
**ПРИБОРОСТРОЕНИЕ ДЛЯ БИОЛОГИИ
И МЕДИЦИНЫ**

УДК 543.07+577.213+681.2.08

© А. Л. Буляница, Н. А. Есикова, А. А. Евстапов, 2022

**ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ КОЛИЧЕСТВЕННОГО
ГЕНЕТИЧЕСКОГО АНАЛИЗА НА ОСНОВЕ АППРОКСИМАЦИИ
КИНЕТИЧЕСКОЙ КРИВОЙ ПОЛИМЕРАЗНОЙ ЦЕПНОЙ РЕАКЦИИ
В РЕАЛЬНОМ ВРЕМЕНИ**

На полимерных микрофлюидных устройствах из поликарбоната и полипропилена реализована количественная полимеразная цепная реакция в реальном времени (ПЦР-РВ). Пороговый цикл определяется на основе точки перегиба функции логистического роста первого порядка, достоверно аппроксимирующей кривую ПЦР при отсутствии мешающих факторов, например наличия пузырей в реакционной камере. Использование статистических критериев (обобщенный критерий Стьюдента, однофакторный дисперсионный анализ) выявило незначимость влияния типа полимера на оценку положения порогового цикла при выбранном ранее алгоритме его поиска. При применении альтернативного алгоритма нахождения порогового цикла на основе построения касательной к кинетической кривой в ряде случаев наблюдается значимое влияние типа полимера на оценку положения порогового цикла и, как следствие, на результат количественного анализа. Предложены и обсуждены алгоритмы обнаружения пузырей в реакционной камере на основе выявления разладки в последовательности измерений, связанные как с оценками параметров аппроксимирующей зависимости, так и с характеристиками временного ряда, сформированного погрешностями аппроксимации.

Кл. сл.: полимеразная цепная реакция в реальном времени, кинетическая кривая, функция логистического роста первого порядка, оценка параметров, однофакторный анализ, восходящая и нисходящая серия, погрешность аппроксимации

ВВЕДЕНИЕ

Количественный генетический анализ, проводимый методом полимеразной цепной реакции в реальном времени (ПЦР-РВ), основывается на выявлении положения характерной точки кинетической кривой — порогового цикла. Наиболее часто в качестве порогового цикла используют либо временное положение точки перегиба кинетической кривой, либо временное положение точки пересечения касательной к кинетической кривой, проведенной в точке перегиба с горизонтальной прямой, задающей характерный уровень сигнала. Характерным уровнем может быть фоновый сигнал или нижняя граница сигнала, практически достоверно отличающегося от уровня фона. Возможны и другие варианты.

Одной из задач работы является проверка возможности замены анализа непосредственно экспериментальных данных (кинетической кривой) на анализ аппроксимирующей зависимости и обоснованности применения в качестве последней функции логистического роста первого порядка. Выбор данной функции связан с тем, что она, во-первых, априорно удовлетворяет качественно-му виду кинетической кривой (относится к семей-

ству сигмоидных функций) и, во-вторых, все ее три параметра имеют понятную содержательную интерпретацию.

Основным оцениваемым параметром аппроксимирующей зависимости является временное положение ее точки перегиба, что необходимо для определения порогового цикла при использовании обоих вышеописанных алгоритмов.

Второй задачей является выявление значимости влияния материала чипа — поликарбоната или полипропилена — на протекание ПЦР и на оценку положения порогового цикла применительно к различным анализируемым пробам, полученным путем различного разбавления исходной пробы. При этом данная задача решалась как для случая применения первого, так и для применения второго алгоритма поиска порогового цикла.

Третья задача связана с выявлением наличия разладки в последовательности измерений, одной из причин которой являлось наличие пузырей в реакционной камере на протяжении нескольких, преимущественно от 3 до 7, циклов ПЦР. Выявление критериев присутствия пузырей, исходя из формы кинетической кривой, является практически важной задачей. Ее решение позволит сделать вывод о допустимости использования полученных

оценок положения порогового цикла и, как следствие, оценить правильность результатов количественного анализа.

Предложенные и обсужденные далее критерии можно отнести к двум группам. Первая группа критериев базировалась на определении параметров аппроксимирующей функции, вторая — на оценке характеристик погрешностей аппроксимации.

УСЛОВИЯ ЭКСПЕРИМЕНТА

Были изготовлены микрочипы из поликарбоната Novattro (SafPlast, Россия) и полипропилена PP 4445S (ТУ 20.16.51-136-05766801-2015, ПАО Нижнекамскнефтехим, Россия) методом термопрессования в гидравлическом прессе ММ-100 (MTDI, Корея) на мастер-форме из нержавеющей стали, изготовленной методом лазерной микрообработки. Для герметизации структур использовались "Пленки для ПЦР плашек полимерные" Р-500 (ООО "ПКФ Современные технологии"). Топология микрочипов представляет из себя 3 камеры с подводящими каналами (рис. 1) [1].

На микрочипах проведена ПЦР-РВ по обнаружению ДНК *сои*, выделенной вручную с использованием набора реагентов М-сорб-ОМ для выделения ДНК и РНК из клинических образцов и объектов окружающей среды. Постановка реакции проводилась для исходной концентрации выделенных нуклеиновых кислот, для разведений на 1, 2 и 3 порядка и для положительного контрольного образца используемого набора "Расте-

ние универсал" в трех повторах для каждого варианта. При работе с данным набором детектирование мишени проводится по красителю R6G, внутреннего положительного контроля — по *Sy5*. Все реагенты предоставлены ООО "Синтол" (Москва). Для сравнения аналогичные измерения проведены в пробирках на АНК-48 (ИАП РАН, СПб).

Измерения в микрочипах проводились на специально созданном макете, обеспечивающем режим термоциклирования для проведения ПЦР-РВ, возбуждение на длинах волн 530, 570 и 685 нм, регистрацию на 580, 630 и 660 нм соответственно. Зона детектирования представляла собой круг диаметром 3 мм. Камеры микрочипов заполнялись реакционной смесью с добавлением пробы ДНК, входы/выходы герметизировались ПЦР-пленкой. Микрочип размещался на нагревательном элементе пленкой вниз.

Одним из существенных недостатков при постановке ПЦР в камере чипа по сравнению с пробирками является образование пузырьков газа при термоциклировании. Оно не оказывает заметного влияния на протекание реакции, однако препятствует проведению корректного оптического детектирования.

АНАЛИЗ КИНЕТИЧЕСКИХ КРИВЫХ ПЦР

Определение положения порогового цикла выполняется на основе поиска точки перегиба кривой ПЦР. Используется аппроксимирующая зависимость логистического роста первого порядка:

$$y = \frac{a}{1 + \exp(-k(x - x_0))}. \quad (1)$$

Здесь a — максимальный сигнал (сигнал насыщения); k — коэффициент, определяющий скорость увеличения сигнала, связанный с эффективностью ПЦР; x_0 — точка перегиба кривой логистического роста первого порядка. Очевидно, что для нормированного сигнала параметр a должен быть близок к 1.

Значимость зависимости положения порогового цикла от материала микрочипа — поликарбоната (pc) и полипропилена (pp) — определяется известными статистическими методами.

1. Обобщенный критерий Стьюдента позволяет дать оценку однородности положения порогового цикла для двух выборок, соответствующих различным материалам.

2. Степень влияния материала на положение порогового цикла также выявляется с помощью однофакторного дисперсионного анализа.

Объемы выборок (число повторений экспериментов) малы (как правило, 3 повтора).

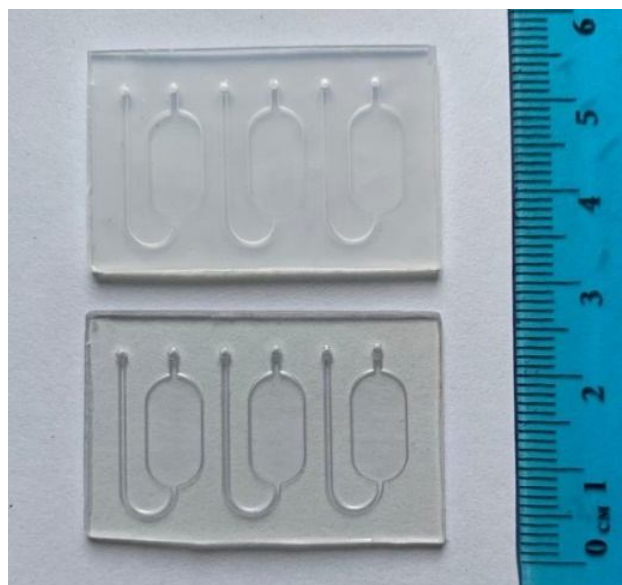


Рис. 1. Изображение микрочипов. Верхний — из полипропилена, нижний — из поликарбоната

Метод одно- или многофакторного анализа хорошо известен [2]. Однако его используют до сих пор, в том числе и в различных прикладных задачах, например, см. [3, 4]. Поскольку даже двухфакторный анализ мы не использовали, то под термином "факторный анализ" полагаем однофакторный анализ. Расчетные формулы и необходимые критерии обсуждаются в [2].

Обозначения пробы: материал полимера определяет буквенную часть — рс или рр, цифра 1, 2 или 3 задает степень разбавления исходной пробы на 1, 2 или 3 порядка, т.е. в 10, 100 или 1000 раз соответственно (символ с на месте цифры — неразбавленная проба).

Проба v_{pk} , или ВПК, представляет собой внутренний положительный контроль. Положение порогового цикла, смещенное более чем на 7 циклов по отношению к пороговому циклу данной пробы, свидетельствует об отсутствии анализируемого генетического материала. В табл. 1 представлены результаты определения положения порогового цикла в пробе внутреннего положительного контроля.

Динамика оценок порогового цикла при использовании различных наборов измерений сигнала пробы внутреннего положительного контроля иллюстрируется данными табл. 2. Оценка положения порогового цикла при использовании всей выборки 28.15 ± 0.04 цикла.

Практически все оценки положения порогового цикла, за исключением сделанных по отсчетам 1–25, имеют приемлемую точность с отклонением средней оценки в пределах нескольких сотых цикла. Принципиально важным является попадание точки перегиба кривой в диапазон выбранных отсчетов сигнала.

Диапазон 1–30 содержит лишь 2 измерения после точки перегиба, и точность оценивания ниже. В остальных случаях оценки положения точки перегиба сопоставимы по точности. Оценка точки перегиба осуществлялась по аппроксимирующей кривой логистического роста первого порядка во всех случаях. При этом при всех вариантах расчета, кроме первых двух, точка перегиба кинетической кривой являлась внутренней точкой, удаленной от границ временного интервала (цикл 1, цикл 40) по крайней мере на 20%-ю часть от длины интервала. Для интервала 1–35 положение порогового цикла 28.09 отстоит от верхней границы примерно на 7 циклов при полной ширине интервала 34 цикла.

Далее в табл. 3 представлены данные расположения точки перегиба кинетической кривой для исходной пробы, разбавленной на 1 порядок (в 10 раз), при использовании микрофлюидных чипов из различных полимерных материалов.

Табл. 1. Расчетные данные по положению порогового цикла

№ п/п	Рс_vpk	Рр_vpk
1	28.15 ± 0.04	27.68 ± 0.05
2	28.15 ± 0.04	28.19 ± 0.04
3	28.24 ± 0.04	28.54 ± 0.03
Среднее	28.18	28.14

Табл. 2. Оценки положения порогового цикла в зависимости от выбора отсчетов информативного сигнала

Диапазон отсчетов	Пороговый цикл	Диапазон отсчетов	Пороговый цикл
1–25	45.66	1–40	28.10 ± 0.04
1–30	28.28 ± 0.24	5–40	28.10 ± 0.04
1–35	28.09 ± 0.06	10–40	28.10 ± 0.04
1–50	28.15 ± 0.04	15–40	28.10 ± 0.03

Табл. 3. Оценки положения точки перегиба кинетической кривой при использовании устройств из различных полимерных материалов

№ п/п	Pc_1	Pp_1
1	25.40 ±0.04	25.42 ±0.02
2	25.55 ±0.04	25.31 ±0.12
3	25.52 ±0.03	26.18 ±0.02
Среднее	25.49	25.64

Табл. 4. Положения точек перегиба кинетических кривых для разных разведений пробы

№ п/п	Pc_c	Pp_c
1	22.00 ±0.03	22.02 ±0.05
2	21.88 ±0.03	22.02 ±0.15
Среднее	21.94	22.02
№ п/п	Pc_2	Pp_2
1	28.22 ±0.04	30.31 ±0.48
2	28.61 ±0.04	29.20 ±0.36
3	28.15 ±0.04	29.97 ±0.11
Среднее	28.33	29.83*
№ п/п	Pc_3	Pp_3
1	31.48 ±0.06	31.12 ±0.03
2	31.71 ±0.03	31.36 ±0.03
3	—	31.75 ±0.04
Среднее	31.59	31.41

Примечание: *недостовечно из-за наличия пузырей (экспертная оценка).

Положение порогового цикла полагаем совпадающим с координатой точки перегиба кинетической кривой. Мы искали точку перегиба аппроксимирующей зависимости (1). Использование процедуры однофакторного дисперсионного анализа позволяет вычислить характерное значение случайной величины F — отношение двух дисперсий, связанных с расхождением средних оценок и дисперсией измерений внутри выборки, соответствующей каждому фактору, удовлетворяющей распределению Фишера с 1 и 4 степенями свободы, исходя из двух градаций фактора (два полимера) и суммарного числа 6 измерений. Этой величине F , примерно равной 0.291, соответствует доверительная вероятность 60%. Это значит, что с такой вероятностью расхождение элементов выборок (значений оценок порогового цикла) не может объясняться влиянием полимера.

Схожие результаты с вероятностями порядка 50% наблюдаются и для оценок порогового цикла при разбавлениях на 3 порядка (F около 0.534). В табл. 4 приведены данные положений точек перегиба кинетических кривых, соответствующих другим разведениям исходной пробы.

Вышеуказанные значения также, в соответствии с методикой однофакторного дисперсионного анализа, свидетельствуют о независимости положения порогового цикла от выбора материала при заданном разведении пробы. Правда, уровень значимости не превышает 50%. Сопоставление положений порогового цикла для проб с разбавлением на 2 порядка не проводилось, поскольку на полипропиленовых чипах были выявлены пузыри.

Далее на рис. 2 и 3 проиллюстрировано распределение параметров аппроксимации a и k применительно к различным кинетическим кривым ПЦР-РВ.

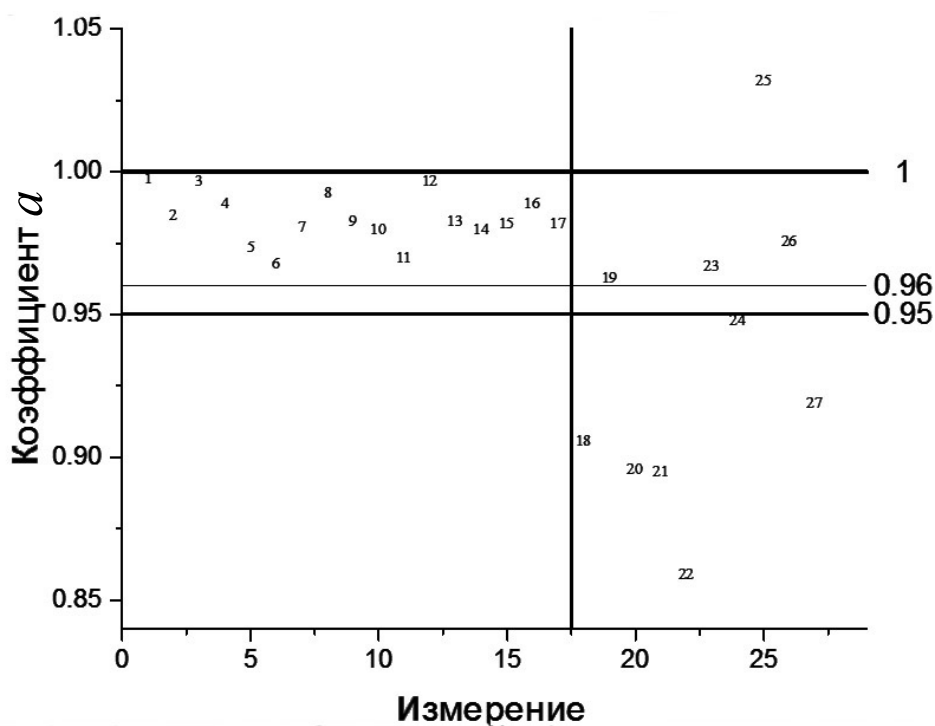


Рис. 2. Распределение оценок параметра a зависимости (1)

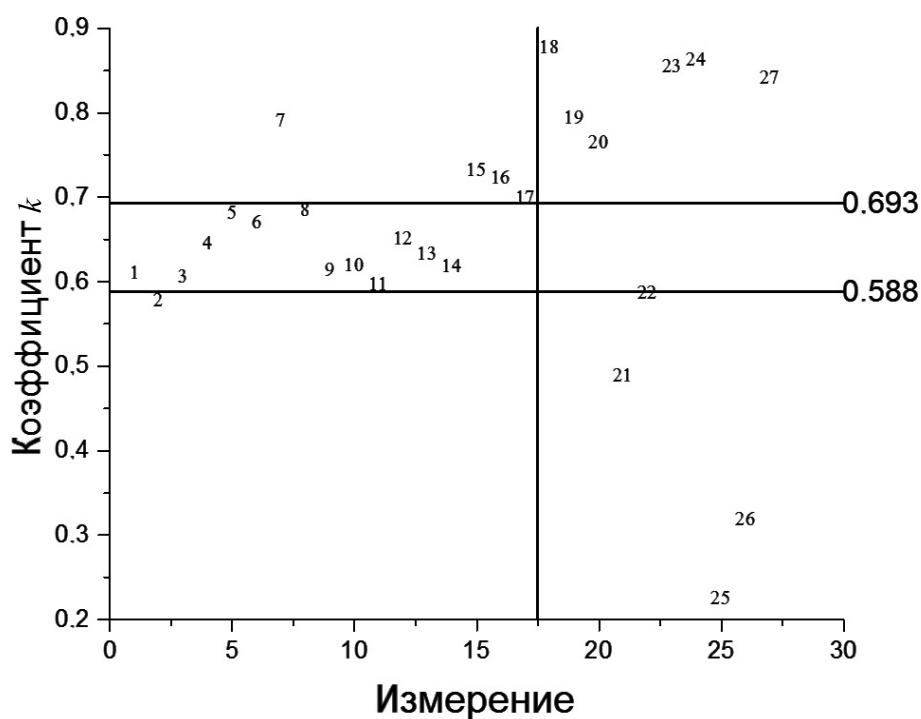


Рис. 3. Зависимость распределения параметра k модели (1)

Изначально сигнал нормирован и приведен к диапазону $[0;1]$. Нулю соответствует наименьшая величина сигнала, единице — наибольшая.

Таким образом, границы диапазона определяются по локальным минимумам и максимумам сигнала без его предварительной обработки.

ФОРМИРОВАНИЕ И ОБСУЖДЕНИЕ КРИТЕРИЕВ НАЛИЧИЯ ПУЗЫРЕЙ В РЕАКЦИОННОЙ КАМЕРЕ

Очевидно, что наличие пузыря влияет на форму кинетической кривой. Изменение формы сигнала или/и скачкообразное изменение его параметров при сохранении вида зависимости относятся к двум типам разладки в последовательности измерений. Выявление разладки основывается либо на оценке параметров аппроксимирующей зависимости (1) с учетом их физического смысла (роли), либо на оценке погрешности аппроксимации. При этом принято допущение, что собственно погрешность аппроксимации пренебрежимо мала, а после компенсации аппроксимирующей детерминированной составляющей информативного сигнала погрешность носит случайный характер. Тогда эта погрешность анализируется на наличие/отсутствие детерминированных составляющих (трендов).

Предложенные и обсужденные далее критерии отсутствия пузырей и, как следствие, доверия к полученным результатам анализа основываются на попадании параметров детерминированной зависимости (1) в диапазон, соответствующий физически обоснованным значениям. В то же время группа критериев, связанная с оценкой параметров помехи, базируется на статистических оценках выборки с отсутствием трендов.

В настоящий момент наличие пузыря и его влияние на форму и параметры кинетической кривой требует исключения соответствующего результата анализа как недостоверного. Во многих случаях степень искажения теоретически обоснованной формы сигнала пробы как сигмоидной функции определяется исключительно на основе субъективной оценки.

Согласно экспертным оценкам, измерения 1–17 соответствуют отсутствию пузырей, 18–27 — наличие пузырей(ей).

На графике (рис. 2) выделена горизонтальная область допустимых значений параметра a при максимальном уровне нормированного сигнала 1. Теоретически оценка этого параметра должна совпадать с максимумом сигнала, он же уровень насыщения.

Жирными горизонтальными прямыми выделена область 0.95–1, в которой может находиться оценка параметра a при достаточно большом отношении сигнал/шум.

Отношение сигнал/шум для рассматриваемых нормированных сигналов пробы при условии корректной компенсации фонового сигнала и отсутствии статистических выбросов находится в диапазоне (50–60) к 1. При этом в качестве характерного значения сигнала рассматривается его среднее

значение, примерно равное 0.5, а в качестве масштаба шума берется среднеквадратичное отклонение погрешности аппроксимации.

Первые 17 измерений, характеризующиеся отсутствием пузырей, полностью попадают в указанную область. При этом из 10 измерений с присутствием пузырей в данный интервал попадает лишь 3 значения из 10. Аналогичные результаты будут получены при повышении нижней границы области допустимых значений параметра a до значения 0.96.

Связь между средней эффективностью ПЦР E и параметром уравнения k имеет вид:

$$E = \exp(k) - 1.$$

Поскольку теоретически обоснованное значение эффективности ПЦР не может превышать 1 (или 100%), то ожидаемое значение параметра k в аппроксимирующем уравнении не должно превышать 0.693.

На рис. 3 горизонтальные прямые соответствуют теоретически обоснованным значениям средней эффективности ПЦР 80% (нижняя прямая со значением 0.588) и 100% (верхняя прямая со значением 0.693). Из 17 измерений при отсутствии пузырей 12 достоверно попадают в указанную область, 5 находятся вне ее (2 из этих 5 измерений практически позиционируются на границе допустимых значений). При этом из 10 измерений с пузырями в данном диапазоне нет ни одного результата (1 практически попадает на границу, а остальные 9 достоверно вне области).

Комбинация этих ограничений позволяет выявить подозрительные с точки зрения наличия пузырей экспериментальные данные.

В некоторых алгоритмах параметр модели (1) x_0 — абсцисса точки перегиба определяет положение порогового цикла. Другим алгоритмом определения порогового цикла является поиск абсциссы точки пересечения касательной, проведенной в точке перегиба ПЦР-кривой, с осью ординат.

Если в аппроксимации (1) параметр a близок 1, то при использовании второго алгоритма положение порогового цикла смещено относительно x_0 и примерно равно $x_0 - 2/k$.

Другие методы поиска параметров зависимости (1) используют не всю выборку (данные по всем циклам 1–40 или 1–50 циклам), а только набор отдельных элементов. Например, ранее в работе [5] была предложена оценочная формула для величины $C = 1 + E = \exp(k)$. Однако очевидно, что предложенная в этой работе четырехточечная формула, использующая измерения сигнала в двух парах равноотстоящих друг от друга отсчетах, условно нумерованных как

$$n, n + k, n + m \text{ и } n + m + k,$$

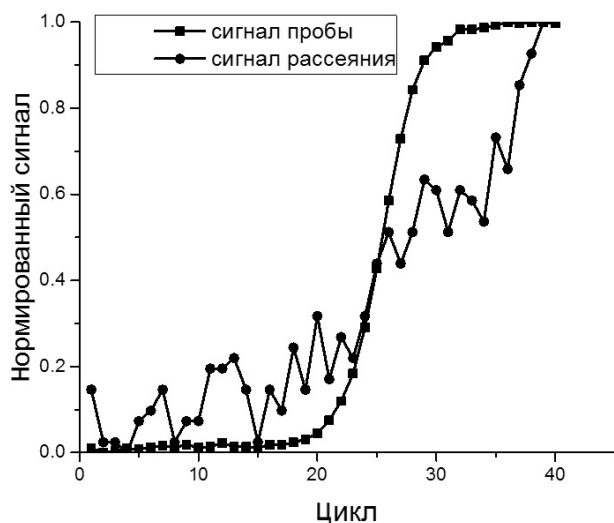


Рис. 4. Нормированные сигналы пробы и рассеяния при отсутствии пузырей (из группы pc_1)

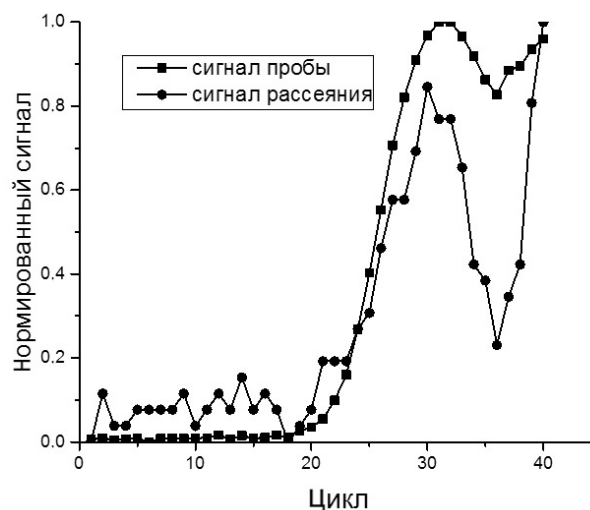


Рис. 5. Нормированные сигналы пробы и рассеяния при наличии пузыря на заключительной стадии (pp_1)

заведомо дает более грубую оценку, чем аппроксимирующая зависимость, которая использует всю выборку измерений, как правило, 40–50.

Также оценка положения точки перегиба кривой может быть выполнена МНК при построении полинома 3-й степени, аналитического вычисления производной второго порядка, приравнивания ее к нулю и нахождения корня решением линейного уравнения. Этот прием был использован при поиске точки перегиба кривой плавления ДНК [6]. Подтверждена ожидаемая закономерность: более точные оценки получаются при использовании отсчетов с линейного участка кривой со значениями нормированного сигнала, близкими к 0.5. При этом лучше брать не ровно 4 точки, а объем выборки должен иметь несколько степеней свободы (не менее 6–8 отсчетов).

Например, для кинетической кривой оценка точки перегиба 25.55 (см. табл. 3, строку 2). Поли-

номиальная аппроксимация кубической параболой по отсчетам 22–27 дает оценку точки перегиба 25.79, что приемлемо для примерного расчета. Здесь приведены типичные зависимости нормированных сигналов пробы и рассеяния при отсутствии пузырей (рис. 4) и при наличии одного пузыря (рис. 5).

Кривая, соответствующая нормированному сигналу рассеяния (см. рис. 4 и 5), может быть адекватно аппроксимирована полиномом второй степени.

Положение порогового цикла, как указано ранее, может определяться с помощью двух алгоритмов. Для исходной пробы, разбавленной в 10 раз (на 1 порядок), в табл. 5 приведены данные оценок положения порогового цикла при анализе на чипах из различных полимерных материалов.

Табл. 5. Расчетные данные по положению порогового цикла в зависимости от метода его поиска

№ п/п	Точка перегиба		Касательная	
	Pc_1	Pp_1	Pc_1	Pp_1
1	25.40	25.42	22.13	22.36
2	25.55	25.31	22.09	22.57
3	25.52	26.18	22.23	22.75
среднее	25.49	25.64	22.15	22.56

Данные табл. 5 свидетельствуют о том, что при оценке положения порогового цикла с использованием построения касательной к кинетической кривой в точке перегиба на основе однофакторного анализа получаем отношение приведенных дисперсий F примерно 11.65. Вывод: с вероятностью 97% различие положений порогового цикла при использовании точки пересечения касательной к перегибу кривой и оси ординат объясняется именно различием материала (поликарбонат или полипропилен). Для алгоритма поиска порогового цикла непосредственно в точке перегиба вывод противоположен: с большой вероятностью фактор выбора материала на положение порогового цикла не влияет.

Другая группа алгоритмов выявления присутствия пузыря может основываться на статистическом критерии восходящих и нисходящих серий погрешности аппроксимации сигналов пробы или рассеяния при достижении высокой достоверности детерминации (коэффициент детерминации заведомо должен превышать 0.9). Один из таких критериев имеет значимость 5% и требует удовлетворения числом восходящих и нисходящих серий двух условий. При общем числе отсчетов (погрешностей аппроксимации), равном N :

$$\begin{cases} ser(N) > \left[\frac{2N-1}{3} - 1.96\sqrt{\frac{16N-29}{90}} \right], \\ \tau_{\max}(N) < \tau^*(N), \end{cases} \quad (2)$$

$$\tau^*(N) = \begin{cases} 5 & N \leq 26, \\ 6 & 26 < N \leq 153, \\ 7 & 153 \leq N \leq 1170. \end{cases} \quad (3)$$

Формула (2) базируется на данных [7, с. 486–488], а ограничения (3) обсуждаются в [8]. Применительно к оценкам информативного сигнала (сигнал пробы) на рис. 4 и 5 объем исходных выборок был 40 отсчетов. Так как для оценки наличия восходящей или нисходящей серий требовалось построить правую разность первого порядка, то объем рассматриваемой выборки разностной погрешности N равен 39. Соответственно, восходящая серия содержит последовательность положительных значений разностной погрешности, нисходящая — отрицательные. Согласно первому условию (2), число серий $ser(N)$ должно быть более 20, согласно второму условию (3), наибольшая длина серии должна быть строго меньше 6, т.е. не более 5.

Наблюдаются следующие тенденции: кривой пробы на рис. 4 соответствует число серий 23 и максимальная длина серии 4. Таким образом,

оба условия (2) и (3) выполнены. Применительно к кривой пробы рис. 5: число серий 19, максимальная длина серии 5. Условие (2) нарушено, условие (3) выполнено.

Аналогичные данные по кривым рассеяния после компенсации детерминированной составляющей — полинома второй степени применительно к рис. 4: число серий 29, максимальная по длине серия 3; применительно к рис. 5: число серий 19, максимальная по длине серия 6.

Если в первом случае (рис. 4) оба условия также выполнены, а это случай отсутствия пузырей, то во втором случае (рис. 5) нарушены оба условия.

Ответ на вопрос, насколько данные критерии присутствия пузырей универсальны, может явиться предметом дальнейшего исследования на основе более репрезентативной статистики. Еще один из критериев можно связать с анализом распределения восходящих или нисходящих фаз (серий) по длине и с расхождением с теоретическими оценками математического ожидания числа серий заданной длины при отсутствии в выборке трендов.

Математическое ожидание числа серий длины d ($d = 1, 2, 3, \dots$) [7] представлено далее формулой (4):

$$N_d = \frac{2(N-d-2)(d^2+3d+1)}{(d+3)!}. \quad (4)$$

При числе отсчетов 39 математические ожидания числа восходящих или нисходящих серий длин 1, 2, ..., 5 равны соответственно 15.00, 6.42, 1.79, 0.38 и 0.07.

Близость реального распределения серий к теоретически обоснованному оценивается с использованием критерия согласия Пирсона (или Хи-квадрат) со следующими поправками и допущениями: а) рассматриваются только три значения — числа серий длины 1, длины 2 и длины 3 или более; б) найденной величине присваивается число степеней свободы 2.5, если она равна 6.3 или более, либо 2 степени свободы при меньших значениях характеристики расхождения. Во втором случае вводится уменьшающий поправочный коэффициент 6/7. Эта классическая методика сформулирована в работе [9]. Особенности распределений фаз по длинам для случаев наличия или отсутствия пузырей иллюстрируются в табл. 6.

Соответственно рассматриваются процентная точка распределения Хи-квадрат с 2 степенями свободы и ее процентные точки [10].

Если уровень вероятности согласия с гипотезой менее 10%, то ее следует отвергать. Следовательно,

Табл. 6. Распределение восходящих и нисходящих фаз по длинам для различных особенностей проведения ПЦР-РВ

Число фаз (серий)	Теория (4)	Проба, рис. 4	Рассеяние, рис. 4	Проба, рис. 5	Рассеяние, рис. 5
$d = 1$	15.00	12	22	10	9
$d = 2$	6.42	7	4	4	5
$d = 3$	1.79	3	3	1	2
$d = 4$	0.38	1	0	2	2
$d = 5$	0.07	0	0	2	0
$d = 6$	0.01	0	0	0	1
Хи-квадрат*		1.64	3.77	4.93	5.04
Вероятность**		44	16	9	8

Примечание: *введен поправочный коэффициент 6/7 [9], а данные по $d = 3, 4, 5$ и 6 просуммированы;

**вероятность округлена до целых процентов.

гипотеза об отсутствии трендов (разладок) после компенсации аппроксимирующей детерминированной составляющей сигнала при наличии пузырей (группа данных рис. 5) должна быть отвергнута. Указанная гипотеза применительно к информативному сигналу пробы при отсутствии пузырей принимается с высоким уровнем значимости (более 40%), что является достаточным для принятия гипотезы об отсутствии трендов в последовательности погрешностей аппроксимации и, как следствие, отсутствия разладки в последовательности исходных измерений.

Следовательно, предложенные алгоритмы принятия решений о наличии пузырей, как связанные с анализом информативного сигнала, так и с погрешностью аппроксимации, с большой вероятностью позволяют делать правильные выводы. Однако насколько полученные результаты специфичны по отношению к выбранному объекту анализа и в какой степени допускают обобщение на другие объекты, требует дополнительного исследования.

ЗАКЛЮЧЕНИЕ

Проведенные серии количественного генетического анализа на основе ПЦР-РВ с использованием полимерных чипов выявили следующие закономерности: при определении порогового цикла как точки перегиба кинетической кривой (логистическая кривая первого порядка) отсутствует значимая зависимость найденных величин от материала чипа (поликарбонат или полипропилен).

В то же время, если оценивать положение порогового цикла по абсциссе точки пересечения касательной к кинетической кривой, построенной в точке ее перегиба, с уровнем фонового (минимального) сигнала, с доверительной вероятностью более 90% фактор материала полимера на положение порогового цикла влияет.

Информативный сигнал (сигнал пробы) при отсутствии пузырей с высокой точностью аппроксимируется логистической кривой первого порядка — одним из типов сигмоидной кривой. При этом коэффициент детерминации превышает 0.99. Также в данных условиях при нормализации сигнала пробы отношение сигнал/шум достигает значений 55–60. Под масштабом (характерной величиной) сигнала понимается среднее значение, под масштабом шума — среднеквадратичное отклонение погрешности аппроксимации.

Проверка критериев наличия пузырей для нормированной кинетической кривой:

а) коэффициент a в аппроксимирующей зависимости существенно ниже 1;

б) нефизическое значение коэффициента k , достоверно превышающее значение 0.70 или существенно меньшее 0.58;

в) на основе анализа восходящих и нисходящих серий в последовательности измерений погрешностей аппроксимации наблюдается невыполнение одного или двух условий отсутствия трендов, а также существенное отклонение длин восходящих и нисходящих серий (фаз) от теоретически обоснованных при отсутствии трендов распределения.

Комбинация предложенных критериев после дополнительной проверки их универсальности (правомерности обобщения на другие биологические объекты или/и другие материалы чипа) может быть включена в программно-математическое обеспечение приборов, что в перспективе исключит участие субъективного фактора (эксперта) при принятии решения о присутствии пузырей в реакционной камере на какой-либо стадии построения кинетической кривой.

Работа выполнена в рамках Государственного задания Министерства науки и высшего образования Российской Федерации № 075-00761-22-00 (тема "Микрофлюидные устройства с интегрированными функциональными микро- и наноразмерными структурами для биологических и медицинских исследований", FFZM-2022-0012).

СПИСОК ЛИТЕРАТУРЫ

1. Есикова Н.А., Гермаш Н.Н., Евстратов А.А. Оперативное изготовление микрочипов для ПЦР-анализа из полимерных материалов в лабораторных условиях // Научное приборостроение. 2020. Т. 30, № 4. С. 21–26. URL: <http://iairas.ru/mag/2020/abst4.php#abst2>
2. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для инженерно-экономических институтов и факультетов. Изд. 4-е, доп. М.: Высшая школа, 1972. 367 с.
3. Тихомиров Н.П., Тихомирова Т.М., Урмаев О.С. Методы эконометрики и многомерного статистического анализа: учебник. М.: Экономика, 2011. 637 с.
4. Плехотников К.Э., Колков С.В. Статистика: учебное пособие. М.: Флинта, 2006. 286 с.
5. Буляница А.Л. Методы оценивания параметров кривой логистического роста. Ч. 1. Оптимизация условий оценивания при наличии аддитивной случайной помехи // Научное приборостроение. 2009. Т. 19, № 3. С. 3–11. URL: <http://iairas.ru/mag/2009/abst3.php#abst1>
6. Белов Д.А., Белов Ю.В., Курочкин В.Е. Новая методика обработки флуоресцентного отклика плавления ДНК // Научное приборостроение. 2018. Т. 28, № 1. С. 3–10. URL: <http://iairas.ru/mag/2018/abst1.php#abst1>
7. Кендалл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976. 736 с.
8. Критерий "восходящих" и "нисходящих" серий. URL: <https://math.semestr.ru/trend/series.php> (дата обращения: 19.04.2022)
9. Wallis W.A., Moore G.H. A significance test for time series analysis // J. Amer. Statist. Ass. 1941. Vol. 36, is. 215. P. 401–409. URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1941.10500577>
10. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. 416 с.

**Институт аналитического приборостроения РАН,
Санкт-Петербург**

Контакты: Буляница Антон Леонидович,
antbulyan@yandex.ru

Материал поступил в редакцию 29.04.2022

INTERPRETATION OF THE RESULTS OF QUANTITATIVE GENETIC ANALYSIS BASED ON THE APPROXIMATION OF THE KINETIC CURVE OF THE POLYMERASE CHAIN REACTION IN REAL TIME

A. L. Bulyanitsa, N. A. Esikova, A. A. Evstrapov

Institute for Analytical Instrumentation of RAS, Saint Petersburg, Russia

A quantitative polymerase chain reaction in real time (real time PCR) has been implemented on polymer microfluidic devices made of polycarbonate and polypropylene. The threshold cycle is determined based on the inflection point of the first-order logistic growth function, which reliably approximates the PCR curve in the absence of interfering factors, for example, the presence of bubbles in the reaction chamber. Use of statistical criteria (generalized Student's criterion, one-way analysis of variance) revealed the insignificance of the influence of the polymer type on the estimation of the position of the threshold cycle in terms of the previously selected algorithm for its search. When using an alternative algorithm for finding the threshold cycle based on the plotting of a tangent to the kinetic curve, in some cases there is a significant influence of the polymer type on the estimation of the position of the threshold cycle and, as a consequence, on the result of quantitative analysis. Proposed and discussed are algorithms for detecting bubbles in the reaction chamber based on the detection of a discrepancy in the measurement sequence associated with both estimates of the approximating dependence's parameters and the characteristics of the time series formed by approximation errors.

Keywords: real-time polymerase chain reaction, kinetic curve, first-order logistic growth function, parameter estimation, single-factor analysis, ascending and descending series, approximation error

INTRODUCTION

Quantitative genetic analysis by polymerase chain reaction in real time (PCR-RT) is based on the identification of the position of the characteristic point of the kinetic curve — the threshold cycle. Most often, the threshold cycle is either the location in time of the inflection point of the kinetic curve, or the location in time of the intersection point of the tangent to the kinetic curve, drawn at the inflection point with a horizontal straight line specifying the characteristic signal level. The characteristic level can be a background signal or a lower limit of a signal that is practically reliably different from the background level. Other options are possible.

One of the tasks of the work is to check the possibility of replacing the analysis of directly experimental data (kinetic curve) for analysis of approximating dependence and the validity of using first-order logistic growth as the last function. The choice of this function is due to the fact that, firstly, it a priori satisfies the qualitative form of the kinetic curve (belongs to the family of sigmoid functions) and, secondly, all three of its parameters have an understandable and meaningful interpretation.

The main parameter of the approximating dependence to be evaluated is the time location of its

inflection point, which is necessary to determine the threshold cycle using both of the above algorithms.

The second task is to identify the significance of the influence of the chip material — polycarbonate or polypropylene — on the PCR process and the assessment of the position of the threshold cycle in relation to various analyzed samples obtained by diluting the original sample in various proportions. This problem was solved both in the case of applying the first and the second algorithms for searching for a threshold cycle.

The third problem is related to the detection of the presence of a discord in the sequence of measurements, one of the reasons was the presence of bubbles in the reaction chamber for several, mainly 3–7, PCR cycles. Identifying the criteria for the presence of bubbles based on the shape of the kinetic curve is an important task. Its solution enables a conclusion regarding the acceptability of applying the obtained estimates of the location of the threshold cycle and, as a result, evaluating the validity of quantitative analysis findings.

The criteria proposed and discussed below can be classified into two groups. The first group of criteria was based on determining the parameters of the approximating function, the second — on the evaluation of approximation error characteristics.

EXPERIMENTAL CONDITIONS

Microchips from Novattro polycarbonate (SafPlast, Russia) and PP 4445S polypropylene (TU 20.16.51-136-05766801-2015, PJSC Nizhnekamskneftekhim, Russia) were manufactured by thermal pressing in the hydraulic press MM-100 (MTDI, Korea) using a master form made of stainless steel made by the laser method. The structures were sealed with Polymer Films for PCR Plates R-500 (OOO PKF Sovremennye Tekhnologii). The topology of microchips consists of 3 chambers with supply channels (Fig. 1) [1].

Fig. 1. Image of microchips.
Upper — made of polypropylene, below — made of polycarbonate

Real time PCR was performed on microarrays to detect soybean DNA manually isolated using the M-sorb-OOM reagent kit for separating DNA and RNA from clinical samples and environmental objects. The reaction was carried out for the initial concentration of isolated nucleic acids, for 1, 2 and 3 fold dilutions and for a positive reference sample from the "Plant universal" kit, used in three replicates for each variant. When working with this set, R6G dye was used for target detection, Cy5 — for internal positive control. All reagents are provided by Sintol LLC (Moscow). For comparison, similar measurements were carried out in test tubes on ANK-48 (IAP RAS, SPb).

Microarray measurements were carried out using a specially designed setup providing a thermal cycling for real time PCR, excitation at wavelengths of 530, 570 and 685 nm, and registration at 580, 630 and 660 nm, respectively. The detection zone was a circle 3 mm in diameter. The microarray chambers were filled with the reaction mixture with the addition of a DNA sample, the inputs/outputs were sealed with a PCR film. The microchip was placed on the heating element with the film down.

One of the significant drawbacks of PCR in the chip chamber compared to test tubes is the formation of gas bubbles during thermal cycling. It does not significantly affect the course of the reaction, but prevents correct optical detection.

ANALYSIS OF PCR KINETIC CURVES

A determination of the position of the threshold cycle is carried out based on the search for the inflection point of the PCR curve. The approximating dependence of logistic growth of the first order is used:

$$y = \frac{a}{1 + \exp(-k(x - x_0))}. \quad (1)$$

Here a is the maximum signal (saturation signal); k is a coefficient determining the rate of signal increase associated with PCR efficiency; x_0 is the inflection point of the logistic growth curve of the first order. Obviously, for a normalized signal, parameter a should be close to 1.

The significance of the dependence of the position of the threshold cycle on the material of the microarray — polycarbonate (pc) and polypropylene (pp) — is determined by known statistical methods:

1. The generalized Student's t-test provides an estimate of the uniformity of the threshold cycle position for two samples corresponding to different materials.

2. The degree of influence of the material on the position of the threshold cycle is also detected by single-factor analysis of variance.

Sample volumes (numbers of repetitions of experiments) are small (usually 3 repetitions).

The method of one- or multifactorial analysis is well known [2]. However, it is still used, including in various application tasks, for example, see [3, 4]. Since we did not even use two-factor analysis, we used the term "factor analysis" to mean one-factor analysis. The calculation formulas and necessary criteria are discussed in [2].

Sample designations: the polymer material defines the letters pc or pp, the number 1, 2 or 3 sets the dilution degree of the initial sample by 1, 2 or 3 orders, i.e. 10, 100 or 1000 times, respectively (symbol c in place of the number — undiluted sample).

The vpk sample, or БПК, is an internal positive control. The position of the threshold cycle, shifted by more than 7 cycles with respect to the threshold cycle of this sample, indicates the absence of analyzed genetic material. Tab. 1 shows the results of determining the position of the threshold cycle in the internal positive control sample.

Tab. 1. Estimated data on the position of the threshold cycle

The dynamics of threshold cycle estimates in terms of using different sets of measurements of the internal positive control sample signal is illustrated in Tab. 2. The estimate of the position of the threshold cycle when using the entire sample is 28.15 ± 0.04 cycle.

Tab. 2. Estimates of threshold cycle position depending on selection of informative signal points

Almost all threshold cycle position estimates, except those made in points 1–25, have acceptable accu-

racy with a deviation of the average estimate within a few hundredths of a cycle. It is critical that the curve inflection point falls within the signal point range.

The range 1–30 contains only 2 measurements beyond the inflection point, and the evaluation accuracy is lower. In other cases, estimates of the inflection point position are comparable in accuracy. The inflection point was estimated using the approximating first-order logistic growth curve in all cases. For all calculation options, except for the first two, the inflection point of the kinetic curve was an internal point remote from the boundaries of the time interval (cycle 1, cycle 40) by at least 20% of the interval length. For interval 1–35, the position of threshold cycle 28.09 is about 7 cycles from the upper boundary, while the total width of the interval is 34 cycles.

Tab. 3 shows the position of the inflection point of the kinetic curve for the initial sample diluted by 1 order (10 times), in the case of using microfluidic chips made of various polymer materials.

Tab. 3. Kinetic curve inflection point position estimates in the case of using devices made of various polymer materials

The position of the threshold cycle is considered to coincide with the coordinate of the inflection point of the kinetic curve. We searched for the inflection point of the approximating dependence (1). The use of the one-way analysis of variance procedure makes it possible to calculate the characteristic value of the random variable F — the ratio of two variances associated with the discrepancy of average estimates to dispersion of measurements within the sample corresponding to each factor satisfying the Fisher distribution with 1 and 4 degrees of freedom, based on two gradations of the factor (two polymers) and the total number of 6 measurements. This value of F , approximately 0.291, corresponds to a confidence probability of 60%. This means that with such probability, the divergence of sample elements (threshold cycle estimate values) cannot be explained by the influence of the polymer.

Similar results with probabilities of about 50% are also observed for threshold cycle estimates of dilutions by 3 orders of magnitude. (F about 0.534). Tab. 4 shows the positions of the inflection points of the kinetic curves corresponding to other dilutions of

Tab. 4. Positions of inflection points of kinetic curves for other sample dilutions

the initial sample.

The above values also, in accordance with the one-way analysis of variance, indicate the independence of the position of the threshold cycle from the choice of material for a given sample dilution. However, the significance level does not exceed 50%. Comparison of threshold cycle positions for samples diluted by 2 orders of magnitude was not carried out, since bubbles were detected on polypropylene chips.

Figs. 2 and 3 illustrate the distribution of approximation parameters a and k with respect to different

Fig. 2. Distribution of parameter estimates a dependence (1)

PCR-RT kinetic curves.

Fig. 3. Dependence of the distribution of the parameter k of the model (1)

Initially, the signal was normalized and reduced to the range $[0; 1]$. The zero corresponds to the smallest value of the signal, the largest signal value corresponds to 1. Thus, the range limits are determined by the local minima and maxima of the signal without pre-processing it.

FORMATION AND DISCUSSION OF CRITERIA FOR BUBBLE PRESENCE IN REACTION CHAMBER

Obviously, the presence of a bubble affects the shape of the kinetic curve. A change in the shape of a signal or/and a jump in its parameters while maintaining the form of dependence are two types of discord in the sequence of measurements. Discord detection is based either on the estimation of the parameters of the approximating dependence (1), taking into account their physical meaning (role), or on the estimation of the approximation error. In this case, the assumption is made that the actual approximation error is negligible, and after compensation for the approximating deterministic component of the informative signal, the error is random. Then this error is analyzed for the presence/absence of deterministic components (trends).

The proposed and discussed criteria for the absence of bubbles and, as a result, confidence in the results of the analysis are based on the parameters of the deterministic dependence (1) falling into the range corresponding to physically justified values. Simultaneously, the set of criteria for estimating interference parameters is based on statistical estimates of the

sample with no trends.

Currently, the presence of a bubble and its influence on the shape and parameters of the kinetic curve requires the exclusion of the corresponding analysis result as unreliable. In many cases, the degree of distortion of a theoretically valid sample waveform as a sigmoid function is determined solely on the basis of subjective evaluation.

According to expert estimates, measurements 1–17 correspond to the absence of bubbles, 18–27 — the presence of a bubble (-s).

The graph (Fig. 2) highlights the horizontal range of acceptable values of the parameter a at the maximum level of the normalized signal 1. Theoretically, the estimate of this parameter should coincide with the signal maximum, which is also the saturation level.

Bold horizontal lines highlight the area 0.95–1, in which the parameter estimate a can be located at a sufficiently large signal-to-noise ratio.

The signal-to-noise ratio for the considered normalized sample signals, provided that the background signal is correctly compensated and there are no statistical outliers, is within the range of (50–60) to 1. In this case, its average value, approximately equal to 0.5, is considered as the characteristic value of the signal, and noise is taken as the standard deviation of the approximation error.

The first 17 measurements, characterized by the lack of bubbles, completely fall into the specified region. Out of 10 measurements with the presence of bubbles, only 3 values fall into this interval. Similar results will be obtained in the case of increasing the lower limit of the range of allowable values of parameter a to 0.96.

The association between mean PCR E efficacy and the equation parameter k is:

$$E = \exp(k) - 1.$$

Since the theoretically reasonable value of PCR efficiency cannot exceed 1 (or 100%), the expected value of the parameter k in the approximating equation should not exceed 0.693.

In Fig. 3, horizontal lines correspond to theoretically reasonable values of the average PCR efficiency of 80% (lower straight line with a value of 0.588) and 100% (upper straight line with a value of 0.693). Of the 17 measurements in the absence of bubbles, 12 reliably fall into the specified area, 5 are outside it (2 of these 5 measurements are practically positioned on the border of permissible values). Out of 10 measurements with bubbles, there is not a single result in this range (1 practically falls on the boundary, and the remaining 9 are reliably outside the region).

The combination of these limitations reveals experimental data that are suspicious in terms of the presence of bubbles.

In some algorithms, the model parameter (1) x_0 — the abscissa of the inflection point — determines the position of the threshold cycle. Another algorithm for determining the threshold cycle is the search for the abscissa of the tangent intersection point drawn at the inflection point of the PCR curve, with the ordinate axis.

If in approximation (1) parameter a is close to 1, then using the second algorithm, the position of the threshold cycle is shifted relative to x_0 and is approximately equal to $x_0 - 2/k$.

Other methods for finding dependence parameters (1) do not use the entire sample (data for all cycles 1–40 or 1–50 cycles), but only a set of individual elements. For example, earlier in [5], an estimation formula was proposed for the value $C = 1 + E = \exp(k)$. However, it is obvious that the four-point formula proposed in this work uses signal measurements in two pairs of equally spaced samples, conditionally numbered as

$$n, n + k, n + m \text{ и } n + m + k,$$

certainly gives a coarser estimate than the approximating dependence, which uses the entire sample of measurements, typically 40–50.

Also, the estimation of the position of the inflection point of the curve can be performed by the method of least squares when constructing a polynomial of the 3rd degree, analytically calculating the second-order derivative, equating it to zero, and finding the root by solving a linear equation. This technique was used in the search for the inflection point of the DNA melting curve [6]. The expected pattern is confirmed: more accurate estimates are obtained by using measurements from the linear section of the curve with normalized signal values close to 0.5. It is better to take not exactly 4 points, and the sample size should have several degrees of freedom (at least 6–8 points).

For example, for the kinetic curve, the inflection point estimate is 25.55 (see Tab. 3, line 2). Polynomial approximation with a cubic parabola over points 22–27 gives an inflection point estimate of 25.79, which is acceptable for a rough calculation. Here are typical dependences of normalized sample signals and

Fig. 4. Normalized sample and scattering signals in the absence of bubbles (from group pc_1)

Fig. 5. Normalized sample and scattering signals in the presence of a bubble at the final stage (pp_1)

scattering in the absence of bubbles (Fig. 4) and in the presence of one bubble (Fig. 5).

The curve corresponding to the normalized scattering signal (see Figs. 4 and 5) can be adequately approximated by a second degree polynomial.

As mentioned above, the position of the threshold cycle can be determined using two algorithms. For the initial sample, diluted 10 times (by 1 order), Tab. 5 shows the estimates of the position of the threshold cycle when analyzing chips made of various polymer materials..

Tab. 5 reveals that estimating the position of the threshold cycle with plotting a tangent to the kinetic curve at the inflection point based on one-way analysis results in a reduced variance ratio F of about 11.65. Conclusion: with a probability of 97%, the difference in the positions of the threshold cycle in the case of using the intersection point of the tangent to the curve inflection and the ordinate axis is explained precisely by the difference in material (polycarbonate or polypropylene). The conclusion is the opposite for the algorithm of finding the threshold cycle directly at the inflection point: the material selection factor is unlikely to affect the position of the threshold cycle.

Another group of algorithms for detecting the presence of a bubble can be based on the statistical criterion of ascending and descending series of errors of approximation of sample signals or scattering when a high certainty of determination is achieved (the determination coefficient must obviously exceed 0.9). One of these criteria is 5% significant and requires satisfaction of two conditions with the number of ascending and descending series. With the total number of points (approximation errors) equal to N :

$$\begin{cases} ser(N) > \left[\frac{2N-1}{3} - 1.96\sqrt{\frac{16N-29}{90}} \right], \\ \tau_{\max}(N) < \tau^*(N), \end{cases} \quad (2)$$

$$\tau^*(N) = \begin{cases} 5 & N \leq 26, \\ 6 & 26 < N \leq 153, \\ 7 & 153 \leq N \leq 1170. \end{cases} \quad (3)$$

Formula (2) is based on data [7, pp. 486–488], and the limitations (3) are discussed in [8]. In relation to the estimates of the informative signal (sample signal) in Figs. 4 and 5, the volume of the initial samples was 40 points. Since it was necessary to plot the right first-order difference to estimate the presence of ascending and descending series, the sample value of the difference error N is 39. Accordingly, the run-up series contains a sequence of positive difference error values, while the runs-down contain negative ones. According to the first condition (2), the number of runs $ser(N)$ should be more than 20, according to the second condition (3), the largest run length should be strictly less

than 6, i.e. not more than 5.

The following trends are observed: the curve for the sample in Fig. 4 corresponds to the number of runs of 23 and a maximum run length of 4. Thus, both conditions (2) and (3) are fulfilled. Applicable to sample curve Fig. 5: number of runs — 19, maximum length of runs — 5. Condition (2) is violated, condi-

Tab. 5. Calculated data on the position of the threshold cycle depending on the method of its search

tion (3) is fulfilled.

Similar data is on scattering curves after compensation of deterministic component — polynomial of the second degree in relation to Fig. 4: number of runs — 29, maximum length of run — 3; for Fig. 5: number of runs — 19, maximum length of run — 6.

In the first case (Fig. 4) both conditions are also met, and this is the case of no bubbles, but in the second case (Fig. 5) both conditions are violated.

The answer to the question of how universal these criteria for the presence of bubbles are can be the subject of further research based on more representative statistics. Another criterion can be associated with the analysis of the distribution of ascending and descending series (phases) by length and with discrepancy with theoretical estimates of the mathematical expectation of the number of series of a given length in the absence of trends in the sample.

The expectation of the number of series of length d ($d = 1, 2, 3, \dots$) [7] is represented by the following formula (4):

$$N_d = \frac{2(N-d-2)(d^2+3d+1)}{(d+3)!}. \quad (4)$$

For the number of samples 39, the mathematical expectations of the number of ascending and descending series of lengths 1, 2, ..., 5 are 15.00, 6.42, 1.79, 0.38 and 0.07, respectively.

The proximity of the actual distribution of series to the theoretically reasonable one is estimated using Pearson's chi-squared test (or χ^2) with the following amendments and assumptions: a) only three values are considered — the number of series of length 1, length 2 and length 3 or more; b) the found value is assigned a degree of freedom 2.5 if it is equal to 6.3 or more, or 2 degrees of freedom for smaller values of the discrepancy characteristic. In the second case, a reducing correction factor of 6/7 is introduced. This classical technique is formulated in work [9]. The features of

Tab. 6. Distribution of ascending and descending phases by lengths for various features of PCR-RT

the phase distributions by length for the presence or

absence of bubbles are illustrated in Tab. 6.

Accordingly, the percentage point of the χ^2 distribution with 2 degrees of freedom and its percentage points [10] are considered.

If the probability level of agreement with the hypothesis is less than 10%, then it should be rejected. Therefore, the hypothesis about the absence of trends (discords) after compensation of the approximating deterministic component of the signal in the presence of bubbles (data group Fig. 5) should be rejected. This hypothesis in relation to the informative signal of the sample in the absence of bubbles is taken with a high level of significance (more than 40%), which is sufficient to accept the hypothesis about the absence of trends in the sequence of approximation errors and, as a result, the absence of a discord in the sequence of initial measurements.

Therefore, the correct conclusions can be drawn with a high probability based on the proposed algorithms for making decisions about the presence of bubbles, both related to the analysis of the informative signal and the approximation error. However, to what extent the obtained results are specific in relation to the chosen object of analysis and to what extent they can be generalized to other objects requires additional research.

CONCLUSION

The quantitative genetic analysis runs based on real-time PCR using polymer chips revealed the following patterns: when defining the threshold cycle as the inflection point of the kinetic curve (first-order logistic curve), there is no significant dependence of these values on the chip material (polycarbonate or polypropylene). If we estimate the position of the threshold cycle by the abscissa of the point of intersection of the tangent to the kinetic curve drawn at its inflection point, with a background level (minimum) signal, with a confidence probability of more than 90%, the polymer material factor affects the position of the threshold cycle.

An informative signal (sample signal) in the absence of bubbles with high accuracy is approximated by a logistic curve of the first order — one of the types of sigmoid curve. The determination factor exceeds 0.99. Also, when normalizing the sample signal under these conditions, the signal-to-noise ratio reaches values of 55–60. The scale (characteristic value) of the signal is the mean value, the noise scale is the root-mean-square deviation of the approximation error.

Checking the criteria for the presence of bubbles for the normalized kinetic curve:

a) coefficient a in approximating dependence is

significantly lower than 1;

b) non-physical value of coefficient k is reliably exceeding 0.70 or significantly less than 0.58;

c) based on the analysis of ascending and descending series in the sequence of measurements of approximation errors, there is a failure to fulfill one or two conditions for the absence of trends, as well as a significant deviation of the lengths of ascending and descending series (phases) from theoretically justified ones in the absence of distribution trends.

The combination of the proposed criteria after additional verification of their versatility (the correctness of generalization to other biological objects or/and other chip materials) can be included in the software and mathematical support of the devices, which in the future will exclude the participation of the subjective factor (human expert) in deciding on the presence of bubbles in the reaction chamber at any stage of kinetic curve construction.

The study was carried out within the framework of the State Assignment of the Ministry of Science and Higher Education of the Russian Federation No. 075-00761-22-00 (subject "Microfluid devices with integrated functional micro- and nanoscale structures for biological and medical research", FFZM-2022-0012).

REFERENCES

1. Esikova N.A., Germash N.N., Evstrapov A.A. [Rapid fabrication of microchips for PCR analysis from polymer materials in the laboratory conditions]. *Nauchnoe Priboroostroenie* [Scientific Instrumentation], 2020, vol. 30, no. 4, pp. 21–26. DOI: 10.18358/np-30-4-i2126 (In Russ.).
2. Gmurman V.E. *Teoriya veroyatnostei i matematicheskaya statistika: uchebnoe posobie dlya inzhenerno-ekonomicheskikh institutov i fakul'tetov. Izd. 4-e, dop* [Probability theory and mathematical statistics: a textbook for engineering and economic institutes and faculties. 4th Suppl. Edit.]. Moscow, Vysshaya shkola Publ., 1972. 367 p. (In Russ.).
3. Tikhomirov N.P., Tikhomirova T.M., Ushmaev O.S. *Metody ehkonometriki i mnogomernogo statisticheskogo analiza: uchebnyk* [Econometrics and multivariate statistical analysis methods: textbook]. Moscow, Ekonomika Publ., 2011. 637 p. (In Russ.).
4. Plokhonikov K.Eh., Kolkov S.V. *Statistika: uchebnoe posobie* [Statistics: educational assistance]. Moscow, Flinta Publ., 2006. 286 p. (In Russ.).
5. Bulyanitsa A.L. [Methods for logistic growth curve parameters estimation. Part. 1. Optimization of estimation conditions at the presence of additive random error]. *Nauchnoe Priboroostroenie* [Scientific Instrumentation], 2009, vol. 19, no. 3, pp. 3–11. URL: <http://iairas.ru/mag/2009/abst3.php#abst1> (In Russ.).
6. Belov D.A., Belov Yu.V., Kurochkin V.E. [New method of DNA melting signal treatment]. *Nauchnoe Priboroostroenie* [Scientific Instrumentation], 2018, vol. 28, no. 1,

- pp. 3–10. DOI: 10.18358/np-28-1-i310 (In Russ.).
7. Kendall M.Dzh., St'yuart A. *Mnogomernyi statisticheskii analiz i vremennye ryady* [Multivariate statistical analysis and time series]. Moscow, Nauka Publ., 1976. 736 p. (In Russ.).
8. Onlain kal'kulyator. *Kriterii "voskhodyashchikh" i "niskhodyashchikh" serii* [Online calculator. "Ascending" and "descending" batches criterion]. URL: <https://math.sememstr.ru/trend/series.php> (accessed: 19.04.2022) (In Russ.).
9. Wallis W.A., Moore G.H. A significance test for time series analysis. *J. Amer. Statist. Ass.*, 1941, vol. 36, is. 215, pp. 401–409. DOI: 10.1080/01621459.1941.10500577
10. Bol'shev L.N., Smirnov N.V. *Tablitsy matematicheskoi statistiki* [Tables of mathematical statistics]. Moscow: Nauka Publ., 1983. 416 p. (In Russ.).

Contacts: Bulyanitsa Anton Leonidovich,
antbulyan@yandex.ru

Article received by the editorial office on 29.04.2022