



Структура и функция биополимеров

УДК 577.324.4

ТЕОРЕТИЧЕСКИЙ АНАЛИЗ АДРЕСОВАННОЙ МОДИФИКАЦИИ ДНК *

М. П. Перельройзен, А. В. Вологодский

Введение. Одной из важных задач, возникающих при изучении первичных структур больших геномов, является задача специфического разрезания ДНК на фрагменты, состоящие из 10^4 — 10^5 пар оснований. Эту задачу в настоящее время трудно решить с помощью рестриктаз, поскольку специфические для них участки расположены в среднем на меньшем расстоянии друг от друга. Одним из возможных подходов к этой проблеме является адресованная модификация однонитевой ДНК [1]. Под этим термином подразумевается, что с однонитевой ДНК-мишенью специфически связывается олигонуклеотид-адрес, имеющий определенную последовательность оснований, к концу которого ковалентно пришит химический реагент, способный модифицировать основания ДНК. Затем по местам такой модификации может быть произведено расщепление ДНК на фрагменты.

Однако вследствие большой гибкости олигонуклеотидной цепи наряду с полностью специфическими комплексами в такой системе возможно связывание адреса и с неполностью комплементарными ему участками мишени. Поскольку число последних в случайной последовательности (которую естественно использовать при анализе возможностей метода) значительно превосходит число полностью комплементарных, возникает вопрос о принципиальной применимости такого подхода к специфическому расщеплению ДНК. В последнее время, главным образом благодаря работам Тиноко с сотр. [2], появились термодинамические характеристики различных несовершенных олигонуклеотидных дуплексов ДНК. Эти данные позволяют провести количественный анализ возможностей специфической модификации ДНК на основе ее связывания с олигонуклеотидными адресами. Такой анализ и является целью настоящей работы.

Формулировка модели. Рассмотрим отрезок ДНК, расположенный между двумя полностью комплементарными олигонуклеотидному адресу участками. Будем считать, что между этими участками нет ни одного, полностью комплементарного адресу, и что длина всего отрезка составляет L нуклеотидных звеньев. Проведем анализ возможности специфической модификации концов этого отрезка при условии, что внутри него не произойдет ни одной неспецифической модификации. Примем, что модификация оснований возможна лишь при связывании олигонуклеотидного адреса с мишенью. Поскольку процессы специфической и неспецифической модификации являются независимыми, вероятность $P(t)$ реализации к моменту времени рассматриваемого события можно представить в виде

$$P(t) = p_s^2(t) \prod_i (1 - p_i(t)), \quad (1)$$

* Представлена членом редколлегии В. И. Ивановым.

где $p_s(t)$ — вероятность специфической модификации определенного конца отрезка к моменту времени t ; $p_i(t)$ — аналогичные вероятности для неспецифических модификаций внутри рассматриваемого отрезка. Произведение берется по всем возможным местам неспецифического связывания.

Специфичность модификации будет зависеть от отношения времени жизни олигонуклеотида в связанном состоянии τ_0 и характерного времени τ_m модификации ДНК реагентом, находящимся в таком комплексе. Ясно, что при

$$\tau_m \ll \tau_0 \quad (2)$$

специфичность модификации будет определяться отношением констант скоростей специфического и неспецифического связывания, поскольку в этом случае первый же акт связывания адреса с матрицей приведет к химической модификации. Как показано в работах [3—5], константа скорости образования дуплекса комплементарными олигонуклеотидами практически не зависит от его длины. Она определяется главным образом временем подхода комплементарных олигонуклеотидов друг к другу. Поэтому константы скоростей специфического и неспецифического связывания адреса также не должны существенно отличаться, и в рассматриваемом предельном случае мы не можем рассчитывать на высокую специфичность модификации.

В обратном случае, т. е. когда

$$\tau_m \gg \tau_0, \quad (3)$$

одному акту химической модификации должно предшествовать много актов образования и развала комплексов. Поэтому специфичность химической модификации будет определяться суммарным временем пребывания в связанном состоянии в специфических и неспецифических комплексах. Иными словами, в этом случае скорость модификации будет пропорциональна равновесным степеням заполнения специфических и неспецифических мест связывания олигонуклеотида с мишенью.

Равновесная степень связывания f олигонуклеотида с рассматриваемым участком равна

$$f = \frac{Kc_a}{1 + Kc_a}, \quad (4)$$

где K — соответствующая равновесная константа связывания; c_a — равновесная концентрация свободных олигонуклеотидов-адресов. Как видно из соотношения (4), величины f , отвечающие специфическому и неспецифическому связыванию, могут существенно различаться лишь в случае

$$Kc_a \ll 1. \quad (5)$$

В этом случае

$$f = Kc_a. \quad (6)$$

Условию (5) всегда можно удовлетворить, подобрав соответствующую температуру опыта и концентрацию адресов. Таким образом, максимальную специфичность адресованной модификации следует ожидать при выполнении условий (3) и (5), что и предполагает наш дальнейший анализ.

Вероятность того, что к моменту времени t произойдет модификация определенного места на рассматриваемом отрезке ДНК, имеет вид

$$p(t) = 1 - \exp(-f\varphi t) = 1 - \exp(-Kc_a\varphi t), \quad (7)$$

где φ — коэффициент, характеризующий реакционную способность модифицирующего реагента. Здесь предполагается для простоты анализа, что олигонуклеотиды находятся в растворе в избытке и изменением их концентрации за рассматриваемые времена можно пренебречь. Выражение (1) при этом можно записать

$$P(t) = [1 - \exp(-K_s c_a \varphi t)]^2 \prod_i \exp(-K_i c_a \varphi t) = [1 - \exp(-K_s c_a \varphi t)]^2 \exp(-K_n c_a \varphi t), \quad (8)$$

где индекс s означает, что соответствующие величины относятся к специфическому, индекс i — к неспецифическому связыванию, а

$$K_n = \sum_i K_i. \quad (9)$$

Нетрудно увидеть, что функция (8) имеет максимум при некотором значении t , которое и является оптимальным временем проведения реакции для получения специфической модификации. Максимальное значение $P(t)$ оказывается равным

$$P_{\max} = \frac{2^2 r^r}{(2+r)^{(2+r)}}, \quad (10)$$

где введено обозначение

$$r = K_n / K_s. \quad (11)$$

Отметим, что, как и следовало ожидать из общих соображений, величина P_{\max} определяется лишь отношением суммы констант неспецифического связывания к константе специфического связывания.

Статистико-механический анализ равновесных констант связывания. Термодинамическую константу связывания олигонуклеотидного адреса с одним полностью комплементарным местом на мишени можно представить в виде [6]

$$K_s = \beta \prod_i s_i, \quad (12)$$

где β — константа нуклеации (образования зародыша) двойной спирали; s_i — константа стабильности i -й пары, определяемая типом данной пары. Произведение берется по всем комплементарным парам образованной структуры.

Для оценки величины K_n определим количество и типы различных, неполностью комплементарных заданному адресу, участков на рассматриваемом отрезке мишени. Мы воспользуемся вероятностным подходом для оценки числа таких участков. Будем считать при этом, что основания в ДНК расположены случайным образом и доля каждого из них составляет $1/4$. Ясно, что в этом случае величина K_n должна быть пропорциональна общей длине отрезка L и может быть представлена в виде

$$K_n(L) = \frac{L}{L_0} K_n(L_0), \quad (13)$$

где L_0 — среднее расстояние между двумя участками специфического связывания адреса, состоящего из n звеньев ($L_0 = 4^n$). Любая конкретная последовательность длины n встречается на отрезке L_0 в среднем один раз, что упрощает вычисления $K_n(L_0)$.

Места неспецифического связывания, дающие заметный вклад в величину K_n , можно разбить на четыре группы.

В группу 1 входят участки длины n , способные образовывать комплексы, содержащие не более одной регулярной спиральной области, т. е. комплексы, в которых неспаренные основания адреса находятся на его концах. Рассмотрим прежде всего случай, когда на одном из концов адреса находится одно неспаренное основание. Это означает, что в данном месте мишени находится одно из трех оснований, некомплементарных рассматриваемому. Для каждого из трех вариантов константа специфического связывания больше константы неспецифического связывания в s_1 или s_n раз в зависимости от того, первое или последнее основание неспаренно. Общий вклад в r для этого случая равен

$$r_1 = 3 \left(\frac{1}{s_1} + \frac{1}{s_n} \right). \quad (14)$$

В случаях, когда на концах адреса имеется произвольное количество неспаренных оснований, вклад таких участков в величину r по аналогии с (14), составляет

$$\begin{aligned} r_1 = & 3 \left(\frac{1}{s_1} + \frac{1}{s_n} \right) + 3^2 \left(\frac{1}{s_1 s_2} + \frac{1}{s_1 s_n} + \frac{1}{s_{n-1} s_n} \right) + \\ & + 3^3 \left(\frac{1}{s_1 s_2 s_3} + \frac{1}{s_1 s_2 s_n} + \frac{1}{s_1 s_{n-1} s_n} + \frac{1}{s_{n-2} s_{n-1} s_n} \right) + \dots \end{aligned} \quad (15)$$

В группу 2 входят участки, имеющие неканонические пары внутри спиральной области. Поскольку наличие даже одной такой пары существенно снижает константу связывания олигонуклеотида, мы ограничимся рассмотрением участков, имеющих лишь одну неканоническую пару. В любом заданном месте спиральной области возможно появление неканонических пар трех типов. Однако в отличие от ситуации в группе 1 дестабилизирующий эффект каждой из этих трех типов пар будет различным. Количе-

ственно мы будем описывать этот эффект введением константы $s_i s_i^*$ вместо s_i в выражение (12) для константы связывания адреса. Величины s_i^* равны

$$s_i^* = \exp(-\Delta G_i/RT), \quad (16)$$

где ΔG_i отвечает изменению свободной энергии дуплекса при замене i -й канонической пары на данную неканоническую. Полный вклад участков, содержащих одну внутреннюю неканоническую пару, в величину r составляет

$$r_2 = \sum_{i=2}^{n-1} (s_{i1}^* + s_{i2}^* + s_{i3}^*), \quad (17)$$

где второй индекс в величине s^*_{ik} отвечает одному из трех типов неканонических пар, возможных на i -м месте. Значения констант s^*_{ik} можно получить на основании данных работы [2] (см. ниже).

В группу 3 входят участки, имеющие одно выпяченное основание в матрице. Верхняя оценка вклада таких участков в величину r может быть получена следующим образом. Равновесную константу полностью специфического связывания (12) необходимо домножить на фактор s^\perp , учитывающий дестабилизацию комплекса за счет вытолкнутого основания. На рассматриваемом отрезке L_0 среднее число таких участков с различной локализацией выпяченных оснований равно $(n-1)$. Таким образом, вклад участков этой группы в величину r составляет

$$r_3 = (n-1) s^\perp. \quad (18)$$

Из-за отсутствия экспериментальных данных мы полагаем, что фактор s^\perp одинаков для разных оснований. В работе [2] он определен лишь для цитозина.

В группу 4 входят участки, вызывающие при образовании комплекса выпячивание одного основания в адресе. По сравнению с группой 3 на единицу уменьшается эффективная длина адреса и, следовательно, число таких участков увеличивается в 4 раза. Среднее число этих участков на рассматриваемом отрезке равно $4(n-2)$. Кроме того, константа связывания в этом случае уменьшается в s_i раз, где индекс i отвечает выпячиваемому основанию. Следовательно,

$$r_4 = 4 \sum_{i=2}^{n-1} s^\perp / s_i. \quad (19)$$

Величины $r_1 + r_4$ дают определяющий вклад в величину r .

Значения параметров s , характеризующих стабильность канонических пар оснований, известны для широкого класса внешних условий [7, 8]. Для 1 М NaCl и $T = 25^\circ\text{C}$ $s_{AT} = 9,2$, $s_{GC} = 52,7$.

Систематическое исследование влияния различных некомплементарных пар на стабильность олигонуклеотидных комплексов впервые было проведено в работе [2]. В ней изучалось плавление комплексов $dCA_3XA_3G \cdot dCT_3YT_3G$, где X, Y = A, C, G, T. В табл. 1 этой работы приведены значения свободных энергий образования всевозможных дуплексов этого ряда. Из этой таблицы можно непосредственно определить величины ΔG_i , входящие в выражение (16) (табл. 1).

Таблица 1

Усредненные по двум возможным ориентациям пары в дуплексе значения ΔG_i , ккал/моль

The ΔG_i values averaged for the two orientations of a given pair in the duplex in kcal/mol

Основание в адресе	Основание в мишени				$(s_{i1}^* + s_{i2}^* + s_{i3}^*) \cdot 10^3$
	A	T	G	C	
A	4,0	0	3,3	4,6	5,80
T	0	4,0	2,8	3,9	12,20
G	4,0	3,6	3,5	0	6,70
C	5,3	4,3	0	5,3	0,80

Таблица 2
 Экстремальные значения $r(L_0)$ (длина $L_0=16384$)
 The extreme values of $r(L_0)$, $L_0=16384$

Адрес	r_1	r_2	r_3	r_4	r
dT ₇	1,166	0,061	0,024	0,009	1,260
dC ₇	0,124	0,004	0,024	0,002	0,154

Константа s^\perp также может быть вычислена на основании данных работы [2], где были найдены свободные энергии образования дуплексов dCA₆G·dCT₆G, dCA₃CA₃G·dCT₆G, dCA₆G·dCT₃CT₃G. Для перечисленных выше условий величина s^\perp равна 0,004.

Из вышесказанных численных значений термодинамических констант видно, что максимальное значение r достигается в случае адреса, состоящего из одних тимидинов, а минимальное — если адрес состоит из одних цитозинов. В табл. 2 приведены значения r для адреса длиной семь нуклеотидных звеньев.

Результаты и обсуждение. Проведенный анализ показывает, что адресованная модификация может в принципе служить основой специфического разрезания молекул ДНК. Еще раз отметим два необходимых условия достижения максимальной специфичности. Во-первых,

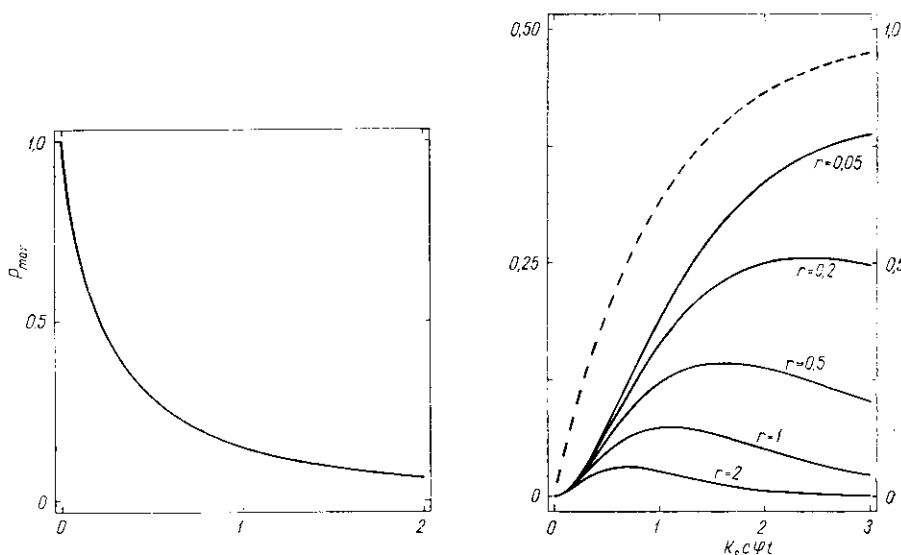


Рис. 1. Зависимость максимального выхода P_{\max} специфической модификации от отношения r суммы констант неспецифического связывания к константе специфического связывания адреса

Fig. 1. The highest ratio of selective and non-selective modification dependence on the ratio of the sum of constants of non-specific binding to the constant of specific binding

Рис. 2. Зависимость P выхода специфической модификации от времени реакции для различных r (сплошные линии). Здесь же приведена зависимость степени модификации участка специфического связывания от времени (пунктирная линия)

Fig. 2. The time dependence of specific modification for different values of r (—). Time dependence of the extent of modification of the specific target (— — —)

характерное время жизни специфического комплекса должно быть значительно меньше характерного времени τ_m модификации ДНК реагентом, находящимся в таком комплексе. Для выполнения этого условия можно использовать очень резкую зависимость τ_0 от температуры и ионной силы [9]. Во-вторых, произведение концентрации адреса на его

равновесную константу специфического связывания должно быть много меньше единицы. Лишь при выполнении этих двух условий специфичность модификации определяется равновесными константами специфического и неспецифического связывания. При этом максимальный выход специфической модификации зависит только от величины r , определяемой уравнением (11). На рис. 1 приведена зависимость $P_{\max}(r)$, отвечающая формуле (10). Видно, что для $r < 0,5$ $P_{\max} > 30\%$. При этом, как показывает анализ, выход любого конкретного фрагмента, возникающего за счет неспецифической модификации, оказывается в выбранных условиях менее 7%. Для типичных значений $r(L_0)$, полученных в предыдущем разделе, всегда найдется область L , при которой $r(L) < 0,5$ (см. соотношение (13)). Таким образом, по крайней мере для этой области длин выход специфической модификации оказывается достаточным для последующего получения специфических фрагментов.

Из выражения (8) следует, что для любого заданного значения r существует определенное оптимальное время реакции, при котором величина P достигает своего максимального значения. Зависимость $P(t)$ для различных значений r приведена на рис. 2. Видно, что максимальное значение P в зависимости от величины r достигается при разной глубине модификации. Выбор оптимального времени реакции является важным условием получения максимальной специфичности модификации. Как следует из приведенных выше данных, подавляющий вклад в величину r дают концевые неканонические пары комплекса. Поэтому для уменьшения значения r концы адреса должны состоять из гуанинов и (или) цитозинов.

В проведенном анализе не учитывалось возможное влияние вторичной структуры однонитевой ДНК на адресованную модификацию. Наличие такой структуры не должно изменить основные результаты работы, поскольку можно считать, что вовлеченные во вторичную структуру области не могут быть в такой системе подвергнуты модификации. Таким образом, вторичная структура может лишь увеличить реальную длину получаемых фрагментов, не меняя их выхода.

THEORETICAL ANALYSIS OF ADDRESSED MODIFICATION OF DNA

*M. P. Perelroyzen, A. V. Vologodskii**

Institute of Bioorganic Chemistry, Siberian Branch
of the Academy of Sciences of the USSR, Novosibirsk

*Institute of Molecular Genetics, Academy of Sciences of the USSR, Moscow

Summary

Specific chemical addressed modification of DNA involves treatment of single-stranded DNA by oligonucleotides complementary to certain target sequences in this DNA and bearing groupings reactive towards the DNA bases. The binding with the partially complementary targets can cause non-specific modification. The selectivity of the addressed modification is the subject of the present investigation. Modification of a fragment is analyzed which is flanked by two target sequences complementary to a given oligonucleotide address and contains no more such exact targets over all its length. Conditions of maximum ratio between specific and non-specific modification are calculated. The probability of the process in which up to the given time both of the target termini will be specifically modified without any non-specific modification of the internal part of a fragment is determined.

1. *Аффинная модификация биополимеров* / Под ред. Д. Г. Кнорре.— Новосибирск: Наука, 1983.—243 с.
2. *Aboul-ela F., Koh D., Tinoco I., Jr.* Base-base mismatches. Thermodynamics of double helix formation for $dCA_3XA_3G+dCT_3YT_3G$ ($X, Y=A, C, G, T$) // *Nucl. Acids Res.*—1985.—13, N 13.—P. 4811—4824.
3. *Craig M. E., Crothers D. M., Doty P.* Relaxation kinetics of dimer formation by self complementary oligonucleotides // *J. Mol. Biol.*—1971.—62, N 2.—P. 383—401.

4. *Pörschke D., Eigen M.* Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic-oligoriboadenylic acid system and of oligoriboadenylic acid alone at acid pH // *Ibid.*— P. 361—381.
5. *Pörschke D., Uhlenbeck O. C., Marlin F. H.* Thermodynamics and kinetics of the helix coil transition of oligomers containing GC base pairs // *Biopolymers.*— 1973.—**12**, N 6.— P. 1313—1335.
6. *Кантор Ч., Шиммель П.* Биофизическая химия.— М.: Мир, 1985.— Т. 3.—534 с.
7. *Веденов А. А., Дыхне А. М., Франк-Каменецкий М. Д.* Переход спираль — клубок в ДНК // *Успехи физ. наук.*— 1971.—**105**, № 3.— С. 479—519.
8. *Wada A., Yabuki S., Husimi Y.* Fine structure in the thermal denaturation of DNA. High temperature-resolution spectrophotometric studies // *CRC Crit. Rev. Biochem.*— 1980.—**9**.— P. 87—144.
9. *Slow relaxation processes in the melting of linear biopolymers: a theory and its application to nucleic acids* / V. V. Anshelevich, A. V. Vologodskii, A. V. Lukashin, M. D. Frank-Kamenetskii // *Biopolymers.*— 1984.—**23**, N 1.— P. 39—58.

Ин-т биоорг. химии Сиб. отд-ния
АН СССР, Новосибирск
Ин-т молекуляр. генетики АН СССР, Москва

Получено 03.03.87