

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ПРОСТРАНСТВЕННЫХ СТРУКТУР ДНК В ИНТЕРФЕЙСАХ КОМПЛЕКСОВ ГОМЕОДОМЕН-ДНК: «ДИККЕРСОНОВЫ ПАРАМЕТРЫ»

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

¹ Институт теоретической и экспериментальной биофизики РАН
г. Пущино, 142290, РФ; e-mail: polrob@mail.ru

² Институт математических проблем биологии РАН – филиал ИПМ им. М.В. Келдыша РАН
г. Пущино, 142290, РФ

³ Московский государственный университет им. М.В. Ломоносова
г. Москва, РФ

⁴ Объединенный институт ядерных исследований
г. Дубна, РФ

⁵ Национальный исследовательский ядерный университет «МИФИ»
г. Москва, РФ

Поступила в редакцию: 11.07.2020

Аннотация. Специфические комплексы гомеодоменов с ДНК образуются при связывании белка с двуспиральной ДНК в В-форме (по широкому желобу). Классификация комплексов белок-ДНК и их интерфейсов по физико-химическим, геометрическим параметрам связана с решением ряда задач вычислительного и статистического характера. Классификация должна дать сведения о взаимосвязи комплексов друг с другом, дать возможность проследить их эволюцию и дать основу для формулировок правил узнавания и построения моделей узнавания. Для решения задач классификации необходимо составить список надлежащих дескрипторов комплексов и их интерфейсов, провести многомерный статистический анализ данных из этого списка. Мы используем геометрические параметры, 3DNA: a software package, пространственных структур ДНК в интерфейсах 75 комплексов гомеодомен-ДНК (PDB). В данной работе представлены результаты многомерного статистического анализа «диккерсоновых параметров» пар азотистых оснований ДНК в интерфейсах комплексов гомеодомен-ДНК.

Ключевые слова: гомеодомен-ДНК, «диккерсоновы параметры», интерфейс, деформация ДНК, классификация.

Процессы ассоциации и диссоциации белков – факторов транскрипции с ДНК - являются основой регуляции транскрипции. Структурные типы факторов транскрипции и их комплексов с ДНК весьма разнообразны и образуют несколько семейств. Одно из них – это семейство гомеодоменов и их комплексов с ДНК. Гомеодомены регулируют дифференцировку клеток. Это обширное семейство организовано весьма просто – относительно небольшой участок ДНК (до 20 пар азотистых оснований) формирует достаточно плотную упаковку с атомными группами α -спирали гомеодомена («узнающая спираль»). И эта же упаковка дает основу для формирования нужной, вполне определенной ориентации атомов ДНК и белка, и образование специфических атомных контактов в том числе через молекулу воды. В этой области тесного соприкосновения белка и ДНК («интерфейс») молекула ДНК деформируется: искажается как «внутренняя геометрия» нуклеотидов, так и форма двойной спирали ДНК (геликоид). Степень деформации зависит от нуклеотидной последовательности ДНК в области интерфейса и уникальна для каждого из комплексов гомеодомен-ДНК. Совокупность всех этих факторов и параметров узнавания обеспечивает надежность, гибкость, точность и характерные времена узнавания. Эти специфические контакты могут давать независимый вклад в аффинность и специфичность. Но это не всегда имеет место, поэтому в некоторых интерфейсах необходимо принимать во внимание зависимость положения контактов в интерфейсах относительно соседних атомных групп белка и ДНК. Мы представили лишь схему образования интерфейсов комплексов гомеодомен-ДНК. Более конкретное и подробное изложение проблемы узнавания белками ДНК (не только факторами транскрипции) приведено в обзоре [1], в котором дан большой список цитированных работ и приведены соответствующие доводы и доказательства.

Анализ пространственной структуры интерфейсов комплексов гомеодомен-ДНК основан на геометрических характеристиках ДНК, полученных с использованием пакета 3DNA, предназначенного для анализа, реконструкции и визуализации трехмерных структур нуклеиновых кислот [2]. В данной работе мы анализируем следующие группы:

1. Параметры, описывающие отклонения положения оснований в паре от идеального их положения- Shear, Stretch, Stagger, Buckle, Propeller, Opening («диккерсоновы параметры»). Здесь Shear и Stretch определяют смещения оснований в плоскости пары, Opening – угол между основаниями в плоскости пары. Три оставшихся параметра характеризуют выход оснований из плоскости. «Идеальным» здесь является положение оснований в абсолютно плоской паре, где все шесть параметров равны [3].

2. Параметры, описывающие пространственное расположение соседних пар оснований, т.е. положение и ориентацию одной пары относительно другой: сдвиги Shift, Slide и Rise по осям Ox, Oy, Oz соответственно и три поворота Tilt, Roll и Twist (параметры твердого тела).

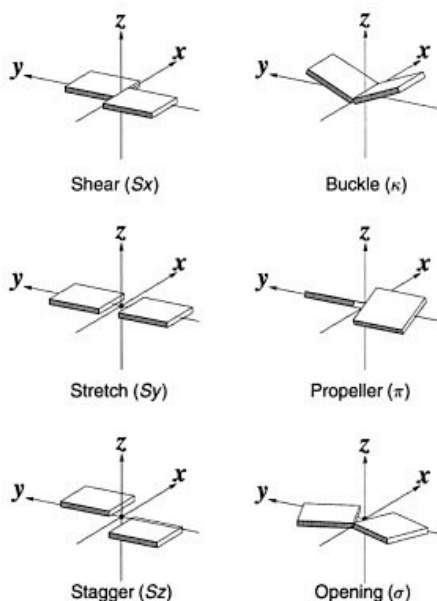


Рисунок 1. «Диккерсоновы параметры»

3. Параметры, описывающие регулярность спирали (спиральные параметры), - смещения по осям x , y и z - displacement, y -displacement, helical rise соответственно, и повороты вокруг осей спирали - inclination (η), tip (θ) и helical twist (Ωh). Две последние группы описывают взаимное положение соседних пар оснований в разных системах координат - в системе отсчета, связанной с димером, и локальной системе спирали. Мы не приводим здесь подробное толкование этих параметров и систем отсчета, т.к. это относится к алгоритму программы и приведено авторами в [2].

4. Кроме того, еще одна группа параметров – это параметры Shift, Slide и Rise по осям Ox , Oy , Oz , соответственно, и три поворота Tilt, Roll и Twist, аналогичные группе два, с тем отличием, что они характеризуют положение соседних оснований в цепочке.

Для анализа мы использовали численные значения «диккерсоновых параметров» пар азотистых оснований (Shear, Stretch, Stagger, Buckle, Propeller, Opening, см. рис. 1) для каждого участка нуклеотидной последовательности ДНК в 75 интерфейсах комплексов гомеодомен-ДНК.

Для выявления влияния качественного фактора – нуклеотидной последовательности фрагмента ДНК в интерфейсе гомеодомен-ДНК, на «диккерсоновы параметры» мы использовали однофакторный дисперсионный анализ, ANOVA, и непараметрические критерии - Крускала-Уоллиса (Kruskal-Wallis test) и медианный критерий Брауна-Муда (Brown-Mood test). Результаты анализа позволяют сделать вывод о том, что влияние нуклеотидной последовательности на переменные Stretch, Stagger, Buckle, Propeller, Opening статистически значимо на уровне $\alpha=0,05$, а на переменную Shear - статистически не значимо. Отметим, что результаты, полученные с помощью

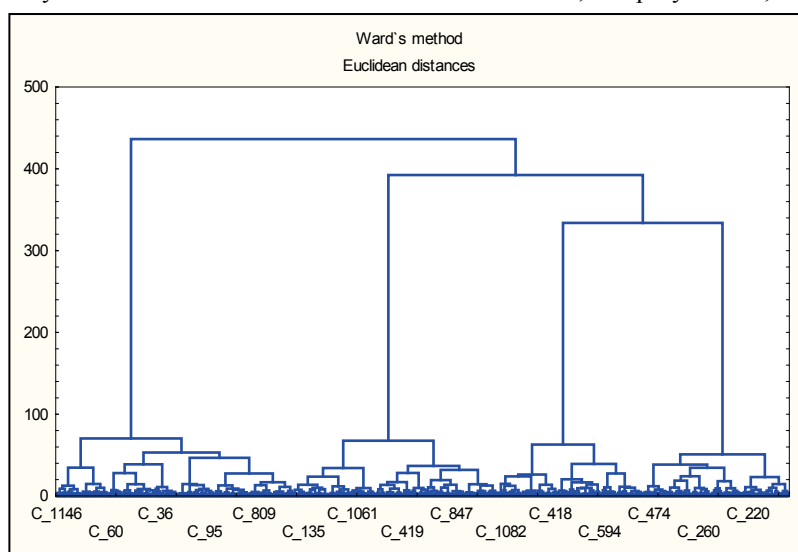


Рисунок 2. Дендрограмма

Classification Matrix					
Rows: Observed classifications					
Columns: Predicted classifications					
Class	Percent Correct	1 p=,3270	2 p=,2753	3 p=,1988	4 p=,1988
1	98,99	394,00	2,00	1,00	1,00
2	100,00	0,00	335,00	0,00	0,00
3	99,59	1,00	0,00	241,00	0,00
4	100,00	0,00	0,00	0,00	242,00
Total	99,59	395,00	337,00	242,00	243,00

Рисунок 3. Классификационная матрица

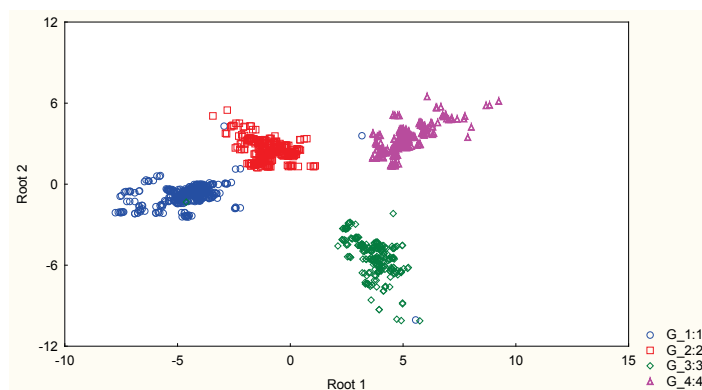


Рисунок 4. Проекция результатов наблюдений на плоскость первых двух канонических переменных

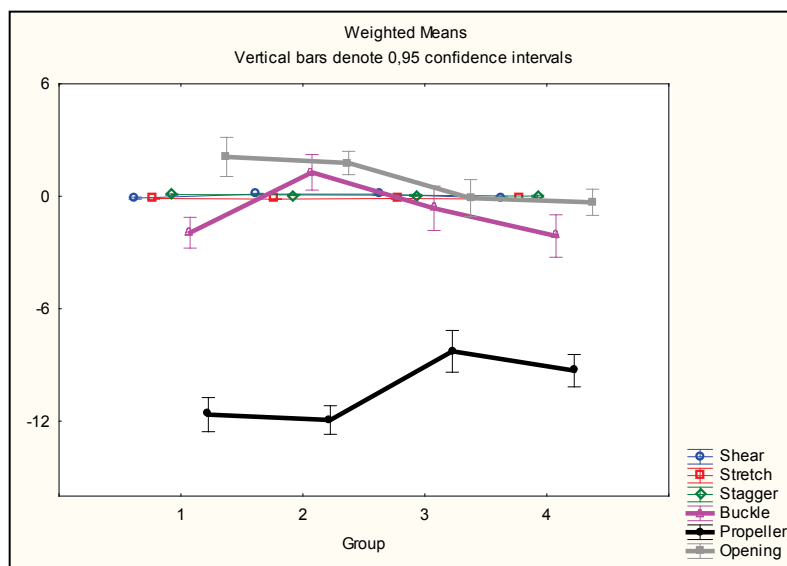


Рисунок 5. Средние значения «диккерсоновых параметров» и 95% доверительные интервалы

однофакторного дисперсионного анализа, подтверждаются и непараметрическими методами. Далее мы использовали методы классификации многомерных данных: кластерный и дискриминантный анализы. Кластерный анализ позволил разбить исходную выборку на четыре однородные группы, класса (рис. 2).

Цель дискриминантного анализа состоит в том, чтобы на основе измерения различных параметров объекта классифицировать его, т. е. отнести к одной из нескольких полученных групп. И определить значения корректных вероятностей классификации. На рисунке 3 приведены результаты дискриминантного анализа, где показано, что вероятность корректной классификации равна ~99%. В дискриминантном анализе используется понятие канонических переменных, которые несут максимальную информацию о взаимном расположении классов относительно друг друга. Поэтому они полезны для геометрического представления результатов дискриминантного анализа. На рисунке 4 приведено графическое представление, т.е. проекция результатов наблюдений на плоскость первых двух канонических переменных. А на рисунке 5 приведены значения средних и 95% доверительные интервалы каждого из «диккерсоновых параметров» от новой переменной, номера класса, Group.

Список литературы / References:

1. Siggers T., Gordan R. Protein–DNA binding: complexities and multi-protein codes. *Nucleic Acids Research*, 2014, vol. 42, no. 4, pp. 2099-2111. DOI: 10.1093/nar/gkt1112.
2. 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, vol. 31, pp. 5108-5121. DOI: 10.1093/nar/gkg680.
3. Xiang-Jun Lu, Wilma K. Olson Characterization of base pair geometry. *Computational Crystallography Newsletter*, 2016, vol. 7, pp. 6-9.
4. Zar J.H. Biostatistical Analysis. *Upper Saddle River, N.J.:* Prentice-Hall, 1999.

MULTIVARIATE STATISTICAL ANALYSIS OF SPATIAL DNA STRUCTURES AT THE INTERFACES OF HOMEODOMEN-DNA COMPLEXES: «DICKERSON PARAMETERS»**¹Polozov R.V., ²Grokhlina T.I., ³Panchenko L.A., ⁴Ivanov V.V.**¹ Institute of Theoretical and Experimental Biophysic, Russian Academy of Sciences*Pushchino, 142290, Russia; e-mail: polrob@mail.ru*² Institute of Mathematical Problems of Biology, Keldysh Institute of Applied Mathematics, Russian Academy of Sciences*Pushchino, 142290, Russia*³ Lomonosov Moscow State University*Moscow, Russia*⁴ Joint Institute for Nuclear Research*Dubna, Russia,*⁵ National Research Nuclear University “MEPhI”*Moscow, Russia*

Abstract. Specific complexes of homeodomains with DNA are formed when the protein binds to double-stranded DNA in the B-form (along the wide groove). The classification of protein-DNA complexes and their interfaces according to physicochemical and geometric parameters is associated with solving a number of computational and statistical problems. Classification should provide information on the relationship of complexes with each other, provide an opportunity to trace their evolution and provide a basis for the formulation of recognition rules and the construction of recognition models. To solve the classification problems, it is necessary to compile a list of appropriate descriptors of the complexes and their interfaces, to conduct a multivariate statistical analysis of the data from this list. We use geometric parameters, 3DNA: a software package, of spatial DNA structures at the interfaces of 75 homeodomain-DNA complexes (PDB). This paper presents the results of a multivariate statistical analysis of “Dickerson parameters” of nitrogenous base pairs of DNA at the interfaces of homeodomain-DNA complexes.

Key words: *homeodomain-DNA, “Dickerson parameters”, interface, DNA deformation, classification.*