Shedding light on the multiple functions of the geminivirus Replication initiator protein

Maio, F.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3

A conserved lysine residue essential for geminivirus replication also controls nuclear localization of the TYLCV Rep protein

Francesca Maio¹, Manuel Arroyo-Mateos¹², Benjamin G. Bobay³⁴⁵, Eduardo R. Bejarano², Marcel Prins¹⁶, and Harrold A. van den Burg¹

¹Molecular Plant Pathology, Swammerdam Institute for Life Sciences, Faculty of Science University of Amsterdam, PO box 94215,1090 GE Amsterdam, The Netherlands.

²Instituto de Hortofruticultura Subtropical y Mediterránea “La Mayora”, Universidad de Málaga-Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), Dept. Biología Celular, Genética y Fisiología, Universidad de Málaga, Campus Teatinos, 29071 Málaga, Spain.

³Duke University NMR Center, Duke University Medical Center, Durham, NC 27710, USA.

⁴Department of Biochemistry, Duke University, Durham, NC 27710, USA.

⁵Department of Radiology, Duke University, Durham, NC 27710, USA.


This Chapter has been published as:

Chapter 3

ABSTRACT

Geminiviruses are ssDNA viruses that infect a wide range of plants. To promote viral replication, geminiviruses manipulate the host cell cycle. The viral protein Rep is essential to reprogram the cell cycle and then initiate viral DNA replication by interacting with a plethora of nuclear host factors. Even though many protein domains of Rep have been characterized, little is known about its nuclear targeting. Here, we show that one conserved lysine in the N-terminal part of Rep is pivotal for nuclear localization of Rep from *Tomato Yellow Leaf Curl Virus* (TYLCV) with two other lysines contributing to its nuclear import. Previous work had identified that these residues are essential for Rep from *Tomato Golden Mosaic Virus* (TGMV) to interact with the E2 SUMO conjugating enzyme (SCE1). We find that mutating these lysines leads to nuclear exclusion of TYLCV Rep without compromising its interaction with SCE1. Moreover, the ability of TYLCV Rep to promote viral DNA replication also depends on this highly conserved lysine independently of its role in nuclear import of Rep. Our data thus reveal that this lysine potentially has a broad role in geminivirus replication, but its role in nuclear import and SCE1-binding apparently differs depending on the Rep protein examined.
INTRODUCTION

Geminiviruses form a large and economically important family of plant viruses (Mansoor et al. 2003; Navas-Castillo et al. 2011; Rojas and Gilbertson 2008). A key signature of these viruses is their circular single-stranded DNA (ssDNA) molecule that is packaged in a twin icosahedral capsid structure (Hesketh et al. 2018; Rojas et al. 2005; Seal et al. 2006). Begomoviruses form the largest genus within the family of geminiviruses. They can have either a bipartite genome (with genomic components known as DNA-A and DNA-B) or a monopartite genome (Fauquet et al. 2003; Fauquet et al. 2008; Zerbini et al. 2017). Begomoviruses are found in the Old World (both genome types) as well as in the New World (mostly bipartite, but monopartite species have now also emerged) (Brown et al. 2015; Macedo et al. 2018; Medina et al. 2018). A begomovirus infection starts when whiteflies feed on phloem sap of non-infected plant during which they transmit virions in the phloem-associated cells. Inside the phloem-associated cells, the ssDNA is released from the virions, copied into double-stranded DNA (dsDNA) and then replicated via rolling-circle or recombination-dependent replication (Hanley-Bowdoin et al. 1999; Jeske et al. 2001; Laufs et al. 1995; Rizvi et al. 2015). Due to their small genome size (2.7 to 3 kb) and limited coding capacity, geminiviruses rely on host cellular processes for their DNA replication by interacting with a wide range of plant proteins (Hanley-Bowdoin et al. 2013; Jeske 2009). As an infection usually starts in differentiated plant cells that have entered the quiescent G0/G1 phase, DNA replication activity is (largely) absent in these recipient host cells. Therefore, geminiviruses must manipulate the cell cycle to promote reentry into the S phase of the cell cycle to allow DNA replication (Egelkrout et al. 2001; Gutierrez et al. 2004; Hanley-Bowdoin et al. 2004; Hanley-Bowdoin et al. 2013; Nagar et al. 1995). Different studies have shown that geminiviruses are able to activate DNA replication by regulation Cyclins in e.g. fission yeast (Hipp et al. 2014), exploiting the translesion synthesis in plants (Richter et al. 2016) and by interacting with transcriptional regulators of the plant cell cycle (Kong et al. 2000).

Since the geminivirus life cycle strongly depends on nuclear activities, most geminiviral proteins are known to translocate to the plant nucleus. For example, the capsid protein (CP) of both monopartite and bipartite begomoviruses is known to contain nuclear import (NLS) and nuclear export signals (NES) (Guerra-Peraza et al. 2005; Kunik et al. 1998; Unseld et al. 2001; Unseld et al. 2004). The transcription activator TrAP and the related protein C2 were shown to contain a basic domain corresponding to residues 17-31 that is important for their nuclear localization (Dong et al. 2003; van Wezel et al. 2001). Also the nuclear shuttle protein BR1 (also called NSP or BV1), which acts in a cooperative manner with the movement protein BL1 (BC1) in the transport of the viral ssDNA genome in host cells, was shown to contain
two NLS motifs in its N-terminal region and one leucine-rich NES motif in the region spanning the residues 177-198 (Sanderfoot and Lazarowitz 1995; Sanderfoot et al. 1996; Ward and Lazarowitz 1999).

Among the few proteins encoded by the geminiviral genome, only one is known to be essential for viral replication, namely the Replication initiator protein (Rep), also called AL1, AC1 or C1. Rep is the most conserved geminiviral protein and exerts a plethora of functions inside the host nucleus (Ruhel and Chakraborty 2018), such as virus-specific recognition of its cognate origin-of-replication (Fontes et al. 1992; Fontes et al. 1994), transcriptional repression of the viral genes (Eagle et al. 1994), binding of dsDNA (Orozco and Hanley-Bowdoin 1998; Singh et al. 2008), DNA nicking, and DNA helicase activity on viral DNA (Choudhury et al. 2006; Clerot and Bernardi 2006; Koonin and Ilyina 1992; Orozco and Hanley-Bowdoin 1996). Moreover, Rep interacts with many proteins. Besides forming homo-oligomers (Choudhury et al. 2006; Orozco et al. 2000), Rep interacts with the viral protein REn (also called AL3, AC3 and C3), which then promotes viral DNA accumulation (Settlage et al. 1996; Settlage et al. 2005), and binds its own viral coat protein (CP), which in turn suppresses both the DNA nicking and ligating activity of Rep (Malik et al. 2005). In addition, the members of the Rep family are known to bind a range of host factors linked to DNA replication, such as the DNA clamp protein PCNA (Proliferating cell nuclear antigen) (Bagewadi et al. 2004; Castillo et al. 2003), Replication factor C (RFC) (Luque et al. 2002), the ssDNA-binding protein RPA (Singh et al. 2007), Histone H3 (Kong and Hanley-Bowdoin 2002), and proteins that control progression of the cell cycle like the Retinoblastoma-related protein (RBR) (Ach et al. 1997; Kong et al. 2000), the Ser/Thr Kinase GRIK (Kong and Hanley-Bowdoin 2002; Shen and Hanley-Bowdoin 2006) and the Ubiquitin conjugating enzyme 2 (UBC2), as well as Histone mono-ubiquitination 1 (HUB1) (Kushwaha et al. 2017). Finally, Rep was shown to interact with the E2 SUMO conjugating enzyme 1 (SCE1) (Arroyo-Mateos et al. 2018; Castillo et al. 2004; Sánchez-Durán et al. 2011) whose activity controls, among others, the SUMO (Small ubiquitin-like modifier) modification of PCNA and RBR (Gali et al. 2012; Rytz et al. 2016).

Sumoylation is a post-translational protein modification that primarily controls nuclear processes by modulating the activity, interactions and/or localization of the modified proteins (Cubéñas-Potts and Matunis 2013; Seeler and Dejean 2003). Covalent attachment of SUMO to target proteins involves a cascade of enzymatic reactions catalyzed by ATP (Chosed et al. 2006; Flotho and Melchior 2013; Gareau and Lima 2010). The last step, the actual attachment of SUMO to specific lysines in target proteins via isopeptide bond formation, is catalyzed by SCE1. The Rep proteins from Tomato golden mosaic virus (TGMV) and Tomato yellow leaf curl Sardinia virus (TYLCSV) interact both with SCE1 (Castillo et al. 2004; Sánchez-Durán et al. 2011).
A conserved lysine residue controls Rep nuclear localization

2011). Studies on Rep from TGMV identified two lysine residues that, when mutated, prevented the interaction between Rep and SCE1 while also inhibiting viral DNA replication, suggesting that this interaction between Rep and SCE1 is potentially directly required for geminiviral replication. Furthermore, Rep was shown to suppress sumoylation of two specific lysines of PCNA (Arroyo-Mateos et al. 2018) (Chapter 2). By now, it is evident that Rep controls reprogramming of the host cell cycle as well as the subsequent initiation of viral DNA replication. In order to carry out these functions, Rep must enter the plant cell nucleus. However, a mechanism for nuclear import remained hitherto unknown for any geminiviral Rep.

We here report strong conservation in the Rep protein family for one of the lysines that is required for Rep$_{TGMV}$ to interact with SCE1. To our surprise, mutating this Lys-to-Ala (KtoA) reduced nuclear import of Rep$_{TYLCV}$, while mutating in concert other lysines needed for the interaction with SCE1 resulted in increased nuclear exclusion of Rep. Moreover, these residues were not essential for Rep$_{TYLCV}$ to interact with SCE1. In reverse, Rep$_{TGMV}$ still entered the nucleus when the corresponding lysines were mutated. Structural modeling of the N-terminal half of Rep revealed that these KtoA mutations largely neutralized a positively charged surface area on Rep$_{TYLCV}$ but not on Rep$_{TGMV}$. This suggested that nuclear import of Rep$_{TYLCV}$ is controlled by this surface area rather than a linear polypeptide—a more typical NLS. Finally, we confirmed that nuclear import of Rep$_{TYLCV}$ is essential for its viral DNA replication activity. This replication activity required these lysines, as previously reported for Rep$_{TGMV}$ (Sánchez-Durán et al. 2011), but this was independent of their role in nuclear import of Rep.

RESULTS

One lysine in the SCE1 binding interface is strongly conserved in the Rep protein family

Previous work had shown that two Reps from distantly related begomoviruses interact with SCE1 from N. benthamiana, i.e. Rep from Tomato yellow leaf curl Sardinia virus (TYLCSV; monopartite Old World clade) and TGMV (bipartite New World clade) (Castillo et al. 2004; Sánchez-Durán et al. 2011). For Rep$_{TGMV}$ this interaction depended on lysines in the N-terminal half (Figure 1A and 1B). In particular, K68 (position $x$) had a major role in SCE1 binding, while the residues K96 ($a$), K102 ($z$) and K107 ($b$) of Rep$_{TGMV}$ had redundant roles in this interaction (Sánchez-Durán et al. 2011). All these lysines appear to be conserved to some extent in the Rep protein family, when looking at set of well-studied mono- and bipartite begomoviruses
Figure 1. Lysine residues involved in SCE1 binding are conserved in Rep proteins from different begomoviruses.
(A) Diagram of REP with its known functional domains; the red line indicates the region of Rep\textsuperscript{TGMV} required for SCE1-binding. (B) Protein sequence alignment of the Reps from different begomoviruses depicting the region corresponding with the residues 40-108 in Rep\textsuperscript{TYLCV}. The full virus names are indicated in materials and methods. The arrows and letters indicate lysines important for Rep\textsuperscript{TGMV} to interact with SCE1 and the additional Lys (y) found in some Rep proteins in this domain. Black arrows point to residues here studied, white arrows point Lys required for viral replication and not targeted for mutagenesis in this work. (continued on next page)
A conserved lysine residue controls Rep nuclear localization

Figure 1. Legend (continued)

“Mono” and “Bi” stand for monopartite and bipartite begomoviruses and OW/NW indicates the origin of the analyzed viruses (Old World, New World). (C) Table indicating the corresponding positions and ConSurf conservation score of the analyzed Lys residues in the here-studied Reps. To each of the here-studied Lys residue a unique color is assigned that is also used in the following figures. (D) Ribbon diagram of the 3D structure model of the N-terminal half of RepTYLCV colored according to the degree-of-sequence conservation using a scale from maroon (highest) to cyan (lowest) based on the ConSurf program (Ashkenazy et al. 2010) using a MSA with 337 Rep protein sequences from different geminiviral species and isolates (Shen et al. 2018).

(Figure 1B and 1C). In addition, some of these Reps share another lysine (position y, not present in RepTGMV) which is positioned three residues downstream the lysine in position x. To calculate the degree of conservation of the lysine residues in this region, we performed a ConSurf analysis on a multiple sequence alignment (MSA) comprising 337 geminiviral Reps (Shen et al. 2018). This analysis revealed that lysine in position y is extremely variable (class 2 out of 9 classes), K102 (z) shows some degree-of-conservation (class 7 out of 9 classes), while K68, K96 and K107 (x, a and b) are highly conserved (class 8 the first two and class 9 the third one) (Figure 1C). This model also revealed that the lysines at positions x, y and z all reside on one side of the protein model (Figure 1D).

Lysine in position x determines nuclear import of RepTYLCV rather than SCE1 binding

To judge whether, in general, the lysine at position x has a key role in Rep for the interaction with SCE1, we decided to study a distant homolog of RepTGMV for its interaction with SCE1. We picked Rep from TYLCV-Alb13 (hereafter referred to as RepTYLCV), as it represents a monopartite OW geminivirus, whereas RepTGMV represents a bipartite NW geminivirus. We introduced single (K67A and K101A), double (K67/101A) and triple Lys-to-Ala (KtoA) mutations (K67/71/101A) in RepTYLCV, which correspond with the positions x, y and z (Figure 1C). Importantly, we did not mutate the lysines in the positions a and b, as this would impair other functions of Rep, e.g. viral replication (Sánchez-Durán et al. 2011), DNA binding and DNA cleavage activity of Rep (Kong et al. 2000; Orozco and Hanley-Bowdoin 1998). We then assessed whether mutating these lysines of RepTYLCV would cause a loss-of-interaction with SCE1 using bimolecular fluorescence complementation (BiFC). We fused the C-terminal half of the super Cyan fluorescent protein (SCFP) (Gehl et al. 2009) to C-terminus of these Rep variants creating a Rep-SCFP\textsuperscript{C} chimera, while the N-terminal half of SCFP was attached to the N-terminus of SCE1 (SCFP\textsuperscript{N}-SCE1). Reconstitution of CFP fluorescence was then used as a proxy for protein-protein
Figure 2. The conserved lysines determine nuclear targeting of Rep\textsuperscript{TYLCV} rather than its interaction with SCE1. (continued on next page)
A conserved lysine residue controls Rep nuclear localization

Figure 2. Legend (continued)

(A) BiFC showing the nuclear localization of the RepTYLCV-SCE1 protein complex exclusively in nuclear foci/bodies (NBs). Image shown represents a typical N. benthamiana epidermal cell (top) and a 4x zoom showing its nucleus (bottom); white arrows highlight BiFC signal in the cytoplasm. Scale bars represent 5 μm. (B) Box plot depicting the number of NBs per nucleus in the cells expressing the indicated Rep variants (x-axis) as BiFC pair together with SCE1. (C) Similar to B, except that the scatter plot represents the size of the NBs. (D) Box plot depicting the CFP fluorescence intensity ratio in the cytoplasm vs. nucleus of the images shown in (A). For all plots, a total number of 16 cells per sample (n=16) was analyzed; in the box plots, “horizontal bars, boxes, whiskers and dots” represent “median, interquartile ranges (IQR), data range from the minimum to the maximum, and each individual value”, respectively; in the scatter plot, “the horizontal bar, whiskers and dots” indicate “median, IQR and each individual value”, respectively. A Kruskal-Wallis statistical test was performed followed by a Dunn’s post-hoc test for each data set; the letters denote statistically different groups (p<0.05). (E) Yeast two-hybrid assay between ADGAL4-SCE1 and BDGAL4-RepTYLCV variants – LW: control plate for yeast growth and –LWH: selection plate with medium to test for the interaction. All mutants tested retained their interaction with SCE1. (F) Subcellular localization of RFP-tagged RepTYLCV variants in N. benthamiana upon transient expression with Agrobacterium. Arrows indicate fluorescence in the cytoplasm; the asterisks mark non/weakly fluorescent nuclei. (G) Box plot depicting the RFP fluorescence intensity ratio in the cytoplasm vs. nucleus of the images shown in (F); a total of 8 cells per sample was analyzed. (H) Analysis of WT Rep-SCFP C protein levels and its variants using an anti-HA immunoblot (HA tag is positioned between Rep and SCFP C ). (I) Immunoblot of the Rep-RFP fusion proteins upon transient expression in N. benthamiana. Proteins were detected with an anti-RFP antibody. To demonstrate equal protein loading, the membranes were stained with Ponceau S.

We noted that the BIFC pair formed by WT RepTYLCV and SCE1 aggregated in nuclear foci, hereafter called nuclear bodies (NBs, Figure 2A). Different from RepTGMV (Sánchez-Durán et al. 2011), we found that the interaction between RepTYLCV and SCE1 remained intact despite the single or double KtoA mutations, i.e. both the average size and number of NBs was similar to that of WT. Only the triple KtoA mutant formed less NBs than WT Rep, i.e. on average it formed 5 NBs per nucleus instead of ±10 for WT (Figure 2B and 2C). Introduction of these KtoA mutations did not impact the Rep protein levels in planta, as all the Rep-SCFP C chimeras accumulated at the expected mass of ~60 kDa at roughly the same protein levels (Figure 2H). A few smaller protein bands were present on the immunoblot, but these products likely reflect undesired protein degradation in the protein extracts and they do not impede our conclusions. Thus, the different KtoA Rep-SCFP C variants are each (relatively) stable in planta.

To independently confirm our pattern of RepTYLCV-SCE1 interactions, we shifted to the yeast two-hybrid (Y2H) assay expressing RepTYLCV as a fusion protein with the GAL4 binding domain (BD) and SCE1 as a fusion with the GAL4 activation domain (AD). We first verified that none of the BD-RepTYLCV variants caused autoactivation of the reporter genes in the Y2H strain (Figure S1A). Also in the Y2H assay, we found that none of the KtoA variants of RepTYLCV had lost its interaction with SCE1
albeit that overall the interaction appears to be relatively weak in the Y2H assay (Figure 2E).

Although these results suggested that these lysines are not essential for Rep^TYLCV and SCE1 to interact, we noted that introduction of the KtoA mutations changed the distribution pattern of the BiFC signal. Whereas the BiFC signal for WT Rep-SCE1 and the K101A variant was nearly exclusively nuclear, the BiFC signal of the other KtoA variants was strongly increased outside the nucleus (Figure 2A and 2D). In particular, the single mutation K67A caused already a profound shift of the BiFC signal out of the nucleus, which was further enhanced when the mutations K71A or K101A were added alone or combination to K67A (Figure 2D).

To determine the cause of this redistribution, we examined the subcellular localization of Rep alone expressing it with a RFP-tag at its C-terminus (Rep-RFP). Both WT Rep^TYLCV and the K101A variant localized primarily to the nucleus. However, the variants K67A, K67/101A and K67/71/101A resided each both in the nucleus and cytoplasm with the triple KtoA mutant accumulating the least in the nucleus (Figure 2F and 2G). A co-localization study of the Rep KtoA triple mutant with an endoplasmic reticulum (ER) marker (ER Green Tracker), confirmed that this RFP signal outside the nucleus does not strongly overlap with the ER and that it, therefore, most likely reflects cytoplasmic accumulation (Figure S1B). Quantification of the RFP fluorescence intensity ratio in the cytoplasm versus the nucleus revealed that mutating K67A had a larger impact on nuclear exclusion than K101A (Figure 2G). Nuclear exclusion was further stimulated when the mutations K101A alone or K71A+K101A were introduced in this K67A variant. Thus, these mutations cause increasingly nuclear exclusion of Rep.

Importantly, in the BiFC experiment each of the tested Rep KtoA variants-SCE1 pairs resided to a large proportion in the nucleus (often in NBs), suggesting that the BiFC interaction itself promoted in part the nuclear import of the protein pair. In fact, SCE1 is known to localize both to the cytoplasm and nucleus of the plant cell (Mazur et al. 2019). Alternatively, we cannot exclude that the Rep-SCFPC fusion (~60 kDa), being a bit smaller than the Rep-RFP fusion (~70 kDa), by itself is already more prone to diffuse passively into the nucleus (Timney et al. 2016). Similar to the BiFC assay, the RFP-tagged Rep variants all accumulated at the expected mass in plant cells (Figure 21). The protein levels were slightly higher for the double and triple KtoA mutants than WT Rep; this likely reflects their increased cytosolic localization, which normally facilitates protein extraction from plant tissue. Only the K67A mutant accumulated less than WT Rep. Nearly every protein sample showed an extra band around 25 kDa, corresponding to free RFP, which was likely released due to (residual) proteolytic activity in the protein extracts.

To confirm that the lysine in position x is not exclusively needed for nuclear import of
A conserved lysine residue controls Rep nuclear localization

Rep from TYLCV Alb-13, we decided to introduce the corresponding single, double and triple KtoA mutations in Rep from TYLCV-Almeria, hereafter called RepTYLCV-Alm (see Figure 1C for the corresponding Lys positions). Again, we examined the subcellular localization of the above described Rep variants by labelling them with RFP and transiently expressing them in N. benthamiana leaves. As noted before, WT RepTYLCV-Alm and the variants K69A (y) and K99A (z) resided each foremost in the nucleus, while K65A (x) accumulated primarily outside the nucleus; adding the mutations K69A (xy) and K99A (xz) alone or in combination (xyz) to K65A further enhanced nuclear exclusion of Rep (Figure S1C and S1D). Quantification of the RFP signal ratio in the nucleus versus the cytosol marked that also in this case the K65A mutation alone had already a strong impact on Rep nuclear exclusion (Figure S1D). Our data thus imply that one or more lysines involved in SCE1 binding also act in nuclear import of Rep from TYLCV strains. We also tested several NLS prediction tools, but none of them predicted a clear NLS signature in the protein sequence of these Reps.

The conserved Lys are required for RepTGMV to bind SCE1 but not its nuclear retention

Previously, Sanchez-Duran and co-workers had shown that in the case of RepTGMV the residues K68+K102 (xz) combined were essential for Rep to interact with SCE1 in planta (Sánchez-Durán et al. 2011). However, our results raised the possibility that the observed ‘loss-of-interaction’ between Rep and SCE1 for the corresponding Rep mutants might be due to their mislocalization to the cytosol. In order to test this notion, we introduced the substitutions K68A (x), K102A (z) and K68/102A (xz) in RepTGMV-RFP and we analyzed their subcellular localization. In contrast to RepTYLCV, the KtoA substitutions did not interfere with nuclear import of RepTGMV, as we saw no RFP signal for any these tested RepTGMV variants in the cytosol (Figure 3A). Then, we tested their interaction with SCE1 using BiFC, i.e. RepTGMV-SCFP+C + SCE1-SCFPN. While WT RepTGMV accumulated to some extent in NBs in this assay, both RepTGMV K68A (x) and K68/102A (xz) had lost their ability to aggregate with SCE1 in NBs, (Figure 3B). Instead, we found a diffuse fluorescence nuclear signal for these two KtoA variants, which suggests that those mutations potentially interfere with SCE1 enzymatic activity (Mazur et al. 2019). Mutating K102 (z) alone did not abolish its interaction with SCE1 in NBs, as both the number and average size of the NBs remained unchanged in comparison to the WT RepTGMV (Figure 3C and 3D). To assess if the absence of NBs correlated with reduced affinity of RepTGMV for SCE1, we performed an Y2H assay. While the WT BD-RepTGMV fusion protein interacted with the AD-SCE1 fusion, no interaction was observed for BD-RepTGMV K68/102A
Figure 3. The lysine residues are not essential nuclear localization of RepTGMV, while in part for RepSLCCNV.

(A) RepTGMV-RFP variants reside exclusively in the nucleus, meaning that introduction of KtoA mutations did not impede the nuclear localization. Scale bars represent 5 μm. (B) BiFC assay showing that WT RepTGMV and K102A interact with SCE1 inside NBs, while K68A and K68/102A mutants yield a diffuse fluorescence signal in the nucleus. (C) Box plot depicting the number of NBs per nucleus in cells expressing the indicated RepTGMV variants (x-axis) as BiFC pair with SCE1. Per sample, 8 cells were analyzed. (D) Similar to C, except that the scatter plot represents the size of NBs (E) Yeast two-hybrid assay between AD GAL4-SCE1 and BD GAL4- fused to RepTGMV WT and K68/102A variants; –LW: control plate for yeast growth; –LWH: selection plate with medium to test for the interaction. (F) Subcellular localization of RFP-tagged RepSLCCNV, WT and KtoA variants, in N. benthamiana upon transient expression using Agrobacterium. White arrows indicate cytoplasmic localization of Rep. Scale bar is 5 μm. (G) Box plot depicting the ratio of the RFP fluorescence intensity in the cytoplasm vs. nucleus of the images shown in (E), per sample 12 cells were analyzed. Conditions are similar to Figure 2.
A conserved lysine residue controls Rep nuclear localization

and AD-SCE1 in this Y2H assay (Figure 3E). Again, none of the BD-Rep\textsuperscript{TGMV} fusions caused autoactivation of the reporter genes when expressed alone in the Y2H strain (Figure S1E).

These results thus confirm the original observations (Sánchez-Durán et al. 2011) that these residues control in the case of Rep\textsuperscript{TGMV} its interaction with SCE1. Also in the BiFC interaction with SCE1, these Rep\textsuperscript{TGMV} variants gave no clear fluorescence signal in the cytoplasm. Thus, these two lysines have distinct functions in these two Rep proteins, i.e. they are required for Rep\textsuperscript{TGMV} to interact with SCE1, while they are not essential for its nuclear accumulation, while the exact opposite was found for Rep\textsuperscript{TYLcv}.

The lysines are also implicated in nuclear localization of Rep\textsuperscript{SLCCNV}

In order to determine whether these lysines also control nuclear targeting of other Reps, we selected a third Rep, i.e. Rep from Squash leaf curl China virus (Rep\textsuperscript{SLCCNV}), as target for mutagenesis based on two criteria: (1) SLCCNV is a bipartite begomovirus like TGMV and (2) it has nearly the same peptide sequence as Rep\textsuperscript{TYLcv} in the selected region, including the additional lysine at position \( y \) (Figure 1B and 1C).

We introduced single (K101A, \( z \)), double (K67/71A, \( xy \)) and triple mutations (K67/71/101A, \( xyz \)) in Rep\textsuperscript{SLCCNV} and expressed them transiently as RFP-fusions in \( N. \) benthamiana. Whereas WT Rep\textsuperscript{SLCCNV}-RFP and the K101A variant (\( z \)) exclusively localized to the nucleus, the RFP signal of the double and triple mutants was visible in the nucleus and cytosol (Figure 3F). Although none of the double/triple KtoA mutations caused a strong nuclear exclusion of Rep\textsuperscript{SLCCNV}, as the one seen before for Rep\textsuperscript{TYLcv}, their RFP signals were significantly increased in the cytoplasm compared to WT (Figure 3G). Together these data argue that the conserved lysines control nuclear localization of different Rep proteins. However, the data also signify that other basic residues likely contribute to this function, as the Rep\textsuperscript{SLCCNV} KtoA variants still accumulated inside the nucleus.

The lysine residues form a positively charged surface patch on the Rep\textsuperscript{TYLcv}-Alm

As an NLS typically consists of a short stretch of positively charged residues, we wondered whether the KtoA substitutions impacted the protein structures of Rep\textsuperscript{TYLcv} and Rep\textsuperscript{TGMV} by changing their electrostatic surface charge. To test this, we performed molecular dynamics simulations of the N-terminal half of Rep\textsuperscript{TYLcv-Alm} and Rep\textsuperscript{TGMV}, both WT and the KtoA variants, using the published NMR structure of Rep\textsuperscript{TYLcv} (residues 4-121) as a template structure (Campos-Olivas et al. 2002). In each
Case we found that the molecular simulations were stable over the run time irrespective of the sequence or mutation used. The modeling revealed that introduction of the KtoA substitutions in Rep$^{TYLCV-Alm}$ largely neutralized a positively charged surface area (Figure 4A), while the positive charge of this surface area remained intact to some degree for Rep$^{TGMV}$ when the K68/102A double mutation was introduced (Figure 4B). The modeling provides, therefore, an explanation why the Rep$^{TGMV}$ KtoA mutants still enter the plant nucleus, while the Rep$^{TYLCV-Alm}$ KtoA mutants are largely excluded from the nucleus.

Rep$^{TYLCV}$ triple KtoA mutant still interacts with SCE1 when reintroduced in the nucleus

Since the triple KtoA mutant of Rep$^{TYLCV}$ was largely excluded from the nucleus

Figure 4. Structural models of Rep$^{TYLCV}$ and Rep$^{TGMV}$ reveal that three lysines together constitute a positively charged surface area.

Electrostatic surface plots of (A) WT Rep$^{TYLCV-Alm}$ (left), Rep$^{TYLCV-Alm}$ K65/99A (center) and Rep$^{TYLCV-Alm}$ K65/69/99A (right) and (B) WT Rep$^{TGMV}$ (left) and Rep$^{TGMV}$ K68/102A (right), (residues 4-121). Electrostatic surface plots of the protein models were created using the Adaptive Poisson-Boltzmann Solver in PyMOL. The blue color denotes positively charged electrostatic surface, while the red color denotes negative charged areas.
A conserved lysine residue controls Rep nuclear localization

Figure 5. Reinforced nuclear retention of RepTYLCV triple KtoA mutant restores its interaction with SCE1 inside NBs.

(A) By fusing an alien NLS to the N-terminus of the Rep triple mutant, it relocalizes to the nucleus, while adding a NES or Myristolation motif (MYR) to N-terminus WT Rep localizes it foremost to the cytoplasm/plasma membrane and less to the nucleus. Arrows highlight Rep-RFP in the cytoplasm; asterisks mark nuclei. Scale bar represents 5 μm (B) Box plot showing the ratio of RFP fluorescence intensity in the cytoplasm vs. nucleus of the images shown in (A), per sample, 12 cells were analyzed. (C) BiFC assay showing nuclear localization of the RepTYLCV fusion variants in complex with SCE1. Fusion of an NLS to Rep triple KtoA mutant restrains the Rep-SCE1 complex to NBs, as seen for WT Rep. (continued on next page)
Figure 5. Legend (continued)

Fusions of a NES and MYR motif to WT Rep causes reconstitution of the BiFC Rep-SCE1 complex in NBs/nucleus and the cytoplasm/plasma membrane, respectively. Arrows indicate BiFC signal in the cytoplasm. Scale bar represents 5 μm. (D) Box plot depicting the number of NBs per nucleus in cells expressing the indicated BiFC pair (x-axis). (E) Similar to D, except that the scatter plot represents the size of NBs. (F) Box plot depicting the ratio of CFP fluorescence intensity in the cytoplasm vs. nucleus of the images shown in (C), a total number of 14 cells was analyzed per sample. Conditions are similar to Figure 2.

(Figure 2F), we examined whether reintroduction of this mutant in the nucleus would restore its interaction with SCE1 inside NBs by fusing the NLS motif of SV40 to the Rep-RFP triple mutant. This would reveal whether the decrease in NBs seen with the BiFC couple ‘SCE1+Rep triple KtoA mutant’ (Figure 2B) was a consequence of (i) mislocalization or (ii) reduced affinity for SCE1. In the latter case, SCE1 apparently aids in trapping Rep in NBs. Attachment of an NLS restored nuclear localization of Rep-RFP K67/71/101A with no residual fluorescence in the cytosol (Figure 5A and 5B). As additional controls, we attached (i) a NES sequence of PKI (Protein kinase inhibitor) and (ii) a myristoylation (MYR) motif of CBL1 (Calcineurin B-like protein 1) to the N-terminus of WT Rep to mimic nuclear exclusion by the triple KtoA mutation. Even though both fusion proteins, NES-Rep and MYR-Rep, localized primarily outside the nucleus, they accumulated still to some extent in the nucleus (Figure 5A and 5B). Next, we expressed the above described NLS/NES variants in a BiFC experiment (NLS/NES-Rep-SCFP with SCFPN-SCE1) to analyze if adding a functional NLS could restore the SCE1 interaction for the triple KtoA mutant in NBs. The enforced nuclear localization of the NLS-Rep triple KtoA mutant restored indeed the BiFC interaction with SCE1 inside NBs (K67/71/101A N-NLS, Figure 5C). The number of NBs was similar for the NLS-Rep triple KtoA variant and WT Rep, but the size of the NBs was increased for the triple mutant compared to WT Rep (Figure 5D and 5E). The NES-Rep and MYR-Rep BiFC fusions also interacted with SCE1 in NBs, similar as WT Rep. This probably correlates with their incomplete nuclear exclusion. In agreement, their BiFC signal was also present outside the nucleus, like the BiFC combination ‘SCE1+Rep triple KtoA mutant’ (Figure 5C and 5F). Combined, these data confirm our hypothesis that the altered BiFC signal of the ‘SCE1+Rep triple KtoA’ couple is largely due to Rep mislocalization rather than disruption of the SCE1-binding interface of Rep. Thus, for RepTYLCV the conserved lysines are not essential for SCE1 binding, but they are critical for its nuclear localization.

Mutating the conserved lysine (x) in RepTYLCV-Alm impairs viral replication activity
A conserved lysine residue controls Rep nuclear localization

In order to examine if the conserved lysines also control viral replication, we took advantage of transgenic *N. benthamiana* plants carrying a 2IRGFP cassette. These plants contain a 35Spro:GFP expression cassette flanked by two direct intergenic repeats (IRs) of TYLCV-Almeria, which allows us to monitor viral replication activity of Rep. Upon transient expression of Rep^TYLCV-Alm^ in these plants, extra-chromosomal circular DNA molecules (ECMs) are formed with a GFP expression cassette leading
to massive GFP protein levels (Figure 6A).

To analyze whether the KtoA substitutions also affect Rep replication activity, we then transiently expressed them as Rep-RFP variants in 2IRGFP plants. As the expression levels vary between infiltrated leaves, we always expressed WT Rep in one half of the leaf as an internal control. At 4 dpi we took UV-images of leaves to examine GFP accumulation and tissue was sampled to quantify the ECM levels. Expression of the variants K69A and K99A resulted in strong GFP signals, as did WT Rep (Figure 6B). Expression of K65A resulted in less GFP signal, but still more than the background signal. Expression of the double and triple KtoA mutants resulted in no increase of the GFP signal over the background signal, indicating that those Rep versions entirely failed to stimulate viral DNA replication. Importantly, both WT Rep-RFP and the variants (including the inactive KtoA variants) accumulated at similar levels at the expected mass in this experiment (Figure 6C). We then quantified the ECM levels in DNA extracted from these leaves using qPCR amplifying a unique fragment only present in circular ECMs. The GFP fluorescence signals seen in the UV images correlated positively with the ECM levels detected in these tissues (Figure 6D). Combined, these results confirm that in the case of RepTYLCV-Alm the K65 (x) residue is required for Rep nuclear localization, while K69 (z) and K99 (y) have an auxiliary role. In addition, the mutation K65A (x) strongly suppressed Rep DNA replication activity.

**Nuclear exclusion of Rep effectively suppresses its viral replication activity**

Next, we determined whether compromised DNA replication activity of the Rep KtoA mutants was due to their nuclear exclusion rather than loss of their nuclear replication activity. To this end, we fused a NES to the C-terminus (C-NES) of RepTYLCV-Alm since, as shown for RepTYLCV (Figure 5A), fusion of a NES to the N-terminus does not result in complete nuclear exclusion. To verify that attachment of a peptide did not already compromise Rep activity, we also generated a non-functional NES variant fusion (C-nes). Rep-NES was nearly completely excluded from the nucleus, while Rep-nes showed a normal nuclear localization similar to WT Rep (Figure 7A and 7B). Upon expression in the 2IRGFP plants, both WT and Rep-nes caused accumulation of GFP and ECMs, while Rep-NES did not (Figure 7C and 7E, C-NES). Both REP fusion proteins (C-NES/C-nes) accumulated at the expected mass based on protein immunoblotting (Figure 7D). Free RFP was visible in all samples, which is likely caused by protein degradation as a result of sample preparation. The band corresponding to Rep-NES was more abundant than the other Reps, which could be explained by its cytosolic localization facilitating its extraction. Thus, in our experimental set-up nuclear exclusion of Rep (Rep-NES)
A conserved lysine residue controls Rep nuclear localization

effectively suppressed ECM mobilization and amplification of GFP replicons from the 2IRGFP cassette.

Reintroduction of Rep\textsuperscript{TYLCV-Alm} triple KtoA variant to the nucleus does not restore viral replication activity

As GFP replicon formation requires Rep nuclear localization, we examined if re-
introduction of the Rep double and triple KtoA variants to the nucleus would restore GFP replicon accumulation. To this end, we fused the SV40 NLS to the C-terminus of the RepTYLCV-Alm double and triple KtoA mutants. As negative control, we fused again a non-functional NLS variant to these Rep variants (‘nls’). RFP-tagged chimeras of these Rep NLS/nls variants were then transiently expressed in N. benthamiana to first determine their localization pattern. NLS attachment drove all three tested Rep KtoA variants back into nucleus, while attachment of the non-functional nls did not (Figure 8A and 8B). Importantly, none of the Rep-NLS KtoA variants increased the GFP protein and ECMs DNA levels in the 2IRGFP plants (Figure 8C and 8E), meaning that despite their nuclear-localization these Rep-NLS mutants were still unable to cause viral DNA replication. Again, the full-length protein of these Rep variants accumulated to similar or higher levels in planta than WT Rep (Figure 8D). Combined, these data signify that these three lysines in RepTYLCV are important for nuclear localization of Rep, while they also act redundantly in viral DNA replication independently of their function in nuclear targeting. However, these lysines are not essential for RepTYLCV to interact with SCE1.

DISCUSSION

Rep is well known to be nuclear-localized, but the protein region responsible to its nuclear targeting was thus far unknown for any geminiviral Rep. Here we reveal that three lysines determine jointly nuclear localization of Rep for two TYLCV strains, to some extent for RepSLCCNV, but not for RepTGMV. Previously, it was shown that (i) a truncated form of Rep from TYLCSV (RepTYLCSV) showed reduced nuclear accumulation (residues 1-120) (Sardo et al. 2011) and that (ii) mutating the basic residues in the N-terminal region (namely residues R2, R5, R7, K11 and H56, L57, H58) of Rep from African cassava mosaic virus (RepACMV) compromised in part its nuclear import (Hong et al. 2003). These latter two reports lacked, however, data on the role of individual residues for nuclear localization. Notably, the same lysines, here-shown to be essential for nuclear localization, were previously shown to be critical for RepTGMV to interact with SCE1 (Sánchez-Durán et al. 2011). Mutating these lysines in RepTYLCV did not abolish its SCE1 interaction (Figure 2). This emphasizes again the multi-functionality of these viral proteins. In agreement with this notion, these three Lys residues were together also essential for viral DNA replication activity in the nucleus, as reintroduction of the triple KtoA RepTYLCV-Alm variant into the nucleus by fusing a functional NLS motif to this variant did not restore viral replication activity (Figure 8).

The in planta localization studies revealed that foremost K67 determines nuclear accumulation of RepTYLCV, with auxiliary roles for K71 and K101. Once all three lysines
A conserved lysine residue controls Rep nuclear localization

Figure 8. Rep KtoA mutants fail to replicate the 2IRGFP DNA cassette even when forced into the nucleus

(A) Subcellular localization of the Rep double and triple KtoA variants with an NLS or a non-functional nls fused to their C-terminus. Arrows highlight Rep-RFP in the cytoplasm; asterisks mark nuclei. Scale bar represents 10 μm. (B) Box plot showing the ratio of RFP fluorescence intensity in the cytoplasm vs. nucleus of the images shown in (A), per sample 10 cells were analyzed. (C) UV image 4 days post agroinfiltration of 2IRGFP N. benthamiana leaves to assess Rep-mediated GFP accumulation by the Rep-NLS/nls KtoA variants (infiltrated Rep constructs are indicated in the drawings). (D) Immunoblot on total protein extracts of agroinfiltrated leaf areas to detect accumulation of the RPF-tagged Rep-NLS/nls KtoA variants (anti-RFP). To demonstrate equal protein loading, the membranes were stained with Coomassie Brilliant Blue. (E) Quantification of the ECMs levels in the 2IRGFP agroinfiltrated leaf areas using real time PCR on total DNA extract obtained from the infiltrated areas.
were mutated, Rep\textsuperscript{TYLCV} was completely nuclear excluded (Figure 2F), meaning that the region spanning the residues 67-101 controls nuclear import of Rep\textsuperscript{TYLCV}. In many cases, a single cluster of 5-7 basic residues already acts as a functional (monopartite) NLS that is recognized by nuclear cargo receptors (Karyopherins) (Radu et al. 1995). Besides a monopartite form, NLS can also adopt a bipartite nature (two clusters of basic residues separated by a 10-12 residue linker) or have a non-canonical signature (Dingwall and Laskey 1991; Kalderon et al. 1984a; Kalderon et al. 1984b; Kosugi et al. 2009; Robbins et al. 1991). The here-identified lysines are likely part of a non-canonical NLS, as none of the existing software tools predicted a classical NLS motif in this region of Rep. In agreement, structural modeling revealed that these lysines of Rep\textsuperscript{TYLCV-Alm} constitute a positive charged surface area, which was neutralized when they were changed to alanines (Figure 4). Additional studies should reveal whether this surface patch is recognized by the nuclear cargo receptors.

In the case of Rep\textsuperscript{SLCCNV}, mutating these three lysines caused only a partial shift of the protein towards the cytoplasm (Figure 3F). This implies that additional residues likely contribute to Rep\textsuperscript{SLCCNV} nuclear targeting. In the case of Rep\textsuperscript{TGMV}, none of these Lys residues was alone or in combination essential for its nuclear localization (Figure 3A). In agreement, the Rep\textsuperscript{TGMV} KtoA variant largely retained its positively charged surface area likely needed for nuclear import. Additional basic residues are, therefore, likely important for Rep\textsuperscript{TGMV} nuclear localization or it might even contain a different second NLS. In line, we noted that the residues 42-46 (KKFIK) of Rep\textsuperscript{TGMV} form a quasi-conserved basic motif that matches with a classical signature of a monopartite NLS, and that other basic residues in this region (e.g. His52, His58 and His60) are also strongly conserved (ConSurf score of 9). We currently have no data whether these latter residues contribute to nuclear localization of Rep\textsuperscript{TGMV}.

Despite being conserved, the here-studied lysines have apparently diversified in their function in Rep from different begomoviruses. De facto, the same lysines were first identified as an interaction site of SCE1 in Rep\textsuperscript{TGMV} (Sánchez-Durán et al. 2011) and our BiFC and Y2H assays confirmed the original observation that K68 and K102 are jointly required for this Rep\textsuperscript{TGMV}-SCE1 interaction (Figure 3B). However, mutating the related Lys in Rep\textsuperscript{TYLCV} did not compromise its interaction with SCE1 in both assays (Figure 2). The modest reduction in the number of NBs formed by the BiFC pair SCE1-Rep\textsuperscript{TYLCV} triple KtoA mutant is fully accounted for by its partial nuclear exclusion, since fusion of an NLS to this variant restored both the size and number of NBs reaching the levels of WT Rep (Figure 5C). These results suggest that the conserved Lys residues, while being essential for the nuclear localization of Rep\textsuperscript{TYLCV}, are not essential for its SCE1 interaction, whereas the exact opposite is true for Rep\textsuperscript{TGMV}. Given the fact that both Reps interact with SCE1 in NBs and the KtoA substitutions in Rep\textsuperscript{TYLCV} did not compromise its recruitment to SCE1-
containing NBs, it is likely that RepTYLCV interacts with SCE1 via a divergent binding site compared to RepTGMV. Moreover, co-localization of Rep with the sumoylation machinery inside NBs was thus far not reported. In the case of SUMO, NB formation depends strictly on SCE1 enzymatic activity (Mazur et al. 2019), again implying that Rep does not inhibit SCE1 activity. This cell biology observation thus confirms the original observation that global SUMO conjugate levels are unchanged when Rep is (transiently) overexpressed in planta (Sánchez-Durán et al. 2011). Clearly, these data warrant additional studies on the function and composition of these Rep-SCE1 NBs during viral infection.

As a demonstration of the multifunctional nature of Rep, we and others (Sánchez-Durán et al. 2011) showed that its conserved lysines also have a function viral replication of TYLCV and TGMV. Using a replication reporter system (2IRGFP plants), we here demonstrate that the KtoA mutations impaired mobilization and/or amplification of the extrachromosomal replicons in the 2IRGFP plants and that this was not rescued when the Rep triple KtoA variant was reintroduced into the nucleus by fusing the SV40 NLS to Rep (Figure 8). The N-terminal part of Rep encompasses a DNA-binding domain with motifs required for initiation of rolling-circle replication (motifs I, II and III and GRS, Figure 1A). When these motifs are mutated in a viral clone, this leads to a non-infectious clone that does not (i) replicate its own viral genome nor (ii) cleave the viral ssDNA (Nash et al. 2011; Orozco and Hanley-Bowdoin 1996; Orozco et al. 1997; Orozco and Hanley-Bowdoin 1998). The lysines here analyzed are located in between motif II (a metal-binding site likely involved in the protein conformation and DNA cleavage), GRS (a motif required for maintaining the relative positioning of motif II) and motif III (with the catalytic site for DNA cleavage). Furthermore, the mutations K96A and K68/102A in RepTGMV impaired TGMV viral replication (Sánchez-Durán et al. 2011). Thus, it is possible that mutating these lysines could have altered these nuclear activities of Rep. Further studies are thus needed to elucidate as well the contribution of these lysines for other nuclear activities of Rep.

MATERIALS AND METHODS

General methods and cloning

All molecular techniques were performed according to standard methods (Sambrook and Russell, 2001). E. coli strain DH5α was used for subcloning. Primers and plasmids used in this work are listed in Table S1 and S2, respectively. All the constructs were generated by PCR amplification using Phusion DNA Polymerase (ThermoFisher), recombination with the Gateway vector pDONR207 (ThermoFisher) using the BP
Chapter 3

Clonase II (ThermoFisher) reaction and subsequent transfer to destination vectors using the Gateway LR clonase II reaction (ThermoFisher). Point mutations were introduced by site-directed mutagenesis using the QuikChange protocol. The corresponding DNA coding sequences of the NLS from SV40 (MLQPKKKRKV), a non-functional mutated nls variant (MLQPNNNNN), a NES from the heat stable protein kinase inhibitor PKI (NELALKAAGADINK), a non-functional mutated nes variant (NELALKAGADANK) or a myristolation motif from CBL1 (MYR, MGCFHKSAAKEF) were fused to coding sequence of Rep via PCR amplification. The order of the fusions (X-REP, REP-X) is indicated in the figures. The constructs for in planta localization were cloned into pGWB654 (with a C-terminal mRFP tag) (Nakamura et al. 2010); for BiFC we used the vectors pDEST-GWSCYCE (C-terminal half of S(CFP)3A [residues 156-239], referred to as SCFP) and pDEST-SCYNEGW (N-terminal half of S(CFP)3A [residues 1-173], SCFPN) (Gehl et al. 2009); for Y2H the inserts were cloned into gateway-compatible vectors pGADT7/pGBK7 (Clontech) (Chien et al. 1991). All inserts were verified with DNA sequencing.

TYLCV 2IRGFP plant generation

In order to obtain the plasmid p2IR-GFP, a fragment of 491 bp containing the intergenic region (IR, nucleotides 2460-2781 and 1 to 169, including the promoter sequences of C2 and CP) of TYLCV-Almeria was amplified from the plasmid pGTYCZ-40 (Morilla et al. 2005). Two sets of primers were used: ILIRupEcoRI and ILIRloEcoRI to obtain the IR flanked by EcoRI sites and primers ILIRupHindIII and ILIRloHindIII to amplify IR flanked by HindIII sites.

The EcoRI-IR-EcoRI fragment was cloned in the EcoRI site of pBINGFP (Morilla et al. 2006). This plasmid contains the expression cassette 35S Promoter-GFP-Nos terminator flanked by a HindIII and EcoRI restriction sites. The orientation of the IR in the recombinant plasmid (pBIRGFP) was determined by PCR using the primer combinations ILIRloEcoRI/pBINX1 and ILIRloEcoRI/pBINX2. Subsequently, the HindIII-IR-HindIII fragment was subcloned in pBIRGFP to yield p2IR-GFP. The primer combinations ILIRupHindIII/pBINX2 and ILIRloHindIII/pBINX2 were used to select the plasmid that contains both IR fragments in the same orientation. Nicotiana benthamiana plants were stably transformed with this construct using Agrobacterium tumefaciens and selected for low GFP expression as described in (Morilla et al. 2004).

Transient protein expression using agro-infiltration

The binary constructs were transformed in A. tumefaciens strain GV3101 (Koncz and Schell 1986) by electroporation. Single colonies were grown overnight to an OD₆₀₀
A conserved lysine residue controls Rep nuclear localization

of 0.8-1.5 in low salt LB medium (1% w/v Tryptone, 0.5% w/v yeast extract, 0.25% w/v NaCl, pH 7.0). Cells were pelleted, washed and resuspended in infiltration medium (1× MS [Murashige and Skoog] salts (Duchefa), 10 mM MES pH 5.6, 2% w/v sucrose, 200 μM acetylsyringone).

For protein localization and BiFC studies, four-week old N. benthamiana leaves were syringe-infiltrated with A. tumefaciens cells at an OD_{600} = 1 for all constructs. When two cultures were co-infiltrated for BiFC analysis, they were mixed at a ratio 1:1 to a final OD_{600} = 1. A. tumefaciens strain carrying the pBIN61 binary vector to express the P19 silencing suppressor (referred to as pBIN61:P19) of Tomato bushy stunt virus (TBSV) (Voinnet et al. 2003) was added to every sample at a final OD_{600} = 0.5. Three days post-infiltration N. benthamiana leaf material was collected for microscopic analysis and protein expression.

For the geminiviral replication activity assay, 3-4 weeks old 2iRGFP N. benthamiana plants were syringe-infiltrated with A. tumefaciens carrying the different constructs at an OD_{600} = 1. GFP expression was visualized 4 days post-infiltration using a UVP Blak-Ray B100AP lamp with an excitation wavelength of 365 nm.

Confocal microscopy and image analysis

ER-Tracker™ Green (BODIPY™ FL Glibenclamide, Thermo Fisher) was syringe-infiltrated into agroinfiltrated leaves in 0.01% Tween20 at a final concentration of 5 μM prior to imaging. A confocal laser scanning microscope, Zeiss LSM510, was used to capture the fluorescent images. The agroinfiltrated N. benthamiana cells were examined 3 dpi using Zeiss c-Apochromat 40X 1.2 water-immersion Korr objective. The fluorescence was detected with the following beam/filter settings: SCFP-excitation at 458 nm (argon laser), primary beam-splitting mirrors 458/514, secondary beam splitter 515 nm, band filter BP 470-500 nm; RFP-excitation at 543 nm (helium-neon laser), primary beam-splitting mirrors 488/543, secondary beam splitter 635 nm, band filter LP 585-615 nm; ER-Tracker Green-excitation at 488 nm (argon laser), primary beam-splitting mirrors 405/488, secondary beam splitter 490 nm, band filter BP 505-550 nm. For all observations, the pinhole was set at 1 Airy unit. For every experimental sample, three independent biological replicates were examined and one representative image is shown.

Images were analyzed and processed with ImageJ Fiji 1.0v software (https://fiji.sc) (Schindelin et al. 2012); NBs were counted and measured using the Analyze Particles tool in ImageJ. To measure the fluorescence signal ratio between cytoplasm and nuclear, a rectangular region-of-interest (ROI) of circa 30 square pixels was drawn, which covered the cytoplasm for approximately 30-40 μm². The mean fluorescence intensities in the selected ROI were calculated for the cytoplasm. The same ROI was
then moved to the nucleus and mean fluorescence intensities were measured for the nuclear signal. The ratio between the mean fluorescence intensities of the cytoplasm and of the nucleus was calculated for every picture. The number of analysed cells for every experiment and statistical information are specified in the figure captions. Data visualization and statistical data analyses were done with Prism 7.0 (GraphPad).

**Yeast two-hybrid assay**

The protocol for Y2H assay is described in (de Folter and Immink 2011). The pGADT7/pGBK7 vectors containing the inserts of interest were transformed into *Saccharomyces cerevisiae* PJ69-4a (James et al. 1996) using the standard lithium acetate/single-stranded DNA/polyethylene glycol 3350 protocol. Transformed colonies were selected on minimal yeast medium (MM) supplemented with amino acid solution lacking L-leucine and L-tryptophan (-LW). To select for protein-protein interactions, three independent colonies were picked for each transformation and resuspended in 100 μl sterile milliQ water; a 10-fold serial dilution series of the resuspended colonies was spotted on MM agar plates –LWH (H stands for L-histidine). Plates were then incubated at 30°C for 3 days prior to scoring.

**Protein extraction and immunoblotting**

Two leaf disks (approximately 50 mg) of *N. benthamiana* leaf material were harvested and snap frozen in liquid nitrogen. The material was homogenized with plastic pestles. Laemmli buffer (0.1 M Tris pH 6.8, 20 % glycerol, 4 % SDS, 100 mM DTT, 0.001% Bromophenol blue) was added to each sample (100 μl of buffer per sample). Tubes were vortexed vigorously and boiled for 10 minutes at 96°C. The extracts were then centrifugated at maximum speed (14,000 rpm at 4°C) for 5 minutes. A total of 10 μl of the protein extract was separated on 10% SDS-PAGE gels and subsequently transferred onto a PVDF membrane. Immunodetection of the proteins was performed according to standard protocols using anti-RFP antibody (Chromotek 6G6; 1:1000) to detect the Rep-RFP fusion proteins, anti-HA (Roche 3F10; 1:2000) antibody for Rep-SCFP<sub>C</sub> fusions as primary antibodies and anti-Rat (Pierce 31470; 1:10000) or anti-Mouse (Pierce 31430; 1:10000) as secondary antibodies. The labelled proteins were visualized using enhanced chemiluminescence (ECL, 0.1 M Tris-HCl pH 8.5, 1.25 mM luminol [Sigma-Aldrich 09253] in DMSO, 0.2 mM p-Coumaric acid [Sigma C9008] in DMSO, 0.01% H<sub>2</sub>O<sub>2</sub>) and visualised using MXBE Kodak films (Carestream). Equal loading of the proteins was confirmed by estimating the total amount of Rubisco in each sample by Ponceau S or Coomassie Brilliant Blue staining of the membrane.
Quantification of extrachromosomical molecules using qPCR

To quantify the level of ECM molecules in the infiltrated 2IRGFP plant leaves, DNA was extracted from approximately 50 mg of plant material and 250 ng of total DNA was used as template per real-time PCR reaction in a QuantStudio3 thermocycler (ThermoFisher). The PCR amplicons were amplified with the Hot FIREPol EvaGreen qPCR (Solis Biodyne) kit. The signal for the ECMs was normalized against plant gDNA using the signal for the 25S rRNA amplicon as an internal reference for each sample (accession no. KP824745.1). The Ct values were corrected for primer efficiencies. All expression data were analyzed using the pipeline in the qBASE+ software (Biogazelle).

Modeling of Rep N-terminal domains of TGMV and TYLCV

Structural models for the N-terminal domain of TGMV, TGMV triple mutant, TYLCV-Alm, and TYLCV-Alm mutant Rep sequences (residues 1-120) were created as described previously (Nash et al. 2011; Shen et al. 2018). Molecular dynamics (MD) simulations of the above-described structural models were performed with the GROMACS 5.1 software package using the AMBER99-SB-ILDN force field and the flexible TIP3 water model individually (Abraham et al. 2015). The initial structures were immersed in a periodic water box of dodecahedron shape (1 nm thickness) and neutralized with counterions. Electrostatic energy was calculated using the particle mesh Ewald method with 0.9-nm cutoff distances for the Coulomb and van der Waals interactions. After energy minimization, the system was equilibrated to 300 K and normal pressure for 100 ps with position restraints for heavy atoms and LINCS constraints for all bonds. The system was coupled to the external bath by the Parrinello-Rahman pressure and temperature coupling. The final MD calculations were performed under the same conditions except that the position restraints were removed and the simulation was run for 100 ns. The last frame of the 100 ns simulation was extracted for each model for electrostatic surface analysis. The PyMOL Molecular Graphics System version 1.8 (Schrodinger) was used for structural analysis and image creation. The electrostatic surface plots of the protein models were created by using the Adaptive Poisson-Boltzmann Solver in PyMOL (Baker et al. 2001; Dolinsky et al. 2004; Lerner and Carlson 2006). The PDB2PQR Web server (Dolinsky et al. 2004) was used with the AMBER99 force field and output naming schemes with a default pH of 7.5.

Multi sequences alignment
Chapter 3

The list with 337 Reps was obtained from (Shen et al. 2018). Multi sequence alignment (MSA) was performed as described in (Shen et al. 2018).

Accession numbers

The here used DNA clones of TYLCV isolate Alb13 Rep (FJ956702.1, kindly provided by Keygene N.V., Wageningen, Netherlands); TYLCV-Almeria Rep (AJ489258); TGMV Rep (NC001507); SLCCNV, Squash leaf curl China virus, Rep (KC222956.1, synthetized by Eurofins Genomics); SCE1 (AT3G57870) has been described previously. For the Rep protein alignment, we used: TYLCCNV, Tomato yellow leaf curl China virus (CAC85509); TYLSCV, Tomato yellow leaf curl Sardinia virus (AAA47955); ToLCNDV, Tomato leaf curl New Delhi virus (CAF04471); ACMV, African cassava mosaic virus (AAD40938.1); CtlCV, Cotton leaf curl virus (KC412251); PGMV, Pepper golden mosaic virus (EF210556); PHYVV, Pepper huasteco yellow vein virus (AAL02410); ToMoV, Tomato mottle virus (AAC32414).

ACKNOWLEDGMENTS

We are grateful to L. Tikovsky and H. Lemereis for taking care of our plants; LCAM-FNWI (Molecular Citology, SILS, University of Amsterdam) for maintenance and help with confocal microscopy. We thank L. Hanley-Bowdoin and W. Shen (NC State University) for sharing the Rep MSA; M. Mazur for cloning and sharing SCE1 clones; T. Nagakawa (Shimane University, Japan) for sharing plasmids; M. Kwaaitaal for insightful discussion. This research was supported by a grant from the Spanish Ministerio de Ciencia y Tecnología (AGL2016-75819-C2-1-R). MA-M was supported by the Topsector T&U program Better Plants for Demands (grant 1409-036 to HvdB), including the partnering breeding companies. FM is financially supported by Keygene N.V. (The Netherlands).

REFERENCES


A conserved lysine residue controls Rep nuclear localization


Chapter 3


James, P., Halladay, J., and Craig, E. A. 1996. 'Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast', Genetics, 144: 1425-36.


Koncz, C., and Schell, J. 1986. 'The promoter of TL-DNA gene 5 controls the tissue-specific expression
A conserved lysine residue controls Rep nuclear localization

of chimaeric genes carried by a novel type of *Agrobacterium* binary vector’, *Molec Gen Genet* 204: 383-96.


(Capsicum annuum) Is a Dead-End Host for Tomato yellow leaf curl virus’, Phytopathology, 95: 1089-97.
A conserved lysine residue controls Rep nuclear localization


Chapter 3

Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. 2003. 'An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus', *Plant J.*, 33: 949-56.
