

УДК 577.3

ХАРАКТЕРНЫЕ ОСОБЕННОСТИ РАСЩЕПЛЕНИЯ ДНК УЛЬТРАЗВУКОМ

© 2009 Ю.Д. Нечипуренко^{1,2*}, М.В. Головкин¹, Д.Ю. Нечипуренко², И.А. Ильичева¹,
Л.А. Панченко³, Р.В. Полозов⁴, С.Л. Гроховский¹¹Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва²Московский государственный университет им. М.В. Ломоносова, физический факультет³Московский государственный университет им. М.В. Ломоносова, биологический факультет⁴Институт теоретической и экспериментальной биофизики РАН

Статья поступила 21 июня 2008 г.

Ключевым моментом при количественном анализе картин электрофоретических гелей высокого разрешения, содержащих большое количество полос, является учет позиционного эффекта — зависимости интенсивности полосы от ее положения на геле. В данной работе развиты два метода, которые позволяют учитывать такой эффект. Оценку зависимости относительных частот разрывов межнуклеотидных связей от типа нуклеотидов, получаемые при анализе картин расщепления фрагментов молекул ДНК под действием ультразвука или химических факторов, проводили при помощи: а) линейно скользящего среднего; б) описания функции позиционной зависимости как полинома третьей степени. Оба метода приводят к сходным результатам. Анализ расщепления различных последовательностей общей длиной более двадцати тысяч нуклеотидов позволил установить относительные частоты разрывов связей во всех 16 динуклеотидах, встречающихся в последовательности ДНК. Установлено, что при ультразвуковом воздействии фосфодиэфирная связь, расположенная между цитозином и гуанином, подвергается расщеплению значительно чаще, чем соответствующие связи между прочими нуклеотидами в последовательности.

Ключевые слова: ультразвуковое расщепление ДНК, конформационная гетерогенность ДНК, электрофорез на геле, позиционная специфичность, футпринтинг.

ВВЕДЕНИЕ

Ранее нами был предложен экспериментальный подход, позволяющий установить зависимость локальной конформации молекулы ДНК от ее нуклеотидной последовательности. Мы показали, что частота разрывов сахарофосфатного остова ДНК под действием ультразвука определяется последовательностью нуклеотидов в ДНК [1] и обнаружили, что способность к разрывам зависит от локальных структурно-динамических свойств молекулы ДНК [2]. Такие свойства являются важным фактором, определяющим регуляцию экспрессии генов (см., например, [3—6]). Недавно был разработан экспериментальный метод, позволяющий обнаружить зависимость частоты разрывов сахарофосфатного остова некоторыми нуклеазами и химическими агентами от конформации ДНК [5]. Актуальной становится задача физического картирования регуляторных участков ДНК, т.е. нахождение участков ДНК, которые узнают факторы генетической экспрессии [6].

Анализируя данные о расщеплении ДНК с известной последовательностью, мы имеем возможность сформировать базу данных расщепления конкретных ее участков и получить информацию о неоднородности локальных структурно-динамических свойств двойной спирали ДНК.

* E-mail: nech99@mail.ru

Для получения данных о расщеплении ДНК нами был использован метод электрофореза в денатурирующем полиакриламидном геле. Расщепляющий химический агент или ультразвук воздействовали на фрагменты двухспиральной ДНК с известной последовательностью длиной несколько сотен пар нуклеотидов. Фрагменты содержали радиоактивную метку на конце одной из цепей. После денатурации фрагментов реакционные смеси наносили на гель. Результаты электрофореза и последующих процедур экспонирования, сканирования и цифровой обработки изображений записывали в виде массивов данных, соответствующих отдельным дорожкам геля. Элементом такого массива для одной дорожки является значение интенсивности полосы N_i , которое пропорционально количеству фрагментов длиной в i нуклеотидов в реакционной смеси.

Обычно на гелях имеется целый ряд дефектов, которые сильно осложняют перевод данных в цифровой формат и последующую обработку. Неоднородности толщины геля, пузырьки воздуха и присутствие разных количеств солей в наносимых образцах приводят к искривлению и деформации дорожек и к смещению полос на разных дорожках относительно друг друга. В исходном меченом фрагменте ДНК часто имеются небольшие примеси других фрагментов. Из-за потерь, вызванных осаждением на стенках пробирок, разные дорожки часто имеют различную суммарную интенсивность полос.

При продолжительной физической или химической обработке меченых фрагментов проявляется эффект "двойного удара", приводящий к увеличению в реакционной смеси доли коротких меченых фрагментов, а при ультразвуковом расщеплении наблюдается еще и эффект "нещепления концов": участки, близкие к концам фрагмента, расщепляются с меньшей частотой. По мере движения в геле вследствие диффузии происходит расширение полос.

Перечисленные факторы указывают на сложность количественного анализа данных о расщеплении молекул ДНК с определенной нуклеотидной последовательностью. Многократно проводя однотипные эксперименты, можно добиться получения "идеального геля", стандартизировать и учесть влияние большинства физических факторов. Однако для молекул ДНК с другими последовательностями нуклеотидов необходимо заново повторять подбор условий. При этом остается нерешенной проблема сравнения экспериментальных данных, полученных в разных сериях экспериментов. Поэтому актуальной задачей является поиск адекватных методов анализа таких экспериментальных данных.

АНАЛИЗ ПОЛИАКРИЛАМИДНЫХ ГЕЛЕЙ

На рис. 1 приведены изображения полиакриламидных гелей, полученные в результате разделения реакционных смесей радиоактивно меченых фрагментов ДНК после химического расщепления по пуринам (а) и после облучения ультразвуком (б) при разном времени воздействия. Для перевода картин расщепления в цифровую форму нами была использована компьютерная программа SAFA, разработанная группой из Стендфордского университета [7]. Эта программа позволяет после корректировки формы треков дорожек геля вычислять интенсивности всех полос на всех дорожках и соотносить эти полосы с заданной последовательностью нуклеотидов.

Химическое расщепление ДНК муравьиной кислотой в присутствии дифениламина [8] вызывает потерю пуринового основания в месте расщепления. Таким образом, полосы на дорожках $I-I0$ (см. рис. 1, а) и дорожке $A+C$ (см. рис. 1, б) — это олигонуклеотиды без концевого пурина.

Подвижность фрагментов нелинейно зависит от их молекулярной массы, суммарного заряда и от плотности геля. При разделении фрагментов ширина полосы зависит от расстояния, которое соответствующий фрагмент проходит в геле. Фрагменты меньшей длины проходят большее расстояние и образуют более широкие полосы, таким образом, интенсивность полос оказывается позиционно-детерминированной. Для оценки этого явления нами были проанализированы результаты химического расщепления фрагментов по пуринам при таком времени реакции, когда явление "двойного удара" пренебрежимо мало. При химической обработке вероятность расщепления конкретного нуклеотида не зависит от его положения во фрагменте ДНК. Однако оказалось, что после разделения в указанных на рис. 1 условиях интенсивности

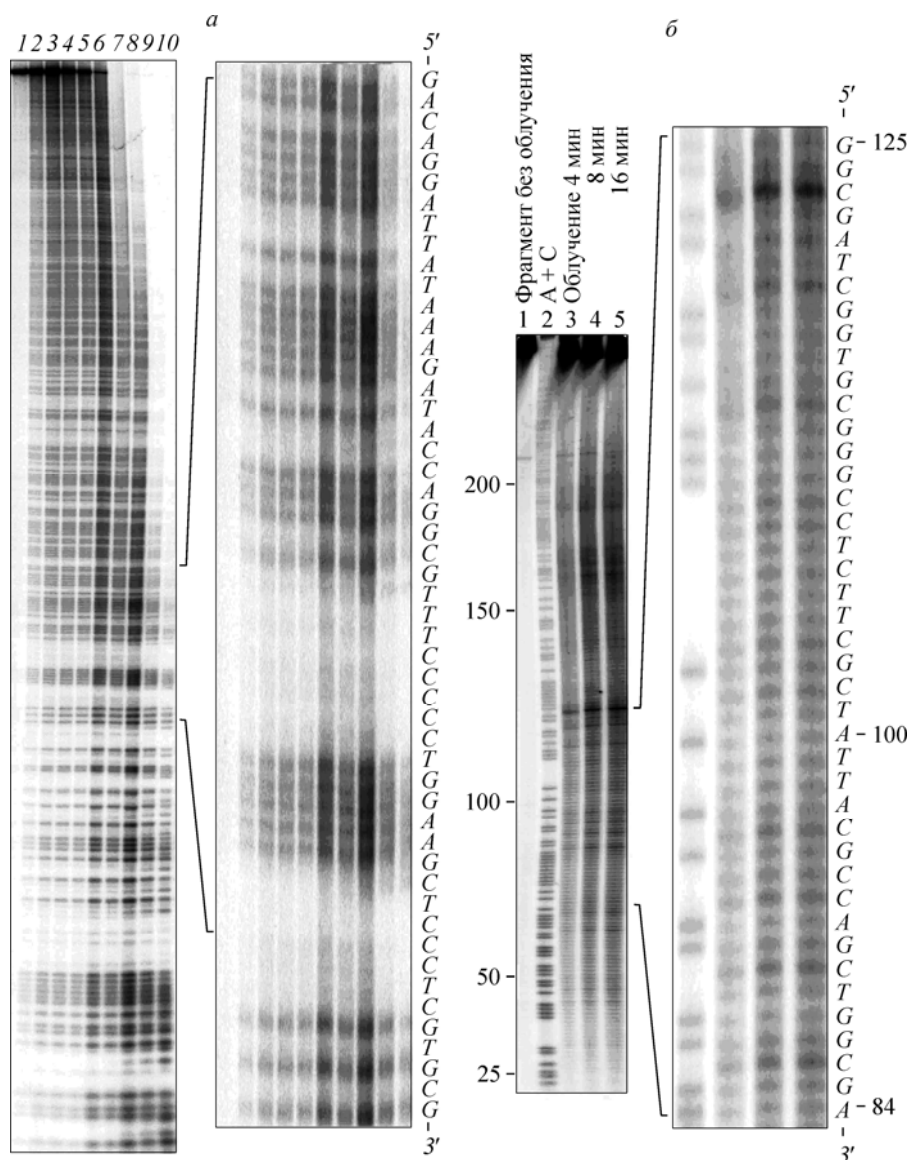


Рис. 1. Профили расщепления фрагментов ДНК в 6%-м денатурирующем полиакриламидном геле: *a* — результаты химического расщепления по пуринам: дорожка 1 — фрагмент без обработки, дорожки 2–10 — обработка муравьиной кислотой в присутствии дифениламина в течение 15, 30 с, 1, 2, 4, 8, 16, 32 и 64 мин соответственно. В правой части рисунка показан увеличенный участок геля и приведена соответствующая ему последовательность; *б* — результаты обработки ультразвуком с частотой 22 кГц: дорожка 1 — фрагмент без обработки, 2 — химическое расщепление по пуринам, 3, 4, 5 — облучение фрагмента ультразвуком 4, 8 и 16 мин соответственно. В правой части рисунка показан увеличенный участок геля и приведена соответствующая ему последовательность

полос в районах длин 50-го и 250-го нуклеотидов сильно различались. Этот факт свидетельствует о том, что имеет место позиционный эффект: при регистрации полос, содержащих одинаковое количество радиоактивной метки, интенсивность засвечивания флуоресцентного экрана зависит от ширины полосы (которая, в свою очередь, определяется положением полосы на геле).

Анализ полос на дорожках, полученных после облучения молекулы ДНК ультразвуком в течение 4, 8 и 16 мин показал (см. рис. 1, *б*), что при увеличении времени облучения интенсивность расщепления увеличилась, однако отношение интенсивностей полос друг к другу почти не изменилось. В этом случае отсутствует "эффект двойного удара", но наблюдается ярко выраженный краевой эффект: концы молекулы ДНК, как видно из рисунка, расщепляются значительно слабее, чем ее середина.

АНАЛИЗ ПРИ ПОМОЩИ "СКОЛЬЗЯЩЕГО ОКНА"

Значение интенсивности полосы N_i на дорожке пропорционально концентрации фрагмента, заканчивающегося i -м нуклеотидом. Чтобы учесть составляющую интенсивности, зависящую от положения полосы на дорожке, введем в рассмотрение величину r_i — относительную интенсивность полосы, соответствующую фрагменту длиной i нуклеотидов среди фрагментов, имеющих длины от $i - k$ до $i + k$ нуклеотидов:

$$r_i = N_i \frac{2k + 1}{N_{i-k} + N_{i-k+1} + \dots + N_{i+k}}, \quad (1)$$

где N_i — интенсивность полосы фрагмента длиной i нуклеотидов; $2k + 1$ — размер "окна", в центре которого находится соответствующая этому фрагменту полоса ($i > k$). Такая процедура позволяет учесть различия в интенсивности дорожек на разных гелях, а также позиционные эффекты. Варьирование величины "окна" показало, что для ультразвукового расщепления оптимальной является величина в 31 нуклеотид ($k = 15$): уменьшение этой величины приводит к увеличению разброса данных, а дальнейшее увеличение практически не сказывается на полученных значениях r_i (но уменьшает число данных, получаемых "на выходе" после проведения такого анализа).

АНАЛИЗ ПРИ ПОМОЩИ "МЕТОДА ПОЛИНОМА"

Зависимость величины интенсивности N_i от длины фрагмента i может быть представлена как полиномиальная. В простейшем случае это полином третьей степени (в пользу чего говорит асимметрия расщепления молекулы ДНК ультразвуком по длине [2]):

$$N_i = R_i (ai^3 + bi^3 + ci + d), \quad (2)$$

где a, b, c, d — коэффициенты полинома; R_i — относительная интенсивность i -й полосы (для "метода полинома").

Коэффициенты полинома a, b, c и d определяются при помощи МНК, в котором минимизируется сумма квадратов отклонения величин $\{N_i\}$ от соответствующих значений полинома ($ai^3 + bi^3 + ci + d$). После того, как коэффициенты полинома определены, из (3) можно вычислить величины R_i :

$$R_i = \frac{N_i}{ai^3 + bi^3 + ci + d}. \quad (3)$$

Далее мы будем называть метод, позволяющий вычислить величины R_i при помощи аппроксимации позиционной зависимости N_i , "методом полинома", а метод, позволяющий вычислять величины r_i при помощи скользящего окна — "методом скользящего окна".

СРАВНЕНИЕ МЕТОДОВ

Для химического расщепления оцифрованные дорожки 2—10 на рис. 1, а после деления на среднюю величину интенсивности полосы и последующей обработки при помощи одного из предложенных методов должны давать сходные значения относительной интенсивности для каждой конкретной полосы. Можно ожидать, что дорожки, которые отличаются разным количеством метки или разным временем экспонирования, будут после применения одного из предложенных методов приводить к близким значениям относительных интенсивностей для всех полос (полностью эти значения не могут совпадать, так как эффект "двойного удара" данными методами может компенсироваться лишь отчасти).

Отличие относительной интенсивности полосы из одной дорожки от относительной интенсивности соответствующей полосы из другой дорожки, обнаруженное после применения одного метода к обоим дорожкам, может свидетельствовать как о загрязнении дорожек, так и о способности данного метода учитывать позиционные эффекты.

На рис. 2 приведены данные для двух дорожек, соответствующих химическому расщеплению (см. рис. 1, а, дорожки 3 и 8) до обработки (а и б) и после применения "метода скользяще-

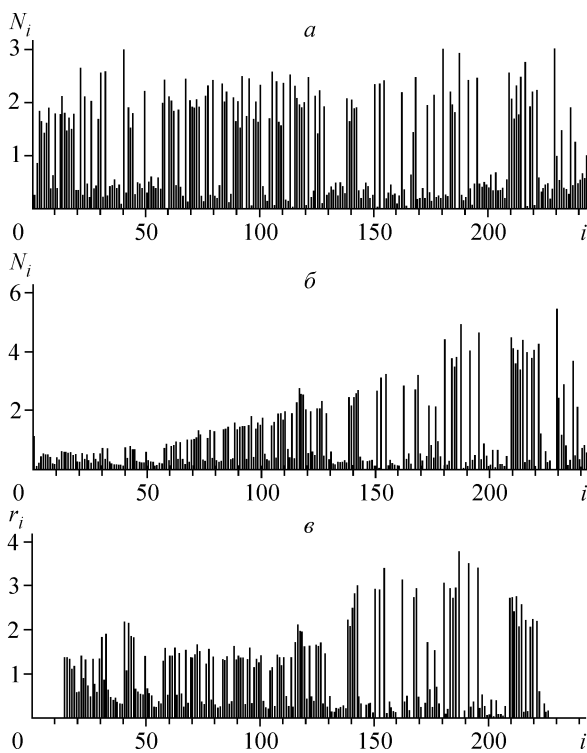


Рис. 2. Интенсивность полос гель-электрофореза для химического расщепления ДНК (см. рис. 1, а):

a и *б* соответствуют временам обработки 30 с и 16 мин, *в* — значения относительных интенсивностей после обработки данных на рис. 2, *б* при помощи метода скользящего среднего

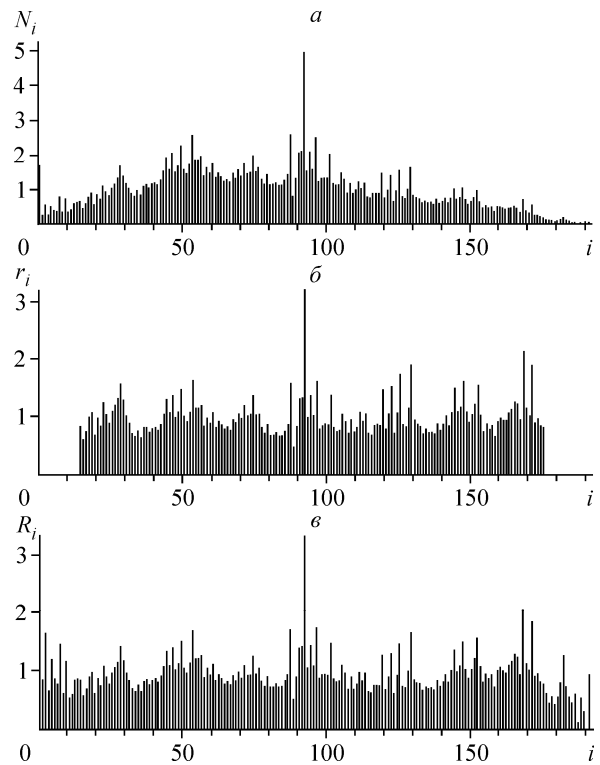


Рис. 3. Интенсивность полос гель-электрофореза для ультразвукового расщепления ДНК (см. рис. 1, б):

a — величины интенсивностей полос дорожки 4 без предварительной обработки, *б*, *в* — значения относительной интенсивности после обработки данных при помощи метода скользящего окна и метода полинома соответственно

го окна". Видно, что этот метод позволяет, по крайней мере частично, компенсировать эффект "двойного удара". На рис. 3 приведены данные, полученные при сканировании дорожки 4 (см. рис. 1, б), соответствующей ультразвуковому расщеплению (*a* — без обработки, *б* и *в* — после учета позиционных эффектов при помощи метода скользящего окна и метода полинома соответственно). Видно, что оба метода приводят примерно к одинаковым результатам.

Мы сравнили эти методы следующим образом: рассмотрели значения относительной интенсивности, затем рассчитали средние значения интенсивности каждой полосы для девяти дорожек, соответствующих химическому расщеплению (см. рис. 1, а), и среднеквадратичные отклонения. Суммируя все отклонения, мы получили величину, характеризующую способность метода "свести" все данные в одну "стандартизованную" дорожку со средними значениями относительной интенсивности полос. Аналогичную процедуру провели для данных по ультразвуковому расщеплению (см. рис. 1, б). Результаты сравнения представлены в табл. 1. Здесь приведены значения средних квадратичных отклонений, полученных при сравнении дорожек при помощи метода скользящего окна и метода полинома:

$$D = \sum_{i=1}^Q \sqrt{\frac{\sum_{j=1}^m (x_{ij} - \langle x_i \rangle)^2}{m(m-1)}}$$

где *m* — число дорожек; *Q* — число рассматриваемых полос в дорожке; величина *x_{ij}* — значение относительной интенсивности *i*-й полосы *j*-й дорожки (для *i*-й полосы *x_{ij}* соответствует от-

Т а б л и ц а 1

Сравнение методов анализа

| Тип воздействия на молекулу ДНК | Число полос | D | |
|---------------------------------|-------------|------------------------|----------------|
| | | Метод скользящего окна | Метод полинома |
| Химическое расщепление | 212 | 21,14 | 23,80 |
| Ультразвук | 162 | 6,02 | 5,99 |

носительной интенсивности r_i для метода скользящего окна и относительной интенсивности R_i для метода полинома); $\langle x_i \rangle$ — среднее по всем m дорожкам значение относительной интенсивности i -й полосы.

Метод скользящего окна после применения к данным по химическому расщеплению для девяти дорожек приводит к несколько меньшему среднеквадратичному отклонению, чем метод полинома (разница составляет примерно 10 %), в то время как применение обоих методов к данным по ультразвуковому расщеплению ДНК дает примерно одинаковые значения таких отклонений. Учет специфичности расщепления ДНК ультразвуком (см. [2]) позволяет при помощи процедуры самосогласования уменьшить среднеквадратичное отклонение для метода полинома на 20 %.

Следует заметить, что расхождение в значениях средних квадратичных отклонений (см. табл. 1) при сравнении методов характеризует не "ошибочность" того или иного метода, а его способность компенсировать позиционные эффекты и стандартизировать данные. Оба метода дают здесь примерно одинаковые результаты.

Недостатком метода скользящего окна является то, что при его применении база начальных данных интенсивностей сужается (с каждого конца фрагмента "пропадают" данные, соответствующее полуширине окна, см. рис. 2, *в* и 3, *б*).

При помощи описанных выше методов на основании анализа данных по расщеплению одного фрагмента ДНК ультразвуком (дорожка 4 на рис. 1, *б*) мы получили также величины относительных частот разрыва фосфодиэфирных связей, расположенных с 5' конца от *A*, *T*, *G* и *C* нуклеотидов (табл. 2).

Заметим, что эти данные коррелируют с данными, полученными нами в работе [2] при оценке относительных частот разрыва связей между разными динуклеотидами при воздействии ультразвука (в этой работе при помощи метода скользящего окна была обработана база данных по расщеплению 2500 динуклеотидов).

РЕЗУЛЬТАТЫ АНАЛИЗА РАСЩЕПЛЕНИЯ ДНК

Выше мы показали, что оба использованных нами метода анализа картин расщепления дают примерно одинаковые результаты. Мы применили метод скользящего среднего с окном в 31 нуклеотид для анализа базы данных по расщеплению 20 588 динуклеотидов. Эксперименты проводились в стандартных условиях 0,1 М NaOAc, pH 6,0 (методика получения и обработки данных описана в работах [1, 2]). В табл. 3 приведены значения относительной частоты разрыва фосфодиэфирных связей в 16 возможных динуклеотидах, встречающихся в ДНК. Суммированы данные для 138 дорожек разных гелей (часть последовательностей фрагментов перекрывалась).

Из приведенной таблицы следует, что динуклеотид 5'-CpG-3' расщепляется при облучении ультразвуком почти в полтора раза чаще, чем все другие динуклеотиды (со значимой статической достоверностью). В целом следует заключить, что фосфодиэфирная связь, следующая за сахаром цитозина, в наибольшей степени подвержена разрыву при ультразвуковой обработке.

Т а б л и ц а 2

Относительные частоты разрыва связей с 5' конца от различных нуклеотидов

| Тип нуклеотида | Число нуклеотидов | Относительные частоты разрыва | |
|----------------|-------------------|-------------------------------|-----------------|
| | | Метод скользящего окна | Метод полинома |
| <i>A</i> | 38 | $0,88 \pm 0,20$ | $0,86 \pm 0,22$ |
| <i>C</i> | 39 | $1,34 \pm 0,43$ | $1,28 \pm 0,45$ |
| <i>G</i> | 55 | $0,95 \pm 0,18$ | $0,92 \pm 0,18$ |
| <i>T</i> | 30 | $0,89 \pm 0,15$ | $0,86 \pm 0,16$ |

Т а б л и ц а 3

Относительные частоты разрыва связей разных динуклеотидов

| Тип нуклеотида | Число динуклеотидов | Среднее значение относительной частоты разрыва | Минимальные и максимальные значения относительной частоты разрыва | Стандартное отклонение |
|----------------|---------------------|--|---|------------------------|
| AA | 1636 | 0,919 | 0,570; 1,289 | 0,129 |
| AC | 1076 | 0,913 | 0,574; 1,242 | 0,128 |
| AG | 1028 | 0,900 | 0,575; 1,266 | 0,124 |
| AT | 1374 | 0,904 | 0,590; 1,257 | 0,119 |
| CA | 1265 | 1,160 | 0,686; 1,829 | 0,209 |
| CC | 1141 | 1,007 | 0,646; 1,416 | 0,144 |
| CG | 1230 | 1,444 | 0,823; 2,627 | 0,334 |
| CT | 1077 | 1,130 | 0,714; 1,718 | 0,198 |
| GA | 1153 | 0,970 | 0,620; 1,359 | 0,133 |
| GC | 1317 | 0,954 | 0,554; 1,368 | 0,146 |
| GG | 1168 | 0,922 | 0,584; 1,312 | 0,145 |
| GT | 1101 | 0,952 | 0,623; 1,327 | 0,126 |
| TA | 1065 | 0,973 | 0,670; 1,297 | 0,120 |
| TC | 1173 | 0,912 | 0,532; 1,262 | 0,131 |
| TG | 1305 | 0,979 | 0,645; 1,361 | 0,126 |
| TT | 1672 | 0,932 | 0,616; 1,275 | 0,127 |

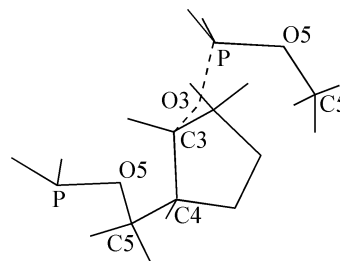
Можно предположить, что это связано с локальными структурно-динамическими свойствами двойной спирали ДНК. На рис. 4 показан участок сахарофосфатного остова ДНК, пунктиром показаны связи, по одной из которых происходит расщепление динуклеотида.

В данной работе рассмотрены методы количественной обработки электрофоретических гелей высокого разрешения, позволяющие устранить позиционно-детерминированные искажения в значениях интенсивностей полос на геле. Нами использованы методы скользящего среднего и описания позиционной зависимости при помощи полинома. Отмечена возможность компенсации явления "двойного удара" в результате применения указанных методов. Получены оценки относительных частот разрыва различных межнуклеотидных связей. Разрывы сахарофосфатного остова молекулы ДНК происходят чаще всего по фосфодиэфирной связи, соединяющей цитозин и гуанин (в направлении 5'-CpG-3').

Экспериментальные материалы и методы изложены в работах [1, 2].

Работа выполнена при финансовой поддержке грантов Российского фонда фундаментальных исследований (07-04-01031, 07-03-00492 и 08-04-01739) и Программы Президиума РАН по молекулярной и клеточной биологии.

Рис. 4. Фрагмент сахарофосфатной цепи ДНК. Пунктиром показаны связи, по которым может происходить расщепление ДНК



СПИСОК ЛИТЕРАТУРЫ

1. Гроховский С.Л. // Мол. биол. – 2006. – **40**. – С. 317 – 325.
2. Гроховский С.Л., Ильичева И.А., Нечипуренко Д.Ю. и др. // Биофизика. – 2008. – **53**. – С. 417 – 425.
3. Нечипуренко Ю.Д., Полозов Р.В., Нечипуренко Д.Ю. и др. Математические модели регуляции экспрессии генов: механические возмущения структуры ДНК / Математика. Компьютер. Образование: Сб. научных трудов. Т.2 / Под ред. Г.Ю. Ризниченко. – М.; Ижевск: НИЦ "Регулярная и хаотическая динамика", 2006. – С. 392 – 402.
4. Nechipurenko Yu.D., Nechipurenko D.Yu., Polozov R.V. et al. DNA-based nanostructures: changes of Mechanical properties of DNA upon ligand binding // Nanomaterials for Application in Medicine and Biology (NATO Workshop in Bonn). Springer. – 2008. – P. 59 – 67.
5. Greenbaum J.A., Parker S.C.J., Tullius T.D. // Genome Res. – 2007. – **17**. – P. 940 – 946.
6. Greenbaum J.A., Pang Bo, Tullius T.D. // Ibid. – P. 947 – 953.
7. Das R., Laederach A., Pearlman S.M. et al. // RNA. – 2005. – **11**. – P. 344 – 354.
8. Tate W.P., Petersen G.B. // Biochem. J. – 1975. – **147**. – P. 439 – 445.