

МОЛЕКУЛЯРНАЯ БИОЛОГИЯ

УДК 577.323

МОДЕЛИРОВАНИЕ СТРУКТУРЫ ДНК-БЕЛКОВЫХ КОМПЛЕКСОВ
С ИСПОЛЬЗОВАНИЕМ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ
ПО РЕЗОНАНСНОМУ ПЕРЕНОСУ ЭНЕРГИИ И ПЕРЕКИСНОМУ ОКИСЛЕНИЮ

Г.А. Армеев, Т.К. Горковец, Д.А. Ефимова, К.В. Шайтан, А.К. Шайтан*

*Кафедра биоинженерии, биологический факультет,
Московский государственный университет имени М.В. Ломоносова;
Россия, 119234, г. Москва, Ленинские горы, д. 1, стр. 12*

** e-mail: alex@molsim.org*

Обсуждаются методы построения трехмерных моделей ДНК в комплексе с белками с помощью компьютерного моделирования и не прямых методов изучения конформации макромолекул. Рассматриваются способы интерпретации экспериментальных данных, полученных непрямыми методами изучения трехмерной структуры биомолекул. Обсуждаются аспекты интеграции такого рода данных в процесс построения молекулярных моделей ДНК-белковых комплексов на основании геометрических характеристик ДНК. Предлагается алгоритм поиска конформации таких комплексов на основе данных экспериментов по ферстеровскому резонансному переносу энергии (FRET) и перекисному окислению ДНК (анализируемому с помощью метода гидроксильного футпринтинга), а также информации о локальной гибкости ДНК. Алгоритм апробирован на примере построения гипотетической модели ДНК в нуклеосоме, связанной с гистоном H1.

Ключевые слова: молекулярное моделирование, ДНК, FRET, футпринтинг, нуклеосома, хроматин.

Структура ДНК-белковых комплексов влияет на течение множества процессов, таких как транскрипция, репликация и репарация. В исследованиях структуры додекамеров ДНК было показано, что реальная конфигурация этой макромолекулы значительно отличается от структуры идеальной двунитевой спирали. Необходимость существования более высоких уровней организации ДНК очевидна, так как геном эукариотического организма не может поместиться в относительно компактном ядре в полностью развернутом виде.

В 1974 г. Р. Корнбергом [1] были открыты структурные единицы компактизации хроматина, которые позже были названы нуклеосомами. Пространственная структура нуклеосом долгое время оставалась неизвестной, и лишь в 1997 г. методом рентгеноструктурного анализа (РСА) была впервые определена структура нуклеосомы с почти атомарным разрешением [2]. Нуклеосома представляет собой октамер белков гистонов, который несет на себе 145–147 нуклеотидных пар. ДНК закручена вокруг октамера, образуя 1,65 витка левозакрученной суперспирали. Белковое ядро нуклеосомы образует цилиндр диаметром 65 Å и высотой 60 Å.

На уровне единичных нуклеосом осуществляется тонкая регуляция работы генетического аппарата клетки. Одним из ярких примеров регуляции экспрессии генов является так называемый нуклеосомный барьер [3]: вероятность прохождения РНК-

полимеразы II сквозь нуклеосому зависит от типа и структуры нуклеосомы. Помимо этого, нуклеосомы взаимодействуют со множеством разнообразных транскрипционных факторов и структурных белков. Существует большое количество гистоновых вариантов, последовательность и структура которых отличается от канонических, что определяет их специфические свойства [4]. Например, в областях центромер хромосом ДНК находится в специфических нуклеосомах, содержащих модифицированный гистон CenH3 (CENP-A).

Нуклеосомы являются ключевыми и фундаментальными единицами упаковки ДНК. Их взаимодействие, а также пространственная конфигурация хроматина определяют процессы реализации и передачи генетической информации. Значительным прорывом, повысившим интерес к изучению нуклеосом, стало получение данных об их атомарном устройстве. На данный момент в банке данных PDB (Protein Data Bank) насчитывается более 100 структур различных нуклеосом. Однако изучение более высоких уровней организации хроматина обычными методами структурной биологии затруднено. На данный момент одной из наиболее крупных структур, полученных с помощью метода РСА, является комплекс из четырех нуклеосом [5]. Хроматиновые фибриллы трудно изучать при помощи метода РСА, так как структуры такого размера крайне сложно кристаллизовать, к тому же места рас-

положения нуклеосом не фиксированы строго [6], а, значит, хромосомы в кристалле будут сильно отличаться друг от друга. Информацию о структуре можно получать косвенно биофизическими и биохимическими методами, например, из данных экспериментов по измерению эффективности ферстеровского резонансного переноса энергии (FRET) и с помощью определения профилей перекисного окисления ДНК (гидроксильного футпринтинга).

Метод FRET основан на явлении безизлучательного переноса энергии с одного красителя на другой. Красители подбирают таким образом, чтобы донор флуоресцировал в области спектра поглощения акцептора. Введение меток возможно путем замены аминокислот в белках либо азотистых оснований в ДНК на их аналоги, несущие на себе флуорофорные группировки. Вероятность переноса энергии значительно зависит от расстояния между флуоресцентными метками; таким образом, измеряя эффективность FRET, можно оценить дистанцию между мечеными участками макромолекулы. В работе [7] таким методом были показаны тонкие различия (5 Å) по данному параметру между канонической и центромерной нуклеосомой. Измерения FRET можно производить в растворе, при этом усредняется сигнал от всех находящихся в нем молекул, но измерения такого рода не позволяют судить о конформационных переходах внутри этих молекул. Помимо этого, измерения можно производить в режиме единичных молекул, собирая флуоресценцию лишь с малого конфокального объема в растворе. Такие измерения позволяют различать конформационные переходы в структуре молекул, определять заселенность разных конформаций (долю молекул, находящихся в определенной конфигурации), а также оценивать времена таких переходов. Таким образом, FRET позволяет получить набор расстояний между парами оснований в меченой макромолекуле. Комбинируя эту информацию с механическими свойствами ДНК, а также с данными гидроксильного футпринтинга, можно судить о трехмерной структуре изучаемых комплексов.

Футпринтинг является полезным методом изучения структуры ДНК в различных комплексах, будь то нуклеосома или же комплекс ДНК с транскрипционными факторами. Метод основан на определении доступности нуклеиновых кислот для атаки гидроксильными радикалами и позволяет изучать ДНК с любой последовательностью с разрешением до одного нуклеотида. Гидроксильные радикалы обладают высокой реактивностью и расщепляют сахарофосфатный остов ДНК [8,9]. Основная реакция заключается в отщеплении водородов от 3'-, 4'- и 5'-углеродов дезоксирибозы с образованием промежуточного продукта, являющегося радикалом, и последующим разрывом сахарофосфатной

цепи ДНК [9]. Данные футпринтинга можно применять для определения мест ДНК, ассоциированных с белками, что дает дополнительную информацию о пространственных ограничениях внутри макромолекул при построении молекулярных моделей.

Моделирование ДНК и нуклеосом можно производить в полноатомном приближении [10] и рассчитывать энергию конформации ДНК при помощи традиционных силовых полей (AMBER, CHARMM и т.д.). Однако поиск оптимальной конфигурации для таких моделей требует значительных вычислительных ресурсов. Для повышения скорости расчетов применяют огрубленные модели. В одной из таких моделей каждая нуклеотидная пара описывается шестью переменными (три трансляционных координаты и три вращательных), как и каждый шаг между нуклеотидными парами [11]. Таким образом, для цепи из N нуклеотидных пар конфигурация ДНК описывается набором из $(12N - 6)$ обобщенных переменных. Такой подход, в отличие от других огрубленных моделей, позволяет однозначно воссоздавать атомарную структуру молекулярного комплекса. Дополнительно в данную модель можно ввести пространственные ограничения, полученные из экспериментов по измерению эффективности FRET и определению профилей гидроксильного футпринтинга, что позволяет восстановить конфигурацию исследуемого объекта в атомарном разрешении.

Материалы и методы

Интерпретация данных FRET одиночных молекул

Расчет расстояний по данным FRET производился в соответствии с формулой:

$$E_{FRET} = \frac{1}{1 + (R/R_0)^6},$$

где R — расстояние между флуоресцентными метками, а R_0 — ферстеровский радиус пары меток (56 Å для пары цианинового 3 и 5). Расстояние, рассчитанное для пары красителей, не соответствует расстоянию между нуклеотидными парами, так как метки закреплены на ДНК при помощи линкерного участка из 10–15 атомов углерода. Для учета смещения были созданы молекулярные модели меток. При помощи метода молекулярной динамики был построен набор их конформеров и определены координаты наиболее вероятного местонахождения меток. Моделирование производилось в программе Gromacs [12] без учета электрических взаимодействий при температуре 500 К. Полученные координаты центров меток в дальнейшем использовались для расчета расстояния между нуклеотидными парами при оптимизации геометрии ДНК.

Обработка данных гидроксильного футпринтинга

Профили футпринтинга аппроксимировали суммой функций Лоренца по методу наименьших квадратов [13]. Пики нумеровали, вероятность разрыва цепи для каждого нуклеотида определяли как площадь под функцией для соответствующего пика.

Расчет энергии деформации ДНК

Энергия деформации рассчитывалась из обобщенных переменных, описывающих конфигурацию ДНК в соответствии с формулой:

$$E_{\text{deform}} = \frac{1}{2} \sum_{i=1}^6 \sum_{j=1}^6 f_{ij} \Delta\theta_i \Delta\theta_j$$

$$\Delta\theta_i = (\theta_i - \theta_i^0),$$

где θ_i — значения обобщенных переменных, а f_{ij} — коэффициенты жесткости ДНК для отклонения обобщенных переменных для каждой нуклеотидной пары [14]. Расчет обобщенных переменных из атомистической структуры и восстановление координат из набора переменных производился при помощи программы 3DNA [15]. Минимизация энергии производилась по методу сопряженных градиентов.

Все расчеты производились на персональном компьютере в программной среде Python с модулем SciPy.

Результаты и обсуждение

Энергию изгиба ДНК можно рассчитывать при помощи силовых полей, которые применяют для моделирования методом молекулярной динамики, однако ресурсоемкость таких расчетов и проблемы с явным учетом электростатических взаимодействий ограничивают применение этого подхода. Другим способом расчета энергии деформации является применение набора эмпирических силовых коэффициентов, которые были получены путем анализа геометрии ДНК из множества различных кристаллических структур [14]. Расчет конформации в обобщенных координатах можно производить быстрее, потому что число переменных в модели значительно меньше.

Так как минимизация происходит в пространстве обобщенных координат конформации ДНК, а расчет расстояний в трехмерном пространстве, то на каждом шаге минимизации требуется производить расчет атомарной модели ДНК. Общий алгоритм поиска конфигурации ДНК, отвечающий

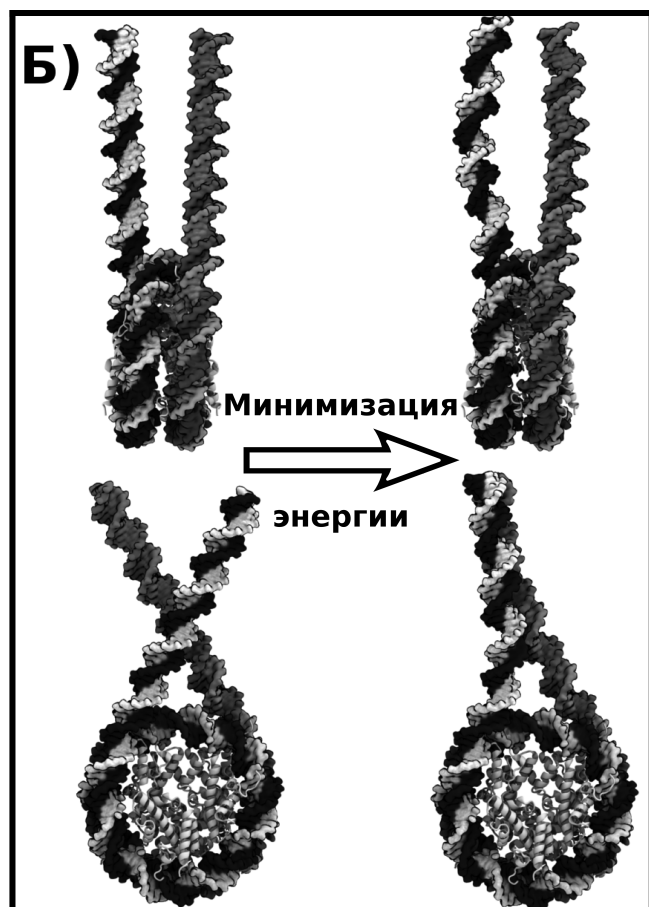
критериям жесткости нуклеиновой кислоты, расстояниям, полученным из FRET, а также ограничениям, полученным из анализа профилей футпринтинга, показан на рис. А.

К недостаткам такого подхода можно отнести отсутствие учета взаимодействий между участками ДНК, что может привести к возникновению стерических перекрытий внутри молекулы. Однако при корректном выборе участков ДНК для минимизации появление таких структур маловероятно.

В качестве объекта для испытания метода была использована нуклеосома с гистонами H1. Этот комплекс на данный момент отсутствует в банке трехмерных структур PDB, но есть работы, где его структура исследуется при помощи ядерного магнитного резонанса [16]. Гистон H1, также называемый линкерным, образует комплекс с нуклеосомами, способствуя более высокому уровню компактизации ДНК. Оказывая такое влияние на размещение/упорядочение ДНК в клетке, гистон H1 играет большую роль в регуляции экспрессии генов. Структурные особенности взаимодействия линкерного гистона и нуклеосомы до сих пор неизвестны и их изучение вызывает большой интерес.

За основу для построения модели была выбрана структура 1KX5 [17] из банка данных PDB, к которой были добавлены прямые линкеры в соответствии с последовательностью, используемой в эксперименте FRET. На основании данных футпринтинга были отобраны участки ДНК, которые не связаны с белком. После проведения минимизации геометрии ДНК с ограничениями расстояний из данных FRET была получена асимметричная структура с параллельным расположением линкерных участков ДНК (рис. Б). Весьма схожие модели были получены в работе [18], где структуру хроматосом исследовали методом гидроксильного футпринтинга и электронной микроскопии. В дальнейшем полученная структура может быть применена в качестве мишени для макромолекулярного докинга гистона H1.

Понимание молекулярных основ функционирования хроматина важно, так как существует проблема ограниченности экспериментальных методик для определения параметров взаимодействий ДНК с гистонами, транскрипционными факторами и структурными белками на молекулярном уровне. Данная проблема является следствием того, что экспериментальные методы вроде РСА, ЯМР и электронной микроскопии в состоянии определить структуры лишь небольших комплексов, которые обладают компактной и упорядоченной структурой. В то же время все больший интерес привлекают к себе структуры, не отличающиеся строгой упорядоченностью и их описание должно вестись не в рамках



одиночных структур, а в рамках набора конформационных ансамблей.

Разработанный метод позволяет создавать модели геометрии ДНК в комплексах с белками на основании интегрирования экспериментальных данных и данных о локальной жесткости ДНК при помощи молекулярного моделирования. Этот метод позволяет также получать атомистические модели и изучать особенности структуры как одиночных нуклеосом, так и их комплексов с другими макромолекулами. На примере хроматосомы показана принципиальная возможность получения геометрии линкерной ДНК, что важно для понимания устройства хроматина.

Авторы работы выражают благодарность М.П. Кирпичникову, В.М. Студитскому, А.В. Феофанову и сотрудникам их лабораторий за предоставленные экспериментальные данные.

Работа выполнена при финансовой поддержке Российского научного фонда (проект № 14-24-00031).

Рисунок. А) Схема алгоритма поиска конформации ДНК с учетом измерений проведенных методами FRET и футпринтинга. Пунктиром выделен блок минимизации энергии при поиске структур. Б) Визуализация результата поиска конформации хроматосомы. Отобрана вторичная структура гистонов, для ДНК показана поверхность, доступная растворителю

СПИСОК ЛИТЕРАТУРЫ

1. Kornberg R.D. Chromatin structure: a repeating unit of histones and DNA // *Science*. 1974. Vol. 184. N 4139. P. 868–871.
2. Luger K., Mäder A.W., Richmond R.K., Sargent D.F., Richmond T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution // *Nature*. 1997. Vol. 389. N 6648. P. 251–260.
3. Studitsky V.M., Clark D.J., Felsenfeld G. Overcoming a nucleosomal barrier to transcription // *Cell*. 1995. Vol. 83. N 1. P. 19–27.
4. Shaytan A.K., Landsman D., Panchenko A.R. Nucleosome adaptability conferred by sequence and structural variations in histone H2A-H2B dimers // *Curr. Opin. Struct. Biol.* 2015. Vol. 32. P. 48–57.
5. Schalch T., Duda S., Sargent D.F., Richmond T.J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre // *Nature*. 2005. Vol. 436. N 7047. P. 138–141.
6. Mueller-Planitz F., Klinker H., Becker P.B. Nucleosome sliding mechanisms: new twists in a looped history // *Nat. Struct. Mol. Biol.* 2013. Vol. 20. N 9. P. 1026–1032.
7. Duyn V., Vinogradov S.A., Lampson M.A., Black B.E., Falk S.J., Guo L.Y., Sekulic N., Smoak E.M., Mani T., Logsdon G.A., Gupta K., Jansen L.E.T., Gregory D. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere // *Science*. 2015. Vol. 348. N 6235. P. 699–703.
8. Hayes J.J., Tullius T.D., Wolffe A.P. The structure of DNA in a nucleosome // *Proc. Natl. Acad. Sci. USA*. 1990. Vol. 87. N 19. P. 7405–7409.
9. Jain S.S., Tullius T.D. Footprinting protein-DNA complexes using the hydroxyl radical // *Nat. Protoc.* 2008. Vol. 3. N 6. P. 1092–1100.
10. Shaytan A.K., Armeev G.A., Goncarencu A., Zhurkin V.B., Landsman D., Panchenko A.R. Combined influence of linker DNA and histone tails on nucleosome dynamics as revealed by microsecond molecular dynamics simulations // *J. Biomol. Struct. Dyn.* 2015. Vol. 33. Suppl 1. P. 3.
11. Dickerson R.E., Bansal M., Calladine C.R., Diekmann S., Hunter W.N., Kennard O., von Kitzing E., Lavery R., Nelson H.C.M., Olson W.K. Definitions and nomenclature of nucleic acid structure parameters // *J. Mol. Biol.* 1989. Vol. 205. N 4. P. 787–791.
12. Pronk S., Páll S., Schulz R., Larsson P., Bjelkmar P., Apostolov R., Shirts M.R., Smith J.C., Kasson P.M., van der Spoel D., Hess B., Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit // *Bioinformatics*. 2013. Vol. 29. N 7. P. 845–854.
13. Shadle S.E., Allen D.F., Guo H., Pogożelski W.K., Bashkin J.S., Tullius T.D. Quantitative analysis of electrophoresis data: novel curve fitting methodology and its application to the determination of a protein–DNA binding constant // *Nucleic Acids Res.* 1997. Vol. 25. N 4. P. 850–860.
14. Olson W.K., Gorin A.A., Lu X.-J., Hock L.M., Zhurkin V.B. DNA sequence-dependent deformability deduced from protein–DNA crystal complexes // *Proc. Natl. Acad. Sci. USA*. 1998. Vol. 95. N 19. P. 11163–11168.
15. Lu X., 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. N 17. P. 5108–5121.
16. Zhou B.-R., Feng H., Kato H., Dai L., Yang Y., Zhou Y., Bai Y. Structural insights into the histone H1-nucleosome complex // *Proc. Natl. Acad. Sci. USA*. 2013. Vol. 110. N 48. P. 19390–19395.
17. Davey C.A., Sargent D.F., Luger K., Maeder A.W., Richmond T.J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution // *J. Mol. Biol. Elsevier*. 2002. Vol. 319. N 5. P. 1097–1113.
18. Syed S.H., Goutte-Gattat D., Becker N., Meyer S., Shukla M.S., Hayes J.J., Everaers R., Angelov D., Bednar J., Dimitrov S. Single-base resolution mapping of H1–nucleosome interactions and 3D organization of the nucleosome // *Proc. Natl. Acad. Sci. USA*. 2010. Vol. 107. N 21. P. 9620–9625.

Поступила в редакцию
15.09.2015

MODELING OF POTEIN-DNA COMPLEXES GEOMETRY UTILISING FRET AND FOOTPRINTING DATA

G.A. Armeev, T.K. Gorkovets, D.A. Efimova, K.V. Shaitan, A.K. Shaytan*

Department of Bioengineering, School of Biology, M.V. Lomonosov Moscow State University,
Leninskiye gory 1–12, Moscow, 119234, Russia;

*e-mail: alex@molsim.org

We discuss the question of constructing three-dimensional models of DNA complexed with proteins on the basis of computer modeling and indirect methods for studying the structure of macromolecules. We consider methods of interpreting the experimental data obtained by indirect methods for studying the three-dimensional structure of biomolecules. We discuss some aspects of the integration of such data in to the process of molecular models building (based on the geometric characteristics of DNA). We propose an algorithm for the estimation of protein-DNA complexes structure on the basis of information about the local DNA flexibility and experimental data obtained by Forster resonance energy transfer (FRET) and hydroxyl footprinting methods. Finally, we use this method for prediction of the hypothetical configuration of DNA in nucleosome bound with histone H1.

Key words: molecular modeling, DNA, FRET, footprinting, nucleosomes, chromatin.

Сведения об авторах:

Армеев Григорий Алексеевич — аспирант, мл. науч. сотр. кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: armeev@molsim.org

Горковец Татьяна Константиновна — аспирант кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: gorkovets@molsim.org

Ефимова Дарья Андреевна — студент кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: dar9468@yandex.ru

Шайтан Константин Вольдемарович — докт. физ-мат. наук, проф. кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-23-74; e-mail: shaytan49@yandex.ru

Шайтан Алексей Константинович — канд. физ-мат. наук, вед. науч. сотр. кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: alex@molsim.org