



УДК 517.21 + 541.69

© 1993 г. Е. Д. Свердлов, О. Д. Ермолаева

ВЫЧИТАЮЩАЯ ГИБРИДИЗАЦИЯ. ТЕОРЕТИЧЕСКИЙ АНАЛИЗ И ПРИНЦИП ЛОВУШКИ

Институт биоорганической химии им. М. М. Шемякина и Ю. А. Овчинникова РАН, Москва

Ключевые слова: вычитающая гибридизация, обогащение.

Проведен теоретический анализ кинетик гибридизации и степеней обогащения последовательностей, отличающих два гомологичных генома или транскрипта.

Показано, что при гибридизации одно- или двухцепочечной ДНК (трейсер), содержащей эту последовательность (мишень), с большим избытком одноцепочечной комплементарной ДНК драйвера, не содержащей мишени, получают значительно более высокие степени обогащения мишени, чем при гибридизации двухцепочечного драйвера и трейсера или одноцепочечного трейсера с двухцепочечным драйвером. Следует сделать вывод, что первый подход может дать путь к эффективной экстракции мишени без искусственного упрощения геномов перед вычитанием. Предложен принцип ловушки, предотвращающий гибриды, образующиеся при первом способе, от диссоциации в процессе РСР-амплификации трейсера, обогащенного в результате вычитания. Сочетание гибридизации одноцепочечных драйвера и трейсера с использованием принципа ловушки может стать основой высокоэффективной экстракции последовательностей, приводящих к фенотипическим различиям.

Вычитающая гибридизация занимает все более важное место в ряду методов, используемых для выделения последовательностей, отличающих один геном от другого, тесно с ним связанного, или мРНК, присутствующих в одном, но отсутствующих в другом типе клеток (для краткого обзора см. [1]).

Хотя теория реассоциации ДНК или гибридизации ДНК с комплементарной РНК достаточно хорошо разработана [2, 3], ее применение к случаю вычитающей гибридизации имеет некоторую специфику. Эта специфика практически не учитывается исследователями, использующими вычитающую гибридизацию для создания обогащенных геномных или кДНК-библиотек. Интуитивные подходы дают худшие результаты, чем можно было бы получить при корректном выборе стратегии, и толкают авторов на усложнения методики, которые иногда изящны, но вполне излишни.

В связи с этим кажется целесообразным проанализировать несколько кинетических схем вычитающей гибридизации, с тем чтобы дать исследователям готовые для использования формулы.

1. Определения

мишень (target) — последовательность, которая должна быть выделена в процессе гибридизации;

трейсер (tracer) — ДНК, содержащая эту последовательность и используемая в недостатке при вычитающей гибридизации;

драйвер (driver) — ДНК, идентичная трейсеру, за исключением содержания мишени. Берется в большом избытке;

C_0^d и C_0^t , М — концентрация фрагмента одноцепочечного драйвера, представленного в ДНК в виде единственной копии, в начале реассоциации и через t с после начала;

C_0^i и C_0^j , М — соответствующие концентрации гомологичного участка (фрагмента) трейсера ($C_0^i \ll C_0^d$);

U_0 и U_i , М — соответствующие концентрации мишени в одноцепочечном состоянии, представленной в геноме в виде одной копии ($U_0 = C_0^i$);

h_i — концентрация образовавшегося к моменту t двухцепочечного фрагмента мишени ($h_i = U_0 - U_i$);

e_i — концентрации образовавшихся к этому моменту двухцепочечных фрагментов трейсера;

$E_s = U_i/C_i^t$ — обогащение мишени в одноцепочечной фракции;

$E_d = h_i/e_i$ — обогащение мишени в двухцепочечной фракции;

R^{**} , $M^{-1} \cdot c^{-1}$ — константа скорости реассоциации цепей. $R \cong 10^6 M^{-1} \cdot c^{-1}$ для фрагментов длиной ≈ 500 звеньев в 0,18 М NaCl при оптимальной температуре.

2. Обогащение при вычитающей гибридизации двухцепочечных драйвера и трейсера *** (метод 1)

Смешаем ДНК драйвера и трейсера так, что $C_0^d \gg C_0^t$, денатурируем и позволим реассоциировать при температуре $T_{\text{онт}} \approx T_{\text{пл}} - 25^\circ C$.

В результате будут идти три процесса:

1) реассоциация мишени с образованием двухцепочечного фрагмента в концентрации h , скорость которой определяется уравнением 1;

2) реассоциация трейсера. Цепи трейсера реассоциируют сами с собой и с цепями драйвера (уравнение 2);

3) реассоциация драйвера, цепи которого реассоциируют сами с собой и с цепями трейсера (уравнение 3).

В результате основное количество трейсера в продуктах реакции будет содержаться в виде двухцепочечного гибрида драйвера и трейсера, а двухцепочечный трейсер будет образовываться за счет реассоциации его цепей (уравнение 4).

$$\frac{dU_i}{dt} = -RU_i^2 \quad (1);$$

$$\frac{dC_i^t}{dt} = -R \cdot (C_i^t)^2 - R \cdot C_i^d \cdot C_i^t \quad (2);$$

* Здесь все концентрации соответствуют концентрациям реассоциирующих цепей, а не мононуклеотидов — компонентов этих цепей. Концентрация мононуклеотидов принята в теоретических расчетах реассоциации [2, 3]. Она может быть получена умножением принятых здесь концентраций на длину ДНК драйвера или трейсера в числе звеньев.

** R не зависит от сложности реассоциирующих молекул ДНК и используется в сочетании с концентрациями цепей. В теории кинетики реассоциации в формулах обычно используют k , $M^{-1} \cdot c^{-1}$ в сочетании с концентрациями мононуклеотидов: $k = R/G$, где G — длины реассоциирующих молекул ДНК, равные $3 \cdot 10^9$ для млекопитающих, $5 \cdot 10^7$ для дрожжей, $5 \cdot 10^6$ для *E. coli*, $(3-5) \cdot 10^4$ для бактериофагов T7 и λ и т. д. [2, 3].

*** ДНК представлены их фрагментами, полученными путем расщепления рестриктными эндонуклеазами при геномном вычитании или кДНК-копиями при вычитании кДНК.

$$\frac{dC_t^d}{dt} = -R \cdot (C_t^d)^2 - R \cdot C_t^d \cdot C_t^i \quad (3); \quad \frac{de_t}{dt} = R (C_t^i)^2 \quad (4).$$

Поскольку $C_0^d \gg C_0^i$ и $U_0 = C_0^i$, уравнения (2) и (3) могут быть упрощены:

$$\frac{dC_t^i}{dt} = -R \cdot C_t^d \cdot C_t^i \quad (2'); \quad \frac{dC_t^d}{dt} = -R \cdot (C_t^d)^2 \quad (3').$$

Решение этих уравнений дает зависимости концентраций от времени:

$$\frac{C_t^d}{C_0^d} = \frac{1}{1 + RC_0^d t} \quad (5); \quad \frac{C_t^i}{C_0^i} = \frac{1}{1 + RC_0^d t} \quad (6);$$

$$\frac{U_t}{U_0} = \frac{1}{1 + RU_0 t} \quad (7); \quad e_t = \frac{R (C_0^i)^2 t}{1 + RC_0^d t} \quad (8);$$

$$h = \frac{R (U_0)^2 t}{1 + RU_0 t} \quad (9); \quad E_s \approx E_d = \frac{h_t}{e_t} = \frac{1 + RC_0^d t}{1 + RU_0 t} \quad (10).$$

При больших временах ($RC_0^d t$ и $RU_0 t \gg 1$) обогащение равно отношению исходных концентраций драйвера и трейсера.

3. Обогащение при вычитающей гибридизации двухцепочечного драйвера и одноцепочечного трейсера $C_0^d \gg C_0^i$ (метод 2)

В этом случае $\frac{dU_t}{dt} = 0$, остальные уравнения соответствуют формулам (2') и (3'), их решения — формулам (5), (6). Обогащение происходит только в одноцепочечной фракции и дается формулой

$$E_s = \frac{U_0}{C_t^i} = 1 + RC_0^d t \quad (11).$$

4. Обогащение при вычитающей гибридизации одноцепочечного драйвера и комплементарного ему одноцепочечного трейсера $C_0^d \gg C_0^i = U_0$ (метод 3)

В этом случае

$$\frac{dU_t}{dt} = 0; \quad \frac{dC_t^d}{dt} = 0; \quad \frac{dC_t^i}{dt} = -(RC_t^d) C_t^i \quad (12).$$

Поскольку драйвер в большом избытке, уравнение (12) может рассматриваться как уравнение реакции псевдопервого порядка, т. е. RC_t^d может рассматриваться постоянным на протяжении реакции. Тогда

$$\frac{C_t^i}{C_0^i} = e^{-RC_0^d t} \quad (13)$$

и обогащение в одноцепочечной фракции

$$E_s = \frac{U_0}{C_t^i} = e^{RC_0^d t} \quad (14).$$

5. Обогащение при вычитающей гибридизации одноцепочечного драйвера с двухцепочечным трейсером $C_0^d \gg C_0^t = U_0$ (метод 4)

В этом случае обогащение в одноцепочечной фракции E_s определяется формулой (15), а обогащение в двухцепочечной фракции — формулой (16).

$$E_s = \frac{e^{RC_0^d t}}{1 + RU_0 t} \quad (15); \quad E_d = \frac{2 \cdot RC_0^d t}{(1 + RU_0 t)(1 - e^{-2RC_0^d t})} \quad (16).$$

6. Сопоставление обогащений при четырех вариантах вычитающей гибридизации с использованием реальных концентраций драйвера и трейсера

Соотношения драйвера и трейсера взяты из статьи Лисицына и соавт. [4]:

$$C_0^d = 40 \text{ мкг в 4 мкл} = 1,11 \cdot 10^{-11} \text{ М,}$$

$$C_0^t = U_0 = 0,5 \text{ мкг в 4 мкл} = 1,39 \cdot 10^{-13} \text{ М}$$

(принято: длина ДНК — $3 \cdot 10^9$ нуклеотидов и молекулярный вес нуклеотидного звена ≈ 300).

Для условий Лисицына и др. [4], упростивших ДНК примерно на порядок перед гибридизацией, эти цифры равны $1,11 \cdot 10^{-10}$ и $1,39 \cdot 10^{-12}$ соответственно.

Сопоставления обогащений, достигаемых в разные интервалы времени ренатурации, для рассмотренных выше трех случаев даны в табл. 1.

Приведенные цифры демонстрируют значительный выигрыш в обогащении, достигаемый при использовании метода 3 в вычитающей гибридизации.

При вычитающей гибридизации двухцепочечного трейсера с одноцепочечным драйвером обогащение в одноцепочечной фракции близко к величинам, получаемым при вычитании одноцепочечных драйвера и трейсера. В двухцепочечной фракции обогащение несколько выше (при больших временах — в 2 раза), чем при вычитании двухцепочечных молекул. Это легко видеть при анализе формул (15) и (16).

За 20 ч, которые использовали авторы работы [4] для вычитающей гибридизации, достигается обогащение всего в 8 раз, тогда как при использовании третьей схемы могло быть достигнуто более чем 2000-кратное обогащение.

Эффект может быть увеличен далее за счет увеличения времени гибридизации. Приводимые цифры представляют собой оценки. Но качественный эффект, позволяющий сравнивать порядки величин, сохраняется.

Отдельной проблемой, возникающей при вычитании кДНК, является неравнопредставленность отдельных генов или мРНК. Например, Галау с соавт. [5] мРНК из печени мыши подразделяет на три фракции: высоко-, средне- и малопредставленных мРНК (табл. 2). (Среднюю длину мРНК принимают равной 2000 звеньям.)

В данном случае обогащение целесообразно определить как отношение суммарной концентрации всех мРНК, точнее кДНК трейсера, перед началом отжига к их суммарной концентрации через время t :

$$E_s = \frac{\sum_i C_{0i}^t}{\sum_i C_{0i}^d}.$$

Для частного случая, приведенного в табл. 2,

$$\sum_i C_{0i}^t = 6400 \cdot C_{03}^t \cdot e^{-RC_{03}^d t} + \frac{250}{21} \cdot 330 \cdot C_{03}^t \cdot e^{-\frac{250}{21} RC_{03}^d t} +$$

Таблица 1

Сопоставление обогащений при разных методах вычитающей гибридизации в зависимости от времени

Время, с	Обогащение		
	метод 1 * ($E_d \approx E_s$)	метод 2 (E_s)	метод 3 (E_s)
1000	1,10	1,11	1,12
10 000	2,08	2,11	3,04
20 000	3,13	5,22	9,231
40 000	5,15	5,44	85,11
50 000	6,12	6,55	258,26
60 000	7,07	7,66	785,25
72 000 ** (20 ч)	8,17	8,99	2957,21
100 000 (28 ч)	10,6	12,1	66177,16

* Этот метод использован Лисицыным и соавт. [4].

** Это время использовано в работе [4].

Таблица 2

Различные классы цитоплазматической poly(A)⁺-мРНК в печени мыши [5]

Тип мРНК	Доля мРНК каждого класса	Число последовательностей каждого класса на клетку	Число молекул каждой последовательности на клетку
Высокопредставленные (C_{11})	0,29	12	7200
Среднепредставленные (C_{12})	0,27	330	250
Низкопредставленные (C_{13})	0,44	6400	21

$$+ \frac{7200}{21} \cdot 12 \cdot C_{03}^i \cdot e^{-\frac{7200}{21} RC_{03}^d t} \quad (17).$$

При этом предполагается, что в драйвере соотношения всех мРНК, кроме мРНК-мишени, сохраняются такими же, как в трейсере.

Через достаточно продолжительное время ($RC_{03}^d t > 0,25$) вторым и третьим членом выражения (17) можно пренебречь. Тогда обогащение выразится простой формулой (17''):

$$\sum_i C_{1i}^i \approx 6400 \cdot C_{03}^i \cdot e^{-RC_{03}^d t} \quad (17'); \quad E_s = \frac{\sum_i C_{0i}^i}{6400 \cdot C_{03}^i \cdot e^{-RC_{03}^d t}} \quad (17'').$$

Если концентрации драйвера и трейсера взяты как для ДНК (см. выше), то

$$C_{03}^d = \frac{0,44}{6400} \sum_i C_{0i}^d = \frac{0,44}{6400} \cdot \frac{10}{300 \cdot 2000} = 1,15 \cdot 10^{-9} \text{ моль цепей/л.}$$

Зависимость обогащения от времени при реассоциации одноцепочечных кДНК
 $(RC_{03}^1 = 10^6 \cdot 1,15 \cdot 10^{-9} = 1,15 \cdot 10^{-3})$

Время, с	10 000	20 000	50 000	100 000
Обогащение	7,17	22,6	717,2	223 656

Концентрация соответствующей фракции трейсера в 80 раз меньше ($C_{03}^1 = 1,44 \cdot 10^{-11}$) и суммарная концентрация трейсера

$$\sum_i C_{0i}^1 = 2,08 \cdot 10^{-7} \text{ моль цепей/л.}$$

Рассчитанные значения обогащения в зависимости от времени см. в табл. 3.

Цифры, приведенные в табл. 1 и 3, показывают, что выбор времени реассоциации чрезвычайно важен для степени обогащения — незначительное, с точки зрения экспериментатора, изменение времени может оказать катастрофическое влияние на результат.

7. Некоторые выводы и принцип ловушки (trapper)

Из приведенных сопоставлений однозначно следует подавляющее кинетическое преимущество гибридизации одноцепочечных цепей драйвера для вычитания как геномов, так и кДНК. Полученные недавно экспериментальные данные по вычитанию одноцепочечных кДНК-библиотек подтверждают этот вывод [6, 7].

Более того, данные табл. 1 позволяют считать, что при таком способе можно обойтись без упрощения сравниваемых геномов, использованного Лисицыным и соавт. [4] для достижения таких же степеней обогащения. Упрощение вносит элемент неопределенности в вычитание и оставляет значительные шансы пропустить искомое различие.

Действительно, за 20 ч обогащение равно 8,17 (табл. 1) с упрощенными примерно на порядок геномами. Приблизительно такой же степени обогащения можно достичь с одноцепочечными геномами без уменьшения сложности за ≈ 50 ч. Но при этом уменьшается риск потери требуемых последовательностей.

Тем не менее подобные степени обогащения неудовлетворительны. Поэтому приходится прибегать к дополнительным приемам для их увеличения. Одним из наиболее эффективных принципов, используемых в последнее время, является специфическая PCR *-амплификация трейсера после его обогащения и повторные вычитания с обогащенным продуктом. Именно поэтому большинство авторов предпочитают метод 1 вычитающей гибридизации (см., например, [4, 8—12]). Образующийся двухцепочечный продукт может быть непосредственно амплифицирован, если к нему присоединяют специфические линкеры.

Тем не менее специфической амплификации можно подвергнуть и одноцепочечный обогащенный трейсер. Для этого его необходимо дифференцировать от образовавшихся в результате взаимодействия с драйвером гибридов.

Одним из основных способов, который используется также и для рутинного вычитания по методу 1, является физическое отделение гибридов от обогащенной одноцепочечной фракции с помощью, например, биотинилированного драйвера и стрептавидина. Стрептавидин после гибридизации будет взаимодействовать с гибридами и с избытком драйвера, но не с трейсером. Однако при этом не очевидна степень очистки.

Мы предлагаем для дифференциации незагибризовавшегося трейсера от

*PCR — полимеразная цепная реакция.

трейсера, вошедшего в процессе гибридизации в состав гибрида, использовать способ, который может быть назван способом ловушки.

Принцип способа заключается в создании таких условий, при которых загибридизовавшийся трейсер либо теряет способность к диссоциации из образовавшегося дуплекса в условиях PCR-амплификации, либо разрушается перед амплификацией.

Недавно один из возможных вариантов такого подхода был использован при вычитающей гибридизации кДНК с мРНК [13]. В этом случае кДНК в гибриде была химически «сшита» с мРНК драйвера. Другим способом может стать введение в драйвер аналогов оснований, увеличивающих стабильность дуплекса — например, 2-аминодезоксиаденозина или 5-метилдезоксицитидина [14, 15]. При этом образующийся гибрид драйвера с трейсером имел бы более высокую температуру плавления по сравнению с продуктом превращения одноцепочечного трейсера в двухцепочечный перед амплификацией. Повышение температуры плавления гибридов предотвратило бы их участие в амплификации, делая амплификацию специфической для трейсера. Эффект предотвращения амплификации при введении в ДНК 5-метилдезоксицитидиновых звеньев наблюдался [16].

Наконец, весьма перспективным может оказаться разрушение трейсера в составе гибрида с помощью эндонуклеазы III *E. coli* или другого фермента, которые специфически гидролизуют двухцепочечные ДНК [17].

Возможны и другие варианты осуществления принципа ловушки. В сочетании с вычитающей гибридизацией с одноцепочечным драйвером, который в этом случае играет роль ловушки, этот принцип способен дать очень высокие степени обогащения без издержек, связанных с упрощениями генома.

СПИСОК ЛИТЕРАТУРЫ

1. Myers R. M. // Science. 1993. V. 259. P. 942—943.
2. Britten R. J., Kohne D. E. // Science. 1968. V. 161. P. 529—540.
3. Britten R. J., Davidson E. H. // Nucleic Acid Hybridization. A Practical Approach/Eds Hames B. D., Higgins S. J. Oxford — Washington D. C.: IRL Press, 1985. P. 3—15.
4. Lisitsyn N., Lisitsyn N., Wigler M. // Science. 1993. V. 259. P. 946—951.
5. Galau G. A., Klein W. H., Britten R. J., Davidson E. H. // Arch. Biochem. and Biophys. 1977. V. 179. P. 584—599.
6. Rubenstein J. L. R., Brice A. E. J., Ciaranello R. O., Denney D., Porteus M. H., Usdin T. B. // Nucl. Acids. Res. 1990. V. 18. P. 4833—4842.
7. Swaroop A., Xu J., Agarwal N., Weissman S. M. // Nucl. Acids. Res. 1991. V. 19. P. 1954.
8. Straus D., Ausubel F. M. // Proc. Nat. Acad. Sci. USA. 1990. V. 87. P. 1889—1893.
9. Wieland I., Bolger G., Asouline G., Wigler M. // Proc. Nat. Acad. Sci. USA. 1990. V. 87. P. 2720—2724.
10. Лисицын Н. А., Розенберг М. В., Лайнер Г. А., Вагнер Л. Л., Потапов В. К., Колесник Т. В., Сverdlov E. D. // Молекулярная генетика, микробиология и вирусология. 1993. N 3. С. 26—29.
11. Zarskiy A. G., Lukyanov S. A., Vasiliev O. L., Smirnov Y. V., Belyavsky A. V., Kazanskaya O. V. // Develop. Biol. 1992. V. 152. P. 373—382.
12. Timblin C., Battey J., Kuehl W. M. // Nucl. Acids Res. 1990. V. 18. P. 1587—1593.
13. Hampson I. N., Pope L., Cowling G. J., Dexter T. M. // Nucl. Acids Res. 1992. V. 20. P. 2899.
14. Cheong C., Tinoco I., Jr., Chollet A. // Nucl. Acids Res. 1988. V. 16. P. 5115—5122.
15. Gill J. E., Mazrimas J. A., Bishop C. C., Jr. // Biochim. et biophys. acta. 1974. V. 335. P. 330—348.
16. Wong K. K., McClelland M. // Nucl. Acids Res. 1991. V. 19. P. 1081—1085.
17. Linn S. M. // Nucleases/Eds Linn S. M., Roberts R. J. Cold. Spring Harb. Lab., 1982. P. 291—309.

Поступила в редакцию
23.VII.1993

E. D. Sverdlov, O. D. Ermolayeva

**SUBTRACTIVE HYBRIDIZATION. THEORETICAL ANALYSIS
AND A PRINCIPLE OF THE «TRAPPER»**

*M. M. Shemyakin and Yu. A. Ovchinnikov Institute of Bioorganic Chemistry,
Russian Academy of Sciences, Moscow*

An analysis of the kinetics of the subtractive hybridization and the extent of enrichment of sequence differences between two otherwise identical genomes or transcripts has been carried out for different conditions. The hybridization of single- or double-stranded DNA (tracer) containing the difference (target) to the large excess of single-stranded complementary DNA not having the sequence (driver) gives a much higher enrichment of the target than hybridization of double-stranded driver and tracer or double-stranded driver with single-stranded tracer. The conclusion is that the first approach can give means for efficient extraction of the target without artificial simplifications of the genomes before subtraction. A principle of the «trapper», preventing the hybrids formed from dissociation in the course of PCR amplification of the target enriched, is proposed as a complement to the single-stranded DNA hybridization in the process of highly efficient subtraction.