

КОЛИЧЕСТВЕННЫЕ МЕТОДЫ АНАЛИЗА РАСЩЕПЛЕНИЯ ДНК УЛЬТРАЗВУКОМ

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

Ключевым моментом при количественном анализе электрофоретических гелей высокого разрешения, содержащих большое количество полос, является учет позиционного эффекта – зависимости интенсивности полосы от её положения на геле. Оценки зависимости относительных частот разрывов межнуклеотидных связей от типа нуклеотидов, получаемые при анализе картин расщепления фрагментов молекул ДНК под влиянием ультразвука или действия химических факторов, производились при помощи двух методов: а) линейно скользящего среднего; б) описания функции позиционной зависимости как полинома третьей степени. Оценки относительных частот разрывов межнуклеотидных связей для разных нуклеотидов приводят к сходным результатам вне зависимости от выбора конкретного метода для учета позиционного эффекта

Введение. После расшифровки геномов ряда организмов следующей по важности задачей является поиск регуляторных участков генов. Места связывания регуляторных белков и РНК-полимераз с ДНК определяются локальной структурой и динамикой молекулы ДНК. Ранее нами был предложен экспериментальный подход, позволяющий установить зависимость локальной конформации молекулы ДНК от её 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 также 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, – то есть нахождение участков ДНК, которые узнают факторы экспрессии [6]. Заметим, что такое узнавание участков ДНК молекулой белка является «непрямым» [7] – молекула белка узнает в целом конформацию участка ДНК и/или электростатический потенциал на некотором расстоянии от его поверхности. Такой механизм узнавания отличается от «прямого узнавания» определенных реакционных центров ДНК определенными атомными группами молекулы белка [8]. Механизмы «непрямого» узнавания могут быть связаны с распределением электростатического потенциала вдоль сигнальных (промоторных) участков ДНК [9, 10].

Анализируя данные о расщеплении ДНК с известной последовательностью, мы имеем возможность сформировать базу данных расщепления конкретных её 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 действовали на образцы двухспиральной ДНК с известной последовательностью длиной несколько сот пар нуклеотидов. Такие образцы содержали радиоактивную метку на конце одной из цепей. После денатурации фрагментов реакционные смеси наносили на гель. Электрофорез и последующие процедуры экспонирования и сканирования давали результаты в виде цифрового файла, содержащего массивы данных, соответствующие отдельным дорожкам геля. Элементами такого массива для одной дорожки являются значения интенсивности полос N_i . Эти интенсивности пропорциональны количеству фрагментов длины i в реакционной смеси, соответствующей данной дорожке.

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 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 разную суммарную интенсивность полос дорожки.

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 ещё и эффект «неспеления 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 с такими молекулами ДНК, 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 остаётся нерешенной проблема сравнения экспериментальных данных, полученных в разных сериях экспериментов. Поэтому актуальной задачей является поиск адекватных методов анализа таких экспериментальных данных.

Оцифровка полиакриламидных гелей. На рис. 1 приведены картины полиакриламидных гелей, полученные в результате разделения реакционных смесей радиоактивно меченых фрагментов ДНК после химического расщепления по пуринам (*a*) 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 Стендфорского университета [11]. Эта программа позволяет после корректировки 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 вызывает потерю пуринового основания в месте расщепления. Таким образом, поло-

сы на дорожках 1-10 (рис. 1а) и дорожке А+G (рис.1б) – это олигонуклеотиды без концевого пурина. Подвижность фрагментов нелинейно зависит от их молекулярной массы, суммарного заряда и от плотности геля.

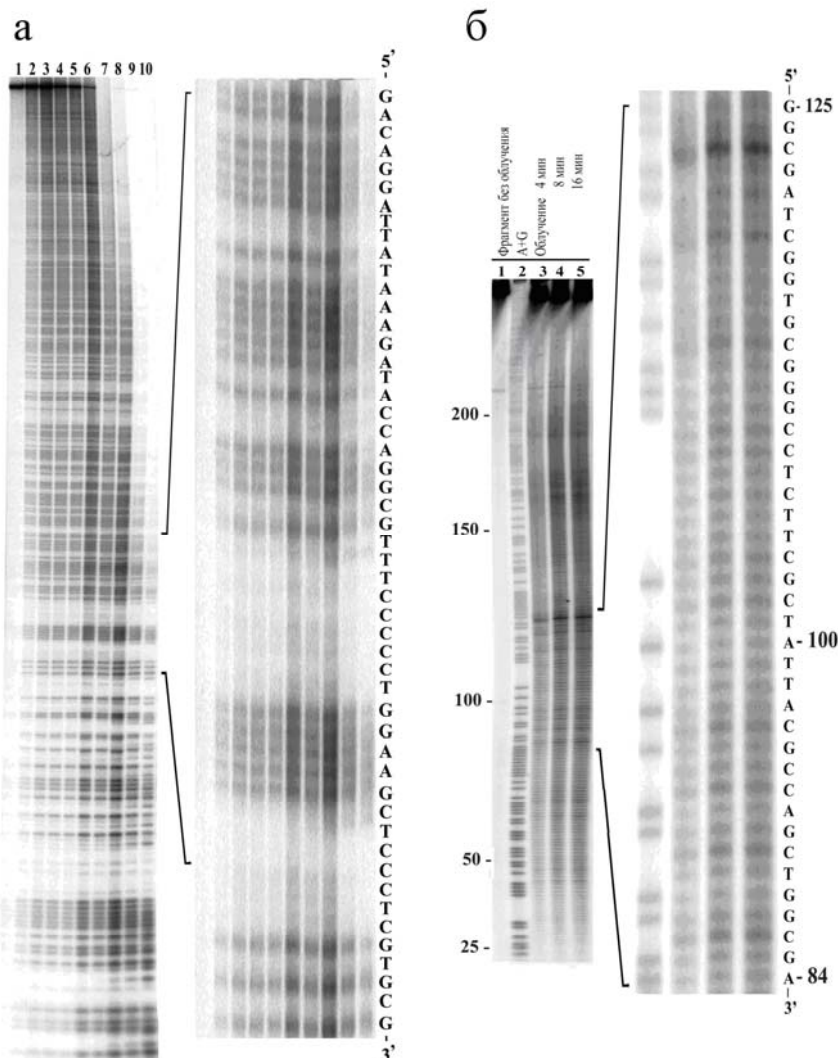


Рис. 1. Профили расщепления фрагментов ДНК в 6%-ном денатурирующем полиакриламидном геле: *а.* Результаты химического расщепления по пуринам: дорожка 1 - фрагмент без обработки; дорожки 2–10 – обработка муравьиной кислотой в присутствии дифениламина в течении 15 сек, 30сек, 1 мин, 2 мин, 4 мин, 8 мин,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

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

Анализ полос на дорожках, полученных после облучения молекулы ДНК ультразвуком в течении 4, 8 и 16 минут 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 её середина.

Анализ при помощи «скользящего окна». Значение интенсивности полосы N_i на дорожке пропорционально концентрации фрагмента, заканчивающегося i -ым 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 оптимальным 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 .

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 [4]):

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

где a, b, c, d – коэффициенты полинома, а R_i – относительная интенсивность i -ой полосы (для «метода полинома»).

Если мы рассматриваем фрагменты с длинами $L, L + 1 \dots M$, то

$$R_i = \frac{(M - L + 1) N_i}{\sum_{j=L}^M N_j}, \quad (3)$$

Таким образом,

$$\sum_{i=L}^M R_i = M - L + 1.$$

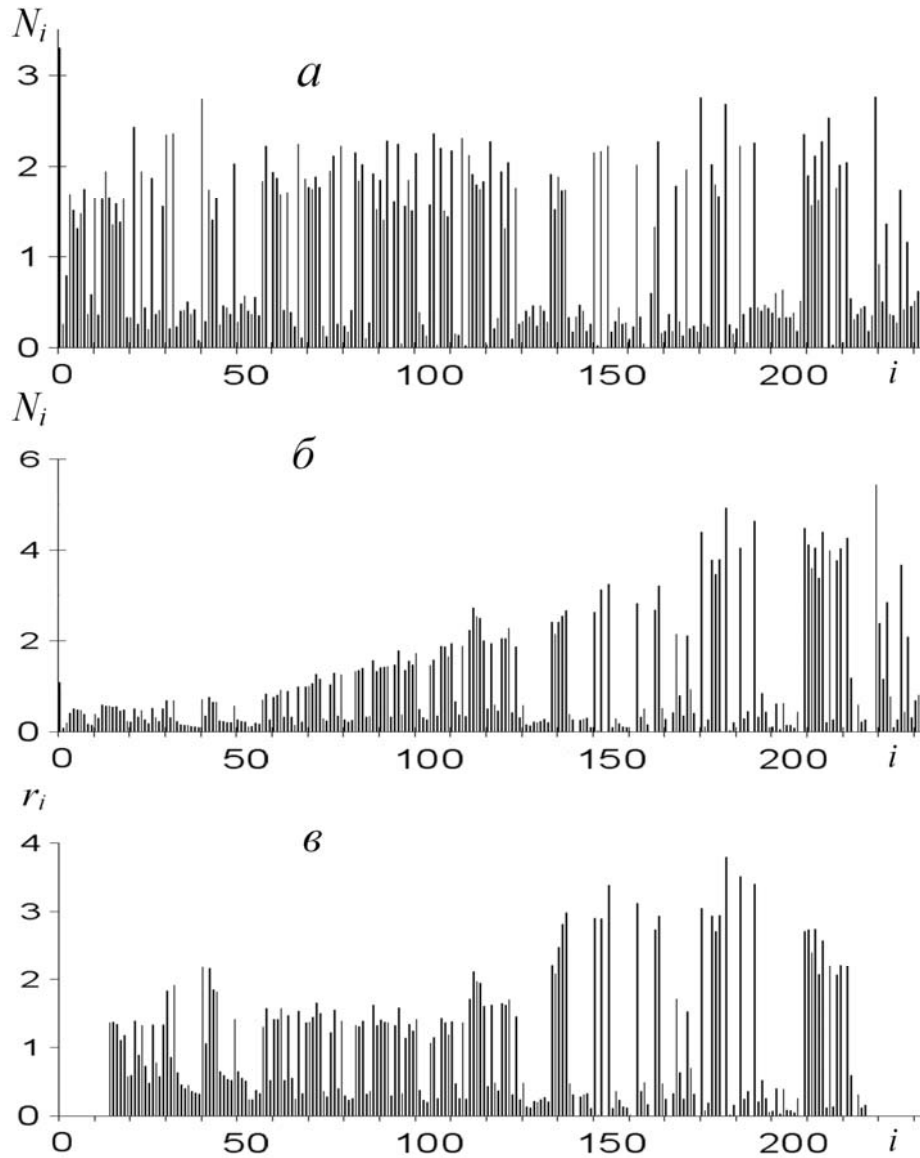


Рис. 2. Значения интенсивностей полос гель-электрофореза, соответствующие химическому расщеплению ДНК (см. рис. 1а). *a* и *б* соответствуют временам обработки 30 секунд и 16 минут. *в* – значения относительных интенсивностей после обработки данных на рис. 2б при помощи «метода скользящего окна»

Коэффициенты полинома *a*, *b*, *c* и *d* можно оценить в предположении минимальности отклонения набора величин $\{N_i\}$ от значений полинома $(a i^3 + b i^2 + c i + d)$. После того, как коэффициенты полинома оценены, из (3) можно вычислить величины R_i :

$$R_i = \frac{N_i}{a i^3 + b i^2 + c i + d},$$

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 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.

Таблица 1. Сравнение методов анализа

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

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 таблице 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 -ой 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.

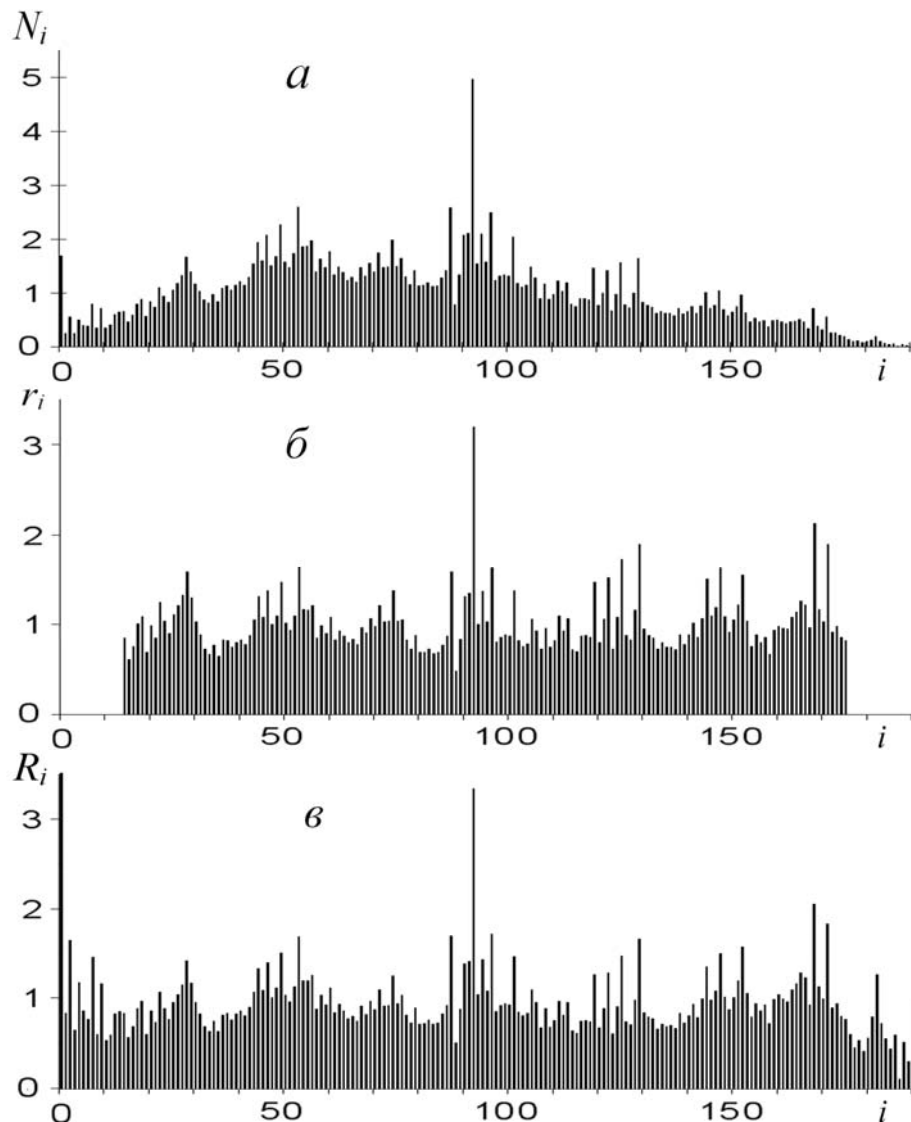


Рис. 3. Значения интенсивностей полос гель-электрофореза, соответствующие ультразвуковому расщеплению ДНК (см. рис. 1б). *а* – величины интенсивностей полос дорожки 4 без предварительной обработки. *б*, *в* – значения относительных интенсивностей, после обработки данных при помощи «метода скользящего окна» и «метода полинома», соответственно. Заметим, что большинство максимумов здесь определяется положением цитозинов

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 к данным по ультразвуковому расщеплению ДНК обоих методов даёт 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 расхождения в значениях средних квадратичных отклонений (Таблица 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, G и С нуклеотидов (см. Таблицу 2).

Таблица 2. Относительные частоты разрыва связей с 5' конца от различных нуклеотидов

Тип нуклеотида	Число нуклеотидов	Относительные частоты разрыва	
		Метод скользящего окна	Метод полинома
А	38	0,88 ± 0,2	0,86 ± 0,22
С	39	1,34 ± 0,43	1,28 ± 0,45
G	55	0,95 ± 0,18	0,92 ± 0,18
Т	30	0,89 ± 0,15	0,86 ± 0,16

Следует заметить, 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 в 2500 нуклеотидов).

Заключение. 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 метод 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 оства молекулы ДНК происходят чаще всего с 5' конца цитозина.

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

СПИСОК ЛИТЕРАТУРЫ

1. *Гроховский С.Л.* Специфичность расщепления ДНК ультразвуком // Мол. биология. 2006., т. 40, с. 317–325.
2. *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, №2 (в печати).
3. *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. 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., 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)*. Springer. 2008. P. 59–67.
5. *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.
6. *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.
7. *Araújo-Bravo M. J., Fujii S., Kono H., Ahmad Sh., Sarai A.* Sequence-dependent conformational energy of DNA derived from molecular dynamics simulations: toward understanding the indirect readout mechanism in protein-DNA recognition // *J. Am. Chem. Soc.* 2005. Vol. 127, P. 16074–16089.
8. *Гурский Г.В., Заседателев А.С.* Точные соотношения, описывающие связывание регуляторных белков и других решеточных лигандов на двухспиральных полинуклеотидах // *Биофизика.* 1978. Т. 23. С. 932–946.
9. *Polozov R.V., Dzhylyadin T.R., Sorokin A.A., Ivanova N.N., Sivozhelezov V.S., Kamzolova S.G.* Electrostatic potentials of DNA. Comparative analysis of promoter and nonpromoter nucleotide sequences // *J Biomol. Struct. Dyn.* 1999. Vol. 16. P. 1135–1143
10. *Polozov R.V., Sivozhelezov V.S., Ivanov V.V., Melnikov Yu.* On a classification of *E.Coli* promoters according to their electrostatic potentials. // *Письма в ЭЧАЯ.* 2005. Т.2 С. 82–90.
11. *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.

QUANTITATIVE METHODS FOR ANALYSIS OF DNA CLEAVAGE BY ULTRASOUND

**Nechipurenko Yu. D., Golovkin M. V., Nechipurenko D. Yu., Il'icheva I. A.,
Panchenko L. A., Polozov R. V., Grokhovsky S. L.**

For quantitative analysis of high resolution gel electrophoresis data with great number of stripes it is necessary to consider the position effect, i.e. the dependence of stripe intensity on its position on the gel. Dependency of relative cleavage rates of DNA internucleotide bonds (measured from patterns of DNA fragments cleavage by ultrasound or chemical factors) on base pair sequence was estimated using two methods: a) Linearly Weighted Moving Average; b) polynomial built by the least-square method. Estimates of relative internucleotide bonds cleavage rates for different nucleotides were similar regardless of particular method used to consider the position effect