

7. Use of «Perceptron» algorithm to distinguish translational initiation sites in *E. coli* / G. D. Stormo, T. D. Scheider, L. Gold, A. Ehrenfucht // Nucl. Acids Res.— 1985.— 13, N 8.— P. 2997—3011.
8. Энхансероподобные структуры в умеренно повторяющихся последовательностях эукариотических геномов / И. А. Шахмурадов, Н. А. Колчанов, В. В. Соловьев, В. А. Ратнер // Генетика.— 1986.— 22, № 3.— С. 357—367.

Ин-т ботаники им. В. Л. Комарова АН АзССР, Баку

Получено 10.05.90

УДК 576.315.42

В. В. Шматченко, А. Б. Бережнев

КАРТИРОВАНИЕ МЕСТ ПРИКРЕПЛЕНИЯ ДНК К ЯДЕРНОМУ СКЕЛЕТУ МЕТОДОМ ГРАФИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ПРОТЯЖЕННЫХ НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Методом графического представления нуклеотидных последовательностей в виде кривых линий, отражающих распределение АТ- и GC-оснований по длине последовательности, выявлены характерные (S-образные) профили кривых, соответствующие местам крепления ДНК к ядерному скелету эукариот. Впервые показана применимость использованного метода для картирования участков связывания ДНК со скелетными структурами ядра. Тем самым продемонстрирована возможность обнаружения функционально однотипных негомологических участков ДНК, что не представляется возможным с помощью других известных компьютерных методов анализа протяженных нуклеотидных последовательностей.

Введение. К настоящему времени установлена первичная структура большого числа генов и прилегающих к ним некодирующих областей ДНК. Поток такой информации нарастает и требует осмысления. Анализ же структурно-функциональной организации генома сегодня несколько отстает от процесса накопления данных о первичных последовательностях. Поэтому определенную ценность представляет всякий свежий подход, проливающий свет на выяснение функциональной значимости тех или иных участков последовательностей.

Широкие возможности в этом плане открывают разнообразные компьютерные методы анализа последовательностей, позволяющие обрабатывать значительные массивы информации и представлять результаты в удобной для осмысления форме, например в виде графиков. Одним из перспективных подходов такого рода является метод графического представления нуклеотидных последовательностей, предложенный Хамори [1], где для визуального анализа предлагается более детальная картина, чем та, которая получается при глобальном анализе содержания GC-оснований. В последнем случае фиксируется лишь суммарный уровень GC- и AT-оснований, имеющий биологический смысл (см., например, [2]). Кроме того, данный метод в нашем исследовании позволяет судить о функциональном подобии негомологических последовательностей ДНК путем установления схожести профилей получаемых с его помощью кривых линий, являясь своего рода графическим аналогом метода выравнивания. От выравнивания метод построения профилей нуклеотидных последовательностей отличает возможность эффективного сопоставления протяженных участков последовательностей, не обладающих к тому же близостью первичных структур, что бывает актуально в ряде случаев. В частности, при анализе хромосомной ДНК на предмет наличия мест прикрепления к ядерному скелету нами был использован данный метод как наиболее адекватный решаемой проблеме.

© В. В. ШМАТЧЕНКО, А. Б. БЕРЕЖНЕВ, 1990

По современным представлениям, ДНК эукариотических клеток организована в виде многочисленных суперскрученных петель, основания которых прикреплены к скелетным структурам ядра (ядерному матриксу) или хромосомы (хромосомному остову) [3—6]. Средний размер петли ДНК составляет приблизительно 50 тысяч пар нуклеотидов (т. п. н.), а районы прикрепления ДНК к ядерному скелету имеют размер несколько сотен пар нуклеотидов [3, 4, 7]. Известно, что места

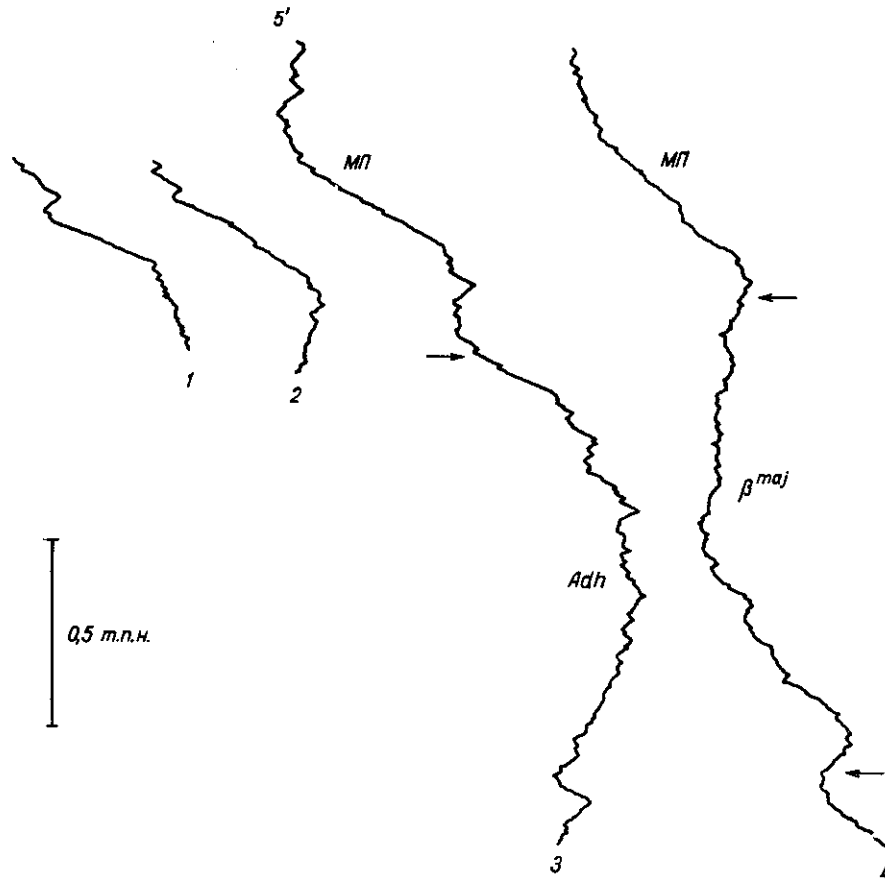


Рис. 1. Графическое представление нуклеотидных последовательностей: часть гена дигидрофолатредуктазы, прочно связанная с ядерным матриксом клеток хомяка (1); 5'-фланкирующая область гена *Sgs4* *D. melanogaster* (2); ген алкогольдегидрогеназы *D. melanogaster* (3); β-глобиновый ген мыши (4). МП — место прикрепления ДНК к ядерному скелету. Стрелками указаны границы генов

Fig. 1. Graphic representation of nucleotide sequences: matrix attachment region of hamster dihydrofolate reductase gene (1); 5'-flanking sequence of *Drosophila Sgs4* gene (2); alcohol dehydrogenase gene of *Drosophila melanogaster* (3); mouse β-globin gene (4). МП — matrix association region. Arrows point the boundaries of genes.

прикрепления обогащены АТ-основаниями и, как правило, располагаются в межгенных некодируемых областях хромосомной ДНК [7—9]. Домениная организация генома необходима для поддержания компактного состояния ДНК и, по-видимому, играет важную роль в процессах репликации и регуляции генной активности [6, 8, 10].

В настоящей работе для картирования областей прикрепления ДНК к ядерному скелету применяли компьютерную программу PROFILE (ИБФМ АН СССР), реализующую алгоритм построения графиков нуклеотидных последовательностей на IBM PC-совместимых компьютерах. В качестве источника последовательностей ДНК мы использовали банк данных GenBank. Были проанализированы нуклеотидные

последовательности организмов, представляющих различные таксономические группы. Показано соответствие характерных S-образных профилей участкам ДНК, отвечающим за связывание с ядерным скелетом.

Материалы и методы. Для построения графиков последовательностей ДНК применяли подход Хамори [1]. Согласно этому методу, нуклеотидные остатки последовательности рассматриваются в качестве фиксированных векторов, состыкованных друг с другом по принципу «голова — хвост». Векторы вычерчивают в трехмерном пространстве кривую линию, полностью отражающую распределение этих нуклеотидных остатков в последовательности ДНК.

В настоящей работе мы использовали вариант метода графического представления последовательностей, в котором из стереометрической пары изображений, дающих в совокупности объемную картину, применяли лишь одно из изображений последовательности ДНК. Подобный метод применяли, в частности, в работе Хамори [11].

Алгоритм графического представления последовательностей путем вычерчивания соответствующих профилей был реализован в ИБФМ АН СССР в виде компьютерной программы PROFILE на языке СИ применительно к IBM PC-совместимым компьютерам.

Последовательности ДНК, начиная с 5'-конца, преобразовывали в кривые, вычерчивание которых начинали с верхней части графика, откладывая влево и вниз отрезки, соответствующие G- и C-основаниям, а вправо и вниз — соответствующие A- и T-основаниям.

Для анализа использовали также программу дот-матричного анализа DNASEQ (ИБ АН СССР) и программу анализа нуклеотидных и аминокислотных последовательностей DM5 (Аризонский университет, США).

Последовательности ДНК получили из 55-го выпуска банка данных GenBank.

При работе использовали компьютер Правец-16А (Болгария). Вычерчивание профилей нуклеотидных последовательностей проводили на плоттере МИКРОНИКА ПР 297 (Болгария) и принтере Robotron K 6314 (ГДР).

Результаты и обсуждение. К настоящему времени в окрестностях различных генов высших организмов экспериментально локализованы участки прикрепления ДНК к ядерному скелету, общим свойством которых является их АТ-богатый состав [7—9, 12, 13]. В ходе работы мы сравнили профили кривых, получаемых в результате графического представления данных участков ДНК.

На рис. 1 кривые 1 и 2 соответствуют участкам последовательностей ДНК, которые крепятся к скелетным структурам: часть гена дигидрофолатредуктазы хомяка [13] (кривая 1) и 5'-фланкирующая область гена *Sgs4 Drosophila melanogaster* [9] (кривая 2).

На рис. 1 также представлены профили нуклеотидной последовательности гена алкогольдегидрогеназы *D. melanogaster* (кривая 3) и β-глобинового гена мыши (кривая 4). Согласно экспериментальным данным [9, 14], 5'-концевые области этих генов прилегают к ядерному матриксу.

Как видно из графика, кривые 1 и 2, а также 5'-концевые участки кривых 3 и 4, соответствующие районам прикрепления ДНК к ядерному скелету, имеют сходный профиль. Похожий профиль мы наблюдали и в ряде других аналогичных случаев. Всего было проанализировано 14 последовательностей ДНК, т. е. практически все известные к настоящему времени последовательности, для которых установлены места крепления к ядерному матриксу (данные не представлены).

Применяемый нами метод дает возможность более детально проанализировать участки, соответствующие местам крепления ДНК к ядерному матриксу, для которых обычно указывается лишь процентное содержание АТ-оснований. Графический метод выявляет детальное распределение АТ- и GC-оснований по длине этих участков: центральный район с высоким уровнем содержания АТ-оснований и более короткие фланкирующие отрезки, относительно равномерно насыщенные как АТ-, так и GC-основаниями. Соответствующие данным участкам профили имеют характерный S-образный вид.

В свою очередь, наличие характерных S-профилей дает возможность визуального выявления потенциальных мест прикрепления ДНК к ядерному скелету. Дополнительным подтверждением соответствия выявляемых участков предполагаемой функции может быть наличие на этих участках сайтов узнавания топоизомеразой II. Поскольку, как известно, основным белковым компонентом ядерного матрикса [15] и

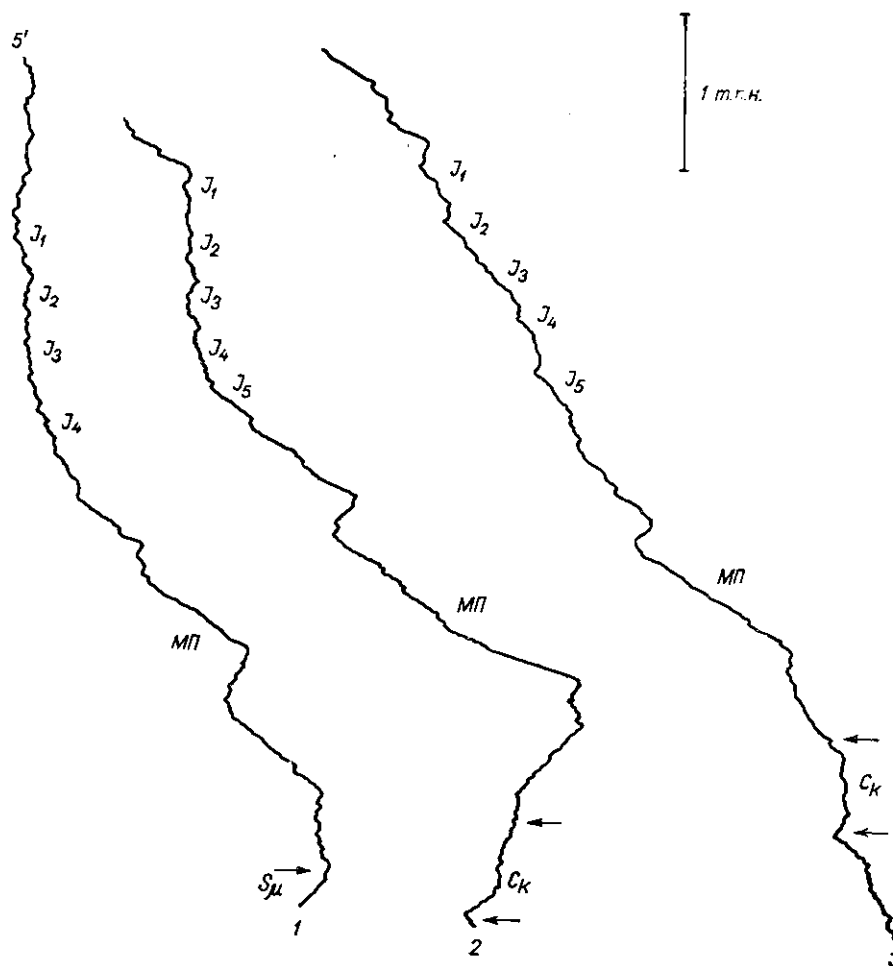


Рис. 2. Сравнение кривых, соответствующих локусам иммуноглобулиновых генов: ген μ -цепи иммуноглобулинов мыши (1); ген κ -цепи иммуноглобулинов кролика (2); ген, кодирующий κ -цепь иммуноглобулинов мыши (3). МП — место прикрепления ДНК к ядерному скелету. Стрелками указаны границы генов

Fig. 2. The comparison of curves corresponding to immunoglobulin gene loci: mouse mu immunoglobulin (1); rabbit kappa immunoglobulin (2); mouse kappa immunoglobulin (3). МП — matrix association region. Arrows point the boundaries of genes.

хромосомного остова [16] является топоизомераза II, а участки ДНК, прилегающие к указанным скелетным структурам, имеют последовательности, которые узнаются этим ферментом.

В целях изучения возможности предсказания потенциально однотипных по выполняемой функции районов на примере мест связывания ДНК с ядерным матриксом сравнивали профили нуклеотидных последовательностей родственных генов. Для этого были выбраны эмбриональные гены иммуноглобулинов мыши с экспериментально установленными местами прикрепления, а также аналогичный ген кролика, где данный функциональный район не установлен.

Графический профиль последовательности ДНК, кодирующей μ -цепь иммуноглобулина мыши (рис. 2, кривая 1), имеет характерный S-образный вид в той части кривой, которая соответствует месту связывания ДНК с ядерным скелетом. Место крепления ДНК в данном случае расположено в большом интроне указанного генетического локуса [12].

Кривая 3 (рис. 2) соответствует графическому представлению части генетического локуса, кодирующего κ -цепь мышечных иммуноглобулинов [17]. Как видно из графика, в пределах данного участка генома в интроне между семействами J-генов и C-геном имеется S-профиль, соответствующий участку прикрепления ДНК к ядерному скелету [8].

Кривая 2 (рис. 2) представляет собой графическое изображение гена κ -цепи иммуноглобулина кролика [18]. На участке кривой, соответствующей большому интрону этого гена (между J5-сегментом и C-геном), обнаруживается характерный S-профиль.

Кроме того, в данном участке ДНК (рис. 2, кривая 2) имеются три сайта узнавания топоизомеразой II (поиск вели по канонической последовательности для топоизомеразы II *D. melanogaster* GTN(A/T)A(T/C)ATTNATNN(G/A) [19]). Хотя экспериментальные данные о соответствии этого участка месту связывания ДНК с ядерным матриксом отсутствуют, можно предположить, что он выполняет именно такую функцию.

Помимо графического построения мы провели сравнение первичных структур участков прикрепления ДНК методом дот-матричного анализа. При заданных условиях сравнения последовательностей (не менее 13 совпадений в окне из 15 оснований) гомология этих участков ДНК не выявлена (данные не приведены).

Таким образом, метод, примененный в настоящей работе, позволяет картировать функционально подобные участки на протяженных последовательностях ДНК (в том числе имеющие различные первичные структуры при одной и той же биологической функции, как это было продемонстрировано нами на примере анализа мест прикрепления ДНК к ядерному скелету). Данный метод может служить основой для скрининга банков нуклеотидных последовательностей путем визуального распознавания функционально значимых участков генома.

MAPPING DNA ATTACHMENT SITES ON NUCLEAR FRAME BY GRAPHIC REPRESENTATION OF LONG DNA SEQUENCES

V. V. Schmatchenko, A. B. Berezhnev

Institute of Biochemistry and Physiology of Microorganisms,
Academy of Sciences of the USSR, Pushchino, Moscow region

Summary

Characteristic (S-form) curves corresponding to nuclear matrix association regions have been revealed by graphically represented DNA sequences. The curves reflect AT and GC distribution along the sequence length. The method is applicable for mapping DNA attachment sites on nuclear frame. Functionally identical nonhomologous DNA regions in long sequences can be detected which is impossible using other computer-assisted methods of nucleotide sequence analysis.

СПИСОК ЛИТЕРАТУРЫ

1. Hamori E., Ruskin J. H curves, a novel method of representation of nucleotide series especially suited for long DNA sequences // *J. Biol. Chem.*—1983.—258, N 2.— P. 1318—1327.
2. Ikemura T., Aota S. Global variation in G+C content along vertebrate genome DNA. Possible correlation with chromosome band structures // *J. Mol. Biol.*—1988.—203, N 1.— P. 1—13.
3. Benyajati C., Worcel A. Isolation, characterization and structure of the folded interphase genome of *Drosophila melanogaster* // *Cell.*—1976.—9, N 3.— P. 393—407.
4. Paulson J. R., Laemmli U. K. The structure of histone depleted metaphase chromosomes // *Ibid.*—1977.—12, N 5.— P. 817—828.

5. Vogelstein B., Pardoll D. M., Coffey D. S. Supercoiled loops and eukaryotic DNA replication // *Ibid.*—1980.—22, N 1.—P. 79—85.
6. Georgiev G. P., Nedospasov S. A., Bakayev V. V. Supranucleosomal levels of chromatin organization // *The Cell Nucleus.*—1978.—6, N 1.—P. 3—34.
7. Mirkovitch J., Mirault M.-E., Laemmli U. K. Organization of the higher-order chromatin loop: Specific DNA attachment site on nuclear scaffold // *Cell.*—1984.—39, N 1.—P. 223—232.
8. Cockerill P. N., Garrard W. T. C. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites // *Ibid.*—1986.—44, N 2.—P. 273—282.
9. Gasser S. M., Laemmli U. K. Cohabitation of scaffolding binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster* // *Ibid.*—46, N 3.—P. 521—530.
10. Dijkwel P. A., Hamlin J. L. Matrix attachment region are positioned near replication initiation sites, genes, and an interamlicon junction in the amplified dihydrofolate reductase domain of chinese hamster ovary cells // *Mol. and Cell. Biol.*—1988.—8, N 12.—P. 5398—5409.
11. Hamori E., Varga G. DNA sequence (H) curves of the human immunodeficiency virus I and some related viral genomes // *DNA.*—1988.—7, N 5.—P. 371—378.
12. Cockerill P. N., Yuen M.-H., Garrard W. T. The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements // *J. Biol. Chem.*—1987.—262, N 11.—P. 5394—5397.
13. Kas E., Chasin L. A. Anchorage of the chinese hamster dihydrofolate reductase gene to the nuclear scaffold occurs in an intragenic region // *J. Mol. Biol.*—1987.—198, N 4.—P. 677—692.
14. Greenstein R. J. Constitutive attachment of murine erythroleukemia cell histone depleted DNA loops to nuclear scaffolding is found in the β -major but not the α 1-globin gene // *DNA.*—1988.—7, N 9.—P. 601—607.
15. Berrios M., Osheroff N., Fisher P. A. In situ localization of DNA topoisomerase II, a major polypeptide component of the *Drosophila* nuclear matrix fraction // *Proc. Nat. Acad. Sci. USA.*—1985.—82, N 13.—P. 4142—4146.
16. *Metaphase* chromosome structure. Involvement of topoisomerase II / S. M. Gasser, T. Laroche, J. Falquet et al. // *J. Mol. Biol.*—1986.—188, N 3.—P. 613—629.
17. Max E. E., Maizel J. V., Jr., Leder P. The nucleotide sequence of a 5,5-kilobase DNA segment containing the mouse κ immunoglobulin *J* and *C* region genes // *J. Biol. Chem.*—1981.—256, N 10.—P. 5116—5120.
18. Emorine L., Max E. E. Structural analysis of a rabbit immunoglobulin κ 2 *J-C* locus reveals multiple deletions // *Nucl. Acids Res.*—1983.—11, N 24.—P. 8877—8890.
19. Sander M., Hsieh T.-S. Double strand DNA cleavage by type II DNA topoisomerase from *Drosophila melanogaster* // *J. Biol. Chem.*—1983.—258, N 15.—P. 8421—8428.

Ин-т биохимии и физиологии микроорганизмов АН СССР,
Пуццно

Получено 11.03.90

УДК 576.315.42

В. В. Волков, А. Ю. Леонтьев

ИССЛЕДОВАНИЕ СИММЕТРИИ ГЕНЕТИЧЕСКИХ ТЕКСТОВ МЕТОДОМ ФУРЬЕ-АНАЛИЗА

В работе предложен метод классификации симметричных структур одноцепочечной ДНК, использующий понятие цветной симметрии. Приводится систематическое перечисление цветных точечных и пространственных групп. Для поиска структур с инверсионной симметрией предлагается использовать фурье-анализ с фильтрацией, основанной на симметричной числовой кодировке нуклеотидной последовательности. Прямые повторы обнаруживаются с помощью фурье-анализа как первого этапа поиска, существенно сокращающего число операций сравнения. Методы опробованы на модельных последовательностях и областях инициации репликации прокариотического генома.

В настоящее время общеизвестно, что такие функциональные сайты, как точка начала репликации, сайты узнавания рестриктаз, сайты терминации транскрипции, энхансеры и другие являются симметричными структурами либо обогатены ими (см., например, [1—4]). Поэтому одним из этапов выявления функциональных сайтов в генетических текстах является поиск симметричных структур в молекуле ДНК.

© В. В. ВОЛКОВ, А. Ю. ЛЕОНТЬЕВ, 1990