

Математические методы анализа электрофоретических картин расщепления ДНК

М. В. Головкин¹, Д. Ю. Нечипуренко², И. А. Ильичева¹,
Л. А. Панченко², Р. В. Полозов³, С. Л. Гроховский¹,
Ю. Д. Нечипуренко^{2,a}

¹ Институт молекулярной биологии им. В. А. Энгельгарда,
119991, г. Москва, ул. Вавилова, д. 32

² Московский государственный университет им. М. В. Ломоносова,
119991, ГСП-1, Москва, Ленинские горы, МГУ, д. 1, стр. 2

³ Институт теоретической и экспериментальной биофизики РАН,
142290, г. Пущино Московской обл., ул. Институтская, д. 3

E-mail: ^a nech99@mail.ru

Получено 30 июня 2009 г.

При анализе картин электрофореза высокого разрешения необходимо учитывать позиционный эффект — зависимость интенсивности полосы от ее положения в геле. Оценки зависимости относительных частот разрывов межнуклеотидных связей от типа нуклеотидов, получаемые при анализе картин расщепления фрагментов молекул ДНК, производились при помощи методов: а) линейно скользящего среднего; б) описания функции позиционной зависимости полиномом третьей степени. Анализ расщепления последовательностей ДНК при помощи линейно скользящего среднего позволил установить относительные частоты разрывов связей во всех 16 динуклеотидах, встречающихся в последовательности ДНК.

Ключевые слова: электрофорез, расщепление ДНК, физическое картирование ДНК

Calculational methods for electrophoretic cleavage pattern analysis of DNA

M. V. Golovkin¹, D. Yu. Nechipurenko², I. A. Il'icheva¹, L. A. Panchenko², R. V. Polozov³,
S. L. Grokhovsky¹, Yu. D. Nechipurenko²

¹ Engelhardt Institute of Molecular Biology RAS, Vavilov str. 32, Moscow 119991, Russia

² Moscow State University, MSU, 1, building 2, GSP-1, Leninskiye Gory, Moscow, 119991, Russia

³ Institute of Theoretical and Experimental Biophysics RAS, Institutskaya str. 3, Puschino, Moscow Region, 142290, Russia

Abstract. — For quantitative analysis of high resolution gel electrophoresis data it is necessary to consider the position effect, i. e. the dependence of each band intensity on its position on the gel. The relative cleavage rates of DNA phosphodiester bonds in all 16 possible steps (calculated using quantitative data on electrophoretic patterns of DNA fragments cleavage) were estimated using two methods: a) Linearly Weighted Moving Average; b) polynomial built by the least-square method. The values obtained for cleavage rates for different dinucleotides were similar regardless of particular method.

Key words: electrophoresis, DNA cleavage, Physical mapping of DNA structure

Citation: *Computer Research and Modeling*, 2009, vol. 1, no. 3, pp. 287–295 (Russian).

© 2009 М. В. Головкин, Д. Ю. Нечипуренко, И. А. Ильичева, Л. А. Панченко, Р. В. Полозов, С. Л. Гроховский, Ю. Д. Нечипуренко

Введение

Зависимость локальных структурно-динамических свойств молекулы ДНК от последовательности нуклеотидов играет важную роль в процессах регуляции генетической экспрессии (см., например, [1–4]). Ранее было показано, что частота разрыва сахарофосфатного остова ДНК под действием ультразвука определяется последовательностью нуклеотидов в ДНК [5–6]. Недавно также был разработан экспериментальный метод, позволяющий исследовать зависимость частоты разрывов сахарофосфатного остова некоторыми нуклеазами и химическими агентами от конформации ДНК [3]. Поскольку известно, что многие регуляторные последовательности ДНК проявляют необычные структурно-динамические свойства, актуальной становится задача физического картирования таких участков ДНК, — т. е. поиск последовательностей, с которыми избирательно связываются факторы генетической экспрессии [4].

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.

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 реакционные смеси наносили на гель. Электрофорез и последующие процедуры экспонирования, сканирования и цифровой обработки изображений давали результаты 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 их перевод 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 смещению полос на разных дорожках друг относительно друга. В исходном меченом фрагменте ДНК часто имеются небольшие примеси других фрагментов. Из-за потерь, вызванных осаждением на стенках пробирок, разные дорожки часто имеют разную суммарную интенсивность полос.

При продолжительной физической или химической обработке меченых фрагментов проявляется эффект «двойного удара», приводящий к увеличению в реакционной смеси доли коротких 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.

Результаты и обсуждение

Анализ полиакриламидных гелей. На рис. 1 приведены картины полиакриламидных гелей, полученные в результате разделения реакционных смесей радиоактивно меченых фрагментов ДНК после химического расщепления по пуринам (а) и после облучения ультразвуком (б)

при разных временах воздействия. Для перевода картин расщепления в цифровую форму нами была использована компьютерная программа SAFA, разработанная группой из Стенфордского 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.

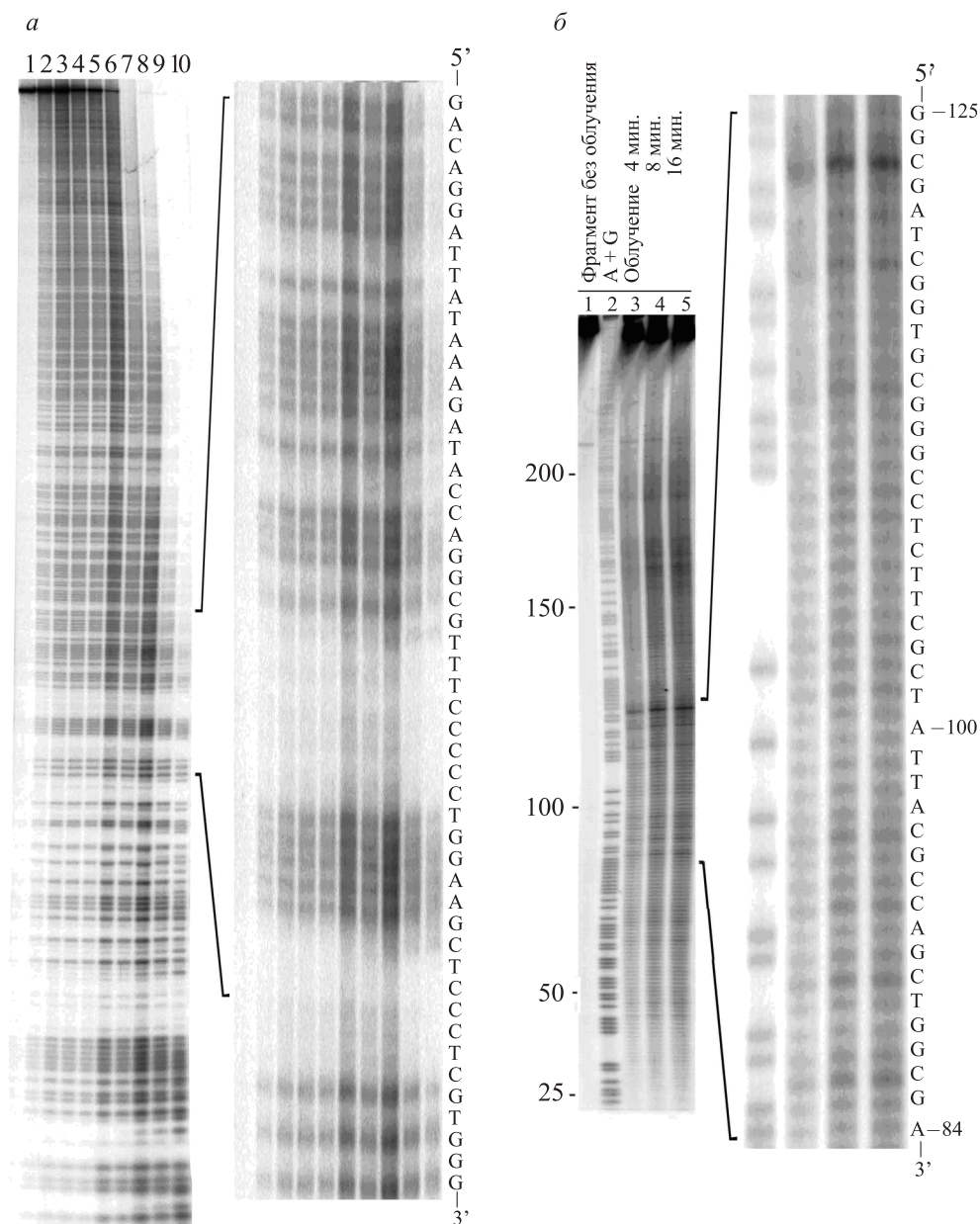


Рис. 1. Профили расщепления фрагментов ДНК в 6 %-ном денатурирующем полиакриламидном геле: *a* — результаты химического расщепления по пуринам: дорожка 1 — фрагмент без обработки; дорожки 2–10 — обработка муравьиной кислотой в присутствии дифениламина в течение 15, 30 сек, 1, 2, 4, 8, 16, 32 и 64 мин 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

Химическое расщепление ДНК муравьиной кислотой в присутствии дифениламина [8] вызывает потерю пуринового основания в месте расщепления. Таким образом, полосы на дорожках 1–10 (рис. 1а) и дорожке А+Г (рис. 1б) — это олигонуклеотиды без концевой пурина.

Подвижность фрагментов нелинейно зависит от их молекулярной массы, суммарного заряда и 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 не зависит от его положения во фрагменте ДНК. Однако оказалось, что после разделения в указанных на рис. 1 условиях интенсивности 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 (рис. 1б), 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 интенсивностях 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 описана полиномом. В простейшем случае, это полином третьей степени (в пользу чего говорит асимметрия расщепления молекулы ДНК ультразвуком [6]):

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

где a, b, c, d — коэффициенты полинома, а 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 ($a i^3 + b i^2 + c i + d$). После того как коэффициенты полинома определены, из (3) можно вычислить величины R_i :

$$R_i = \frac{N_i}{ai^3 + bi^2 + ci + d}.$$

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. 1а после деления на среднюю величину интенсивности полосы и последующей обработки при помощи одного из предложенных методов должны давать сходные относительные 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 обеим 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 (рис. 1а, дорожки 3 и 8) до обработки (а и б) и после применения «метода скользящего 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. 1б), соответствующей ультразвуковому расщеплению (а — 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.

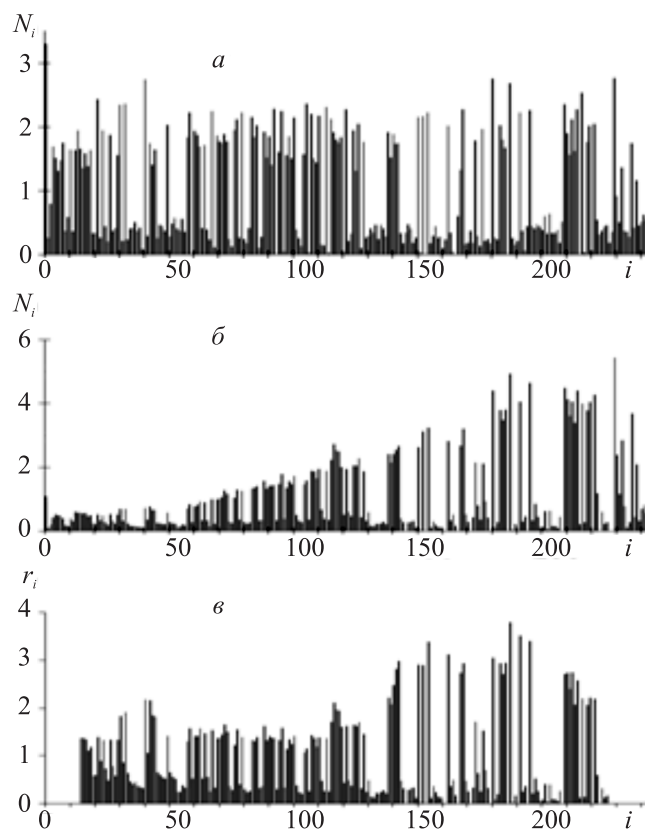


Рис. 2. Интенсивности полос гель-электрофореза для химического расщепления ДНК (см. рис. 1а). а и б соответствуют временам обработки 30 сек и 16 мин; в — значения относительных интенсивностей после обработки данных на рис. 2б при помощи метода скользящего среднего

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

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

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 относительной 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.

Таблица 1

Сравнение методов анализа

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

«Метод скользящего окна» после применения к данным по химическому расщеплению для девяти дорожек приводит к несколько меньшему среднеквадратичному отклонению, чем метод полинома, 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. [6]) позволяет при помощи процедуры самосогласования уменьшить среднеквадратичное отклонение для метода полинома на 20 %.

Следует заметить, что расхождения в значениях средних квадратичных отклонений (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. 2в и рис. 3б).

При помощи описанных выше методов на основании анализа данных по расщеплению одного фрагмента ДНК ультразвуком (дорожка 4 на рис. 1б) мы получили также величины относительных частот разрыва фосфодиэфирных связей, расположенных с 5' конца от А, Т, Г и С нуклеотидов (см. табл. 2).

Заметим, что эти данные коррелируют с данными, полученными нами в работе [6] 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 2 500 динуклеотидов).

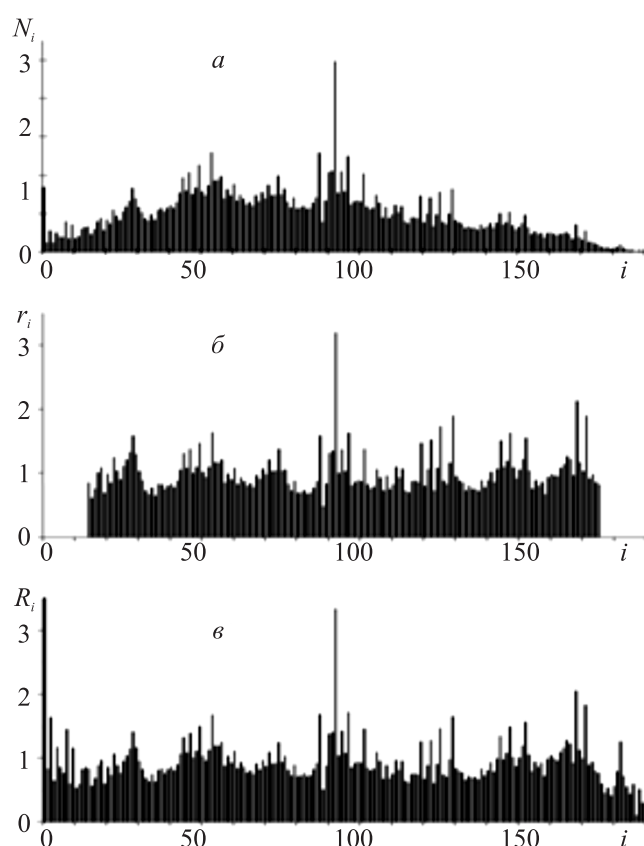


Рис. 3. Интенсивности полос гель-электрофореза для ультразвукового расщепления ДНК (см. рис. 1б). *a* — величины интенсивностей полос дорожки 4 без предварительной обработки; *б, в* — значения относительных интенсивностей после обработки данных при помощи «метода скользящего окна» и «метода полинома» соответственно

Таблица 2

Относительные частоты разрыва связей с 5' конца от различных нуклеотидов

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

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, соответствующих интенсивностям расщепления 20 588 динуклеотидов. Эксперименты проводились в стандартных условиях 0,1 М NaOAc pH 6.0 (методика получения и обработки данных описана в работах [5, 6]). В табл. 3 приведены относительные частоты разрыва фосфодиэфирных связей в 16 возможных динуклеотидах, встречающихся в ДНК. Суммированы данные для 138 дорожек разных гелей (часть последовательностей фрагментов перекрывалась).

Таблица 3

Относительные частоты разрыва связей разных нуклеотидов

	Число динуклеотидов	Среднее значение относительной частоты разрыва	Минимальные и максимальные значения относительной частоты разрыва	Стандартное отклонение
AA	1 636	0,919	0,570; 1,289	0,129
AC	1 076	0,913	0,574; 1,242	0,128
AG	1 028	0,900	0,575; 1,266	0,124
AT	1 374	0,904	0,590; 1,257	0,119
CA	1 265	1,160	0,686; 1,829	0,209
CC	1 141	1,007	0,646; 1,416	0,144
CG	1 230	1,444	0,823; 2,627	0,334
CT	1 077	1,130	0,714; 1,718	0,198
GA	1 153	0,970	0,620; 1,359	0,133
GC	1 317	0,954	0,554; 1,368	0,146
GG	1 168	0,922	0,584; 1,312	0,145
GT	1 101	0,952	0,623; 1,327	0,126
TA	1 065	0,973	0,670; 1,297	0,120
TC	1 173	0,912	0,532; 1,262	0,131
TG	1 305	0,979	0,645; 1,361	0,126
TT	1 672	0,932	0,616; 1,275	0,127

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. Можно предположить, что это связано с локальными структурно-динамическими свойствами двойной спирали ДНК.

Заключение

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 метод описания позиционной зависимости при помощи полинома. Показано, что оба метода приводят к незначительно отличающимся результатам, а также продемонстрирована возможность компенсации явления «двойного удара» в результате применения указанных методов. При помощи метода «скользящего окна» получены оценки относительных частот разрыва различных межнуклеотидных связей.

Экспериментальные материалы и методы изложены в работах [5, 6].

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

Список литературы

1. Нечипуренко Ю. Д., Полозов Р. В., Нечипуренко Д. Ю., Ильичева И. А., Воробьев Е. А., Гроховский С. Л. и Гурский Г. В. Математические модели регуляции экспрессии генов: механические возмущения структуры ДНК // Математика. Компьютер. Образование: Сб. науч. тр. / Под ред. Г. Ю. Ризниченко. М.–Ижевск: НИЦ «Регулярная и хаотическая динамика», 2006. Т. 2. С. 392–402.
2. Nechipurenko Yu. D., Nechipurenko D. Yu., Polozov R. V., Grokhovsky S. L., Gursky G. V. DNA-based nanostructures: changes of Mechanical properties of DNA upon ligand binding // *Nanomaterials for Application in Medicine and Biology (NATO Workshop in Bonn)*. 2008. Springer. P. 59–67.
3. Greenbaum J. A., Parker S. C. J. and Tullius T. D. Detection of DNA structural motifs in functional genomic elements // *Genome Res.* 2007. Vol. 17. P. 940–946.
4. Greenbaum J. A., Pang Bo and Tullius T. D. Construction of a genome-scale structural map at single-nucleotide resolution // *Genome Res.* 2007. Vol. 17. P. 947–953.
5. Гроховский С. Л. Специфичность расщепления ДНК ультразвуком // *Мол. биология*. 2006. Т. 40. С. 317–325.
6. Гроховский С. Л., Ильичева И. А., Нечипуренко Д. Ю., Панченко Л. А., Полозов Р. В., Нечипуренко Ю. Д. Локальные неоднородности структуры и динамики двухспиральной ДНК: исследование при помощи ультразвука // *Биофизика*. 2008. Т. 53. С. 417–425.
7. Das R., Laederach A., Pearlman S. M., Herschlag D., Altman R. B. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments // *RNA*. 2005. Vol. 11. P. 344–354.
8. Tate W. P., Petersen G. B. Stability of pyrimidine oligodeoxyribonucleotides released during degradation of deoxyribonucleic acid with formic acid — diphenylamine reagent // *Biochem J.* 1975. June, Vol. 147. P. 439–445.

