Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9534–9537, December 1990
Cell Biology

A bacterial peptide acting as a plant nuclear targeting signal: The amino-terminal portion of *Agrobacterium* VirD2 protein directs a β-galactosidase fusion protein into tobacco nuclei

(protein transport/T-complex/transformation/virulence genes)

A. HERRERA-ESTRELLA, M. VAN MONTAGU, AND K. WANG*

Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium

Contributed by M. Van Montagu, September 10, 1990

ABSTRACT ***Agrobacterium tumefaciens* is a soil bacterium capable of transferring DNA to the genome of higher plants. Of the virulence region-encoded proteins of the tumor-inducing (Ti) plasmid of *A. tumefaciens*, the VirD1 and VirD2 proteins are essential for T-DNA transfer to plant cells. These two proteins have been shown to be directly responsible for the formation of T-strands. VirD2 was also shown to be firmly attached to the 5′ termini of T-strands; these facts have led to its postulation as a pilot protein in the T-DNA transfer process and as a nucleus-targeting signal in plants. We have constructed a chimeric gene by fusing the *virD2* gene and the *Escherichia coli lacZ* gene. Cell fractionation and electron microscopy studies with transgenic tobacco plants containing the VirD2–LacZ fusion protein indicate that the first 292 amino acids of VirD2 are able to direct the cytoplasmic protein β-galactosidase to the plant nucleus. This provides an example of cross-kingdom nuclear localization between two free-living organisms: a bacterial peptide is capable of acting as a eukaryotic (plant) nuclear targeting signal.**

Agrobacterium tumefaciens is a soil bacterium capable of transferring DNA to the genome of higher plants (1). Among the virulence region-encoded proteins of its tumor-inducing (Ti) plasmid, only two polypeptides, VirD1 and VirD2, encoded by the 5′ half of the *Agrobacterium virD* locus, are known to participate directly in the generation of T-strand DNA molecules (1–4). Further insight into these polypeptides has shown (5, 6) that VirD1 possesses a DNA-relaxing activity and the VirD2 is a strand-specific and sequence-specific endonuclease. Recent studies (7–11) indicate that the VirD2 product becomes tightly associated with the 5′ termini of T-strand and nicked molecules. All of the processes summarized above have evolved to produce a structure, the T-DNA transfer complex, capable of penetrating bacterial and recipient plant cell plasma membranes and possibly cell walls to introduce new genetic information into the plant genome. Major functions that may be attributed to the proteins of the transfer complex (encoded by the *virD* and *virE* operons) include protection of the T-strand molecule (8), recognition of transmembrane routes from the bacterium to the plant cell (1), targeting of the transfer complex to the plant nucleus, and efficient integration of the transferred DNA into the plant genome.

MATERIALS AND METHODS

Plasmid Construction and Plant Transformation. Plasmid pTR-lacZ was constructed as follows. A 3.1-kilobase *Nco* I–*Bam*HI fragment containing the *Escherichia coli lacZ* gene (a gift from T. Teeri, University of Helsinki, Finland) was cloned behind the plant promoter T_R2' (12) at the *Xba* I site in vector pAH1. pAH1 is a derivative of pC27-3 (13) in which an ATG codon within polylinkers was deleted. The selectable marker gene *nptII*, encoding neomycin phosphotransferase II, is driven by the other half of this dual promoter (T_R1'). To construct pTR-vd-lacZ, the same *lacZ* fragment was first translationally fused to an 881-base-pair *Xmn* I fragment carrying 65% of the 5′ portion of the *virD2* gene. This chimeric gene was then cloned at the *Xba* I site in pAH1 as in pTR-lacZ.

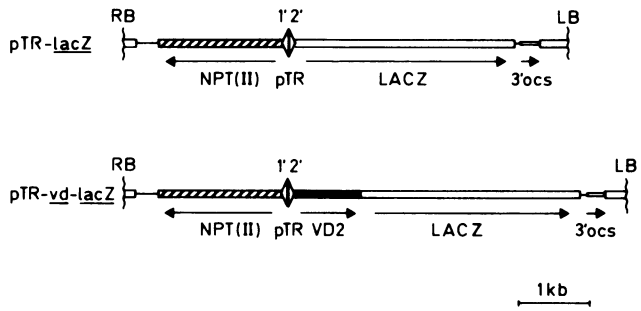

FIG. 1. Constructs used to determine the function of the VirD2 polypeptide in nuclear localization. LB and RB, left and right borders of the T-DNA; 3′ocs, 3′ terminator of the octopine synthase gene; kb, kilobase.

Plasmids pTR-lacZ and pTR-vd-lacZ (see Fig. 1) were mobilized into *A. tumefaciens* harboring pGV2260 (14) by triparental crosses using pRK2013 (15) as a helper plasmid. Leaf disc of *Nicotiana tabacum* (W38) was used for plant transformation as described (16).

Isolation of Nuclei and Assay of β-Galactosidase Activity in Subcellular Fractions. Nuclei were isolated according to Mazzolini *et al.* (17). Samples were treated with 2% glutaraldehyde prior to the assay. Activity was measured using *o*-nitrophenyl β-D-galactoside as substrate, as described by Miller (18) except that the pH of the buffer was 7.4.

For the electrophoretic analysis, equivalent amounts of protein from the three subcellular fractions (nuclear, organellar, and cytoplasmic) of the TR-VD-LACZ and TR-LACZ transgenic plants were subjected to SDS/PAGE (19). The gel was then used to assay for β-galactosidase activity as described (20).

Protease Protection. Nuclei of the TR-VD-LACZ and TR-LACZ transgenic plants were incubated with Pronase (125 μg/ml; Calbiochem-Behring) in 10 mM Tris·HCl, pH 7.6/2 mM $CaCl_2$ either with 0.2 M sucrose (intact nuclei) or without sucrose (and vigorously mixed with a Vortex to break the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: X-Gal, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside.

*Present address: Garst Seed Company, Research Department, Box 500, Slater, IA 50224.

nuclei) for 3 hr at 37°C. Untreated nuclei were incubated in the same conditions but no Pronase was added.

Electron Microscopy. Tissue sections and nuclei were reacted with the chromogenic substrate 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) after fixation with 2% glutaraldehyde (20) and were embedded in LR white acrylic resin (London Resin, London) (T. Hürek, unpublished work). β-Galactosidase activity is detected by the formation of electron-dense crystals of indigo when X-Gal is used as substrate.

RESULTS

A major obstacle in studying the role of the VirD2 protein at late stages in the T-DNA transfer process is the very limited amount of this protein associated with the T-strand (T-complex) which is theoretically transferred to the plant cell. We approached this problem by introducing and expressing the *virD2* gene in the plant. Fig. 1 illustrates the plasmid constructs used for plant transformation. A *virD2–lacZ* chimeric gene was cloned behind the wound-inducible plant promoter T_R2' (12, 20) to generate plasmid pTR-vd-lacZ. The control construct pTR-lacZ shares all the features of pTR-vd-lacZ except for the *virD2* gene fused to the *lacZ* gene. The *lacZ* gene (encoding β-galactosidase) was chosen for this purpose due to its versatility as a reporter gene (20) and its proven value in studies of protein accumulation in nuclei (21, 22).

Transgenic plants carrying *lacZ* or the *virD2–lacZ* fusion were induced to express the corresponding proteins by preparing protoplasts. Nuclei were isolated from the protoplasts. The various cellular fractions were assayed for β-galactosidase activity. As shown in Table 1, equivalent amounts of β-galactosidase activity were found in the cytosolic fraction of plants transformed with pTR-vd-lacZ and pTR-lacZ. However, a striking difference was observed when the amounts of activity accumulated in the nuclear fraction of the two plants were compared. The specific activity accumulated in the nuclei of the TR-VD-LACZ plants was ≈3-fold higher than that found in the cytosol, whereas in the TR-LACZ plant the specific activity in the nuclei was about 5-fold lower than that of the cytosolic fraction. Almost no activity was found in the organellar fractions (chloroplasts, mitochondria, peroxisomes, etc.) of both plants. These data strongly indicate that the 292 amino-terminal amino acids from VirD2 promote the localization of β-galactosidase activity in the nucleus and not in any other organelle.

Protein samples from the different cellular fractions were used to analyze the pattern of β-galactosidase activity after separation by SDS/PAGE. A very distinct band of enzymatic activity was detected in the nuclear fraction of the TR-VD-LACZ plant (Fig. 2, lane 2). This band, corresponding to the size predicted for the VirD2–LacZ fusion protein, was still present after the nuclei were washed with 1 M NaCl (data not shown), indicating its strong association with them. A weakly active band was also observed in the organellar fraction (lane 4). This was very likely due to contamination with VirD2–LacZ released from broken nuclei during isolation. As expected, no detectable activity was found in the nuclear (lane 1) and organellar (lane 3) fractions of the TR-LACZ transgenic plant. Both plants showed active bands in their cytosolic fraction (lanes 5 and 6). The β-galactosidase-active band detected in the TR-LACZ plant corresponded to that of the bacterial enzyme (lane 5 vs. lane 8). Interestingly, a smear of activity was detected in the cytosolic fraction of the TR-VD-LACZ plant (lane 6); its highest band corresponded with the position of the VirD–LacZ fusion protein whereas the lowest had an electrophoretic mobility similar to that of the bacterial LacZ. This could be explained if the VirD2 portion of the protein is especially sensitive to degradation by protease when present in the cytosol and is protected when it migrates into the nucleus. This hypothesis would also explain the high level of activity detected in the cytosol of the TR-VD-LACZ plant (Table 1). That no endogenous LacZ activity could be detected in any of the fractions is due to inactivation during the fractionation procedure (data not shown).

Table 1. β-Galactosidase activity in subcellular fractions of TR-LACZ and TR-VD-LACZ plants

Fraction	Activity, units	
	TR-LACZ	TR-VD-LACZ
Cytosolic	526.98	377.98
Organellar	34.85	72.62
Nuclear	112.00	1036.41

Activity was measured as described in *Materials and Methods*; 1 unit produces 1 nmol of nitrophenol produced per mg of protein per min. Organellar fraction included mitochondria, chloroplasts, peroxisomes, etc.

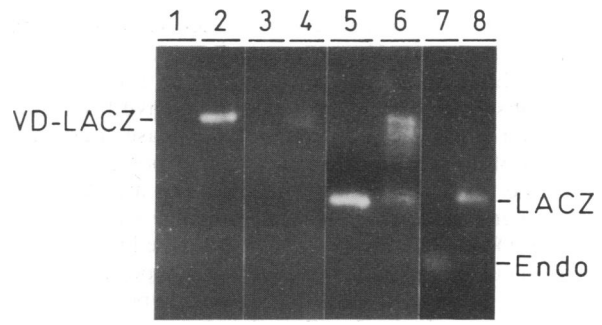

FIG. 2. Electrophoretic analysis of β-galactosidase activity in subcellular fractions. Lanes 1 and 2, nuclear fractions; lanes 3 and 4, organellar fractions; lanes 5 and 6, cytosolic fractions. Even numbers represent the fractions obtained from TR-VD-LACZ plants, the odd numbers those of the TR-LACZ plants. Total protein of an untransformed tobacco plant (lane 7) and of an *E. coli* strain carrying the bacterial β-galactosidase gene (lane 8) was used as reference. VD-LACZ, the VirD2–LacZ hybrid protein; LACZ, the bacterial β-galactosidase; Endo, the endogenous β-galactosidase in total extracts of a nontransformed tobacco plant.

Protease protection is widely used to demonstrate that proteins have fully traversed membranes from one cellular compartment into another. Using this approach, we found that the β-galactosidase activity associated with nuclei was partially protected from degradation by Pronase (Table 2). When isolated nuclei of the TR-VD-LACZ plant were Pronase-treated in a buffer containing an osmoprotectant (intact nuclei), only 36% of the original activity was lost after the

Table 2. β-Galactosidase activity of protease-treated nuclei

Nuclei	Activity	
	Units	% retained
TR-VD-LACZ		
Untreated	424.33	100.00
Intact	270.89	63.83
Broken	153.43	36.15
TR-LACZ		
Untreated	63.80	100.00
Intact	22.33	35.00
Broken	23.92	37.50

Nuclei of the TR-VD-LACZ and TR-LACZ transgenic plants either were treated with protease directly after isolation (intact), or had their membranes disrupted prior to treatment (broken), or did not receive the protease treatment (untreated) (see *Materials and Methods*). β-Galactosidase activity was assayed and is expressed as described in Table 1.

treatment. Nuclei resuspended in a buffer without osmoprotectant and broken by Vortex mixing lost 64% of the original activity. The incomplete protection of the β-galactosidase activity observed with intact nuclei may have been due to penetration of the nuclear envelope by Pronase, a mixture of proteases of undetermined mass. Another possibility is that part of the fusion protein might have been attached to the nuclear membrane and thus have remained exposed to the protease. On the other hand, the incomplete loss of activity in broken nuclei may have been due to the reported tolerance of LacZ to proteolytic attack (23). This hypothesis is supported by data obtained with isolated nuclei of TR-LACZ plants (Table 2), where the low activity associated with this fraction was reduced to 35% and 37.5% of the original when treated with Pronase before and after breakage, respectively.

Electron microscopy studies of tissue sections of the transgenic plants showed that the fusion protein from the TR-VD-LACZ plant was located in the nucleus (Fig. 3*A*), as suggested by the protease protection and cell fractionation experiments. However, an important part of the β-galactosidase activity appeared to have accumulated in the nuclear membrane of both the control TR-LACZ plant and the TR-VD-LACZ plant (Fig. 3 *A* and *B*). The nonspecific accumulation of indigo precipitate at the nuclear envelope has been observed in other cases (T. Teeri, personal communication). It is not known whether this phenomenon is due to accumulation of β-galactosidase at the nuclear membrane or to accumulation of the reaction product formed in the cytoplasm. Since the cell fractionation experiments had shown 10-fold higher β-galactosidase activity in nuclear fractions of the TR-VD-LACZ plant than in nuclear fractions of the control TR-LACZ plant, further electron microscopy studies were carried out using isolated nuclei. In nuclei from the TR-VD-LACZ plant, the indigo precipitate was found both associated with the nuclear envelope and in the nucleoplasm (Fig. 3*C*). In contrast, no reaction product was found in nuclei of the control plant, either in the membrane or in the nucleoplasm (Fig. 3*D*). These data suggest that the association of β-galactosidase activity with the nuclear envelope observed in complete cells (Fig. 3 *A* and *B*) was most likely nonspecific and due to manipulation. Although a considerable amount of precipitate was detected in the cytosol of both plants, no specific accumulation of activity was observed in any other organelle from either transgenic plant. No indigo precipitate was detected in untransformed control plants (data not shown).

DISCUSSION

The experiments reported here demonstrate the ability of a truncated VirD2 protein from *A. tumefaciens* to direct nuclear localization of the bacterial protein β-galactosidase in plants. The ability of peptides to target proteins into the nucleus has been extensively studied in animal cells and yeast (24–26). Recent studies have led to a proposed consensus signal sequence for nuclear targeting, Lys-(Lys or Arg)-Xaa-(Lys or Arg) (27). Analysis of the VirD2 protein sequence revealed the presence of three conserved peptides in *A. tumefaciens* and *A. rhizogenes* that resemble the proposed signal (27). Two of these peptides are located at the carboxyl-terminal part of the nopaline (pTiC58) VirD2 protein, at positions 417–420 (Lys-Arg-Pro-Arg) and 431–434 (Lys-Arg-Glu-Arg). The third potential signal is located at positions 31–34 (Arg-Lys-Gly-Arg). The truncated protein used in this work contained only the putative signal from the amino-terminal part of VirD2 (positions 31–34). This peptide alone is capable of promoting accumulation of β-galactosidase activity in the nuclei. Although the putative signal sequence (Arg-Lys-Gly-Arg) contained in this peptide has an arginine at the first position instead of the lysine proposed for the

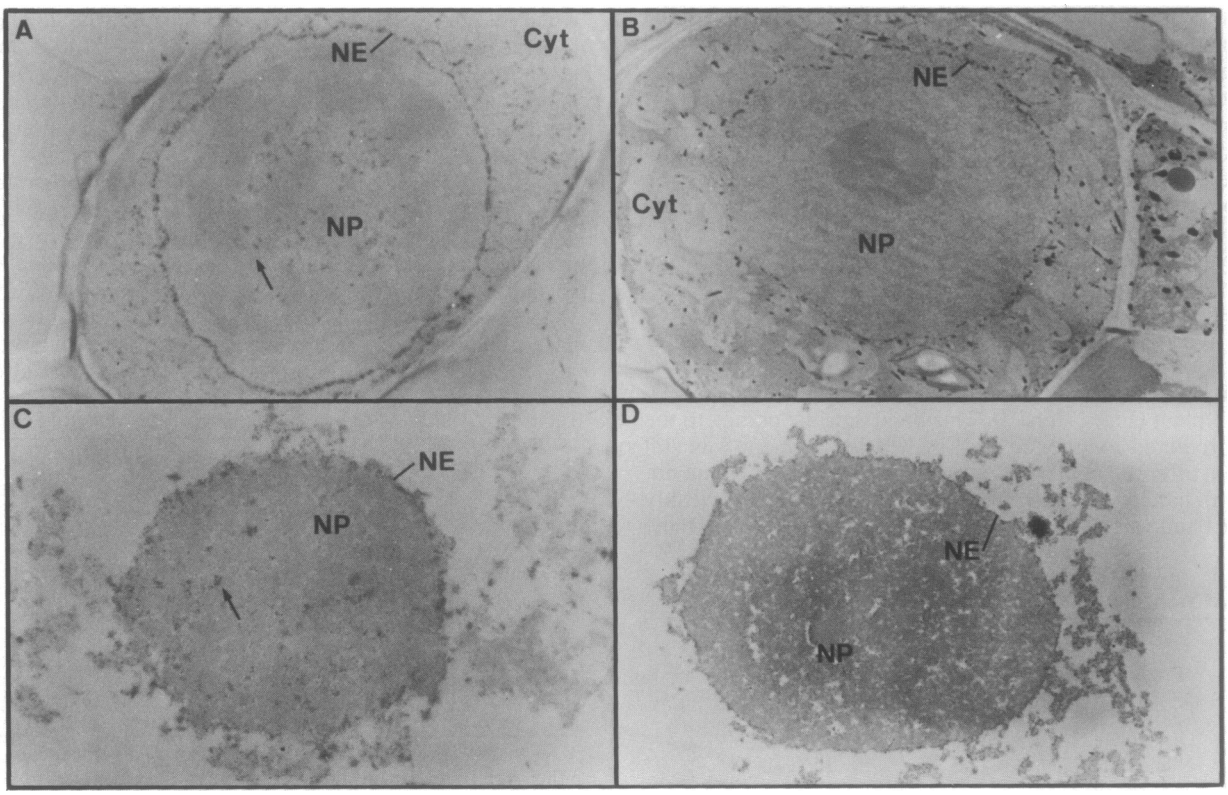

FIG. 3. Electron microscopic analysis of β-galactosidase activity in tissue sections of TR-VD-LACZ (*A*) and TR-LACZ (*B*) transgenic plants and in isolated nuclei from TR-VD-LACZ (*C*) and TR-LACZ (*D*) plants. Examples of the indigo-precipitated crystals formed by hydrolysis of the β-galactosidase substrate X-Gal are indicated by arrows. Cyt, cytoplasm; NE, nuclear envelope; NP, nucleoplasm.

consensus signal described for animal cells, it has been shown for the simian virus 40 targeting signal that such a substitution yields a peptide that is active as a nuclear targeting signal (28). Our data do not eliminate the possibility that the remaining two postulated signals at the carboxyl terminus could act as auxiliary nuclear targeting signals. In a similar way, the polyoma virus large tumor antigen employs two different peptides to direct nuclear accumulation (29). Moreover, increasing the number of signals per molecule increases the relative uptake by nuclei of albumin–peptide conjugates and the effective size of the channels available for translocation in animal cells (30). Therefore, the presence of three putative nuclear targeting signals in the VirD2 protein from *A. tumefaciens* might lead to a higher efficiency of transport through the nuclear envelope, which presumably would be required for the transport of the T-strand.

Our data, together with other recent observations, allow us to suggest a sequence of events leading to T-DNA transfer from bacteria to plants. It is initiated by the T-DNA border cleavage by the *virD*-encoded endonuclease activity (3, 4, 31). The second step is most probably the replication of the bottom strand in a polar way, 5′ to 3′ from the nicked border, thereby producing a single-stranded molecule (2) with VirD proteins associated at the 5′ terminus (7–11). This T-complex subsequently associates with nonspecific single-stranded DNA-binding proteins, such as VirE2 (32, 33), resulting in the formation of an intermediate form. These proteins would protect the T-strand from attack by nucleases (11) and might participate in active transport through the cytoplasmic membrane in conjunction with membrane-associated proteins such as those encoded by the *virB* loci (34, 35). The presence of the VirD2 protein at the 5′ terminus of the T-complex molecules, its function as a nuclear targeting signal, and its endonuclease activity strongly implicate its involvement in subsequent T-DNA integration into the plant genome. These data are supported by the fact that after integration, the T-DNA molecules have well-defined borders at the right-hand end of the T-DNA but not at the left (36–38).

Further investigation is required to determine whether the postulated signal peptide (Arg-Lys-Gly-Arg) is actually responsible for the nuclear accumulation that has been observed and whether the carboxyl-terminal part of the VirD2 protein plays a role in targeting to nuclei. A question of major interest is whether the VirD2 protein is responsible for promoting T-DNA transport into the plant nucleus and whether it plays a role in integration of the T-DNA into the plant nuclear genome.

We thank Drs. A. Caplan, C. Castresana, and L. Herrera-Estrella for critical reading of the manuscript and M. De Cock, V. Vermaercke, S. Van Gijsegem, and K. Spruyt for preparation of the manuscript and figures. We are grateful to R. De Rijcke for the excellent electron microscopy work and to Dr. T. Hürek for invaluable advice. A.H.-E. is indebted to the Commission of the European Communities for a training grant.

1. Zambryski, P. (1988) *Annu. Rev. Genet.* **22,** 1–30.
2. Stachel, S. E., Timmerman, B. & Zambryski, P. (1987) *EMBO J.* **6,** 857–863.
3. Yanofsky, M. F., Porter, S. G., Young, C., Albright, L. M., Gordon, M. P. & Nester, E. W. (1986) *Cell* **47,** 471–477.
4. Albright, L. M., Yanofsky, M. F., Leroux, B., Ma, S. & Nester, E. W. (1987) *J. Bacteriol.* **169,** 1046–1055.
5. Ghai, J. & Das, A. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3109–3113.
6. Soliman, M. & Das, A. (1990) *J. Bacteriol.*, in press.
7. Herrera-Estrella, A., Chen, Z.-M., Van Montagu, M. & Wang, K. (1988) *EMBO J.* **7,** 4055–4062.
8. Ward, E. R. & Barnes, W. M. (1988) *Science* **242,** 927–930.
9. Young, C. & Nester, E. W. (1988) *J. Bacteriol.* **170,** 3367–3374.
10. Howard, E. A., Winsor, B. A., De Vos, G. & Zambryski, P. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4017–4021.
11. Dürrenberger, F., Crameri, A., Hohn, B. & Koukolíková-Nicola, Z. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9154–9158.
12. Velten, J., Velten, L., Hain, R. & Schell, J. (1984) *EMBO J.* **3,** 2723–2730.
13. Alliotte, T., Zhu, L. H., Van Montagu, M. & Inzé, D. (1988) *Plasmid* **19,** 251–254.
14. Deblaere, R., Bytebier, B., De Greve, H., Deboeck, F., Schell, J., Van Montagu, M. & Leemans, J. (1985) *Nucleic Acids Res.* **13,** 4777–4788.
15. Van Haute, E., Joos, H., Maes, M., Warren, G., Van Montagu, M. & Schell, J. (1983) *EMBO J.* **2,** 411–418.
16. Budar, F., Deboeck, F., Van Montagu, M. & Hernalsteens, J.-P. (1986) *Plant Science (Lucknow, India)* **46,** 195–206.
17. Mazzolini, L., Vaeck, M. & Van Montagu, M. (1989) *Eur. J. Biochem.* **178,** 779–787.
18. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
19. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
20. Teeri, T. H., Lehväshlaiho, H., Franck, M., Uotila, J., Heino, P., Palva, E. T., Van Montagu, M. & Herrera-Estrella, L. (1989) *EMBO J.* **8,** 343–350.
21. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. (1984) *Cell* **39,** 499–509.
22. Hall, M. N., Hereford, L. & Herskowitz, I. (1984) *Cell* **36,** 1057–1065.
23. Ullman, A. & Perrin, D. (1970) in *The Lactose Operon*, eds. Beckwith, J. R. & Zipser, D. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 143–172.
24. Chelsky, D., Ralph, R. & Jonak, G. (1989) *Mol. Cell. Biol.* **9,** 2487–2492.
25. Richardson, W. D., Mills, A. D., Dilworth, S. M., Laskey, R. A. & Dingwall, C. (1988) *Cell* **52,** 655–664.
26. Silver, P. A., Keegan, L. P. & Ptashne, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5951–5955.
27. Wang, K., Herrera-Estrella, A. & Van Montagu, M. (1990) *J. Bacteriol.* **172,** 4432–4440.
28. Kalderon, D., Richardson, W. D., Markham, A. T. & Smith, A. E. (1984) *Nature (London)* **311,** 33–38.
29. Richardson, W. D., Roberts, B. L. & Smith, A. E. (1986) *Cell* **44,** 77–85.
30. Dworetzky, S. I., Lanford, R. E. & Feldherr, C. M. (1988) *J. Cell Biol.* **107,** 1279–1287.
31. Wang, K., Stachel, S., Timmerman, B., Van Montagu, M. & Zambryski, P. (1987) *Science* **235,** 587–591.
32. Gietl, C., Koukolíková-Nicola, Z. & Hohn, B. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 9006–9010.
33. Citovsky, V., De Vos, G. & Zambryski, P. (1988) *Science* **240,** 501–504.
34. Thompson, D. V., Melchers, L. S., Idler, K. B., Schilperoort, R. A. & Hooykaas, P. J. J. (1988) *Nucleic Acids Res.* **16,** 4621–4636.
35. Ward, J. E., Akiyoshi, D. E., Regier, D., Datta, A., Gordon, M. P. & Nester, E. W. (1988) *J. Biol. Chem.* **263,** 5804–5814.
36. Zambryski, P., Depicker, A., Kruger, K. & Goodman, H. (1982) *J. Mol. Appl. Genet.* **1,** 361–370.
37. Holsters, M., Villarroel, R., Gielen, J., Seurinck, J., De Greve, H., Van Montagu, M. & Schell, J. (1983) *Mol. Gen. Genet.* **190,** 35–41.
38. Slightom, J. L., Jouanin, L., Leach, F., Drong, R. F. & Tepfer, D. (1985) *EMBO J.* **4,** 3069–3077.