

ТОЧНОСТЬ КОЛИЧЕСТВЕННОГО АНАЛИЗА ДНК С ИСПОЛЬЗОВАНИЕМ ПОЛИМЕРАЗНОЙ ЦЕПНОЙ РЕАКЦИИ В РЕАЛЬНОМ ВРЕМЕНИ

© 2013 г. Д. Г. Сочивко, А. А. Фёдоров, Д. А. Варламов,
В. Е. Курочкин, академик Р. В. Петров

Поступило 24.12.2012 г.

DOI: 10.7868/S0869565213110261

В ходе решения многих задач биологии и медицины возникает необходимость определить присутствие специфических последовательностей нуклеиновых кислот в образцах. Для этого применяют метод полимеразной цепной реакции (ПЦР), представляющей собой цепную ферментативную реакцию размножения анализируемого фрагмента ДНК, реализуемую в виде последовательных циклов удвоения присутствующих в реакции фрагментов.

Помимо наличия исследуемой специфической последовательности ДНК, зачастую нужно также определить ее концентрацию в образце. На настоящее время наиболее распространенным методом количественного анализа нуклеиновых кислот является проведение ПЦР с регистрацией продуктов реакции в режиме реального времени (ПЦР-РВ). Если в процессе реакции производить измерение концентрации ее продуктов на каждом цикле, то мы получим кинетическую кривую ПЦР, имеющую дискретный по времени характер. При этом, задав некоторое пороговое детектируемое значение количества продукта A_t , получим то, что реакции с различным исходным количеством фрагмента A_0 достигнут порогового уровня за $N = \lg_2(A_t/A_0)$ циклов амплификации. Значение N для данной реакции называют пороговым циклом реакции. Очевидно, что две пробы, отличающиеся по исходной концентрации фрагмента в два раза, будут отличаться по пороговому циклу в ПЦР-РВ на единицу [1].

Как и любой метод количественного анализа, ПЦР-РВ позволяет проводить измерения в опре-

деленном динамическом диапазоне и с ограниченной точностью. Динамический диапазон ПЦР-РВ ограничен снизу объемом пробы и покрывает все концентрации, встречающиеся в практических приложениях — от единичных копий исследуемой последовательности до концентраций 10^9 копий на пробу. Точность количественного анализа при помощи определения порогового цикла ПЦР-РВ ограничена несколькими факторами. Прежде всего имеются физические ограничения точности измерения концентрации продуктов реакции, проводимой в приборе для проведения ПЦР-РВ, за счет флуориметрической детекции. Как следствие, пороговый цикл реакции определяется с некоторой погрешностью, зависящей от совершенства оптической системы и методов обработки данных. В то же время существуют факторы, природа которых не зависит от инженерной реализации ПЦР-РВ.

В данной работе рассмотрены два таких источника погрешности количественного анализа методом ПЦР-РВ. Первым источником является случайная вариация количества копий, отбираемых из образца и вносимых в реакцию. Вторым источником погрешности является стохастическая природа самой ПЦР, связанная с ограниченной эффективностью реакции, обычно меньшей единицы [1, 2]. В работе промоделирована типичная ситуация анализа образца, содержащего некоторую концентрацию целевых последовательностей ДНК. Анализ состоит из отбора пробы, постановки ПЦР-РВ с получением кинетической кривой, определения порогового цикла реакции и расчета количества копий целевой ДНК в отобранной пробе. Преимуществом моделирования в данном случае является отсутствие влияния на результат приборных погрешностей, а также возможность многократного повторения анализа для получения статистических оценок.

Пусть после процедуры выделения ДНК из образца имеется исходный объем анализируемой жидкости, содержащий n копий ДНК. Для постановки реакции ПЦР-РВ используется некоторая

“ЗАО Синтол”, Москва

Институт аналитического приборостроения
Российской Академии наук, Санкт-Петербург

Всероссийский научно-исследовательский институт
сельскохозяйственной биотехнологии

Российской сельскохозяйственной академии, Москва

Институт иммунологии, Москва

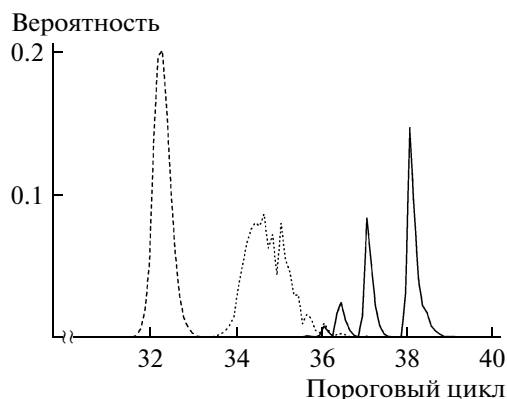


Рис. 1. Распределение вероятностей значений порогового цикла для концентраций 1, 10, 50 копий исходной ДНК на реакцию (справа налево).

часть α данного объема – проба. Концентрация образца определяется как $C = n\alpha$ копий на реакцию.

Принимая положение о равномерности пространственного распределения молекул (об отсутствии каких-либо взаимодействий между молекулами ДНК) в растворе, получим, что вероятность каждой отдельной частицы оказаться в данном объеме равна значению α . Число частиц m , оказавшихся в анализируемой пробе, будет описываться биномиальным распределением, поскольку в данном случае имеется классическая схема испытаний по Бернулли с вероятностью успеха единичного события – α . Таким образом, вероятность получить m частиц в анализируемой пробе $p(m)$ описывается законом

$$p(m) = \frac{n!}{(n-m)!m!} \alpha^m (1-\alpha)^{n-m}.$$

Реализация данного алгоритма была выполнена на основе программы, моделирующей процесс ПЦР-РВ для заданного количества исходных копий ДНК и числа циклов реакции, представленной в работе [2]. Использовали следующие параметры модели, характеризующие типичную реальную реакцию ПЦР-РВ: начальная вероятность удвоения $p_0 = 0.95$, максимальное количество копий $N_{\max} = 10^{12}$, параметр формы кривой $k = 0.7$. Пороговый цикл определяли линейной интерполяцией при значении порога 10^{11} копий (10% от максимального количества). Указанная программа была дополнена модулем, генерирующим исходное число копий на реакцию по задаваемым значениям n и α , с помощью генератора случайной величины с биномиальным законом распределения, полученного методом обращения функции распределения [3]. Значение n для образца с концентрацией C определяли как $n = C/\alpha$. Значение α было принято равным 0.01.

В большинстве случаев исходный образец содержит двухцепочную ДНК, однако в процессе выделения молекула ДНК может денатурировать, в результате чего комплементарные цепи разойдутся и будут внесены в реакцию независимо. Такая ситуация встречается, например, если выделение ДНК проводится при помощи сорбентов, и для элюции ДНК применяют дистиллированную воду. Эта возможность была учтена в модели. Исходное количество каждой из комплементарных цепей ДНК могло задаваться одинаково (метод 1, соответствующий образцу интактной двухцепочной ДНК) или же задаваться независимо, путем повторного запуска данного генератора для одного и того же значения (метод 2, соответствующий образцу денатурированной ДНК).

Для перевода значений порогового цикла N в концентрации использовали калибровочную зависимость, полученную путем анализа усредненных значений порогового цикла для образцов с количеством 100 и 1000 копий на реакцию: $C = 1.934^{(38.26 - N)}$.

Статистический анализ проводили по результатам 50000 реализаций модели при фиксированных параметрах. Результаты ранжировали, доверительный интервал при вероятности 95% определяли по значениям порогового цикла (концентрации ДНК) точек с рангом 1250 и 48750.

Рисунок 1 иллюстрирует изменение формы распределений вероятностей значений порогового цикла с увеличением концентрации анализируемого фрагмента ДНК в образце. Распределение для образца с исходной концентрацией одна копия на пробу демонстрирует наличие нескольких пиков, соответствующих попаданию в реакцию 1, 2, 3, 4 копий. Амплитуда этих пиков, а также количество реализаций с нулевым количеством копий описываются распределением Пуассона. Ширина каждого пика определяется разбросом значений порогового цикла, связанным со стохастическим характером ПЦР при вероятности удвоения менее 1 [1]. Расстояние между соседними пиками уменьшается экспоненциально по мере роста количества копий на реакцию, и постепенно отдельные пики начинают перекрываться. Этот момент можно видеть на примере распределения для образца с концентрацией 10 копий на пробу, в правой части которого еще видны отдельные пики, левая же часть является практически гладкой. Для больших концентраций распределение может быть описано почти симметричной гауссовой кривой (примером является кривая для образца с концентрацией 50 копий на пробу).

На рис. 2 показано изменение среднего значения порогового цикла, а также верхней и нижней границ 95-процентного доверительного интервала в зависимости от концентрации образца для двух методов генерации, соответствующих образцам с нативной двухцепочной либо денатурирован-

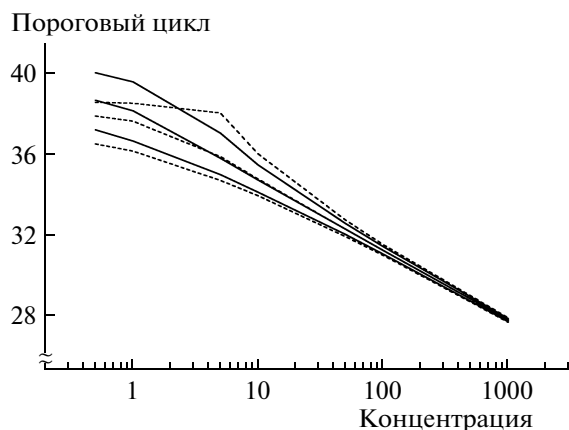


Рис. 2. Зависимость средних значений порогового цикла и доверительного интервала при вероятности 95% от концентрации образца. Пунктирная линия — метод 1, сплошная линия — метод 2 (см. основной текст). Здесь и на рис. 3, 4 концентрация — среднее количество НК в анализируемых пробах.

ной одноцепочной ДНК. Очевидно, что анализ денатурированных образцов обеспечивает более высокую точность определения концентрации, что связано с фактически удвоенным количеством фрагментов ДНК, попадающих в реакцию, в результате чего случайная вариация их числа снижается. Изгиб кривых в области единичных значений концентрации связан с дискретным характером измерения — реакция протекает при внесении хотя бы одной копии фрагмента (либо одиночной денатурированной цепи, что эквивалентно половине копии).

Точность количественного анализа ДНК обычно повышают путем проведения повторного анализа образца. В практических применениях ПЦР-РВ для каждого анализируемого образца ставят обычно 2–3 реакции с одинаковой концентрацией пробы и затем усредняют их результаты. Такой формат анализа был исследован при помощи моделирования. Для этого 50000 последовательных реализаций модели были разбиты на группы по 3 либо 10 последовательных реализаций, в каждой группе было вычислено среднее значение параметра и проведен статистический анализ результатов. Необходимо заметить, что для корректного проведения такого анализа, как в модельных экспериментах, так и на практике, результат каждого измерения должен быть переведен из порогового цикла в концентрацию, и при усреднении должны быть включены нулевые значения, соответствующие отсутствию реакции в пробе. В противном случае средние оценки будут систематически смещены для диапазона малых (единичные копии на пробу) концентраций.

С целью наглядного сопоставления результатов для разных концентраций образца доверительные интервалы при вероятности 95% были

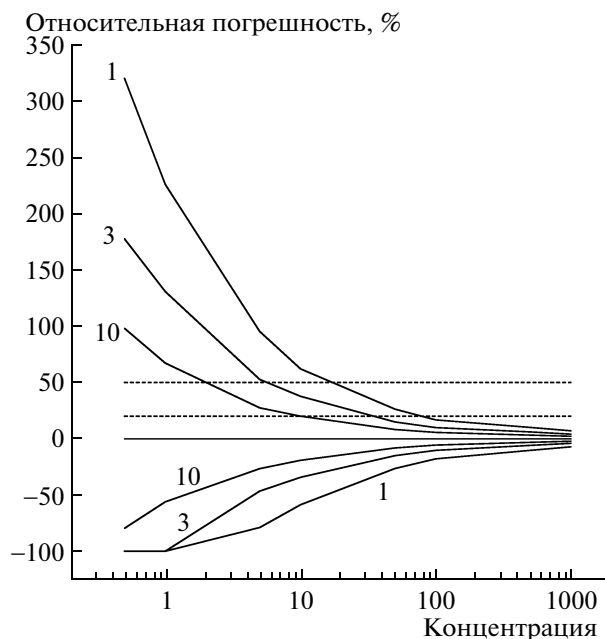


Рис. 3. Зависимость относительной погрешности анализа ПЦР-РВ от концентрации образца для различного числа усредняемых значений (указаны на кривых). Пунктирные линии соответствуют погрешности 20 и 50%.

выражены в единицах относительной погрешности по аналогии с коэффициентом вариации. На рис. 3 показана зависимость относительной погрешности количественного анализа от концентрации образца при проведении нескольких измерений. На основании этих результатов можно оценить минимальные концентрации образцов, для которых относительная погрешность не превышает 20%, как 77 копий на пробу для одиночного измерения, 25 копий на пробу для трехкратного измерения и 14 копий на пробу для 10-кратного измерения. Та же оценка для погрешности менее 50% — 24, 7 и 2 копии на пробу соответственно. В целом можно сказать, что образцы с концентрацией, заведомо превышающей 100 копий на пробу, можно анализировать с приемлемой для практики точностью даже в одном измерении. С другой стороны, если концентрация исследуемого фрагмента может быть нулевой и практически значимо даже присутствие единичных копий, то необходимо использовать не менее трех повторных измерений. Заметим, что при этом повышается не только точность количественного анализа, но и надежность отрицательного результата.

Был предложен принципиально отличный от ПЦР-РВ метод количественного анализа ДНК на основе ПЦР, в котором анализируемая проба разбивается на множество микрообъемов, в каждом из них проводится ПЦР и регистрируется резуль-

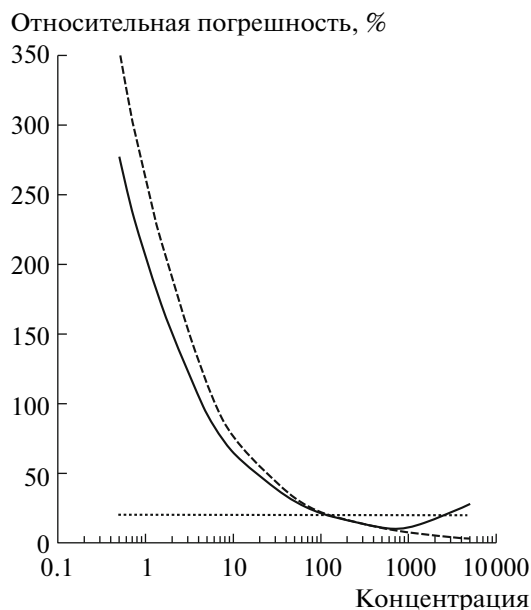


Рис. 4. Зависимость относительной погрешности анализа ПЦР-РВ (штриховая линия) и цПЦР (сплошная линия) от концентрации образца. Пунктирная линия соответствует погрешности 20%.

тат реакции — положительный либо отрицательный. Этот метод был предложен для количественного анализа аллельных вариантов последовательности ДНК и получил название цифровой ПЦР (цПЦР) [4]. Метод реализован, например, в картриджах производства компании “Fluidigm” (США), в которых проба объемом 4.59 мкл распределяется между 765 ячейками, при этом в одном картридже имеются блоки для независимого анализа 12 проб [5]. Динамический диапазон цПЦР очевидным образом ограничен снизу объемом анализируемой пробы, а сверху — количеством ячеек, поскольку с ростом концентрации образца растет вероятность попадания в одну ячейку нескольких копий исследуемого фрагмента ДНК. Точность анализа методом цПЦР рассмотрена в работе [5]. Для анализа в реакторе, содержащем N ячеек, из которых M ячеек дали положительный результат при постановке реакции, вычисляют оценку вероятности положительного результата для каждой ячейки как $p = \frac{M}{N}$. Распределение целевых фрагментов ДНК по ячейкам описывается как пуассоновский процесс, в результате которого в отдельную ячейку может попасть несколько фрагментов ДНК. Точность анализа определяют путем вычисления доверительного интервала для оценки p

$$p_{95\%} = p \pm 1.96 \sqrt{\frac{p(1-p)}{N}}$$

Поскольку цПЦР определяет только сам факт присутствия фрагмента в ячейке, а не его количе-

ство, то для оценки концентрации фрагментов на ячейку C по вероятности p используют уравнение

$$C = -\ln(1 - p).$$

Аналогичным образом оценки границ доверительного интервала вероятности $p_{95\%}$ пересчитывают в оценки для концентрации $C_{95\%}$.

Нами было проведено сопоставление точности анализа с применением методов ПЦР-РВ и цПЦР. Для этого приняли, что проба одного и того же объема вносится в пробирку для ПЦР-РВ и в картридж для цПЦР, содержащий 1000 ячеек. Затем определяли границы доверительного интервала при вероятности 95% для каждого метода и рассчитывали относительную погрешность определения концентрации. Результаты анализа для проб с различной концентрацией приведены на рис. 4. Можно видеть, что по сравнению с ПЦР-РВ цПЦР обеспечивает несколько большую точность при концентрациях менее 100 копий на пробу, и происходит существенное падение точности при концентрациях более 1000. Концентрации, превышающие 10000 копий на пробу (т.е. более 10 копий на ячейку), не могут быть проанализированы с помощью цПЦР в принципе, поскольку количество незаполненных ячеек при этом стремится к нулю.

При низких концентрациях анализируемого фрагмента ДНК точность метода цПЦР практически полностью определяется статистикой испытаний по схеме Бернулли. Различие в точности методов при небольших концентрациях связано с влиянием на результаты анализа стохастической природы ПЦР-РВ. В то же время даже при анализе единичных копий это влияние несущественно по сравнению с вариацией, порождаемой случайным процессом при отборе пробы. Более того, при концентрациях выше 100 копий на пробу вклад стохастического процесса реакции в общую погрешность анализа не выявляется вообще, даже при очень низкой эффективности реакции, что было подтверждено нами путем моделирования реакций с эффективностями 0.9–0.6.

Динамический диапазон цПЦР может быть повышен за счет увеличения объема анализируемой пробы (нижняя граница определяемых концентраций) и числа ячеек (верхняя граница), что требует изменения конструкции картриджа. При этом аналитические характеристики ПЦР-РВ могут быть улучшены в той же мере значительно более простым путем — проведением анализа в повторях, что может быть реализовано любым пользователем ПЦР-РВ независимо от производителей приборов и тест-систем.

Нам представляется, что наиболее перспективным направлением развития количественного анализа ДНК будет создание методов, сочетающих технологии цПЦР и ПЦР-РВ таким образом, что реакция проводится в массиве ячеек, реги-

страция же реакции ведется по принципу ПЦР-РВ в каждой ячейке. Этот принцип реализован, например, в системе OpenAmpa, разработанной компанией “BioTrove” (США) [6].

Работа выполнена при финансовой поддержке Программы фундаментальных исследований Президиума РАН № 24 “Фундаментальные основы технологий наноструктур и наноматериалов”, проект “Исследования кинетики и аналитических характеристик полимеразной цепной реакции в реальном времени”, а также Минобрнауки России по мероприятию 5.2 ФЦП “Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса на 2007–2013 годы” (ГК № 16.552.11.7032).

СПИСОК ЛИТЕРАТУРЫ

1. Социвко Д.Г., Федоров А.А., Курочкин В.Е., Петров Р.В. // ДАН. 2010. Т. 434. № 2. С. 265–268.
2. Социвко Д.Г., Федоров А.А., Лавров В.В., Курочкин В.Е., Петров Р.В. // ДАН. 2011. Т. 439. № 5. С. 696–699.
3. Рыжиков Ю. И. Имитационное моделирование. Теория и технологии. СПб.: КОРОНА принт, 2004. 384 с.
4. Vogelstein B., Kinzler K.W. // Proc. Nat. Acad. Sci. USA. 1999. V. 96. № 16. P. 9236–9241.
5. Dube S., Qin J., Ramakrishnan R. // PLoS One. 2008. V. 3. Iss. 8. e2876.
6. Morrison T., Hurley J., Garcia J., et al. // Nucl. Acids Res. 2006. V. 34. Iss. 18. e123.