p60^{src}, thus supporting the proposal of Neil *et al.*[26]. Some similarity between p60^{src} and the *mos* gene product of Mo-MSV was reported[27] and similarity was also found between two transforming gene products coded by Ha- and Ki-MSV[28].

Protein kinase activity phosphorylating tyrosine residues is a common feature of the transforming gene products of all avian sarcoma viruses so far tested. Therefore, it would be very interesting to know the amino acid sequences of the gene products of other avian sarcoma viruses such as Fujinami virus and PRCII.

As previously reported[2,8], there are DNA sequences homologous to those of *yes* and *src*, respectively in normal chick cells (c-*yes* and c-*src*), and these cellular sequences have been well conserved during evolution, being found in fish and in humans, suggesting some important function for these genes[9]. Therefore, the similarity of p90$^{gag\text{-}yes}$ and p60^{src} at a protein level implies similarity of the gene products coded by c-*yes* and c-*src*, and suggests that c-*yes* and c-*src* diverged from a common prototype sequence at a very early stage of vertebrate evolution.

C-*src* was shown to be expressed in protein p60$^{c\text{-}src}$ in normal cells[10,11] and c-*yes* was shown to be transcribed into mRNA (M.Y., unpublished data). We have no data that could explain why different genes coding for proteins with similar structure and function are conserved and expressed in normal cells, although some speculations are possible. For example, expression of these cellular genes might depend on different types of differentiation or differences in the cellular distributions, and thus functions, of the proteins.

The structural similarity between p90$^{gag\text{-}yes}$ and p60^{src} does not necessarily imply that the mechanisms of cellular transformation by these viruses are the same. Similarity in the carboxy-terminal domain could explain the similar protein kinase activity[29], but, differences in the amino-terminal domain suggest a different distribution of the proteins within cells, and thus that they phosphorylate different target proteins, since the amino-terminal domain of p60^{src} is reported to be involved in anchorage of the protein to the plasma membrane[30].

We thank Dr Haruo Sugano for his useful discussions. This work was partly supported by Grant-in-Aid for Cancer Research from the Ministry of Education, Science and Culture of Japan.

Received 11 January; accepted 24 March, 1982.

1. Bishop, J. M. *A. Rev. Biochem.* **47**, 35–88 (1978).
2. Yoshida, M., Kawai, S. & Toyoshima, K. *Nature* **287**, 653–654 (1980).
3. Neil, J. C., Breitman, M. L. & Vogt, P. K. *Virology* **108**, 98–110 (1981).
4. Hanafusa, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3009–3013 (1980).
5. Lee, W.-H. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 2018–2022 (1980).
6. Gysdael, J., Neil, J. C. & Vogt, P. K. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2611–2615 (1981).
7. Shibuya, M., Hanafusa, T., Hanafusa, H. & Stephenson, J. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6536–6540 (1980).
8. Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. *Nature* **260**, 170–173 (1976).
9. Spector, D., Varmus, H. E. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4102–4106 (1978).
10. Spector, D., Baker, B., Varmus, H. E. & Bishop, J. M. *Cell* **13**, 381–386 (1978).
11. Collet, M. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2021–2024 (1978).
12. Kawai, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 6199–6203 (1980).
13. Feldman, R. A., Hanafusa, G. & Hanafusa, H. *Cell* **22**, 757–765 (1980).
14. Neil, J. C., Ghysdael, J. & Vogt, P. K. *Virology* **109**, 223–228 (1981).
15. Levinson, A. D., Opperman, H., Levinthow, L., Varmus, H. E. & Bishop, J. M. *Cell* **15**, 561–572 (1978).
16. Powson, T. *et al. Cell* **22**, 767–775 (1980).
17. Czernilofsky, A. P. *et al. Nature* **287**, 198–203 (1980).
18. Itohara, S., Hirata, K., Inoue, M., Hatsuoka, M. & Sato, A. *Gann* **69**, 825–830 (1978).
19. Yoshida, M., Kawai, S. & Toyoshima, K. *J. Virol.* **38**, 430–437 (1981).
20. Hirt, N. *J. molec. Biol.* **26**, 365–369 (1967).
21. Zasloff, M., Ginder, G. P. & Felsenfeld, G. *Nucleic Acids Res.* **5**, 1139–1152 (1978).
22. Williams, B. G. & Blattner, F. R. *J. Virol.* **29**, 555–575 (1979).
23. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
24. Collet, M. S., Erikson, E., Purchio, A. F., Brugge, J. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3159–3163 (1979).
25. Hunter, T. & Sefton, B. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1311–1315 (1980).
26. Neil, J. C., Ghysdael, J., Vogt, P. K. & Smart, J. E. *Nature* **291**, 675–677 (1981).
27. Beveren, C. V. *et al. Nature* **289**, 258–262 (1981).
28. Ellis, R. W. *et al. Nature* **292**, 506–511 (1981).
29. Levinson, A. D., Courtneidge, S. A. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1624–1628 (1981).
30. Courtneidge, S. A., Levinson, S. A. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3783–3787 (1980).
31. Czernilofsky, A. P. *et al. Nucleic Acids Res.* **8**, 2967–2984 (1980).
32. Swanstrom, R., Delorbe, W. J., Bishop, J. M. & Varmus, H. E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 124–128 (1981).

LETTERS TO NATURE

Magnetic fields, convection and solar luminosity variability

W. C. Livingston

Kitt Peak National Observatory, Tucson, Arizona 85726, USA

Model calculations universally indicate that heat transport from the solar interior is almost entirely by convection in the outer envelope. Spiegel and Weiss[1] have recently discussed how magnetic fields can intrude into this envelope to affect convection and thus introduce transients into the Sun's luminosity. I demonstrate here from Fourier transform spectrometer observations of Fraunhofer line asymmetry that granular convection is retarded in the presence of surface magnetism. I then present full disk observations which suggest a lessening of convection over the past 5 yr. As this time interval coincides with the rise from a minimum to a maximum of solar activity, with the presumed injection of new magnetic flux from the interior, the resolved disk and full disk observations are consistent. Finally I note the continued temporal decrease in the spectroscopic temperature of the low photosphere and consider how a change of convection and the lowering temperature may be related.

The convective process is plainly manifest at the Sun's surface by the granulation pattern. Rising bright granules, transporting excess energy which is then radiated into space, are surrounded by the cooler intergranule gas which is falling. This causes a velocity–brightness correlation which in turn gives rise to an asymmetry in Fraunhofer lines observed near disk centre[2,3].

Spectral line asymmetries are traditionally represented by plotting the shifts of the bisectors of the line profile (as in Fig. 1). These bisectors have a characteristic 'C' shape, the middle of which is blueshifted in wavelength relative to the laboratory value. The upper end of the C is less blueshifted, as increased density low in the photosphere, coupled with the condition of flow continuity, results in lower convective velocities. The lower end is also less blueshifted, because of the cessation of upward velocity at the top of the granulation layer as represented by the line core. Detailed calculations by Nordlund[4] provide a quantitative basis for the above hypothesis.

A Fourier transform spectrometer (FTS) is preferred for bisector observations; it has a symmetric response profile, is capable of giving absolute wavelengths, records many lines simultaneously and is easily fed with unfocused (integrated) sunlight[5]. In a day-long sequence of full disk 1-m FTS measurements, mid C values typically display a wavelength variance of ~ 0.1 mÅ, corresponding to a velocity noise of 5 m s^{-1}. For historical reasons, some of the data presented here were obtained with a 13.5-m grating spectrometer. The greatest problem in this spectrograph is thermal drift which introduces a time-dependent asymmetric instrumental profile. Velocity noise is also about 10 times worse than with the FTS, but by averaging many records the noise in the data becomes tolerable.

Figure 1 shows line bisectors for a region of high magnetic flux compared with a non-magnetic area nearby. The spatial resolution is purposely low, about 1×2 arc min, to average over local conditions. This record is typical of some 50 regions

0028-0836/82/200208—02$01.00
© 1982 Macmillan Journals Ltd