

## СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ КИНЕТИЧЕСКИХ КРИВЫХ ПОЛИМЕРАЗНОЙ ЦЕПНОЙ РЕАКЦИИ

© 2011 г. Д. Г. Сочивко, А. А. Фёдоров, В. В. Лавров, Д. А. Варламов,  
В. Е. Курочкин, академик Р. В. Петров

Поступило 28.04.2011 г.

Полимеразная цепная реакция (ПЦР) лежит в основе множества молекулярных методов, применяемых для качественного и количественного анализа специфических последовательностей нуклеиновых кислот. Процесс ПЦР представляет собой цепную реакцию размножения анализируемого фрагмента ДНК, который можно представить в виде дискретного ветвящегося случайного процесса с участием нескольких типов частиц. В работе [1] предложена модель такого процесса. Модель описывала циклическое размножение частиц двух типов, каждая из которых генерирует частицу другого типа с определенной вероятностью  $P_{12}$  или  $P_{21}$ . Показано, что такая модель порождает распределения вероятностей для количества копий после заданного количества циклов процесса, зависящие от значений вероятности генерации. В предложенной модели вероятности принимались за постоянные величины, в результате размножение частиц протекало неограниченно экспоненциально. Несмотря на то что такое приближение справедливо лишь для начальных циклов реакции, удалось показать, что стохастическая природа реакции не препятствует точному определению числа исходных копий. В процессе реальной ПЦР накопление фрагмента постепенно замедляется и его количество, регистрируемое в виде кинетической кривой методом ПЦР в реальном времени (ПЦР-РВ), выходит на плато. Мы разработали стохастическую модель ПЦР, которая позволила получить кинетические кривые, максимально приближенные к наблюдаемым в эксперименте.

Экспериментальные кинетические кривые ПЦР, полученные в процессе амплификации одиночных исходных копий ДНК, выходят на плато после 45–50 циклов реакции, что позволяет приблизительно оценить предельное количество копий, присутствующих в реакции, как  $2^{45}$ . При мо-

делировании такого количества частиц методом Монте-Карло объем вычислений не позволяет накопить достаточную статистику за приемлемое время. В то же время, если имеется  $N$  частиц, каждая из которых генерирует одну частицу с вероятностью  $p$ , то мы имеем дело с биномиальным распределением вероятностей для количества генерируемых частиц. Известно, что при больших значениях  $N$  биномиальное распределение может быть приближено нормальным с математическим ожиданием  $Np$  и дисперсией  $Np(1-p)$  [2]. Тогда для получения случайного количества частиц, образованных в результате данного шага размножения при известных  $N$  и  $p$ , достаточно генерировать случайную величину, имеющую нормальное распределение с заданными параметрами. Для этого был использован простой алгоритм [3], позволяющий получить такую случайную величину с целью ускорения расчетов. В течение первых циклов размножения модель работает по алгоритму, описанному в [1], а после достижения заданного порогового количества частиц переходит к генерации случайной гауссовой величины, представляющей количество новых частиц, полученных в текущем цикле.

Результаты имитационных экспериментов для новой модели проанализированы при различных значениях порога:  $10^5$ ,  $10^4$ ,  $10^3$  и  $10^2$  частиц. Соответствующие функции распределения плотности:  $f_5(n)$ ,  $f_4(n)$ ,  $f_3(n)$ ,  $f_2(n)$ . Также проведено сравнение с результатами имитационных экспериментов на модели [1], которая в данном случае соответствует бесконечно большому значению порога —  $f_\infty(n)$ . Результаты анализа приведены на рис. 1 и в табл. 1. Визуальное сопоставление плотностей вероятности свидетельствует об эффективности предложенной модели. С помощью критерия Колмогорова–Смирнова [4] проведен анализ идентичности функции распределения  $F_\infty(n)$ , соответствующей функции плотности вероятности  $f_\infty(n)$ , и функций распределения  $F_5(n)$ ,  $F_4(n)$ ,  $F_3(n)$ ,  $F_2(n)$ , соответствующих функциям плотности  $f_5(n)$ ,  $f_4(n)$ ,  $f_3(n)$ ,  $f_2(n)$ . Критическое значение числа

---

ЗАО “Синтол”, Москва

Институт аналитического приборостроения  
Российской Академии наук, Санкт-Петербург  
Институт иммунологии, Москва

**Таблица 1.** Параметры распределения для результатов реализации имитационного алгоритма с различными величинами порогового значения количества частиц

Характеристика	$f_{\infty}(n)$	$f_5(n)$	$f_4(n)$	$f_3(n)$	$f_2(n)$
Математическое ожидание	632061.7	632109.3	632010.2	630302.5	629233.1
Дисперсия	71744.4	71997.3	71807.9	71496.2	71391.9
Максимальное значение $ F_{\infty}(n) - F_i(n) $	0	0.0018	0.0011	0.0163	0.0257

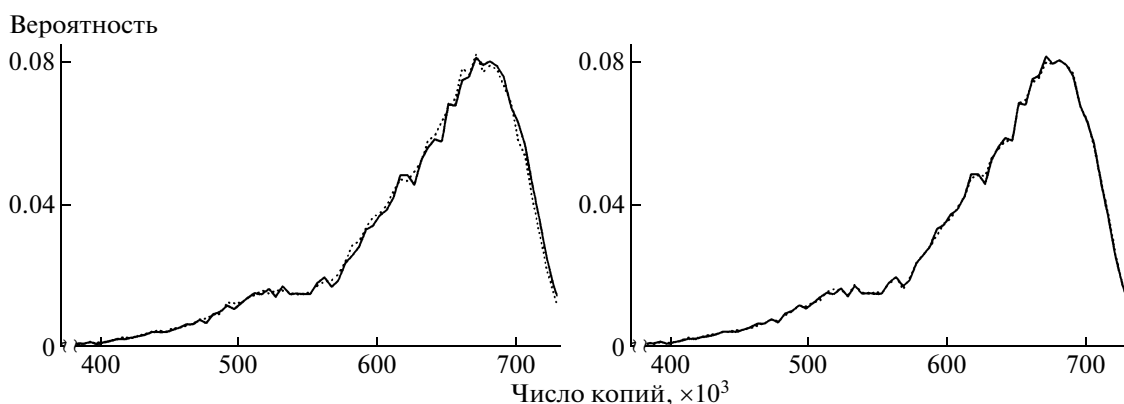
Колмогорова–Смирнова  $\lambda_{КС}$  для уровня значимости  $\alpha = 0.05$  вычислялось по формуле [5]:

$$\lambda_{КС} = 1.36 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}, \quad (1)$$

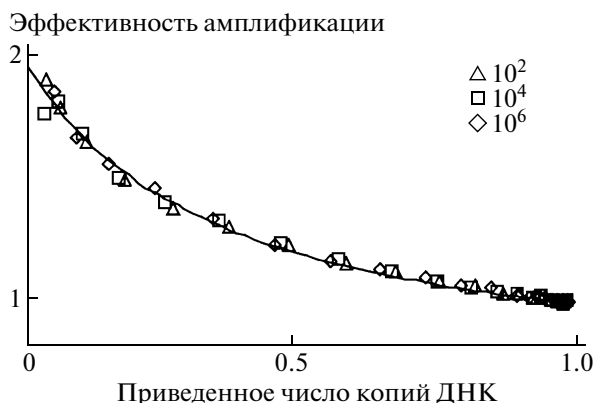
где  $n_1$  и  $n_2$  – объем выборок. В нашем случае  $\lambda_{КС} = 0.0061$ . Как видно из табл. 1, максимальное различие функций распределения  $|F_{\infty}(n) - F_i(n)|$  не превысило значения  $\lambda_{КС}$  только для  $F_5(n)$  и  $F_4(n)$ . Следовательно, для заданного уровня значимости можно утверждать, что только при значениях порога  $10^4$  и  $10^5$  получаемые функции распределения не имеют существенных отличий от функции контрольного распределения  $F_{\infty}(n)$ .

Вероятность синтеза новой копии амплифицируемого фрагмента на отдельном цикле ПЦР может быть оценена по отношению амплитуды кинетической кривой реакции на предыдущем и данном цикле. Примем предположение о том, что эта вероятность зависит только от числа копий ДНК, присутствующих в реакции. Тогда форма кинетической кривой зависит только от начальной концентрации фрагмента и для любой экспериментальной кривой можно построить зависимость вероятности синтеза от количества продукта реакции, приведенного к максимальному. Эту зависимость можно назвать кривой эффективности ПЦР. Примеры таких кривых, полученных в

постановках одной и той же реакции амплификации с различными исходными количествами копий, приведены на рис. 2. Очевидно, что эти кривые практически совпадают и могут быть аппроксимированы единой функцией. В исследовании [6], посвященном анализу ПЦР с точки зрения химической кинетики, была предложена максимально упрощенная функция, имеющая единственный параметр. С нашей точки зрения, недостатком этой функции является фиксированная начальная вероятность удвоения, равная единице. Этот факт связан с необходимостью принять время протекания каждого цикла амплификации за бесконечность, что было продиктовано требованием простоты полученного аналитического решения системы кинетических уравнений. Кроме того, предложенная авторами [6] функция построена с допущением неизменной концентрации праймеров по ходу реакции, а поэтому она не описывает плато реакции и, соответственно, не применима для описания заключительных циклов ПЦР. Поскольку для целей нашей работы не важен физический смысл используемой функции, то удобно взять такую функцию, которая в явном виде содержит начальную вероятность удвоения и максимальное количество продукта реакции, соответствующее плато кинетической



**Рис. 1.** Плотность распределения количества копий после 20 циклов репликации одной исходной частицы. Вероятность репликации 0.95. Количество модельных экспериментов 50 000. Сплошная линия –  $f(n)$ , пунктирная слева –  $f_5(n)$ , справа –  $f_3(n)$ .



**Рис. 2.** Зависимость эффективности амплификации от количества образовавшихся копий ДНК. Маркерами обозначены три набора экспериментальных данных, полученных в ходе ПЦР с различным исходным количеством копий. Сплошная линия – кривая аппроксимации функцией (2).

кривой. Аппроксимация кривых эффективности проводилась с помощью следующей функции:

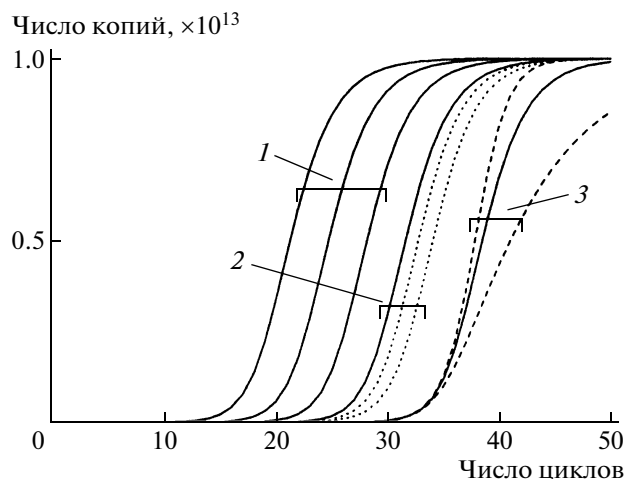
$$p(n) = p_0 - \frac{n}{\frac{1}{p_0} - k + kn}, \quad n = \frac{N}{N_{\max}}, \quad (2)$$

где  $p_0$  – начальная вероятность удвоения,  $n$  – приведенное количество копий,  $N$  – абсолютное количество копий,  $N_{\max}$  – максимальное количество копий,  $k$  – параметр, определяющий форму кривой.

Чтобы применить функцию эффективности в модели, необходимо оценить максимальное количество копий фрагмента, достигаемое в реальных реакциях амплификации. Для этого были проведены постановки ПЦР с последовательными разведениями образца ДНК, аналогично методу лимитирующих разведений, и выбраны кривые амплификации с максимальным значением порогового цикла. Такие кривые должны соответствовать реакциям, начинающимся с единичной исходной копии ДНК. Кривые были нормированы по амплитуде таким образом, чтобы начальный уровень сигнала был равен нулю, а максимальный уровень на плато реакции – единице. Затем проведена аппроксимация рекурсивной функцией вида:  $n_i = n_{i-1} + n_{i-1}(1 + p(n_{i-1}))$ , где  $n_i$ ,  $n_{i-1}$  – приведенные количества копий на циклах  $i$  и  $i-1$ ,  $p(n_{i-1})$  – значение функции (2) для  $n_{i-1}$ .

В процессе аппроксимации варьировали параметры  $k$  и  $n_0$ . Значение  $n_0$ , соответствующее начальному приведенному количеству копий, составило примерно  $10^{-13}$ , что соответствует максимальному количеству  $10^{13}$  копий, синтезируемых в ходе реакции.

Полученная функция эффективности (2) была включена в модель с параметрами  $p_0$ ,  $k$  и  $N_{\max}$ . Это позволило определять вероятность размножения



**Рис. 3.** Кинетические кривые ПЦР, полученные при различных значениях параметров модели. Группа 1:  $p_0 = 0.95$ ,  $k = 0.7$ ,  $N_0 = 10^7$ ,  $10^6$  и  $10^5$  (слева направо); группа 2:  $k = 0.7$ ,  $N_0 = 10^4$ ,  $p_0 = 0.95$ ,  $0.9$  и  $0.85$  (слева направо); группа 3:  $p_0 = 0.95$ ,  $N_0 = 10^2$ ,  $k = 0.5$ ,  $0.7$  и  $0.9$  (слева направо).

на каждом цикле реакции в зависимости от имеющегося количества частиц.

Поскольку в модели участвуют два типа частиц, соответствующие двум комплементарным цепям амплифицируемого фрагмента, для каждого из двух типов имеется своя функция эффективности. Вообще, функция эффективности для разных типов цепей может иметь различный характер, зависящий от кинетики связывания соответствующего праймера и элонгации образовавшегося дуплекса. Кроме того, вполне вероятно, что при больших концентрациях цепей эффективность реакции зависит от общего количества этих цепей. Тем не менее принимая, что на начальном участке роста кривой концентрации цепей не слишком велики, мы посчитали, что реакции не влияют друг на друга и зависимость кинетики от концентрации сходна для обоих типов цепей. Таким образом, в модели использовались равные значения коэффициентов функции эффективности для двух типов частиц, кроме значений  $p_0$ , для которых хотелось оценить влияние на ход реакции.

Для оценки работы предложенной модели была создана программа, позволяющая проводить расчет количества копий частиц на каждом цикле амплификации при различных значениях параметров модели. Программа и ее описание доступны на веб-странице [www.syntol.ru/science/stochasticPCR/main.html](http://www.syntol.ru/science/stochasticPCR/main.html).

На рис. 3 приведены кинетические кривые, полученные при различных реализациях модели. Можно видеть, что общая форма кривых соответствует наблюдаемой в реальных экспериментальных условиях, что определяется заложенной в мо-

дель функцией эффективности. Поведение кривых, полученных при различном начальном количестве частиц  $N_0$  (рис. 3, группа 1), демонстрирует регулярный сдвиг в направлении увеличения порогового цикла при снижении начальной концентрации в полном соответствии с поведением экспериментальных кривых ПЦР в сериях разведений образца. Для двух фиксированных начальных количеств продемонстрированы эффекты вариации параметров модели на форму кинетической кривой. В одном случае (рис. 3, группа 2) изменяли начальную вероятность удвоения  $p_0$ , в результате чего по мере уменьшения параметра кривая смещалась в направлении больших циклов, в то же время ее форма не изменялась. Это явление соответствует изменению угла наклона калибровочной прямой, обычно наблюдаемой в ПЦР при низкой эффективности реакции, вызванной, например, неоптимальным тепловым режимом ее проведения. В другом случае (рис. 3, группа 3) варьировали параметр  $k$  функции эффективности. Это привело к изменению формы кривой при отсутствии смещения в области ее начального роста. Подобные изменения формы характерны для реакций, протекающих в условиях недостаточной активности полимеразы либо в присутствии ингибиторов ПЦР.

Далее было проанализировано статистическое поведение модели. Для этого проведены повторные реализации модели при фиксированных значениях параметров. По полученным кинетическим кривым вычислены значения порогового цикла при значении порога  $10^{12}$  копий и оценены их распределения вероятностей. Для каждого распределения определяли значения математического ожидания и стандартного отклонения. В табл. 2 приведены зависимости параметров распределения порогового цикла от вероятности  $p_0$  и от параметра  $k$  функции эффективности. Можно видеть, что интервал разброса значений порогового цикла сильно зависит от начальной вероятности удвоения  $p_0$  и значительно менее от параметра  $k$ . В то же время, если оценить доверительный интервал порогового цикла с вероятностью 99.7% как трехкратную величину стандартного отклонения, то его величина не превосходит  $\pm 0.1$  цикла даже для весьма неэффективных реакций ( $p_0 = 0.7$ ) и составляет пренебрежимо малую величину для реакций, которые обычно используются в практике (при  $p_0 \sim 0.95$ ).

Итак, предложенная в данной работе модель позволяет генерировать реалистичные кинетические кривые ПЦР, а также анализировать влияние

**Таблица 2.** Влияние параметров модели  $p_0$  и  $k$  на распределение значений порогового цикла

Параметр	Пороговый цикл	Стандартное отклонение, $\times 10^{-3}$
$p_0 (k = 0.7)$		
1	30.13	0.02
0.98	30.53	3.5
0.95	31.23	4.9
0.9	32.46	8.0
0.8	35.43	12.1
0.7	39.2	16.1
$k (p_0 = 0.95)$		
0.9	31.58	5.7
0.8	31.33	5.3
0.7	31.23	5.0
0.6	31.18	4.9
0.5	31.15	4.7

различных параметров реакции на поведение этих кривых. Возможность быстрых повторных реализаций расчета позволяет исследовать влияние стохастических эффектов на результаты реакции. В совокупности полученные инструменты могут быть использованы для дальнейшего изучения кинетики ПЦР, а также установления теоретических пределов ее аналитической чувствительности.

Работа выполнена при поддержке Программы фундаментальных исследований Президиума РАН № 22, раздел “Бионанотехнологии”. Работа выполнена с использованием оборудования Центра коллективного пользования “ВНИИСБ” Всероссийского научно-исследовательского института сельскохозяйственной биотехнологии.

#### СПИСОК ЛИТЕРАТУРЫ

1. Социвко Д.Г., Федоров А.А., Курочкин В.Е., Петров Р.В. // ДАН. 2010. Т. 434. № 2. С. 265–268.
2. Феллер В. Введение в теорию вероятностей и ее приложения. М.: Мир, 1964. Т. 1. 499 с.
3. Marsaglia G., Bray T.A. // SIAM Rev. 1964. V. 6. P. 260–264.
4. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. 416 с.
5. Шеннон Р. Имитационное моделирование систем — искусство и наука. М.: Мир, 1978. 420 с.
6. Boggy G.J., Woolf P.J. // PLoS ONE. 2010. V. 5. Iss. 8. e12355.