Modeling RNA interference in mammalian cells

Giulia Cuccato1†, Athanasios Polynikis2†, Velia Siciliano1, Mafalda Graziano1, Mario di Bernardo2,3, Diego di Bernardo1,3*

Abstract

Background: RNA interference (RNAi) is a regulatory cellular process that controls post-transcriptional gene silencing. During RNAi, double-stranded RNA (dsRNA) induces sequence-specific degradation of homologous mRNA via the generation of smaller dsRNA oligomers of length between 21-23nt (siRNAs). siRNAs are then loaded onto the RNA-induced Silencing multiprotein Complex (RISC), which uses the siRNA antisense strand to specifically recognize mRNA species which exhibit a complementary sequence. Once the siRNA loaded-RISC binds the target mRNA, the mRNA is cleaved and degraded, and the siRNA loaded-RISC can degrade additional mRNA molecules. Despite the widespread use of siRNAs for gene silencing, and the importance of dosage for its efficiency and to avoid off target effects, none of the numerous mathematical models proposed in literature was validated to quantitatively capture the effects of RNAi on the target mRNA degradation for different concentrations of siRNAs. Here, we address this pressing open problem performing in vitro experiments of RNAi in mammalian cells and testing and comparing different mathematical models fitting experimental data to in silico generated data. We performed in vitro experiments in human and hamster cell lines constitutively expressing respectively EGFP protein or tTA protein, measuring both mRNA levels, by quantitative Real-Time PCR, and protein levels, by FACS analysis, for a large range of concentrations of siRNA oligomers.

Results: We tested and validated four different mathematical models of RNA interference by quantitatively fitting models’ parameters to best capture the in vitro experimental data. We show that a simple Hill kinetic model is the most efficient way to model RNA interference. Our experimental and modeling findings clearly show that the RNAi-mediated degradation of mRNA is subject to saturation effects.

Conclusions: Our model has a simple mathematical form, amenable to analytical investigations and a small set of parameters with an intuitive physical meaning, that makes it a unique and reliable mathematical tool. The findings here presented will be a useful instrument for better understanding RNAi biology and as modelling tool in Systems and Synthetic Biology.

Background

RNA interference (RNAi) is a well characterized regulatory mechanism in eukaryotes [1-3] as well as a powerful tool for understanding gene function, thanks to the discovery that synthetic small interfering RNA oligomers (siRNAs) can efficiently induce RNAi in mammalian cells [4,5]. RNAi has also been used extensively as a novel “biological part” to design synthetic biological circuits in synthetic biology [6,7]. Artificial gene silencing has the potential to become a major genetic-기술ased therapeutic tool for viral infections [8,9], cancer [10] or inherited genetic disorders [11-13].

Despite its widespread experimental application, the best way to quantitatively model RNA interference is still under debate. In systems and synthetic biology, mathematical models are essential to carry out in silico investigations of biological pathways, or novel synthetic circuits. The aim of this work is to find the most appropriate quantitative mathematical model that can correctly describe the RNAi phenomenon in mammalian cells, for varying concentrations of the siRNA oligomers.

A schematic representation of the RNA interference mechanism is illustrated in Figure 1. In step 1, the presence of double stranded RNA (dsRNA) elicits a response in the cell mediated by the Dicer enzyme,
which binds and cleaves the dsRNA into fragments of 21-23 base pairs, called small interfering RNA (siRNA). In step 2, siRNAs are loaded onto a multiprotein complex called RNA Induced Silencing Complex (RISC) and one strand (the passenger strand) is discarded and degraded [14], while the guide strand remains within RISC and serves as a template in the silencing reaction. In step 3, the guide strand assembles into a functional siRNA-RISC complex, which contains the siRNA bound to the Ago protein [1]. Target mRNAs are then recognized by Watson-Crick base pairing [1] and bound by the siRNA-RISC complex. Finally, in step 4, mRNA degradation is induced, target mRNA is dissociated from the siRNA, and the
siRNA-RISC complex is released to process further mRNA targets [15]. Here, we will focus on quantitatively modeling step 2 to step 4 of the RNA interference process using, as an experimental tool, synthetic siRNA oligomers.

Results and Discussion

In order to model the effects of RNA interference on mRNA expression levels at different concentration of siRNA oligomers, we carried out in-vivo experiments of RNA interference on two mammalian cell-lines stably expressing the EGFP protein or the tTA protein, respectively.

In the first set of experiments (set I), Human Embryonic Kidney cells stably expressing EGFP (HEK293-EGFP cell-celline), were transfected with varying quantities of synthetic siRNA oligomers directed against the EGFP mRNA in the range of 0 to 200 pmol. Figure 2 shows the ratio of EGFP mRNA levels between treated cells and negative control cells (i.e. transfected with a non-nonpecific siRNA oligomer) measured 48 hours after transfection. Error-bars represent the standard-error from three biological replicates for each point. Similarly, in the second set of experiments (set II), EGFP protein levels were measured via FACS analysis 60 hours after transfection of HEK293-EGFP cells with synthetic siRNA oligomers directed against the EGFP mRNA, or with non-nonpecific siRNA oligomers as negative controls. The ratio between EGFP protein levels in siRNA-treated cells versus negative controls, for the same siRNA quantities as in the set I experiments, is reported in Figure 3. In the third set of experiments (set III), we tested Chinese Hamster Ovary cells (CHO AA8) stably expressing the tetracycline-tetracyclineegulated transactivator tTA at a low level, using the same protocol used previously for HEK cells. Cells were transfected with varying quantities of synthetic siRNA oligomers against the tTA mRNA in the range of 0 to 200 pmol. Since the tTA is not fluorescent, no FACS measurement were performed in this cell-celline. In Additional file 1 Figure 1 we shows the ratio of tTA mRNA levels between cells treated with the silencing tTA oligomer and cells treated with the negative control (non-targeting shuffled siRNA).

RNAi Modeling

We were interested in formulating a model that can quantitatively describe the effects of varying quantities of siRNA oligomers onto the degradation of the target mRNA species, and of its corresponding protein product. A general dynamical model describing transcription of the mRNA species, its siRNA-mediated degradation, and translation of its protein products, can be described by a system of ordinary differential equations (ODEs). Let \( X_{m}, X_{p} \) and \( X_{s} \) be the mRNA, protein and siRNA concentrations, respectively. The evolution of their time-timeependent concentrations can be described by the following ODEs:

\[
\frac{dX_{m}}{dt} = k_{m} - d_{m}X_{m} - \delta (X_{m}, X_{s}) \\
\frac{dX_{p}}{dt} = k_{T}X_{m} - d_{p}X_{p}
\]

The parameter \( k_{m} \) represents the transcription rate from the promoter that transcribes the mRNA species targeted by the siRNA oligomer; \( d_{m} \) represents the basal degradation rate of the mRNA species. RNAi can be considered as a mechanism that enhances the degradation of the targeted mRNA, therefore the function \( \delta (X_{m}, X_{s}) \) is an extra degradation term that represents the rate at which mRNAs are degraded due to RNAi. This
function, \( \delta(X_m, X_s) \), depends on both the mRNA and siRNA levels, \( X_m \) and \( X_s \) respectively. The parameter \( k_T \) is the protein translation rate, whereas \( d_p \) represents the basal protein degradation rate. At least four different models have been proposed in the literature, for the RNA interference mechanism [15-17]. Each of these models is based on the general approach described by equations (1) but each has a different functional form for \( \delta(X_m, X_s) \). Table 1, lists all the different models studied here with a description of their corresponding parameters.

**Model 1: The stoichiometric model**

One possible way of modeling the effects of RNAi on the mRNA degradation is to consider a stoichiometric reaction between the siRNAs and mRNAs. Let \( \text{siRNA}^+ \) denote the concentration of the siRNA-RISC complex, namely the fraction of the siRNAs that are loaded into RISC complexes (step 2 of Figure 1). Then the stoichiometric reaction between the siRNA* and the mRNA (step 3 and 4 of Figure 1) can be described as follows:

\[
[\text{siRNA}^+] + [\text{mRNA}] \leftrightarrow [\text{mRNA} - \text{siRNA}^+] \rightarrow \emptyset. \quad (2)
\]

Namely, the siRNA-loaded RISC binds to the complementary mRNA and then both are degraded. According to this model, following the law of mass action, we predict that the siRNA mediated degradation will be proportional to the product of the concentration of siRNA oligomers and the targeted mRNA species:

\[
\delta(X_m, X_s) = k_1 X_m X_s \quad (3)
\]

where parameter \( k_1 \) represents the proportionality constant. In this modeling approach, the siRNA-RISC complex is assumed not to be recycled, but the RISC needs to be reloaded before it can degrade another mRNA molecule (i.e. in this model the dashed line linking step 4 to step 3 in Figure 1 is not taken into account). Indeed, this model was suggested for RNA interference in prokaryotes [16,18-20]. RNAi in prokaryotes has important differences with RNAi in eukaryotes. One of these is that the interaction between the siRNA and its mRNA target is non-nontalytic in nature [16,21]. This is not the case in mammalian cells. Once the mRNA is cleaved, the siRNA-RISC complex is dissociated from it, in an ATP-dependent manner [22] and it is free to process further targets [1,14,23]. Additionally, in some organisms, such as C.elegans, the primary dsRNA trigger induces synthesis of secondary siRNAs (if the target mRNA is present) through the action of RdRP enzymes, strengthening and perpetuating the silencing response [1,24].

**Model 2: Stoichiometric model with co-operativity**

This model is a straightforward extension of Model 1, which additionally takes into account the presence of multiple sites on the targeted mRNA where the siRNA-loaded RISC can bind. Model 1 can be easily extended to include co-operativity:

\[
[\text{mRNA}] + [\text{siRNA}^+] + ... + [\text{siRNA}^+] \leftrightarrow \text{mRNA} - h_2 \text{siRNA}^+ \rightarrow \emptyset \quad (4)
\]

As before, the rate of RNAi-driven degradation, can be easily obtained applying the law of mass action:

\[
\delta(X_s, X_m) = k_2 X_s^h_2 X_m \quad (5)
\]

where \( k_2 \) is the proportionality constant and \( h_2 \) is the number of siRNA target sites. This model was suggested in [17] for modeling RNAi by miRNAs, since it is experimentally proven that multiple sites for the same miRNA can boost target mRNA repression [25]. This model however suffers from the same limitations as Model 1.

**Table 1 The different models RNA interference models for the RNAi-induced mRNA degradation rate \( \delta(X_s, X_m) \) and their corresponding parameters**

| Model | \( \delta(X_m, X_s) \) | \( k_1 \): Rate of mRNA-siRNA* complex formation |
|-------|--------------------------|-----------------------------------------------|
| Model 1 | \( k_1 X_m X_s \) | |
| Model 2 | \( k_2 X_s^h_2 X_m \) | \( k_2 \): Rate of mRNA-siRNA* complex formation, \( h_2 \): Number of siRNA target sites |
| Model 3 | \( \frac{c_1 k_1 h_2 X_s}{c_3 + k_3 h_3 X_m} X_m \) | \( k_3 \): Rate of mRNA-siRNA* complex formation, \( c_3 \): Cleavage and dissociation rate of mRNA-siRNA*, \( h_3 \): Number of siRNA target sites |
| Model 4 | \( \frac{d_4 X_s^h_4}{\theta_4^h_4 + X_s^h_5} X_m \) | \( d_4 \): Maximal degradation rate of the mRNA due to RNAi, \( \theta_4 \): Michaelis-Menten like constant, \( h_4 \): Number of siRNA target sites |
Model 3: Enzymatic model
A detailed model of RNAi specific for mammalian cells was proposed by Malphettes et al. [15]. This model accounts for the catalytic nature of RNAi in mammalian cells, thus modeling step 2, step 3 and step 4 of Figure 1, including the recycling of the RISC complex (dashed line from step 4 to step 3). Specifically, it assumes that once the cleavage and degradation of the targeted mRNA is completed, the siRNA-RISC complex (siRNA*) dissociates from its mRNA target, and it is free to degrade further mRNA molecules. Additionally, this model also considers that the targeted mRNA may have a multiple number, h₃, of siRNA binding sites. The siRNA-RISC complex (siRNA*) can bind to any site on the mRNA to form an intermediate mRNA-siRNA* complex, which can either accommodate further siRNA-RISC complexes on any other free binding sites, or cleave the target mRNA and dissociate from the cleavage products. The reaction of the complex formation of the target mRNA with the siRNA-RISC complex is described as follows (for details refer to [15]):

\[
[\text{siRNA}^*] + [\text{mRNA}] \leftrightarrow [\text{mRNA} - \text{siRNA}^*].
\] (6)

The model considers the following reaction between an intermediate mRNA-siRNA* complex with another siRNA-RISC complex (for all \(i \in [1 : h₃]\)):

\[
[\text{siRNA}^*] + [\text{mRNA} - \text{siRNA}^*] \leftrightarrow [\text{mRNA} - \text{siRNA}^*].
\] (7)

The generic cleavage and degradation reaction of the mRNA by any interacting siRNA-RISC complex (\(\forall i \in [1 : h₃]\)) is represented by:

\[
[\text{siRNA}^*] + [\text{mRNA}] \rightarrow \text{cleaved mRNA} + [\text{siRNA}^*].
\] (8)

In [15], the authors show that the rate of RNAi-driven mRNA degradation, \(\delta (Xₚ, Xₗ)\), is given by:

\[
\delta (Xₚ, Xₗ) = \frac{c₃k₃h₃}{c₃ + k₃h₃Xₗ} XₚXₗ
\] (9)

where \(k₃\) is the rate of mRNA-siRNA* complex formation for a single siRNA target site (reaction 7) and \(c₃\) is the cleavage and dissociation rate of mRNA-siRNA* complex (reaction 8). This functional form, for a constant \(Xₗ\), is a classic Michaelis-Menten enzymatic reaction, as can be observed by simply defining \(Vₗ = c₃Xₗ\) and \(Kₗ = \frac{c₃}{k₃h₃}\), and rewriting Eq. (9) as:

\[
\delta (Xₚ, Xₗ) = \frac{VₗXₗ}{Kₗ + Xₗ}
\] (10)

This is in perfect agreement with the experimental finding on the enzymatic activity of both non-nonmammalian and mammalian RISC on mRNA degradation, as reported in [23,26,27], where the product of the reaction (degraded mRNA) was measured at varying the concentration of the substrate (non-degraded mRNA \(Xₗ\)) for a constant amount of the RISC enzyme.

Model 4: Phenomenological model
In [17] the authors proposed a standard Hill-kinetic model to describe the post-transcriptional effects of miRNAs on the gene expression. miRNAs are processed by the cell to produce dsRNAs which then follow the typical RNAi pathway, schematically described in Figure 1 [15,28,29]. By considering a Hill-type enzymatic model with an Hill coefficient \(h \geq 1\), the model can be extended to account, either, for multiple binding sites of the siRNA on the same target mRNA, or, for the cooperativity of protein complexes involved in RNAi [17]. Differently from the other three model presented above, this model is a phenomenological model not derived from specific biochemical reactions:

\[
\delta (Xₚ, Xₗ) = \frac{d₄Xₚ^{h₄}}{\theta₄^{h₄} + Xₚ^{h₄}} Xₗ
\] (11)

This model, despite being phenomenological has interesting properties. The kinetic parameters \(d₄\) and \(\theta₄\) depend on the efficiency of siRNA binding to its sites on the target mRNA [17]: \(d₄\) represents the maximal degradation rate of the mRNA due to RNA interference; \(\theta₄\) the concentration of siRNA oligomers needed to achieve half of the maximal degradation rate. The above equation implies that for \(Xₚ \ll \theta₄\), the increase in the RNAi mediated degradation is linear with \(Xₚ^{h₄}\) (namely,
it will be identical to Model 1 for \( h_3 = 1 \), or to Model 2 for \( h_3 = h_2 \), while it saturates at higher levels of \( X_s \), reaching the maximal degradation rate \( d_4 \), differently from Model 3.

**Parameter Identification**

The four models were fitted to the three mRNA and protein experimental datasets (I, II and III), by searching for the parameter values for which the model-generated data best fitted the experimental data, according to a squared error measure. The results of the fitting procedure for each of the models, together with the optimized values of their parameters, are given in Table 2 for EGFP and in Additional file 1 Table A1 for tTA.

The fitting results for the mRNA levels (set I) are shown in Figure 4. Model 4 gives a significant smaller error than the other three models (see Table 2), in fact, compared to Models 1 and 3, the error of Model 4 is two orders of magnitude smaller. Model 2, gives a better fitting than Models 1 and 3, however it is worse than Model 4. The optimized value for the parameter \( d_4 \) in Model 4, is \( d_4 = 0.0081 \text{ min}^{-1} \), indicating that the strength of siRNA mediated mRNA degradation is comparable to the strength of basal mRNA degradation (since its value is in the same order of magnitude as the degradation rate of the EGFP mRNA, namely \( d_m = 0.0173 \text{ min}^{-1} \) [7]).

Note also that the parameters found for Model 2 include a coefficient \( h_2 = 0.126 \), hence less than unity. Since \( h_2 \) describes the number of siRNA binding sites on the targeted mRNA, it should be greater than, or equal to, 1 in order to have a clear biological interpretation. However, if we constrain this parameter to be greater or equal to one, then the model optimizes at the value of \( h_2 = 1 \), which makes Model 2 identical to Model 1.

We have observed that in all numerical simulations, Models 1 and 3 are almost indistinguishable. The large optimized value of parameter \( c_3 \) of Model 3 (namely \( c_3/k_m = 1.33 \times 10^4 \)), the low value of parameter \( k_3 h_3 = 1.40 \times 10^{-4} \) suggest \( c_3 \gg k_3 h_3 X_m \), and hence the function \( \delta (X_s, X_m) \) of Model 3 can be approximated by \( k_3 h_3 X_m \), which is nothing else than Model 1 (notice that the parameter optimization for Model 1 gives \( k_1 = k_3 h_3 \) in Table 2). Therefore, in this parameter space,

**Table 2 Numerical fitting results of the four models for in vitro experimental data for the EGFP protein and mRNA**

| Experiment on EGFP mRNA levels | Fit Err. | Pred. Err. | Parameters |
|-------------------------------|---------|------------|------------|
| Model 1                       | 1.00    | 0.98       | \( k_1 = 1.38 \times 10^7 \text{ (pmol min)}^{-1} \), \( h_2 = 0.126 \), |
| Model 2                       | 0.13    | 0.12       | \( k_2 = 5.00 \times 10^6 \text{ (pmol}^2 \text{ min)}^{-1} \), \( c_3/k_m = 1.33 \times 10^3 \text{ a.u.} \), |
| Model 3                       | 1.00    | 0.05       | \( k_3 h_3 = 1.40 \times 10^7 \text{ (pmol min)}^{-1} \), \( d_4 = 8.1 \times 10^4 \text{ min}^{-1} \), \( h_4 = 4.47 \) |
| Model 4                       | 0.04    | 0.05       | \( h_4 = 4.47 \) |

| Experiment on EGFP protein levels | Fit Err. | Pred. Err. | Parameters |
|-----------------------------------|---------|------------|------------|
| Model 1                           | 0.97    | 0.81       | \( k_1 = 9.90 \times 10^5 \text{ (pmol min)}^{-1} \), |
| Model 2                           | 0.46    | 0.39       | \( k_2 = 1.10 \times 10^5 \text{ (pmol}^2 \text{ min)}^{-1} \), \( h_2 = 0.456 \), |
| Model 3                           | 1.00    | 0.12       | \( k_3 h_3 = 1.20 \times 10^5 \text{ (pmol min)}^{-1} \), \( c_3/k_m = 1.33 \times 10^3 \text{ a.u.} \), |
| Model 4                           | 0.21    | 0.12       | \( d_4 = 8.6 \times 10^4 \text{ min}^{-1} \), \( h_4 = 4.49 \) |

The relative value of the error (Fit Error) and the relative value of the prediction error (Pred. Err.) for each model are given, together with the corresponding optimized values of its parameters. The unit of measurements are reported for the dimensional parameters. (a.u. stands for arbitrary units of concentration).
Model 3 is almost identical to Model 1. Similar results were obtained when searching for the parameters which best fitted the measured protein levels (set II), shown in Figure 5. Model 4 is again the one with the smallest error. Model 1 and 3 are again unable to capture efficiently the experimental data. Model 2 is better than Models 1 and 3, but it is still worse than Model 4.

When we fitted the models to the third experimental dataset (set III), which was performed on a different cell-line, with both a different target mRNA and a different siRNA oligomer, Model 4 still performed better than the others with the smallest error, although Model 2 was a close match. Model 3 and 4 were again behaving very similarly and had the largest error. It should be noted that as it happened with the previous fitting results, also in this case Model 2 has a Hill coefficient smaller than unity ($h_2 = 0.126$).

Finally we observed, as shown in Additional file 1 Figure A1, that the experimental error for set III experiment was larger when compared to the EGFP experiment (set I). This was caused by the relative low expression of the $tTA$ in the CHO cell lines, when compared to the $EGFP$ expression in HEK cells, which made measurements more noisy.

**Assessing the model predictive ability**

The models described above have the same number of unknown parameters to be learned (Methods), but for Model 4, which has one extra parameter. To be sure that the improved performance of Model 4 in describing the experimental data was not due to overfitting, we computed for each model and for each experimental dataset, the prediction error, which allows to assess the generalisation performance of the models [30]. We followed a "leave-one-out" cross validation strategy, where for each model and for each dataset, the parameter identification procedure was repeated each time removing one of the experimental points and then predicting the missing point with the identified parameters. We thus could estimate a prediction error for each model in each experimental dataset. This value is reported in Table 2 and Additional file 1 Table A1. Model 4 is again the one with the smallest prediction error, once again confirming this model superior performance in describing the data.

**Conclusions**

Our findings show that the simple Hill function described by Model 4 is sufficient to quantitatively describe the effect of RNA interference, at the mRNA and protein level, in mammalian cells in vitro, for varying concentration of siRNA oligomers.

One significant feature of Model 4 is that it can predict the saturation effect of the RNAi process that we observed experimentally. We considered the possibility that this saturation could be in fact due to the inability of the cell to uptake high concentration of siRNA oligomers, however recent experiments [27], prove that uptake of siRNA oligomers in cells is linear with the concentration of siRNA oligomers transfected, at least in the concentration range we used. Additionally, Khan et al in [31], observed upregulation of mRNA targets of endogenous micro-RNA when transfecting siRNA oligomers in mammalian cells. In order to explain this effect, they suggested a saturation of the RISC complex (or other necessary small RNA processing or transport machinery).

It has been demonstrated in [23,26,27] that the enzymatic activity of RISC can be efficiently modeled in-initro as a classic Michaelis-Menten reaction, where the target mRNA is the substrate, the siRNA-loaded RISC is the active enzyme (at a constant concentration), and the product is the degraded mRNA. This is one feature that Model 4 does not capture; namely for a fixed amount of siRNA-loaded RISC (i.e. $X_s$), Eq.(11) should approximate the Michaelis-Menten in Eq.(10), instead Model 4 becomes simply proportional to $X_{m}$, as Model 1 and 2. Nevertheless, Model 4 approximates very well the experimental data. We believe this happens because
enzymatic reactions have a typical $K_M$ much greater that the physiological concentration of their substrate ($X_m$) [23], and the same happens for the RISC complex [23,26]. In this condition, the Michaelis-Menten equation becomes $V_m \frac{X_m}{K_M + X_m} \approx \frac{V_{M4}}{K_M} X_m$ and therefore it is linear in $X_m$, as predicted by Model 4. When the siRNA concentration varies, Model 4 predicts that the parameters $\frac{V_m}{K_M}$ will change as a function of $X$, as described by Eq.(11).

The three parameters of Model 4 have a straightforward biological interpretation, and their values can be easily tuned to accommodate for different efficiencies of RNAi. For example, the parameter $d_4$ can be used to weigh the degradation due to the RNAi compared to the endogeneous mRNA degradation, and its strength, i.e. what is the maximal degradation rate that can be achieved. $\theta_4$ quantifies the siRNA oligomers concentration needed to achieve half of the maximal degradation of the targeted mRNA. The $h_4$ coefficient can accommodate for multiple target sites on the same mRNA, or for the cooperativity of the RISC complex.

Clearly, the RNAi process is very complex and no one-on-one-one relationship can be found between parameters of Model 4 and RNAi biological components. Nevertheless, it has been shown in [27] that between $10^4$ and $10^5$ siRNA oligomers per cell (corresponding to a concentration in the range 10 pM-100 pM) are sufficient to reach half-halfaximal mRNA target degradation. Model 4 predicts that half-halfaximal degradation is achieved for an amount of siRNA oligomers equal to $\theta_4$. The value of this parameter when fitting mRNA levels (Table 2 and Additional file 1 Table A1) is $\theta_4 = 0.1$ pmol despite of the different cell-cellines and mRNA-siRNA pairs tested (EGFP and tTA). This value corresponds to a concentration of 50 pm in our experimental setting, hence in good agreement with the previously reported range. Altogether these observations suggest that the quantity $\theta_4$ could be cell-cellype, mRNA, and siRNA independent.

It is estimated that the concentration of active RISC in a cell is about 3 - 5 nM [23,26,32]. Taking into account that the volume of a mammalian cell is in the range $10^{-13} L - 10^{-12}L$, then we can estimate that the number of active RISC in a cell is in the range $10^3 - 10^5$. The above observations suggest that saturation begins when the number of siRNA oligomers in a cell becomes comparable to the number of RISC molecules.

We observed that the parameters of Model 4 estimated when fitting protein levels (set II experiments) are very close to the ones estimated when fitting mRNA levels (set I experiments). Namely, the optimized values of $d_4$ and $h_4$ are very similar for both experimental data. This is important since these are two independent biological experiments. This proves the mathematical robustness of Model 4. The only parameter changing between the two sets of experiment is $\theta_4$, which represents the concentration of siRNA oligomers needed to achieve half of the maximal degradation rate ($d_4$). This is reflected in Figure 2and Figure 3, where it is clear that saturation is achieved at about 1 pmol for the mRNA data (Figure 2) and at about 20 pmol for the protein data (Figure 3). This difference may be due to biological variability, or to the simplified model of protein translation dynamics we used (steady-state approximation).

We also confirmed that Model 4 is cell-celline-independent, mRNA-independent, and siRNA-independent, since it can accurately describe the RNA interference process on a different cell-celline (CHO) expressing a different mRNA (tTA), silenced by a different siRNA oligomer.

Interestingly, the difference in Model 4 parameters, when testing a different mRNA-siRNA pair (i.e. tTA versus EGFP), shows that only $d_4$ (the maximal degradation rate) and $h_4$ (the cooperativity) change significantly, suggesting that these two parameters can be used to describe changes in siRNA-mRNA silencing specific strength, whereas $\theta_4$ may be kept constant.

Recently it has been proposed that siRNA and microRNA efficacy, defined as the percentage decrease in the target mRNA level due to the silencing reaction, could be limited due to mRNA abundance [33] or to mRNA degradation rate [34].

Model 4 predicts that the percentage decrease in target mRNA level (obtained from Eq. (13) simply dividing by $k_m/d_m$) is indeed sensitive to $d_m$ (the target mRNA degradation rate), with a higher degradation rate corresponding to a weaker effect of the silencing reaction, and vice-viceersa. This result is in line with the experimental observation described in Larsson et al [34]. In addition, according to Model 4, the transcription rate $k_m$ of the target mRNA does not have any influence on the silencing reaction efficacy. The target mRNA abundance, in absence of the silencing reaction, is simply obtained from Eq. 1 as $k_m/d_m$. Smaller $d_m$ will correspond to a higher mRNA abundance (for a constant $k_m$) therefore a correlation between mRNA abundance and sensitivity to mRNA can be found [33], but this is only an indirect effect mediated by the degradation rate, at least according to our model. Our conclusion is that siRNA-mediated degradation in mammalian cells can always be best represented as an enzymatic reaction described by an Hill function, whose parameters have to be tuned to the specific siRNA-mRNA pair.
The models discussed so far consider the average behavior of a population of cells. In the case of singe-singcell experiments, these models might not be efficient enough due to their deterministic nature and will not be able to capture any stochastic effects.

Since RNA has a plethora of functional properties and plays many roles in regulating gene expression, it has been used in a number of different studies as a tool for elucidating gene functions. In fact with RNAi it is possible to selectively knock-knockedown any gene and even modulate its dosage [35]. RNA has also been used in the design of therapeutic molecules as well as metabolic reprogramming [36]. The potential uses of this versatile molecule are still very much under study, but their effectiveness depend on many variables such as, the concentration of the silencing reagent, the transfection techniques, the cell type used and the target type selection. In the present study we biologically validated for the first time a mathematical model (Model 4) that has a simple mathematical form, amenable to analytical investigations and a small set of parameters with an intuitive physical meaning that can be used both by the computational and the experimental community interested in the analysis and application of RNA interference.

Methods

RNA interference by small interfering oligonucleotides (siRNA)
The sequence of the 21-mer siRNA double-doublestranded oligomers targeting EGFP was identical to the one reported in [37]. This siRNA targets the coding sequence of the EGFP gene starting at position 237 from the ATG, on the target sequence AAGCAGCAGACTTCCTTCAAG. The siRNA HPLC purified, with sequence GCAGCACUUCUUCAAGtt (concentration 100 μM) was synthesized by Ambion. As a negative control we used, in all experiments a shuffled sequence non targeting siRNA from Dharmacon. A 21-mer siRNA oligonucleotide was designed against the coding sequence of tetracycline-tetracyclinecontrolled transactivator (tTA) using the Ambion technology platform. Custom siRNA HPLC purified with sequence GGUUUAAACAACCCGUAAACtt (concentration 100 μM) were synthesized by Ambion on the target sequence AAGGTTCACACCCCGTAAAAC starting at position 57 from the ATG in the tTA gene coding sequence.

Cell culture and transfection
HEK 293 stably expressing EGFP (kindly provided by Mara Alfieri) were maintained at 37°C in a 5% CO2-humidified incubator. HEK 293 cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM, GIBCO BRL) supplemented with 10% heat-heatnactivated fetal bovine serum (FBS, Invitrogen) and 1% antibiotic/antimycotic solution (GIBCO BRL). CHO AA8 Tet-Off Cell Line (Clontech) stably expressing the tetracycline-tetracyclinecontrolled transactivator (tTA) were maintained at 37degC in a 5% CO2-humidified incubator. CHO cells were cultured in alpha-MEM (GIBCO BRL) supplemented with 10% heat-heatnactivated fetal bovine serum (FBS, Invitrogen) and 1% antibiotic/antimycotic solution (GIBCO BRL). Cells were seeded at a density of 300.000 per well in a 6 wells multi-multiwell and transfected 1 day after seeding using Lipofectamine 2000 (Invitrogen) according to manufacturer’s instructions with siRNA (Silencer Custom siRNA, 100 μM, Ambion) in a range of quantities from 0.001 pmol to 200 pmol (total concentration).

The amounts of transfected siRNA oligomers were: 0, 0.001, 0.01, 0.05, 0.1, 0.5, 1.0, 10.0, 20.0, 40.0, 60.0, 80.0, 100.0 and 200.0 pmol in a total of 2 mL of medium (so the final concentrations of siRNA oligomers were 5 × 10⁻⁴, 5 × 10⁻³, 2.5 × 10⁻², 5 × 10⁻¹, 5.0, 10.0, 20.0, 30.0, 40.0, 50.0, and 100 nM respectively). Each experiment was performed in biological triplicates, and the resulting standard deviations were computed and reported in each graph. One day post-postransfection, the media and ligand were replaced. Transfected cells were collected 48 hours post-postransfection for RNA extraction and subsequent analysis. FACS analysis was performed 60 hours after transfection.

RNA extraction and Real-time PCR
Total RNA extraction from 35 mm culture plates was performed using the Qiagen RNeasy Kit (Qiagen) according to manufacturers instructions. Retro-transcription of 1 μg of the total RNA extracted was performed using the QuantiTect® Reverse Transcription Kit (Qiagen), according to manufacturers instructions. Quantitative real-realime PCR was performed using a LightCycler (Roche Molecular Biochemicals, Mannheim, Germany) to analyze the amplification status of EGFP and tTA. Amplification of the genes was performed from the cDNA obtained from the total RNA and using the LightCycler DNA Master SYRB Green I kit (Roche Molecular Biochemicals). Primer sequences for Human GAPDH and Chinese Hamster GAPDH (used as reference genes) were designed by Primer 3.0 http://frodo.wi.mit.edu/ (Forward primer Human GAPDH: GAAGGTGAAGGTCGGAGTGA; Reverse primer Human GAPDH: GAGATGCTGTAGGGATTTC; Forward primer Hamster GAPDH : ACCCAGAAGACTGTTGATAG; Reverse primer Hamster GAPDH: GGTACGAGGTAAGGATTTC). Primer sequences for EGFP and tTA were also designed with Primer 3.0 (Forward primer EGFP: ACGACGCGAATCTAAGACCC; Reverse primer EGFP: GCATCGACTTCAGAGAAG; Forward primer tTA: ACAGCCGAATTAGCTGCTT; Reverse primer tTA: ACCTAGCTTTCTGGGAGGTT).
relative amounts of genes were compared with the GAPDH reference genes and calculated using the Principle of Relative Quantification Analysis according to the standard formula $2^{-\Delta C_{T}}$. To confirm the specificity of the amplification signal, we considered the primer dissociation curve in each case.

**FACS analysis**

Cells from 35 mm culture plates were trypsinized, filtered and subjected to Fluorescence-Activated Cell Sorting (FACS) analysis 60 hours posttransfection in a Becton Dickinson FACS Aria.

**Models**

In the context of the specific in vitro experiments we carried out, we can make the following assumptions to derive the mathematical model:

1. Cells express the target mRNA at a constant rate $k_m$, which corresponds to the maximum transcription rate of the promoter.
2. We assume that the siRNA oligomers will be quickly loaded into the RISC and that step 2 of Figure 1 takes place in much shorter time scale than steps 3 and 4.

Therefore, the steps 2 - 4 of RNA interference mechanism as shown in Figure 1 can be described by Equations 1. The negative control experiments involved the addition of non-nonpecific siRNA oligomers, which are not complementary to the target mRNA and therefore are not able to trigger the RNA interference mechanism. Namely, $\delta (X_m, X_s) = 0$. Therefore, the equations corresponding to the negative control experiments are:

$$\text{[mRNA]: } \frac{dX_m}{dt} = k_m - d_m X_m,$$

$$\text{[protein]: } \frac{dX_p}{dt} = \frac{k_T}{d_p} X_m - d_p X_p,$$  \hspace{1cm} (12)

**Steady-state equations**

For the numerical fitting of the in vitro experiments we used the steady state equations for the mRNAs or proteins. For example, for the in vitro experiments on RNA levels, the experimental period of 48 hours before extracting the RNA is considered long enough for the mRNAs to approach their equilibrium value. In order to solve for the mRNA or protein steady state we assume that siRNA concentration remains constant through the 48 hours of the in vitro experiments. In general, the siRNA-RISC complex, is considered very stable and one can assume that the degradation of siRNA is so slow that it does not have any effect on the overall dynamics. The steady state equations for the mRNA concentrations of the four models are:

$$\text{Model 1: } \dot{X}_m = \frac{k_m}{d_m + k X_m},$$

$$\text{Model 2: } \dot{X}_m = \frac{k_m}{d_m + \sqrt{2} \gamma X_m},$$

$$\text{Model 3: } \dot{X}_m = \frac{\beta + \sqrt{\beta^2 + 4 \delta k d_m h_s}}{2 dp} X_m,$$

$$\text{Model 4: } \dot{X}_m = \frac{k_M X_m}{d_m d_s + (d_s + d_m) X_m}.$$  \hspace{1cm} (13)

The corresponding mRNA equilibrium of the negative control experiments is simply $X_m = k_m/d_m$ (for all the models since $\delta (X_m, X_s) = 0$). Therefore, when fitting the ratio of the mRNA levels between positive and negative control, we multiply equations (13) by $d_m/k_m$. For models 1, 3 and 4 this results in the cancellation of term $k_m$, making the numerical fitting independent of the strength of the promoter. However, this is not the case for Model 3 (because these terms cannot be cancelled out). Throughout the numerical optimization, the degradation rate of mRNA EGFP was fixed at the value of $d_m = 0.0173 \text{ min}^{-1}$, which corresponds to a half-life of 40 minutes, as estimated in [7]. We used the same value of $d_m$ also when fitting the experimental dataset III for the tTA mRNA since this is in the reported range for this mRNA as well [7]. In order to optimize Model 3 with the smallest possible number of parameters, we clustered its 4 different parameters ($k_3, h_3, c_3, k_m$) in order to have only two optimized quantities: $k_3 h_3$ and $c_3/k_m$. For the in vitro experiments in protein levels, we fitted numerically the protein steady-state equations. The equilibrium concentration of protein is given by:

$$X_p = \frac{k_T}{d_p} \tilde{X}_m,$$  \hspace{1cm} (14)

where $\tilde{X}_m$ is the mRNA equilibrium, which is different for each model (equations 13). Additionally, the protein steady-state of the negative control model is:

$$X_p = \frac{k_T k_m}{d_p d_m}.$$  \hspace{1cm} (15)

For the numerical fitting of the ratio of protein levels between negative and positive control, one needs to divide equation (14) by equation (15).

**Parameter fitting and Prediction Error**

For the numerical fitting of the mRNA levels from in vitro experiments, we used the following error function:
\[ \sum_{i} \left( \frac{Y_{i \text{ model}} - Y_{i \text{ data}}}{SE^i} \right)^2 \]  

where \( N \) is the number of experimental points, \( Y_{i \text{ data}} \) is the experimental measurement of experiment \( i \), \( Y_{i \text{ model}} \) is the model prediction for experiment \( i \) and \( SE^i \) is the standard error of experiment \( i \). A genetic algorithm implemented in the “Genetic Algorithm and Direct Search Toolbox” of Matlab (the \textit{ga} command) was then used to find the parameters which minimised the error function.

The absolute value errors of each model were then normalised against the largest error. Namely, the error of Model 3 (which in both case was the largest one) was set to 1 and all the other errors of the remaining three models, were normalised against error of Model 3.

The Prediction Error (PE) for each experimental dataset was computed by repeating the parameter fitting procedure described above, but this time using a leave-one-out cross-validation procedure. The PE was then computed as the average error (Eq. 16) between the predicted value and the experimental value across all the experimental points. As done for the error function, we then computed a relative value for the PE in order to compare the performance across the different models by normalising against the maximum PE across the four models, and reported it in Table 2 and in Additional file 1 Table A1.

Please observe that in the case of the \( t \)TA mRNA the cost function used was as in Eq. 16 but without dividing by the standard error \( SE \) since in this case the experimental data were more noisy and the \( SE \) could not be estimated accurately.

**Additional material**

**Additional file 1**: Supplementary material for Modeling RNA interference in mammalian cells. Results and fitting of in vitro experiments on hamster ovary cell line (CHO) constitutively expressing \( t \)TA protein. We measured mRNA levels, by quantitative Real-Time PCR for a large range of concentrations of siRNA oligomers, from 0.001 pmol to 200 pmol (total concentration). The amounts of transfected siRNA oligomers were: 0, 0.001, 0.01, 0.05, 0.1, 1.0, 10.0, 20.0, 40.0, 60.0, 80.0, 100.0 and 200.0 pmol in a total of 2 mL of medium (so the final concentrations of siRNA oligomers were: 5 \( \times \) 10\(^{-7}\), 5 \( \times \) 10\(^{-8}\), 2.5 \( \times \) 10\(^{-8}\), 2.5 \( \times \) 10\(^{-9}\), 5 \( \times \) 10\(^{-10}\), 5.0, 10.0, 20.0, 30.0, 40.0, 50.0, and 100 nM respectively). Each experiment was performed in biological triplicates, and the resulting standard deviations are computed and reported in each graph. In Additional file 1 Table A1, numerical fitting results and predicted error for the four models, in Additional file 1 Figure A1, graphic representation of the numerical fitting.

**Acknowledgements**

Thanks to Laura Pisapia for the FACS analysis, S. Giovane and F. Menolascina for technical support and to Mara Alfieri for providing the Hek-EGFP stable clone. Funding for this work has been provided by the Italian Ministry of Research Grant ITALBIONET to DdB.

**Author details**

1Telethon Institute of Genetics and Medicine (TIGEM), Naples, Italy.
2Department of Engineering Mathematics, University of Bristol, Bristol, UK.
3Department of Computer and Systems Engineering, University of Naples Federico II, Naples, Italy.

**Authors’ contributions**

GC and VS designed the experiments. GC, VS and MG performed the experiments. AP developed the models and AP and DdB conducted simulations. AP, GC and DdB drafted the manuscript. DdB and MG conceived and supervised the collaboration and overall strategy of the project and edited the manuscript. All authors have read and approved the final manuscript.

**Received:** 7 October 2010 **Accepted:** 27 January 2011

**Published:** 27 January 2011

**References**

1. Carthew R, Sontheimer E: Origins and Mechanisms of miRNAs and siRNAs. Cell 2009, 136:642-655.
2. Bartel D: MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell 2004, 116:281-297.
3. Cullen BR: RNAi the natural way. Nat Genet 2005, 37:1163-1165.
4. Brummelkamp TR, Bernards R, Agami R: A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. Science 2002, 296(5567):550-553 [http://www.sciencemag.org/cgi/content/abstract/296/5567/550].
5. Elbashir S, Harborth J, Lendecke Y, Yalcin A, Weber K, Tuschl T: Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 2001, 411:494-498.
6. Deans TL, Cantor CR, J CJ: A Tunable Genetic Switch Based on RNAi and Repressor Proteins for Regulating Gene Expression in Mammalian Cells. Cell 2007, 130:363-372.
7. Tsigges M, Marquez-Lago T, Stelling J, Fussenegger M: A tunable synthetic mammalian oscillator. Nature 2009, 457:309-312.
8. Bank S, Brik V: Prospects of RNA interference therapy in viral diseases: update 2006, Expert Opin Biol Ther 2006, 6:115-160.
9. Fulton A, Peter S, Perkins G, Jarosinski K: Effective Treatment of Respiratory Alphaherpesvirus Infection Using RNA Interference. PLoS ONE 2009, 4:e4118.
10. Takeshita F, Ochiya T: Therapeutic potential of RNA interference against cancer. Cancer Sci 2006, 97:689-96.
11. Leung R, Whitaker P: RNA interference: from gene silencing to gene-gene specific therapeutics. Pharmacol Ther 2005, 107:223-39.
12. Aagaard L, JJ R: RNAi therapeutics: Principles, prospects and challenges. Adv Drug Del Rev 2007, 59:75-86.
13. Shetye K, Suchita A, Nishanta M, Vibha R: RNA interference: Emerging diagnostics and therapeutic tools. Bioch and Bioph Res Commun 2009, 386:273-277.
14. Filipowicz W: RNAi: the nuts and bolts of the RISC machine. Cell 2005, 122:17-20.
15. Malphettes L, Fussenegger M: Impact of RNA interference on gene networks. Metab Eng 2006, 8:672-683.
16. Levine E, Zhang Z, Kuhlman T, Hwa T: Quantitative characteristics of gene regulation by small RNA. PLoS Biol 2007, 5:e229.
17. Khanin R, Vincelli P: Computational Modeling of Post-Transcriptional Gene Regulation by MicroRNAs. J Comput Biol 2008, 15:305-316.
18. Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, Margalit H: Regulation of gene expression by small non-coding RNAs: a quantitative view. Mol Syst Biol 2007, 3:138.
19. Mehta P, Goyal S, Wingreen S: A quantitative comparison of siRNA-based and protein-proteinased gene regulation. Mol Syst Biol 2008, 4:221.
20. Mitarai N, Andersson AMC, Krishna S, Semsey S, Sneppen K: Efficient degradation and expression prioritization with small RNAs. Phys Biol 2007, 4:164-171.

21. Massé E, Escorcia FE, Gottesman S: Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. Genes Dev 2003, 17:2374-2383.

22. Hutvagner G, Zamore P: A microRNA in a Multiple-Turnover RNAi Enzyme Complex. Science 2002, 297:2056-2060.

23. Haley B, Zamore P: Kinetic analysis of the RNAi enzyme complex. Nat Struct & Mol Biol 2004, 11:7.

24. Meister G, Tuschi T: Mechanisms of gene silencing by double-stranded RNA. Nature 2004, 431:338-342.

25. Rajewsky N: microRNA target predictions in animals. Nat Rev Genet 2006, 38:58-513.

26. Martinez J, Tuschi T: RISC is a 5’ phosphomonoester-phosphomonoester-producing RNA endonuclease. Genes Dev 2004, 18:979-980.

27. Overhoff M, Wünsche W, Szakiel G: Quantitative detection of siRNA and single-stranded oligonucleotides: relationship between uptake and biological activity of siRNA. Nucleic Acids Res 2004, 32:170.

28. Doench J, Petersen C, Sharp P: siRNAs can function as miRNAs. Genes Dev 2003, 17:438-442.

29. Zeng Y, Yi R, Cullen B: MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci 2003, 100:9779-9784.

30. Hastie T, Tibshirani R, Friedman J: The elements of statistical learning: data mining, inference, and prediction New York: Springer, 2009.

31. Khan AA, Betel D, Miller ML, Sander C, Leslie CS, Marks DS: Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. Nat Biotechnol 2009, 27:549-555.

32. Brown KM, Chu Cy, Rana TM: Target accessibility dictates the potency of human RISC. Nat Struct Mol Biol 2005, 12:469-470.

33. Arvey A, Larsson E, Sander C, Leslie C, Marks D: Target mRNA abundance dilutes microRNA and siRNA activity. Molecular systems biology 2010, 6.

34. Larsson E, Sander C, Marks D: mRNA turnover rate limits siRNA and microRNA efficacy. Molecular Systems Biology 2010, 6.

35. Hemann M, Fridman J, Zilfou J, Hernando E, Paddison P, Cordon-Cardo C, Hannon G, Lowe S: An epi-epileptic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes in vivo. Nature Genetics 2003, 33:396-400.

36. Chen Y, Jensen M, Smolke C: Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. Proc Natl Acad Sci 2010, 1-6.

37. Chiu Y, Rana T: RNAi in human cells: basic structural and functional features of small interfering RNA. Mol Cell 2002, 10:49-561.

doi:10.1186/1752-0509-5-19

Cite this article as: Cuccato et al.: Modeling RNA interference in mammalian cells. BMC Systems Biology 2011 5:19.