

## УЛЬТРАЗВУКОВОЕ РАСЩЕПЛЕНИЕ ДНК: АНАЛИЗ СТРУКТУРНО-ДИНАМИЧЕСКИХ ХАРАКТЕРИСТИК РЕГУЛЯТОРНЫХ УЧАСТКОВ ГЕНОМА И ОШИБОК СЕКВЕНИРОВАНИЯ

© 2020 г. И.А. Ильичева\*, М.В. Ходыков\*, Л.А. Панченко\*\*, Р.В. Полозов\*\*\*, Ю.Д. Нечипуренко\*

\*Институт молекулярной биологии им. В.А. Энгельгардта РАН, 119991, Москва, ул. Вавилова, 32

\*\*Биологический факультет Московского государственного университета имени М.В. Ломоносова, 119991, Москва, Ленинские горы, 1/12

\*\*\*Институт теоретической и экспериментальной биофизики РАН, 142290, Пушкино Московской области, ул. Институтская, 3

E-mail: imb\_irina@rambler.ru, nech99@mail.ru

Поступила в редакцию 27.02.2020 г.

После доработки 27.02.2020 г.

Принята к публикации 28.02.2020 г.

Приведен краткий обзор одной из сторон многогранной научной деятельности выдающегося ученого-экспериментатора С.Л. Гроховского — открытия им нового физико-химического явления специфичности ультразвукового расщепления межнуклеотидных связей в двухспиральной В-ДНК. Это явление позволило связать эффекты расщепления с конформационно-динамическими характеристиками нуклеотидных последовательностей в В-ДНК и дало возможность проводить биоинформатический анализ геномов на новой основе. Функциональные участки ДНК обладают определенными конформационно-динамическими характеристиками, что позволяет ставить и решать задачи эпигенетики и канцерогенеза.

*Ключевые слова:* ультразвуковое расщепление ДНК, конформационно-динамические характеристики ДНК, специфичность расщепления ДНК, эпигенетика, биоинформатический анализ генома.

DOI: 10.31857/S0006302920030096

*Посвящается памяти Сергея Львовича Гроховского*

*Жизнь человека есть самодвижущийся круг, который из бесконечно малого расходится во все стороны, в новые и все большие круги, не имеющие конца.*

*Р.У. Эмерсон*

В 2018 г. ушел из жизни удивительный человек и выдающийся экспериментатор — Сергей Львович Гроховский. Талантливый химик-синтетик, он синтезировал большое число различных лигандов, специфично связывающихся с ДНК. Эти работы широко известны. Они дали основание для построения моделей связывания противоопухолевых соединений и других лекарственных средств с ДНК [1–3]. Гроховский также довел до совершенства методы ультразвукового футпринтинга ДНК. Две последние его работы [4, 5] — яркое тому свидетельство. Они были опубликованы уже после того, как автора не стало.

В 2006 г. виртуозное владение экспериментом позволило Сергею Львовичу при облучении двойной спирали ДНК ультразвуком обнаружить зависимость частоты расщепления межнуклеотидных связей от нуклеотидной последовательности [6].

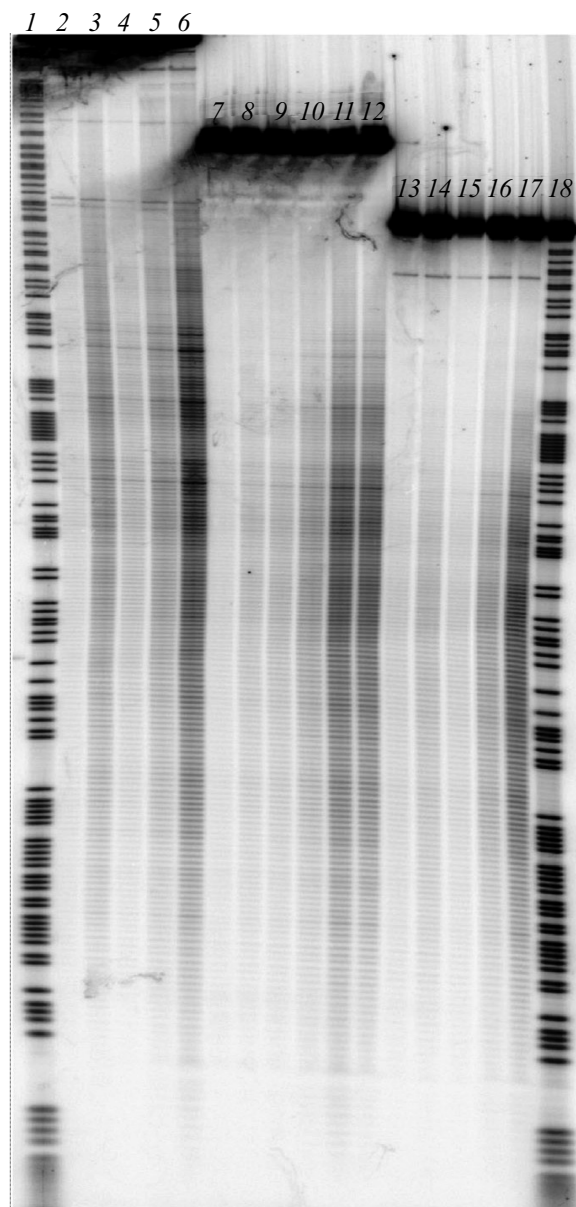
Ранее считалось, что позиции ультразвукового расщепления межнуклеотидных связей в двойной спирали ДНК расположены вдоль ДНК случайно и не зависят от нуклеотидной последовательности. Точность определения нуклеотидных последовательностей геномов методами NGS (next generation sequencing — секвенирование но-

*Сокращения:* Ключевые слова: ультразвуковое расщепление ДНК, конформационно-динамические характеристики ДНК, специфичность расщепления ДНК, эпигенетика, биоинформатический анализ генома.

вого поколения) существенно зависит от справедливости этого предположения. Гроховский провел детальное исследование процесса ультразвукового расщепления ДНК. Он облучал рестриктные фрагменты плазмид ультразвуком с частотой 22 и 44 кГц. Разработанная им схема эксперимента включала два этапа: а) предварительную пришивку на 3'-конец рестриктного фрагмента плазмидной ДНК одного (любого из четырех возможных) дополнительного нуклеотида, который был мечен по фосфору ( $^{33}\text{P}$ ); б) анализ полученных фрагментов расщепления при помощи электрофореза в полиакриламидном геле. Полученный результат был неожиданным и противоречил принятым представлениям о случайности позиций разрывов межнуклеотидных связей. Электрофорез меченых по 3'-концу фрагментов ДНК плазмид фиксировал в полиакриламидном геле полосы почернения разной интенсивности. Каждая из полос отвечала фрагменту известной нуклеотидной последовательности (рис. 1).

Методики, которые были разработаны для выявления влияния нуклеотидной последовательности на частоту расщепления плазмидных ДНК, детально описаны в работах [7–11]. Оцифровка интенсивностей полос на гелях позволила получить значения относительных частот расщепления динуклеотидов ( $R$ ) и тетрануклеотидов ( $T$ ) и провести статистический анализ позиций расщепления [7, 8]. Здесь мы приводим значения этих величин в таблице и на рис. 2 соответственно. Эти данные показывают различия частот ультразвуковых разрывов в зависимости от типа динуклеотида и его ближайшего окружения (нуклеотидного контекста).

Мы проанализировали различные версии объяснения явления зависимости частот ультразвукового расщепления ди- и тетрануклеотидов от последовательности нуклеотидов при облучении раствора двуспиральной ДНК. Многие физические свойства двойной спирали ДНК зависят от ее нуклеотидной последовательности. Термодинамическая стабильность дуплексов ДНК описывается в рамках модели «ближайших соседей». Согласно этой модели значения свободных энергий стэкинга дуплексов динуклеотидов  $\text{CpG}$  и  $\text{GpC}$  очень близки. Однако частоты расщепления этих динуклеотидов существенно различаются. Другая версия – влияние резких изгибов оси двойной спирали (например, появление кинков) также не объясняет эти отличия. Такие изгибы наиболее часто встречаются на участках с высоким содержанием АТ-пар, тогда как частоты расщепления ДНК на этих участках близки средним значениям. В этой связи важно также отметить, что совершенно иной характер ультразвукового расщепления наблюдается при облучении повре-



**Рис. 1.** Результат расщепления фрагмента ДНК в 6%-м денатурирующем полиакриламидном геле после обработки ультразвуком с частотой 22 кГц: дорожки 1, 18 – химическое расщепление по пуринам; дорожки 2, 7 и 13 – облучение фрагментов в течение 2 мин; дорожки 3, 8 и 14 – облучение фрагментов в течение 2 мин в присутствии 50% глицерола; дорожки 4, 9 и 15 – облучение фрагментов в течение 4 мин; дорожки 5, 10 и 16 – облучение фрагментов в течение 8 мин; дорожки 6, 11 и 17 – облучение фрагментов в течение 16 мин; дорожка 12 – облучение фрагментов в течение 16 мин в присутствии 0.5 М тиомочевины.

жденной ДНК, которая содержит одноцепочечные разрывы («ники»). Эти разрывы вызывают резкие изгибы оси спирали. В этом случае характерная для регулярной двуспиральной ДНК специфичность расщепления отсутствует [12].

Относительные частоты ультразвукового расщепления динуклеотидов

Динуклеотид	$N$	$R$	$d$
AA	1636	0.919	0.129
AC	1076	0.913	0.128
AG	1028	0.900	0.124
AT	1374	0.904	0.119
CA	1265	1.160	0.209
CC	1141	1.007	0.144
CG	1230	1.444	0.334
CT	1077	1.130	0.198
GA	1153	0.970	0.133
GC	1317	0.954	0.146
GG	1168	0.922	0.145
GT	1101	0.952	0.126
TA	1065	0.973	0.12
TC	1173	0.912	0.131
TG	1305	0.979	0.126
TT	1672	0.932	0.127

Примечание. Используются обозначения:  $N$  – объем выборки;  $R$  – среднее значение;  $d$  – стандартное отклонение выборочного среднего.

Мы проанализировали литературные данные, включающие ЯМР-исследования малых фрагментов ДНК, рентгеноструктурные работы и конформационные расчеты ДНК [8]. Связь частот ультразвукового расщепления разных ди- и тетра-нуклеотидов с особенностями их локальной конформационной динамики удалось обнаружить при анализе данных ЯМР и соотношении их с особенностями карт потенциальной энергии нуклеозидов и нуклеотидов [8]. Конформации цикла  $\beta$ -D-дезоксирiboзы в двойной спирали В-ДНК находятся в динамическом равновесии. Положения атомов кольца по отношению к его средней плоскости постоянно меняются. Эти конформационные движения, называемые псевдо-вращением (переходы  $N \leftrightarrow S$  между двумя низкоэнергетическими областями), сопряжены с

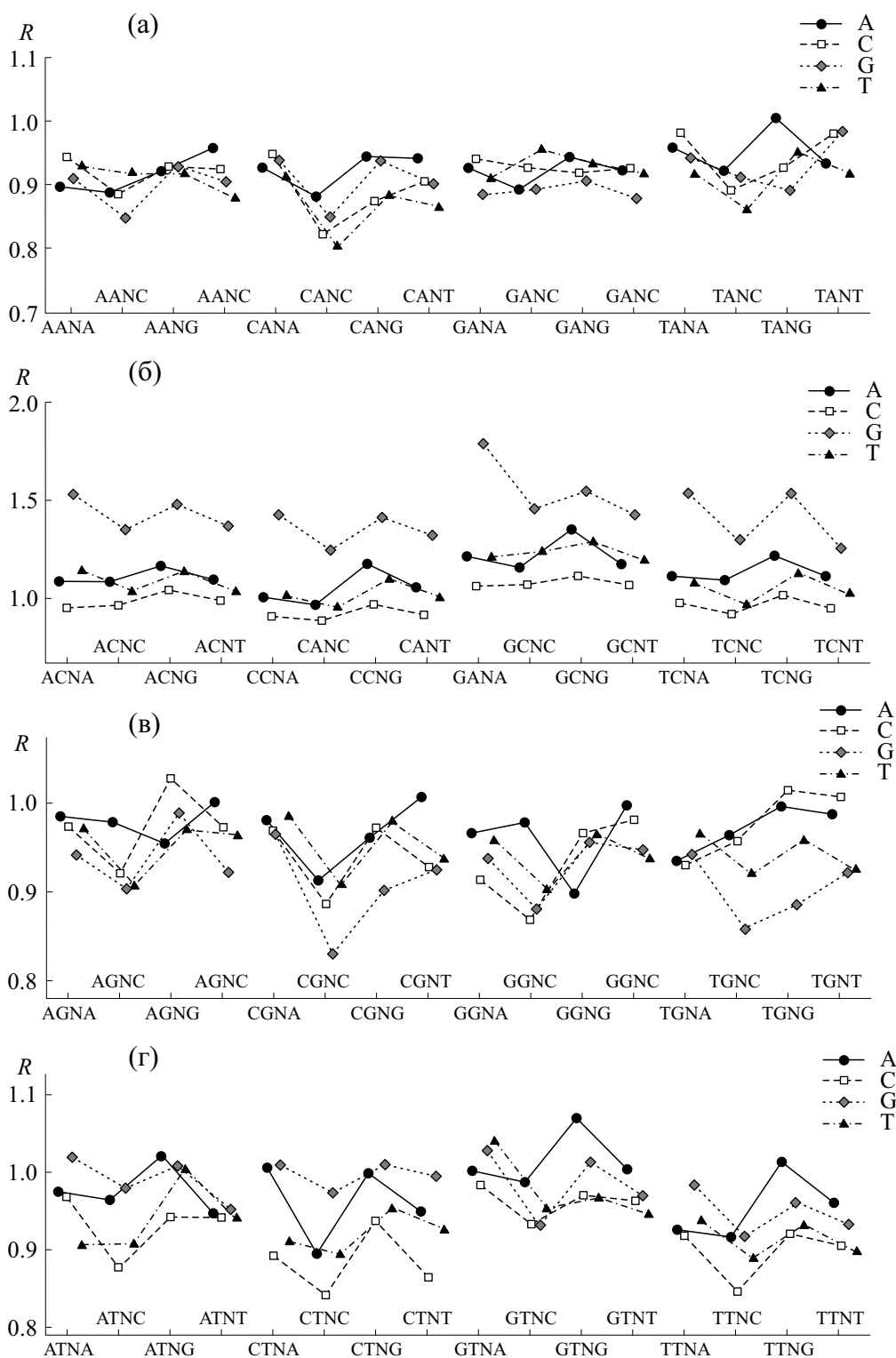
конформационными движениями вокруг связей  $C3'-O3'$  и  $O3'-P$  и с движениями оснований вокруг гликозидной связи. Низкоэнергетические области ( $S$ , *high-anti*) совпадают у трех дезоксирибонуклеотидов ( $dA$ ,  $dG$ ,  $dT$ ) и отличаются от низкоэнергетической области  $dC$ , которая смещена в сторону ( $N$ , *anti*). Именно после дезоксицитидина ультразвуковой разрыв межнуклеотидных связей происходит наиболее часто. Мы предположили, что частоты разрыва межнуклеотидной связи зависят от времени пребывания дезоксирибозы в  $N$ -области псевдо-вращения.

Ультразвуковое расщепление ДНК имеет механохимическую природу. Физическая модель, которая описывает динамику растяжения фрагмента ДНК, оказавшегося в потоке жидкости с высоким градиентом скоростей в результате коллапса кавитационных пузырьков, возникающих при ультразвуковом облучении, была построена в работе [13]. У фрагмента ДНК, оказавшегося вблизи схлопывающегося кавитационного пузырька, глобальная ось двойной спирали ориентируется вдоль потока. Вероятность механохимического расщепления связи  $C3'-O3'$  зависит от степени ее растяжения в потоке. Поэтому важным фактором в процессе расщепления является ориентация этой связи по отношению к глобальной оси спирали ДНК.

Предложенная модель дает качественное описание расщепления ДНК. Однако количественное описание механизма расщепления ДНК – все еще нерешенная задача. Трудности ее решения заключаются в том, что мы не знаем в полной мере систему определяющих параметров (их тип и число), характеризующих процесс расщепления. Как известно, нахождение системы определяющих параметров представляет очень важный этап постановки задачи.

Важно отметить, что частоты расщепления асимметричных комплементарных ди- и тетра-нуклеотидов различны. Например, значение  $R$  для динуклеотида  $CpA$  равно 1.13, а для комплементарного ему динуклеотида  $TrG$  оно равно 0.90; значение  $T$  для тетра-нуклеотида  $ApCpGpA$  равно 1.54, а для комплементарного ему тетра-нуклеотида  $TrCpGpT$  – 1.26. Это означает, что величины относительных частот расщепления динуклеотидов позволяют охарактеризовать каждую из комплементарных цепей в отдельности [8].

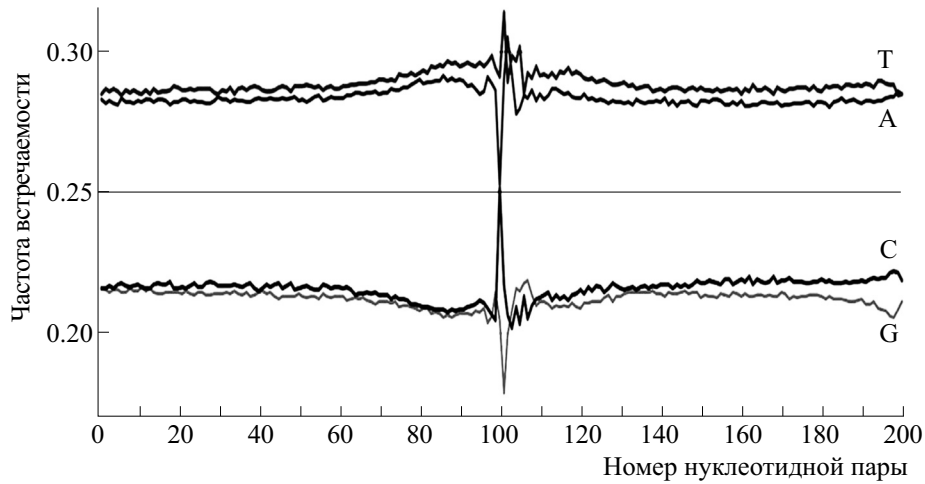
Далее нами был проведен анализ фрагментации геномных ДНК разными методами (ультразвук, небулизация, Коварис). Основой механизма расщепления ДНК всеми этими методами является механохимическая реакция. Библиотеки фрагментов для NGS-секвенирования доступны, и мы воспользовались ими, чтобы определить частоты разных позиций расщепления ДНК [14]. Частоты встречаемости каждого из четырех нук-



**Рис. 2.** Средние значения относительных частот ультразвукового расщепления рестриктных фрагментов плазмидных ДНК.  $R$  - частота расщепления, по оси абсцисс представлены тетрануклеотиды.

леотидов на участке в 200 пар ( $\pm 100$  пар) с центром в позиции фрагментации показаны на рис. 3. Нуклеотидные последовательности фраг-

ментов можно наложить на нуклеотидную последовательность референсного генома и таким образом оценить частоты разрывов в ди- и тетра-



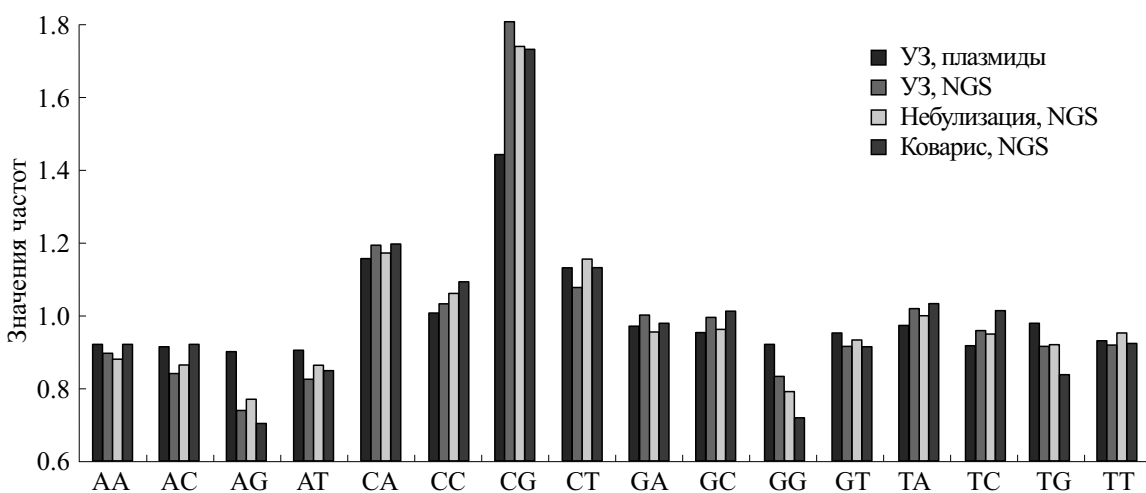
**Рис. 3.** Частоты встречаемости каждого из четырех дезоксирибонуклеотидов (A, C, G и T) на участке в 200 нуклеотидных пар ( $\pm 100$  нуклеотидных пар) с центром в позиции расщепления.

нуклеотидах при фрагментации геномной ДНК. Нами проведено сравнение относительных частот расщепления 16 динуклеотидов, полученных на основании данных NGS-фрагментации геномных ДНК этими тремя методами, с относительными частотами, полученными в результате анализа данных ультразвукового расщепления плазмидных ДНК (рис. 4) [14, 15].

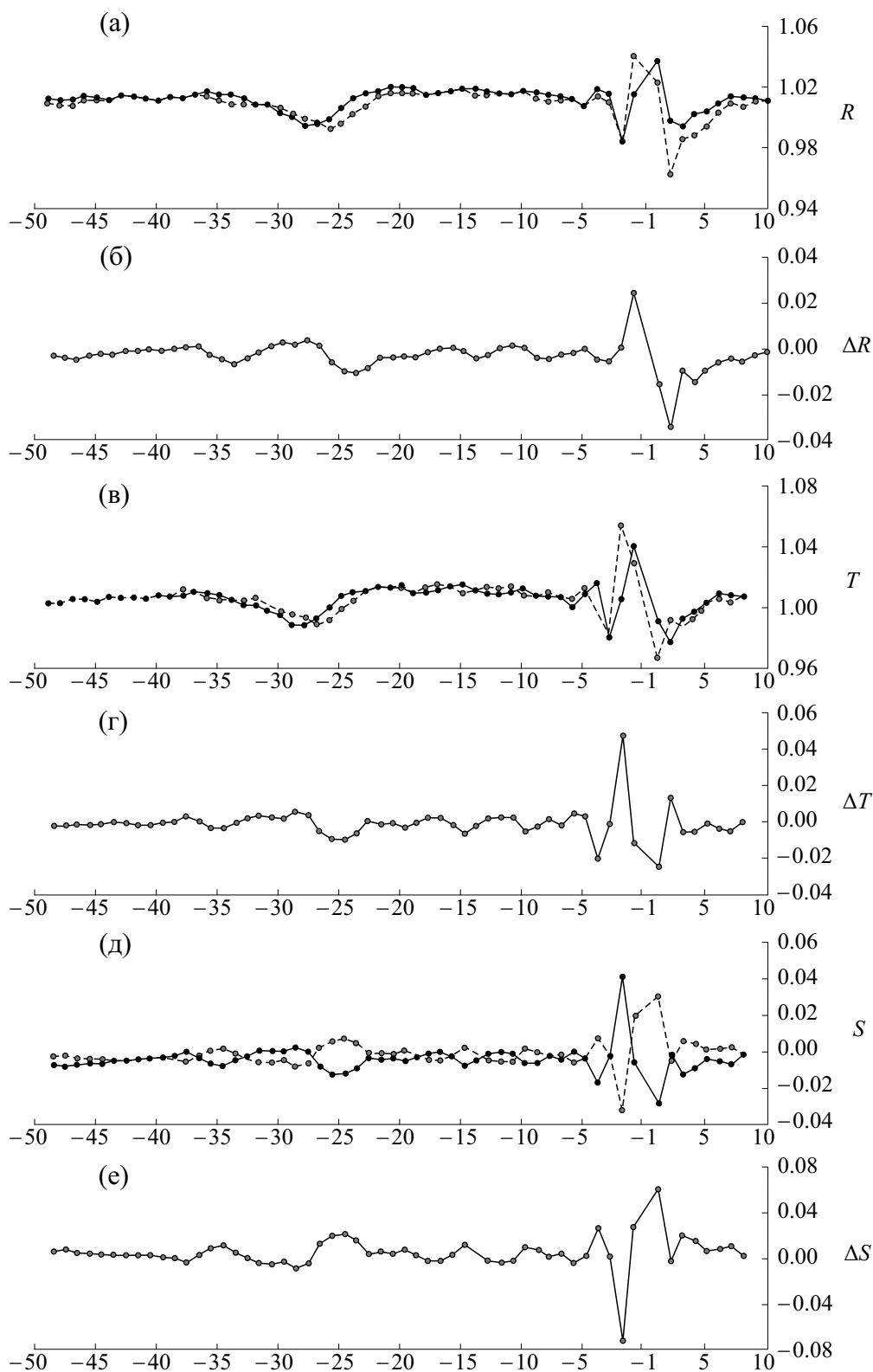
В геномной ДНК расщепление показывает примерно такую же зависимость от последовательности нуклеотидов, что и в плазмидных ДНК. Единственный динуклеотид, частота расщепления которого в геномной ДНК значительно выше, чем в плазмидных ДНК, — это динуклеотид CpG. Возможно, это связано с тем, что в геномных ДНК определенная часть CpG-динуклеотидов метилирована. Наличие ме-

тильной группы в 5-м положении кольца цитозина играет важную роль в эпигенетике. Мы предположили, что метилированные динуклеотиды CpG расщепляются чаще, чем неметилированные [15]. Это предположение было подтверждено в работах [16, 17], где проведен анализ относительных частот встречаемости метилированного и неметилированного CpG-динуклеотидов в позициях расщепления геномных ДНК. Для этого были использованы данные бисульфитного секвенирования. Оказалось, что частота расщепления метилированных CpG примерно в полтора раза выше неметилированных.

Величины  $R$  и  $T$  характеризуют структурно-конформационную динамику локального участка ДНК, далее мы будем называть их характеристиками СКД. Мы полагаем, что структурно-кон-



**Рис. 4.** Значения средних частот расщепления 16 динуклеотидов, полученных на основании данных NGS-фрагментации геномных ДНК тремя методами, в которых используется ультразвуковое облучение, и значения относительных частот ультразвукового облучения плазмидных ДНК.



**Рис. 5.** Профили СКД-характеристик промоторов генома человека: (а) — профили величины  $R$  в кодирующей (сплошная линия) и некодирующей (пунктирная линия) цепях; (б) — разность величин  $R$  для кодирующей и некодирующей цепей; (в) — профили величины  $T$  в кодирующей (сплошная линия) и некодирующей (пунктирная линия) цепях; (г) — разность значений  $T$  для кодирующей и некодирующей цепей; (д) — профили величины  $S$  в кодирующей (сплошная линия) и некодирующей (пунктирная линия) цепях; (е) — разность величин  $S$  кодирующей и некодирующей цепей.

формационная динамика участков геномных ДНК является важным фактором регуляции экспрессии генов. Для проверки этого предположения мы воспользовались базой EPD (New section of the Eukaryotic Promoter Database, <http://epd.vital-it.ch>), которая содержит нуклеотидные последовательности различных организмов. Вначале мы проанализировали профили последовательностей нуклеотидов в промоторах, рассматриваемые как их текстовые характеристики, и обнаружили в них существенные различия [18]. Затем мы построили так называемые «обобщенные профили» изменений шести механических и структурных параметров (энергии стэкинга, параметров Roll и Slide, изменения энергии дуплексов при отклонении от равновесных значений, а также энергии изгиба двойной спирали в сторону узкого желоба). Эти параметры характеризуют локальные особенности структуры ДНК (использовалась база данных <http://diprodb.fli-leibniz.de>). Термином «обобщенные профили» мы называем профили, полученные суммированием профилей индивидуальных промоторов, которые представлены в базе организмов определенного вида. Также мы построили «обобщенные профили» частот расщепления ди- и тетра-нуклеотидов – СКД-характеристики  $R$  и  $T$ , и ввели параметр  $S = (T - R)/T$ , величина которого характеризует влияние ближайших соседей динуклеотида (контекст) на его механохимическое расщепление. Поскольку мы связываем интенсивность механохимического расщепления с конформационными движениями, то полагаем, что при  $S < 0$  контекст ослабляет их интенсивность, а при  $S > 0$  усиливает. Профили СКД-характеристик ДНК строили как для кодирующей цепи, так и для комплементарной цепи. Все они представлены в работе [18]. Эти профили позволили сделать следующий важный вывод: несмотря на различия нуклеотидных последовательностей промоторов у разных организмов, при рассмотрении их, как текстов, механические, структурные и конформационные характеристики двойной спирали ДНК на этих участках сходны. Эти характеристики позволяют выделить промоторы среди других участков нуклеотидных последовательностей ДНК. Наиболее заметны особенности профилей в окрестности начала транскрипции, которая может захватывать от шести до восьми нуклеотидных пар. Все характеристики на этом участке меняются скачкообразно на каждом шаге, что свидетельствует о потере регулярности двойной спирали на этом участке. Отсутствие строгой регулярности облегчает перестройку структуры при связывании регуляторных белков и формировании «открытого комплекса», необходимого для транскрипции.

Другим участком, выделяющимся на профилях СКД характеристик последовательностей промоторов, является так называемый ТАТА-

бокс, с которым непосредственно связывается регуляторный белок ТВР – окрестности позиции (–30) относительно начала транскрипции. Для всех организмов характер изменений структурных и физико-химических параметров на этом участке схож. Значения механических и структурных характеристик указывают на расширение малой бороздки и увеличение жесткости двойной спирали на изгиб в сторону малой бороздки. Детальное описание особенностей профилей для каждой из характеристик дано в работе [18]. На рис. 5 приведены профили характеристик  $R$ ,  $T$ ,  $S$  промоторов человека. Позиции от –32 до –24 пар оснований относительно начала транскрипции, т.е. позиции ТАТА-бокса отличаются малыми значениями  $R$  и  $T$ , что может свидетельствовать об уменьшении интенсивности конформационных движений в цепях ДНК на этом участке. Профиль характеристики  $S$  указывает на периодические изменения интенсивностей конформационных движений в обеих комплементарных цепях, которые происходят в противофазе вплоть до позиции –3 пары оснований. В позиции ТАТА-бокса различие в амплитудах колебаний значений  $S$  у комплементарных цепей наибольшее. Итак, СКД-характеристики нуклеотидной последовательности геномов указывают на характерные структурные и конформационные особенности двойной спирали в промоторных областях.

В последнее время появились многочисленные указания на ошибки в результатах секвенирования геномов [19, 20]. Большинство из них связано с неточной оценкой встречаемости GC-пар оснований. Заметное увеличение частот механохимического расщепления после дезоксицитидина в двойной спирали может быть одним из источников погрешностей получаемых нуклеотидных последовательностей. Разработка методов, которые могли бы учитывать частоты встречаемости нуклеотидов на 5'-концах фрагментов («ридов»), используемых при NGS-секвенировании, и проводить на их основе корректировку результатов – это еще одно направление, которое стало возможным благодаря открытию С.Л. Гроховским явления специфичного к нуклеотидной последовательности механохимического расщепления ДНК.

Открытие, сделанное Гроховским, дает новые возможности развития и разработки методов изучения особенностей структуры и конформационной динамики ДНК (см., например, работу [21]), которые влияют на генную экспрессию и, тем самым, определяют такие процессы, как канцерогенез и эпигенез.

## БЛАГОДАРНОСТИ

Авторы считают своим приятным долгом поблагодарить Д.Ю. Нечипуренко и Л.А. Урошлеву за полезные обсуждения.

## ФИНАНСИРОВАНИЕ РАБОТЫ

Работа поддержана Программой фундаментальных исследований государственных Академий наук на 2013–2020 годы (темы №№ 01201363818 и 01201363820) и Программой по молекулярной и клеточной биологии Президиума РАН № 01201456592, а также Российским фондом фундаментальных исследований (проект № 18-07-00354а).

## КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

## СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит описания каких-либо исследований с использованием людей и животных в качестве объектов.

## СПИСОК ЛИТЕРАТУРЫ

- G. V. Gursky, et al., Cold Spring Harbor Symp. Quant. Biol. **47**, 367 (1983).
- А. С. Заседателев, А. Л. Жузе, К. Циммер, С. Л. Гроховский и др., Докл. АН СССР **231** (4), 1006 (1976).
- S. L. Grokhovsky, A. N. Surovaya, G. Burckhardt, et al., FEBS Lett. **439** (3), 346 (1998).
- С. Л. Гроховский, Молекуляр. биология **52** (4), 705 (2018).
- С. Л. Гроховский, Молекуляр. биология **52** (5), 846 (2018).
- С. Л. Гроховский, Молекуляр. биология **40** (2), 317 (2006).
- С. Л. Гроховский, И. А. Ильичева, Д. Ю. Нечипуренко и др., Биофизика **53** (3), 417 (2008).
- S. L. Grokhovsky, I. A. Il'icheva, D. Yu. Nechipurenko, et al., Biophys. J. **100** (1), 117 (2011).
- Ю. Д. Нечипуренко, М. В. Головкин, Д. Ю. Нечипуренко и др., Журн. структур. химии **50** (5), 1040 (2009).
- S. L. Grokhovsky, I. A. Il'icheva, D. Yu. Nechipurenko, et al., in *Gel Electrophoresis – Principles and Basics*, ed. by S. Magdeldin (InTech, 2012).
- S. L. Grokhovsky, I. A. Il'icheva, D. Yu. Nechipurenko, et al., in *Advances in Engineering Research*, ed. by V. M. Petrova (Nova Science Publishers, 2014), v. 8, pp. 213–236.
- I. A. Il'icheva, D. Y. Nechipurenko, and S. L. Grokhovsky, J. Biomol. Structure and Dynamics **27** (3), 391 (2009).
- Д. Ю. Нечипуренко, И. А. Ильичева, М. В. Ходыков и др., Биофизика **59** (6), 1061 (2014).
- M. S. Poptsova, I. A. Il'icheva, D. Y. Nechipurenko, L. A. Panchenko, et al., Sci. Rep. **4**, 4532 (2014).
- Ю. Д. Нечипуренко, И. А. Ильичева, М. С. Попцова и С. Л. Гроховский, Актуальные вопросы биологической физики и химии **2** (1), 339 (2017).
- Ю. Д. Нечипуренко, И. А. Ильичева, Э. Т. Абдуллаев и др., Актуальные вопросы биологической физики и химии **3** (4), 884 (2018).
- L. A. Uroshlev, E. T. Abdullaev, I. R. Umarova, et al., Sci Rep. (2020) (in press).
- I. A. Il'icheva, M. V. Khodikov, M. S. Poptsova, et al., BMC Genomics **17** (1), 973 (2016).
- Y. Benjamini and T. P. Speed, Nucleic acids research, **40** (10), e72 (2012).
- J. R. Wang, B. Quach, and T. S. Furey, BMC Bioinformatics **18** (1), 357 (2017).
- D. Semyonov, Yu. Nechipurenko, Bioessays (2020) (in press).

## Ultrasonic DNA Cleavage: Analysis of Conformational-Dynamic Features of Regulatory Regions in the Genome and Sequencing Errors

I.A. Il'icheva\*, M. V. Khodikov\*, L.A. Panchenko\*\*, R.V. Polozov\*\*\*, and Yu.D. Nechipurenko\*

\*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, ul. Vavilova 32, Moscow, 119991 Russia

\*\*Faculty of Biology, Lomonosov Moscow State University, Leninskie Gory 1/12, Moscow, 119991 Russia

\*\*\*Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Institutskaya ul. 3, Pushchino, Moscow Region, 142290 Russia

The article is dedicated in memory of Dr. Sergei Lvovich Grokhovsky, a distinguished scientist and an experimenter. This paper contains a brief review of one aspect of the multifaceted scientific activity of Dr. S.L. Grokhovsky whose work contributed immensely to the understanding of a new physicochemical event such as the specificity of ultrasound-induced cleavage of internucleotide bonds in double-stranded B-DNA. His discovery has afforded a view of the cleavage effects within the conformational and dynamic characteristics of nucleotide sequences in B-DNA and an opportunity to conduct bioinformatical analysis of genomes on a new plane. Functional DNA regions are characterized by distinct conformational-dynamic features. We demonstrate the effectiveness of this approach for the analysis of correlation of these conformational and dynamic features with the epigenetic development and carcinogenesis.

*Keywords: ultrasonic DNA cleavage, conformational-dynamic characteristics of DNA, specificity of DNA cleavage, epigenetics, bioinformatics analysis of genome data*