A model for the equilibrium binding of the intercalating dye SYBR Green I during DNA amplification

© D.R. Ibragimov[1], A.A. Fedorov[1], D.A. Varlamov[2]

[1] Institute of Analytical Instrument Making, Russian Academy of Sciences, St. Petersburg, Russia
[2] All-Russia Research Institute of Agricultural Biotechnology, Moscow, Russia
E-mail: danil_ib@mail.ru

Received May 5, 2025
Revised July 1, 2025
Accepted July 7, 2025

A model of the polymerase chain reaction (PCR) with a fluorescent signal generated by the intercalating dye SYBR Green I is proposed. The DNA amplification model describes the reaction kinetics occurring during the annealing and elongation stages. In the model, the fluorescence intensity is determined by the amount of bound dye, which is calculated using the dissociation constant of the equilibrium reaction between the dye and double-stranded DNA. The model demonstrated high accuracy in approximating experimental data, as well as the ability to predict the initial number of DNA molecules in the sample.

Keywords: real-time PCR model, SYBR Green I dye, Runge–Kutta methods, Nelder–Mead method.

DOI: 10.61011/TPL.2025.12.62794.8073

Polymerase chain reaction (PCR) is an indispensable molecular biology technique used for the qualitative and quantitative analysis of nucleic acids. It is based on the amplification of specific DNA regions in repeated temperature cycles. Its real-time modification (RT-PCR) allows one to determine target DNA concentrations using a plot of the fluorescent reporter signal associated with the accumulation of DNA copies. To estimate the concentration in the sample under study [1], one needs to determine the point at which the fluorescent signal reaches a given threshold level C_T and construct calibration dependence $C_T / \lg N_0$ for samples with known concentration N_0. However, the accuracy of this method depends on the difference in amplification efficiency of the reference and the sample. In actual samples, the reaction efficiency is determined by many factors: temperature variations, pipetting errors, presence of inhibitory impurities, etc. [2]. These are the reasons why a mathematical apparatus allowing one to evaluate and factor in the reaction efficiency for each sample is needed to improve the accuracy of quantitative assessment [3]. A number of mathematical models, which may be divided into approximation and kinetic ones, have been proposed as a means to solve this problem. Approximation models do not take the nature of a reaction into account and offer various mathematical functions for data approximation with the purpose of determining more accurately the reaction characteristics and DNA concentration in the sample. Four- and five-parameter sigmoid functions [4,5] are the models of this kind that are used most often. Unlike approximation models, kinetic ones characterize the internal processes of PCR with kinetic and equilibrium reaction equations; however, virtually no models providing an adequate approximation of experimental RT-PCR curves are available at present. The kinetic approach to PCR modeling has an advantage in that it characterizes actual events in a reaction mixture, which allows one to introduce various processes and factors potentially improving the accuracy of model description. The simplest model neglects the contribution of primers at the annealing stage [6], and the PCR description provided by it has significant errors [7]. More detailed models take into account primer dimer formation [8,9], the efficiency of double-stranded DNA denaturation [8], the addition of nucleotides at the elongation stage [9,10], and the temperature regime [10] and characterize the action of the DNA polymerase enzyme using the Michaelis–Menten equation [8] or a detailed description of the kinetics of its interaction with the reaction products [9,10].

Either intercalating dyes (IDs) or hybridization probes are used as fluorescent reporters for detection of PCR products. Dyes, such as SYBR Green I and EvaGreen, are inserted between nucleotide pairs, enhancing the fluorescence quantum yield by several orders of magnitude. Hybridization probes contain a dye and a fluorescence quencher in their structure. As a result of probe hybridization and subsequent enzymatic cleavage, they are separated in space, which allows for detection of the dye signal. Since the physical mechanisms of their operation differ, RT-PCR curves obtained with IDs and probes also differ. Existing kinetic models ignore this, assuming a linear relationship between fluorescence and the amount of accumulated DNA [6–9]. This simplification is unjustified, since differences in signals between the methods lead to discrepancies between theoretical and experimental data, which is confirmed by the difference in fluorescence curves of IDs and probes [11]. Thus, a method for converting the amount of DNA into fluorescence intensity with account for the specifics of the detection technique is needed to construct an adequate RT-PCR model.

In the present study, we propose an RT-PCR model that takes into account the binding of ID molecules to accumulated double-stranded DNA. The amount of bound

ID determines the intensity of the detected fluorescent signal. Reactions proceeding at the annealing stage are considered in the proposed model. The efficiency of denaturation, which involves „melting“ double-stranded DNA into complementary single-stranded DNA (ssDNA) S and S', is assumed to be 100 %. At the annealing stage, the ssDNA formed at the denaturation stage is hybridized with specific primers denoted as Pr and Pr'. The resulting matrix–primer complexes are denoted as T and T'. In addition, the ssDNA formed at the denaturation stage may expel the primers and hybridize with each other. This product of the so-called reverse hybridization reaction is designated as P_2. Hybridization is considered to be irreversible and primer dimer formation is neglected. Under these assumptions, the annealing stage is characterized by the following reactions:

$$S + Pr \xrightarrow{k_1} T, \tag{1}$$

$$S' + Pr' \xrightarrow{k_1} T', \tag{2}$$

$$S + T' \xrightarrow{k_2} P_2 + Pr', \tag{3}$$

$$S' + T \xrightarrow{k_2} P_2 + Pr, \tag{4}$$

$$T + T' \xrightarrow{k_2} P_2 + Pr + Pr', \tag{5}$$

where k_1 is the rate constant of primer-ssDNA complex formation and k_2 is the hybridization rate constant.

The completion of a complementary ssDNA fragment by DNA polymerase enzyme E, which results in the formation of a copy of the original DNA fragment (P_1), is regarded as a one-step process. This stage is characterized by the following reaction:

$$E + T \overset{k_3,k_{-3}}{\longleftrightarrow} ET \xrightarrow{k_{cat}} E + P_1, \tag{6}$$

where k_3, k_{-3} are the rate constants of forward and reverse reactions of formation of the enzyme–substrate complex and k_{cat} is the enzymatic synthesis constant.

Since reaction (6) is enzymatic, the Michaelis–Menten model of enzymatic synthesis may be used (under certain assumptions) to describe the elongation stage:

$$\frac{dP_1}{dt} = \frac{V_m C_T}{K_M + C_T}, \tag{7}$$

where $V_m = k_{cat} C_E$ is the maximum reaction rate and $K_M = (k_{-3} + k_{cat})/k_3$ is the Michaelis constant that is equal to the concentration of the substrate (ssDNA and primer complex) at which the reaction rate is two times lower than the maximum one.

The law of mass action is used is used to characterize the kinetics of reactions (1)–(5). We assume that the processes at two complementary chains are symmetrical (i. e., both chains have the same values of corresponding reaction parameters). This assumption implies that the concentrations of complementary single-stranded fragments are the same in the model system, which helps simplify the system. Taking Eq. (7) into account, we then obtain the following system of differential equations for both primed and unprimed reaction components:

$$\frac{dC_S^i}{dt} = -k_1 C_S^i C_{Pr}^i - k_2 \left(C_S^i\right)^2 - k_2 C_T^i C_S^i, \tag{8}$$

$$\frac{dC_{Pr}^i}{dt} = -k_1 C_S^i C_{Pr}^i + k_2 \left(C_T^i\right)^2 + k_2 C_T^i C_S^i, \tag{9}$$

$$\frac{dC_T^i}{dt} = k_1 C_S^i C_{Pr}^i - k_2 \left(C_T^i\right)^2 - k_2 C_T^i C_S - \frac{V_m C_T^i}{K_M + C_T^i}, \tag{10}$$

$$\frac{dC_{P_2}^i}{dt} = k_2 \left(C_S^i\right)^2 + 2k_2 C_T^i C_S^i + k_2 \left(C_T^i\right)^2, \tag{11}$$

$$\frac{dC_{P_1}^i}{dt} = \frac{V_m C_T^i}{K_M + C_T^i}, \tag{12}$$

where i is the amplification cycle number.

Solving this system at each PCR cycle, one may obtain the entire DNA amplification curve. The initial conditions at the first cycle correspond to the initial concentrations of ssDNA $C_S^0(0)$ and primers $C_{Pr}^0(0)$. At subsequent cycles, the initial concentration of ssDNA is increased by the amount of product obtained at the end of the previous cycle, ensuring intensification of the signal in the process of amplification:

$$C_S^i(0) = C_S^{i-1}(0) + C_{P_1}^{i-1}(\tau) + C_{P_2}^{i-1}(\tau), \tag{13}$$

where τ is the duration of annealing and elongation stages.

The initial primer concentration for the next cycle is equal to the amount of primer remaining at the end of the previous one:

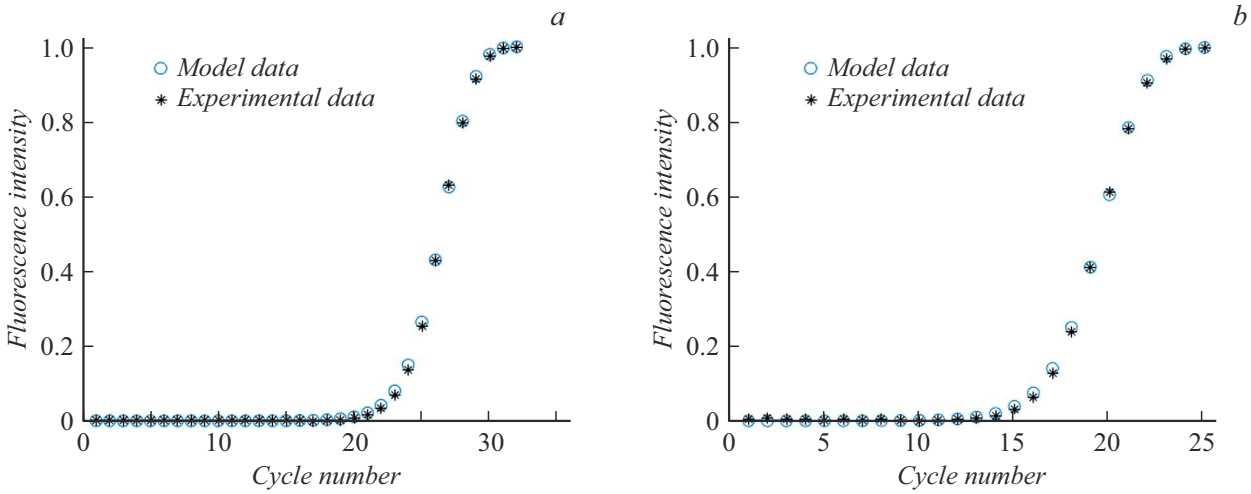
$$C_{Pr}^i(0) = C_{Pr}^{i-1}(\tau). \tag{14}$$

The initial concentrations of other components are zero for all cycles.

The process of formation of a complex of DNA and ID molecules is analyzed within the equilibrium approximation. The number of ID molecules bound to a particular DNA depends primarily on the DNA length and the dye/DNA concentration ratio, which changes in the course of PCR. We assume for simplicity that the amount of bound ID may be estimated using a certain averaged dissociation constant K_d of the equilibrium reaction that remains the same for all cycles. In this case, concentration of bound ID C_{PI} at the ith PCR cycle with a known DNA concentration and concentration of free dye C_I may be determined using the expression for the dissociation constant of the equilibrium reaction of dye and DNA binding

$$C_{PI}^i = \frac{\left(C_{P_1}^i(\tau) + C_{P_2}^i(\tau)\right) C_I}{C_{P_1}^i(\tau) + C_{P_2}^i(\tau) + K_d}. \tag{15}$$

System (8)–(12) was solved in MATLAB using the fourth- and fifth-order Runge–Kutta methods, which were found to be optimal in terms of accuracy and speed. Constants V_m and K_M were estimated based on the experimental



Theoretical and experimental RT-PCR plots. *a* — 100-fold diluted sample; *b* — undiluted sample. The plots are normalized to the maximum value.

data on kinetics of Taq polymerase: K_M and k_{cat} were estimated as $\sim 1\,\mu$M and $\sim 32\,s^{-1}$, respectively [12]. If the k_{cat} value and the enzyme concentration are known, V_m may be estimated. The values of other reaction constants (k_1, k_2, and K_d) were chosen using the Nelder–Mead method.

Experimental RT-PCR plots were obtained using synthetic DNA with a length of 97 nucleotides and a test system for its quantitative RT-PCR analysis produced by NPF Sintol. The concentration of model DNA in the test solution was determined via digital PCR with a Sniper DQ24 system and was close to 8200 DNA molecules in $1\,\mu$l (the volume taken for analysis). The forward and reverse primer sequences were CACATATTTACACAATGGCAAAGC and CTGAACACACATTATTACTCCGAA, respectively. The initial concentration of each primer was $C_{Pr}(0) = 200$ nM. The used amount of the SYBR Green I fluorescent dye was $0.07\,\mu$l of a 100x solution, which corresponds approximately to a molar concentration of $0.28\,\mu$M [13]. The amount of Taq polymerase in the reaction mixture was 5 units. The overall reaction mixture volume was $25\,\mu$l. RT-PCR curves were obtained using a Bio-Rad CFX amplifier for the original and 100-fold diluted samples in five replicates. The PCR cycle consisted of preliminary denaturation and 45 cycles of heating to 95 °C (15 s) and cooling to 62 °C (30 s). No increase in fluorescent signal values was observed for negative control samples within 50 reaction cycles, indicating a lack of nonspecific reactions due to primer dimerization.

The results of approximation of experimental RT-PCR curves by the proposed model are shown in the figure. The presented data reveal that the proposed model approximates accurately the curve of accumulation of the fluorescent intercalating dye signal at both high and low DNA concentrations in the sample.

The approximation error for each of the ten analyzed curves was estimated as a sum of squared differences between the experimental and model data and was found to be less than 0.1 %. The initial amount of DNA predicted by the model did not exceed 7 % of the average value estimated using digital PCR, which was 8200 molecules, for five 100-fold diluted samples. In the case of undiluted samples, the deviation did not exceed 15 % of the expected amount. The values of constants k_1, k_2, and K_d fitted by the model were $1.1 \cdot 10^6\,M^{-1} \cdot s^{-1}$, $1.02 \cdot 10^6\,M^{-1} \cdot s^{-1}$, and $6.6 \cdot 10^{-8}$ M, which corresponds to the range of data presented in literature [10,14].

Thus, the present study is the first to propose a working model of RT-PCR where the SYBR Green I intercalating dye is used to detect the reaction product. The model demonstrated high accuracy in approximating experimental data and the capacity to estimate the DNA concentration in the test sample, which is indicative of viability of the proposed approach to RT-PCR modeling. A number of factors exerting an influence on the reaction kinetics, such as the length of DNA, its sequence, the formation of primer dimers, polymerase inactivation, dye attenuation, denaturation efficiency, etc., were neglected in this model. The introduction of these factors into RT-PCR modeling is the subject of further research.

Funding

This study was carried out under the state assignment of the Ministry of Science and Higher Education of the Russian Federation (project FFZM-2025-0002).

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] D.V. Rebrikov, G.A. Samatov, D.Yu. Trofimov, P.A. Semenov, A.M. Savilova, I.A. Kofiadi, D.D. Abramov, *PTsR v real'nom vremeni*, Ed. by D.V. Rebrikov (Laboratoriya znanii, M., 2023), pp. 114–118 (in Russian).

[2] A. Ruiz-Villalba, J.M. Ruijter, M.J.B. van den Hoff, Life, **11** (6), 496 (2021). DOI: 10.3390/life11060496

[3] J.M. Ruijter, R.J. Barnewall, I.B. Marsh, A.N. Szentirmay, J.C. Quinn, R. van Houdt, Q.D. Gunst, M.J.B. van den Hoff, Clin. Chem., **67** (6), 829 (2021). DOI: 10.1093/clinchem/hvab052

[4] R.G. Rutledge, Nucl. Acids Res., **32** (22), e178 (2004). DOI: 10.1093/nar/gnh177

[5] A.N. Spiess, C. Feig, C. Ritz, BMC Bioinformatics, **9**, 221 (2008). DOI: 10.1186/1471-2105-9-221

[6] G.J. Boggy, P.J. Woolf, PLoS One, **5** (8), e12355 (2010). DOI: 10.1371/journal.pone.0012355

[7] G. Cobbs, BMC Bioinformatics, **13**, 203 (2012). DOI: 10.1186/1471-2105-13-203

[8] J.L. Gevertz, S.M. Dunn, C.M. Roth, Biotechnol. Bioeng., **92** (3), 346 (2005). DOI: 10.1002/bit.20617

[9] S. Mehra, W.-S. Hu, Biotechnol. Bioeng., **91** (7), 848 (2005). DOI: 10.1002/bit.20555

[10] K. Marimuthu, C. Jing, R. Chakrabarti, Biophys. J., **107** (7), 1731 (2014). DOI: 10.1016/j.bpj.2014.08.019

[11] A.A. Fedorov, D.G. Sochivko, D.A. Varlamov, Tech. Phys., **65** (9), 1516 (2020). DOI: 10.1134/S1063784220090169.

[12] W. Chen, B. Chen, X. Li, G. Xu, L. Yang, J. Wu, H. Yu, FEBS J., **291** (13), 2876 (2024). DOI: 10.1111/febs.17091

[13] F. Mao, W.Y. Leung, X. Xin, BMC Biotechnol., **7**, 76 (2007). DOI: 10.1186/1472-6750-7-76

[14] A.I. Dragan, R. Pavlovic, J.B. McGivney, J.R. Casas-Finet, E.S. Bishop, R.J. Strouse, M.A. Schenerman, C.D. Geddes, J. Fluoresc., **22** (4), 1189 (2012). DOI: 10.1007/s10895-012-1059-8

Translated by D.Safin