

УДК: 577.3

Конформационно-динамические свойства ДНК и подходы к физическому картированию генома

Ю. Д. Нечипуренко^{1,2 a}, Д. Ю. Нечипуренко², И. А. Ильичева¹,
М. В. Головкин¹, Л. А. Панченко³, Р. В. Полозов⁴, С. Л. Гроховский¹

¹ Институт молекулярной биологии им. В. А. Энгельгардта РАН,
Россия, 119991, Москва, ул. Вавилова, 32

² Московский государственный университет им. М. В. Ломоносова, Физический факультет
Россия, 119899, Москва, Воробьевы Горы

³ Московский государственный университет им. М. В. Ломоносова, Биологический факультет,
Россия, 119899, Москва, Воробьевы Горы

⁴ Институт теоретической и экспериментальной биофизики,
Россия, 142290, Московская область, Пущино, Московской области, ул. Институтская, д. 3

^a E-mail: nech@imb.ru

Получено 15 ноября 2010 г.

Ранее нами был предложен и развит метод исследования ДНК, основанный на расщеплении двунитевых фрагментов ДНК под действием ультразвука. Были получены относительные частоты расщепления фосфодиэфирной связи во всех 16 динуклеотидах. Увеличение базы проанализированных данных до 20 тысяч нуклеотидов позволило получить относительные частоты расщепления для 256 тетрануклеотидов. Эти величины количественно характеризуют влияние последовательности нуклеотидов на конформационную динамику сахарофосфатного остова. Сейчас известны определённые типы гетерогенности сахарофосфатного остова ДНК, связанные с расщеплением ДНК различными химическими агентами или ДНКазой I. Гетерогенность свойств молекулы ДНК может быть использована для физического картирования генома, то есть определения участков, отвечающих за регуляцию генетической экспрессии.

Ключевые слова: ультразвуковое расщепление ДНК, динамика молекулы ДНК, подвижность молекулы ДНК, d(CpG) динуклеотид

DNA conformational dynamics: approach to the physical mapping of genome

Nechipurenko Yu. D.^{1,2 a}, Nechipurenko D. Yu.², Il'icheva I. A.¹, Golovkin M. V.¹,
Panchenko L. A.³, Polozov R. V.⁴, Grokhovsky S. L.¹

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

² Dep. of Physics, Moscow State University, Moscow, 119992, Russia

³ Dep. of Biology, Moscow State University, Moscow, 119992, Russia

⁴ Institute of Theoretical and Experimental Biophysics RAS, Puschino, 142290, Russia

Abstract. – Recently we have developed a new method for studying DNA based on ultrasound - induced cleavage of DNA sugar-phosphate backbone. Relative cleavage rates of the phosphodiester bonds in all 16 dinucleotides have been determined. The increased amount of data sampling (of more than 20 000 nucleotides) made it also possible to obtain cleavage rates in all 256 possible tetranucleotides. These values quantitatively characterize sequence effects on conformational dynamics of DNA sugar phosphate backbone. Same type of DNA heterogeneity have been discovered and studied using its chemical cleavage induced by various chemical agents and DNase I. The presence of essential heterogeneity in structural properties of DNA might be a key for physical mapping of the genomes, i.e. determining the structural profiles being responsible for DNA recognition by gene expression regulation machinery.

Keywords: ultrasonic cleavage of DNA, sequence effects on DNA structure, DNA flexibility, d(CpG) dinucleotide

Citation: *Computer Research and Modeling*, 2010, vol. 2, no. 4, pp. 419–428 (Russian).

Работа выполнена при финансовой поддержке гранта РФФИ (08-04-01739) и Программы Президиума РАН по молекулярной и клеточной биологии.

© 2010 Юрий Дмитриевич Нечипуренко, Дмитрий Юрьевич Нечипуренко, Ирина Алексеевна Ильичева, Михаил Викторович Головкин, Лариса Андреевна Панченко, Роберт Валентинович Полозов, Сергей Львович Гроховский

Введение

В настоящее время для компьютерного анализа стали доступны нуклеотидные последовательности геномов множества организмов. Большое значение приобретает развитие эффективных методов обнаружения функциональных сайтов расшифрованных последовательностей [Liolios et al, 2006; Wang et al, 2006]. Несмотря на значительные успехи методов биоинформатики, задача точного определения регуляторных участков в геноме остается одной из важных проблем молекулярной биологии [Bajic et al, 2004]. Определение промоторных областей в ДНК, то есть участков, отвечающих за регуляцию работы генов, важно не только для обнаружения новых генов, но и для локализации областей генома, в которых следует проводить экспериментальный и теоретический поиск сайтов, наиболее важных для регуляции транскрипции [Bajic et al, 2006].

Известно, что свойства промоторных участков ДНК, как правило, сильно отличаются от свойств других областей генома [Florquin et al, 2005]. Большинство существующих алгоритмов поиска промоторных участков в геномах опирается на свойства нуклеотидных последовательностей уже известных промоторов, используют такие методы как дискриминантный анализ, цепи Маркова и искусственные нейронные сети [Sonnenburg et al, 2006]. Программы, реализующие алгоритмы такого рода, требуют больших объемов данных для «обучения» и являются зачастую ориентированными на определенный вид организмов, а также далеко не всегда способны быстро работать при анализе на уровне целого генома [Bajic et al, 2004; Sonnenburg et al, 2006].

В последние годы успешно развиваются новые подходы к поиску регуляторных участков в геномах, основанные на использовании данных о контекстной зависимости различных физико-химических характеристик ДНК [Florquin et al, 2005; Abelev et al, 2008; Dineen et al, 2009]. Основной идеей таких подходов является теоретическое построение и последующий анализ профиля, характеризующего изменение вдоль ДНК определенного физико-химического параметра. При этом используются самые разные локальные характеристики ДНК – гибкость, энергия стэкинга, температура плавления и многие другие. Зависимость используемых физических характеристик ДНК от нуклеотидной последовательности, как правило, описывается в рамках ди-, три- и тетрануклеотидных приближений. Существенную роль играет вырожденность таких «физических кодов» по нуклеотидной последовательности ДНК: разные последовательности могут приводить к схожим профилям изменения локального структурного параметра ДНК. Вырожденность такого рода позволяет предложить новые критерии для анализа функциональных участков генома [Parker et al, 2009]. Описанные подходы представляют собой важное дополнение к общепринятым методам биоинформатики.

Стоит отметить, что при описании характерных свойств регуляторных последовательностей ДНК различные физико-химические характеристики молекулы в определенном смысле могут дополнять друг друга [Florquin et al, 2005]. В процессах взаимодействия ДНК с регуляторными и структурными белками важную роль играют локальные конформационно-динамические характеристики молекулы ДНК: в конечном итоге белок «узнает» не определенную последовательность пар оснований в ДНК, а структуру молекулы, как правило, меняя её локальную геометрию [Becker et al, 2006]. Описано множество вариантов ДНК-белкового узнавания [Becker et al, 2006; Rohs et al, 2009], и для разных классов белков ключевыми факторами, определяющими «прочность» связывания, могут быть самые разные локальные характеристики ДНК: гибкость, геометрия спирали, стабильность дуплекса, подвижность определенных молекулярных групп и т. д. Как известно, все эти параметры в ДНК зависят от нуклеотидной последовательности, причем влияние последовательности на локальные характеристики ДНК может выходить далеко за рамки ди- и тетрануклеотидного описания [Faiger et al, 2006]. Таким образом, сложность и многообразие процессов регуляции генетической экспрессии требуют учёта неоднородности разных физических параметров молекулы ДНК при описании характерных свойств её функциональных сайтов.

В данной работе для теоретического анализа геномных последовательностей мы предлагаем использовать новую величину – относительную частоту ультразвукового расщепления. Ра-

нее нами был предложен и развит новый метод исследования ДНК, основанный на изучении картин её ультразвукового расщепления, полученных методом гелевого электрофореза. Анализ экспериментальных данных позволил выявить зависимость частоты разрывов сахарофосфатного остова ДНК от последовательности пар оснований [Гроховский, 2006]. Влияние локальной последовательности нуклеотидных пар на частоту расщепления фосфодиэфирной связи описывается в рамках ди- и тетрануклеотидного приближений. Ранее нами были получены относительные частоты расщепления межнуклеотидных связей в каждом из 16 динуклеотидов [Гроховский и др., 2008; Головкин и др., 2009; Нечипуренко и др., 2009]. Увеличение экспериментальных данных до выборки в более чем 20 тысяч нуклеотидов позволило установить относительные частоты расщепления также для всех 256 тетрануклеотидов.

Полученные величины относительных частот расщепления ДНК коррелируют с конформационной подвижностью дезоксирибозы, примыкающей к месту разрыва [P'icheva et al, 2009], тем самым отражая локальную конформационную подвижность сахарофосфатного остова ДНК: чем больше относительная частота ультразвукового расщепления данной фосфодиэфирной связи, тем более подвижным является сахарофосфатный остов в соответствующем участке молекулы.

Следует отметить, что полученные нами величины позволяют различать особенности структурной динамики сахарофосфатного остова в каждой из цепей ДНК: показано, что величины относительной частоты расщепления фосфодиэфирных связей в комплементарных динуклеотидах могут значимо различаться [Гроховский и др., 2008]. Поэтому при исследовании геномных последовательностей появляется возможность анализировать конформационно-динамические особенности обеих цепей двойной спирали ДНК.

Материалы и методы

Экспериментальные материалы и методы изложены в работах [Гроховский, 2006; Гроховский и др., 2008]. В работах [Гроховский и др., 2008; Головкин и др., 2009; Нечипуренко и др., 2009] опубликованы значения относительных частот ультразвукового расщепления ДНК для 16 динуклеотидов ($dN_n pdN_{n+1}$), (p – фосфатная группа, d – дезоксирибоза), в которых разрыв сахарофосфатного остова происходит между нуклеотидами pdN_n и pdN_{n+1} (по $C3' - O3'$ связи). В таблице 1 приведены полученные нами значения относительных частот расщепления центральной фосфодиэфирной связи для тетрануклеотидов

$$\{N_{n-1}N_nN_{n+1}N_{n+2}\}.$$

Результаты

Используя значения, приведенные в таблице 1 мы рассчитали профили ультразвукового расщепления для последовательностей 648 промоторов генома человека (MYC кластер, взятый из базы данных Transcriptional Regulatory Element Database:

<http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=home>)

Все последовательности изначально были выровнены: рассматривался блок длиной в 1 000 нуклеотидов, содержащий 700 нуклеотидов до точки старта транскрипции и 300 – после.

Для каждой последовательности восстанавливался теоретический профиль относительных частот ультразвукового расщепления, после чего полученные профили сглаживались методом линейной фильтрации с окном в 10 нуклеотидов. Отметим, что для построения профилей использовались обе цепи, то есть бралась суммарная частота расщепления по обоим цепям.

На рис. 1 приведен профиль ультразвукового расщепления, полученный усреднением 648 профилей.

Таблица 1. Выборочные характеристики относительных частот ультразвукового расщепления центральной фосфодиэфирной связи в 256 тетрануклеотидах

Обозначения:

 N – объём выборки; \bar{R} – средняя величина; S – стандартное отклонение; $S_{\bar{R}}$ – стандартное отклонение среднего

	N	\bar{R}	S	$S_{\bar{R}}$
AAAA	232	0.894	0.132	0.009
AAAC	151	0.883	0.139	0.011
AAAG	93	0.915	0.107	0.013
AAAT	96	0.951	0.142	0.013
CAAA	114	0.921	0.169	0.012
CAAC	80	0.877	0.118	0.015
CAAG	77	0.938	0.142	0.015
CAAT	77	0.935	0.126	0.015
GAAA	119	0.921	0.113	0.012
GAAC	49	0.887	0.117	0.019
GAAG	88	0.936	0.113	0.014
GAAT	132	0.917	0.127	0.011
TAAA	108	0.952	0.125	0.013
TAAC	73	0.916	0.097	0.015
TAAG	55	0.997	0.155	0.018
TAAT	95	0.929	0.124	0.013

	N	\bar{R}	S	$S_{\bar{R}}$
AAGA	70	0.905	0.098	0.012
AAGC	91	0.845	0.142	0.015
AAGG	72	0.924	0.133	0.016
AGAT	79	0.899	0.108	0.012
CAGA	81	0.935	0.146	0.016
CAGC	100	0.846	0.136	0.014
CAGG	82	0.932	0.128	0.014
CAGT	92	0.897	0.137	0.014
GAGA	50	0.880	0.118	0.017
GAGC	59	0.889	0.109	0.014
GAGG	48	0.914	0.117	0.017
GAGT	50	0.874	0.107	0.015
TAGA	41	0.935	0.082	0.013
TAGC	50	0.907	0.114	0.016
TAGG	40	0.886	0.145	0.023
TAGT	26	0.975	0.197	0.039

	N	\bar{R}	S	$S_{\bar{R}}$
AACA	96	0.938	0.112	0.011
AACC	77	0.881	0.131	0.015
AACG	100	0.923	0.128	0.013
AACT	78	0.919	0.127	0.014
CACA	58	0.943	0.151	0.020
CACC	85	0.820	0.130	0.014
CACG	46	0.871	0.174	0.026
CACT	73	0.900	0.116	0.014
GACA	59	0.935	0.166	0.022
GACC	34	0.921	0.117	0.020
GACG	83	0.913	0.156	0.017
GACT	61	0.914	0.129	0.016
TACA	38	0.986	0.094	0.015
TACC	79	0.887	0.113	0.013
TACG	71	0.921	0.101	0.012
TACT	43	0.973	0.114	0.017

	N	\bar{R}	S	$S_{\bar{R}}$
AATA	110	0.925	0.121	0.012
AATC	65	0.913	0.117	0.014
AATG	98	0.913	0.127	0.013
AATT	129	0.875	0.115	0.010
CATA	75	0.910	0.137	0.016
CATC	67	0.801	0.189	0.023
CATG	70	0.879	0.103	0.012
CATT	88	0.863	0.108	0.011
GATA	91	0.907	0.083	0.009
GATC	92	0.950	0.120	0.012
GATG	77	0.929	0.113	0.013
GATT	76	0.913	0.107	0.012
TATA	67	0.913	0.129	0.016
TATC	89	0.858	0.114	0.012
TATG	69	0.946	0.106	0.013
TATT	115	0.914	0.132	0.012

Таблица 1 (продолжение)

	N	\bar{R}	S	$S_{\bar{R}}$
АСAA	57	1.103	0.176	0.023
АСAC	53	1.108	0.177	0.024
АСAG	86	1.183	0.213	0.023
АСAT	67	1.116	0.181	0.022
ССAA	73	1.025	0.178	0.021
ССAC	45	0.989	0.149	0.022
ССAG	96	1.190	0.213	0.022
ССAT	81	1.073	0.168	0.019
GCAA	126	1.223	0.204	0.018
GCAC	71	1.174	0.215	0.026
GCAG	89	1.356	0.263	0.028
GCAT	73	1.186	0.203	0.024
TCAA	93	1.127	0.171	0.018
TCAC	94	1.112	0.144	0.015
TCAG	84	1.231	0.266	0.029
TCAT	80	1.129	0.200	0.022

	N	\bar{R}	S	$S_{\bar{R}}$
ACCA	78	0.975	0.112	0.013
ACCC	72	0.986	0.147	0.017
ACCG	79	1.064	0.143	0.016
ACCT	45	1.012	0.101	0.015
CCCA	45	0.930	0.093	0.014
CCCC	107	0.912	0.115	0.011
CCCG	103	0.992	0.147	0.014
CCCT	59	0.941	0.138	0.018
GCCA	118	1.080	0.153	0.014
GCCC	50	1.085	0.223	0.032
GCCG	48	1.130	0.130	0.019
GCCT	61	1.086	0.150	0.019
TCCA	53	0.999	0.127	0.018
TCCC	88	0.943	0.113	0.012
TCCG	70	1.035	0.149	0.018
TCCT	70	0.966	0.130	0.016

	N	\bar{R}	S	$S_{\bar{R}}$
ACGA	86	1.537	0.331	0.036
ACGC	96	1.362	0.309	0.032
ACGG	61	1.483	0.283	0.036
ACGT	53	1.381	0.369	0.051
CCGA	63	1.432	0.309	0.039
CCGC	91	1.257	0.220	0.023
CCGG	91	1.417	0.311	0.033
CCGT	55	1.328	0.246	0.033
GCGA	100	1.783	0.577	0.058
GCGC	87	1.462	0.358	0.038
GCGG	102	1.549	0.334	0.033
GCGT	72	1.427	0.324	0.038
TCGA	47	1.543	0.371	0.054
TCGC	104	1.300	0.282	0.028
TCGG	65	1.543	0.265	0.033
TCGT	58	1.263	0.227	0.030

	N	\bar{R}	S	$S_{\bar{R}}$
ACTA	44	1.160	0.210	0.032
ACTC	74	1.064	0.134	0.016
ACTG	58	1.157	0.243	0.032
ACTT	69	1.060	0.158	0.019
CCTA	23	1.031	0.145	0.030
CCTC	73	0.980	0.128	0.015
CCTG	73	1.118	0.157	0.018
CCTT	67	1.029	0.120	0.015
GCTA	40	1.227	0.166	0.026
GCTC	90	1.256	0.227	0.024
GCTG	112	1.303	0.216	0.020
GCTT	88	1.210	0.225	0.024
TCTA	31	1.098	0.068	0.012
TCTC	56	0.992	0.131	0.018
TCTG	101	1.149	0.195	0.019
TCTT	81	1.042	0.188	0.021

	N	\bar{R}	S	$S_{\bar{R}}$
AGAA	96	0.986	0.117	0.012
AGAC	30	0.979	0.137	0.025
AGAG	48	0.956	0.144	0.021
AGAT	69	1.003	0.127	0.015
CGAA	100	0.982	0.205	0.020
CGAC	84	0.914	0.150	0.016
CGAG	52	0.962	0.102	0.014
CGAT	61	1.008	0.223	0.029
GGAA	102	0.968	0.123	0.012
GGAC	26	0.979	0.136	0.027
GGAG	31	0.898	0.175	0.031
GGAT	98	0.998	0.155	0.016
TGAA	88	0.936	0.124	0.013
TGAC	87	0.964	0.117	0.013
TGAG	77	0.997	0.106	0.012
TGAT	105	0.987	0.123	0.012

	N	\bar{R}	S	$S_{\bar{R}}$
AGCA	95	0.975	0.136	0.014
AGCC	46	0.922	0.136	0.020
AGCG	73	1.029	0.136	0.016
AGCT	84	0.974	0.145	0.016
CGCA	99	0.969	0.115	0.012
CGCC	106	0.886	0.173	0.017
CGCG	92	0.974	0.171	0.018
CGCT	82	0.929	0.146	0.016
GGCA	75	0.913	0.128	0.015
GGCC	68	0.868	0.121	0.015
GGCG	104	0.967	0.169	0.017
GGCT	58	0.982	0.144	0.019
TGCA	88	0.931	0.134	0.014
TGCC	55	0.958	0.143	0.019
TGCG	94	1.016	0.174	0.018
TGCT	87	1.008	0.155	0.017

Таблица 1 (продолжение)

	N	\bar{R}	S	$S_{\bar{R}}$
AGGA	37	0.943	0.100	0.016
AGGC	70	0.905	0.150	0.018
AGGG	89	0.989	0.153	0.016
AGGT	46	0.921	0.155	0.023
CGGA	89	0.962	0.112	0.012
CGGC	69	0.831	0.133	0.016
CGGG	63	0.902	0.159	0.020
CGGT	100	0.925	0.148	0.015
GGGA	76	0.939	0.165	0.019
GGGC	71	0.881	0.155	0.018
GGGG	84	0.958	0.160	0.018
GGGT	71	0.948	0.121	0.014
TGGA	54	0.944	0.152	0.021
TGGC	93	0.859	0.142	0.015
TGGG	65	0.885	0.134	0.017
TGGT	93	0.922	0.130	0.013

	N	\bar{R}	S	$S_{\bar{R}}$
AGTA	36	0.972	0.083	0.014
AGTC	38	0.909	0.170	0.028
AGTG	59	0.972	0.099	0.013
AGTT	112	0.965	0.125	0.012
CGTA	60	0.987	0.081	0.010
CGTC	57	0.908	0.165	0.022
CGTG	32	0.981	0.178	0.031
CGTT	91	0.940	0.138	0.014
GGTA	87	0.959	0.123	0.013
GGTC	30	0.904	0.139	0.025
GGTG	88	0.966	0.134	0.014
GGTT	103	0.938	0.118	0.012
TGTA	76	0.967	0.126	0.014
TGTC	62	0.921	0.127	0.016
TGTG	83	0.960	0.135	0.015
TGTT	89	0.925	0.138	0.015

	N	\bar{R}	S	$S_{\bar{R}}$
ATAA	103	0.974	0.109	0.011
ATAC	94	0.963	0.148	0.015
ATAG	49	1.020	0.164	0.023
ATAT	98	0.945	0.136	0.014
CTAA	30	1.005	0.077	0.014
CTAC	19	0.895	0.073	0.017
CTAG	24	0.998	0.065	0.013
CTAT	64	0.948	0.128	0.016
GТАА	93	1.001	0.127	0.013
GTAC	43	0.986	0.130	0.020
GТАG	28	1.068	0.157	0.030
GTAT	92	1.002	0.100	0.010
TТАА	103	0.925	0.111	0.011
TТАC	84	0.915	0.106	0.012
TТАG	53	1.012	0.113	0.016
TТАT	86	0.960	0.104	0.011

	N	\bar{R}	S	$S_{\bar{R}}$
ATCA	104	0.967	0.165	0.016
ATCC	79	0.877	0.096	0.011
ATCG	69	0.942	0.101	0.012
ATCT	64	0.941	0.110	0.014
CTCA	95	0.892	0.109	0.011
CTCC	73	0.842	0.166	0.019
CTCG	51	0.937	0.107	0.015
CTCT	71	0.864	0.123	0.015
GTCA	70	0.983	0.129	0.015
GTCC	36	0.932	0.080	0.013
GTСG	48	0.970	0.132	0.019
GTCT	30	0.962	0.144	0.026
TТCA	81	0.917	0.115	0.013
TТCC	97	0.845	0.131	0.013
TТCG	100	0.920	0.136	0.014
TТCT	99	0.905	0.136	0.014

	N	\bar{R}	S	$S_{\bar{R}}$
ATGA	85	1.017	0.105	0.011
ATGC	82	0.979	0.116	0.013
ATGG	66	1.006	0.123	0.015
ATGT	80	0.951	0.137	0.015
CTGA	89	1.008	0.141	0.015
CTGC	101	0.973	0.139	0.014
CTGG	82	1.008	0.109	0.012
CTGT	71	0.993	0.135	0.016
GTGA	86	1.026	0.125	0.013
GTGC	73	0.929	0.139	0.016
GTGG	55	1.011	0.121	0.016
GTGT	49	0.968	0.124	0.018
TТGA	97	0.983	0.125	0.013
TТGC	82	0.917	0.118	0.013
TТGG	102	0.960	0.132	0.013
TТGT	110	0.932	0.118	0.011

	N	\bar{R}	S	$S_{\bar{R}}$
ATTA	107	0.908	0.105	0.010
ATTC	95	0.908	0.122	0.013
ATTG	95	1.004	0.116	0.012
ATTT	105	0.938	0.123	0.012
CTTA	48	0.910	0.108	0.016
CTTC	90	0.895	0.098	0.010
CTTG	67	0.954	0.119	0.014
CTTT	97	0.926	0.152	0.015
GTTA	81	1.041	0.145	0.016
GTTC	66	0.954	0.133	0.016
GTTG	111	0.966	0.127	0.012
GTTT	138	0.945	0.125	0.011
TТТА	93	0.937	0.117	0.012
TТTC	127	0.890	0.133	0.012
TТTG	117	0.932	0.119	0.011
TТTT	236	0.897	0.140	0.009

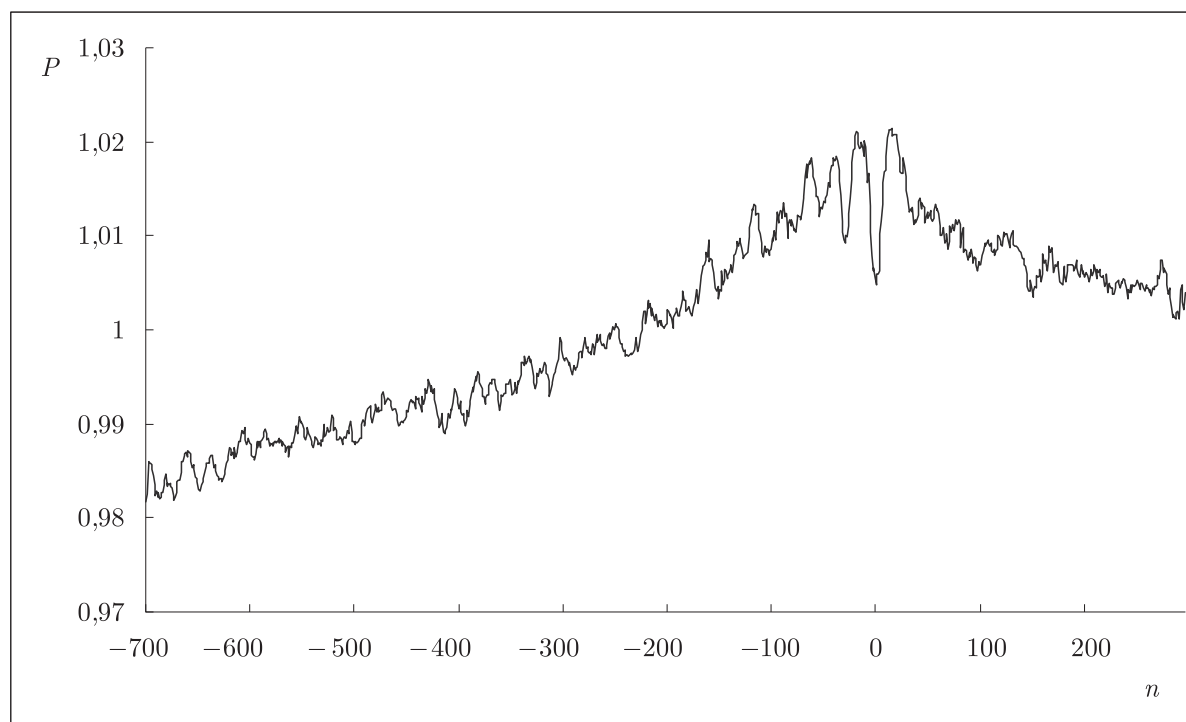


Рис. 1. Зависимость относительной частоты ультразвукового расщепления (P) от позиции соответствующей фосфодиэфирной связи (n) по отношению к точке начала транскрипции (0). Приведен средний для 648 последовательностей профиль ультразвукового расщепления

Из рис. 1 видно, что рассматриваемая нами характеристика имеет явную аномалию в районе старта транскрипции. Полученный профиль соответствует обнаруженным ранее результатам (возрастание к месту сайта старта транскрипции частоты встречаемости цитозинов и гуанинов в последовательности [Abeel et al, 2008; Cao et al, 2008]) и отражает преимущественное расщепление фосфодиэфирной связи, следующей за цитозином.

Для того, чтобы проанализировать конформационно-динамические характеристики разных участков генома, мы использовали следующую характеристику:

$$S(n) = \frac{P(N_{n-1}N_nN_{n+1}N_{n+2}) - P(N_nN_{n+1})}{P(N_{n-1}N_nN_{n+1}N_{n+2})}$$

здесь $P(N_{n-1}N_nN_{n+1}N_{n+2})$ – относительная частота расщепления центральной фосфодиэфирной связи тетра nukлеотида $N_{n-1}N_nN_{n+1}N_{n+2}$, а $P(N_nN_{n+1})$ – средняя относительная частота расщепления динуклеотида N_nN_{n+1} .

Эта величина характеризует степень влияния фланкирующих нуклеотидов на ультразвуковое расщепление фосфодиэфирной связи центрального динуклеотида в составе тетра nukлеотида и может служить мерой «подвижности» сахарофосфатного остова молекулы ДНК, динуклеотидный состав которой задан.

При $S(n) < 0$ фланкирующие нуклеотиды приводят к уменьшению ультразвукового расщепления (а следовательно, к уменьшению подвижности сахарофосфатного остова), а при $S(n) > 0$, наоборот, – к увеличению ультразвукового расщепления по сравнению со средним значением для данного динуклеотида (к увеличению подвижности остова).

На рис. 2 представлен усреднённый профиль, полученный аналогично первому рисунку, но для величины $S(n)$, соответствующей кодирующей нити.

Из рис. 2 видно, что область до сайта старта транскрипции в среднем характеризуется величинами $S(n) < 0$, причем переход к более высоким значениям $S(n)$ достигается на сравнительно

но небольшом участке – в области от -130 до $+1$ нуклеотида, после чего величины $S(n)$ принимают значения, близкие к нулевому.

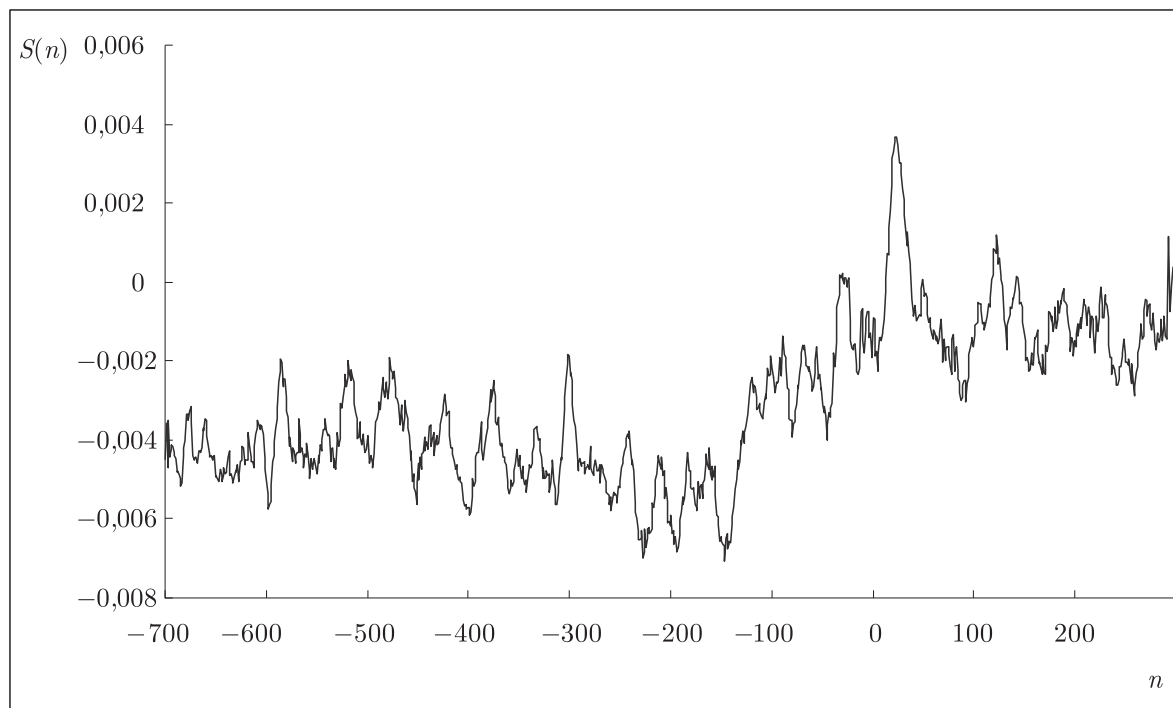


Рис. 2. Зависимость относительного изменения частоты ультразвукового расщепления (S) при переходе от динуклеотидного к тетра nukлеотидному приближению от положения соответствующей фосфодиэфирной связи (n) по отношению к точке начала транскрипции (0). Приведен средний для 648 последовательностей профиль величины $S(n)$

Обсуждение

Задачи картирования, аннотации и классификации функциональных участков геномов по первичной структуре ДНК в настоящее время в первом приближении решены. Однако для понимания механизмов экспрессии генов необходимо поставить аналогичные задачи по конформационно-динамическим характеристикам ДНК, развить методы их решения и дать молекулярно-биологическую интерпретацию полученным результатам. В свою очередь, решение этих задач может дать основу для установления физико-химических механизмов экспрессии генов. Разработка подходов и методов для анализа конформационно-динамических состояний молекулы ДНК в участках, имеющих различные функции, представляет собой актуальную задачу.

Биоинформационные методы анализа генома используют текстовый подход для определения функциональных участков. В настоящее время известно, что наряду с узнаванием белками ДНК посредством образования водородных связей с основаниями («прямое узнавание») важную роль играет «непрямое узнавание» – за счет конформационных изменений сахарофосфатного остова, причем сходные конформации могут образовывать разные последовательности нуклеотидов.

Предложенный нами подход к анализу конформационно-динамических характеристик в зависимости от нуклеотидной последовательности, а также созданная база данных расщепления ультразвуком всех возможных тетра nukлеотидов дают новые инструменты для предсказания промоторных участков, позиционирования положения нуклеосом и поиска других функциональных участков ДНК.

Полученный результат позволяет предположить, что в промоторных участках ДНК человека нуклеотиды эволюционно отобраны таким образом, что их влияние ведёт к уменьшению

гибкости сахарофосфатного остова. Уменьшение гибкости промоторных участков ДНК может иметь своей целью увеличение доступности регуляторных сайтов для факторов, регулирующих экспрессию генов, так как по мере уменьшения гибкости участков ДНК, как известно, происходит уменьшение вероятности образования комплексов ДНК с гистонами [Pedersen et al, 1998].

Уменьшение гибкости и стабильности ДНК в промоторных участках было неоднократно подчеркнуто различными авторами и активно используется в пакетах программ по предсказанию сайтов старта транскрипции в геномных последовательностях [Florquin et al, 2005; Abeel et al, 2008; Dineen et al, 2009; Cao et al, 2008]. Как уже отмечалось, ключевым звеном многих алгоритмов является построение профилей различных физико-химических характеристик ДНК. Как правило, такие алгоритмы используют параметризацию, полученную в рамках динуклеотидного приближения.

Проведенный анализ явно свидетельствует о наличии эффектов последовательности в промоторных участках генома человека, проявляющихся на уровне тетра-нуклеотидного приближения, а также о возможности выявления этих эффектов при помощи полученных данных по расщеплению фрагментов ДНК ультразвуком [Grokhovsky et al, 2011].

Список литературы

- Abeel T., Saeys Y., Bonnet E., Rouzé P., Van de Peer Y. Generic eukaryotic core promoter prediction using structural features of DNA // *Genome Research*. 2008. Vol. 18. P. 310–23.
- Bajic V. B., Tan S. L., Suzuki Y., Sugano S., Tan S. L., Suzuki Y., Sugano S., Suzuki Y., Sugano S., Sugano S. Promoter prediction analysis on the whole human genome // *Nat. Biotechnol.* 2004. Vol. 22. P.1467–1473.
- Becker N. B., Wolff L., Everaers R. Indirect readout: detection of optimized subsequences and calculation of relative binding affinities using different DNA elastic potentials // *Nucleic Acids Res.* 2006. Vol. 34. P. 5638–5649.
- Cao X. Q., Zeng J, Yan H. Structural property of regulatory elements in human promoters // *Phys Rev E Stat Nonlin Soft Matter Phys.* 2008. Vol. 77(4 Pt 1). P. 041908.
- Dineen D. G., Wilm A., Cunningham P., Higgins D. G. High DNA melting temperature predicts transcription start site location in human and mouse // *Nucleic Acids Res.* 2009. Vol. 37(22). P.7360–7.
- Faiger, H., Ivanchenko, M., Cohen, I., Haran, T. E. TBP flanking sequences: Asymmetry of binding, long-range effects and consensus sequences // *Nucleic Acids Res.* 2006. Vol. 34. P. 104–119.
- Florquin K., Saeys Y., Degroeve S., Rouzé P., Van de Peer Y. Large-scale structural analysis of the core promoter in mammalian and plant genomes // *Nucleic Acids Res.* 2005. Vol. 33(13). P. 4255–64.
- Grokhovsky S. L., Il'icheva I. A., Nechipurenko D. Yu, Golovkin M. V., Panchenko L. A., Polozov R. V., Nechipurenko Yu. D. Sequence-Specific Ultrasonic Cleavage of DNA // *Biophys. J.* 2011. Vol. 100(1), P. 117–25.
- Il'icheva I. A., Nechipurenko D. Y., Grokhovsky S. L. Ultrasonic cleavage of nicked DNA // *J. Biomol Struct Dyn.* 2009. Vol. 27(3). P. 391–8.
- Liolios K., Tavernarakis N., Hugenholtz P., Kyripides N. C. The Genomes On Line Database (GOLD) v.2: A monitor of genome projects worldwide // *Nucleic Acids Res.* 2006. Vol. 34. P. D332–D334.
- Parker S. C., Hansen L., Abaan H. O., Tullius T. D., Margulies E. H. Local DNA topography correlates with functional noncoding regions of the human genome // *Science.* 2009. Vol. 324(5925). P. 389–92.
- Pedersen A. G., Baldi P., Chauvin Y., Brunak S. DNA Structure in Human RNA Polymerase II Promoters // *J. Mol. Biol.* 1998. Vol. 281. P. 663–673.

- Rohs R., West S. M., Liu P., Honig B.* Nuance in the double-helix and its role in protein–DNA recognition // *Curr. Opin. in Struct. Biol.* 2009. Vol. 19. P. 171–177.
- Solovyev V., Kosarev P., Seledsov I., Vorobyev D.* Automatic annotation of eukaryotic genes, pseudogenes and promoters // *Genome Biol.* 2006. Vol. 7. P. 11–12.
- Sonnenburg S. O., Zien A., Rättsch G.* ARTS: Accurate recognition of transcription starts in human // *Bioinformatics.* 2006. Vol. 22. P. 472–480.
- Wang G., Zhang W.* A steganalysis-based approach to comprehensive identification and characterization of functional regulatory elements // *Genome Biol.* 2006. Vol. 7. R49. doi: 10.1186/gb-2006-7-6-r49.
- Головкин М. В., Нечипуренко Д. Ю., Ильичева И. А., Панченко Л. А., Полозов Р. В., Гроховский С. Л., Нечипуренко Ю. Д.* Математические методы анализа электрофоретических картин расщепления ДНК // *Компьютерные исследования и моделирование* 2009. Т. 1. С. 287–295.
- Гроховский С. Л.* Специфичность расщепления ДНК ультразвуком // *Мол. биология.* 2006. Т. 40. С. 317–325.
- 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.
- Нечипуренко Ю. Д., Головкин М. В., Нечипуренко Д. Ю., Ильичева И. А., Панченко Л. А., Полозов Р. В., Гроховский С. Л.* Характерные особенности расщепления ДНК ультразвуком // *Журнал структурной химии* 2009. Т. 50. С. 1040–1047.