

Классификация комплексов семейств белков: гомеодомены – ДНК, цинковые пальцы – ДНК. Статистический анализ структур ДНК в интерфейсах комплексов гомеодоменов – ДНК

Грохлина Т.И.¹, Панченко Л.А.², Полозов Р.В.³, Сивожелезов В.С.⁴, Иванов В.В.^{5,6}

¹ИМПБ РАН – филиал ИПМ им. М.В. Келдыша РАН, Пуццино, Россия

²Биологический факультет МГУ им. М.В. Ломоносова, Москва, Россия

³Институт теоретической и экспериментальной биофизики РАН, Пуццино, Россия

⁴Институт биофизики клетки РАН, Пуццино, Россия

⁵Объединенный институт ядерных исследований, Дубна, Россия

⁶Национальный исследовательский ядерный университет “МИФИ”, Москва, Россия

grokhlina@mail.ru

В этой работе мы обсудим подходы и методы решения проблемы классификации комплексов белок-ДНК. Для решения задачи предполагается использовать методы вычислительной геометрии и статистического анализа многомерных данных, а в качестве объекта исследования – два семейства факторов транскрипции и их комплексы с ДНК: гомеодомены и цинковые пальцы. Вычислены значения переменных, характеризующих геометрию нуклеотидов двойной спирали ДНК в ее комплексе с белком: «диккерсоновы параметры», двугранные углы, водородные связи и т. д. (на основе данных о 19 структурах комплексов из PDB). На первом этапе решается задача о воздействии качественного фактора – фрагмента нуклеотидной последовательности ДНК в интерфейсе белок-ДНК, имеющего 19 уровней, на переменные Shear, Stretch, Stagger, Buckle, Propeller, Opening. Результаты статистического анализа позволяют сделать вывод о том, что это влияние на Stretch, Stagger, Propeller и Opening – статистически значимо, а на Shear и Buckle – статистически не значимо. Известно, что угловой параметр, Propeller, коррелирует с гибкостью ДНК и существенно меняется при связывании гомеодомена с ДНК. Ни для одного из параметров не найдено комплекса со статистически значимым максимумом или минимумом, что указывает на консервативность влияния связывания гомеодомена на структуру ДНК.

Ключевые слова: структура интерфейсов, комплексы ДНК-белок, гомеодомен, цинковый палец, классификация, многомерный статистический анализ.

Classification of DNA complexes of homeodomain and zinc-finger protein families: statistical analysis of DNA structures in interfaces of homeodomain-DNA complexes

Grokhlina T.I.¹, Panchenko L.A.², Polozov R.V.³, Sivozhelezov V.S.⁴, Ivanov V.V.^{5,6}

¹IMPB RAS – Branch of KIAM RAS, Pushchino, Russia

²Lomonosov Moscow State University

³Institute of Theoretical and Experimental Biophysics, RAS, Pushchino, Russia

⁴Institute of Biophysics of the Cell, RAS, Pushchino, Russia

⁵Joint Institute for Nuclear Research, Dubna, Russia

⁶National Research Nuclear University “MEPhI”, Moscow, Russia

We will herein discuss approaches and methods for solving the classification problem of protein-DNA complexes. We propose to apply methods of computational geometry, cluster analysis, discriminant analysis, factor analysis, and principal component analysis, with two families of transcription factors and their complexes with DNA, homeodomains and zinc fingers, as research objects. The values of variables characterizing nucleotide geometries of the DNA double helix in DNA complexes with proteins are calculated. These include "Dickerson parameters", dihedral angles, hydrogen bonds, etc., based on data from 19 PDB structures of the complexes. At the first stage, the impact is analyzed of a 19-level qualitative factor, the nucleotide sequence of the DNA fragment in the protein-DNA interface, on the Shear, Stretch, Stagger, Buckle, Propeller, and Opening variables. Statistical analysis allowed us to conclude that such influence on Stretch, Stagger, Propeller and Opening is statistically

significant, while that on the Shear and Buckle it is statistically insignificant. The Propeller angular parameter is known to correlate with the flexibility of DNA and changes significantly upon homeodomain binding with DNA. None of the parameters identified a complex with the statistically significant maximum or minimum, which indicates conservatism of the impact of homeodomain binding on the structure of DNA.

Key words: interface structures, DNA-protein complexes, homeodomain, zinc finger, classification, multivariate statistical analysis.

Введение

Регуляция процессов транскрипции осуществляется белками, факторами транскрипции, посредством их связывания со специфическими участками ДНК [1–3]. (Другой способ регуляции – короткие одноцепочечные РНК[4]).

В этих белках квазинезависимые домены, а именно они «узнают» ДНК, соединены гибкими и относительно короткими аминокислотными цепочками – линкерами. Нельзя, однако, считать, что линкеры только соединяют домены. Для ряда факторов транскрипции, например, ключевым для функции является фосфорилирование их линкеров [5, 6]. В других случаях линкеры влияют на узнавание ДНК доменами [7]. Более того, иногда линкеры участвуют в узнавании наравне с «основными» доменами, и тогда их считают отдельными линкерными доменами [8, 9]

Факторы транскрипции могут иметь в своей структуре от одного до нескольких доменов и распознавать ими специфические участки ДНК. Этот «модульный принцип» (функционально, домены – «модули») организации белков посредством комбинации доменов придает биомолекулярному узнаванию достаточно большое разнообразие типов и надлежащую точность. [10–12]

Эти характерные черты узнавания приводят к постановке задач об упорядочении комплексов белок-ДНК, т.е. к задачам классификации их по физико-химическим и геометрическим параметрам, определяющим структуры комплексов и их интерфейсы (пространственная область специфических контактов атомов и атомных групп ДНК и белка).

Это одна из причин того, что в настоящей работе мы подчеркиваем важность классификации интерфейсов с обязательным учетом их структурных и физико-химических параметров.

Классификация должна нам дать сведения о взаимосвязи структур комплексов, возможность проследить их эволюцию и, что особенно важно, дать основу для нахождения общих правил узнавания. Классификация должна помочь также в установлении новых соотношений между комплексами, которые позволили бы унифицировать всю огромную область структур молекул, определяющих существо процесса транскрипции. Это предполагает обработку большого объема данных как экспериментальных,

так и теоретических.

Для решения задач классификации необходимо найти систему геометрических параметров, список дескрипторов, определяющих структуру комплексов и их интерфейсов, и применить методы многомерного статистического анализа – дисперсионного, кластерного, факторного, дискриминантного.

Такова схема исследования, которой мы предполагаем следовать.

Ранее классификация комплексов белок-ДНК по геометрическим параметрам проводилась исходя из линейных комбинаций векторных произведений $C^{\alpha}-C^{\beta}$ векторов ДНК-узнающих аминокислот и нормалей к плоскостям пар оснований. Эти параметры были введены в работе [13] и названы авторами «стерическими соотношениями». Сами авторы не проводили классификации интерфейсов, а только заметили, что стерические соотношения очень разнятся у комплексов с ДНК белков разных семейств. Как следствие, никаких структурных или физико-химических закономерностей, общих для ДНК-белковых комплексов, ожидать нельзя.

Мы обратили внимание, что, по крайней мере, у одного семейства ДНК-узнающих белков, а именно у гомеодоменов, стерические соотношения очень близки, и стали искать правила узнавания внутри этого семейства. Сначала мы обнаружили эти правила путем визуального анализа репрезентативных наборов комплексов ДНК-гомеодомен [14, 15], а потом подтвердили и обобщили эти правила на всех доступных на тот момент комплексах, создав базу данных соответствующих интерфейсов ANTPC [16].

Стерические соотношения, введенные в работе [13], послужили основой классификации ДНК-белковых интерфейсов в работе [17], но в ней промежуточной стадией классификации было выравнивание аминокислотных последовательностей ДНК-узнающих белков, так что в значительной степени классификация интерфейсов была сведена к классификации аминокислотных последовательностей узнающих спиралей гомеодоменов.

Между тем, именно на примере гомеодоменов выяснилось [18], что такого рода классификация, основанная на их аминокислотных последовательностях, встречает значительные трудности. Оказалось практически невозможным провести строгую эволюционную классификацию гомеодоменов при объединении семейств в классы (внутри семейства гомеодоменов принята иерархия

индивидуальные гомеодомены – семейства – классы – супраклассы). Таким образом, нуклеотидная последовательность гомеодомена недостаточна даже для классификации самих гомеодоменов, не говоря уже об их комплексах с ДНК, при том, что именно связывание с ДНК и диссоциация от нее – основная биологическая функция гомеодоменов.

По-видимому, возникшие трудности говорят о неудачном выборе параметров, определяющих пространственную структуру интерфейсов комплексов. Действительно, нормаль к плоскости пары оснований не принадлежит к числу параметров, определяющих локальную геометрию ДНК, поэтому попытка классификации интерфейсов на этой основе свелась к геометрическим характеристикам узнающей спирали гомеодомена. Ее конформация при связывании с ДНК не меняется [19] и определяется главным образом аминокислотной последовательностью. Это, видимо, и проявилось в фактической неудаче классификации интерфейсов белок-ДНК в работе [12].

В нашей работе мы используем другие, более физически обоснованные, параметры локальной структуры ДНК, в надежде, что они окажутся более удачным выбором параметров и переменных, определяющих пространственную структуру интерфейсов комплексов.

Геометрические характеристики ДНК были получены с использованием пакета 3DNA, предназначенного для анализа, реконструкции и визуализации трехмерных структур нуклеиновых кислот [20]: переменные и параметры, описывающие положения азотистых оснований в комплементарных парах, локальную структуру двойной спирали ДНК (водородные связи, торсионные углы и т.д.).

Материалы и методы

Для анализа мы использовали численные значения диккерсоновых параметров (Shear, Stretch, Stagger, Buckle, Propeller, Opening, см. рис. 1.) для каждого из 19-ти фрагментов нуклеотидной последовательности в интерфейсе белок-ДНК.

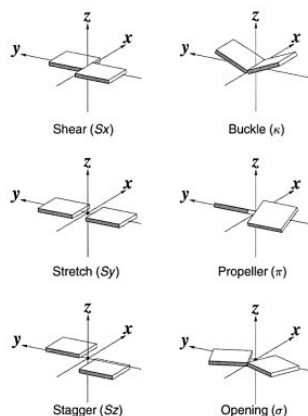


Рис. 1. Диккерсоновы параметры.

На первом этапе мы решаем задачу о воздействии качественного фактора – нуклеотидной последовательности фрагмента ДНК в интерфейсе белок-ДНК, на диккерсоновы параметры. Для решения применялись как однофакторный дисперсионный анализ, ANOVA, так и непараметрические критерии – Крускала–Уоллиса (Kruskal–Wallis test) и медианный критерий Брауна–Муда (Brown–Mood test) [21]. Отметим, что результаты, полученные с помощью однофакторного дисперсионного анализа, подтверждаются и непараметрическими методами.

Значения переменной C , обозначающей на рисунках тип комплекса соответственно его PDB коду, перечислены в таблице 1. Там же представлены длины нуклеотидной цепи, числа пар оснований и наименования ДНК-узнающего белка или белков, участвующих в комплексе. Все эти параметры и определяют тип комплекса, неявно включающий и узнаваемую в нем нуклеотидную последовательность, и последовательность узнающей ее спирали белка (не показаны).

Таблица 1. Исходные данные (кристаллография, ЯМР) комплексов гомеодоменов с ДНК

Код PDB	C^1	Код цепибелка ²	Имя белка по базе данных Uniprot	Длина н.п. ДНК ³	Число п.о. ДНК ³
1	2	3	4	5	6
1AHD	1	P	Homeotic protein antennapedia	14	14
1AKH	2	A	Mating-type protein A1	21	19
1AKH	2	B	Mating-type protein ALPHA2		
1APL	3	C	Mating-type protein ALPHA2	21	19
1APL	3	D	Mating-type protein ALPHA2		
1AU7	4	A	Pituitary-specific positive transcription factor 1	25	24
1AU7	4	B	Pituitary-specific positive transcription factor 1		
1B72	5	A	Homeobox protein Hox-B1	20	19
1B72	5	B	Pre-B-cell leukemia transcription factor 1		
1B8I	6	A	Homeotic protein ultrabithorax	15	13
1B8I	6	B	Homeobox protein extradenticle		
1CQT	7	A	POU domain, class 2, transcription factor 1	15 ⁴	27 ⁴
1CQT	7	B	POU domain, class 2, transcription factor 1	15	27
1CQT	7	I	POU domain class 2-associating factor 1	15	27
1CQT	7	J	POU domain class 2-associating factor 1	15	27
1DU0	8	A	Segmentation polarity homeobox protein engrailed	21	20
1DU0	8	B	Segmentation polarity homeobox protein engrailed		
1E3O	9	C	POU domain, class 2, transcription factor 1	11	11
1FJL	10	A	Segmentation protein paired	14	14
1FJL	10	B	Segmentation protein paired		
1FJL	10	C	Segmentation protein paired		
1GT0	11	C	POU domain, class 2, transcription factor 1	24	23

1GT0	11	D	Transcription factor SOX-2		
1		2	3	4	5
1IG7	12	A	Homeobox protein MSX-1	13	13
1JGG	13	A	Segmentation protein even-skipped	10	10
1JGG	13	B	Segmentation protein even-skipped	10	10
1NK3	14	P	Homeobox protein vnd	16	13
1OCT	15	C	POU domain, class 2, transcription factor 1	15	14
1HDD	16	C	Segmentation polarity homeobox protein engrailed	21	20
1HDD	16	D	Segmentation polarity homeobox protein engrailed		
1HF0	17	A	POU domain, class 2, transcription factor 1	22	22
1HF0	17	B	POU domain, class 2, transcription factor 1		
1IC8	18	A	Hepatocyte nuclear factor 1-alpha	21	20
1IC8	18	B	Hepatocyte nuclear factor 1-alpha		
1K61	19	A	Mating-type protein ALPHA2	21	20
1K61	19	B	Mating-type protein ALPHA2		
1K61	19	C	Mating-type protein ALPHA2		
1K61	19	D	Mating-type protein ALPHA2		
1LE8	20	A	Mating-type protein A1	20	19
1LE8	20	B	Mating-type protein ALPHA2		

- Примечания: 1. Значение переменной *C* используется в наших расчетах как тип комплекса.
2. PDB файл может содержать несколько белковых молекул, которые могут быть как несколькими экземплярами одного фактора транскрипции, так и различными факторами транскрипции. Молекулы идентифицируются одиночными латинскими буквами.
3. Число пар оснований может быть меньше длины нуклеотидной последовательности (н.п.) ДНК, так как не все основания образуют пары. Число пар оснований используется в статистических расчетах как число наблюдений.
4. В этом примере число пар оснований оказалось больше длины н.п. ДНК, так как в кристалле оказалось две одинаковых двойных цепи ДНК.

Результаты и обсуждение

Результаты анализа (рис. 2–5) позволяют сделать вывод о том, что влияние нуклеотидной последовательности на переменные Stretch, Stagger, Propeller, Opening статистически значимо на уровне $\alpha = 0.05$, а на переменные Shear и Buckle – статистически незначимо. На всех рисунках приведены значения выборочных средних и стандартные ошибки для рассматриваемых переменных.

Чтобы выяснить, для какого типа комплекса выборочное значение каждой из рассматриваемых переменных является максимальным (или минимальным), использовались параметрический, Тьюки–Крамера, и непараметрический, Крускала–Уоллиса, методы множественного сравнения [22]. Проведенный анализ показал, что ни для одной из переменных нельзя выделить комплекс с

максимальным (или минимальным) значением переменной.

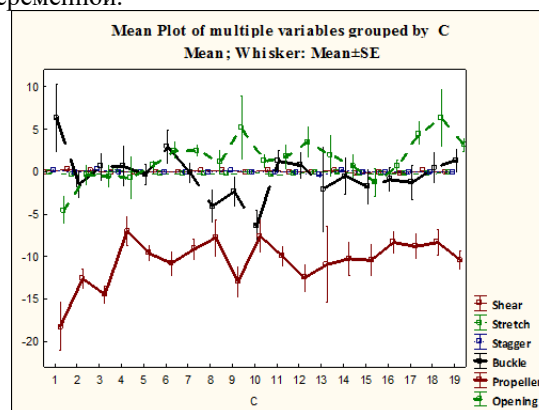


Рис. 2. Выборочные средние и стандартные ошибки переменных Shear, Stretch, Stagger, Buckle, Propeller, Opening.

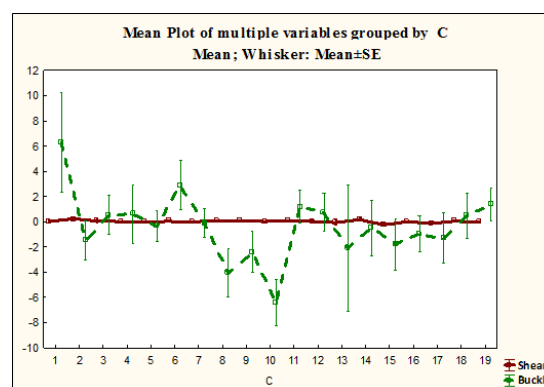


Рис. 3. Выборочные средние и стандартные ошибки переменных Buckle, Opening.

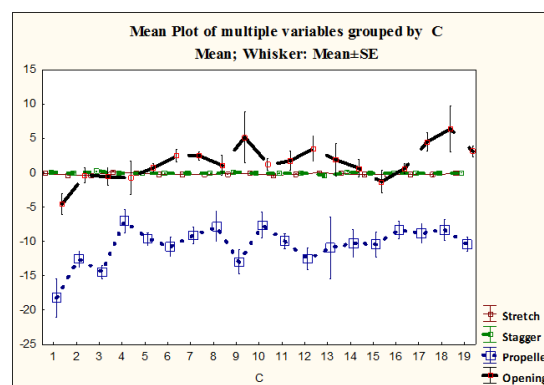


Рис. 4. Выборочные средние и стандартные ошибки переменных Stretch, Stagger, Propeller, Opening.

Отсутствие выраженных максимальных и минимальных значений каждого из параметров у комплексов говорит о консервативности структурных изменений, происходящих в ДНК при ее специфическом связывании с гомеодоменом. Этого следовало ожидать, учитывая тот факт, что гомеодомены принадлежат к числу наиболее консервативных из ДНК-связывающих белков [13–16]. Таким образом, мы отработаем нашу вычислительную методику на консервативных гомеодоменах. В дальнейшем будем применять ее

для других классов факторов транскрипции, прежде всего цинковых пальцев [23].

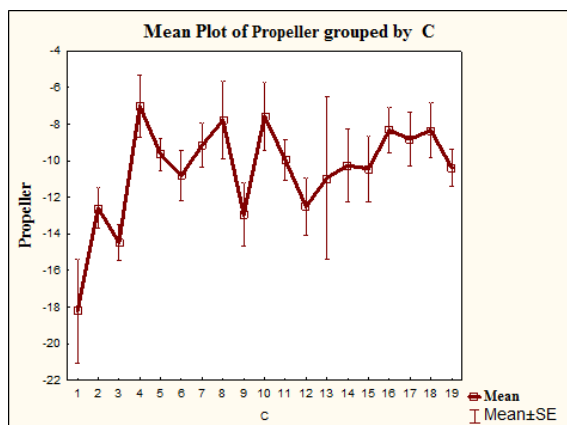


Рис. 5. Выборочные средние и стандартные ошибки переменной Propeller.

Наибольшую биологическую значимость представляет последний рисунок, поскольку переменная Propeller, как известно, коррелирует с гибкостью нуклеотидной цепи [24].

В экспериментальном исследовании, где методом молекулярной биологии SELEX синтезируются нуклеотиды, связывающие заданный белок, показано, что именно гомеодомены, в отличие от цинковых пальцев, уменьшают численные значения Propeller до существенно отрицательных [25]. Исследование комплексов цинковых пальцев с ДНК поэтому представляется естественным продолжением настоящей работы.

Заключение

Настоящая работа представляет собой первую попытку статистического анализа структур ДНК-белковых комплексов на основе именно тех геометрических параметров ДНК, которые отражают особенности ее локальной структуры. Полученные для гомеодоменов результаты в целом указывают на эволюционную консервативность воздействия гомеодоменов на структуру узнаваемой ими ДНК, что задает своего рода реперную точку для исследования белок-нуклеинового узнавания в целом, так как сами гомеодомены – наиболее консервативное из семейств ДНК-узнающих белков. Вторая, в некотором смысле противоположная, ожидаемая реперная точка – семейство цинковых пальцев из числа наиболее эволюционно изменчивых.

Другое (методическое) направление дальнейших исследований – включение в расчеты и статистический анализ геометрических характеристик, описывающих взаимодействия между соседними парами оснований ДНК в дополнение к рассмотренным здесь параметрам индивидуальных пар оснований.

Результаты настоящего и последующих исследований должны помочь нам выбрать как меру близости конформационных изменений ДНК при ее

комплексообразовании с белками, так и наиболее адекватные методы классификации ДНК-белковых комплексов в целом, прежде всего комплексов ДНК с белками-факторами транскрипции.

Благодарности

Работа поддержана грантом РФФИ 17-07-01331а.

Список литературы

1. Bondos S.E., Tan X.X. Combinatorial transcriptional regulation: the interaction of transcription factors and cell signaling molecules with homeodomain proteins in *Drosophila* development. *Crit. Rev. Eukaryot. Gene. Expr.* 2001. V. 11. P. 145–171.
2. Nemer G., Nemer M. Regulation of heart development and function through combinatorial interactions of transcription factors. *Ann. Med.* 2001. V. 33. P. 604–610. doi: [10.3109/07853890109002106](https://doi.org/10.3109/07853890109002106).
3. Ohsawa R., Kageyama R. Regulation of retinal cell fate specification by multiple transcription factors. *Brain. Res.* 2008. V. 1192. P. 90–98. doi: [10.1016/j.brainres.2007.04.014](https://doi.org/10.1016/j.brainres.2007.04.014).
4. Kashyap V., Rezende N.C., Scotland K.B., Shaffer S.M., Persson J.L., Gudas L.J., Mongan N.P. Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs. *Stem. Cells Dev.* 2009. V. 18. P. 1093–1108. doi: [10.1089/scd.2009.0113](https://doi.org/10.1089/scd.2009.0113).
5. Kamato D., Burch M.L., Piva T.J., Rezaei H.B., Rostam M.A., Xu S., Zheng W., Little P.J., Osman N. Transforming growth factor- β signalling: role and consequences of Smad linker region phosphorylation. *Cellular signalling.* 2013. V. 25. P. 2017–2024. doi: [10.1016/j.cellsig.2013.06.001](https://doi.org/10.1016/j.cellsig.2013.06.001).
6. Hough C., Radu M., Doré J.J. Tgf-beta induced Erk phosphorylation of smad linker region regulates smad signaling. *PLoS One.* 2012. V. 7. P. e42513. doi: [10.1371/journal.pone.0042513](https://doi.org/10.1371/journal.pone.0042513).
7. Xu H.E., Rould M.A., Xu W., Epstein J.A., Maas R.L., Pabo C.O. Crystal structure of the human Pax6 paired domain–DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes & development.* 1999. V. 13. P. 1263–1275.
8. Hüntelmann B., Staab J., Herrmann-Lingen C., Meyer T. A conserved motif in the linker domain of STAT1 transcription factor is required for both recognition and release from high-affinity DNA-binding sites. *PLoS One.* 2014. V. 9. P. e97633. doi: [10.1371/journal.pone.0097633](https://doi.org/10.1371/journal.pone.0097633).
9. Wiesler S.C., Weinzierl R.O. The linker domain of basal transcription factor TFIIB controls distinct

- recruitment and transcription stimulation functions. *Nucleic Acids Res.* 2011. V. 39. P. 464–474. doi: [10.1093/nar/gkq809](https://doi.org/10.1093/nar/gkq809).
10. Vermeirssen V., Barrasa M.I., Hidalgo C.A., Babon J.A., Sequerra R., Doucette-Stamm L., Barabási A.L., Walhout A.J. Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network. *Genome research.* 2007. V. 17. P. 1061–1071. doi: [10.1101/gr.6148107](https://doi.org/10.1101/gr.6148107).
 11. Ward J.J., Thornton J.M. Evolutionary models for formation of network motifs and modularity in the *Saccharomyces* transcription factor network. *PLoS computational biology.* 2007. V. 3. P. e198. doi: [10.1371/journal.pcbi.0030198](https://doi.org/10.1371/journal.pcbi.0030198).
 12. Siggers T., Reddy J., Barron B., Bulyk M.L. Diversification of transcription factor paralogs via noncanonical modularity in C2H2 zinc finger DNA binding. *Molecular cell.* 2014. 55. P. 640–648. doi: [10.1016/j.molcel.2014.06.019](https://doi.org/10.1016/j.molcel.2014.06.019).
 13. Pabo C.O., Nekludova L. Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *Journal of molecular biology.* 2000. V. 301. P. 597–624. doi: [10.1006/jmbi.2000.3918](https://doi.org/10.1006/jmbi.2000.3918).
 14. Chirgadze Y.N., Zheltukhin E.I., Polozov R.V., Sivozhelezov V.S., Ivanov V.V. Binding regularities in complexes of transcription factors with operator DNA: homeodomain family. *Journal of Biomolecular Structure and Dynamics.* 2009. V. 26. P. 687–700. doi: [10.1080/07391102.2009.10507282](https://doi.org/10.1080/07391102.2009.10507282).
 15. Chirgadze Y.N., Sivozhelezov V.S., Polozov R.V., Stepanenko V.A., Ivanov V.V. Recognition rules for binding of homeodomains to operator DNA. *Journal of Biomolecular Structure and Dynamics.* 2012. V. 29. P. 715–731. doi: [10.1080/073911012010525019](https://doi.org/10.1080/073911012010525019).
 16. Грохлина Т.И., Зрелов П.В., Иванов В.В., Полозов Р.В., Чиргадзе Ю.Н., Сивожелезов В.С. База данных аминокислотно-нуклеотидных контактов в комплексах ДНК с белками семейства гомеодоменов. *Письма в журнал «Физика элементарных частиц и атомного ядра».* 2013. Т. 15. С. 755–762.
 17. Siggers T.W., Silkov A., Honig B. Structural alignment of protein-DNA interfaces: insights into the determinants of binding specificity. *Journal of molecular biology.* 2005. V. 345. P. 1027–1045. doi: [10.1016/j.jmb.2004.11.010](https://doi.org/10.1016/j.jmb.2004.11.010).
 18. Holland P.W., Booth H.A.F., Bruford E.A. Classification and nomenclature of all human homeobox genes. *BMC biology.* 2007. V. 5. P. 47.
 19. Miyazono K., Zhi Y., Takamura Y., Nagata K., Saigo K., Kojima T., Tanokura M. Cooperative DNA-binding and sequence-recognition mechanism of aristaless and clawless. *EMBO J.* 2010. V. 29. P. 1613–1623. doi: [10.1038/emboj.2010.53](https://doi.org/10.1038/emboj.2010.53).
 20. Lu X.-J., Olson W.K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 2003. V. 31. P. 5108–5121. doi: [10.1093/nar/gkg680](https://doi.org/10.1093/nar/gkg680).
 21. Мятлев В.Д., Панченко Л.А., Ризниченко Г.Ю., Терехин А.Т. *Теория вероятностей и математическая статистика. Математические модели.* М.: Юрайт, 2017. 321 с. ISBN 978-5-534-01698-7.
 22. Zar J.H. *Biostatistical Analysis: 4thed.* N.J.: Prentice-Hall, Inc. Upper Saddle River, 1999. 998 p.
 23. Chirgadze Y.N., Boshkova E.A., Polozov R.V., Sivozhelezov V.S., Dzyabchenko A.V., Kuzminsky M.B., Stepanenko V.A., Ivanov V.V. The electrostatic role of the Zn-Cys2His2 complex in binding of operator DNA with transcription factors: mouse EGR-1 from the Cys2His2 family. *J. Biomol. Struct. Dyn.* 2018. V. 7. P. 1–14. doi: [10.1080/07391102.2017.1404937](https://doi.org/10.1080/07391102.2017.1404937).
 24. Hancock S.P., Gane T., Cascio D., Rohs R., Di Felice R., Johnson R.C. Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. *Nucleic Acids Res.* 2013. V. 41. P. 6750–6760. doi: [10.1093/nar/gkt357](https://doi.org/10.1093/nar/gkt357).
 25. Dror I., Golan T., Levy C., Rohs R., Mandel-Gutfreund Y. A widespread role of the motif environment in transcription factor binding across diverse protein families. *Genome research.* 2015. V. 25. P. 1268–1280. doi: [10.1101/gr.184671.114](https://doi.org/10.1101/gr.184671.114).