

## О ПРИРОДЕ ДОМИНИРОВАНИЯ ОЛИГОМЕРНЫХ (dA:dT)<sub>n</sub> ТРЕКОВ В СТРУКТУРЕ ГЕНОМОВ ЭУКАРИОТ

© 2016 г. А.А. Самченко, С.С. Киселев, А.В. Кабанов,  
М.С. Кондратьев, В.М. Комаров

Институт биофизики клетки РАН, 142290, Пуцино Московской области, ул. Институтская, 3  
E-mail: komarov@icb.psn.ru

Поступила в редакцию 11.07.16 г.

Проведен сравнительный частотный анализ встречаемости олигомерных W- и S-треков в геномах различных видов эукариот, отличающихся своим GC-составом. В исследованных эукариотических геномах – *Dictyostelium discoideum* (GC ~ 25,7%), *Caenorhabditis elegans* (GC ~ 36,9%), *Arabidopsis thaliana* (GC ~ 38,0%), *Drosophila melanogaster* (GC ~ 38,8%), *Homo sapiens* (GC ~ 40,0%), *Gallus gallus* (GC ~ 50,0%), *Leishmania major* (GC ~ 59,1%) – было выявлено доминирование мононуклеотидных и смешанных (A/T)<sub>n</sub>-последовательностей по сравнению с (G/C)<sub>n</sub>-последовательностями. С использованием результатов квантово-химических расчетов дано теоретическое обоснование важной роли имеющихся различий в исходном, скрытом структурном полиморфизме водородного связывания уотсон-криковских AT- и GC-пар в формировании этого явления. Сделано заключение о том, что пониженный структурный полиморфизм AT-пар по сравнению с высоким, четырехкратным полиморфизмом основного состояния GC-пар может выступать наиболее вероятным источником предпочтительности и надежности «использования» природой AT-пар в геномной ДНК многих видов организмов.

*Ключевые слова:* геномы эукариот, доминирование AT-пар, квантово-химический расчет.

Считается, что одним из ключевых параметров структурно-функциональной организации молекул нуклеиновых кислот является содержание гуанин-цитозиновых уотсон-криковских пар. При этом, хотя в геномах разных видов организмов GC-состав может заметно отличаться [1], тем не менее, в кодирующих последовательностях его процентный вклад всегда наблюдается повышенным по отношению к другим областям генома [2,3]. Происхождение этого явления многие склонны связывать с исходно большей термодинамической стабильностью GC-пар по сравнению с AT-парами, поскольку в их комплементарном спаривании принимают участие три, а не две водородные связи, как это имеет место в AT-парах. Известны многочисленные гипотезы относительно условий закрепления выделяемой особенности строения генома у живых организмов: устойчивость к УФ-радиации [4], термоадаптация [5], размер генома [6], длина кодирующей последовательности [7], давление окружающей среды [8], особенности мутационного процесса [9], скорость транскрипции [10].

Однако такой сравнительно упрощенный подход в установлении специфики структурной организации геномов как про-, так и эукариот

представляется нам весьма ограниченным и внутренне противоречивым.

Во-первых, в приводимых доводах о важной роли процентного состава GC-пар в организации генома, например у микроорганизмов, чаще всего акцентируют внимание на широком диапазоне наблюдаемых вариаций этого вклада – от 16 до 75% [11,12], что действительно впечатляет. Вместе с тем имеющаяся статистика распределения организмов по вкладам уотсон-криковских пар в структуре порядка 1700 прокариотических хромосом и плазмид показывает еще и другое. Максимум распределения приходится на те микроорганизмы, которым свойственен весьма высокий, порядка 65%, вклад уотсон-криковских AT-, а не GC-пар оснований в составе своих хромосом [13].

Во-вторых, для геномов эукариотических организмов уже давно известно, что равновесие между AT- и GC-парами оказывается сильно смещенным, как это ни странно, не в сторону GC-, а в сторону AT-пар (с превышением на 40–80%!) [14]. Как правило, это характерно для многоклеточных эукариот, чьи геномы являются AT-богатыми [15].

В-третьих, в связи с быстрым развитием технологий секвенирования появляется все больше и больше фактов, указывающих на свое-

образии вкладов последовательностей, содержащих только основания А или Т, либо только G или C, и часто обозначаемых как W- (weak) и S- (strong) треки [16], в организацию структуры генома [17,18]. Хотя функции большинства из найденных повторяющихся и уникальных нуклеотидных последовательностей четко пока не определены, тем не менее определенные корреляции в частотах появления W- и S-треков в кодирующих и в регуляторных областях геномов тоже наблюдаются [19–24].

Ранее нами для довольно большого ряда структур архейных и эубактериальных хромосом (порядка 400 представителей) впервые было показано [25,26], что в 75% случаев характерным оказывается преобладание  $oligo(dA)_n$ - и  $oligo(dT)_n$ -повторов над  $oligo(dG)_n$ - и  $oligo(dC)_n$ -повторами (где  $n \geq 5$ ), причем такое доминирование сохраняется даже для хромосом, в составе которых GC-пар явно больше половины (а именно, где содержание GC-пар варьирует от 50 до 65%). Похожие результаты обнаруживаются в литературе и для геномов некоторых организмов эукариот [27,28].

Вполне вероятно, что выявляемая во многих случаях тенденция к повышенной распространенности в молекулах ДНК треков с AT-парами, по крайней мере для треков короткой длины, отражает некую важную особенность организации генетических процессов как у прокариот, так и у эукариот, указывая тем самым на определяющую роль уотсон-криковского спаривания AT-оснований в формировании этих процессов, обусловленную, по-видимому, некоторым нетривиальным своеобразием внутреннего строения самих аденин-тиминных пар. В этом плане установление обобщенной картины доминирования последовательностей из AT-пар в структуре геномов разнообразных видов эукариот с последующим обоснованием наиболее вероятной физической первопричины возникновения такого явления представляется нам чрезвычайно актуальным.

## ЦЕЛЬ РАБОТЫ

Цель данного исследования двоякая. Во-первых, стояла задача выполнить анализ встречаемости W- и S-треков переменной длины в геномах эукариот, обладающих разным GC-составом, и показать, что доминирование олигомерных (A/T) $_n$ -последовательностей в структуре ДНК здесь действительно является общим явлением.

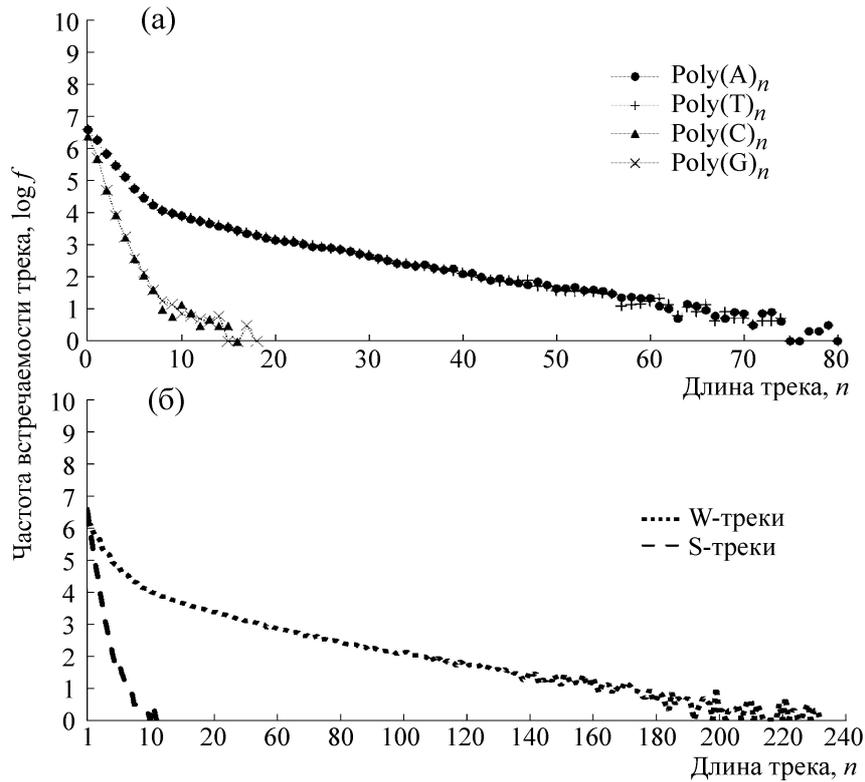
Во-вторых, используя результаты наших предыдущих теоретических работ о существо-

вании различий в неоднозначности (в структурном полиморфизме) формы водородного связывания уотсон-криковских AT- и GC-пар [29–32], мы пытались обосновать наиболее вероятный физический механизм спонтанного зарождения и накопления пространственных неоднородностей и дефектов в регулярной структуре двойной цепи ДНК. Основываясь на данных более строгих квантово-химических расчетов, мы предполагали подтвердить наличие исходной, пониженной неоднозначности пространственной формы комплементарного спаривания AT-пар по сравнению с полиморфизмом GC-пар, которая может выступать источником предпочтительности и надежности «использования» природой AT-пар в структурно-функциональной организации геномной ДНК любых организмов.

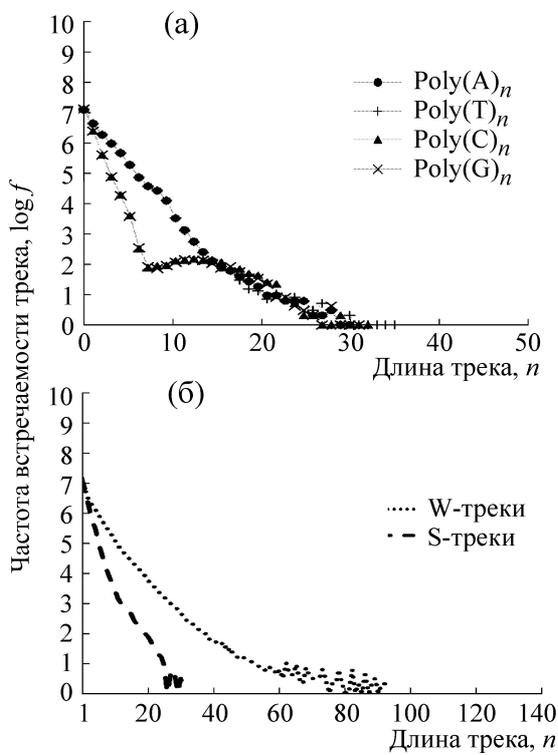
## РЕЗУЛЬТАТЫ

**Частоты встречаемости олигонуклеотидных треков.** Был выполнен анализ частот встречаемости удлиняющихся мононуклеотидных ( $A_n$ ,  $T_n$ ,  $G_n$ ,  $C_n$ ) и смешанных W- и S-треков в геномах различных видов эукариот, взятых из базы GenBank [33,34]. Были выбраны семь представителей эукариот с последовательно увеличивающимся процентным GC-составом генома: *Dictyostelium discoideum* (25,7%); *Caenorhabditis elegans* (36,9%); *Arabidopsis thaliana* (38,0%); *Drosophila melanogaster* (38,8%); *Homo sapiens* (40,0%); *Gallus gallus* (50,0%), *Leishmania major* (59,1%) [1]. Оценивались частоты появления указанных выше треков постепенно увеличивающейся длины  $n$  от 1 до  $n_{\max}$ , где  $n_{\max}$  – максимальная длина нуклеотидного трека, обнаруживаемого в геноме. Все расчеты по обработке данных геномных структур проводились нами на основе разработанной собственной компьютерной программы, особенности алгоритма которой описаны в разделе «ПРИЛОЖЕНИЕ».

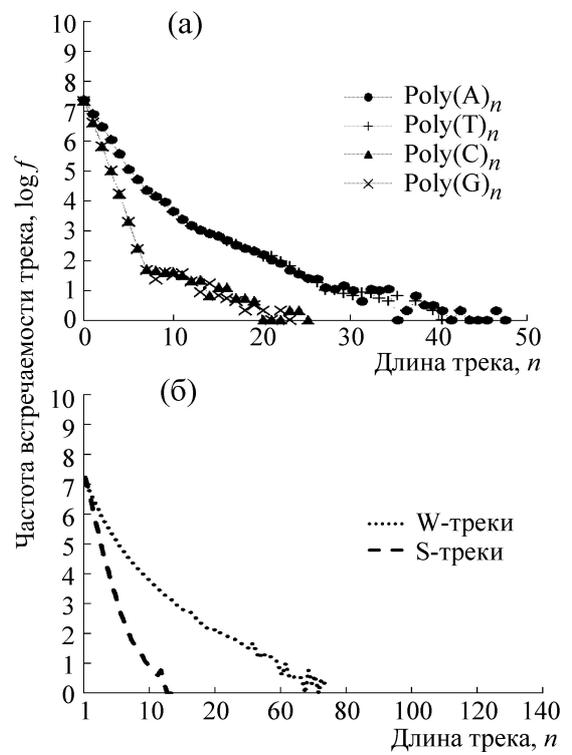
Было обнаружено, что в исследованных геномах эукариот частота встречаемости  $f$  мононуклеотидных ( $A_n$  и  $T_n$ ) и смешанных (A/T) $_n$  W-треков как функция длины трека практически всегда, как это и предполагалось нами, оказывается доминирующей над встречаемостью мононуклеотидных ( $G_n$  и  $C_n$ ) и смешанных (G/C) $_n$  S-треков (рис. 1–7, на представленных графиках из-за очень большого диапазона значений частот встречаемости величины  $f$  представлены в логарифмическом масштабе). Кроме того, в наиболее интересном и важном случае – случае поведения удлиненных W-треков смешанной природы в структуре ДНК – обнаруживаются повторы длиной в сотню, а то и в



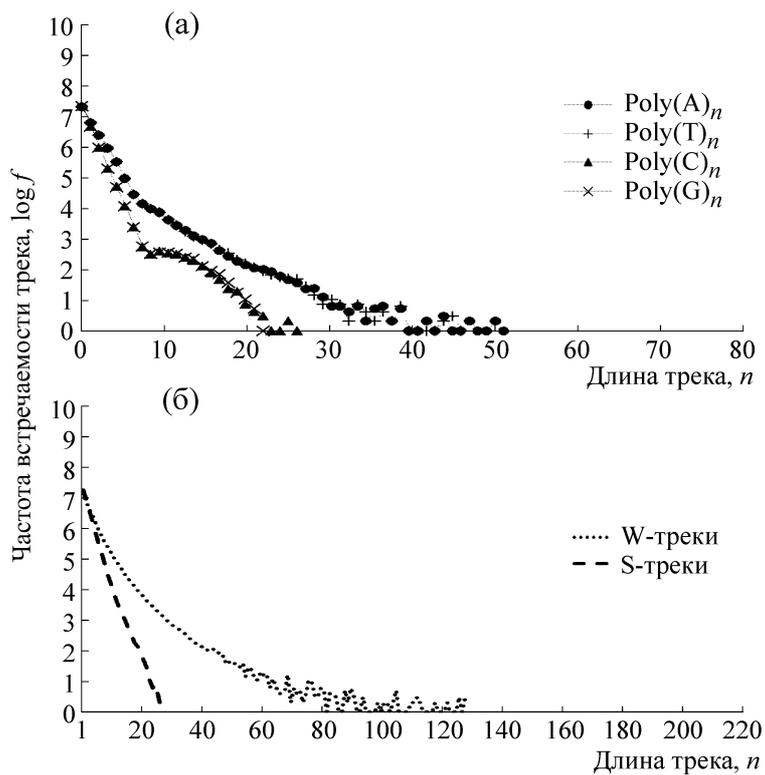
**Рис. 1.** Частоты встречаемости мононуклеотидных треков (а) и смешанных W- и S-треков (б) в геноме *Dictyostelium discoideum*.



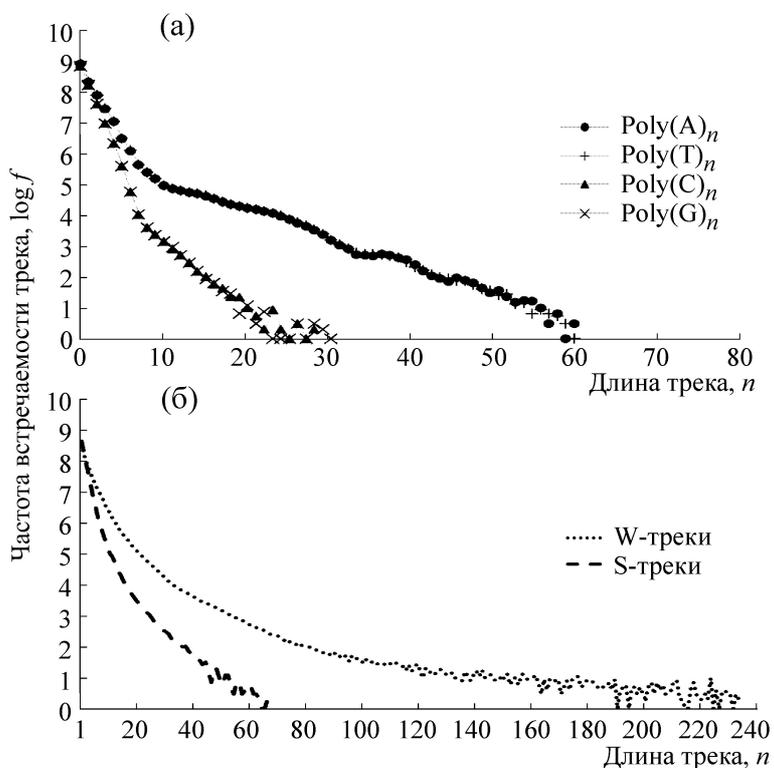
**Рис. 2.** Частоты встречаемости мононуклеотидных треков (а) и смешанных W- и S-треков (б) в геноме *Caenorhabditis elegans*.



**Рис. 3.** Частоты встречаемости мононуклеотидных треков (а) и смешанных W- и S-треков (б) в геноме *Arabidopsis thaliana*.



**Рис. 4.** Частоты встречаемости мононуклеотидных треков (а) и смешанных W- и S-треков (б) в геноме *Drosophila melanogaster*.



**Рис. 5.** Частоты встречаемости мононуклеотидных треков (а) и смешанных W- и S-треков (б) в геноме *Homo sapiens*.

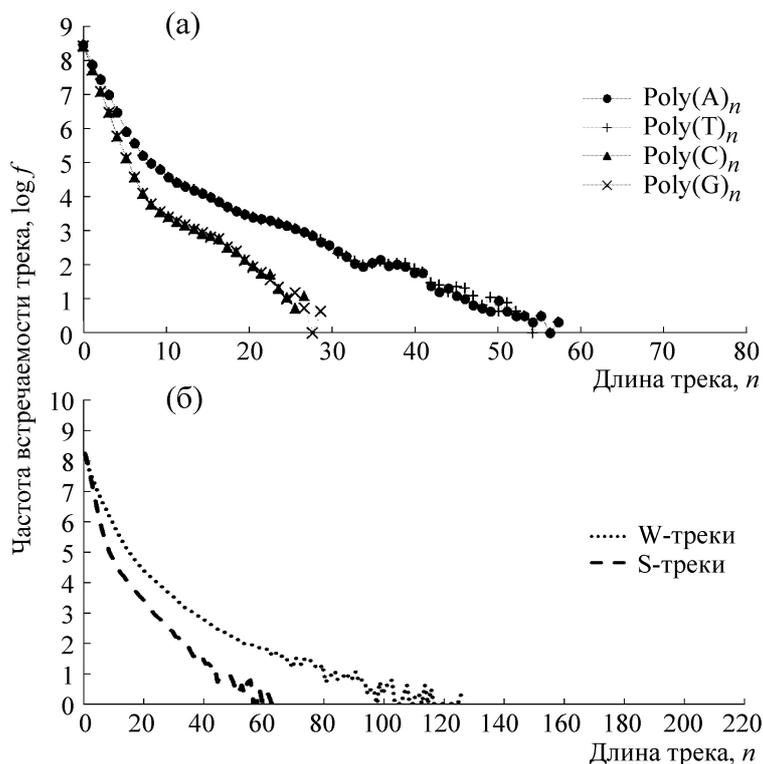


Рис. 6. Частоты встречаемости мононуклеотидных треков (а) и смешанных W- и S-треков (б) в геноме *Gallus gallus*.

несколько сотен уотсон-криковских АТ-пар нуклеотидов (пример геномов *D. discoideum* и *H. sapiens*, рис. 1 и 5). В то же время для смешанных S-треков частота появления удлиненных последовательностей сходится практически к нулю уже где-то в районе длин двух-трех десятков GC-пар! Такой резкий характер спада встречаемости протяженных последовательностей из пар GC-природы оказывается свойственным для всех исследованных нами геномов эукариот.

Весьма выразительно выглядит иллюстрация обнаруженного явления при конкретном количественном сопоставлении частот появления W- и S-треков одинаковой длины на примере генома человека. В табл. 1 представлены данные по частотам встречаемости  $\text{oligo}(\text{dA})_n$ - и  $\text{oligo}(\text{dT})_n$ -последовательностей и по аналогичным характеристикам у  $\text{oligo}(\text{dG})_n$ - и  $\text{oligo}(\text{dC})_n$ -нуклеотидных последовательностей в функции увеличения длины цепи  $n$ . В четвертом столбце представлены отношения этих частот (величина  $k$ ) для каждой длины  $n$  трека.

Как видно из табл. 1, даже для простых мононуклеотидных треков не столь большой длины, например для треков из 10 пар нуклеотидов, доминирование встречаемости аденин-

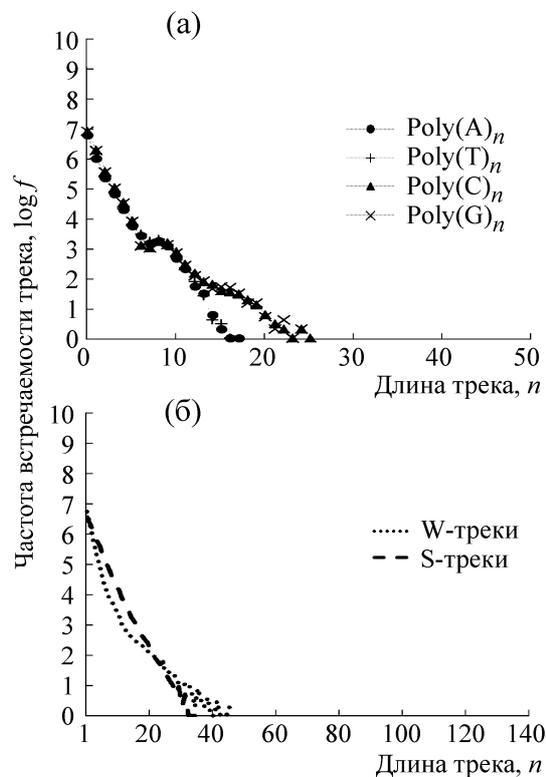


Рис. 7. Частоты встречаемости мононуклеотидных треков (а) и смешанных W- и S-треков (б) в геноме *Leishmania major*.

**Таблица 1.** Частоты встречаемости мононуклеотидных  $A_n(T_n)$ - и  $G_n(C_n)$ -треков в геноме человека

Длина трека $n$	Суммарная частота встречаемости мононуклеотидных $A_n$ - и $T_n$ -треков	Суммарная частота встречаемости мононуклеотидных $G_n$ - и $C_n$ -треков	Отношение частот встречаемости треков $k = f(A/T)_n / f(G/C)_n$
1	792380825	655948993	1,20
2	208810372	166761379	1,25
3	82093595	44115960	1,86
4	30787551	10747687	2,86
5	12026221	2548282	4,72
6	3585147	504420	7,11
7	1499230	77644	19,31
8	541642	15497	34,95
9	310016	5958	52,03
<b>10</b>	<b>201669</b>	<b>3556</b>	<b>56,71</b>
11	123264	2175	56,67
12	96609	1365	70,77
13	85154	814	104,61
14	75933	466	162,95
15	67816	261	259,83
16	58167	158	368,14
17	47052	107	439,74
18	37807	68	555,98
19	32056	47	682,04
<b>20</b>	<b>27191</b>	<b>26</b>	<b>1045,81</b>
21	24041	20	1202,05
22	21982	8	2747,75
23	19312	9	2145,78
24	16678	8	2084,75
25	13695	3	4565,00
26	10613	0	$\infty$
27	8315	6	1385,83
28	6682	1	6682,00
29	5116	5	1023,20
<b>30</b>	<b>3751</b>	<b>2</b>	<b>1875,50</b>
31	2384	1	2384,00
39	674	1	674,00
49	74	1	74,00
68	1	0	$\infty$
69	1	0	$\infty$
70	0	1	0,0
72	1	0	$\infty$
73	1	0	$\infty$
79	1	0	$\infty$
81	0	1	0,0
83	1	0	$\infty$
90	1	0	$\infty$

**Таблица 2.** Максимальные длины  $n_{\max}$  (п.н.), протяженных W- и S-треков в геномах эукариот (цифры в скобках – номера хромосом)

Треки	<i>D. discoideum</i>	<i>C. elegans</i>	<i>A. thaliana</i>	<i>D. melanogaster</i>	<i>H. sapiense</i>	<i>G. gallus</i>	<i>L. major</i>
W	440 (2)	684 (2)	186 (6)	936 (4)	1067 (16)	340 (2)	64 (26)
S	22 (2)	32 (3)	31 (4)	43 (2)	261 (8)	129(19)	38 (35)

тиминовых структур над гуанин-цитозинными уже оказывается значительным. Различие частот их появления составляет почти два порядка,  $k \approx 60$ . Для случая более протяженных треков, длиной  $n = 30$  п.н., эти отличия вырастают еще больше, увеличиваясь резко на несколько порядков,  $k \approx 1,9 \cdot 10^3$ ! Похожие закономерности в асимметрии встречаемости АТ- и GC-пар оказываются справедливыми, если обратиться к приведенному ранее рис. 5 и для треков смешанной природы. Полученные нами данные не противоречат наблюдениям по частотному анализу встречаемости олигонуклеотидных последовательностей в геноме человека hg 19, недавно представленным в работе [35].

Довольно интересными выглядят результаты для лейшмании (*L. major*), обладающей весьма большим процентным GC-содержанием (~60%). Для нее оказывается не столь явным доминирование смешанных последовательностей (A/T)<sub>n</sub>-типа над (G/C)<sub>n</sub>-треками, свойственное всем остальным эукариотам. Здесь, согласно рис. 7, частоты встречаемости S- и W-треков в районе длин до 30 пар нуклеотидов оказываются довольно близкими и только в асимптотике проявляется конкретное преимущество распространенности цепочек смешанного типа из АТ-пар.

Мы полагаем, что в данном случае нашло отражение всей специфики хромосомной организации *L. major* [33]. Прежде всего, и это главное, здесь исходно наблюдается значительный дефицит АТ-пар. Кроме того, протяженность самого генома невелика (~30 Мб) по сравнению с размерами других проанализированных нами геномов эукариот (геномы *Homo sapiense* и *Gallus gallus*, например, имеют размеры свыше 1200 Мб). По-видимому, наложение этих ограничений сивелировало общую картину асимметрии встречаемости АТ- и GC-пар в ДНК данного организма (рис. 7).

Эти же обстоятельства сказались, по нашему мнению, и на резком сокращении наблюдаемой здесь максимальной длины протяженных W-треков. Так, согласно расчетам (табл. 2), в *L. major* наибольшая длина W-трека составила всего 64 нуклеотидных АТ-пары. Это самый короткий из всех протяженных W-треков, най-

денных нами в ряду геномов эукариот. В геноме человека, например, имеющего размеры почти на два порядка больше, встречаются довольно протяженные W-треки, состоящие из 1067 пар нуклеотидов! Что касается предельного размера S-трека в геноме *L. major*, то он здесь также оказался малым и составил в длину всего 38 пар нуклеотидов.

Таким образом, все представленные в данном разделе результаты подтверждают общую картину повышенной распространенности АТ-треков по сравнению с треками GC-природы в геномах различных видов эукариот и указывают на существование некоей фундаментальной первопричины отличий в строении АТ- и GC-пар оснований, которая в структурно-функциональной организации молекулы ДНК порождает такие интересные масштабные, «накопительные» эффекты.

**Модель скрытого структурного полиморфизма уотсон-криковских АТ- и GC-пар.** Для объяснения возможной физической первопричины, лежащей в основе описанного выше явления, мы воспользовались результатами своих предыдущих теоретических работ по исследованию особенностей структурной организации изолированных нуклеотидных ДНК-комплексов. Ранее нами на основе полуэмпирических квантово-химических расчетов было выявлено существование скрытого структурного полиморфизма в водородном связывании нуклеотидов, если только в их Н-спаривании принимает участие хотя бы одна аминокислота [29–32]. Причина этого явления – в особенностях электронного строения экзоциклических аминокислотных оснований.

Как известно, атом азота NH<sub>2</sub>-группы в структуре молекул с алифатическими или сопряженными связями всегда реализуется в форме, близкой к sp<sup>3</sup>-гибридному, пирамидальному характеру строения своих валентных связей. Соответственно, азотистые основания исходно имеют склонность к неплоскому по аминокислотной группе строению, которое подтверждается экспериментально и воспроизводится во многих теоретических исследованиях [37–41]. Эта же «пирамидальность» строения аминокислотных оснований формирует также и некопланарность оди-

ночных уотсон-криковских АТ- и GC-пар [42,43].

В дополнение к этим обстоятельствам инверсионная бистабильность геометрии  $\text{NH}_2$ -группы определяет еще одну, более фундаментальную, на наш взгляд, особенность водородного связывания пар, а именно – указанную выше скрытую неоднозначность формы уотсон-криковского спаривания оснований. Оценки показали [29], что для АТ-пары характерным оказывается двукратное вырождение энергии основного состояния с реализацией геометрии правого или левого «пропеллерового» разворота плоскостей оснований за счет разного наклона N–H-связей аминогруппы. Для случая уотсон-криковских GC-пар выявляется повышенный, четырехкратный структурный полиморфизм водородного связывания оснований. Здесь из-за участия в процессе двух пирамидальных  $\text{NH}_2$ -групп возможна реализация в основном состоянии двух «пропеллероподобных» и двух «ступенькообразных» вариантов спаривания оснований.

Такое своеобразие водородных связей азотистых оснований, как отмечено в работе [41], приводит к тому, что в общем случае для известного круга канонических, неправильных и «хугстеновских» пар справедливой оказывается не одна единственная планарная геометрия их основного состояния [45], а 81 устойчивая форма H-спаривания оснований. Поэтому изначально высокая, четырехкратная неоднозначность формы уотсон-криковского связывания гуанина и цитозина может выступать, на наш взгляд, постоянным источником спонтанного зарождения неоднородностей и нарушения строгого пространственного изоморфизма пар в регулярной структуре двойной спирали ДНК.

Вполне вероятно, что GC-пары как менее «предсказуемый» элемент в воспроизведении исходной геометрии комплементарного спаривания, несмотря на повышенную энергию своего водородного связывания, были попросту «минимизированы» природой в организации геномов. Достигалась таким способом надежность протекания в клетке генетических процессов репликации, транскрипции – где как раз крайне важна однозначность сохранения пространственной формы комплементарного спаривания нуклеотидов.

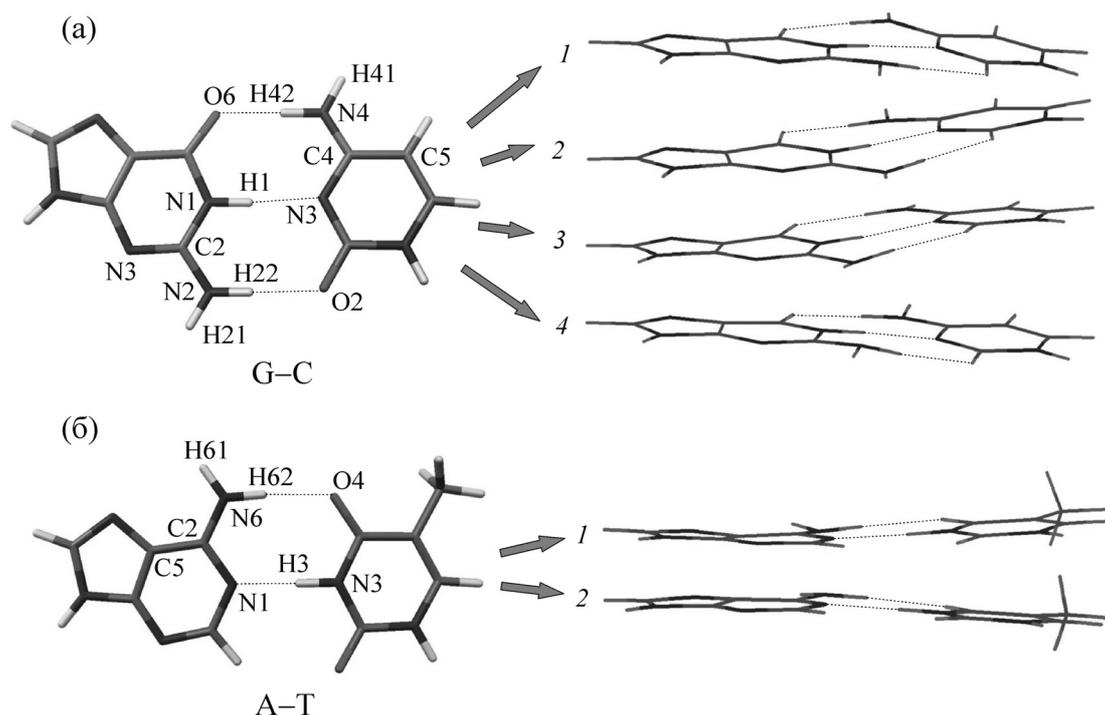
Прямые экспериментальные данные о некопланарности и неоднозначности формы изолированных АТ- и GC-пар в настоящее время, к сожалению, пока отсутствуют по причине слабой термоустойчивости оснований в газовой фазе. Однако существующие рентгеноструктур-

ные данные по монокристаллам АТ- и GC-пар [46–49], а также результаты газофазных спектральных исследований искусственного аналога уотсон-криковской пары 2-аминопиридин·2-пиридон [50] убедительно свидетельствуют, что по крайней мере некопланарность пар оснований, связываемых аминогруппами, – достаточно распространенное явление.

С целью дальнейшего теоретического обоснования относительной устойчивости выявленного, скрытого структурного полиморфизма комплементарного спаривания оснований мы провели более строгие, по сравнению с нашими предыдущими полуэмпирическими оценками, квантово-химические расчеты уотсон-криковских пар. В данном случае был выполнен оптимизационный конформационный анализ этих пар с использованием *ab initio* MP2/6-311G-приближения на базе программного пакета PC-GAMESS/FIREFLY [51]. Для оценки величин энергетических барьеров для перехода пар из одной неплоской конформации в другую был использован подход так называемого «расчета в одной фиксированной точке» (single point calculations) в варианте MP2/6-311G(d,p)//MP2/6-311G-приближения.

Найденная таким образом геометрия неплоских форм уотсон-криковских АТ- и GC-пар представлена на рис. 8. В табл. 3 приведены результаты оценки углов пропеллера и сдвига плоскостей оснований в этих структурах. При этом полученные неплоские формы, как и предполагается, были проверены на их устойчивость. Соответствующий расчет спектра гармонических колебаний не выявил здесь мнимых значений частот. Уместно также заметить, что, согласно оптимизационным расчетам, все пропеллероподобные и ступенькообразные формы АТ- и GC-пар реализовывались, вообще говоря, с некоторым изломом (buckling) плоскостей оснований в паре. Однако угловые величины этих изломов оказались не столь велики и составляли  $\leq 5^\circ$ . Поэтому особое внимание мы на этом не стали заострять, указывая характеристичными в табл. 3 только параметры пропеллероподобного разворота и ступенькообразного сдвига плоскостей оснований в парах.

Полученные в результате этого расчета значения по углу пропеллера  $\sim 10\text{--}15^\circ$  и по предпочтительности энергии неплоской геометрии над плоской, составляющей примерно  $\sim 1,0$  ккал/моль, в целом подтвердили наши предыдущие данные о существовании некопланарности и неоднозначности формы H-связывания уотсон-криковских пар. Для АТ-пары, таким образом, остается справедливым двукратный, а для GC-пары – четырехкратный поли-



**Рис. 8.** Исходные плоские и оптимизированные методом MP2/6-311G структуры уотсон-криковских АТ- и GС-пар: «пропеллероподобные» АТ(WCu) (а) и АТ(WCd) (б) структуры аденин-тиминового пары; «пропеллероподобные» GС(WCdd) (в) и GС(WCuu) (г) и «ступенькообразные» GС(WCud) (д) и GС(WCdu) (е) структуры гуанин-цитозинового пары.

морфизм структуры комплементарного Н-связывания оснований.

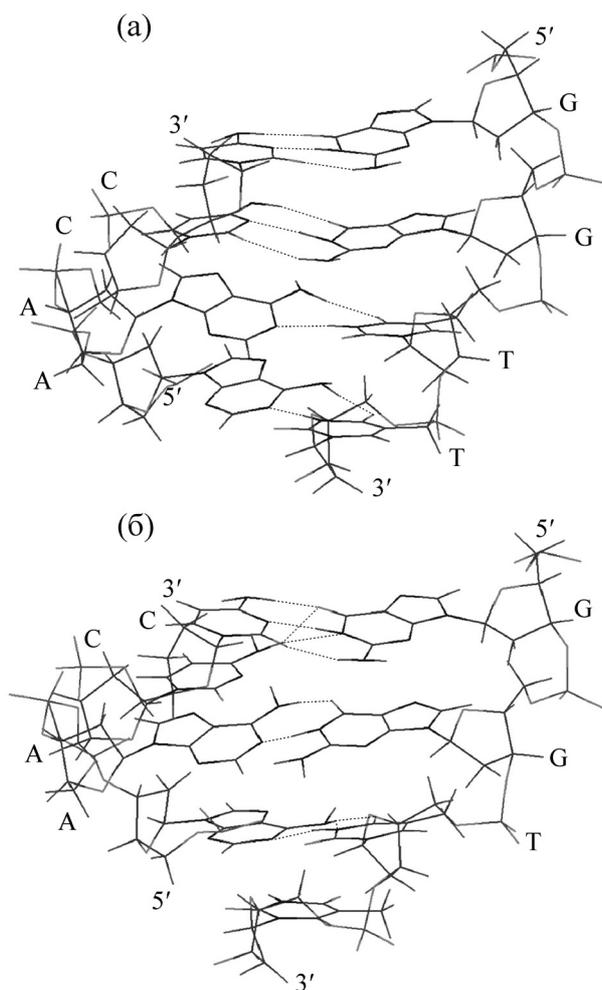
«Пропеллерообразный» характер геометрии одиночных АТ- и GС-пар отмечался иногда и ранее в ряде теоретических исследований других авторов [52–55]. Однако в данном разделе статьи мы делаем акцент именно на существовании отличий в общепринятом «пространственном изоморфизме» уотсон-криковских АТ- и GС-пар и особенно на наличии скрытого, повышенного структурного полиморфизма GС-пар по сравнению с полиморфизмом АТ-пар. Все

это, по нашему мнению, может оказывать значительное влияние на процесс формирования неоднородностей, а также на процесс спонтанного зарождения «неправильных» пар азотистых оснований (точечных мутаций) в регулярной структуре двойной спирали ДНК.

Для идентификации некопланарных форм уотсон-криковских (WC) пар с различным наклоном N–Н-связей пирамидальных аминогрупп оснований (с ориентацией вниз, «down», и вверх, «up», этих связей относительно плоскости основания) мы воспользовались ранее введенными

**Таблица 3.** Теоретически рассчитанные значения углов пропеллера и сдвига плоскостей оснований в уотсон-криковских парах

Структура пары оснований	Угол пропеллера (град) или сдвиг плоскостей оснований (Å)	Разница в энергии между плоской и неплоской структурами, ккал/моль
АТ, плоская	0°	0
АТ(WCu), пропеллер	–9,2°	–1,80
АТ(WCd), пропеллер	9,2°	–1,80
GС, плоская	0°	0
GС(WCdd), пропеллер	–15,05°	–1,09
GС(WCuu), пропеллер	15,05°	–1,09
GС(WCud), ступенька	0,55 Å	–1,08
GС(WCdu), ступенька	–0,55 Å	–1,08



**Рис. 9.** Компьютерное моделирование вариантов водородного связывания азотистых оснований в структуре В-формы тетрамерного ДНК-дуплекса ААСС. (а) – Вариант классического комплементарного Н-связывания оснований с участием неплоских форм АТ(WCd), АТ(WCd), СG(WCdu) и СG(WCdu) (см. текст); (б) – вариант водородного связывания оснований после спонтанного изменения «вверх» («up»)-ориентации N–H-связей NH<sub>2</sub>-групп у обоих аденинов и у одного из цитозинов в цепи 5'-ААСС-3'. При этом в цепи 5'-GGTT-3' реализована спонтанная ориентация «вниз» («down») N–H-связей NH<sub>2</sub>-группы только у второго гуанина.

обозначениями [31,56]. Так, уотсон-криковская АТ-пара со структурой «отрицательного пропеллера» представляется нами как АТ(WCd)-форма. Она отражает наиболее выгодную по энергии форму с наклоном вниз («down») N–H-связей аминогруппы аденина относительно плоскости основания (рис. 8а). Соответственно, АТ-пара со структурой «положительного пропеллера» представляется как форма АТ(WCu) и отражает другую устойчивую, симметричную форму предыдущего варианта с ориентацией вверх («up») N–H-связей аминогруппы аденина

в паре (рис. 8б). Аналогичным способом для случая уотсон-криковской GC-пары определяются четыре некопланарные формы основного состояния:

– две симметричные, «пропеллероподобные» формы, обозначаемые как GC(WCdd) и GC(WCuu) и отвечающие «down-down»- и «up-up»-ориентации N–H-связей аминогрупп гуанина и цитозина соответственно (рис. 8в и 8г);

– две симметричные, «ступенькообразные» формы – GC(WCdu) и GC(WCud), соответственно с «down-up»- и «up-down»-ориентацией N–H-связей аминогрупп гуанина и цитозина (рис. 8д и 8е).

**Особенности организации водородного связывания оснований в ДНК-дуплексе.** С целью иллюстрации возможных последствий спонтанной перестройки ориентации связей пирамидальных NH<sub>2</sub>-групп уотсон-криковских пар в структуре комплементарной двойной спирали В-формы ДНК, нами с использованием методов молекулярной механики и полуэмпирического квантово-химического РМЗ приближения были смоделированы два варианта организации Н-связывания оснований в конкретной двойной последовательности ДНК. В качестве примера был рассмотрен изолированный тетрамерный дуплекс, где в роли матричной цепи выступала последовательность вида 5'-ААСС-3', а в роли синтезированной *de novo* цепи – соответственно последовательность 5'-GGTT-3'.

Первый случай (рис. 9а) моделировал структуру такого фрагмента двойной спирали с определенной исходной формой комплементарного спаривания АТ- и GC-пар. Так, в матричной цепи 5'-ААСС-3' аминогруппы всех оснований рассматривались находящимися в форме с наклоном «вниз» («down») своих валентных N–H-связей относительно плоскости основания, в то время как в цепи *de novo* 5'-GGTT-3' наклон N–H-связей NH<sub>2</sub>-групп обоих гуанинов соответствовал ориентации «вверх» («up») этих связей относительно плоскости основания. Таким способом моделировали вариант классического комплементарного сопряжения цепочек ДНК с участием уотсон-криковских пар, в котором реализовывались некопланарные формы АТ(WCd), АТ(WCd), СG(WCdu) и СG(WCdu) с конкретным пропеллероподобным и ступенькообразным характером Н-связывания оснований.

Второй случай (рис. 9б) моделировал ситуацию Н-связывания оснований в том же дуплексе, но в варианте внезапной инверсионной перестройки пирамидальной структуры аминогрупп у некоторых пар азотистых оснований.

Энергетический барьер подобного типа перестройки  $\text{NH}_2$ -групп с изменением пространственной формы водородного связывания, как было отмечено выше, не столь велик и имеет порядок  $\sim 1,0$  ккал/моль. Он ненамного отличается от величины тепловой энергии окружающей среды  $kT \sim 0,6$  ккал/моль, поэтому данный процесс в обычных условиях достаточно вероятен. Предполагалось, что в матричной цепи 5'-ААСС-3' у обоих аденинов и у одного из цитозинов произошло спонтанное изменение наклона N-H-связей  $\text{NH}_2$ -групп на ориентацию «вверх» («up») этих связей. В *de novo* цепи 5'-GGTT-3' предполагалась перестройка наклона N-H-связей  $\text{NH}_2$ -группы «вниз» («down») только у второго гуанина. Как видно из рис. 9б, в таком варианте внезапных, спонтанных перестроек в исходной форме некоторых пар оснований, как и ожидалось, оптимальные некопланарные водородные связи и укладка оснований в двойной цепи уже реализуются в другом, весьма сложном виде. Прежде всего, оказывается «выпавшим» из уотсон-криковского АТ-спаривания один из тиминов. Кроме того, возникло бифуркационное водородное связывание обоих цитозинов матричной цепи с одним из гуанинов второй цепи. Но самое главное – произошла структурная мутация и образовалась новая пара хугстеновского типа АГ. И этот вариант укладки нуклеотидов реализовался на той же самой комплементарной двойной последовательности d(ААСС)-d(GGTT), что и в первом случае.

Достаточно очевидно, что подобным способом формируются и другие разнообразные структурные перестройки водородных связей в указанном фрагменте двойной цепи ДНК, которые мы, однако, здесь не иллюстрируем. Отметим только, что их число довольно велико и резко увеличивается именно за счет многообразия вариантов вкладов от исходно высокого, четырехкратного полиморфизма структуры присутствующих здесь ГС-пар.

В качестве экспериментального подтверждения гипотезы о неизбежном доминировании уотсон-криковских ГС-пар в подобного рода мутационных процессах следует указать на недавнюю работу [57]. В ней на довольно большом экспериментальном материале впервые была отмечена корреляция скорости точечных спонтанных мутаций в генах человека с содержанием ГС-пар. Как ни странно, эта скорость действительно оказывается линейно возрастающей с ростом процентного ГС-состава генов! Но именно это и следовало ожидать согласно рассмотренным выше аргументам.

Таким образом, представленный материал позволяет утверждать о важном влиянии особенностей скрытого пространственного полиморфизма уотсон-криковских пар на процесс спонтанного нарушения канонического водородного связывания оснований и образования «дефектов» в регулярной структуре двойной спирали ДНК.

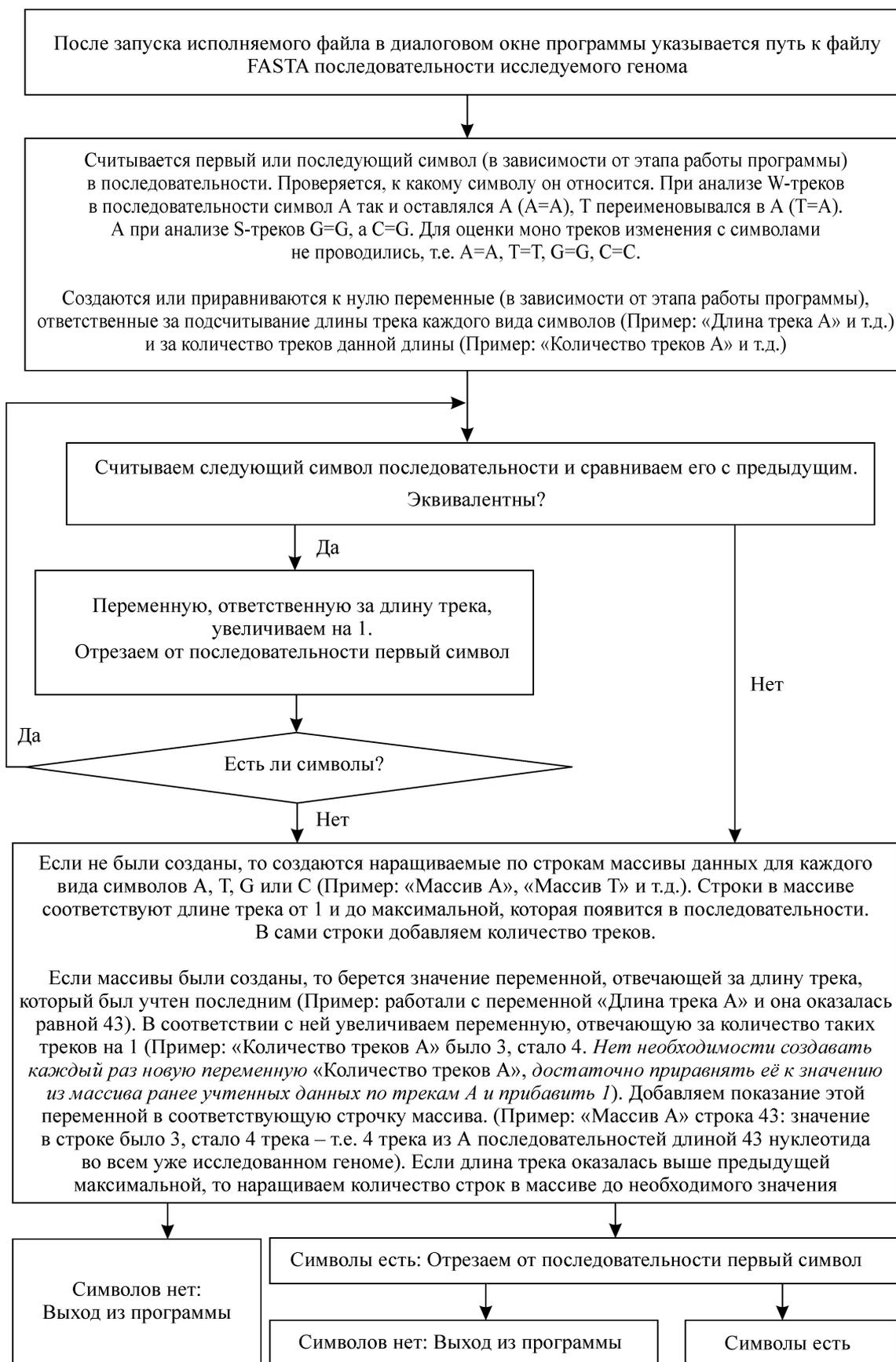
## ЗАКЛЮЧЕНИЕ

Проведенный в работе сравнительный частотный анализ встречаемости уотсон-криковских пар в составе геномов разных видов эукариот, отличающихся своим ГС-составом, выявил общую картину доминирования у них распространенности  $(A/T)_n$ -треков по сравнению с треками ГС-природы. Это указывает на определяющую роль комплементарного АТ-типа спаривания оснований в организации генетических процессов, обусловленную неким внутренним своеобразием строения АТ-пар.

Опираясь на результаты современных квантово-химических расчетов и имеющиеся экспериментальные данные, мы обосновали существование исходных отличий в некопланарности и неоднозначности формы водородного связывания в структуре одиночных уотсон-криковских АТ- и ГС-пар. Сделан вывод о том, что принципиальным отличием здесь выступает пониженный структурный полиморфизм АТ-пары по сравнению с высоким, четырехкратным полиморфизмом геометрии основного состояния ГС-пары. Данное обстоятельство рассматривается наиболее вероятным источником предпочтительности и надежности «использования» природой АТ-пар в структурно-функциональной организации геномной ДНК любого вида организма.

## ПРИЛОЖЕНИЕ

Частотный анализ встречаемости различного вида нуклеотидных последовательностей в двойных цепочках геномных ДНК из базы GenBank (формат FASTA) был проведен нами с помощью разработанной собственной программы для персонального компьютера, особенности алгоритма которой приводятся ниже в виде сокращенной блок-схемы. Для простоты представления в данном алгоритме не отображен процесс распараллеливания вычислений, а также исключены некоторые мелкие детали.



## СПИСОК ЛИТЕРАТУРЫ

1. Y. Zhou, J. W. Bizzaro, and K. A. Marx, *BMC Genomics* **5**, 95 (2004).
2. A. Eyre-Walker and L. D. Hurst, *Nature Rev. Genetics* **2** (7), 549 (2001).
3. J. Bohlin, E. Skjerve, and D. W. Ussery, *PLoS Comput. Biol.* **4** (4), e1000057 (2008)
4. C. E. Singer and B. N. Ames, *Science* **170**, 822 (1970).
5. H. Musto, H. Naya, A. Zavala, et al., *FEBS Lett.* **573**, 73 (2004).
6. H. Musto, H. Naya, A. Zavala, et al., *Biochem. Biophys. Res. Commun.* **347**, 1 (2006).
7. J. L. Oliver and A. Marin, *J. Mol. Evol.* **43**, 216 (1996).
8. K. U. Foerstner, C. von Mering, S. D. Hooper, and P. Bork, *EMBO Rep.* **6**, 1208 (2005).
9. N. Sueoka, *J. Mol. Evol.* **37**, 137 (1993).
10. G. Kudla, L. Lipinski, F. Caffin, et al., *PLoS Biol.* **4**, e180 (2006).
11. H. Nishida, *Curr. Issues Mol. Biol.* **15**, 19 (2013).
12. H. Wu, Z. Zhang, S. Hu, and J. Yu, *Biol. Direct* **7** (2), 1 (2012).
13. D. W. Ussery, T. M. Wassenaar, and S. B. Borini, *Computing for Comparative Microbial Genomics* (Springer-Verlag, London, 2009).
14. S. Karlin and J. Mrazek, *Proc. Natl. Acad. Sci. USA* **94**, 10227 (1997).
15. G. Bernardi, *Structural and Evolutionary Genomics. Natural Selection in Genome Evolution* (Elsevier, Amsterdam, The Netherlands, 2004).
16. A. Cornish-Bowden, *Nucl. Acids Res.* **13**, 3021 (1985).
17. K. A. Marx, S. T. Hess, and R. D. Blake, *J. Biomol. Struct. & Dyn.* **11**, 57 (1993).
18. T. Coenye and P. Vandamme, *DNA Research* **12**, 221 (2005).
19. J. A. Subirana and X. Messeguer, *J. Theor. Biol.*, **283**, 28 (2011).
20. K. A. Marx, Y. Zhou, and I. Q. Kishawi, *J. Biomol. Struct. & Dyn.* **23**, 429 (2006).
21. J. A. Subirana and X. Messeguer, *Nucl. Acids Res.* **38** (4), 1172 (2010)
22. G. Yagil, *J. Mol. Evol.* **37**, 123 (1993).
23. G. Yagil, *Genomics* **87**, 591 (2006).
24. B. Shomer and G. Yagil, *Nucl. Acids Res.* **27**, 4491 (1999).
25. С. С. Киселев, В. М. Комаров, И. С. Масулис и О. Н. Озолин, *Компьют. исследов. моделир.* **2** (2), 183 (2010).
26. С. С. Киселев, *Дис. ... канд. биол. наук (ИБК РАН, Пущино, 2012).*
27. F. Piazza and P. Lio, *Physica A* **347**, 472 (2005).
28. K. J. Dechering, K. Cuelenaere, R. N. H. Konings, and J. A. M. Leunissen, *Nucl. Acids Res.* **26**, 4056 (1998).
29. В. М. Комаров и Н. Г. Мевх, *Журн. физ. химии* **69** (8), 1419 (1995).
30. В. М. Комаров, *Биофизика* **43** (6), 967 (1998).
31. V. M. Komarov, *J. Biol. Phys.* **24**, 167 (1999).
32. A. V. Kabanov and V. M. Komarov, *Intern. J. Quantum. Chem.* **88** (5), 579 (2002).
33. <http://www.ncbi.nlm.nih.gov/genbank/>.
34. [ftp://ftp.ncbi.nlm.nih.gov/genomes/H\\_sapiens](ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens).
35. V. Murray, *Comput. Biol. Chem.* **54**, 13 (2015).
36. E. A. Worthey, A. Schnauffer, G. Aggarwal, et al., *Nucl. Acids Res.* **31** (14), 4201 (2003).
37. R. D. Brown, D. N. Godfrey, D. McNaughton, and A. P. Pierlot, *J. Am. Chem. Soc.* **111**, 2308 (1989).
38. R. D. Brown, D. N. Godfrey, D. McNaughton, and A. P. Pierlot, *Chem. Phys. Lett.* **156**, 61 (1989).
39. В. М. Комаров и Р. В. Полозов, *Биофизика* **35** (2), 367 (1990).
40. J. Sponer and P. Hobza, *J. Phys. Chem.* **98**, 3161 (1994).
41. J. Sponer and P. Hobza, *Int. J. Quant. Chem.* **57**, 959 (1996).
42. V. M. Komarov and R. V. Polozov, *Z. Naturforsch.* **42c**, 1080 (1990).
43. V. M. Komarov, R. V. Polozov, and G. G. Konoplev, *J. Theor. Biol.* **155**, 281 (1992)..
44. В. М. Комаров, А. В. Кабанов, Ю. А. Лазарев, и А. В. Шаповалов, *Математика. Компьютер. Образование* **6** (2), 405 (1999).
45. P. Hobza and C. Sandorfy, *J. Am. Chem. Soc.* **109**, 1302 (1987)
46. C. C. Wilson, *Nucl. Acids Res.* **15**, 8577 (1987).
47. C. C. Wilson and P. Tollin, *Nucleosides & Nucleotides* **6**, 643 (1987).
48. C. C. Wilson, *Nucleosides & Nucleotides* **9**, 479 (1990).
49. J. Jursa and J. Kypr, *Gen. Physiol. Biophys.* **12** (5), 401 (1993).
50. J. R. Roscioli and D. W. Pratt, *Proc. Natl. Acad. Sci. USA* **100**, 13752 (2003).
51. <http://classic.chem.msu.su/gran/games/>
52. P. Hobza and J. poner, *Chem. Rev.* **99** (11), 3247 (1999).
53. V. I. Danilov and V. M. Anisimov, *Biopolym. Cell.* **20** (1–2), 71 (2004).
54. V. I. Danilov, D. M. Hovorun, and N. Kurita, *Біополімери і клітина* **21** (1), 70 (2005).
55. A. R. Srinivasan, R. R. Sauers, M. O. Fenley, et al., *Biophys. Rev.* **1**, 13 (2009).
56. V. M. Komarov, *Biophysics* **43** (6), 917 (1998).
57. V. M. Schaibley, M. Zawistowski, D. Wegmann, et al., *Genome Res.* **23**, 1974 (2013).

## On the Nature of Domination of Oligo (dA:dT)<sub>n</sub> Tracts in the Structure of Eukaryotic Genomes

A.A. Samchenko, S.S. Kiselev, A.V. Kabanov, M.S. Kondratjev, and V.M. Komarov

*Institute of Cell Biophysics, Russian Academy of Sciences,  
ul. Institutskaya 3, Pushchino, Moscow Region, 142290 Russia*

A comparative frequency analysis of the occurrence of oligomeric W- and S-tracts in the genomes of different eukaryotic species, distinguished by their GC-content, was carried out. In studied eukaryotic genomes – *Dictyostelium discoideum* (GC ~ 25.7%), *Caenorhabditis elegans* (GC ~ 36.9%), *Arabidopsis thaliana* (GC ~ 38.0%), *Drosophila melanogaster* (GC ~ 38.8%), *Homo sapiens* (GC ~ 40.0%), *Gallus gallus* (GC ~ 50.0%), *Leishmania major* (GC ~ 59.1%) – mononucleotide and mixed (A/T)<sub>n</sub>-sequences were found to dominate compared to (G/C)<sub>n</sub>-sequences. Based on the results of quantum chemical calculations, the theoretical argument is given in support of the important role of the existence of differences in the initial, hidden structural polymorphism of hydrogen bonding of Watson–Crick AT- and GC-pairs in the formation of this phenomenon. It was concluded that reduced structural polymorphism of AT-pairs in comparison with high-throughput 4-fold polymorphism of the basic state of GC-pairs may be the most likely source of the observed preference and reliability of Nature’s “use” of AT-pairs in genomic DNA in numerous types of organisms.

*Key words: eukaryotic genomes, domination of AT-pairs, quantum chemical calculations*