FEBS Letters 362 (1995) 116–120 FEBS 15293

DNA replication specificity of TYLCV geminivirus is mediated by the amino-terminal 116 amino acids of the Rep protein

Isabelle Jupin*, François Hericourt, Bianca Benz, Bruno Gronenborn

Institut des Sciences Végétales, CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette Cedex, France

Received 2 January 1995; revised version received 13 February 1995

Abstract Geminiviruses are plant DNA viruses replicating by a rolling circle mechanism. We have investigated the specificity of replication origin recognition of two different isolates of tomato yellow leaf curl virus (TYLCV). Here, we show that TYLCV-Sardinian and -Israeli replication proteins display a high degree of specificity for their respective origins. The DNA sequences recognized are located on the left part of the intergenic region whereas the amino-terminal 116 amino acids of the Rep protein determine the specificity of origin recognition.

Key words: Geminivirus; TYLCV; DNA replication; Rolling circle; Specificity of origin recognition

1. Introduction

Geminiviruses are a family of plant viruses with a small, circular, single-stranded DNA genome [1]. Their genome is replicated in the nuclei of infected cells via double-stranded intermediates by a rolling circle mechanism analogous to that used by ssDNA phages and plasmids [2,3]. Geminiviruses depend largely on host factors to mediate their transcription and replication. They encode a small number of overlapping ORFs, but only a single viral protein (Rep), encoded by the AL1 or C1 gene, is required for their replication [4,5]. The ORFs are arranged in two divergent clusters separated by an intergenic region (IR) of about 300 nt. The intergenic region contains a GC-rich inverted repeat present in all geminiviruses, that has the potential to form a stem–loop structure [1]. The Rep protein shares no homology to any known DNA polymerase, but is related to proteins involved in the initiation of DNA replication of ssDNA bacterial plasmids [6]. It was recently shown that this protein possesses a nicking-closing activity and initiates rolling circle replication by a site-specific cleavage within the loop of the conserved structure [7]. In addition, the specific binding of AL1 protein to a sequence in the IR has been reported, and its binding site has been localized to the left side of the IR [8–10]. The interaction of Rep with this high affinity binding site is essential for viral replication [10] as well as for transcriptional repression of its own gene [11,12]. In spite of the functional conservation between the *trans*-acting replication factors encoded by geminiviruses, these proteins show specificity for replication of their cognate genome [13,14]. We have begun to analyse this replication specificity using two isolates of tomato yellow leaf curl virus (TYLCV). These viruses have a common functional organization and 77% overall sequence identity [15,16]. Here, we show that the Sardinian (-Sar) and Israeli (-Is) isolates of TYLCV exhibit strict specificity in the interaction between the *trans*-acting Rep protein and its *cis* targets, required for their replication in vivo. This specificity is determined by the first 116 amino acids of the Rep protein and the left part of the IR.

*Corresponding author.

Abbreviations: aa, amino acid; nt, nucleotide; ss, single-stranded; IR, intergenic region; ORF, open reading frame; BGMV, bean golden mosaic virus; SqLCV, squash leaf curl virus; TGMV, tomato golden mosaic virus; TYLCV, tomato yellow leaf curl virus.

2. Materials and methods

2.1. Plasmid constructs

The plasmid pTY-*Sst*14, containing the DNA of the Sardinian isolate of TYLCV cloned into the unique *Sst*I site of pUC118, has been described previously [15]. A modified version of pTY-*Sst*14 with a deletion in the polylinker was obtained by digestion with *Kpn*I and *Pst*I restriction enzymes, filling in with T4 DNA polymerase and re-ligation, to create pTYΔKP. The Israeli isolate of TYLCV cloned into the unique *Sph*I site of pTZ18, pTYH20.6, was kindly provided by H. Czosnek [16]. Nucleotide numbering is according to [15]. For the sake of clarity, the same numbering was adopted for the clone pTYH20.6, the first nucleotide of the conserved TAATATTAC being defined as base 1. All DNA manipulations were performed using standard techniques [17,18]. Site-directed mutagenesis was performed using uridinylated ssDNA [19].

2.1.1. Exchange of the IR. In order to exchange the intergenic regions of the two isolates, restriction sites were engineered on both sides of the IR. In pTYΔKP, a unique *Bam*HI restriction site naturally occurs at nt 152, 2 amino acids after the initiation codon of the ORF V1. A *Bsp*EI restriction site was introduced by site-directed mutagenesis at nt 2635, 4 amino acids downstream of the initiation codon of the ORF C1, using the oligonucleotide AATGCCAAGATCCGGACGTTTTAGTATC to create the clone 8D. Identical restriction sites were introduced at similar positions in the clone pTYH20.6 using the oligonucleotide CAACATGCCTCGTT**CCGGA**AAAATATATGCC that creates a *Bsp*EI site at nt 2617 and the oligonucleotide CATTAAGAAGTGGATCCCACATATTGC that introduces a *Bam*HI site at nt 160 to create the clone 1E.

With the exception of the *Bsp*EI site in the clone 1E that introduces 2 amino acid changes in the sequence of the TYLCV-Is Rep protein, replacing Leu^3 by a Ser and Phe^4 by a Gly, none of the other changes alter the amino acid sequence of the proteins encoded by C1 and V1 ORFs.

The clone 2O that corresponds to TYLCV-Sar coding regions with an -Is IR was obtained by exchanging the *Bsp*EI–*Bam*HI restriction fragment between 8D and 1E. This results in the production of a wild-type -Sar C1 ORF.

To obtain the reciprocical exchange corresponding to TYLCV-Is coding regions with a -Sar IR, the TYLCV-Is genome was first recloned in the *Sph*I site of a pUC118 vector where the *Bam*HI site from the polylinker had been first filled in with the Klenow fragment of DNA polymerase I and re-ligated, so that the resulting clone 9FI only contains the *Bam*HI site artificially introduced into the V1 ORF. The clone 5EN was then obtained by exchanging the *Bsp*EI–*Bam*HI fragments between 9FI and 8D. Due to the presence of the *Bsp*EI site, the clone 5EN contains 2 amino acids mutated compared to the wild-type TYLCV-Is C1 ORF. These 2 amino acid changes were switched back

SSDI 0014-5793(95)00221-9

to the wild-type TYLCV-Is C1 ORF sequence using the oligonucleotide AATGCCAAGATTATTTAAAATATAT to produce the clone 6FL.

To ensure that the mutagenesis procedure had not caused any further mutation, the intergenic regions of the clones 8D, 1E, 2O, 5EN and 6FL were entirely sequenced by the chain termination method using T7 DNA polymerase on double-stranded templates and specific primers.

2.1.2. Subdivision of the IR. To create clones with chimeric IR, we took advantage of the *Ssp*I site present in the conserved nonanucleotide sequence between the inverted repeats. Since this site is not unique in the TYLCV genome, we used the three-point ligation technique to assemble restriction fragments. Left and right halves of TYLCV-Sar IR were obtained by digestion of the clone 8D with the restriction enzymes *Bsp*EI and *Ssp*I, and *Ssp*I and *Bam*HI, respectively, and purification of the corresponding 172 and 152 nt fragments on 4% acrylamide gels. Similarly, left and right halves of TYLCV-Is IR were obtained from the clone 1E by digestion with the same pairs of enzymes and purification of the corresponding 175 and 158 nt fragments.

The clone 10FM was thus obtained by ligation between the 5579 nt *Bsp*EI–*Bam*HI fragment from 8D, the 175 nt *Bsp*EI–*Ssp*I fragment from 1E and the 152 nt *Ssp*I–*Bam*HI fragment from 8D. In a similar way, the clone 41EP was obtained by ligation between the 5579 nt *Bsp*EI–*Bam*HI fragment from 8D, the 172 nt *Bsp*EI–*Ssp*I fragment from 8D and the 158 nt *Ssp*I–*Bam*HI fragment from 1E.

The intergenic regions of the 10FM and 41EP clones were entirely sequenced by the chain termination method using T7 DNA polymerase on double-stranded templates.

2.1.3. Domain exchange within the Rep protein. In order to exchange domains of the Rep protein between the two isolates, restriction sites were first engineered in the ORF C1. In pTYΔKP, a unique *Bgl*II restriction site naturally occurs at nt 1523, in the ORF C2, downstream of the stop codon of the ORF C1. An *Xho*I restriction site was introduced by site-directed mutagenesis at nt 2242, 123 amino acids after the initiation codon of the ORF C1, using the oligonucleotide GGACGATCTGCTCGAGGAGGACAACAG to create the clone 9FQ. Identical restriction sites were introduced at similar positions in the clone pTYH20.6 using the oligonucleotide GCGTGTAGATCTAGACTGTGG that creates a *Bgl*II site at nt 1543, and the oligonucleotide GCAGATCAGCTCGAGGAGGTCAGC that introduces a *Xho*I site at nt 2265, 121 amino acids after the initiation codon of the ORF C1, to create the clone 6FP.

The clone 4FB was obtained by exchanging the *Xho*I–*Bgl*II restriction fragment between 9FQ and 6FP. 3FA was then obtained by exchanging the IR between 4FB and 2O using the *Bst*BI and *Bam*HI restriction sites.

2.2. Replication in tomato suspension cells

Protoplasts derived from a tomato suspension culture of *Lycopersicon esculentum*X *L. pennellii* were obtained as previously described [20]. The cloned viral DNA was linearized at the cloning site (*Sst*I for the constructs pTYΔKP, 2O, 10FM, 41EP, 3FA, 4FB, and *Sph*I for pTYH20.6 and 6FL) and 10 μg was used for each transfection of 2.5×10^5 protoplasts. Transfection and culture conditions were done according to [20].

2.3. Isolation and characterization of viral DNA forms

The isolation of DNA from protoplasts was done according to [21]. The presence of viral DNA was probed by Southern blotting on Hybond N (Amersham) using radiolabeled TYLCV DNA. Isolate-specific probes were obtained using fragments corresponding to the intergenic region of each isolate (fragment *Bsp*EI–*Bam*HI from 8D and 1E). For the constructs 10FM and 41EP, the 1371 nt fragment *Bgl*II–*Bam*HI from pTYΔKP covering the V1, V2 and C3 ORFs was used as a probe. After hybridization and washing under stringent conditions (0.2× SSC, 0.1% SDS at 65°C), the membrane was dehybridized and then rehybridized with a probe corresponding to the other isolate.

3. Results and discussion

Comparison of the nucleotide sequences of TYLCV-Sar and -Is revealed that these viruses share about 77% sequence identity. They display the same genetic organization, induce similar symptoms, and infect the same host plants [15,16]. However, it was not known whether TYLCV-Sar and -Is display selectivity in interaction between the *cis*- and *trans*-acting factors required for their replication.

To address this question, hybrid viruses containing the IR of the heterologous isolate were constructed. For this purpose, unique restriction sites *Bsp*EI and *Bam*HI were introduced on the right and left borders of the IR and the fragments corresponding to the IR were exchanged between the two isolates. This led to the clone 2O, corresponding to a TYLCV hybrid genome with all wild-type ORFs from TYLCV-Sar but an -Is IR (Fig. 1), and to the reciprocical clone 5EN. Due to the introduction of the *Bsp*E1 restriction site, 5EN does not encode for a wild-type -Is C1 ORF but contains two mutated amino acids, Leu^3 being replaced by Ser and Phe^4 by Gly. To rule out a potential effect of the mutations on the function of the Rep protein, the clone 6FL (Fig. 1), in which these amino acids were switched back to TYLCV-Is wild-type sequence, was constructed by site-directed mutagenesis. DNA of these constructs was introduced into protoplasts derived from tomato suspension cells and assayed for their capacity to replicate autonomously. Total DNA was isolated after a 7 day culture period and analyzed by Southern blot by sequentially probing with isolate-specific radiolabeled DNA under conditions that prevented cross-hybridization between the -Sar and -Is DNA sequences. Double- and single-stranded forms of replicated viral DNA were readily detected for TYLCV-Sar (pTYΔKP) and -Is (pTYH20.6) (Fig. 2A, lanes 1 and 4). In contrast, the hybrid constructs 6FL and 2O were not able to replicate (Fig. 2A, lanes 2 and 3).

These results established that TYLCV-Sar and -Is Rep proteins display a high degree of specificity for their respective origins, hence, these two viruses are an appropriate system to investigate the molecular basis of Rep-mediated replication specificity.

To more precisely delineate the *cis* elements necessary for that specificity, the IR was subdivided into two parts using the *Ssp*I restriction site located in the stem of the conserved putative stem–loop structure. The mutants 10FM and 41EP, both encoding wild-type TYLCV-Sar ORFs but containing hybrid intergenic regions (Fig. 1), were tested for their ability to replicate in the tomato protoplasts system. As shown in Fig. 2B, the mutant 41EP, in which the left half originated from TYLCV-Sar, replicated efficiently whereas 10FM did not. Since the DNA sequence forming the putative stem–loop structure is identical between the two TYLCV isolates it does not represent a specificity determinant, and the *cis*-acting sequences mediating replication specificity are located within the 146 nt sequence encompassing the left half of the IR.

These findings are in agreement with previous studies carried out with squash leaf curl virus (SqLCV) and tomato golden mosaic virus (TGMV), [13], or TGMV and bean golden mosaic virus (BGMV) [14] that revealed the incompatibility between the replication factors and a heterologous DNA component. These studies had localized the specificity determinant of the SqLCV replicon to a region of about 90 nt that encompasses the stem–loop structure and 60 nt of 5′ upstream sequence. Recently, Fontes et al. [14] demonstrated that the stem–loop structure does not contribute an element of specificity origin recognition, despite its requirement for replication. In addition, they showed that the sequence previously identified as the high affinity AL1 binding site [10] is necessary but not sufficient to direct specific origin recognition in vivo, and they suggest the

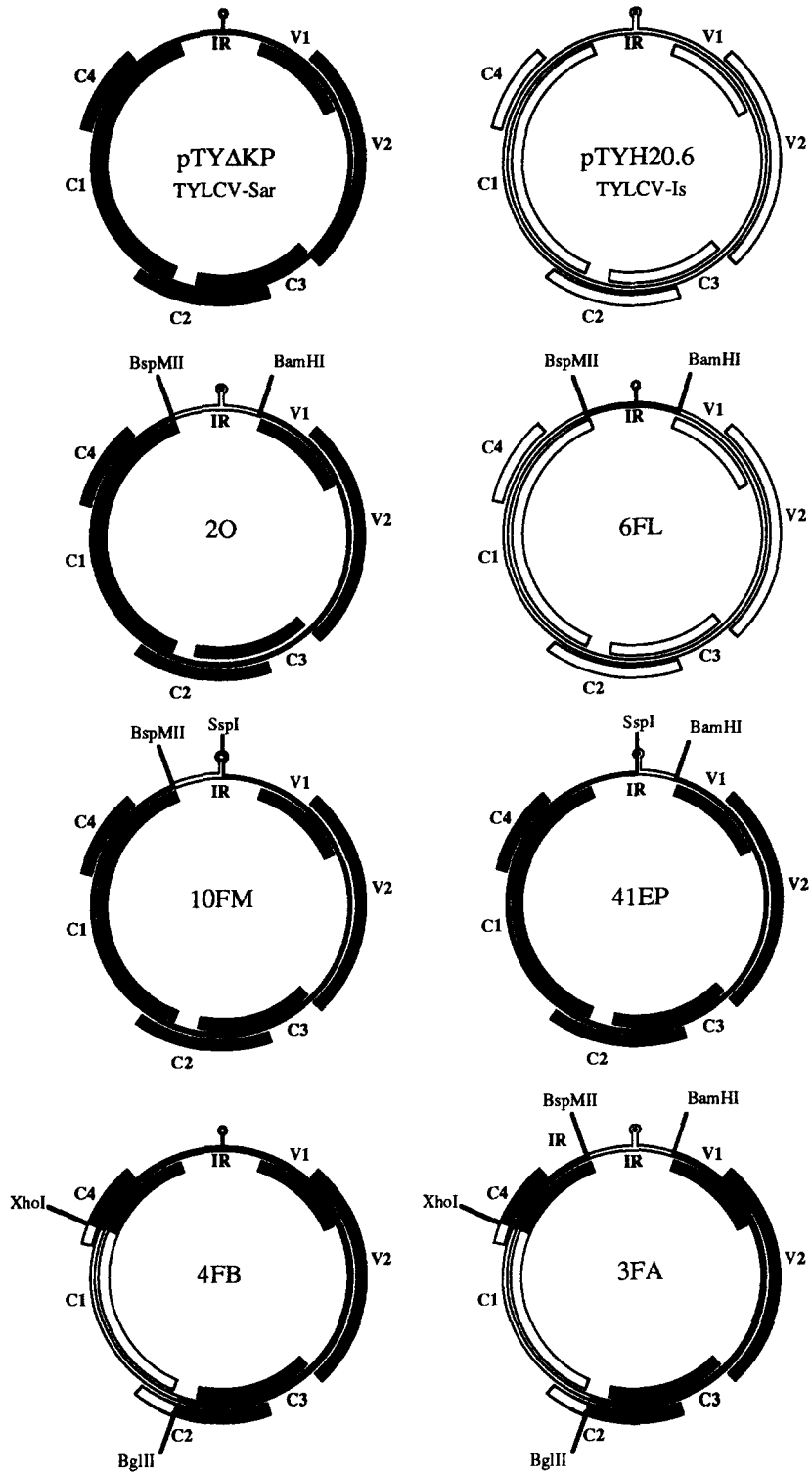

Fig. 1. Schematic diagrams of TYLCV strains and mutants. The position of the six ORFs are indicated along with restriction sites used to exchange DNA fragments. The TYLCV-Sar genome is indicated by black bars whereas the TYLCV-Is genome is indicated by open bars. The conserved stem–loop and the intergenic region (IR) are indicated.

presence of a second specificity determinant of replication located between the AL1 binding site and the stem–loop. More recently, Argüello-Astorga et al. [22] identified iterative sequence motifs arranged similarly between phylogenetically related groups of geminiviruses, and suggested their involvement as specific binding sites for the geminiviral replication-associated proteins. Additional experiments will be required to further define the *cis*-acting sequences required for specific in vivo recognition and precisely determine whether these iterative sequences participate in specificity.

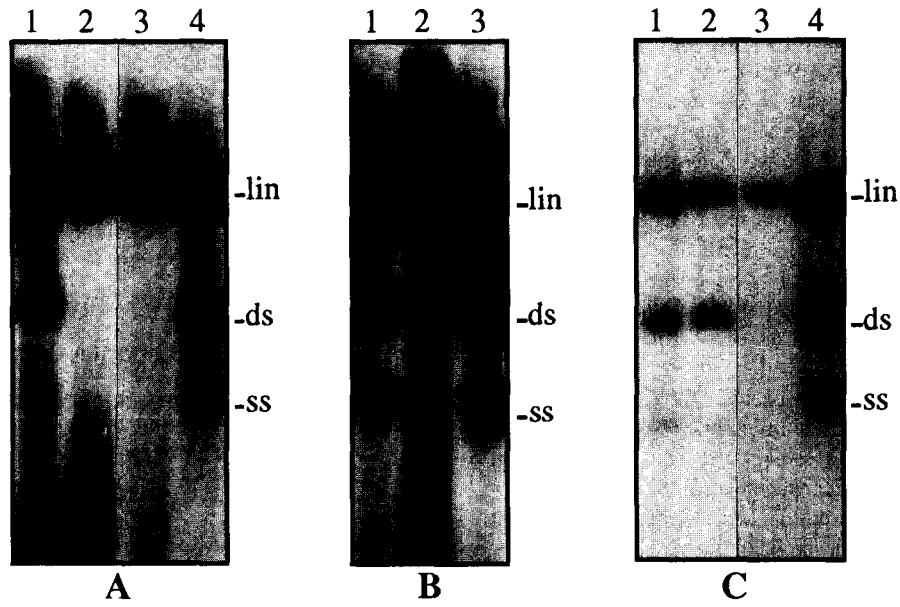

Fig. 2. Replication of TYLCV mutants in a single cell-based replication assay. (A) Specificity of interaction between the IR and the *trans*-acting factors. Tomato protoplasts were transfected with linear DNA from plasmids pTYΔKP (lane 1), 6FL (lane 2), 2O (lane 3) and pTYH20.6 (lane 4). 7 days after transfection, total DNA was isolated and subjected to Southern blot analysis using TYLCV-Sar IR (lanes 1 and 2) or -Is IR (lanes 3 and 4) specific probes. (B) The *cis*-acting sequences involved in origin recognition specificity are located on the left half of the IR. Tomato protoplasts were transfected with linear DNA from clones pTYΔKP (lane 1), 10FM (lane 2), and 41EP (lane 3). 7 days after transfection, total DNA was isolated and subjected to Southern blot analysis using a TYLCV-Sar specific probe. (C) The specificity is mediated by the N-terminal 116 amino acids of the TYLCV Rep protein. Tomato protoplasts were transfected with linear DNA from clones pTYΔKP (lane 1), 4FB (lane 2), 3FA (lane 3) and pTYH20.6 (lane 4). 7 days after transfection, total DNA was isolated and subjected to Southern blot analysis using TYLCV-Sar IR (lanes 1 and 2) or -Is IR (lanes 3 and 4) specific probes. The positions of linear (input) DNA (lin), supercoiled double-stranded DNA (ds) and single-stranded DNA (ss) are indicated.

The previous results demonstrate that the Rep proteins of the closely related geminiviruses TYLCV-Sar and -Is are not interchangeable, even though there is strong sequence homology and functional equivalence between these proteins. The Rep proteins of these two isolates share 76% amino acid sequence identity (Fig. 3) and presumably have the same tertiary structure. They show a high degree of similarity with consensus motifs of replication initiator proteins of bacterial plasmid families [6], as well as DNA-dependent ATPases [23] (Fig. 3).

In order to define the protein domain involved in the specific recognition of the origin, we exchanged domains between the TYLCV-Sar and -Is Rep proteins. This homolog-scanning mutagenesis is useful for identifying sequences that cause functional variation among homologous proteins [24]. The Rep protein was divided into two parts at amino acid 123 (amino acid 121 in TYLCV-Is) by the introduction of a unique *Xho*I restriction site, neutral for the Rep amino acid sequence. This part of the protein is conserved between the TYLCV Rep proteins as well as among the Rep proteins of bipartite geminiviruses, and hence can not be part of the specificity determining domain. The *Xho*I and *Bgl*II restriction sites were used to replace the C-terminal part of the Rep protein of the constructs pTYΔKP and 2O by the one of pTYH20.6. Due to the amino acid identity between the Rep proteins from the two isolates, this manipulation produced constructs in which the 116 N-terminal amino acids were identical to TYLCV-Sar Rep protein (Fig. 1). Due to the location of the *Xho*I and *Bgl*II sites, this exchange also affected the overlapping ORFs C2 and C4 (Fig. 1). However, previous studies ([20, and unpublished results) have shown that none of the proteins encoded by these ORFs contribute to TYLCV replication.

As shown in Fig. 2C, lane 2, the mutant 4FB, in which this chimeric Rep protein is associated with a TYLCV-Sar IR, replicated autonomously in protoplast-derived cells of tomato. This demonstrates the functionality of the chimeric protein, and proves that the domain exchange did not cause any structural impairment. On the other hand, when the same hybrid Rep protein is associated with an -Is IR, as in the clone 3FA

Fig. 3. Schematic representation of the TYLCV Rep protein. Amino acids identical between TYLCV-Sar and -Is Rep proteins are represented by black bars. Regions I (aa 15–22), II (aa 55–63) and III (aa 100–109) represent homologies with consensus Rep motifs [6], the putative nicking Tyr^{103} is indicated by an asterisk. Regions A (aa 217–231), B (aa 258–263), B′ (aa 271–287) and C (aa 298–305) represent homologies with DNA-dependent ATPase [23]. The domain responsible for origin specific recognition is indicated.

(Fig. 2C, lane 3), no replication could be detected. The chimeric Rep protein is thus still specific for a particular IR, demonstrating that the specific recognition of the origin is conferred by the 116 N-terminal amino acids. This N-terminal domain contains several motifs conserved among the geminiviral Rep proteins in particular a 'PHLH' motif, possibly involved in metal ion coordination [6], as well as the putative DNA-nicking Tyr^{103} [23] (Fig. 3). Due to the amino acid identity between the two isolates, the chimeric protein is 91% identical to the TYLCV-Is Rep protein, differing at 32 amino acid positions, but is still specific for the -Sar IR. Further experiments will define which of these amino acid differences are responsible for the origin recognition specificity.

It is likely that the sequence-specific DNA binding of the replication protein to its cognate intergenic region contributes to the replication specificity. However, the Rep–ori interaction differs from the common site-specific protein–DNA interactions because, in addition, it entails an enzymatic activity, the introduction of a site-specific nick in the viral strand [7]. In addition, sequence similarities had prompted speculation about a potential helicase activity of the geminivirus Rep proteins [23]. Indeed, TYLCV Rep protein has recently been shown to possess ATPase activity in vitro that is required for viral replication (unpublished results). Whether the specificity of origin recognition is mediated through direct DNA–protein interaction between the left half of the IR and the 116 N-terminal amino acids of the Rep protein, or whether it is modulated according to its catalytic activities, or through the interaction between Rep and host or other viral-encoded proteins, remains to be established.

Acknowledgements: We thank Françoise Jouanneau for expert technical assistance in protoplast preparation, Henryk Czosnek for providing the clone pTYH20.6 and the tomato suspension culture, Jürgen Laufs and Jeff Leung for comments on the manuscript, and Jim Morrison for constant support during transfection experiments.

References

[1] Lazarowitz, S.G. (1992) Crit. Rev. Plant Sci. 11, 327–349.
[2] Saunders, K., Lucy, A. and Stanley, J. (1991) Nucleic Acids Res. 19, 2325–2330.
[3] Stenger, D.C., Revington, G.N., Stevenson, M.C. and Bisaro, D.M. (1991) Proc. Natl. Acad. Sci. USA 88, 8029–8033.
[4] Rogers, S.G., Bisaro, D.M., Horsch, R.B., Fraley, R.T., Hoffmann, N.L., Brand, L., Elmer, J.S. and Lloyd, A.M. (1986) Cell 45, 593–600.
[5] Elmer, J.S., Brand, L., Sunter, G., Gardiner, W.E., Bisaro, D.M. and Rogers, S.G. (1988) Nucleic Acids Res. 16, 7043–7060.
[6] Ilyina, T.V. and Koonin, E.V. (1992) Nucleic Acids Res. 20, 3279–3285.
[7] Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S.G., Schell, J. and Gronenborn, B. (1995) Proc. Natl. Acad. Sci. USA (in press).
[8] Fontes, E.P.B., Luckow, V.A. and Hanley-Bowdoin, L. (1992) Plant Cell 4, 597–608
[9] Thömmes, P., Osman, T.A.M., Hayes, R.J. and Buck, K.W. (1993) FEBS Lett. 319, 95–99.
[10] Fontes, E.P.B., Eagle, P.A.,Sipe, P.S., Luckow, V.A. and Hanley-Bowdoin, L. (1994) J. Biol. Chem. 269, 8459–8465.
[11] Sunter, G., Hartitz, M.D. and Bisaro, D.M. (1993) Virology 195, 275–280.
[12] Eagle, P.A., Orozco, B.M. and Hanley-Bowdoin, L. (1994) Plant Cell 6, 1157–1170.
[13] Lazarowitz, S.G., Wu, L.C., Rogers, S.G. and Elmer, J.S. (1992) Plant Cell 4, 799–809.
[14] Fontes, E.P.B., Gladfelter, H.J., Schaffer, R.L., Petty, I.T.D. and Hanley-Bowdoin, L. (1994) Plant Cell 6, 405–416.
[15] Kheyr-Pour, A., Bendahmane, M., Matzeit, V., Accotto, G.P., Crespi, S. and Gronenborn, B. (1991) Nucleic Acids Res. 19, 6763–6769.
[16] Navot, N., Pichersky, E., Zeidan, M., Zamir, D. and Czosnek, H. (1991) Virology 185, 151–161.
[17] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology, Wiley, New York.
[18] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[19] Kunkel, T.A. (1985) Proc. Natl. Acad. Sci. USA 82, 488–492.
[20] Jupin, I., de Kouchkovsky, F., Jouanneau, F. and Gronenborn, B. (1994) Virology 204, 82–90.
[21] Matzeit, V., Schaefer, S., Kamman, M;, Schalk, H-J., Schell, J. and Gronenborn, B. (1991) Plant Cell 3, 247–258.
[22] Argüello-Astorga, G.R., Guevara-Gonzàlez, R.G., Herrera-Estrella, L.R. and Rivera-Bustamante, R.F. (1994) Virology 203, 90–100.
[23] Koonin, E.V. (1993) Nucleic Acids Res. 21, 2541–2547.
[24] Cunningham, B.C. and Wells, J.A. (1989) Science 244, 1081–1085.