Research article

Mathematical modelling of OAS2 activation by dsRNA and effects of dsRNA lengths

Deokro Lee[1], Amit Koul[2], Nikhat Lubna[2], Sean A. McKenna[2] and Stéphanie Portet[1,*]

[1] Department of Mathematics, University of Manitoba, Winnipeg, MB R3B 2N2, Canada

[2] Department of Chemistry, University of Manitoba, Winnipeg, MB R3B 2N2, Canada

* **Correspondence:** Email: Stephanie.Portet@umanitoba.ca; Tel: +1 204 474 6933; Fax: +1 204 474 7611.

Abstract: The activation of 2'-5'-oligoadenylate synthetase (OAS) enzymes by direct interaction with viral double-stranded RNA (dsRNA) is a key part of the innate immune response to viral infection. A downstream effect of the OAS-dsRNA interaction is to degrade the single-stranded RNA to prevent the spread of the virus. The activation of OAS2, one of the members of the OAS family, depends on dsRNA length. Combining *in vitro* experiments and mathematical modelling, we test different hypotheses for the OAS2 activation mechanisms by its cofactor dsRNA. After model calibration and selection, the cooperative binding of multiple OAS2 to a single dsRNA is shown to best represent the effect of its cofactor length on enzyme activity.

1. Introduction

Host detection of viral pathogens by the innate immune response is an effective first line of defence against infection. In eukaryotic cells, the innate immune system relies heavily on protein recognition of evolutionarily conserved viral double-stranded RNA (dsRNA) to stimulate interferon production that prepares neighbouring cells for attack [1]. The interferon response then up-regulates the transcription of additional antiviral proteins to amplify the response. A key family of interferon-stimulated antiviral enzymes are the 2'-5'-oligoadenylate synthetases (OAS). OAS enzymes specifically bind viral dsRNA and through a complex cascade attenuate viral replication. An understanding of how these enzymes discriminate viral from host dsRNA remains a key unresolved question.

OAS enzymes require viral dsRNA as a cofactor to polymerise ATP into chains of unusual 2'-

5' linked oligoadenylates (2-5A) [2]. These chains bind to and activate RNase L, which degrades cellular and viral RNA thereby restricting viral propagation [3]. In humans, three genes (OAS1/2/3) each produce enzymes with one, two and three repeats of a 346 amino acid core OAS unit [4]. Each version has different dsRNA affinity, builds different 2-5A chain lengths, and responds to different viral infections. Interaction between multiple OAS enzymes (oligomerisation) appears to be central to these differences, presumably to respond to different viral dsRNA lengths. OAS1, comprising a single domain, forms monomer, dimers and tetramers, and will bind dsRNA of any length longer than 17 base pairs (bp) [5]. OAS2 can form a dimer and requires a minimum of 35 bp [6]. OAS3 is not known to oligomerize and requires at least 50 bp dsRNA for activation [7]. While high-resolution structures of OAS enzymes show that dsRNA is responsible for dramatic reorganisation of the enzyme active site, little is known regarding the impact of dsRNA binding on oligomerisation or the detailed catalytic mechanism.

OAS enzymes are differentially expressed in specific cell types, and their presence in different sub-cellular fractions such as mitochondria, nuclear, and rough/smooth microsomal fractions suggest that these proteins may have distinct functional roles. OAS1, 2, and 3 differ significantly in their response to common members of the Flaviviridae family of human viral pathogens including Yellow fever, Dengue fever, Japanese encephalitis, Tick-borne encephalitis and West Nile virus. Preliminary data from [6] suggests that OAS1 and OAS2 may have distinct substrate specificity in terms of dsRNA length, which may explain their divergent roles in antiviral processes.

The OAS enzymatic mechanism is complex as it involves dsRNA cofactors of variable length and structure, an ATP substrate that is polymerised into chains of variable length, and non-covalent OAS self-association. As a result of this complexity, the detailed biochemical process is not completely defined. In this work, we focus on OAS2 activation by dsRNA. Based on *in vitro* experimental data published in [6], the process appears to depend on concentrations and lengths of dsRNA; however, limitations of the experimental data that can only be accurately recorded at the endpoint of enzymatic activity prevents complete definition of the process. Hence, mathematical modelling is used to investigate possible OAS2 activation mechanisms and the effects of dsRNA lengths. Plausible biochemical scenarios are translated to mathematical models and their responses are compared to *in vitro* experimental data to test different hypotheses. In total, three kinetic models are derived. After calibration of models to data, a model selection method is used to determine the model that best represents the data. Furthermore, analysing parameter estimates allows us to propose a mechanism accounting for the role of dsRNA length in OAS2 activation.

2. Experimental data

OAS2 *enzymes* require dsRNA as a *cofactor* for enzymatic activity. In the presence of *substrate* ATP, OAS2 binds to dsRNA and the enzyme polymerises ATP into chains of 2-5A leaving pyrophosphate (PPi) as a byproduct [2]. *In vitro* experimental data considered in this work are concentrations of PPi recorded at six time points (including the initial time, $t_0 = 0$ min). The byproduct concentration is measured at $t_1 = 5$ min, $t_2 = 10$ min, $t_3 = 15$ min, $t_4 = 20$ min and $t_5 = 30$ min for one concentration $c_O = 0.3 \mu g/mL$ of OAS2, and five concentrations of dsRNA are considered with seven different lengths [6]. The five dsRNA concentrations c_j are $c_1 = 0.5 \mu g/mL$, $c_2 = 1 \mu g/mL$, $c_3 = 2 \mu g/mL$, $c_4 = 4 \mu g/mL$ and $c_5 = 8 \mu g/mL$. The seven lengths L_k considered for dsRNA are

$L_1 = 40\text{bp}$, $L_2 = 50\text{bp}$, $L_3 = 60\text{bp}$, $L_4 = 70\text{bp}$, $L_5 = 80\text{bp}$, $L_6 = 90\text{bp}$ and $L_7 = 120\text{bp}$. The ATP concentration used in the experiments is $1\text{mg/mL} \gg c_O$. Experimental data used in this work have been published in [6]; for details on *in vitro* experiments yielding them refer to [6].

3. Modelling

In the *in vitro* experiments considered [6], the *substrate*, which is ATP, is in abundance. Hence, the contribution of substrate to enzyme kinetics is not of interest in this work; no equation describes ATP concentration in models. The OAS2 enzyme consists of two OAS domains [8]. As shown in Koul *et al.* [6], a minimal length of 35bp for dsRNA is required to activate OAS2, and both OAS2 domains must be bound simultaneously by dsRNA to lead to product formation. Furthermore, the binding and unbinding of enzymes and cofactors are reversible reactions; neither enzyme nor cofactor are consumed during these reactions nor the formation of product and byproduct. Finally, as previously mentioned, the experimental data only monitors the concentration over time of the byproduct (PPi) for a given enzyme concentration and different cofactor concentrations and lengths. In the experimental data, the formation of byproduct is sensitive to both cofactor concentrations and lengths.

Based on these observations and available *in vitro* experimental data three mathematical models are developed to investigate the activation of OAS2 by dsRNA and the role of the dsRNA length on the activation (Figure 1).
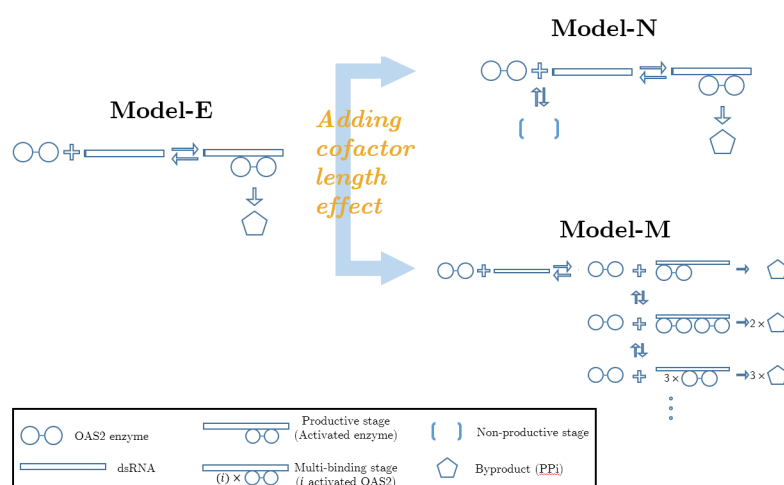

Figure 1. Three models considered to investigate OAS2 activation by dsRNA.

3.1. Model-E

The first model considers an enzymatic kinetics describing only the kinetics of the enzyme E and cofactor R as the substrate is in abundance. The activated enzyme or productive stage D, which is formed by the OAS2 and the cofactor bound to the two domains of OAS2, is not consumed in the formation of byproduct P. Assumptions of Model-E are as follows (Figure 1)

$$E + R \underset{k_{1m}}{\overset{k_1}{\rightleftharpoons}} D$$

$$\emptyset \xrightarrow{k_p D} P$$
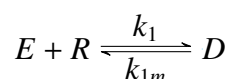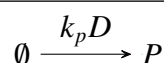
and the model equations are

$$
\begin{aligned}
\frac{dE}{dt} &= -k_1 ER + k_{1m} D, \\
\frac{dR}{dt} &= -k_1 ER + k_{1m} D, \\
\frac{dD}{dt} &= k_1 ER - k_{1m} D, \\
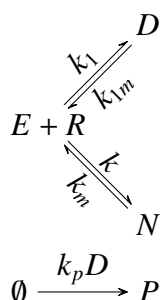\frac{dP}{dt} &= k_p D,
\end{aligned}
\tag{3.1}
$$

where k_1 is the rate of association of an OAS2 and dsRNA, it describes the rate of production of the compound D or activated enzyme. To reduce the model complexity, binding of the two domains of OAS2 to the same dsRNA is not explicitly described and is approximated as an one-step binding. The rate k_{1m} is the dissociation rate of the enzyme and cofactor. Note that in Model-E once a dsRNA molecule is bound with an OAS2, it cannot accommodate another OAS2 enzyme. The parameter k_p is the rate of formation of the byproduct by the activated enzyme. Parameters are listed in Table 1. Model-E is considered with non-negative initial conditions $E(0) = c_O$, $R(0) = c_R$, $D(0) = 0$ and $P(0) = 0$, where the positive constants c_O and c_R are the initial concentrations of OAS2 and dsRNA, respectively.

Table 1. Parameters used in models. (Top) Fixed parameters whose values are set from experimental conditions [6]. (Bottom) Free parameters whose values are estimated by calibrating models responses to experimental data.

Parameter	Description (units)	Model-
c_O	Initial concentration of OAS2 (0.3μg/mL)	E, N, M
c_R	Initial concentration of dsRNA (0.5, 1, 2, 4 and 8μg/mL)	E, N, M
L	Length of dsRNA (40, 50, 60, 70, 80, 90 and 120 bp)	M
$2R_O$	Diameter of OAS2 (35 bp)	M
k_1	Association rate of an OAS2 enzyme to an unoccupied dsRNA ((μ g/mL)$^{-1}$.min^{-1})	E, N, M
k_{1m}	Dissociation rate of the OAS2 enzyme and dsRNA (min^{-1})	E, N ,M
k_i	Association rate of the ith enzyme to an dsRNA ((μ g/mL)$^{-1}$.min^{-1})	M with $i = \{2, \cdots, n\}$, where $n = \lfloor \frac{L}{2R_o} \rfloor$
k_{im}	Dissociation rate of ith enzyme from an dsRNA (min^{-1})	M with $i = \{2, \cdots, n\}$, where $n = \lfloor \frac{L}{2R_o} \rfloor$
k	Association rate of a non-productive stage ((μ g/mL)$^{-1}$.min^{-1})	N
k_m	Dissociation rate of non-productive stage (min^{-1})	N
k_p	Rate of byproduct formation (min^{-1})	E, N, M

3.2. Model-N

A way to represent the impact of the dsRNA length on OAS2 activation is to consider that some interactions are not complete [9] as the two domains of OAS2 have to bind to dsRNA to produce an active OAS2. Hence, the formation of compounds in which both domains are not properly attached to cofactor is hypothesised. These compounds are called non-productive compounds, their specific forms are not described. Based on this assumption, the non-productive binding stage N that cannot result in the formation of the product and byproduct is added to Model-E. After a non-productive binding, OAS2 can detach from dsRNA and attach to dsRNA again. The assumptions for Model-N are as follows (Figure 1)
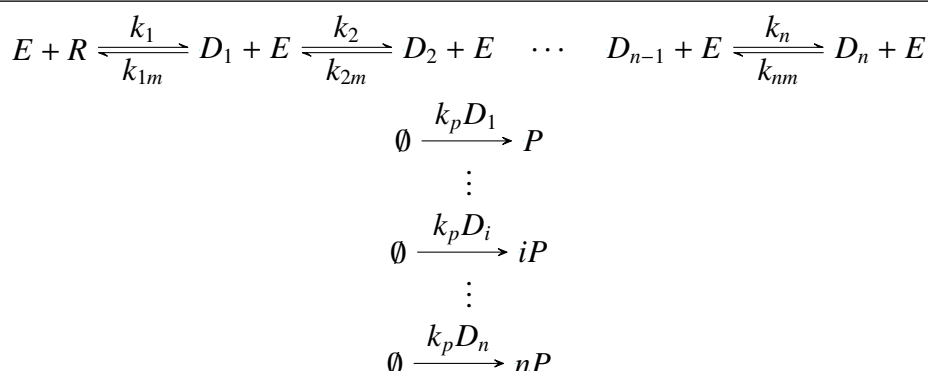

yielding the model equations

$$
\begin{aligned}
\frac{dE}{dt} &= -k_1 ER + k_{1m} D - kER + k_m N, \\
\frac{dR}{dt} &= -k_1 ER + k_{1m} D - kER + k_m N, \\
\frac{dD}{dt} &= k_1 ER - k_{1m} D, \\
\frac{dN}{dt} &= kER - k_m N, \\
\frac{dP}{dt} &= k_p D,
\end{aligned}
\tag{3.2}
$$

where the additional parameters k and k_m are the association and dissociation rates of the non-productive compound N, respectively (Table 1). Model-N is considered with the initial conditions $E(0) = c_O$, $R(0) = c_R$, $D(0) = N(0) = 0$ and $P(0) = 0$.

3.3. Model-M

Another possible effect of cofactor length on OAS2 activation is the occurrence of multi-binding of OAS2 on a same dsRNA. dsRNA is not consumed nor changed by the OAS2 activation; OAS2 interacts with dsRNA, facilitates product formation, and then detaches from dsRNA. In the *in vitro* experiments considered, the longer the dsRNA used, the more product formation observed at a given OAS2 concentration [6]. Hence, another reasonable scenario is that longer dsRNA able to accommodate more than one OAS2 (presumably > 70 bp) can be bound with more than one OAS2 simultaneously, leading to an increased product formation. Hence, the multi-binding assumption that allows up to n OAS2 bound to one dsRNA simultaneously, is summarised as follows (Figure 1):

$$E + R \underset{k_{1m}}{\overset{k_1}{\rightleftharpoons}} D_1 + E \underset{k_{2m}}{\overset{k_2}{\rightleftharpoons}} D_2 + E \quad \cdots \quad D_{n-1} + E \underset{k_{nm}}{\overset{k_n}{\rightleftharpoons}} D_n + E$$

$$\emptyset \xrightarrow{k_p D_1} P$$
$$\vdots$$
$$\emptyset \xrightarrow{k_p D_i} iP$$
$$\vdots$$
$$\emptyset \xrightarrow{k_p D_n} nP$$

where $n = \left\lfloor \frac{L}{2R_O} \right\rfloor$, with L the length of dsRNA and R_O the radius of an OAS2 enzyme. The compound D_i represents i activated OAS2 attached to the same dsRNA with $i \in \{1, \cdots, n\}$. As shown in [6], the activity of OAS1, which is the monomeric equivalent of the catalytic domain of OAS2, does not depend on cofactor length. Therefore, the rate of formation of byproduct per activated enzyme is assumed to be constant and does not depend on the number of OAS2 bound to the same dsRNA. The compound D_i produces i byproducts at rate k_p.

The kinetics of OAS2 in presence of dsRNA of length L is then described by the following system of $n + 3$ equations:

$$\begin{aligned}
\frac{dE}{dt} &= -k_1 ER + k_{1m} D_1 + \sum_{i=1}^{n-1} \left[k_{(i+1)m} D_{i+1} - k_{(i+1)} D_i E \right], \\
\frac{dR}{dt} &= -k_1 ER + k_{1m} D_1, \\
\frac{dD_1}{dt} &= k_1 ER - k_{1m} D_1 + k_{2m} D_2 - k_2 D_1 E, \\
&\vdots \\
\frac{dD_j}{dt} &= k_j D_{j-1} E - k_{jm} D_j + k_{(j+1)m} D_{j+1} - k_{(j+1)} D_j E, \\
&\quad j \in \{2, 3, \cdots, n-1\}, \\
&\vdots \\
\frac{dD_n}{dt} &= k_n D_{n-1} E - k_{nm} D_n, \\
\frac{dP}{dt} &= k_p \sum_{i=1}^{n} i D_i,
\end{aligned} \tag{3.3}$$

where the additional parameters k_i with $i > 1$ are the rates of subsequent association of one OAS2 to a dsRNA already associated with $i - 1$ OAS2. The rates k_{im} with $i > 1$ are the dissociation rates of one OAS2 from a dsRNA bound with i OAS2. The initial conditions for Model-M are $E(0) = c_O$, $R(0) = c_R$, $D_i(0) = 0$ with $i \in \{1, \cdots, n\}$ and $P(0) = 0$. Note that when dsRNA used in experiments have a length L smaller or equal to 70bp, then Model-M is considered with $n = 1$ which is equivalent to Model-E.

4. Experimental data versus models

Before calibrating models to data, their behaviour is characterised. Mathematical analysis of models is carried out and presented in [10]; all models are well-posed. Furthermore, reduced models obtained by decoupling the P equation from systems (3.1), (3.2) or (3.3) are shown to have an unique stable positive steady state (results not shown; see [10]).

To calibrate models, numerical solutions of P in systems (3.1), (3.2) or (3.3) are compared to *in vitro* experimental data, published in [6], that track the concentration of PPi over time for different concentrations of dsRNA c_j $(= c_R)$ of a given length L_k $(= L)$. The experimental data are shown in Figures 2 or 5 as red errors bars.

Once each model is calibrated to experimental data, the best model amongst the three models considered is selected based on the data, goodness of the fit and model parsimony using the corrected Akaike Information Criterion (AICc) [11, 12].

4.1. Model calibration

Experimental data (in red in Figures 2 or 5) show that the formation of byproduct increases with the concentration and length of the cofactor. Although the aim of the work is to find a mechanism that can explain the formation of byproduct for all lengths and concentrations of dsRNA simultaneously, we wanted first to verify if models can accommodate changes in concentrations for a given length and if the effect of dsRNA length on the production of P can be identified through the variation of the parameter estimate values obtained when models are calibrated to each length independently. Thus, two strategies are used to estimate free parameter values listed in Table 1. First, model responses are fitted to experimental data obtained with the five concentrations of dsRNA for a given length at the five time points. The second strategy considers all five concentrations for seven lengths of dsRNA at the five time points.

In the first strategy, for the length L_k with $k \in \{1, \cdots, 7\}$, the cost function to be minimised is:

$$RSS_{L_k}(\boldsymbol{p}) = \sum_{i=1}^{m_i} \left(\sum_{j=1}^{m_j} \left(P_{c_j,L_k}(\boldsymbol{p}, t_i) - D_{t_i,c_j,L_k} \right)^2 \right), \tag{4.1}$$

where $\boldsymbol{p}$ is the free parameter vector, $P_{c_j,L_k}(\boldsymbol{p}, t_i)$ is the predicted concentration of PPi (solution for P of a given model) at time t_i for the concentration of dsRNA $c_R = c_j$ of length $L = L_k$. The experimental concentrations of PPi are D_{t_i,c_j,L_k} at time t_i for the concentration of dsRNA $c_R = c_j$ of length $L = L_k$. The parameters $m_j (= 5)$ and $m_i (= 5)$ are the numbers of dsRNA concentrations and time points considered in the study, respectively. In this first calibration strategy, parameters are estimated on 25 observation points by minimising (4.1).

In the second strategy, parameters are estimated over 175 observation points and the cost function is then:

$$RSS(\boldsymbol{p}) = \sum_{k=1}^{m_k} \left(\sum_{i=1}^{m_i} \left(\sum_{j=1}^{m_j} \left(P_{c_j,L_k}(\boldsymbol{p}, t_i) - D_{t_i,c_j,L_k} \right)^2 \right) \right), \tag{4.2}$$

where $m_k (= 7)$ is the number of dsRNA length considered in the study.

To find the estimates $\hat{p}$ for free parameter values for both strategies, the optimisation problem $RSS(\hat{p}) = \min_{p} RSS(p)$ is numerically solved using the genetic algorithm developed in the R package GA [13, 14]. To obtain confidence intervals for estimates $\hat{p}$, the genetic algorithm is run several times and let converge to obtain several plausible points in the multidimensional parameter space. The likelihood of a particular estimate $\hat{p}$ is a measure of how well the data supports that particular value of parameters. Maximising the likelihood is equivalent to minimising the negative of the log-likelihood function. Assuming independent and normally distributed additive measurement errors, the negative log-likelihood of the estimates $\hat{p}$ for a given model can be roughly approximated by $RSS_{L_k}(\hat{p})$ for the first or $RSS(\hat{p})$ for the second calibration strategy [15]. Hence, the likelihood of different estimates obtained with the heuristic approach is assessed using RSS values.

4.2. Model selection

For each calibration strategy (one length or all lengths), the AICc of model $i \in$ {Model-E, Model-N, Model-M} is computed as follows:

$$AICc_i = M \ln \left(\frac{RSS^i(\hat{p}_{Best})}{M} \right) + \frac{2K_i M}{M - K_i - 1}, \tag{4.3}$$

where K_i is the number of estimated parameters (number of free parameters in mathematical model i + 1), M is the number of observations. For the one length strategy, M is equal to 25 whereas, for the all lengths strategy, M is 175. The constant $RSS^i(\hat{p}_{Best})$ is the smallest value of the errors obtained in multiple runs of the genetic algorithm in calibrating model i for a given strategy. For each calibration strategy, the best model is the one with the smallest AICc.

To help the interpretation of results, the Akaike weights are then computed [15, 16]. The Akaike weight w_i of model i is

$$w_i = \frac{\exp \left(\frac{-\Delta_i}{2} \right)}{\sum_{q=1}^{Q} \exp \left(\frac{-\Delta_q}{2} \right)}, \tag{4.4}$$

where $\Delta_i = AICc_i - \min_{i} AICc_i$ and $Q(= 3)$ is the number of models considered in the study. The Akaike weights can be interpreted as the probability of the models to be the most likely given the experimental data and set of models considered. A model can be considered as the single best model if its weight w_i satisfies $w_i > 0.95$.

5. Results

Fitting results obtained with the first and second strategy are shown in Figures 2 and 5, respectively. Parameter estimates, which are obtained with the first strategy, are shown in multidimensional parameter spaces with information related to their likelihood in Figure 3 and distributions and confidence intervals (95% of values obtained by calibrating models) of parameter values are shown in Figure 4. Details on parameter estimates obtained with the second calibration strategy are given in Figure 6. Results of model selection for both strategies are given in Table 2.
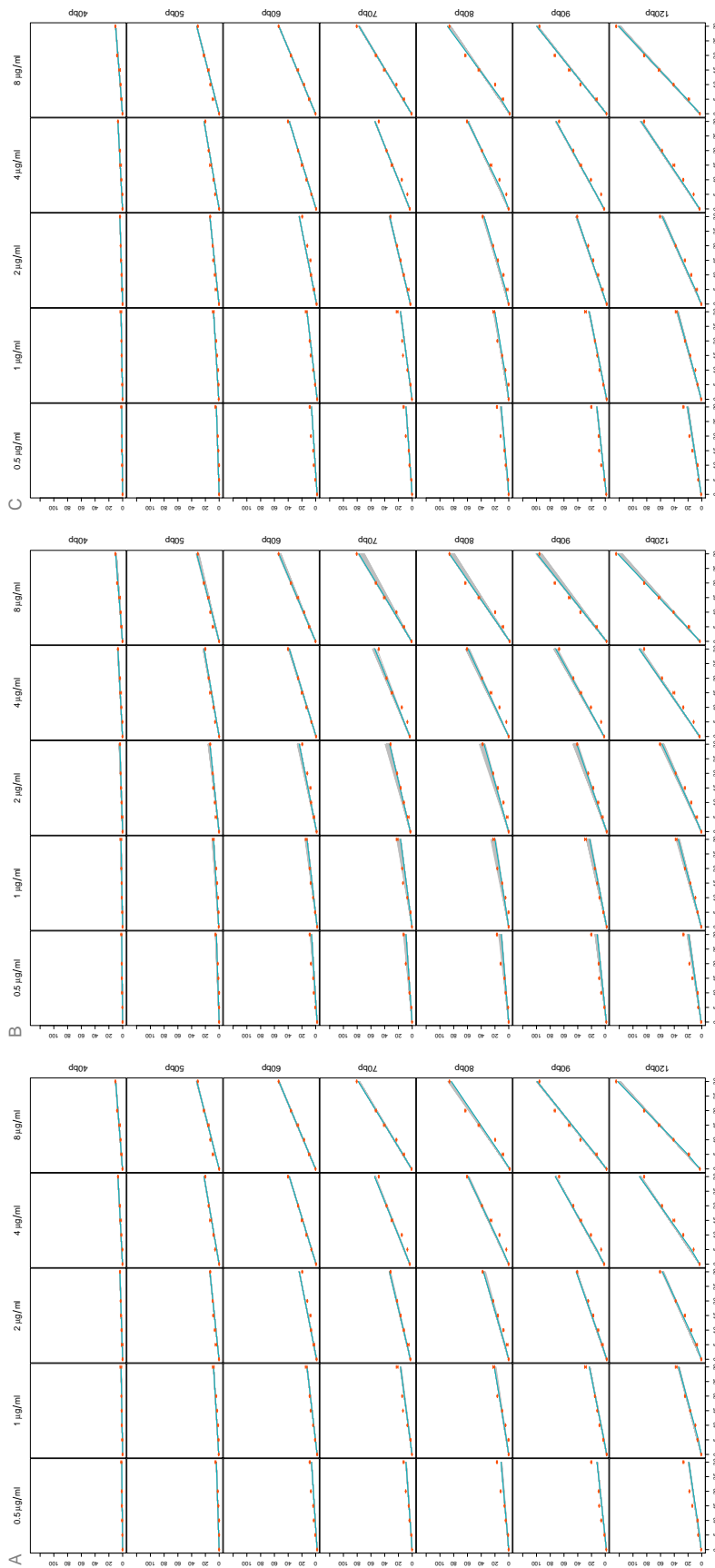
Figure 2. Fitting results obtained with the first strategy (all concentrations for one length). (A) Model-E. (B) Model-N. (C) Model-M. In each panel, each subfigure shows the PPi concentrations (in $\mu g/mL$) over time (in minutes) for the five concentrations of dsRNA arranged in columns and the seven lengths arranged in rows. Red error bars and dots are the experimental data, the concentrations of PPi at the six time points. Grey curves represent all model solutions using parameter value estimates obtained with the global optimisation method. Blue curves are the best fits with errors $RSS_{L_k}(\hat{p}_{Best})$.

5.1. First calibration strategy

When each length of dsRNA is considered independently, all models well represent the formation of PPi over time for all concentrations of dsRNA of a given length (Figure 2). Models accommodate changes in the cofactor concentration.

For the three models, values of the rate of production of PPi, k_p, are getting larger as the dsRNA length increases (Figures 3 and 4). When only one OAS2 is allowed on a single dsRNA molecule, there exists a linear relationship between the dissociation constant of enzyme and dsRNA k_{1m}/k_1 and the rate k_p (Figure 3). Interestingly, for the shortest length 40bp, values for k_{1m}/k_1 are larger than k_p values whereas this trend is reversed for lengths longer than 60bp. For short dsRNA, the binding of enzyme and cofactor is the limiting mechanism whereas for the longer dsRNA the limiting factor is the formation of product.

Model-E appears to be the most identifiable. Intervals of parameter estimates obtained for Model-E are quite narrow and fitting errors do not vary much (Figures 3 and 4A). The genetic algorithm always converges to similar parameter values with similar errors when calibrating Model-E for individual length data.

On the contrary, intervals for parameter estimates obtained for Model-N are more disperse and errors span over wider ranges (Figures 3 and Figure 4B). For Model-N, similar parameter value distributions (Figure 4B) with more clustered estimates in parameter spaces (Figure 3) are obtained as the length gets larger than 70bp. Interestingly, for long lengths (more than 80bp), the formation of non-productive compounds is mostly found less favourable than the formation of productive compounds ($k_m/k > k_{1m}/k_1$) (Figure 3). For long lengths (more than 80bp), unproductive states are short-lived in comparison to productive states.

For Model-M, when dsRNA are long enough to allow the binding of multiple OAS2 to the same dsRNA (from 80 to 120bp), we can observe that the dissociation constants k_{im}/k_i for the subsequent bindings of OAS2 are smaller than those of prior ones, $k_{1m}/k_1 > k_{2m}/k_2 > k_{3m}/k_3$ (Figures 3 and 4C). This result indicates that subsequent bindings are more favourable than the prior ones, which suggests a cooperative dynamics for OAS2 activation.

Keeping in mind that for lengths smaller or equal to 70bp Model-M and Model-E are equivalent, Model-E is found to be the best model using AICc when dsRNA are shorter than or equal to 70bp (Table 2). For 80 and 90bp long dsRNA, the confidence set of models is Model-E and Model-M. Finally, for the 120bp length, the simplest model Model-E is selected as the unique best model. However, note that if only the goodness of fit ($RSS(\hat{\boldsymbol{p}}_{\boldsymbol{Best}}) = RSS_{L_k}(\hat{\boldsymbol{p}}_{\boldsymbol{Best}})$ in Table 2) is used as a criterion the most complex model Model-M is the best model for dsRNA longer than 70bp, whereas the three models are indistinguishable for dsRNA shorter or equal to 70bp.
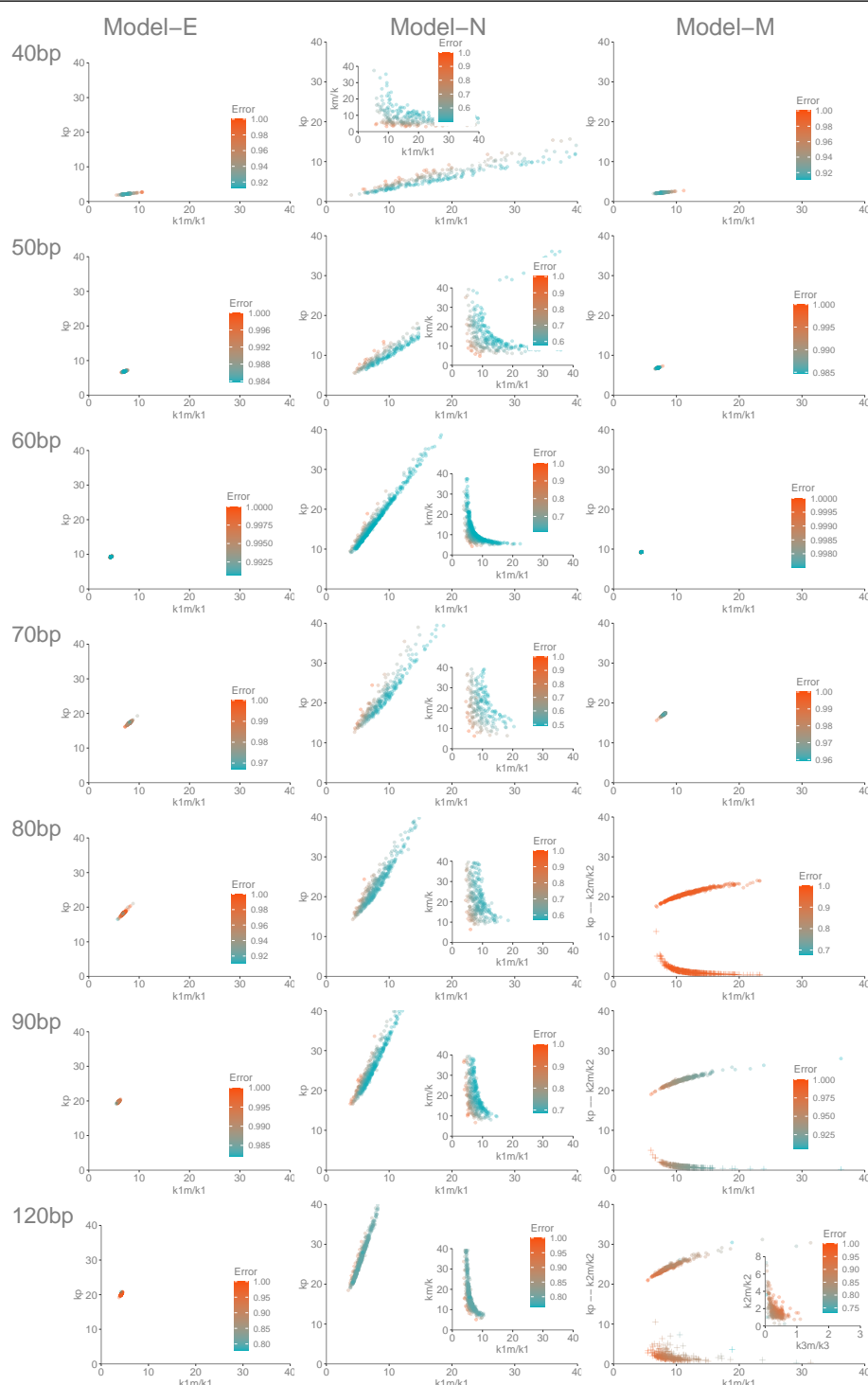
Figure 3. Parameter estimates with first strategy. Each dot represents one estimate $\hat{\boldsymbol{p}}$ obtained by calibrating models with the heuristic optimisation method minimising (4.1). For Model-M and lengths from 80 to 120 bp, in the same graph, dots represent estimates in the $k_{1m}/k_1 - k_p$ plane and + in the $k_{1m}/k_1 - k_{2m}/k_2$ plane. Colours are defined as $RSS_{L_k}(\hat{\boldsymbol{p}})/\max(RSS_{L_k}(\hat{\boldsymbol{p}}))$. Blue codes for the smallest "relative error" (the most likely parameter values given the considered data) and red codes for largest values of relative errors. Units for k_p are min^{-1} and for dissociation constants $k_{\bullet m}/k_{\bullet}$ μg/mL.
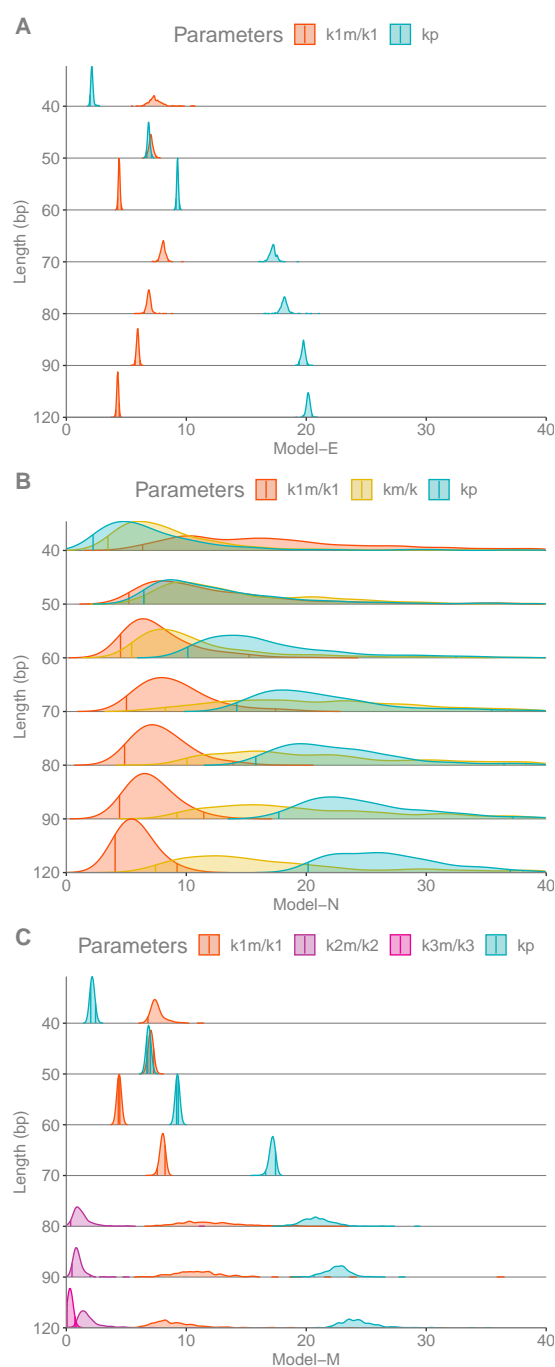
Figure 4. Distributions of parameter values obtained with the first strategy of model calibration when each length is considered independently: (A) Model-E, (B) Model-N and (C) Model-M. Vertical lines in distributions indicate 95% confidence intervals for the given parameter. Units for k_p are min^{-1} and for dissociation constants $k_{\bullet m}/k_{\bullet}$ $\mu\text{g/mL}$.

Table 2. Model selection results. $RSS(\hat{p}_{Best})$ are the smallest errors obtained among the several runs of the genetic algorithm to minimise (4.1) for the one length calibration strategy or (4.2) for the all lengths calibration strategy. K is the number of estimated parameters used in the computation of AICc defined in (4.3) (number of parameter in the mathematical model + 1). AICc weights for each model and calibration strategy are computed using (4.4). Bold indicates the best models.

dsRNA length		40bp	50bp	60bp	70bp	80bp	90bp	120bp	All
RSS($\hat{p}_{Best}$)	**Model-E**	6.133	55.40	86.56	198.3	330.5	372.0	276.8	30753.3
	Model-N	6.113	55.27	86.56	200.6	352.1	375.4	341.0	30779.6
	Model-M	6.131	55.40	86.56	199.9	247.6	341.6	234.7	9685.4
K	**Model-E**	4	4	4	4	4	4	4	4
	Model-N	6	6	6	6	6	6	6	6
	Model-M	4	4	4	4	6	6	8	8
AICc	**Model-E**	25.13	29.89	41.05	61.77	74.54	77.50	70.11	912.80
	Model-N	18.54	36.50	47.71	68.73	82.79	84.39	81.99	917.22
	Model-M	25.14	29.89	41.05	61.98	73.99	82.03	80.98	719.25
AICc weight	**Model-E**	**0.49**	**0.49**	**0.49**	**0.52**	**0.43**	**0.88**	**0.993**	9×10^{-43}
	Model-N	0.02	0.02	0.02	0.01	0.01	0.03	0.003	1×10^{-43}
	Model-M	**0.49**	**0.49**	**0.49**	**0.47**	**0.56**	**0.09**	0.004	**1**

## 5.2. Second calibration strategy

When all dsRNA lengths are considered simultaneously, Model-E and Model-N well represent the experimentally observed formation of PPi over time for all concentrations of dsRNA of intermediate lengths 60-80bp, whereas they fail to accurately represent byproduct formation for shorter and longer lengths (Figure 5A-B). As with the first calibration strategy, estimates for shared dissociation constants and k_p are found to be of the same order of magnitude for Model-E and Model-N (Figure 6). Despite the additional experimental data considered with the second calibration strategy, Model-N still appears to be highly non-identifiable.

On the other hand, Model-M gives a good representation of all dsRNA concentrations regardless of dsRNA length (Figure 5C). Despite requiring the largest number of parameters, Model-M is also selected as the unique best model when compared to all data (Table 2). Furthermore, note that the use of additional data in the second strategy improves the identifiability of Model-M: when considering all experimental data, narrower confidence intervals for estimates are obtained and estimate values are more clustered in multidimensional parameter spaces (Figure 6).

Finally, for the best model, Model-M, all estimates found for k_{1m}/k_1 and k_p are linked by the same linear relationship with mostly $k_p < k_{1m}/k_1$. Furthermore, there exists the same linear relationship between all estimates found for k_{1m}/k_1 and k_{2m}/k_2 ensuring $k_{1m}/k_1 > k_{2m}/k_2$ (Figure 6). Estimates for dissociation constants of the activation by the same dsRNA of a second (k_{2m}/k_2) and a third (k_{3m}/k_3) OAS2 have similar values. The dissociation constant of the first enzyme (k_{1m}/k_1) is more than ten times larger than the second (k_{2m}/k_2) and third (k_{3m}/k_3). The second and third enzymes have higher affinity (which is the inverse of the dissociation constant) than the first enzyme. This suggests that subsequent bindings of OAS2 are facilitated by the first binding and that there is cooperative binding of OAS2 to a single dsRNA molecule.
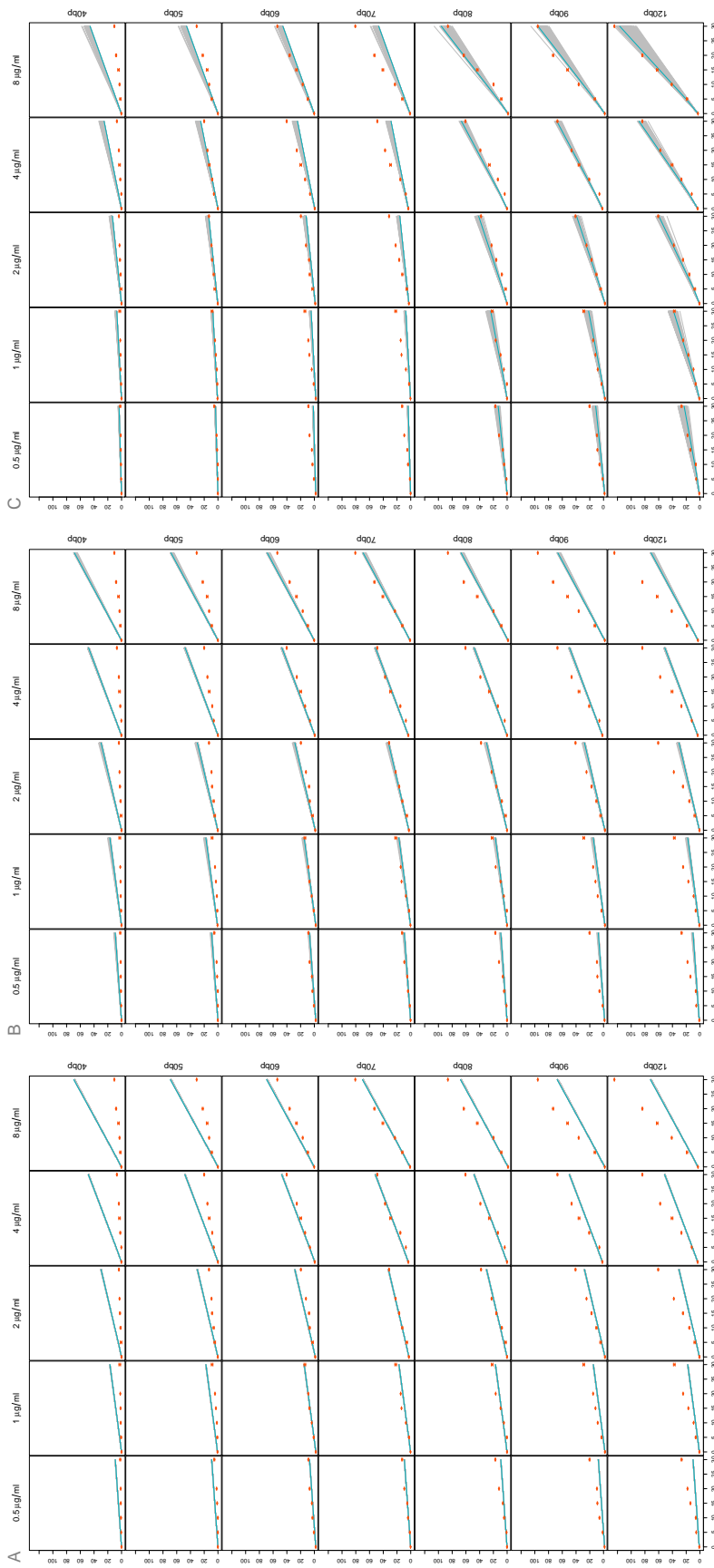
Figure 5. Fitting results obtained with the second strategy (all concentrations for all length). (A) Model-E. (B) Model-N. (C) Model-M. In each panel, each subfigure shows the PPi concentrations (in μg/mL) over time (in minutes) for the five concentrations of dsRNA arranged in columns and the seven lengths arranged in rows. Red error bars and dots are the experimental data, the concentrations of PPi at the six time points. Grey curves represent all model solutions using parameter value estimates obtained with the global optimisation method. Blue curves are the best fits with errors $RSS(\hat{p}_{Best})$.
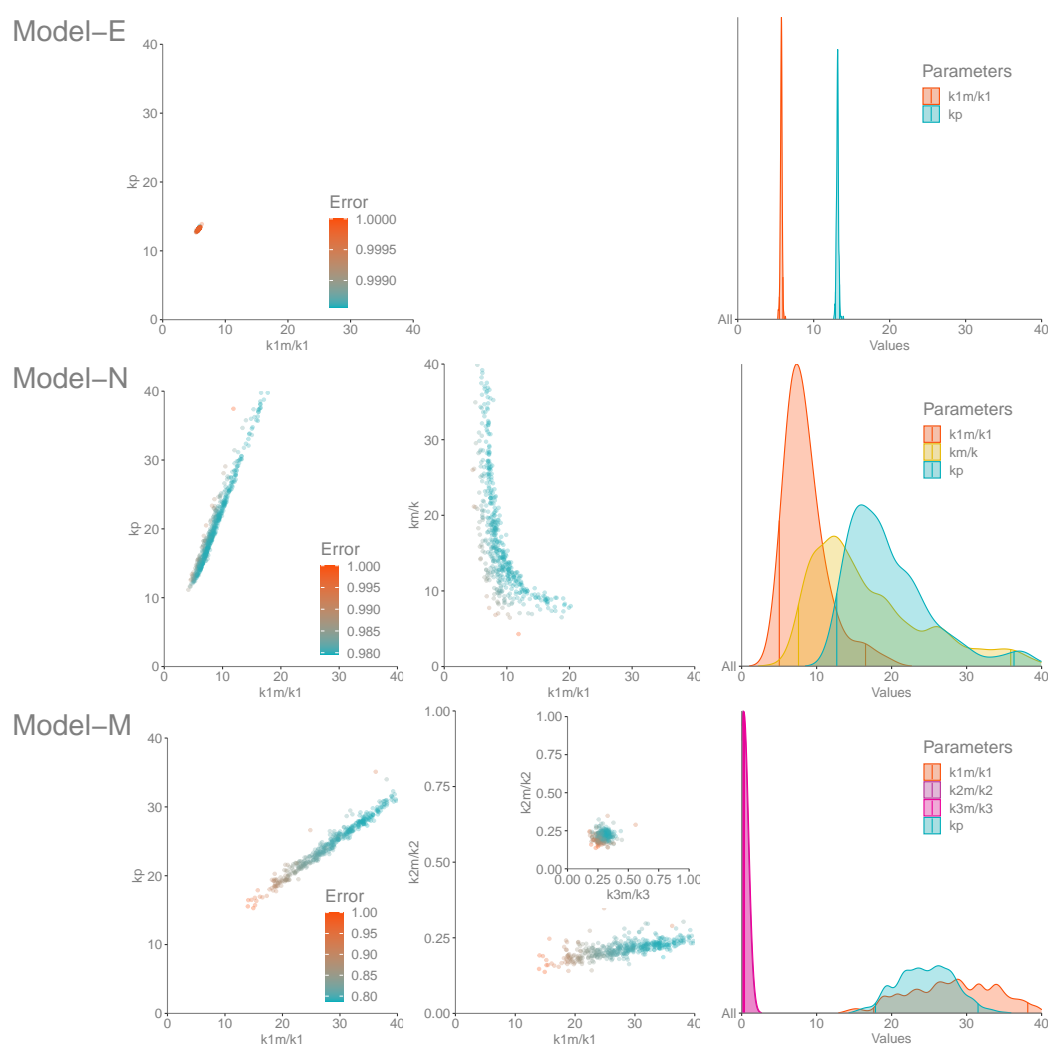
Figure 6. Results of model calibration to all concentrations and lengths data: (first row) Model-E, (second row) Model-N and (third row) Model-M. (First and second columns) Parameter estimates in multidimensional parameter spaces: each dot represents one of estimates $\hat{\boldsymbol{p}}$ obtained by minimising (4.2). Colors, which are defined as $RSS(\hat{\boldsymbol{p}})/\max(RSS(\hat{\boldsymbol{p}}))$, are used to assess estimates likelihood. Blue codes for the smallest "relative error" (the most likely parameter values given the considered data) and red codes for largest values of relative errors. (Last column) Distributions of parameter values with 95% confidence intervals delimited by vertical lines. Units for k_p are min^{-1} and for dissociation constants $k_{\bullet m}/k_{\bullet}$ μg/mL.

6. Conclusions

In this work, using mathematical modelling we investigate the effect of dsRNA length on the activation of OAS2; dsRNA serve as cofactors to enable the enzymatic activity of OAS2. Based on experimental data, our working hypothesis is that the activity of OAS2 increases as its cofactor length increases. Three models are designed: a simple enzymatic activity (Model-E), enzymatic activity with reversible inhibition (Model-N) and activation of multiple OAS2 by a single dsRNA (Model-M). In

Model-E, the cofactor length is not explicitly accounted for, whereas Model-N considers the possibility of the formation of a reversible unproductive RNA-protein interaction that results from the binding of only one domain of OAS2 to dsRNA due to its length. Model-M more explicitly incorporates the cofactor length by allowing a single dsRNA of sufficient length (more than 70bp) to bind to and activate multiple OAS2 enzymes. Moreover, two strategies are adopted to calibrate models. First, dsRNA lengths data are considered independently that yields an implicit length-dependency of the model parameters and gives us another tool to characterise the effect of the cofactor length on the different parts of the kinetics. In the second calibration strategy, all dsRNA concentrations and lengths data are considered simultaneously as the aim of our study is to propose a scenario that accommodates changes in both cofactor concentrations and lengths.

The power/strength of our study lies on the consideration of multiple models tested against experimental data with various strategies and then compared via model selection; see also [17, 18]. When considering kinetics for a single dsRNA length, the simplest enzymatic model Model-E is selected as the best model; however, when considering the kinetics for all dsRNA lengths simultaneously, the multiple binding model Model-M is the best model. Interestingly, a closer look at the estimates of parameters found for Model-M suggests that the multiple binding of OAS2 to the same dsRNA follows a cooperative regime (affinity of the activation of subsequent OAS2 enzymes is much higher than this of the first activation). Model-N, with transient inhibition, is never found to best explain the data even though it gives a good representation when lengths are considered independently. Note that a more complex version of Model-N with transitions from the non-productive compound N to productive compound D was also considered. As it provided similar results than Model-N but was even more penalized by the extra parameters, we did not present it in the work. However, the consideration of the formation of unproductive stages and its poor performance allow us to conclude that this scenario is not a mechanism that drives OAS2 activation by dsRNA.

Hence, based on our analysis, we can propose a mechanism for OAS2 activation that accounts for the impact of the length of its cofactor: multiple OAS2 can bind to the same dsRNA, when its length permits, and subsequent bindings are facilitated which underlies an enzyme with a cofactor cooperativity. This hypothesis now has to be experimentally verified, but this seems to be in preliminary agreement with the recent observation emphasising stable self-association of OAS2 as central to its mechanism [8].

Here, only three models are presented but more models have been developed in the base for this paper [10]. To the best of our knowledge, these mathematical models are the first models developed for the activation of OAS2 by dsRNA. Finally, other modelling approaches can be considered to account for the effect of cofactor length on the enzyme activation such as the use of length-dependent binding rate for enzymes to cofactors [19, 20].

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. A. Karpala, T. Doran, B. Andrew, Immune responses to dsrna: implications for gene silencing technologies, *Immunol. Cell Biol.*, **83** (2005), 211–216.

2. C. E. Samuel, Antiviral actions of interferons, *Clin. Microbiol. Rev.*, **14** (2001), 778–809.

3. B. Dauber, G. Heins, T. Wolff, The influenza b virus nonstructural ns1 protein is essential for efficient viral growth and antagonizes beta interferon induction, *J. Virol.*, **78** (2004), 1865–1872.

4. A. G. Hovanessian, J. Justesen, The human 2'-5'oligoadenylate synthetase family: unique interferon-inducible enzymes catalyzing 2'-5' instead of 3'-5' phosphodiester bond formation, *Biochimie*, **89** (2007), 779–788.

5. J. Donovan, M. Dufner, A. Korennykh, Structural basis for cytosolic double-stranded rna surveillance by human oligoadenylate synthetase 1, *P. Natl. Acad. Sci. USA*, **110** (2013), 1652–1657.

6. A. Koul, S. Deo, E. P. Booy, G. Orriss, M. Genung, S. A. McKenna, Impact of double-stranded RNA characteristics on the activation of human 2'-5'-oligoadenylate synthetase 2 (OAS2), *Biochem. Cell Biol.*, **98** (2020), 70–82.

7. J. Donovan, G. Whitney, S. Rath, A. Korennykh, Structural mechanism of sensing long dsrna via a noncatalytic domain in human oligoadenylate synthetase 3, *P. Natl. Acad. Sci. USA*, **112** (2015), 3949–3954.

8. A. Koul, D. Gemmill, N. Lubna, M. Meier, N. Krahn, E. P. Booy, et al. Structural and hydrodynamic characterization of dimeric human oligoadenylate synthetase 2, *Biophys. J.*, **118** (2020), 2726–2740.

9. J. G. Olsen, K. Teilum, B. B. Kragelund, Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness, *Cell. Mol. Life Sci.*, **74** (2017), 3175–3183.

10. D. Lee, *Mathematical modelling of oas2 activation by dsrna and effects of dsrna lengths*, Master's thesis, University of Manitoba, Winnipeg, Manitoba, Canada, 2020.

11. C. M. Hurvich, C.-L. Tsai, Regression and time series model selection in small samples, *Biometrika*, **76** (1989), 297–307.

12. N. Sugiura, Further analysis of the data by Akaike's Information Criterion and the finite corrections, *Commun. Stat.-Theor. M.*, **7** (1978), 13–26.

13. L. Scrucca, GA: A package for genetic algorithms in R, *J. Stat. Softw.*, **53** (2013), 1–37.

14. L. Scrucca, On some extensions to GA package: hybrid optimisation, parallelisation and islands evolution, *The R Journal*, **9** (2017), 187–206.

15. S. Portet, A primer on model selection using the Akaike information criterion, *Infectious Disease Modelling*, **5** (2020), 111–128.

16. K. P. Burnham, D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*, Springer, second edition, 2002.

17. S. Portet, A. Madzvamuse, A. Chung, R. E. Leube, R. Windoffer, Keratin dynamics: modeling the interplay between turnover and transport, *PLOS ONE*, **10** (2015), e0121090.

18. M. Jacquier, S. Kuriakose, A. Bhardwaj, Y. Zhang, A. Shrivastav, S. Portet, et al. Investigation of novel regulation of n-myristoyltransferase by mammalian target of rapamycin in breast cancer cells, *Scientific Reports*, **8** (2018), 1–11.

19. J. Goyette, C. S. Salas, N. Coker-Gordon, M. Bridge, S. A. Isaacson, J. Allard, et al. Biophysical assay for tethered signaling reactions reveals tether-controlled activity for the phosphatase SHP-1, *Sci. Adv.*, **3** (2017), e1601692.

20. S. Portet, Dynamics of in vitro intermediate filament length distributions, *J. Theor. Biol.*, **332** (2013), 20–29.