Modeling tobacco mosaic virus proliferation in protoplasts

Jose C. Merchuk, Asterio Sánchez-Mirón, Sebastian Asurmendi, and Mordechai Shacham

Abstract—The ***tobacco mosaic virus*** (TMV) is one of the most studied viruses. It is frequently used as a model in the research of virus-host interactions. The interest in understanding the mechanism of its proliferation stems basically from the field of agriculture, due to the detrimental effect this virus has on several crops. In addition to this direct application, virus-mediated protein expression systems, which are well established for the synthesis of foreign proteins in animal cell cultures, are now being applied to plants and plant cells by means of plant viral vectors. The use of transformed roots for the propagation of viral vectors has also been proposed. This work presents a mechanistic model describing the transient process of TMV multiplication in a protoplast (a wall-deprived cell). It aims to be a mathematical tool able to simulate the transient behavior of the main molecular pools taking part in the process, which will be useful for exploring, understanding and predicting the dynamics of a host-virus system.

The variables considered are the pools of the main molecules taking part in the viral replication process. The basic balance equations for the cellular pools are presented and a satisfactory fit of the model to the experimental data is shown. The presented model is a necessary step toward the formulation of a basic mechanistic model for the systemic propagation of the virus in a plant tissue. It may be extended in many directions as to the optimization of a system for the production of a foreign protein, to the simulation of manipulation of the virus-cell interaction by external factors, to the mechanism of gene silencing or to the prediction of co-infection dynamics.

.

***Keywords*—**Tobacco mosaic virus, mechanistic model, simulation, virus propagation

I. Introduction

The tobacco mosaic virus (TMV) is one of the most studied plant viruses. It is a model virus for the study of plant-virus interactions and replication processes in the positive-strand RNA genome virus family. Even though a large amount of data has been gathered using whole plants, the complexity of the plant-virus interaction makes an understanding of the underlying process quite difficult. In order to reduce this complexity, several approaches have been taken, one of them being the expression of single genes of the virus by means of transgenic plants for the specific study of the function of one or a few genes at once [1]. Another approach is to reduce the complexity of the host by using single cells (protoplasts) [2]. The advantage of this approach is that plant-tissue cultures are an alternative, fast developing technology that allows the manipulation and control of all the environmental conditions. In such an environment, close-ups of the fundamental processes of virus replication may be studied in more detail. The present paper is focused on developing a model that may serve as a useful mathematical tool for this approach.

In addition to basic scientific interest, there is also a growing practical interest in the potential applications of virus-replication studies for the commercial production of bio-pharmaceutical proteins [3]-[8].

Currently, plant biotechnology relies on two processes for the delivery and expression of heterologous genes in plants: stable genetic transformation and transient infection with viral vectors. The most successful techniques use transient expression with the use of virus replicons or a combined technique with virus vectors delivered via agrobacteria (magnifection) [8]. The commercial and academic interest in using plant-cell virus systems as alternative mechanisms for the production of foreign proteins (vaccines, antibodies, enzymes, etc.), is due to a series of advantages of these systems, mainly the lower risk of contamination with mammalian products [3], [5], [9]-[10]. Thus, it is of great importance to reliably establish the restrictions and limitations of viral infections that will later allow the maximization of protein production by means of the aforementioned viral systems. A mathematical representation of the process that takes into account the main variables may be instrumental in reaching this goal. The model presented here is a step in this direction.

Kinetic models, that study the TMV multiplication in protoplasts from the perspective of population dynamics, have been proposed by several authors [11]-[18]. A number of structured virus modeling attempts (considering the variations of some of the intracellular components) may be found that focus on: human rhinovirus (HVR) and human immunodeficiency virus (HIV-1) [19]; bacteriophages T3 and T7 [20]-[21] and phage Qβ [21]-[29], among others. Buchholz and Schneider [20] introduced the incorporation of lag times in their kinetic mechanisms. A mathematical model of virus trafficking, including the binding, uptake and nuclear accumulation of baculovirus in suspended insect cells, was presented by [30]-[31]. Eigen [22] put forth the hypercycle

Jose C. Merchuk is with the Ben-Gurion University of the Negev, Beer-Sheva, Israel. (phone: 972-544-711023; fax: 97286472916; E-mail: jcm@bgu.ac.il).

Asterio Sánchez-Mirón is with the University of Almeria, Almeria, Spain (e-mail:asmiron@ual.es)

Sebastian Asurmendi is with the Instituto de Biologia, CICVyA, INTA, Argentina. (e-mail: asurmendi.sebastian@inta.gob.ar).

Mordechai Shacham is with the Ben-Gurion University of the Negev, Beer-Sheva, Israel). (e-mail: shacham@bgu.ac.il)

concept and modeled the replication of the Qβ bacteriophage in a simplified scheme, based on the competition of the replicase and the RNA for the available ribosomes. Later on, the DNA trafficking in the cytosol, viewed spatially, was provided by Holcman [32]. Ji and Luo [33] proposed a mathematical model representing the proliferation of TMV and potato virus Y (PVY) replication, assembly and translation in a plant cell.

The model presented here aims to simulate the dynamics of TMV replication inside a protoplast, following the temporal behavior of the genomic RNA positive and negative strands, R_+ and R_-, of the sub-genomic coat protein (R_C), the coat protein itself (P_C), the protein replicase (A_r) and the new generated viruses (V). The goal is to establish a basic tool, one that is able to predict the kinetics of virus production as a function of the main intracellular pools. While the protoplast is an independent and autonomous system, the understanding of the mechanism of viral propagation in this elementary system is a necessary step on the way to simulating viral propagation in a whole cell and the tissue around it. Such a model would enable the prediction of the expected changes in the behavior of the system after engineered modifications of the wild type and might also be extended to predict the effect of externally-driven changes in the pool of cellular proteins on the rate of intracellular virus accumulation [21].

In the present paper, we do not attempt to model the intricate movements of the viral components within the host cell, in which the cytoskeleton plays a major role. Rather, we try to represent the behavior of the main measurable variables by a series of kinetic expressions that engulf the entire dynamics of the process. Thus, we refer to the 'cellular pool' of a certain component without considering its spatial distribution in the cell. This approach has been taken in most of the published models [11]-[17]; [19]-[20],[22], [30]-[31], [33], with few exemptions [32] and [34]. A clear advantage of modeling using the cellular pools as building blocks is that practically all the experimental data on the dynamics of virus replication have been published in these terms, that is, on a per cell basis.

Applying the above assumptions, the accumulation of TMV in a cell may be depicted as follows. It is assumed that the host elements are readily available excess in the cytosol and that they do not influence the kinetics of the process, and also that the cellular elements do not have a regulating role in the propagation of the virus. This scheme considers the infection of a protoplast followed by a series of steps that will be detailed below, leading to the constitution of a viral replication complex (VRC) [35], which is assumed to engulf many of the replication elements, among them genomic and sub-genomic RNAs, replicase and movement protein ($P_{M)}$ [35]. The VRC-like structure is exported to the neighboring cells via the plasmodesmata, accelerating the infection there [36]. In the present model, we simplified the scheme, since we are considering the infection process only in the protoplast. Neither the sub-genomic RNA (R_M), nor P_M and VCR are considered, leading to the scheme shown in Fig 1.

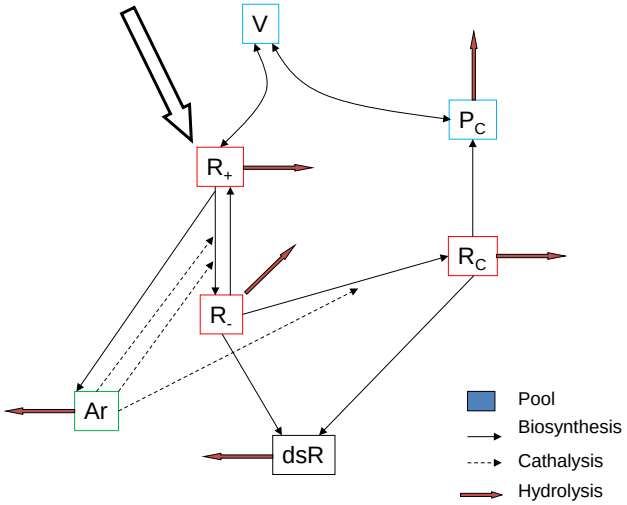

Fig. 1 Schematic representation of the proliferation of TMV in a protoplast indicating the pools considered in the mathematical model.

It is known that the TMV virion contains a single capsule made of m molecules of P_C, and one molecule of the genomic RNA, R+. In the present scheme, the infection process starts at time zero (time of infection, TOI) with the introduction into the protoplast of a single viral particle, which is almost immediately stripped of its protein coat; it appears, therefore, as a positive strand RNA, R_+, which can first act as a template for the synthesis of viral replicase and then of the negative strands, R_- Therefore, the effective multiplicity of infection is MOI=1 [37]**.** In the next paragraph we will present an overview of the process. The specific kinetic equations that were used in the simulation will be presented after that.

.

II. Assembly and Disassembly of P_C

P_C proteins are assembled in three main classes of aggregates: the 4S or A-protein composed of monomers, dimmers (P_{C2}) and trimers (P_{C3}); 20S disks or helices of approximately 38 units; and extended virion-like rods [38]. The assembly of the virion starts with the insertion of a stem-loop structure of the RNA (origin of assembly, OAS) into a 20S unit [39].This new aggregate is connoted as V_{20} and it is assumed that its generation is the controlling step in the process. The assembly proceeds with the addition of P_C aggregates, both in the 5' and in the 3' directions. The assembly of virion V may thus be represented as [39]:

$$
\begin{aligned}
&P_C + P_C \rightarrow P_{C2} \\
&P_{C2} + P_C \rightarrow P_{C3} \\
&xP_{C3} + yP_{C2} + zP_C \rightarrow 20S \\
&20S + R_+ \longrightarrow V_{20}(control) \\
&V_{20} + d\,20S + eP_C + fP_{C2} + gP_{C3} \rightarrow V \\
&V \longleftrightarrow mP_C + R_+
\end{aligned}
\qquad (1)
$$

The first step after infection is the co-translational disassembly in the 5'➔3' direction, uncoating the coding region of the replicase. This allows for the translation of replicase A_r, which starts to be synthesized [40]. This step is shown in the following form--one that is not a representation of a stoichiometric reaction, but rather indicates which viral elements, among those considered here as model variables, play active roles:

$$R_+ \longrightarrow R_+ + Ar \tag{2}$$

In expression (2), neither the cellular elements required for the synthesis and polymerization leading to A_r nor the ribosomes are shown, consistent with our assumption that such elements are present in excess in the cytoplasm.

Next, the replicase enables the replication of the minus strand of the RNA, R-, from the R_+ and also the replication of the R_+ from the R-, as indicated by (3) and (4), respectively:

$$R_+ + Ar \longrightarrow R_+ + Ar + R_- \tag{3}$$

$$R_- + Ar \longrightarrow R_- + Ar + R_+ \tag{4}$$

The replicase also allows for the synthesis of the sub-genomic mRNAs, R_C and R_M, using R- as a template:

$$R_- + Ar \longrightarrow R_- + Ar + R_C \tag{5}$$

$$R_- + Ar \longrightarrow R_- + Ar + R_M \tag{6}$$

The translations of P_C and -P_M are indicated as:

$$R_C \longrightarrow R_C + P_C \tag{8}$$

$$R_M \longrightarrow R_M + P_M \tag{9}$$

All the species may undergo hydrolysis in some measure:

$$R_+ \longrightarrow a \tag{10}$$

$$R_- \longrightarrow a \tag{11}$$

$$R_C \longrightarrow a \tag{12}$$

$$R_M \longrightarrow a \tag{13}$$

$$P_C \longrightarrow a \tag{14}$$

$$P_M \longrightarrow a \tag{15}$$

$$Ar \longrightarrow a \tag{16}$$

where the degraded products are indicated in a generic form as α.

The antisense RNA binding between the R- and the R_C [33] is given by:

$$R_- + R_C \longrightarrow dsRNA \tag{17}$$

As mentioned above, once R_+ is uncoated, it can act as a template for both translation and replication. But these two activities (2 and 3) cannot occur simultaneously on the same strand, since the ribosomes move on the strand in the 5'➔3' direction and the replicase copies in the opposite direction, 3'➔5' [40]. Moreover, it is known that the replication producing a negative strand of RNA, R-, is rapidly completed, approximately 6 hours after infection, as indicated by the level of R-, which then remains constant [41]. The replicase, on the other hand, continues to accumulate, indicating that the translation machinery remains active. Thus far, the detailed mechanism that determines which fraction of the R_+ will be devoted to translation and which to replication during the first stage of TMV infection is not fully understood in this model [18], [29], [40]; such control is assigned to an arbitrary function, θ, meant to produce the rapid increase of R- and its stabilization after a short time. This function may be defined as the ratio of the R_+ dedicated to the translation (via the ribosome producing A_r) to the total R_+ (also including the R_+ dedicated to the replication which yields R-):

$$\theta = \frac{R_+ replication}{R_+ total} \tag{18}$$

The function θ does not appear in the qualitative equations describing the assumed mechanism presented here, but does have an important role in the kinetic equations presented in the next section. Its mathematical expression must be inferred.

Considering that the experimental data show that R- increases after infection and stabilizes after 6-8 hours (indicating that no more of it is being synthesized after that), the following form is proposed for the replication fraction. θ, as a function of R:

$$\theta = 1 - \left(\frac{R_-}{R_{-\max}} \right)^2 \tag{19}$$

where R_{-max} is the maximal value of the negative strand known from the experimental data. This equation yields a value of unity for the initial time of infection and approaches approximately zero after 6-8 h, stopping the synthesis of R-.

III. Cellular Pool Balances

The formulation of this model is based on the transient balances of the cellular pools of free R_+, R-, A_r, R_C, P_C and dsRNA.

The balance of the positive sense RNA is formulated as:

$$\frac{dR_+}{dt} = kV - k_1 P_C R_+ - k_{11} R_+ + \frac{k_2 R_-}{1 + k_3 Ar^2} - r_{\exp} \tag{20}$$

where R_+, R-, V, P_C, R- and A_r are given in units representing the number of molecules per cell. The left-hand side of (20) is the rate of accumulation of free (non-encapsulated) R_+ molecules in the cell. The first term on the right-hand side of the equation represents the net rate of R_+ increase due to virus disassembly. As will be seen later, this term is negligible after the TOI. The second term represents the rate of decrease of the free R_+ pool due to the virus assembly reaction (1). It is known that the controlling step in the series of successive reactions sketched in (1) is the insertion of a stem-loop RNA structure (OAS) into a 20S unit [39]. Here, it is assumed that the different coat-protein aggregates are in equilibrium and, therefore, the total concentration of P_C is proportional to the concentration of the 20S aggregates. This concurs, on the

practical level, with assumptions made by other researchers [33]. The third term is the rate of hydrolysis (9) and the fourth term is the rate of R_+ synthesis by replication from R_- (4), where the possible regulation of this rate by the replicase, A_r, is assumed. The fifth term, r_{exp}, represents the rate of export of VRC, and is presented here only to complete the picture, but in the simulations this term was considered nil. Since we are considering protoplasts, and not cells that are part of a plant tissue, this term is zero in our calculations. The balance of the pool of negative strands of RNA may be written as:

$$\frac{dR_-}{dt} = \frac{k_4 \theta R_+}{1 + k_5 A_r} - k_{12} R_- - k_{16} R_C R_- \quad (21)$$

Equation (21) states that the accumulation of R_- is balanced by three terms: the generation by replication, given by (3) with the addition of the function θ in (19); the disappearance by hydrolysis in (11); and the antisense binding in (17).

Balancing the cellular pool of the viral replicase gives:

$$\frac{dA_r}{dt} = \frac{k_6 (1-\theta) R_+}{1 + k_7 A_r} - k_{13} A_r \quad (22)$$

The first term on the right-hand side represents the rate of generation of A_r by translation (2), which is proportional to the fraction (1-θ) of the positive thread. The kinetic formality chosen allows for the eventual stabilization of A_r, after a certain concentration of enzyme is reached. The following term corresponds to the disappearance of A_r by hydrolysis (16).

Balancing the cellular pool of the sub-genomic mRNAs, R_C, gives:

$$\frac{dR_C}{dt} = \frac{k_8 R_-}{1 + k_9 R_C} - k_{16} R_C R_- - k_{14} R_C \quad (23)$$

where the first term on the right-hand side denotes generation by (5), the second--the possibility of loss of R_C from the pool due to antisense binding (17), and the third term--disappearance by hydrolysis (12).

Balancing the cellular pool of the sub-genomic mRNAs, R_M, would have a similar form, but as stated above, it bears no relevance to the present case and the pools of the rest of the components may be determined without taking it into account.

Balancing the cellular pool of the coat protein shows that the accumulation of P_C equals the sum of the terms for: generation by translation (8); consumption by virus assembly (1); and hydrolysis (14). Here, m is the number of P_C molecules in a viral particle:

$$\frac{dP_C}{dt} = k_{10} R_C R_+ R_+ - m k_1 P_C R_+ - k_{15} P_C \quad (24)$$

The kinetic expression used for the generation term is based on the strong dependency on R_+, as seen from the experimental data. It seems reasonable that when a larger amount of free R_+ is available, the generation of P_C will be more rapid.

The profile giving the accumulation of virus in the cell is obtained from:

$$\frac{dV}{dt} = k_1 P_C R_+ \quad (25)$$

The amount of RNA captured by antisense binding may be evaluated from:

$$\frac{d[dsRNA]}{dt} = k_{16} R_C R_- \quad (26)$$

This equation, with appropriate modifications, will be a key part in the simulation of silencing in the cell (future work, not part of the present paper).

The balances of the cellular pools of the primary infected cell, given by differential equations (19-26), describe the dynamic behaviors of variables R_+, R_-, A_r, R_C, P_C, V and dsRNA. The initial conditions required for solving this mathematical system should represent the situation in which a virus is mechanically introduced into a cell through the surface of the tissue (through a wound) and disassembly takes place over a short period of time (several minutes), which is negligible with respect to the time constants of the infection process as a whole. We normalize the cellular concentrations by taking as reference the initial amount of infecting R_+:

At t = 0, $R_+ = 1$; $R_- = A_r = R_C = R_M = P_C = P_M = V = dsRNA = 0$ (27)

The present model shows the viral replication in a protoplast as a function of seven of the main cellular pools related to the process and, therefore, provides a more detailed description of the process than previous published models do. It allows for the prediction of the dynamic behavior of those pools and may serve as a base for interventions or manipulations at the cellular level at certain points of time in a biotechnological process.

IV. Experimental Database

We have not found published experimental data on the absolute amounts of all the species simulated in the present work; only partial information is available at this time. The data used for the calibration of our mathematical model were derived from those of [41]-[43]. Since none of them quantified all of the species involved in the infection process, some manipulations were done in order to integrate all their reported information into one set of coherent data. Because a viral particle contains a single RNA+ thread, the starting point is the data by [42], where the number of viral particles per cell is shown. As in the aforementioned papers, the same virus is used. The evolution of these data over time has a shape similar to that of the data for 'infectivity' reported by [43].

In general, and because of the analytical methods used, the data reported for R_+ include its total pool, whether it is present as a free molecule or as a viral particle [42]. It is also known that at 20 hours after infection most of the R_+ is already encapsulated. Therefore, at any given time, the difference between the total R_+ pool and the virus particles will give the amount of free R_+ molecules. For the rest of the RNA species, less information is available and only the work of Ishikawa [41] reports the time-evolution of the relative amounts of R_C and R_- respective to R_+. As such, the experimental data of [41] were used to obtain quantitative values of the total amounts of

different RNA species in a cell. To transform the relative molar amounts or R_+ in [41] into absolute amounts, it was accepted that, by the end of the infection process, 90% of the R_+ molecules had been encapsulated [42]. Consequently, the last of the data for R_+ shown by Ishikawa [41] was given a value equal to 90% of Aoki's [42] viral particles ($1.05 \cdot 10^6$). Based on this value, the R_+ concentrations were calculated proportionally.

Once calculated, the total concentrations of R_+, the values of R_C and R_- were calculated proportionally from the data by Ishikawa et al.:

$$R_C = R_+ \cdot \frac{RMAR_C}{RMAR_+}$$

In these expressions, RMAR represents the relative molar amount used by Ishikawa [41]. R_- was calculated in a similar way. The total amounts of AR and P_C were obtained from Table 1 in [43]. Note that when Sakai and Takebe [43] published these data, it was still unknown that the 140000 m.w. protein was the viral replicase. The identification of this entity as the replicase gene came several years later [41], [44]. Those authors did not express their results as the numbers of molecules per protoplast, but reported the measured radioactivity (in counts-per-minute or cpms) per 10^5 protoplasts. In order to obtain the actual cellular concentrations, we had to convert those cpms into numbers of molecules. The authors used a ^{14}C labeled Leucine with a total activity of 298 Ci/mol. According to the GeneBank, the coat protein of TMV (accession number P03571.2) has 12 leucines and the A_r (accession number NP056764) has 109 leucines. The technique used for the quantification of the radioactivity emitted by these molecules was the one devised by Tishler and Epstein [46]. Sakai and Takebe [43] did not report the efficiency of the quantification of radioactivity in macromolecules. We assumed it was the same as the one given by [46], which was 80% for ^{14}C.

Thus, the cellular concentration of the total P_C was calculated as:

$$P_C = \frac{\text{Xcpm}}{\text{Efficiency}} \times \frac{1\text{Ci}}{2.22 \cdot 10^{12}\,\text{cpm}} \times \frac{1\,\text{mole Leucine}}{298\text{Ci}} \times 6.023 \cdot 10^{23} \times \frac{\text{Leucine}}{\text{mole Leucine}} \times \frac{1\,\text{P}_C}{12\text{Leucine}} \times \frac{1}{10^5\,\text{Protoplasts}}$$

where Xcpm is the radioactivity data in [43]. A similar equation was written for A_r, as well.

Finally, we must estimate the cellular concentrations of free P_C and R_+, assuming that both species are either free or in viral particles. We do not take into account PC multimers that are also not considered in our kinetic model, nor are there any available data on their concentrations. We had to make some assumptions in order to get an estimate of the free P_C and R_+. The cellular concentration of R_+ was calculated as the difference between the total R_+ and the number of viral particles at any given time. The free P_C in the protoplast was taken to be the difference between the total P_C number, obtained as shown above, and the number of viral particles multiplied by m (i.e., the number of P_C molecules in the capsid = 2130).

These calculations enabled the assembly of a single set of data shown in Table (1), giving the evolution over time of most of the variables appearing in the present model. Although the data we are using were obtained from various researchers who used different methods (and they might be subject to some estimation error), they were all obtained with the same virus/cell system. As such, we consider them to be the best available data set of this kind.

V. Computational Approach

The model presented in (19)-(27) was used to determine the numerical values of the kinetic parameters: k_1, k_2, etc. fitting the experimental data shown in Table 1. At short times (t < 3 h), the values of most of the variables could not be detected by the experimental methods employed. Consequently, only the data corresponding to t = 3 h = 1080 s and higher were used to determine the parameter values.

The parameter estimation problem may be formulated [47] as follows:

$$\min_{\theta} \Phi(\varphi) = \sum_{\mu=1}^{n} \sum_{i=1}^{m} \left(x_{\mu,i} - x_{\mu,i}^c \right)^2 * W_i \tag{28}$$

subject to:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, \varphi, t)$$
$$\mathbf{x}(0) = \mathbf{x}_0 \tag{29}$$

where **F** is a system of m ordinary differential equations (20-25), **x** is a vector of m dynamic variables (R_-, R_+, A_r , R_C, P_C and V, m = 6), **φ** is a vector of p parameters (k1, k2,…, p = 16), $\mathbf{x}_\mu$ is a vector of m observed values at the μ^{th} data point, $\mathbf{x}_\mu^c$ is a vector of m calculated variable values at the μ^{th} data point, and **W** is a vector of m weighting factors.

Here, the weighting factors, W_i, are used to bring the contributions of the variables of the different orders of magnitude (like R_- and A_r, see Table 1) to similar levels in the objective function. Consequently, weighting factors of the form:

$$W_i = 1 / \max_{\mu} \left| x_{\mu,i} \right| \tag{30}$$

were used. This formulation actually scales all the variables in the objective function to the [-1, 1] interval (hereafter called 'normalization').

The calculated values of the variables $\mathbf{x}_{\mu i}^c$ are obtained by numerically integrating the differential equations from $t_{\mu-1}$ to t_μ for μ = 2, 3, … 22. The integrations of the differential equations were usually carried out by the MATLAB[1] *ode45* function, based on an explicit Runge-Kutta (4, 5) formula--the Dormand-Prince [48] pair. The initial, estimated parameter values that were far from optimal often caused the equations to become 'stiff'. The MATLAB library function *ode15s* was

[1]MATLAB is a product of MathWorks, Inc., <http://www.mathworks.com>

used to integrate the 'stiff' systems, based on the backward difference formula (BDF) method of Gear [49].

The minimization defined by (28) was carried out using the MATLAB *nlinfit* function, based on the Levenberg-Marquardt (LM) algorithm [50]-[51]. This is a gradient-based method, in which derivatives of the objective function and of the constraints must be evaluated.

VI. Numerical results

The minimization of the objective function, when allowing for the variation of all the parameters (p = 16), caused computational difficulties, such as rank deficiency of the matrix of partial derivatives, signaling very low sensitivity of the objective function to some of the parameters. To eliminate such difficulties, the values of some of the low-impact parameters were assumed and others were found by trial and error. The kinetic constants for the hydrolysis of all the proteins (k_{11}, k_{12}, k_{13} and k_{15}) have been given the same value, 10^{-8} [s^{-1}], as assumed by others [33]. The contribution of the hydrolysis term in the differential equations is a very slow consumption of the protein, only detectable over long periods of time. The kinetic constant for the antisense RNA binding between the R_- and the R_C has been arbitrarily taken to be $k_{16}=10^{-14}$ [cell.s^{-1}.molecule^{-1}]. There are no quantitative data on this; it is only known that the amounts of R_- and $\underline{R_C}$ involved are very low. Equation (26) is, therefore, included only to improve the description of the kinetics from a qualitative perspective.

The optimal values of the remaining eight parameters were identified by the minimization of the objective function defined by (28) and (29). These optimal values are shown in Table 2. The sum of the squares of the errors in this solution is: $\Phi(\theta) = 1.1$.

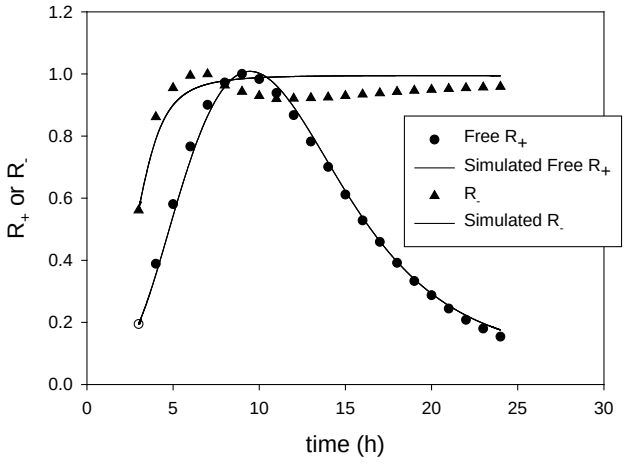

Fig. 2: Pools of the positive (R_+) and the negative (R_-) strands of RNA in a protoplast as obtained from the mathematical model using the optimal parameters, compared to the corresponding values in the experimental data set.

Table 2. Optimal values of the kinetic constants.

Parameter	Optimal value
k_1	$7.99*10^{-13}$
k_2	0.021
k_3	$2.34*10^{-13}$
k_4	$1.50*10^{-4}$
k_5	$1.19*10^{-6}$
k_6	0.022
k_7	$1.84*10^{-9}$
k_8	0.013
k_9	$5.28*10^{-11}$
k_{10}	$1.10*10^{-17}$

VII Interpretation of the results

Figs. 2-4 show the profiles in the cellular pools of the molecules considered as variables in our model during the period from 3-24 h. Figure 2 shows the two RNA strands, R_+ and R_-.

Both concentrations are presented as fractions of their maximal values during the time interval considered.For the positive sense RNA, the agreement between the model and the data is excellent. The model does not fully represent the complex shape of the R_- curve, but the predicted values are, nevertheless, very close to the data-base figures over the entire time range. The deviation is on the order of 10%, which is small in relation to the experimental errors found in our data-bank sources.

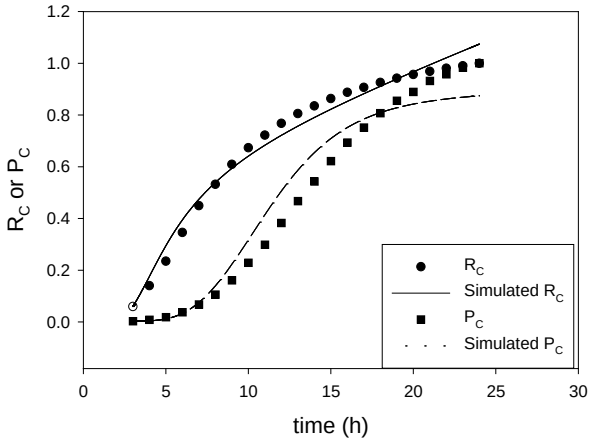

Fig. 3: Pools of the sub-genomic mRNA for coat protein (R_C) and the coat protein (P_C) as obtained from the mathematical model using the optimal parameters, compared to the corresponding values in the experimental data set.

Figure 3 presents the results for pools of sub-genomic mRNA coat-protein translation, R_C, and the pool of coat protein itself, P_C. Both sets of data are satisfactorily fit in shape and in numerical value until approximately 20 h, which is most of the range.

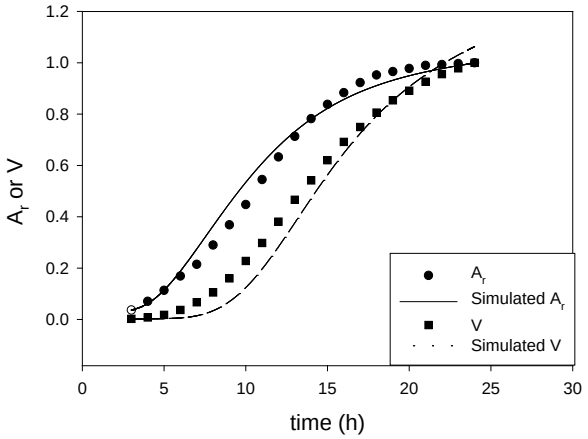

Fig. 4: Pools of the protein replicase (A_r) and the viruses (V) as obtained from the mathematical model using the optimal parameters, compared to the corresponding values in the experimental data set.

The maximal deviations at the highest points of time were still around 10%. Figure 4 shows the profiles of the replicase pool and of the newly assembled viruses. Again, both the correct shape of the curve and a very good approximation to the numerical values is observed, except for the last hours of the run, for times longer than 20 h.

While, in general, the fit of the experimental data by the proposed model is quite satisfactory, the deviations obtained for times longer than 20 h seem to indicate that the simplified kinetic schemes included in our model are unable to describe the extreme complexity of the real process over the whole time range under consideration. However, further refinement of our model would inevitably lead to the addition of new variables and more kinetic constants, which would diminish its reliability; currently sufficient experimental values are lacking for proper validation.

Our model provides also a profile of the pool of dsRNA. Those are not presented here because currently there are no available experimental data to compare. The calculated results show a sigmoid curve, as expected from the profiles of R_C and R- that generate the dsRNA, as indicated in (26). The absolute value of this pool depends on k_{16}, which cannot be evaluated, as stated previously. The use of antisense molecules, targeted at specific mRNAs, is a proven instrument for the study of virus growth dynamics at the molecular level [21] and of gene silencing [52]-[54] and we feel that even the qualitative description of this profile may be of interest.

The overall fit of our model to the data-set may, therefore, be considered satisfactory; this strongly supports the ability of this model to represent the actual behavior of the virus in a host cell and provides a strong base for its extension to the simulation of systemic propagation in plant tissue.

VII. Conclusions

A mathematical model describing the dynamics of the biochemical reactions in a protoplast after infection by TMV has been formulated. The variables considered are: the cellular pools of the positive and negative strands of RNA; the fundamental enzyme replicase; the coat protein; the sub-genomic RNA for the R_C; and the assembled virus. Therefore, a coherent mechanistic description of the intracellular process is obtained. It was found that a simple function of the negative RNA pool provides a satisfactory description of the fraction of R_+ that is devoted to translation and of that devoted to replication. Experimental data from several sources were integrated into a single set, used for the evaluation of the kinetic constants. The satisfactory fit of the presented mathematical model to the set of data indicates that it may serve as a useful tool that can be instrumental in a more general description of the systemic propagation of TMV in vegetal tissue and other applications.

References

[1] G. Conti, M. C. Rodríguez, C. A. Manacorda, and S. Asurmendi, "Transgenic expression of TMV capsid and movement proteins modulate plant basal defense and biotic stress responses in *Nicotiana tabacum*," *Mol. Plant-Microbe Interact.*, vol. 25, no. 10, Oct. 2012, pp. 1370-1384.

[2] A. A. Bazzini, S. Asurmendi, H. E. Hopp, and R. N. Beachy, "Tobacco mosaic virus (TMV) and potato virus X (PVX) coat proteins confer heterologous interference to PVX and TMV infection, respectively," *J. Gen. Virol.*, vol. 87, no. 4, Apr. 2006, pp. 1005-1012.

[3] P. M. Doran, "Foreign protein production in plant tissue cultures," *Curr. Opin. Biotechnol.*, vol. 11, no. 2, Apr. 2000, pp. 199-204.

[4] J. M. Sharp, and P. M. Doran, "Characterization of monoclonal antibody fragments produced by plant cells," *Biotechnol. Bioeng.*, vol. 73, no. 5, Jun. 2001, pp. 338-346.

[5] J. M. Sharp, and P. M. Doran, "Strategies for enhancing monoclonal antibody accumulation in plant cell and organ cultures," *Biotechnol. Progr.*, vol. 17, no. 6, Nov.-Dec. 2001, pp. 979-992.

[6] A. K. Pavlou, and J. M. Reichert, "Recombinant protein therapeutics: Success rates, market trends and values to 2010," *Nat. Biotechnol.*, vol. 22, 2004, pp. 1513-1519.

[7] Y. Gleba, V. Klimyuk, and S. Marillonnet, "Viral vectors for the expression of proteins in plants," *Curr. Opin. Biotechnol.*, vol. 18, no. 2, Mar. 2007, pp. 134-141.

[8] C. Lico, Q. Chen, and L. Santi, "Viral vectors for production of recombinant proteins in plants," *J. Cell. Physiol.*, vol. 216, no. 2, Aug. 2008, pp. 366-377.

[9] F. S. Shadwick, and P. M. Doran, "Infection, propagation, distribution and stability of plant virus in hairy root cultures," *J. Biotechnol.*, vol. 131, no. 3, Sep. 2007, pp. 318-329.

[10] F. S. Shadwick, and P. M. Doran, "Propagation of plant viruses in hairy root cultures: A potential method for in vitro production of epitope vaccines and foreign proteins," *Biotechnol. Bioeng.*, vol. 96, no. 3, Feb 2007, pp. 570-583.

[11] M. A. Nowak, and C. R. M. Bangham, "Population dynamics of immune responses to persistent viruses," *Science: New Series*, vol. 272, no. 5258, Apr. 1996, pp. 74-79.

[12] M. A. Nowak, S. Bonhoeffer, A. M. Hill, R. Boehme, H. C. Thomas, and H. McDade, "Viral dynamics in hepatitis B virus infection," *PNAS USA*, vol. 93, no. 9, Apr. 1996, pp. 4398-4402.

[13] S. Bonhoeffer, R. M. May, G. M. Shaw, and M. A. Nowak, "Virus dynamics and drug therapy," *PNAS USA*, vol. 94, no. 13, Jun. 1997, pp. 6971-6976.

[14] C. R. M. Bangham, *et al.*, "Genetic control and dynamics of the cellular immune response to the human T-cell leukaemia virus, HTLV-I," *Phil. Trans. R. Soc. Lond.* B, vol. 354, no. 1384, 1999, pp. 691-700.

[15] D. Wodarz, M. A. Nowak, and C. R. M. Bangham, C.R.M., "The dynamics of HTLV-I and the CTL response," *Immunol. Today*, vol. 20, no. 5, May 1999, pp. 220-227.

[16] C. R. M. Bangham, R. M. May, and M. A. Nowak, "The dynamics of virus infection: Preface," *Phil. Trans. R. Soc. Lond.* B, vol. 355, no. 1400, Aug. 2000, p. 1005.

[17] G. Enden, Y. H. Zhang, and J. C. Merchuk, "A model of the dynamics of insect cell infection at low multiplicity of infection," *J. Theor. Biol.*, vol. 237, no. 3, Dec. 2005, pp. 257-264.

[18] F. Martínez, J. Sardanyés, S. F. Elena, and J. A. Daròs, "Dynamics of plant RNA virus intracellular accumulation: Stamping machine vs. geometric replication," *Genetics*, vol. 188, 2011, pp. 637-646.

[19] T. J. Wickham, M. L. Shuler, and D. A. Hammer, D.A., "A simple model to predict the effectiveness of molecules that block attachment of human rhinoviruses and other viruses," *Biotechnol. Progr.*, vol. 11, no. 2, Mar.-Apr. 1995, pp. 164-170.

[20] F. Buchholtz, and F. W. Schneider, "Computer simulation of T3/T7 phage infection using lag times," *Biophys. Chem.*, vol. 26, no. 2-3, May 1987, pp. 171-179.

[21] D. Endy, D. Kong, and Y. John, "Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage T7," *Biotech. Bioeng*., vol. 55, no. 2, 1997, pp. 375-389.

[22] M. Eigen, C. K. Biebricher, M. Gebinoga, and W. C. Gardiner, "The hypercycle coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage," *Biochemistry*, vol. 30, no. 46, Nov. 1991, pp. 11005-11018.

[23] B. Reddy, and J. Yin, "Quantitative intracellular kinetics of HIV type 1," *AIDS Res. Hum. Retroviruses*, vol. 15, no. 3, Feb. 1999, pp. 273-283.

[24] R. Srivastava, L. You, J. Summers, and J. Yin, "Stochastic vs. deterministic modeling of intracellular viral kinetics," *J. Theor. Biol.*, vol. 218, no. 3, Oct. 2002, pp. 309-321.

[25] L. You, P. F. Suthers, and J. Yin, "Effects of *Escherichia coli* physiology on growth of phage T7 *in vivo* and *in silico*," *J. Bacteriol.*, vol. 184, no. 7, Apr. 2002, pp. 1888-1894.

[26] Y. Sidorenko, and U. Reichl, "Structured model of influenza virus replication in MDCK cells," *Biotechnol. Bioeng.*, vol. 88, no. 1, Oct. 2004, pp. 1-14.

[27] K.-I. Lim, T. Lang, V. Lam, and J. Yin, "Model-based design of growth-attenuated viruses," *PLoS Comput. Biol.*, vol. 2, no. 9, Sep. 2006, pp. 1001-1016.

[28] H. Dahari, R. M. Ribeiro, C. M. Rice, and A. S. Perelson, "Mathematical modeling of subgenomic hepatitis C virus replication in Huh-7 cells," *J. Virol*., vol. 81, no. 2, Jan. 2007, pp. 750-760.

[29] J. Sardanyés, R. V. Solé, and S. F. Elena, "Replication mode and landscape topology differentially affect RNA virus mutational load and robustness," *J. Virol*., vol. 83, no. 23, Dec. 2009, pp. 12579-12589.

[30] K. U. Dee, and M. L. Shuler, "A mathematical model of the trafficking of acid-dependent enveloped viruses: Application to the binding, uptake, and nuclear accumulation of baculovirus," *Biotechnol. Bioeng.*, vol. 54, no. 5, Jun. 1997, pp. 468-490.

[31] K. U. Dee, D. A. Hammer, and M. L. Shuler, "A model of the binding, entry, uncoating, and RNA synthesis of Semliki Forest virus in baby hamster kidney (BHK-21) cells," *Biotechnol. Bioeng.*, vol. 46, no. 5, Jun. 1995, pp. 485-496.

[32] D. Holcman, "Modeling DNA and virus trafficking in the cell cytoplasm," *J. Stat. Phys.*, vol. 127, no. 3, May 2007, pp. 471-494.

[33] F. Ji, and L. Luo, "A hypercycle theory of proliferation of viruses and resistance to the viruses of transgenic plant," *J. Theor. Biol.*, vol. 204, no. 3, Jun. 2000, pp. 453-465.

[34] R. B. Schinazi, "A spatial stochastic model for virus dynamics," *J. Stat. Phys.*, vol. 128, no. 3, Aug. 2007, pp. 771-779.

[35] S. Asurmendi, R. H. Berg, J. C. Koo, and R. N. Beachy, "Coat protein regulates formation of replication complexes during tobacco mosaic virus infection," *PNAS USA*, vol. 101, no. 5, Feb. 2004, pp. 1415-1420.

[36] S. Kawakami, Y. Watanabe, and R. N. Beachy, "Tobacco mosaic virus infection spreads cell to cell as intact replication complexes," *PNAS USA*, vol. 101, no. 16, Apr. 2004, pp. 6291-6296.

[37] M. P. Zwart, A. Daròs, and S. Elena, " One is enough: in vivo effective population size in dose-dependent for plant RNA virus," Plos Pathog. vol. 7, no. 7. Jul. 2011, pp. 1-12.

[38] J. N. Culver, "Tobacco mosaic virus assembly and disassembly: Determinants in pathogenicity and resistance," *Annu. Rev. Phytopathol.*, vol. 40, Feb. 2002, pp.287-308.

[39] J. N. Culver, G. Stubbs, and W. O. Dawson, "Structure-function relationship between tobacco mosaic virus coat protein and hypersensitivity in *Nicotiana sylvestris*," *J. Mol. Biol.*, vol. 242, no. 2, Sep. 1994, pp. 130-138.

[40] K. W. Buck, "Replication of tobacco mosaic virus RNA," *Phil. Trans. R. Soc. Lond*. B, vol. 354, no. 1383, Mar. 1999, pp. 613-627.

[41] M. Ishikawa, T. Meshi, T. Ohno, and Y., Okada, "Specific cessation of minus-strand RNA accumulation at an early stage of tobacco mosaic virus infection," *J. Virol.*, vol. 65, no. 2, Feb. 1991, pp. 861-868.

[42] S. Aoki, and I., Takebe, "Replication of tobacco mosaic virus RNA in tobacco mesophyll protoplasts inoculated *in vitro*," *Virology*, vol. 65, no. 2, Jun. 1975, pp. 343-354.

[43] S. Aoki, and I., Takebe, "Protein synthesis in tobacco mesophyll protoplasts induced by tobacco mosaic virus infection," *Virology*, vol. 62, no. 2, Dec. 1974, pp. 426-433.

[44] M. Ishikawa, T. Meshi, F. Motoyoshi, N. Takamalsu, and Y. Okada, "*In vitro* mutagenesis of the putative replicase genes of tobacco mosaic virus," *Nucl. Acids Res.*, vol. 14, no. 21, Nov. 1986, pp. 8291-8305.

[45] G. J. Hills, *et al.* "Immunogold localization of the intracellular sites of structural and nonstructural tobacco mosaic virus proteins," *Virology*, vol. 161, no. 2, 1987, pp. 488-496.

[46] V. P. Tishler, and C. J. Epstein, "A convenient method of preparing polyacrylamide gels for liquid scintillation spectrometry," *Anal. Biochem.*, vol. 22, no. 1, Jan. 1968, pp. 89-98.

[47] M. Shacham, and N. Brauner, N., "Application of stepwise regression for dynamic parameter estimation," *Comp. & Chem. Eng.*, vol. 69, 2014, pp. 26-38.

[48] C. W. Gear, *Numerical initial value problems in ordinary differential equations*. Englewood Cliffs, NJ: Prentice Hall, 1971.

[49] J. A. Nelder, and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, 1965, pp. 308-313.

[50] J. R. Dormand, and P. J. Prince, "A family of embedded Runge-Kutta formulae," *J. Comp. Appl. Math.*, vol. 6, no. 1, Mar. 1980, pp. 19-26.

[51] G. A. F. Seber, and C. J. Wild, *Nonlinear regression*. Hoboken, NJ: Wiley-Interscience, 2003.

[52] T. Sijen, and J. M. Kooter, "Post-transcriptional gene-silencing: RNAs on the attack or on the defense?" *Bioessays*, vol. 22, no. 6, Jun. 2000, pp. 520-531.

[53] T. Sijen, *et al.* "On the role of RNA amplification in dsRNA-triggered gene silencing," *Cell*, vol. 107, no. 4, Nov. 2001, pp. 465-476.

[54] C. T. Bergstrom, E. McKittrick, and R. Antia, "Mathematical models of RNA silencing: Unidirectional amplification limits accidental self-directed reactions," *PNAS USA*, vol. 100, no. 20, Sep. 2003, pp. 11511-11516.

Apendix: Table 1. Elaborated experimental data.

Var. No. → Point No.	time (s)	1 R_-	2 R_+	3 R_C	4 V	5 P_c	6 A_r
1	10800	2495	2.00E+05	1.60E+05	2.84E+03	2.99E+05	3.64E+07
2	14400	3194	3.49E+05	2.60E+05	8.96E+03	9.42E+05	6.98E+07
3	18000	3793	5.59E+05	4.99E+05	2.01E+04	2.11E+06	1.12E+08
4	21600	4092	7.99E+05	6.99E+05	4.13E+04	4.34E+06	1.67E+08
5	25200	4033	9.38E+05	8.78E+05	7.46E+04	7.85E+06	2.12E+08
6	28800	3933	1.06E+06	1.06E+06	1.17E+05	1.23E+07	2.86E+08
7	32400	3753	1.13E+06	1.15E+06	1.79E+05	1.88E+07	3.64E+08
8	36000	3593	1.19E+06	1.20E+06	2.54E+05	2.67E+07	4.42E+08
9	39600	3593	1.24E+06	1.35E+06	3.32E+05	3.49E+07	5.38E+08
10	43200	3653	1.29E+06	1.43E+06	4.25E+05	4.47E+07	6.25E+08
11	46800	3713	1.34E+06	1.52E+06	5.19E+05	5.46E+07	7.04E+08
12	50400	3753	1.38E+06	1.58E+06	6.04E+05	6.35E+07	7.72E+08
13	54000	3793	1.40E+06	1.66E+06	6.91E+05	7.27E+07	8.27E+08
14	57600	3793	1.42E+06	1.72E+06	7.71E+05	8.11E+07	8.72E+08
15	61200	3773	1.43E+06	1.78E+06	8.35E+05	8.79E+07	9.11E+08
16	64800	3753	1.44E+06	1.81E+06	8.98E+05	9.44E+07	9.40E+08
17	68400	3713	1.42E+06	1.83E+06	9.51E+05	1.00E+08	9.53E+08
18	72000	3673	1.40E+06	1.83E+06	9.93E+05	1.04E+08	9.65E+08
19	75600	3593	1.38E+06	1.82E+06	1.03E+06	1.09E+08	9.77E+08
20	79200	3494	1.34E+06	1.81E+06	1.06E+06	1.12E+08	9.80E+08
21	82800	3294	1.30E+06	1.78E+06	1.09E+06	1.15E+08	9.84E+08
22	86400	2994	1.24E+06	1.76E+06	1.11E+06	1.17E+08	9.87E+08