

Специфичность контактов в комплексах белок-ДНК

А. А. Анашкина^{1,a}, Н. Г. Есипова¹, Е. Н. Кузнецов², В. Г. Туманян¹

¹ Институт молекулярной биологии им. В. А. Энгельгарда,
119991, г. Москва, ул. Вавилова, д. 32

² Институт проблем управления им. В. А. Трапезникова РАН,
117997, ГСП-7, В-342, г. Москва, ул. Профсоюзная, д. 65

E-mail: ^anasty@imb.ac.ru

Получено 14 апреля 2009 г.

С помощью разбиения Вороного–Делоне определены контакты между белками и нуклеиновыми кислотами в комплексах белок-ДНК. Показано, что приблизительно треть всех контактов составляют контакты с положительно заряженными аминокислотами Arg и Lys, 32,3 %. Ser и Thr дают в сумме следующий по величине вклад — 15 %. Asn образует 6 % контактов, также как и Gly. Вклад каждой из остальных аминокислот не превышает 5 %. Установлено, что статистически значимыми являются контакты типа Asp-G, Trp-C, Glu-C, Asp-C и His-T. Предложен общий механизм участия отрицательно заряженных остатков в специфических взаимодействиях белок-ДНК.

Ключевые слова: разбиение Вороного–Делоне, комплексы белок-ДНК

Contact specificity in protein-DNA complexes

A. A. Anashkina¹, N. G. Esipova¹, E. N. Kuznetsov², V. G. Tumanyan¹

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

² Russian Academy of Sciences Institute of Control Sciences, Profsoyuznaya str. 65, Moscow GSP-7, B-342, 117997, Russia

Abstract. — In this work we investigated contacts between proteins and nucleic acids by Voronoi-Delaunay tessellation. About one third of all contacts are contacts between nucleotides and positively charged residues Arg and Lys, 32,3 %. Ser and Thr together take part in 15 % of contacts. Asn forms 6 % of contacts, as well as Gly. Contribution of each other residue type does not exceed 5 %. Statistically significant are contacts Asp-G, Trp-C, Glu-C, Asp-C and His-T. General mechanism of charged residues participation in specific protein-DNA interactions is suggested.

Key words: Voronoi–Delaunay tessellation, protein-DNA complexes

Citation: *Computer Research and Modeling*, 2009, vol. 1, no. 3, pp. 281–286 (Russian).

Введение

Сайт-специфические взаимодействия между ДНК и белками играют ключевую роль в большинстве биологических процессов, включая транскрипцию, трансляцию и рекомбинацию [1]. Очевидно, что в создание комплекса ДНК-белок вносят свой вклад множество различных факторов [2, 3, 4]. Однако общие характеристики специфического связывания можно оценить методами статистического анализа частоты контактов между специфическими остаток/основание [5].

В данной работе мы исследовали контакты между белками и нуклеиновыми кислотами, определенные с помощью разбиения Вороного–Делоне [6, 7]. Этот непараметрический метод позволяет по атомным координатам определить контакты элементов разных уровней строения биомакромолекул, однозначно и математически строго [8]. Для анализа были отобраны 1109 структур комплексов белок-ДНК в PDB. Проведя построение Вороного–Делоне, мы определили контакты между атомами белков и ДНК, а на их основе контакты между соответствующими аминокислотами и нуклеотидами. Напомним, что контактом между атомами мы называем общую грань полиэдра Вороного ненулевой площади, а контакт между аминокислотой и нуклеотидом определяется как совокупность общих граней полиэдров Вороного составляющих их атомов. Площадь такого контакта определяется как сумма площадей граней составляющих его атомных контактов.

Модель

Анализ специфичности контактов между аминокислотными остатками и нуклеотидами ДНК построен на гипотезе о независимости признаков: значение, принятое признаком А, не влияет на вероятности возможных значений признака В, т. е. $P(A_i B_j) = P(A_i) \cdot P(B_j)$ для всех A_i и B_j .

Пусть $n_{i\bullet} = \sum_j n_{ij}$ — сумма всех проявлений признака А, а $n_{\bullet j} = \sum_i n_{ij}$ — признака В,

$n = \sum_i \sum_j n_{ij}$ — общее число наблюдений. По теореме Бернулли, при $n \rightarrow \infty$:

$$\frac{n_{ij}}{n} \rightarrow p_{ij}; \quad \frac{n_{i\bullet}}{n} \rightarrow p_{i\bullet}; \quad \frac{n_{\bullet j}}{n} \rightarrow p_{\bullet j};$$

поэтому для независимых признаков число ожидаемых контактов:

$$n_{ij} \cong n_{i\bullet} \cdot n_{\bullet j} / n.$$

В качестве величины, отражающей соотношение между числом наблюдаемых и ожидаемых контактов, был выбран индекс представленности контакта g_{ij} как отношение числа наблюдаемых контактов c_{ij} (вычисленных по разбиению Вороного) и ожидаемых значений n_{ij} , определенных для таблиц сопряженности:

$$g_{ij} = \frac{c_{ij}}{n_{ij}}.$$

Результаты

Аминокислотный состав интерфейсов белок-ДНК. На рис. 1 представлен аминокислотный состав ДНК-связывающих белков и интерфейсов белок-ДНК. Заметное увеличение доли аминокислот в аминокислотном составе интерфейсов белок-ДНК по сравнению с аминокислотным

составом ДНК-связывающих белков, практически вдвое, наблюдается для положительно заряженных остатков Arg и Lys, а также заметное увеличение наблюдается для содержащих гидроксильную группу Ser и Thr. Соответственно, заметное уменьшение доли аминокислот наблюдается для отрицательно заряженных аминокислот Asp и Glu, а также для неполярных аминокислот Ala, Cys, Ile, Leu, Met, Phe, Pro.

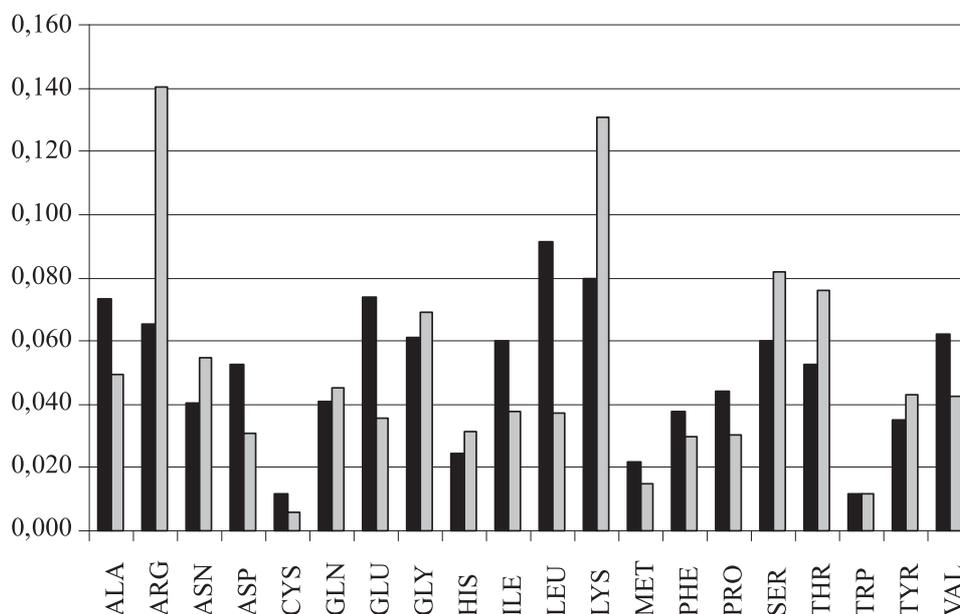


Рис. 1. Аминокислотный состав интерфейсов белок-ДНК в долях. Черные столбики — доли аминокислот в составе ДНК-связывающих белков. Серые столбики — доли аминокислот в интерфейсе белок-ДНК

Контакты и индексы представленности. В табл. 1 приведены числа контактов между аминокислотными остатками и нуклеотидами ДНК, а также индексы представленности g_{ij} . Можно видеть, что приблизительно треть всех контактов составляют контакты с положительно заряженными аминокислотами Arg и Lys, 32,3 %. Ser и Thr дают следующий по величине вклад — 15 %. Asn образует 6 % контактов, также как и Gly. Вклады остальных аминокислот не превышают 5 %.

Рассматривая индексы представленности g_{ij} , отражающие отношение числа наблюдаемых контактов к числу ожидаемых (табл. 1), мы наблюдаем очень интересный перекоп для Asp: «нежелание» взаимодействовать с тимином и огромное предпочтение к гуанину. Итак, согласно индексам представленности табл. 1 для аминокислот, мы можем выделить следующие предпочтения: для аденина — Pro(1,12). Для тимина — His(1,24), а также еще Met, Phe, Val. Для гуанина — Asp(1,47), с меньшим предпочтением Met. Для цитозина — Trp(1,46), Glu(1,33) и Asp(1,23), а также Gln, Cys, Ala.

Итак, количественная оценка контактов аминокислота — нуклеотид в ДНК-белковых комплексах дала интересный результат: наибольший индекс представленности обеспечивают Asp и Glu. Это важно, так как отражает значение отрицательных зарядов для нахождения специфического места посадки белка на ДНК [7].

Исследование отрицательно заряженных аминокислот на поверхности комплексов белок-ДНК показало, что в области связывания в 83 % случаев (интерфейсов) они присутствуют в количестве от одного до десяти. Среднее число отрицательных зарядов на интерфейс составляет 2,57. Интерфейсы с наибольшими количествами отрицательных зарядов (от 8 до 10) образуются такими ДНК-связывающими белками, как различные эндонуклеазы, ДНК лигаза I, обратная транскриптаза, ДНК полимеразы T7, I и IV.

Таблица 1

**Контакты между аминокислотными остатками и нуклеотидами ДНК
и индексы представленности g_{ij}**

	Контакты					Индексы представленности			
	A	T	G	C	Σ	A	T	G	C
ALA	1 283	1 456	1 313	1 268	5 320	0,95	0,98	0,97	1,13
ARG	6 092	6 192	6 463	4 698	23 445	1,02	0,94	1,08	0,95
ASN	2 013	2 284	1 746	1 499	7 542	1,05	1,08	0,91	0,94
ASP	660	476	1 162	804	3 102	0,84	0,55	1,47	1,23
CYS	149	183	165	155	652	0,90	1,00	0,99	1,13
GLN	1 561	1 564	1 349	1 486	5 960	1,03	0,94	0,89	1,18
GLU	869	893	1 036	1 091	3 889	0,88	0,82	1,05	1,33
GLY	1 933	2 235	1 772	1 552	7 492	1,02	1,06	0,93	0,98
HIS	953	1 379	1 014	627	3 973	0,94	1,24	1,00	0,75
ILE	1 090	1 215	1 011	827	4 143	1,04	1,05	0,96	0,95
LEU	1 069	1 196	882	829	3 976	1,06	1,07	0,87	0,99
LYS	4 527	4 654	4 291	3 630	17 102	1,04	0,97	0,98	1,01
MET	383	599	545	359	1 886	0,80	1,13	1,13	0,90
PHE	852	1 144	953	645	3 594	0,93	1,13	1,04	0,85
PRO	961	1 016	712	701	3 390	1,12	1,07	0,82	0,98
SER	2 298	2 667	2 527	1 862	9 354	0,97	1,02	1,06	0,95
THR	2 432	2 831	2 244	2 003	9 510	1,01	1,06	0,93	1,00
TRP	276	330	294	401	1 301	0,83	0,90	0,89	1,46
TYR	1 287	1 480	1 505	1 157	5 429	0,93	0,97	1,09	1,01
VAL	1 256	1 461	1 053	863	4 633	1,07	1,12	0,89	0,89

Заключение

Разбиение Вороного–Делоне позволило получить однозначную картину распределения контактов в интерфейсах ДНК-белковых комплексов. Это дает возможность выделить те контакты, которые важны для организации специфического взаимодействия ДНК-белок. Это дает надежду на установление специфической топологии распределения зарядов на интерфейсах ДНК-связывающих белков и ДНК в качестве элементов иерархически организованного кода ДНК-белкового узнавания. В этом смысле мы предложили определенные выводы о природе неспецифических и специфических взаимодействий ДНК-белок путем анализа результирующего геометрического ансамбля, который возникает после завершения процесса посадки белка на ДНК.

Примечательно, что доля положительно заряженных аминокислот в интерфейсах белок-ДНК выше, чем в составе ДНК-связывающих белков. Возможно, происходит «аккумуляция» положительно заряженных аминокислот на поверхности связывания с ДНК. Кроме того, наличие в этой же области отрицательно заряженных аминокислот, предпочтительно связывающихся с GC-парами, означает, что для возникновения кодовых соотношений AT пары должны чередоваться с GC-парами специфическим образом.

Очевидно, что связывание полианионной ДНК зависит от кластеров положительно заряженных атомов в ДНК-связывающем белке. Для положительных зарядов мы наблюдаем предпочтительность связывания с гистидином только у тимина, а с триптофаном только у цитозина. Контакты с аргинином и лизином распределены равномерно по всем нуклеотидам, при этом содержание гистидина и триптофана невелико. Таким образом, на данном этапе анализа контактов ДНК-белок мы говорим о специфичности взаимодействия с нуклеотидами только для отрицательно заряженных аминокислот. Ориентация ДНК-связывающего белка относительно ДНК определяется положительно заряженными остатками Lys и Arg, суммарным зарядом белка, а также его дипольным моментом [9, 10]. Мы полагаем, что отрицательно заряженные остатки в связывающем центре также могут способствовать разделению белка и ДНК, необходимому на различных стадиях экспрессии генов.

То обстоятельство, что в числе специфических контактов между белками и нуклеиновыми кислотами обнаружены отрицательно заряженные белковые радикалы аминокислот, подтверждается имеющимися литературными данными. Так, квантово-химическими методами Бернар Пюльман [11] показал, что участки GC-пар, выходящие в бороздки, заряжены более положительно, чем остальные аналогичные участки в районах AT-пар двуспиральной ДНК. Этот факт позволил нам сделать выводы о том, что, в то время как большая часть неспецифических взаимодействий осуществляется за счет положительно заряженных аминокислот, специфические взаимодействия осуществляются за счет отрицательных зарядов. Тогда пространственное расположение GC-пар в последовательности нуклеотидов определяет важную часть архитектуры специфических взаимодействий. При этом на белке такую роль «наводчиков» будет играть расположение боковых радикалов аспарагиновой и глутаминовой кислот. В такого рода явлениях, при связывании белков на ДНК, несомненно, важную роль играют расположение и геометрия положительных и отрицательных зарядов.

Работа выполнена при поддержке грантов Российского фонда фундаментальных исследований 07-04-01765а и 08-04-01770а.

Список литературы

1. *Harrington R. E.* DNA curving and bending in protein-DNA recognition // *Molecular Microbiology*. 1992. Vol. 6, № 18. P. 2549–2555.
2. *Brennan R. G., Roderick S. L., Takeda Y., Matthews B. W.* Protein-DNA conformational changes in the crystal structure of a lambda Cro-operator complex // *Proceedings of National Academy of Science USA*. 1990. Vol. 87, № 20. P. 8165–8169.
3. *Jordan S. R., Pabo C. O.* Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions // *Science*. 1988. Vol. 242, № 4880. P. 893–899.
4. *Schultz S. C., Shields G. C., Steitz T. A.* Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees // *Science*. 1991. Vol. 253, № 5023. P. 1001–1007.
5. *Benos P. V., Lapedes A. S., Stormo G. D.* Is there a code for protein-DNA recognition? Probabilistic // *Bioessays*. 2002. Vol. 24, № 5. P. 466–475.
6. *Anashkina A., Kuznetsov E., Esipova N., Tumanyan V.* Comprehensive statistical analysis of residues interaction specificity at protein-protein interfaces // *Proteins*. 2007. Vol. 67, № 4. P. 1060–1077.
7. *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. Т. 53, № 3. С. 402–406.

8. *Медведев Н. Н.* Метод Вороного–Делоне в исследовании структуры некристаллических систем / НИЦ ОИГГМ СО РАН. Новосибирск, 2000.
9. *Kamashev D. E., Ebraldze K. K., Esipova N. G., Mirzabekov A. L.* Electrostatic interaction and differences in specific and nonspecific interactions of the Lac-repressor with DNA // *Biofizika*. 1994. Vol. 39, № 6. P. 1101–1103.
10. *Ahmad S., Sarai A.* Moment-based prediction of DNA-binding proteins // *Journal of Molecular Biology*. 2004. Vol. 341, № 1. P. 65–71.
11. *Pullman B.* Electrostatics of Polymorphic DNA // *Journal of Biomolecular Structure & Dynamics*. 1983. Vol. 1, № 3. P. 773–794.