Howe, Eds. (American Society for Microbiology, Washington, DC, 1989), pp. 163–184.
33. D. Pruss, F. D. Bushman, A. P. Wolffe, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5913 (1994); H.-P. Müller and H. E. Varmus, *EMBO J.* **13**, 4704 (1994).
34. T. A. Baker, M. Mizuuchi, H. Savilahti, K. Mizuuchi, *Cell* **74**, 723 (1993); T. A. Baker, E. Kremenstova, L. Luo, *Genes Dev.* **8**, 2416 (1994).
35. K. Mizuuchi, *Annu. Rev. Biochem.* **61**, 1011 (1992).
36. C. M. Joyce and T. A. Steitz, *ibid.* **63**, 777 (1994); S. M. Linn, R. S. Lloyd, R. J. Roberts, Eds., *Nucleases* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 2, 1993); J. F. Davies II, R. J. Almassy, Z. Hostomska, R. A. Ferre, Z. Hostomsky, *Cell* **76**, 1123 (1994); M. R. Sawaya, H. Pelletier, A. Kumar, S. H. Wilson, J. Kraut, *Science* **264**, 1930 (1994).
37. P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8924 (1988); L. S. Beese and T. A. Steitz, *EMBO J.* **10**, 25 (1991).
38. A. G. Murzin and C. Chothia, *Curr. Opinion Struct. Biol.* **2**, 895 (1992).
39. R. Craigie, K. Mizuuchi, F. D. Bushman, A. Engelman, *Nucleic Acids Res.* **19**, 2729 (1991); D. J. Hazuda, J. C. Hastings, A. L. Wolfe, E. A. Emini, *ibid.* **22**, 1121 (1994); C. Vink, M. Banks, R. Bethell, R. H. A. Plasterk, *ibid.*, p. 2176.
40. M. Cushman and P. Sherman, *Biochem. Biophys. Res. Commun.* **185**, 85 (1992); S. Carteau, J. F. Mouscadet, H. Goulaouic, F. Subra, C. Auclair, *ibid.* **192**, 1409 (1993); M. R. Fesen, K. W. Kohn, F. Leteurtre, Y. Pommier, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2399 (1993); A. Mazumder, D. Cooney, R. Agbaria, A. Gupta, Y. Pommier *ibid.* **91**, 5771 (1994).
41. Z. Otwinowski, in *Data Collection and Processing*, L. Sawyer, N. Isaacs, S. Bailey, Eds. (Science and Engineering Research Council, Warrington, United Kingdom, 1993), pp. 56–62.
42. W. Furey and S. Swaminathan, *Phases—A program package for the processing and analysis of diffraction data for macromolecules*. Poster, American Crystallographic Association meeting, 1990.
43. H. Wu, J. W. Lustbader, Y. Liu, R. E. Canfield, W. A. Hendrickson, *Structure* **2**, 545 (1994).
44. V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, R. M. Sweet, *Nature* **262**, 219 (1993).
45. B. C. Wang, *Methods Enzymol.* **115**, 90 (1985).
46. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Cryst.* **A47**, 110 (1991).
47. A. T. Brünger, *X-PLOR Version 3.1. A system for X-ray crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
48. A. T. Brünger, *Nature* **355**, 472 (1992).
49. P. J. Kraulis, *J. Appl. Cryst.* **24**, 946 (1991).
50. M. Carson, *ibid.*, p. 958.
51. We thank P. Rice and K. Mizuuchi for advice; C. Ogata and P. Rice for assistance in collecting the MAD data at NSLS, Brookhaven; R. Ghirlando for determining the dynamic light scattering properties of the protein; and S. Landry for directing our attention to the similarity between RNase H and ATPase domains. The structural study reported here was based on a long-term effort on the part of a number of people. In particular, we thank F. Bushman, P. Wingfield, and I. Palmer for developing protein purification protocols and supplying a number of integrase derivatives in the early stages of the project; P. Sun for initial crystallization attempts; and M. Carmichael, S. Hosseini, and R. Madabhushi for technical assistance. Supported in part by the NIH Intramural AIDS Targeted Antiviral Program. Coordinates of the structure at the present stage of refinement have been deposited in the Protein Data Bank under accession number 1ITG/T5588.

13 October 1994; accepted 15 November 1994

Splicing of the *rolA* Transcript of *Agrobacterium rhizogenes* in *Arabidopsis*

Armando Magrelli, Kerstin Langenkemper, Christoph Dehio,* Jeff Schell, Angelo Spena†

The *rolA* gene encoded on the Ri plasmid A4 of *Agrobacterium rhizogenes* is one of the transferred (T_L-DNA) genes involved in the pathogenesis of hairy-root disease in plants. The function of the 100–amino acid protein product of *rolA* is unknown, although its expression causes physiological and developmental alterations in transgenic plants. The *rolA* gene of *A. rhizogenes* contains an intron in its untranslated leader region that has features typical of plant pre–messenger RNA introns. Transcription and splicing of the *rolA* pre–messenger RNA occur in the plant cell.

The *rolA* gene from the Ri plasmid A4 of *A. rhizogenes* is one of the T_L-DNA genes transferred from the bacterium to the plant, and it is involved in the pathogenesis of hairy-root disease (*1*). Although transferred DNA (T-DNA)–encoded genes of *A. tumefaciens* can be transcribed and translated in bacterial extracts (*2*), we are not aware of data reporting bacterial transcription of T_L-DNA–encoded genes of *A. rhizogenes*. Expression takes place in transformed plant cells, and the *rolA* gene by itself causes plant developmental alterations, including dwarfism (due to reduced growth and internode distance) and wrinkled leaves (due to reduced growth of the midrib and of vascular tissue) (*3*).

Conflicting results concerning the initiation of transcription of the *rolA* gene [mapped by primer extension to position −29 from the ATG initiation codon (*4*) or to position −100 (*5*)] led us to analyze the structure of the transcript in more detail. We did reverse transcription–polymerase chain reaction (RT-PCR) with polyadenylated [poly(A)$^+$] RNA extracted from *Arabidopsis thaliana* plants transgenic for the *rolA* gene (line 23) (*5*), using as a 5′ primer an oligonucleotide spanning nucleotides −100 to −82 (primer a in Fig. 1) and as 3′ primer an oligonucleotide spanning nucleotides +300 to +282 of the *rolA* gene (primer b in Fig. 1). The polyadenylation site is at position +530 (*5*). The PCR produced fragments of two lengths, differing by 76 nucleotides (Fig. 1). DNA sequence analysis indicated that, in *Arabidopsis*, *rolA* transcripts have either an untranslated leader region (ULR) of at least 100 bases (I in Fig. 1) or a ULR of 24 bases (II in Fig. 1). The shorter class of transcripts is identical to the long one except for the deletion of 76 bases in the ULR of the *rolA* mRNA. The deletion starts with the sequence GT at position −76 and ends with AG at position −3. The dinucleotides GT (at the 5′ end) and AG (at the 3′ end) are known to delimit eukaryotic introns and to be invariable parts of the splice sites (*6*). Thus, the two classes of RT-PCR products represent the unspliced and spliced mRNA of the *rolA* gene. The reported discrepancy could be partially explained by the fact that mapping experiments have used poly(A)$^+$ RNA extracted from tobacco (*4*) and *Arabidopsis* (*5*) plants. The efficiency of *rolA* pre–messenger RNA (pre-mRNA) splicing differs in these two plant species (*7*).

To further characterize the spliced *rolA* mRNA, 12 independent RT-PCR products, corresponding to spliced mRNAs, were cloned and sequenced. Ten clones had a ULR of 27 bases (III in Fig. 1B), whereas in the remaining two clones, the ULR was 24 bases long (II in Fig. 1B). This result defines two classes of spliced transcripts that were generated by the use of alternative 5′ splice sites: the GT at position −73 for class I and the GT at position −76 for class II. Consequently, splicing of *rolA* pre-mRNA removes an intron of 73 to 76 nucleotides, which is in agreement with a minimum length of 70 to 73 nucleotides reported for efficient splicing of introns in plants (*8*). Furthermore, the *rolA* intron has a 71% AT content and the AG 3′ splice site is preceded by a T-rich region, features considered typical of plant pre-mRNA introns (*9*). The two 5′ exon-intron junctions show homology to the plant consensus sequence ($G_{72}G_{100}T_{99}A_{70}A_{55}G_{65}T_{49}$) (*6*).

Many mutant alleles are caused by mutations that interfere with RNA splicing (*10*). Thus, to confirm the molecular data, *rolA* alleles from seven independent null mutants isolated by ethylmethane sulfonate mutagenesis of an *Arabidopsis* line transgenic for the *rolA* gene (line 23) (*5*) were cloned by PCR (5′ primer: AACGCTTCAATACGGTGAG; 3′ primer: AATACGCACGTGGCTGGCGGTCTT) and sequenced. Four out of seven were single point mutations leading either to amino acid substitutions [Arg at position 37 to Trp in mutant line 23-4(1); Pro at position 40 to Ser in mutant line 23-1; Pro at position 40 to Leu in mutant line 23-6] or to change

Max-Plank-Institut für Züchtungsforschung, Carl-von-Linné Weg 10, 50829 Cologne, Germany.

*Present address: Institut Pasteur, 28 rue du Dr Roux, 75724 Paris, Cedex 15, France.
†To whom correspondence should be addressed.

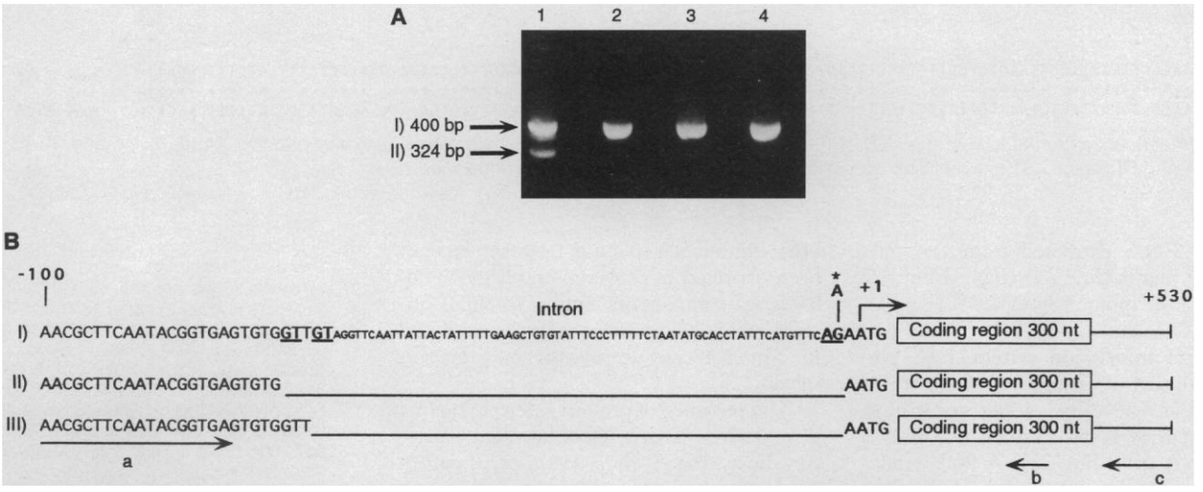

Fig. 1. (**A**) Agarose (3.5%) (NuSieve) gel electrophoresis of RT-PCR products obtained from poly(A)$^+$ RNA extracted from *Arabidopsis* plants transgenic for the *rolA* gene (*21*). Lane 1, line 23 (transgenic for the *rolA* gene); lane 2, null mutant line 23-2(1); lane 3, null mutant line 23-3; lane 4, null mutant line 23-5. RT-PCR done with poly(A)$^+$ RNA extracted from wild-type *Arabidopsis* plants did not give any amplification product. (**B**) DNA sequence of *rolA* poly(A)$^+$ transcripts. I: DNA sequence (*22*) of the ULR of the *rolA* gene RT-PCR product, showing initiation of transcription (−100), initiation of translation (+1), and the transition from G to A (asterisk) at position −2 found in the *rolA* gene of the null mutant lines 23-2(1), 23-3, and 23-5. II: DNA sequence of the ULR from class II spliced transcripts. III: DNA sequence of the ULR from class I spliced transcripts. Nt, nucleotides.

of the ATG initiation codon to ATA (mutant line 23-7). Three independent null mutant lines [23-2(1), 23-3, and 23-5 in Fig. 2A] had a single point mutation changing the G of the AG 3′ splice site to A (Fig. 1B). Northern (RNA) blot analysis of poly(A)$^+$ RNA extracted from these three independent mutants showed steady-state concentrations of *rolA* mRNA that were identical to that of the original *rolA* transgenic line (Fig. 2B), which demonstrates that the transition from G to A does not affect the accumulation of *rolA* transcripts. To rule out the possibility that genes other than *rolA* were mutated, the region from −100 to +330 from each mutant allele was expressed in transgenic tobacco plants under the control of the 35S promoter (*7*). When expressed in tobacco, the chimeric genes derived from the mutants did not show the spliced mRNA band (*7*).

Single base mutations in the consensus sequence at the 3′ splice site of introns are known either to block or to drastically curtail RNA splicing (*10*). Accordingly, RT-PCR done with mRNA extracted from *rolA* transgenic *Arabidopsis* (line 23) and its null mutant lines 23-2(1), 23-3, and 23-5 showed that the unspliced form was present in all cases analyzed, but the spliced product was missing in the null mutants (Fig. 1A). Thus, these three null mutants are defective in *rolA* pre-mRNA splicing. The failure in expressivity of the traits could be due to sequestration of unspliced mRNAs to the nucleus, making them unavailable for translation (*11*). However, because mRNAs from genes without introns or with alternative splicing are transported to the cytoplasm, splicing is not a strict requirement for production of cytoplasmic mRNAs (*12*). Furthermore, it has been shown (*13*) that a small proportion of pre-mRNAs are translated in vivo and that deletion of sequences important for spliceosomal assembly causes transport of pre-mRNA to the cytoplasm. Therefore, splicing might be required for efficient translation of *rolA* mRNA. This hypothesis was tested by translating in vitro poly(A)$^+$ RNA extracted from *Arabidopsis* plants transgenic for the *rolA* gene and from its splicing-defective mutants. A peptide with the expected apparent molecular mass (14 kD) (Fig. 3) is detectable among the immunoprecipitated translation products of

Fig. 2. (**A**) Phenotypes of *Arabidopsis* plants. Wild type (left), *rolA* transgenic line 23 (middle), and null mutant line 23-5 (right). Mutant lines 23-2(1) and 23-3 had a phenotype identical to the ones of 23-5 and of wild-type plants. (**B**) Northern blot analysis of poly(A)$^+$ RNA extracted from *Arabidopsis* plants. Lane 1, null mutant line 23-5; lane 2, null mutant line 23-3; lane 3, null mutant line 23-1; lane 4, line 23; and lane 5, wild type.

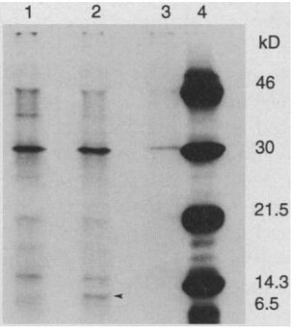

Fig. 3. Polyacrylamide gel electrophoresis (17.5%) showing the products, immunoprecipitated with affinity-purified antibody to *rolA* and protein A–Sepharose, of an in vitro translation in a wheat germ system of poly(A)$^+$ RNA extracted from *Arabidopsis* null mutant line 23-5 (lane 1) and from line 23 (lane 2) (*23*). Lane 3, wild type. Lane 4, Rainbow ^{14}C methylated protein marker (Amersham). Approximately equal amounts of protein (wild type, 1 million cpm; line 23, 1.3 million cpm; line 23-5, 1.5 million cpm) were immunoprecipitated (*23*). The gel was treated with Enhancer and exposed for 40 days. The arrow in lane 2 indicates protein produced by *rolA*.

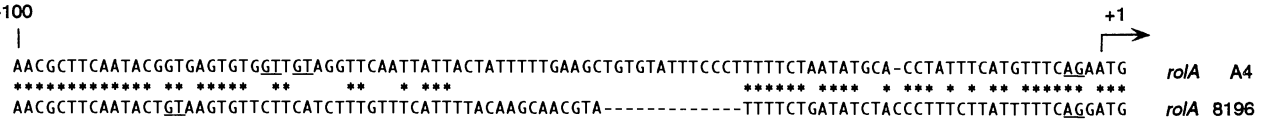

Fig. 4. DNA sequence comparison of the ULR of the *rolA* genes from Ri plasmid A4 and Ri plasmid 8196 (GenBank accession number M60490). Experimentally tested (A4) and putative (8196) 5′ GT and 3′ AG splicing sites are underlined.

poly(A)$^+$ RNA extracted from the *rolA* transgenic plants (line 23) (Fig. 3) but not from the null mutant lines 23-5 (Fig. 3), 23-2(1) (*14*), and 23-3 (*14*). Thus, in a wheat germ translation system, *rolA* pre-mRNA is either not translated at all, or it is translated at a level below detection limit (that is, at least 10 to 20 times less than is *rolA* mRNA from line 23). A 90% reduction in the level of expression of the *rolA* gene is sufficient to restore normal growth in transgenic tobacco plants (*4*).

Our molecular and genetic analysis has shown that: (i) The *rolA* gene from the Ri plasmid A4 contains an intron in its transcribed but untranslated leader region. (ii) This intron is spliced by use of a unique 3′ AG site and at least two alternative GT 5′ sites. (iii) Mutations in the 3′ splice site are null mutants in *Arabidopsis*, which indicates that splicing is required for expression of the *rolA* gene. (iv) Unspliced *rolA* pre-mRNA is translated at least 10 to 20 times less efficiently than is spliced *rolA* mRNA in an in vitro translation system.

Sequence comparison of the ULR from *rolA* genes from two different strains of *A. rhizogenes* (Fig. 4) shows conservation of the AG 3′ splice site. A putative 5′ splice site is found 70 bases upstream of the AG 3′ splice site in a highly conserved region. Thus, the sequence comparison indicates that the biological importance of the *rolA* intron is most likely not limited to the *rolA* gene from the Ri plasmid A4. Unspliced *rolA* pre-mRNA would probably be sequestered in the nucleus and therefore not be available for translation. However, the failure to translate *rolA* pre-mRNA in an in vitro system suggests that even if transported to the cytoplasm, *rolA* pre-mRNA will not be efficiently translated in vivo. Reduced translation of unspliced *rolA* pre-mRNA might result from the inhibitory effect of two out-of-frame ATGs upstream of the initiation codon, in a manner similar to that previously reported (*15*). These ATGs are part of the intron and are therefore no longer present in the spliced mRNA.

Introns of the self-splicing group I type have been found in the mRNAs of bacteriophages, such as T4 and SPO1, and they are relatively common in the tRNA genes of bacteria, including *Agrobacterium* (*16*). Introns of the self-splicing group II have been found in chromosomal bacterial genes (*16*). Pre-mRNA splicing, however, has not been observed in prokaryotes, although eubacterial components similar to small nuclear ribonucleoproteins have been found in *Synechococcus leopoliensis* and *Bacillus subtilis* (*17*).

The presence of pre-mRNA splicing in a gene derived from a bacterial plant pathogen shows that *A. rhizogenes* uses, as part of its strategy to induce hairy-root disease, a mechanism of posttranscriptional control (pre-mRNA splicing) that is so far restricted to eukaryotic genes (*18*). The data presented show that *rolA* gene expression and splicing take place in plant cells; however, we cannot exclude the possibility that the *rolA* gene is also expressed in the bacterial cell.

The *rolA* pre-mRNA intron might have evolved from prokaryotic DNA (*18*). Alternatively, the finding that the intron present in the pre-mRNA of the *rolA* gene has features typical of plant introns might be interpreted as suggesting that either the *rolA* gene or its intron is of plant origin. Although intron mobility is known to result from mechanisms such as homing and transposition (*18*), there is no independent evidence of transfer of genetic information from plants to bacteria.

REFERENCES AND NOTES

1. F. F. White, B. H. Taylor, G. A. Huffman, M. P. Gordon, E. W. Nester, *J. Bacteriol.* **164**, 33 (1985); J. L. Slightom, M. Durand-Tardif, L. Jouanin, D. Tepfer, *J. Biol. Chem.* **261**, 108 (1986).
2. G. Schröder *et al.*, *EMBO J.* **2**, 403 (1983); A. Janssens, G. Engler, P. Zambryski, M. Van Montagu, *Mol. Gen. Genet.* **195**, 341 (1984).
3. V. Sinkar, F. Pythoud, F. F. White, E. W. Nester, M. P. Gordon, *Genes Dev.* **2**, 688 (1988); T. Michael and A. Spena, in Agrobacterium *Protocols, Methods in Molecular Biology*, K. Gartland and M. Davey, Eds. (Humana Press, Clifton, NJ, in press).
4. M. Carneiro and F. Vilaine, *Plant J.* **3**, 785 (1993).
5. C. Dehio and J. Schell, *Mol. Gen. Genet.* **241**, 359 (1993).
6. C. Csank, F. M. Taylor, D. W. Martindale, *Nucleic. Acids Res.* **18**, 5133 (1990).
7. A. Magrelli, K. Langenkemper, C. Dehio, J. Schell, A. Spena, unpublished results.
8. G. J. Goodall and W. Filipowicz, *Plant Mol. Biol.* **14**, 727 (1990).
9. ———, *Cell* **58**, 473 (1989).
10. P. A. Sharp, *Science* **235**, 766 (1987); *Cell* **77**, 805 (1994).
11. J. A. Steitz, *et al.*, in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*, M. Birnstiel, Ed. (Springer-Verlag, Heidelberg, Germany, 1988), pp. 115–158.
12. R. A. Padgett, P. J. Grabowski, M. M. Konarska, S. Seiler, P. A. Sharp, *Annu. Rev. Biochem.* **55**, 1119 (1986).
13. P. Legrain and M. Rosbash, *Cell* **57**, 573 (1989).
14. A. Magrelli, K. Langenkemper, C. Dehio, J. Schell, A. Spena, data not shown.
15. M. Kozak, *J. Cell. Biol.* **115**, 887 (1991).
16. F. Chu, G. F. Maley, F. Maley, M. Belfort, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3049 (1984); B. M. Sjoberg *et al.*, *EMBO J.* **5**, 2031 (1986); H. Goodrich-Blair, V. Scarlato, J. M. Gott, M. X. Xu, D. A. Shub, *Cell* **63**, 417 (1990); M. Q. Xu, S. D. Kathe, H. Goodrich-Blair, S. A. Nierzwicki, D. A. Shub, *Science* **250**, 1566 (1990); B. Reinhold-Hurek and D. A. Shub, *Nature* **357**, 173 (1992); J. Ferat and F. Michel, *ibid.* **364**, 358 (1993).
17. S. A. Kovacs *et al.*, *J. Bacteriol.* **175**, 1871 (1993).
18. A. M. Lambowitz and M. Belfort, *Annu. Rev. Biochem.* **62**, 587 (1993).
19. T. Maniatis, E. F. Frisch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).
20. PCRs were done with the use of complementary DNAs reverse-transcribed from 1 μg of poly(A)$^+$ RNA as templates (primer dT 15mer), and amplification used primer a (5′-AACGCTTCAATACGGTGA-3′) and primer b (3′-CTTTGTTTGGATGCCCTA-5′) (Fig. 1B).
21. Poly(A)$^+$ RNA was extracted with DynaBeads (Dynal) according to supplier instructions. Agarose-formaldehyde gel (1.2%) electrophoresis, Northern blot analysis, and hybridization were done according to standard protocols (*19, 20*). The blot was hybridized to a probe spanning the whole coding region of the *rolA* gene.
22. Direct sequencing of purified products (Fig. 1A) was done with a femtomole sequencing kit (Promega) with the use of primer a. Automated DNA sequence analysis (ALF and ABI) was done on 12 independent clones obtained by subcloning RT-PCR products corresponding to spliced mRNAs in pBluescript II KS$^+$. Primers used in RT-PCR were a and c (primer c: 3′-TCCACACAGCAACAACAACTAT-5′) (Fig. 1B).
23. In vitro translation was done with a translation kit (Boehringer Mannheim) according to the instructions provided by the supplier. After determination of the incorporated (acid-insoluble) radioactivity, approximately equal amounts (1 to 1.5 million cpm.) of proteins were incubated in 100 μl of homogenizing buffer [90 mM Hepes (pH 7.5), 30 mM dithiothreitol, and 2% SDS] at 95°C for 5 min. After centrifugation, the supernatant was diluted with 1.3 ml of renaturation buffer [50 mM Hepes (pH 7.5), 5 mM EDTA, 150 mM NaCl, and 1% Triton X-100] supplemented with leupeptin (20 μg/ml) and phenylmethylsulfonyl fluoride (1 mM). In order to adsorb unspecific binding proteins, 10 mg of protein A–Sepharose in phosphate-buffered saline was added, the mixture was incubated for 1 hour, and the protein was removed by centrifugation. The supernatant was incubated in a mobicolumn (MOBITEC, Göttingen, Germany) with 50 μg of affinity-purified anti-*rolA* immunoglobin G for 12 hours at 4°C. Subsequently, 10 mg of protein A–Sepharose was added and the mixture was incubated for 1 hour. After centrifugation, the pellet retained on the filter was washed with renaturation buffer supplemented with 0.1% SDS. Loading buffer (50 μl) was added to each sample, and the protein was extracted at 95°C for 5 min and then loaded on a 17.5% protein gel. The gel was fixed, treated with Enhancer (NEN, Boston, MA), and exposed for 40 days.

24. We thank R. Heynkes and B. Fiala for DNA sequencing and T. Michael (Norwich) for critical reading of this manuscript. A.S. wishes to thank T. Hohn (Basel) for his encouragement in regard to the *rol* intron hypothesis.

14 July 1994; accepted 14 October 1994