

Взаимодействия белок–ДНК: статистический анализ межатомных контактов в большой и малой бороздках

А.А. Анашкина¹, Е.Н. Кузнецов², А.В. Батяновский³, Л.А. Урошлев⁴, В.Г. Туманян¹, Н.Г. Есипова¹

¹ Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, Россия

² Институт проблем управления им. В.А. Трапезникова Российской академии наук, Москва, Россия

³ Институт биофизики и клеточной инженерии НАН Беларуси, Минск, Республика Беларусь

⁴ Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия

Взаимодействия между белком и ДНК по существу лежат в основе всех процессов, происходящих в живой клетке. Понимание принципов специфического распознавания сайтов ДНК позволит понять, как управляются эти процессы, и даст возможность сознательно вмешиваться в управление ими. В работе изучены методом Вороного–Делоне контакты белок–ДНК на атомном уровне в структурах 3518 комплексов из PDB (все имеющиеся на май 2017 г.), содержащих как белковые цепи, так и ДНК. Метод не содержит параметров и позволяет однозначно выявлять непосредственные контакты между атомами и характеризовать каждый контакт, помимо расстояния между атомами, площадью контакта, определяемой соответствующей гранью полиэдра Вороного. Показано, что большая часть контактов образуется между атомами белка и атомами сахарофосфатного остова ДНК (72.9%). На контакты с атомами нуклеиновых оснований, выходящих в бороздки ДНК, приходится для большой бороздки 17.0% и для малой бороздки 10.1% от всех атомных контактов. Суммарно на взаимодействия между атомами белка и атомами нуклеиновых оснований приходится 27.1% всех атомных контактов. Анализ площади доступной поверхности атомов большой и малой бороздок показал, что она коррелирует с числом контактов (коэффициенты линейной корреляции 0.94 и 0.93 соответственно), однако атомы нуклеиновых оснований, образующие водородные связи, контактируют чаще, чем этого можно было ожидать из статистических соображений. Показано, что конформационно-стабильные пептиды достаточно часто встречаются в областях связывания с ДНК. Анализ остатков в предопределенной конформации в 3518 комплексах белок–ДНК выявил 159 аминокислотных остатков в предопределенной конформации β-изгиба типа I, 15 остатков в конформации β-изгиба типа I', 6 остатков в конформации β-изгиба типа II. Остатков в конформации β-изгиба типа II' найдено не было. Анализ контактов показал, что такие остатки практически не образуют контактов с ДНК. Контакты с атомами нуклеиновых оснований найдены только в двух гомологичных структурах 3qca и 3qe9, где атомы треонина образуют контакты с атомами нуклеиновых оснований AT-пары.

Ключевые слова: комплексы белок–ДНК; разбиение Вороного–Делоне; взаимодействия белок–ДНК; специфичность узнавания.

Protein-DNA interactions: statistical analysis of interatomic contacts in major and minor grooves

А.А. Анашкина¹, Е.Н. Кузнецов²,
А.В. Батяновский³, Л.А. Урошлев⁴,
В.Г. Туманян¹, Н.Г. Есипова¹

¹ Engelhardt Institute of Molecular Biology RAS, Moscow, Russia

² V.A. Trapeznikov Institute of Control Sciences RAS,
Moscow, Russia

³ Institute of Biophysics and Cell Engineering of NAS of Belarus,
Minsk, Belarus

⁴ Vavilov Institute of General Genetics RAS, Moscow, Russia

The interactions between protein and DNA in essence underlie all processes in a living cell. Understanding the principles of specific recognition of DNA sites will open the way to understand how these processes are controlled and to interfere in their operation. In the paper we studied contacts between the protein and DNA at the atomic level in the structures of all the 3518 protein-DNA complexes available in PDB by the Voronoi–Delaunay tessellation method. The method unambiguously defines contacts between atoms without any parameters, and characterizes each contact by the distance between atoms and the contact area, which is determined by the corresponding face of the Voronoi polyhedron. It was shown that most contacts are formed between the protein atoms and the sugar-phosphate backbone of the DNA (72.9%). The contact with the atoms of the nucleic bases emerging into the grooves of DNA is 17.0% for a major groove and 10.1% for all atomic contacts for a minor groove. Totally, the interaction between protein atoms and nucleic base atoms accounts for 27.1% of all contacts. Analysis of the accessible surface area of atoms in the major and the minor grooves showed a correlation with the number of contacts (coefficient of linear correlation 0.94 and 0.93, respectively), however, nucleic acid atoms forming hydrogen bonds make contacts more often than may be expected from statistical considerations. It was shown that conformationally stable peptides occur sometimes in the binding regions with DNA. Analysis of the residues in a predefined conformation in 3518 protein-DNA complexes revealed 159 amino acid residues in a predefined β-bend type I conformation, 15 residues in the conformation of β-bend type I', and 6 residues in the conformation of β-bend type II. No residues in the conformation of β-bend type II' were

found. Analysis of contacts showed that such residues virtually do not form contacts with DNA. Contacts with nucleic base atoms are found only in the two homologous structures 3qea and 3qe9, where threonine atoms form contacts with atoms of nucleotide bases of the AT-pair.

Key words: protein-DNA complexes; Voronoi–Delaunay tessellation; protein-DNA interaction; specificity of recognition.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Анашкина А.А., Кузнецов Е.Н., Батяновский А.В., Урошлев Л.А., Туманян В.Г., Есипова Н.Г. Взаимодействия белок-ДНК: статистический анализ межатомных контактов в большой и малой бороздках. Вавиловский журнал генетики и селекции. 2017;21(8):887-894. DOI 10.18699/VJ17.309

HOW TO CITE THIS ARTICLE:

Anashkina A.A., Kuznetsov E.N., Batianovskii A.V., Uroshlev L.A., Tumanyan V.G., Esipova N.G. Protein-DNA interactions: statistical analysis of interatomic contacts in major and minor grooves. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2017;21(8):887-894. DOI 10.18699/VJ17.309 (in Russian)

Специфические взаимодействия между белками и соответствующими сайтами ДНК лежат в основе фундаментальных биологических процессов, происходящих в живой клетке. Изменение специфичности связывания может привести к серьезным нарушениям биологической функции и, как следствие, к заболеваниям организма. Так, изменение специфичности связывания мутированных транскрипционных факторов с ДНК наблюдается при разных типах рака (Filipova et al., 2002). В отличие от процессов с участием молекулы-адаптера, как в случае генетического кода, взаимоотношения между мономерными единицами (аминокислотами, нуклеотидами) в интерфейсах белок-ДНК не могут быть произвольными, а подчиняются определенным физико-химическим ограничениям. Гипотеза о том, что эти взаимодействия могут быть описаны неким псевдокодом, пусть гораздо менее определенным, чем генетический код, до сих пор не подтверждена и не опровергнута (Choo, Klug, 1997; Pabo, Nekludova, 2000; Venos et al., 2002), но для отдельных классов белков оказалось возможным выявить некий шаблон взаимодействий белков и ДНК и высказать соображения о причинах специфичности и влияния мутаций на специфичность (Wolfe et al., 2000).

Рассмотрим кратко историю анализа белково-нуклеиновых взаимодействий. В начале в связи с малыми объемами информации по комплексам белок-ДНК все данные суммировались, так как только таким образом можно было сформировать достаточно объемный для статистического анализа материал. Своеобразный итог этого этапа подвела работа Торнтон (Luscombe et al., 2001), в которой изучались 129 комплексов белок-ДНК. Были рассмотрены взаимодействия разных типов: вандерваальсовы контакты, ионные мостики, водородные связи. Уже тогда авторы различали неспецифические, «универсальные» взаимодействия, ответственные за связывание с ДНК в целом, и взаимодействия, «зависящие от контекста», отвечающие за распознавание последовательности нуклеиновых оснований ДНК. В дальнейших исследованиях структура комплексов белок-ДНК анализировалась для определенных типов белков, связывающихся с ДНК. Во-первых, уже были накоплены достаточные объемы данных по белковым семействам. Во-вторых, исследователи исходили из того, что взаимодействие является специфическим для данного класса связывающихся с ДНК белков (например,

цинковых пальцев (Wolfe et al., 2000)). Между тем вопрос об общих закономерностях взаимодействий белок-ДНК оставался открытым.

В нашей работе мы поставили задачу анализа пространственных контактов между атомами в структурах комплексов белок-ДНК и поиска структурных детерминант специфичности. Для начала необходимо было решить, какие данные суммируются (что дает очевидные преимущества в смысле статистики), а какие, напротив, должны быть рассмотрены отдельно в силу их специфической природы. Некоторые авторы полагают, что следует отдельно рассматривать комплексы белок-ДНК с разной специфичностью. Так, выделяют высоко-, мульти- и неспецифические комплексы (Cologna, Guo, 2016). При этом под высокоспецифическими понимают белки, которые распознают единственный сайт на ДНК. Мультиспецифические белки распознают несколько сходных сайтов. Неспецифические белки связывают ДНК независимо от последовательности. Соответственно, комплексы разделены на группы по типу распознающего белка (Cologna, Guo, 2016). Мы полагаем, что такое деление является достаточно условным. Если же оно имеет под собой реальные физико-химические основы, то анализ интерфейсов должен объективным образом выявить такое разделение. Во избежание перепредставленности тех или иных взаимодействий, некоторые авторы составляют также репрезентативную выборку, исключая из рассмотрения комплексы с гомологичными белковыми цепями. Ранее в работе (Anashkina et al., 2007) мы показали, что минимальный для изучения набор должен содержать около 460 интерфейсов для оценки аминокислотного состава интерфейсов с точностью 2 % при достоверности 95 %. Поэтому мы полагаем, принимая во внимание объем существующих данных по белок-ДНК комплексам, что надо рассматривать все имеющиеся в распоряжении исследователей экспериментальные структуры комплексов. Это необходимо для повышения достоверности исследования. Таким образом, мы отбирали комплексы из PDB по наличию в структурах комплексов как белковых цепей, так и ДНК. Однако принципиальным моментом является то, что эти данные должны быть позиционированы относительно большой и малой бороздок и сахарофосфатного остова ДНК, поскольку большая и малая бороздки разделены в пространстве атомами сахарофосфатного остова,

и смешивать данные, полученные для разных бороздок, нелогично.

Материалы и методы

Заметим, что ДНК имеет сольватную оболочку, кроме того, в бороздках ДНК находится «структурированная» вода (Привалов, 1958) и ионы. Далеко не во всех комплексах положения атомов кислорода воды и ионов локализованы. По этой причине опосредованные контакты (например, мостик через воду) не рассматривались.

Разбиение Вороного–Делоне. Контакты между белками и ДНК были определены на основе анализа координат атомов пространственной структуры 3518 комплексов белок-ДНК методом разбиения Вороного–Делоне (Medvedev, 1986; Анашкина и др., 2008; Anashkina et al., 2013), аналогично тому, как это было ранее сделано для белок-белковых взаимодействий (Anashkina et al., 2007). Соответственно, термином «контакт» мы будем называть далее общую грань полиэдра Вороного, полученного в результате разбиения Вороного–Делоне. Каждый контакт характеризуется типом пары атомов, расстоянием между этими атомами и площадью общей грани полиэдра Вороного.

Результатом проведенного разбиения являются таблицы контактов между атомами аминокислот и нуклеотидов, а также между более крупными пространственными единицами – аминокислотными остатками и нуклеотидами. Список использованных в исследовании комплексов белок-ДНК и полные таблицы контактов между атомами белка и ДНК приведены в Приложении¹. Программа для построения разбиения написана на языке C++, ее исходный код доступен по запросу авторам статьи через электронную почту.

Конформационно-стабильные пептиды – короткие аминокислотные последовательности, которые включены в различные белковые глобулы преимущественно в одном конформационном состоянии. Методология выявления таких последовательностей изложена в работах (Батяновский, Власов, 2008; Батяновский и др., 2009) и сводится к сравнению конформационных состояний по совокупностям двугранных углов. При наличии преимущественной (более 60 % всех случаев) группы близких конформаций утверждается наличие конформационной стабильности у данной короткой последовательности.

Предопределенные конформации. Конформационно-предопределенные конформации – это локальные области полипептидной цепи белка, конформация которых определяется контекстом, т. е. участками полипептидной цепи справа и слева от этой области (Урошлев и др., 2015; Торшин и др., 2016). Типичным примером такого рода конформаций являются β-изгибы полипептидной цепи. Фиксация остатков β-шпильки, примыкающих к изгибу, резко ограничивает число конформационных вариантов вследствие образования псевдоцикла. Механизм «навязывания» изгибу, соединяющему два антипараллельных β-тяжа, определенной конформации объясняет образование β-изгибов I, I' и II типов, а также β-изгиба II' типа с напряженным в стереохимическом смысле элементом.

¹ См. Приложение по адресу:
<http://www.bionet.nsc.ru/vogis/download/pict-2017-21/appx12.pdf>

Результаты

В настоящей работе мы поставили цель подробно проанализировать весь имеющийся экспериментальный материал по белок-ДНК комплексам в поисках закономерностей связывания белков для каждой из бороздок отдельно, выделяя общие и уникальные черты в обоих случаях. На момент проведения исследования в банке данных PDB было найдено 3518 комплексов белок-ДНК. На каждой стадии анализа были выявлены межатомные взаимодействия, ответственные за специфическое распознавание (распознавание последовательности нуклеотидов), и сделано заключение в пользу или против существования подобия кода узнавания на уровне взаимодействия атомов или мономеров нуклеиновых кислот и белков.

На рис. 1 показано разделение атомов нуклеиновых оснований ДНК на большую и малую бороздки. В большую бороздку выходят атомы N9, C8, N7, C5, C6, N6 аденина, N1, C6, C5, C8, C4, O4 тимина, N9, C8, N7, C5, C6, O6 гуанина, N1, C6, C5, C4, N4 цитозина. В малую бороздку выходят атомы N9, C4, N3, C2, N1 аденина, N1, C2, O2, N3 тимина, N9, C4, N3, C2, N2 гуанина и N1, C2, O2 цитозина. Заметим, что атомы N9 аденина и гуанина и атомы N1 тимина и цитозина могут выходить как в малую, так и в большую бороздку (см. рис. 1), однако трехмерное строение молекулы ДНК таково, что эти атомы не доступны для контакта со стороны малой бороздки. Таким образом,

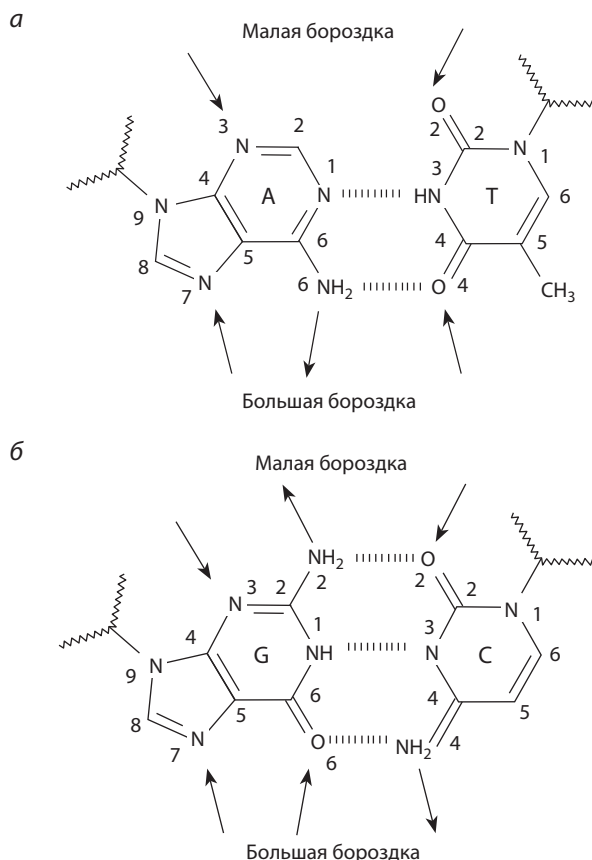


Рис. 1. Атомы ДНК, выходящие в большую и малую бороздки: а – пара АТ; б – пара ГЦ.

Стрелками показаны атомы, являющиеся донорами и акцепторами потенциальных водородных связей.

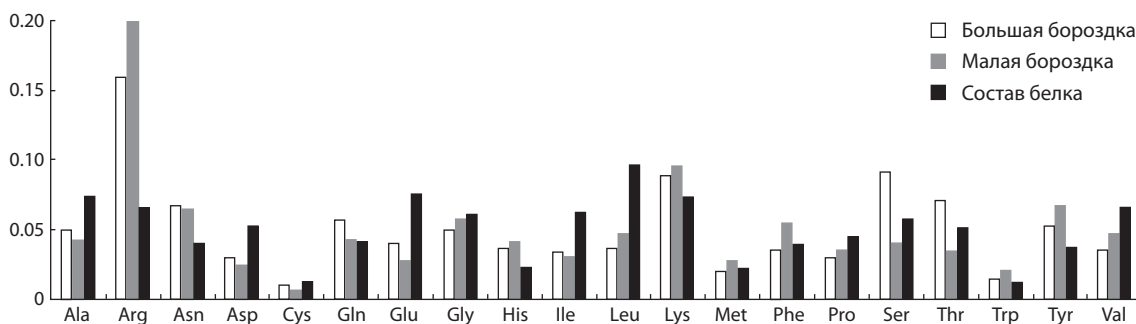


Рис. 2. Аминокислотный состав (в долях) интерфейсов белок-ДНК по большой и малой бороздкам, а также состав белков в целом.

Таблица 1. Количество контактов атомов боковых радикалов и основной цепи белков с атомами большой и малой бороздок, дезоксирибозной и фосфатной групп ДНК*

Атомы белка	Атомы ДНК			
	большой бороздки	малой бороздки	дезоксирибозы	фосфатной группы
Основная цепь	33380 (2.6)**	18818 (1.8)	145773 (11.2)	108173 (8.6)
Боковые радикалы	181801 (14.4)	85026 (8.3)	444666 (33.6)	246468 (19.5)

* В расчетах использована выборка из 3518 комплексов белок-ДНК.

** В скобках указано процентное содержание от общего числа контактов.

Таблица 2. Количество контактов атомов аргинина с атомами ДНК малой бороздки, рассчитанное на выборке 3518 комплексов белок-ДНК

Атомы белка	Атомы нуклеиновых оснований ДНК														Сумма
	DA_C4	DA_N3	DA_C2	DA_N1	DT_N3	DT_O2	DT_C2	DG_C4	DG_N3	DG_C2	DG_N2	DC_O2	DC_C2		
ARG_C	1	14	19	8	3	24	0	2	6	0	19	27	0	123	
ARG_CA	8	88	21	11	12	49	2	65	31	26	50	80	9	452	
ARG_CB	22	205	172	32	45	246	52	27	42	7	110	151	24	1135	
ARG_CD	65	463	278	61	64	424	45	44	157	81	266	265	79	2292	
ARG_CG	28	245	173	36	54	268	44	65	97	52	133	141	30	1366	
ARG_CZ	36	277	229	62	87	310	41	64	157	35	300	193	28	1819	
ARG_N	7	163	140	12	10	127	5	7	26	25	64	67	12	665	
ARG_NE	79	353	241	75	94	353	85	101	209	77	290	272	99	2328	
ARG_NH1	253	1263	949	120	167	1187	184	190	606	134	799	731	133	6716	
ARG_NH2	207	1132	872	97	231	1415	161	188	645	125	988	958	183	7202	
ARG_O	25	75	70	29	28	104	21	39	77	11	57	110	5	651	

Примечание. Серым фоном выделены контакты атомов азота гуанидиновой группы.

контакты этих атомов могут быть отнесены только к большой бороздке. Атомы дезоксирибозы и фосфатной группы полагаем одинаковыми для всех нуклеотидов.

Для начала следует упомянуть, что области связывания белок-ДНК обогащены положительно заряженными аргинином, лизином и гистидином и образующими водородные связи аспарагином, серином, тирозином, в то время как содержание отрицательно заряженных и гидрофобных аминокислот понижено (рис. 2). Также можно видеть, что наблюдаются различия в аминокислотном составе областей взаимодействия белка с большой и малой бороздкой. Наиболее ярко это различие наблюдается для серинов и треонинов, которых много в большой бороздке и суще-

ственно меньше в малой. Тирозин и аргинин, напротив, чаще встречаются в малой бороздке, чем в большой.

Мы разделили весь массив контактов между атомами белков и ДНК на три категории: контакты с атомами нуклеиновых оснований по большой бороздке, по малой бороздке и контакты с атомами сахарофосфатного остова ДНК. Согласно полученным данным, большая часть контактов (53.1 %) образуется между атомами боковых радикалов аминокислот белка и сахарофосфатным остовом ДНК. Контакты боковых радикалов аминокислот белка с атомами нуклеиновых оснований, выходящих в бороздки ДНК, составляют 14.4 % в большой бороздке и 8.3 % в малой (от всех атомных контактов). Атомы основной

Таблица 3. Наиболее частые контакты атомов белка с атомами большой и малой бороздок, а также сахарофосфатного остова

Большая бороздка			Малая бороздка			Сахарофосфатный остов		
Тип атома белка	Контакты с ДНК		Тип атома белка	Контакты с ДНК		Тип атома белка	Контакты с ДНК	
	число	%		число	%		число	%
ARG_NH2	16469	7.4	ARG_NH2	7202	6.7	ARG_NH2	43071	4.5
ARG_NH1	13251	13.3	ARG_NH1	6716	13.0	ARG_NH1	37252	8.4
LYS_NZ	6857	16.4	LYS_NZ	3583	16.3	LYS_NZ	33617	11.9
ARG_NE	6361	19.3	PHE_CZ	2366	18.5	ARG_CD	24686	14.5
ARG_CZ	6329	22.1	ASN_ND2	2342	20.7	ARG_NE	23775	17.0
SER_OG	5508	24.6	ARG_NE	2328	22.8	SER_OG	22962	19.4
ASN_ND2	5057	26.9	ARG_CD	2292	25.0	SER_CB	20718	21.6
ARG_CD	4961	29.1	LYS_CE	2239	27.0	THR_CG2	20579	23.7
GLN_NE2	4378	31.1	PHE_CE1	2183	29.1	GLY_CA	20486	25.9
GLN_OE1	4266	33.0	GLY_CA	1919	30.9	LYS_CE	20382	28.0
ASN_OD1	4131	34.8	TYR_OH	1918	32.7	THR_OG1	20331	30.2
THR_OG1	4100	36.7	ARG_CZ	1819	34.3	LYS_CD	19627	32.2
THR_CG2	3694	38.3	LEU_CD1	1720	35.9	ARG_CZ	17349	34.0
LYS_CE	3661	40.0	PHE_CE2	1640	37.5	ARG_CG	17285	35.8
ALA_CB	3542	41.6	VAL_CG1	1486	38.9	LYS_CB	15656	37.5
SER_CB	3426	43.1	LEU_CD2	1480	40.2	ALA_CB	15401	39.1
TYR_OH	3076	44.5	VAL_CG2	1391	41.5	ARG_CB	15005	40.7
GLY_CA	2891	45.8	ARG_CG	1366	42.8	LYS_CG	14285	42.2
ARG_CG	2826	47.1	ALA_CB	1353	44.1	ASN_ND2	14250	43.7
LYS_CD	2756	48.3	LYS_CD	1314	45.3	GLY_N	13934	45.1
GLU_OE2	2201	49.3	ILE_CD1	1280	46.5	TYR_OH	13928	46.6
GLU_OE1	2188	50.3	GLN_NE2	1259	47.6	GLY_O	13516	48.0
			TYR_CE2	1256	48.8	LYS_O	10699	49.1
			PRO_CG	1251	50.0	GLN_NE2	10250	50.2

Примечание. По каждой из трех категорий приведены контакты, суммарно составляющие более 50 % от числа контактов в данной категории.

цепи также образуют контакты с ДНК, однако в меньшем количестве (табл. 1).

Частота контактов атомов в большой, малой бороздках и с сахарофосфатным остовом

Мы полагаем, что все контакты можно разделить на два класса: специфические контакты, обеспечивающие ключевые взаимодействия, и неспецифические, образующиеся вследствие пространственного сближения атомов, расположенных вблизи от атомов, участвующих в специфических контактах. Например, представим себе водородную (или ионную) связь между атомом белка и атомом ДНК в сайте взаимодействия. Разбиение Вороного–Делоне «увидит» в такой ситуации контакт между этими двумя атомами. В то же время разбиение Вороного–Делоне выявит много других контактов, образованных соседними по структуре атомами белка и ДНК вследствие их сближения в пространстве. Пусть эта водородная (ионная) связь будет взаимодействием «специфическим», участвующим именно в распознавании сайта ДНК, тогда разбиением Вороного–Делоне мы получаем смесь «специфических» и «неспецифических» контактов. В случае применения метода определения контактов между атомами по критерию расстояния ситуация еще сильнее ухудшается, поскольку

дополнительно возникает неопределенность, вызванная неточностью и сильной зависимостью результатов от выбора величины критического расстояния.

Рассмотрим контакты атомов аминокислоты аргинина, а именно контакты этой аминокислоты с атомами нуклеиновых оснований, выходящих в малую бороздку (табл. 2). Из табл. 2 видно, что большая часть контактов (65.6 %) образуется с участием атомов азота гуанидиновой группы.

Таким образом, можно считать контакты атомов азота NH1 и NH2 гуанидиновой группы наиболее значимыми, а контакты других атомов аргинина вторичными, образующимися вследствие пространственного сближения атомов азота NH1 и NH2 гуанидиновой группы со своими атомами-мишенями.

Рассматривая контакты каждой аминокислоты, можно выделить для нее наиболее важные атомы, образующие много контактов. Таким же образом, выделяя наиболее часто контактирующие атомы, можно рассматривать атомы ДНК. Результат такого двойного отбора приведен в Приложении.

Кроме того, мы отобрали наиболее часто контактирующие атомы белка в каждой из трех рассматриваемых категорий контактов (табл. 3). Для всех категорий получилось, что наиболее часто контактируют атомы азота

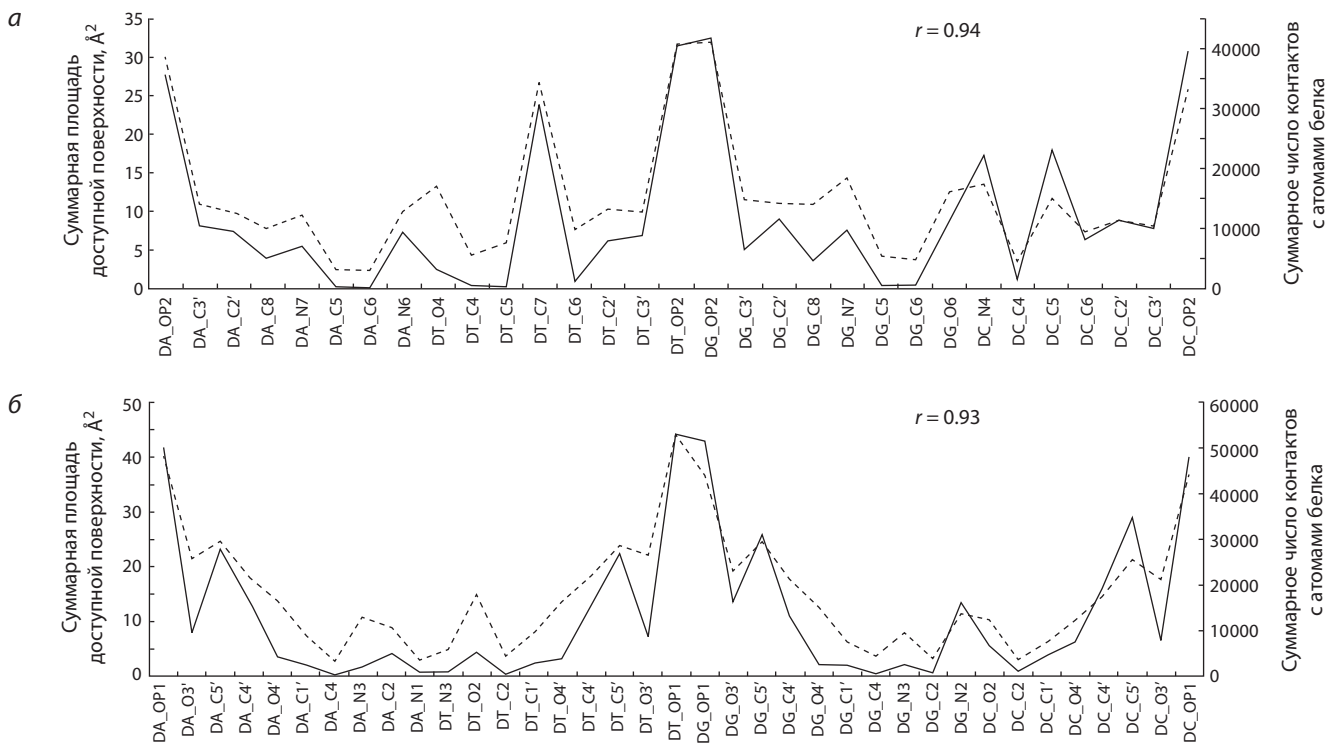


Рис. 3. Суммарное количество контактов и площадь доступной поверхности атомов большой (а) и малой (б) бороздок.

Сплошной линией показана площадь доступной поверхности, штриховой – количество контактов, образуемых атомами. Группы оснований ДНК, выходящие в большую и малую бороздки, расположены в соответствии с их порядком в пространственной структуре основания.

гуанидиновой группы аргинина и аминогруппы лизина, также следует отметить атомы амидной группы аспарагина и глутамина. В большой бороздке часто контактируют гидроксильные группы серина, треонина, тирозина и атомы карбоксильной группы глутаминовой кислоты.

В малой бороздке часто контактирует гидроксильная группа тирозина, наблюдается также большое число контактов с гидрофобными атомами фенилаланина, лейцина, валина, изолейцина, аланина, глицина. С сахарофосфатным остовом часто контактируют атомы аргинина, лизина, серина, треонина, глицина, глутамина, причем как атомы полярных групп, так и гидрофобные (см. табл. 3).

Сравнение частоты контактов с площадью доступной поверхности для атомов ДНК большой и малой бороздок

Мы проанализировали, насколько соответствует частота контактов с атомами ДНК, выходящими в большую и малую бороздку, доступности этих атомов растворителю (рис. 3).

Сравнение показало, что, несмотря на очень хорошее соответствие количества контактов и площади доступной поверхности (коэффициенты корреляции 0.94 и 0.93 для атомов большой и малой бороздок соответственно), некоторые атомы образуют больше контактов, чем можно было бы ожидать из площади доступной поверхности. Это атомы кислорода и азота DA_N7, DT_O4, DG_N7, DG_O6, выходящие в большую бороздку, и атомы DA_N3, DT_O2, DG_N3, DC_O2, выходящие в малую бороздку. Для таких атомов превышение может быть более чем двукратным

(см. рис. 3). Атомы сахарофосфатного остова и углероды краев оснований, выходящих в бороздки, не проявляют подобной тенденции. Естественно предположить, что наблюдаемая картина связана с возможностью или невозможностью образования водородной связи.

Конформационно-стабильные пептиды в комплексах белок-ДНК

Анализ расположения конформационно-стабильных пептидов в структуре ДНК-связывающих белков показал, что такие пептиды встречаются в разных местах белка, а также в участках взаимодействия между белковыми субъединицами в составе белкового комплекса и в области взаимодействия с ДНК. На рис. 4 приведен пример разметки конформационно-стабильных пептидов в структуре комплекса белок-ДНК (1a02).

Вопрос расположения конформационно-стабильных пептидов в области взаимодействия с ДНК и их роли в распознавании ДНК требует более подробного дальнейшего анализа.

Взаимодействия остатков в предопределенной конформации с ДНК

В изученных комплексах белок-ДНК было выявлено 159 аминокислотных остатков в предопределенной конформации β -изгиба типа I, 15 остатков в конформации β -изгиба типа I', 6 остатков в конформации β -изгиба типа II. Остатков в конформации β -изгиба типа II' найдено не было. Всего был обнаружен 91 межатомный контакт между аминокислотами в предопределенной конформа-

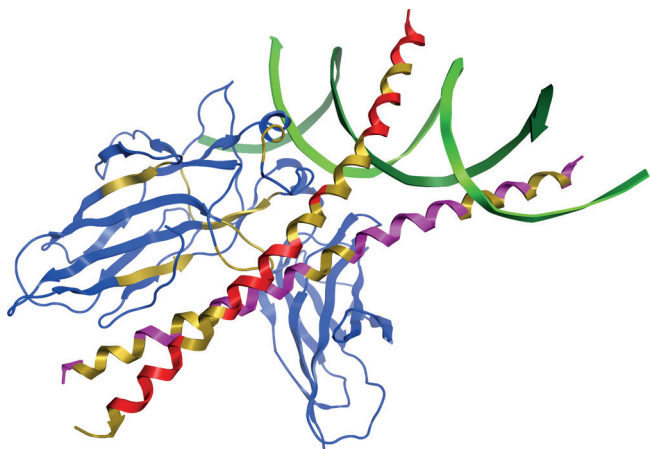


Рис. 4. Структура комплекса ДНК-связывающих доменов белков NFAT (синий), FOS (розовый) и JUN (красный) с ДНК (зеленый и темно-зеленый) (код Protein Data Bank 1a02). Желтым цветом выделены конформационно-стабильные пептиды.

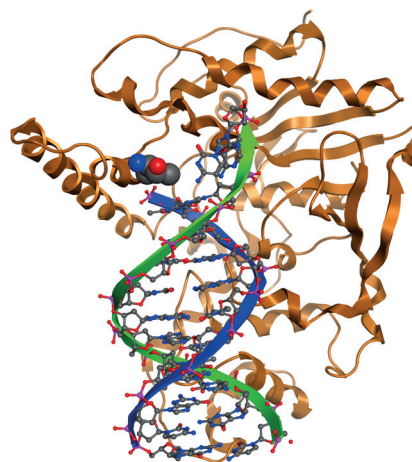


Рис. 5. Остаток T120 в предопределенной конформации β-изгиба типа I образует контакты с атомами нуклеиновых оснований пары АТ в структуре 3qeа.

ции и атомами ДНК: 72 типа I и 19 типа II. Почти все эти контакты образованы с атомами сахарофосфатного остова ДНК. Контакты с атомами нуклеиновых оснований найдены только в двух гомологичных структурах 3qeа и 3qe9 (34 контакта). В этих структурах остаток треонина в предопределенной конформации I типа взаимодействует с атомами нуклеиновых оснований АТ пары (рис. 5).

Таким образом, аминокислотные остатки в предопределенной конформации основной цепи могут принимать участие во взаимодействии с ДНК, однако частота таких событий достаточно мала.

Обсуждение

Можно надеяться, что выяснение принципов специфического распознавания сайтов ДНК в перспективе откроет путь к управлению важнейшими процессами в живых системах. Очевидным методическим подходом для исследований в рамках этой проблемы является анализ экспериментально определенных структур белково-нуклеиновых комплексов. Ранее нами была создана программа для построения пространственного разбиения Вороного–Делоне на системе точек, соответствующих расположению атомов в комплексах белок-ДНК (Anashkina et al., 2007). В рамках этого однозначного математического построения контактирующие атомы определяются как соседние вершины разбиения Делоне или как атомы, имеющие общие грани полиэдров Вороного ненулевой площади. Во всех работах других исследователей определение контактирующих атомов проводится на основе критерия расстояния. Однако, как показано нами ранее, это ведет к систематическим ошибкам: недоопределению числа контактов при выборе недостаточно большого критического расстояния и переопределению числа контактов при выборе слишком большого критического расстояния. В последнем случае фактически между выделенными контактирующими атомами может находиться один атом или более. Особенно трудно контролировать вхождение молекулы воды в такой искусственно увеличенный контакт.

В последнее время основное внимание исследователей обращено на изучение отдельных семейств распознающих ДНК белков. Нашей задачей было включить в рассмотрение всю совокупность имеющихся данных с целью получения характеристик, общих для разных распознающих схем, причем суммирование данных сопровождалось дифференциацией по бороздкам. Аналогичная работа была проделана в свое время Торнтон с соавторами (Luscombe et al., 2001), но статистика взаимодействий отдельно по бороздкам ими не рассматривалась. В цитируемой работе было исследовано 129 комплексов белок-ДНК, в то время как в нашей работе – 3518 комплексов. Для изучения были отобраны все имеющиеся в PDB на весну 2017 г. комплексы белок-ДНК. Мы проанализировали контакты между атомами белка и ДНК в структурах белок-ДНК комплексов методом Вороного–Делоне, что позволило исключить не прямые взаимодействия, которые обычно попадают в поле зрения при использовании критерия расстояния, тем самым «зашумляя» картину. В предположении, что существуют специфические контакты, управляющие распознаванием, и неспецифические, образующиеся вследствие сближения атомов в пространстве, мы выделили наиболее часто контактирующие атомы белка и ДНК. Так, например, контакты всего лишь трех атомов азота гуанидиновой группы аргинина образуют с сахарофосфатным остовом 10.9 %, с малой бороздкой 15.1 % и с большой бороздкой 16.2 % контактов от общего числа контактов в каждой категории соответственно (см. табл. 3).

Наши вычисления показали, что три десятка типов атомов из 13 аминокислот образуют более 50 % контактов во всех трех категориях контактов: по большой бороздке, по малой бороздке и с сахарофосфатным остовом (см. табл. 3). Мы пришли к выводу, что число контактов атомов нуклеотидов с атомами белка хорошо коррелирует с площадью доступной поверхности (коэффициент линейной корреляции по большой бороздке 0.94, по малой – 0.93). Однако для атомов кислорода и азота DA_N7, DT_O4, DG_N7, DG_O6, выходящих в большую бороздку, и

атомов DA_N3, DT_O2, DG_N3, DC_O2, выходящих в малую бороздку, показано, что они контактируют чаще, чем можно ожидать исходя из площади доступной поверхности. Для таких атомов превышение может быть более чем двукратным (см. рис. 3). Это атомы кислорода и азота нуклеиновых оснований, выходящие в малую и большую бороздку и способные к формированию водородных связей, будучи донорами и/или акцепторами таких связей.

Что касается роли конформационно-стабильных пептидов в структуре ДНК-связывающих белков, то по нашим наблюдениям такие пептиды достаточно часто встречаются в разных местах белка, а также в участках взаимодействия между белковыми субъединицами в составе белкового комплекса и в области взаимодействия с ДНК, что требует более подробного анализа.

Анализ аминокислотных остатков в предопределенной конформации основной цепи показал, что эти остатки могут принимать участие во взаимодействии с ДНК, однако частота таких событий достаточно мала. Полученные в работе данные могут оказаться полезными, в частности, для проектирования рекомбинантных белков, распознающих последовательность ДНК.

Благодарности

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проекты № 15-04-99605а, 17-04-02105а, 16-54-00219 Бел_а) и Белорусского республиканского фонда фундаментальных исследований (проект № Б16Р178).

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Список литературы

- Анашкина А.А., Туманян В.Г., Кузнецов Е.Н., Галкин А.В., Есипова Н.Г. Геометрический анализ ДНК-белковых взаимодействий на основе метода Вороного–Делоне. *Биофизика*. 2008;53(3): 402-406.
- Батяновский А.В., Власов П.К. Короткие фрагменты белковой глобулы с преобладающей конформацией. *Биофизика*. 2008;53(4): 556-561.
- Батяновский А.В., Есипова Н.Г., Шноль С.Э. О взаимном расположении коротких конформационно-стабильных олигопептидов в структуре глобулярных белков. *Биофизика*. 2009;54:1137-1143.

- Привалов П.Л. К вопросу о состоянии и роли воды в биологических системах. *Биофизика*. 1958;3(6):738-743.
- Торшин И.Ю., Урошлев Л.А., Есипова Н.Г., Туманян В.Г. Дескриптивная статистика «запрещенных» областей и различных типов вторичной структуры белка в контексте исследования «перекрыченных» бета-шпилек. *Биофизика*. 2016;61(1):11-17.
- Урошлев Л.А., Торшин И.Ю., Батяновский А.В., Есипова Н.Г., Туманян В.Г. Запрещенные конформации полипептидной цепи на примере β -изгиба β -шпильки в SH3-доме α -спектрина. *Биофизика*. 2015;60(1):5-14.
- Anashkina A.A., Kuznetsov E.N., Batyanovskii A.V., Gnuchev N.V., Tumanyan V.G., Esipova N.G. Classification of amino acids based on comparative analysis of contacts in DNA-protein complexes and specific DNA-protein interactions. *Biophysics*. 2013;58(6): 766-770.
- Anashkina A.A., Kuznetsov E.N., Esipova N.G., Tumanyan V.G. Comprehensive statistical analysis of residues interaction specificity at protein-protein interfaces. *Proteins: Struct. Funct. Bioinf*. 2007; 67(4):1060-1077. DOI 10.1002/prot.21363.
- Benos P.V., Lapedes A.S., Stormo G.D. Is there a code for protein-DNA recognition? *Probab(ilistical)ly...* *BioEssays*. 2002;24(5):466-475. DOI 10.1002/bies.10073.
- Choo Y., Klug A. Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol*. 1997;7(1):117-125. DOI 10.1016/S0959-440X(97)80015-2.
- Corona R.I., Guo J.-T. Statistical analysis of structural determinants for protein-DNA-binding specificity. *Proteins*. 2016;84(8):1147-1161. DOI 10.1002/prot.25061.
- Filippova G.N., Qi C.-F., Ulmer J.E., Moore J.M., Ward M.D., Hu Y.J., Loukinov D.I., Pugacheva E.M., Klenova E.M., Grundy P.E., Feinberg A.P., Cleton-Jansen A.M., Moerland E.W., Cornelisse C.J., Suzuki H., Komiyama A., Lindblom A., Dorion-Bonnet F., Neiman P.E., Morse H.C., Collins S.J., Lobanenkov V.V. Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. *Cancer Res*. 2002;62(1):48-52.
- Luscombe N.M., Laskowski R.A., Thornton J.M. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res*. 2001;29(13):2860-2874. DOI 10.1093/nar/29.13.2860.
- Medvedev N.N. The algorithm for three-dimensional Voronoi polyhedra. *J. Comput. Physics*. 1986;67(1):223-229. DOI 10.1016/0021-9991(86)90123-3.
- Pabo C.O., Nekludova L. Geometric analysis and comparison of protein-DNA interfaces: Why is there no simple code for recognition? *J. Mol. Biol*. 2000;301(3):597-624. DOI 10.1006/jmbi.2000.3918.
- Wolfe S.A., Nekludova L., Pabo C.O. DNA recognition by Cys₂His₂ zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct*. 2000;29: 183-212. DOI 10.1146/annurev.biophys.29.1.183.