

of receptor binding site can be obtained as a mould of the hydrophobic part of the molecule, immersed into lipid phase, in the so-called «biologically active» conformation. Therefore, it is possible to make analogs with enhanced affinity to receptor; this possibility is exemplified by newly-constructed carba-analog of angiotensin.

СПИСОК ЛИТЕРАТУРЫ

1. Никифорович Г. В., Галактионов С. Г., Чипенс Г. И. Конформации пептидных биорегуляторов.— М.: Медицина, 1983.— 192 с.
2. Surface properties of an amphiphilic peptide hormone and its analog: corticotropin releasing factor and sauvagine / S. H. Lau, J. Rivier, W. Vale et al. // Proc. Nat. Acad. Sci. USA.— 1983.— 80, N 23.— P. 7070—7074.
3. Gysin B., Schwyzer R. Liposome-mediated labelling of adrenocorticotropin fragments parallels their biological activity // FEBS Lett.— 1983.— 158, N 1.— P. 12—16.
4. Theoretical conformational analysis of oxytocin molecule / G. V. Nikiforovich, V. I. Leonova, S. G. Galaktionov, G. I. Chipens // Int. J. Pept. Prot. Res.— 1979.— 13, N 3.— P. 363—373.
5. Hopfinger A. J. Battershell R. D. Application of SCAP to drug design. 1. Prediction of octanol-water partition coefficients using solvent-dependent conformational analysis // J. Med. Chem.— 1976.— 19, N 5.— P. 569—573.
6. Структура и функции низкомолекулярных пептидов.— Рига: Зинатне, 1980.— 327 с.

Всесоюз. науч.-исследоват. и проектно-конструкт. ин-т микробиол. пр-в, Минск

Получено 28.03.90

УДК 577.323:681.3

Н. И. Луккина, Т. Р. Сойдла

ПОИСК ГИГАНТСКИХ ДНК-БЕЛКОВЫХ КОМПЛЕКСОВ ПРИ ПОМОЩИ ЭВМ

Созданы две программы для выявления и визуализации пространственной асимметрии ДНК. С их помощью анализировали 2 мкм плазмиду дрожжей-сахаромицетов. Выявлен участок особой ориентации G—C-пар. Расположение этого участка совпадает с локализацией нуклеосомного ДНК-белкового комплекса, выявленного ранее в экспериментальной работе. Созданные программы могут быть использованы для поиска крупных ДНК-белковых комплексов.

Введение. Начало секвенирования огромных геномов эукариот ставит перед исследователем заманчивый, но трудный вопрос: существуют ли правила, которые позволяют на основе одной лишь нуклеотидной последовательности выявить участки ДНК, взаимодействующие с регуляторными, а, возможно, и другими нуклеосомными белками. Известно, что белки узнают на ДНК самые разные последовательности, которые невозможно свести к одному или нескольким элементарным консенсус-последовательностям. Список белков, взаимодействующих с ДНК, уже долгие годы остается очень и очень неполным [1]. Поэтому основной вопрос состоит в том, можно ли создать алгоритм для поиска участков ДНК, узнаваемых белками, при условии, что нам не известно, какие нуклеотидные последовательности эти белки узнают на ДНК.

При попытке создания наших программ поиска ДНК-белковых комплексов мы исходили из двух положений. Во-первых, что взаимодействие ДНК с белком чаще всего ограничено только одной стороной ДНК. Во-вторых, что на узнаваемой белком стороне ДНК нуклеотидный состав отличается от усредненных данных по рассматриваемому району. Таким образом, мы искали пространственно-ориентированные аномалии нуклеотидного состава ДНК.

© Н. И. ЛУККИНА, Т. Р. СОЙДЛА. 1990

Шаг спирали В-формы ДНК равен примерно 10 нуклеотидам [2]. Таким образом, в комплексе ДНК — белок могут чередоваться с периодичностью в 10 нуклеотидов участки, связанные с белком, и свободные (например, 5 нуклеотидов, занятых белком, 5 — свободных и т. д.). Экспериментальная оценка шага спирали дает 10,55 для свободной ДНК [3] и 10,15—10,25 — для ДНК в составе нуклеосомы [4]. Поэтому в данной работе мы проверили, можно ли на природных молекулах эукариотической ДНК обнаружить аномалии нуклеотидного состава ДНК, имеющие периодичность 10,0—10,6 нуклеотида. Для выявления этих аномалий использовали программы двух типов, которые будут описаны ниже. В данной статье мы приводим результаты исследования при помощи этих программ нуклеотидной последовательности так называемой 2-микронной (2 мкм) плазмиды эукариотического микроорганизма дрожжей *Saccharomyces cerevisiae* [5]. Наши программы позволяют уверенно выявлять комплексы ДНК — белок, занимающие 200—400 пар оснований (т. е. в 1,5—2 раза больше нуклеосомы).

Материалы и методы. Нуклеотидная последовательность 2 мкм плазмиды была переписана из банка нуклеотидных последовательностей EMBL. Программы для статистической обработки написаны на языках Fortran и Assembler для ЭВМ СМ-3 (операционная система ДЭС-400). Созданные программы, работающие совместно с пакетом статистических подпрограмм, позволяют прорисовывать на графопостроителе распределение нуклеотидов в произвольном наборе «окешек». Параметрически задаются 1) шаг выборки нуклеотидов; 2) шаг пропуска нуклеотидов; 3) окно вывода и 4) коррекция на шаг витка спирали. Первая программа показывает разницу в количестве одного из нуклеотидов: А, Т, С или G в выбранных и пропущенных последовательностях. В данной статье приводятся данные при следующих параметрах: выбор 5, пропуск 5, окошко 5. То есть выбирали 5 нуклеотидов, пропускали следующие 5, снова выбирали 5 нуклеотидов и т. д. Сравнивали $5 \times 5 = 25$ выбранных с 25 пропущенными нуклеотидами. Регистрировали разницу в количестве исследуемого нуклеотида в выбранных и пропущенных 25 нуклеотидных интервалах. Первое такое сравнение дает первую точку на графике. Вторая точка получается при сдвиге окошка на 10 нуклеотидов, затем снова на 10 нуклеотидов и т. д. до конца плазмиды. Следует другой вариант — такое же продвижение окошек, но при сдвиге первого выбора на 1 нуклеотид и т. д. Понятно, что в итоге получаются пять вариантов — разных попарных сравнений двух сторон ДНК. Для учета реального шага спирали (>10 , в данной работе 10,1—10,6) вводили коррекцию — выбирали «лишние» нуклеотиды. Во второй программе последовательность представляется графически в виде раскручивающейся спирали, нуклеотиды (А, Т, G, С в разных вариантах) отображаются в виде сегментов этой спирали.

Результаты и обсуждение. При помощи первой созданной программы анализировали распределение аденина, тимина, гуанина и цитозина по двум сторонам двойной спирали 2 мкм плазмиды дрожжей. Мы проверяли всю нуклеотидную последовательность плазмиды для четырех нуклеотидов при двух оценках шага спирали: 10,4 и 10,0. Оказалось, что одна аномалия в распределении нуклеотидов между двумя сторонами ДНК заметно выделяется на фоне других. Эта аномалия касается распределения G и С при шаге спирали, равном 10,4. Для С (рис. 1) расположение этой аномалии следующее. Максимальная аномалия в варианте 6,1 — с 2885-го по 3380-й нуклеотид. В 5,10 эта аномалия занимает с 2990-го по 3460-й нуклеотид, в 4,9 — с 2990-го по 3440-й. Для G (рис. 2) аномалия имеет большую протяженность в вариантах 6,1 (с 2920-го по 3550-й) и 5,10 (с 3110-го по 3500-й). Для А и Т (а также для всех нуклеотидов при оценке шага спирали 10,0) даже наибольшие аномалии намного короче рассмотренной выше. Для участка, где расположена данная аномалия (с 2600-го по 3600-й нуклеотид), проверяли все оценки шага спирали между 10,30 и 10,44 с интервалом 0,01, а также 10,5 и 10,6. Для оценок 10,5 и 10,6 протяженных аномалий нет. Максимальные оценки размеров аномалии для С были найдены при оценке шага спирали 10,38 — с 2915-го по 3530-й нуклеотиды (размер 615 нуклеотидов), а для G при 10,35 — с 2770-го

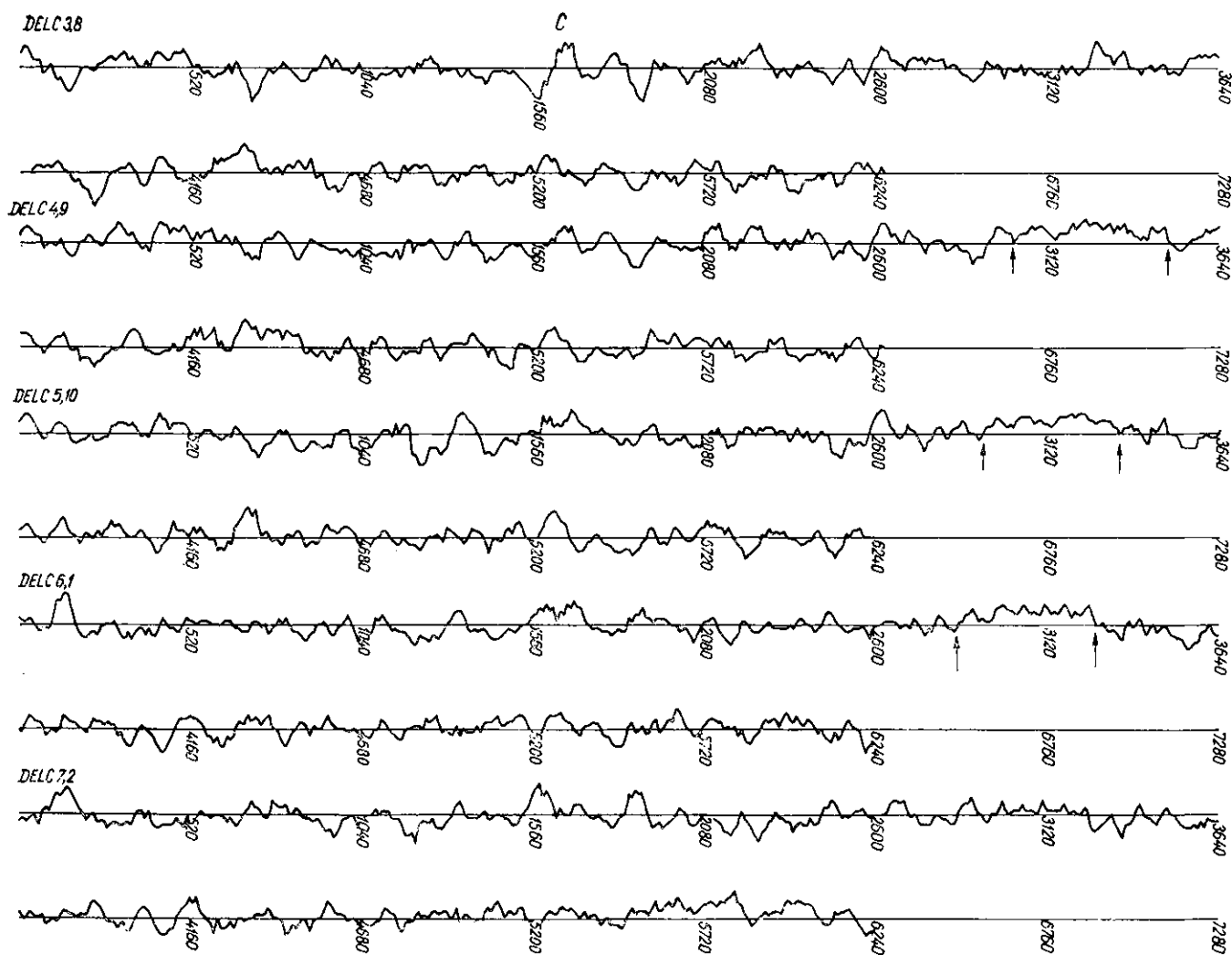


Рис. 1. Распределение С по длине 2 мкм плазмиды дрожжей (форма А). Показаны 5 попарных сравнений. 1, 2, 3 и т. д.— начало чтения соответственно с 1, 2, 3-го и т. д. нуклеотидов последовательности; 6, 1 — сравнение вариантов 6 и 1 и т. д. По вертикали показана разница в количестве С в двух перекрывающихся нуклеотидных окошках (5×5); горизонтальная линия — нулевая отметка для этого сравнения. Вдоль горизонтальной оси дана шкала в нуклеотидах 2 мкм плазмиды. Детали расчетов см. «Материалы и методы». Стрелки указывают аномалию, обсуждаемую в тексте

Fig. 1. Distribution of C along the 2 μm plasmid of yeast (A form). 5 pairwise comparisons are shown. 1, 2, 3 etc.—reading starts with first (second, third etc.) nucleotide of the sequence. 6, 1.— comparison of variants 6 and 1 etc. Vertical line shows difference in «C» quantity in two over-lapping 5×5 nucleotide windows. Horizontal line marks zero for this comparison. On horizontal line it is shown a scale in nucleotides of 2 μm plasmid. Performing of calculations see in Materials and methods. Arrows show the anomaly discussed in the text.

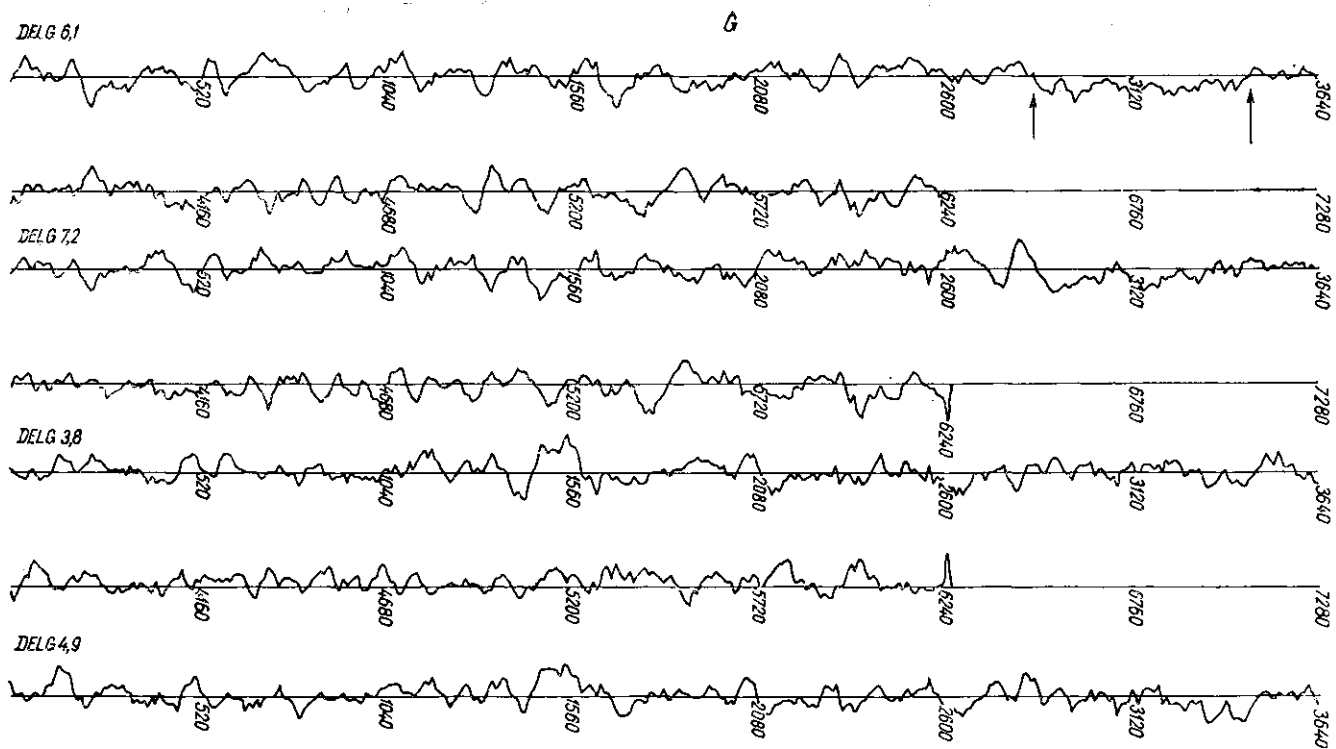


Рис. 2. Распределение G по длине 2 мкм плазмиды дрожжей (Форма А). Обозначения, как на рис. 1
Fig. 2. Distribution of G along the 2 μm plasmid of yeast (A form)
Details see Fig. 1.

по 3520-й нуклеотиды (размер 750 нуклеотидов). Сумма четырех наибольших оценок аномалии (две для С и две для G) максимальна при оценке шага спирали 10,37 (также и сумма шесть (3+3) наибольших оценок).

Если вернуться к рис. 1 и 2, то легко заметить, что следующая по размерам аномалия с 5060-го по 5430-й (вариант 6,1) для С имеет значительную длину только при одном попарном сравнении, т. е. по данным только этих результатов не исключено, что они представляют

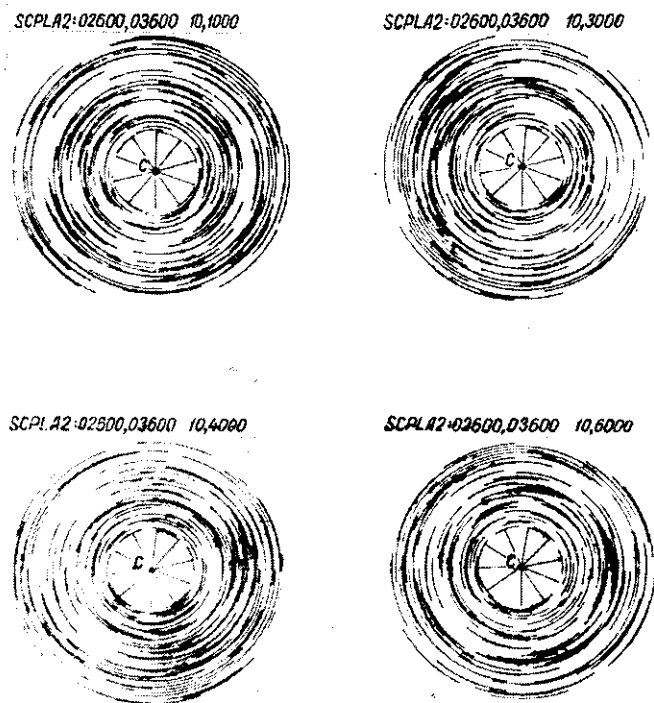


Рис. 3. Распределение С с 2600-го по 3600-й нуклеотид 2 мкм плазмиды дрожжей. ДНК показана в виде раскручивающейся спирали. Каждый остаток С — сегмент этой спирали. Сверху указаны использованные оценки шага спирали

Fig. 3. Distribution of C from 2600 to 3600 nucleotide of the 2 μ m plasmid of yeast. DNA is shown as an unwinding helix. Each C remain is shown as a segment of the helix. The helix lead estimates used, are shown above the helices.

собой статистические флуктуации. Проверка этих районов при разных оценках шага спирали позволяет судить о природе таких аномалий. Здесь отметим, что даже очень консервативная оценка позволяет уверенно (над уровнем чисто статистических колебаний) выявлять аномалии размером более 300—400 нуклеотидов, которые проявляются в более чем 1 варианте сравнения и(или) в оценках для более чем 1 нуклеотида.

Район, где располагается найденная нами аномалия (с 2600-го по 3600-й), мы проверяли при помощи второй нашей программы для 15 разных оценок шага спирали от 10,0 до 10,8. На рис. 3 видно, что для С аномалия практически не наблюдается при оценке шага спирали 10,1 или 10,6, но хорошо видна при оценках 10,3—10,4. Интуитивно ясно, что вторая программа более наглядна и позволяет анализировать пространственно-ориентированные аномалии в структуре ДНК при минимуме допущений о размерах и расположении этих аномалий.

При сравнении полученных нами данных с подробной картой 2 мкм плазмиды [6], видно, что выявленная нами аномалия (минимум с 2900-го по 3500-й нуклеотид) перекрывает локус REP3 (STB), участвующий в сегрегации плазмиды (участок с 2950-го по 3300-й) и регули-

рующей транскрипцию (участок с 2650-го по 2950-й) близлежащих генов [7]. Известно, что этот участок лишен нуклеосомной структуры [8]. Имеются предварительные данные о том, что этот район занят (вероятно, весьма лабильным) ДНК-белковым комплексом *in vivo* [9]. Таким образом, как первая, так и вторая наша программы позволяют выявить по крайней мере некоторые крупные ДНК-белковые комплексы.

Оценка 10,4 (точнее 10,37) для шага спирали у выявленного ДНК-белкового комплекса вполне соответствует ожидаемому для комплекса

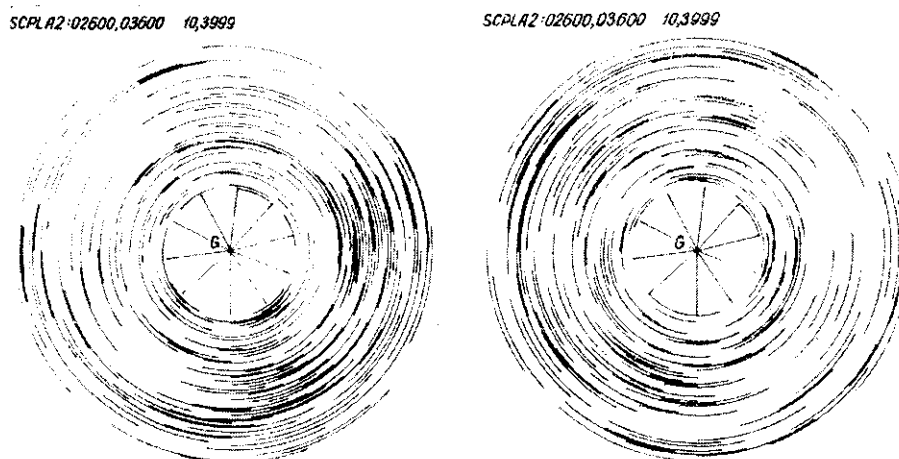


Рис. 4. Распределение С и G при оценке шага спирали 10,3999 с 2 600-го по 3 600-й нуклеотид 2 мкм плазмиды дрожжей. Обозначения, как на рис. 3

Fig. 4. Distribution of C and G on the yeast 2 μm plasmid at helical repeat estimate 10.3999. Details see Fig. 3.

более крупного, чем нуклеосома. Эта оценка промежуточна между данными для свободной ДНК (10,55) и ДНК в составе нуклеосомы (10,15—10,25) (см. выше).

Найденный участок имеет в своем составе 5-кратный неточный повтор 62—63-нуклеотидной консенсус-последовательности. Учитывая нашу оценку шага спирали, каждая 62—63-нуклеотидная последовательность достаточно точно соответствует шести полным оборотам спирали ДНК. Действительно, $6 \cdot 10,4 = 62,4$; $6 \cdot 10,37 = 62,22$, в то время как $6 \cdot 10,55 = 63,30$ и $6 \cdot 10,2 = 61,2$. Таким образом, консенсус-последовательность одинаково ориентирована во всех повторах. Кроме области повтора, выявленная нами аномалия охватывает около 250 нуклеотидов окружающей ДНК.

Как можно убедиться из данных рис. 1 и рис. 2, аномалии для G и C имеют противоположные знаки. Вторая программа, созданная нами, позволяет визуализировать, что в данной области мы имеем случай особой пространственной ориентации G—C-пар (рис. 4).

Несколько лет назад появилась широко известная статья Пташне [10] с классификацией ДНК-белковых комплексов, захватывающих достаточно далеко отстоящие участки ДНК. Описываются убедительные примеры сближения участков ДНК, связанных с белками, так что в промежутке ДНК свободна и выпетливается (looping). Автор при этом считает, что нет убедительных примеров механизма «наползания» (oozing), при котором ДНК покрывается цепочкой молекул регуляторных белков и образуется (возможно, лабильный) крупный ДНК-белковый комплекс. Легко заметить, что описанный нами случай попадает, скорее всего, в последнюю категорию и может в дальнейшем служить моделью для изучения «наползания».

В заключение хотелось бы отметить, что созданные нами программы могут быть в ближайшее время значительно улучшены. На основе полученных данных можно сформулировать одно предсказание. Нам кажется, что с улучшением экспериментальных методов будет увеличиваться количество зарегистрированных случаев, где ДНК оказывается на значительном расстоянии покрыта взаимодействующими молекулами (регуляторных) белков (см. также [1]).

A COMPUTER-AIDED SEARCH FOR GIANT DNA-PROTEIN COMPLEXES

N. I. Lukina, T. R. Soidla

Leningrad Polytechnical Institute;
Leningrad Nuclear Physics Institute

Summary

Two simple programs to detect extensive spatial asymmetries of DNA were created. The programs were used to study the 2 μ m plasmid of *Saccharomyces* yeast. A nearly 600 bp. uninterrupted region of special spatial orientation of G-C base pairs was found covering the REP 3 locus. The anomaly was obvious when estimating the helix lead approx. 104.4 (but not at 1.10.0 or 10.6). It is supposed that the region corresponds to a giant non-nucleosomal DNA-protein complex supposedly of an «oozing» mechanism according to the classification by Ptashne (1986). The complex is also indicated by some recent experimental results. The new programs can be used for searching giant DNA-protein complexes.

СПИСОК ЛИТЕРАТУРЫ

1. Johnson P. F. Eukaryotic transcriptional regulatory proteins // *Annu. Rev. Biochem.*— 1988.— 58.— P. 799—839.
2. Зенгер В. Принципы структурной организации нуклеиновых кислот.— М.: Мир, 1987.— 584 с.
3. Goulet I., Zivanovic Y., Prunell A. Helical repeat of DNA in solution. The V curve method // *Nucl. Acids Res.*— 1987.— 15, N 7.— P. 2803—2821.
4. Satchwell S. C., Drew H. R., Travers A. A. Sequence periodicities in chicken nucleosome core DNA // *J. Mol. Biol.*— 1986.— 191, N 4.— P. 659—675.
5. Hartley J. L., Donelson J. E. Nucleotide sequence of the yeast plasmid // *Nature.*— 1980.— 286, N 5776.— P. 860—865.
6. Сойдла Т. Р., Невзглядова О. В. 2 мкм плазида дрожжей-сахаромисетов // *Молекуляр. биология.*— 1987.— 21, № 6.— С. 1467—1479.
7. Murray J. A. H., Cesareni G. Functional analysis of the yeast plasmid partition locus STB // *EMBO J.*— 1986.— 5, N 12.— P. 3391—3399.
8. Fagrelus T. J., Livingston D. M. Location of DNAase I sensitive cleavage sites in the yeast 2 μ m plasmid DNA chromosome // *J. Mol. Biol.* — 1984.— 173, N 1.— P. 1—13.
9. Veit B. E., Fangman W. L. Chromatin organization of the *Saccharomyces cerevisiae* 2 μ m plasmid depends on plasmid-encoded products // *Mol. and Cell. Biol.*— 1985.— 5, N 9.— P. 2190—2196.
10. Ptashne M. Gene regulation by proteins acting nearby and at a distance // *Nature.*— 1986.— 322, N 6081.— P. 697—701.

Ленингр. политехн. ин-т
Ленингр. ин-т ядер. физики АН СССР

Получено 11.12.89