

УДК 577.21:577.323:581

## НУКЛЕОСОМНАЯ ОРГАНИЗАЦИЯ В САТЕЛЛИТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЯХ ДНК РАСТЕНИЙ

© 2012 г. В.Н. Бабенко<sup>1</sup>, К.О. Куташев<sup>1</sup>, В.Ф. Матвиенко<sup>2</sup>

<sup>1</sup> Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: bob@bionet.nsc.ru;

<sup>2</sup> Государственное образовательное учреждение высшего профессионального образования Новосибирский государственный университет, Новосибирск, Россия

Поступила в редакцию 4 ноября 2011 г. Принята к публикации 16 декабря 2011 г.

Сателлитная ДНК (сателлит), представленная повторяющимися tandemными последовательностями, является гетерохроматином с плотной нуклеосомной упаковкой. Мы искали среди контекстных и конформационных характеристик ДНК детерминирующие факторы, которые способствуют этой упаковке. Мы разработали статистику, основанную на величине изгиба ДНК, для определения характера предпочтения позиционирования нуклеосом в tandemных повторах. Было выявлено статистически значимое отличие профиля изгиба в tandemных повторах от случайной модели. Используя преобразование Фурье, мы выяснили, что не менее 50 % мономеров имеют периодичность профиля изгиба ДНК в районе 170 п.н., что соответствует длине ДНК, обернутой вокруг нуклеосомы, при рассмотрении двух первых гармоник в профиле изгиба ДНК. Мы определили, что в растениях схема позиционирования нуклеосом в субтеломерных сателлитах значительно отличается от аналогичной схемы в центромерных сателлитах.

**Ключевые слова:** нуклеосома, позиционирование, Фурье преобразование, изгиб ДНК, tandemные повторы.

### Введение

Период 2008–2010 гг. ознаменовал завершение 30-летнего поиска кода нуклеосомного позиционирования (Trifonov, 1980, 2011), с одной стороны, и генерацию большого числа последовательностей ДНК, связывающихся с нуклеосомой, высокопроизводительным методом (Schones *et al.*, 2008; Valouev *et al.*, 2008, 2011), с другой стороны. Было опубликовано множество статей и компьютерных приложений по предсказанию нуклеосомного позиционирования (Gupta *et al.*, 2008; Kaplan *et al.*, 2009, 2010; Xi *et al.*, 2010). Стоит отметить, что нуклеосомный код базируется в основном на специфике изгиба ДНК (Sussman, Trifonov, 1978; Schones *et al.*, 2008; Trifonov, 2011). Поскольку связывание нуклеосомы с ДНК стохастично по своей природе, сравнение существующих полногеномных программ предсказания позиционирования нуклеосом выявляет их существенные различия

(Tanaka, Nakai, 2009). Основной причиной этих несоответствий является достаточно низкая энергия связывания нуклеосом (Locke *et al.*, 2010; Valouev *et al.*, 2011), отражающая легкость удаления и восстановления связывания нуклеосомы для функционирования хроматина (Deal *et al.*, 2010). Более того, код нуклеосомного позиционирования часто перекрывается с другими информационными сигналами ДНК, такими, как репликационные, транскрипционные и трансляционные коды (Cohanin, Haran, 2009). Учитывая вышеизложенные обстоятельства, мы сфокусировались на определенных геномных районах, а именно – сателлитных повторах, где схема нуклеосомной упаковки хроматина играет основную функциональную роль.

Для анализа мы использовали данные из базы данных PlantSat (Macas *et al.*, 2002). Как видно из рис. 1, существует явное предпочтение мономеров к длине 140–170 п.н. и 330–350 п.н. В статье было предположено, что это предпо-

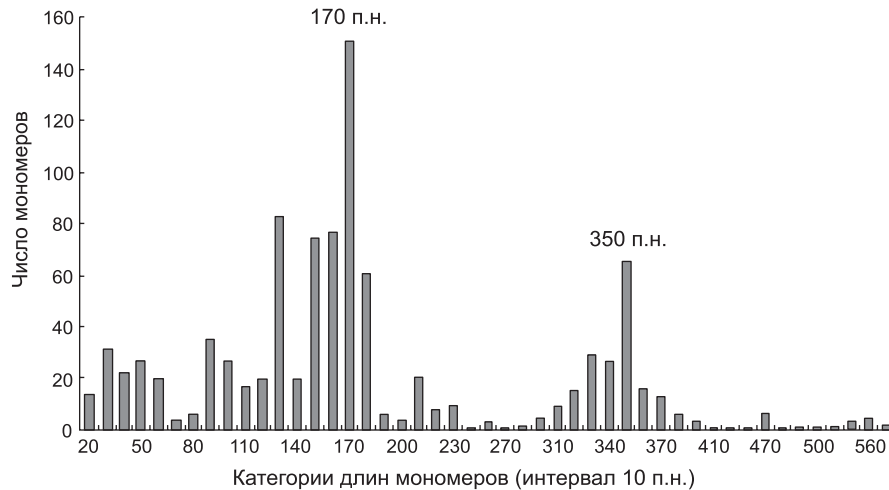


Рис. 1. Распределение длин мономеров из базы данных PlantSat (Macas *et al.*, 2002).

чение отражает длину обертывания ДНК вокруг нуклеосомы 140–180 п.н. (Macas *et al.*, 2002), но гипотеза не была подтверждена ни статистически, ни экспериментально. Более того, больше 60 % мономеров не попадают в эти интервалы по длине. Если для распределения длин мономеров в базе данных отобразить единственного представителя семейства для удаления избыточности, то профиль распределения длин становится еще менее иллюстративным. Цель данной работы – выяснить причины такой вариации длины и ее возможные последствия для позиционирования нуклеосом.

### Материалы и методы

Мы использовали выборку из базы данных PlantSat (версия 2008 г.), содержащей 140 видов растений и 975 последовательностей мономеров пяти основных классов по локализации: центромерных – 258, парацентромерных – 121, интеркалярных – 218, диспергированных – 14 и нуклеолярных (NOR) – 24, остальные – без локализации (Macas *et al.*, 2002). Последовательности были сгруппированы в 165 семейств повторов (по гомологии).

Мы использовали программу Phase, доступную на сайте <http://www.mgs.bionet.nsc.ru/mgs/gnw/nucleosom/>, для оценки паттерна распределения нуклеосом вдоль tandemно повторенных последовательностей. Алгоритм Phase использует неслучайное превышение (в зависимости

от ожидаемых по случайным причинам) динуклеотидов, фазированных (идентичных) на периоде 10–11 п.н., что влияет на адаптивный изгиб ДНК при оборачивании вокруг нуклеосомы. В целом алгоритм достаточно точно отражает профиль изгиба ДНК по отношению к нуклеосомному связыванию (Deal *et al.*, 2010; Locke *et al.*, 2010). Таким образом, для предсказания нуклеосомного потенциала далее использовался один фактор, а именно вышеназванный профиль изгиба ДНК, основанный на периодичности динуклеотидных частот.

Мы также применили дискретное преобразование Фурье (ДПФ) к профилям изгиба ДНК к последовательно повторенным мономерным последовательностям длиной 10 тыс. п.н. с помощью оригинальной программы, написанной на языке Perl.

Для вычисления коэффициентов преобразования Фурье использовались формулы из книги Феллера (Feller, 1971):

$$a_k = \frac{2}{n} \sum_{i=1}^n X_i \cos(\omega_k), \quad b_k = \frac{2}{n} \sum_{i=1}^n X_i \sin(\omega_k),$$

где  $\omega_k = \frac{2\pi k}{n}$  – частота Фурье,  $k = -[\frac{n-1}{2}], \dots,$

$[\frac{n}{2}]$ ,  $n$  – длина последовательности ДНК,  $[i]$  – целая часть от  $i$ ,  $X_i$  – значение профиля изгиба ДНК в позиции  $i$ .

В этом случае периодограмма выглядит как:

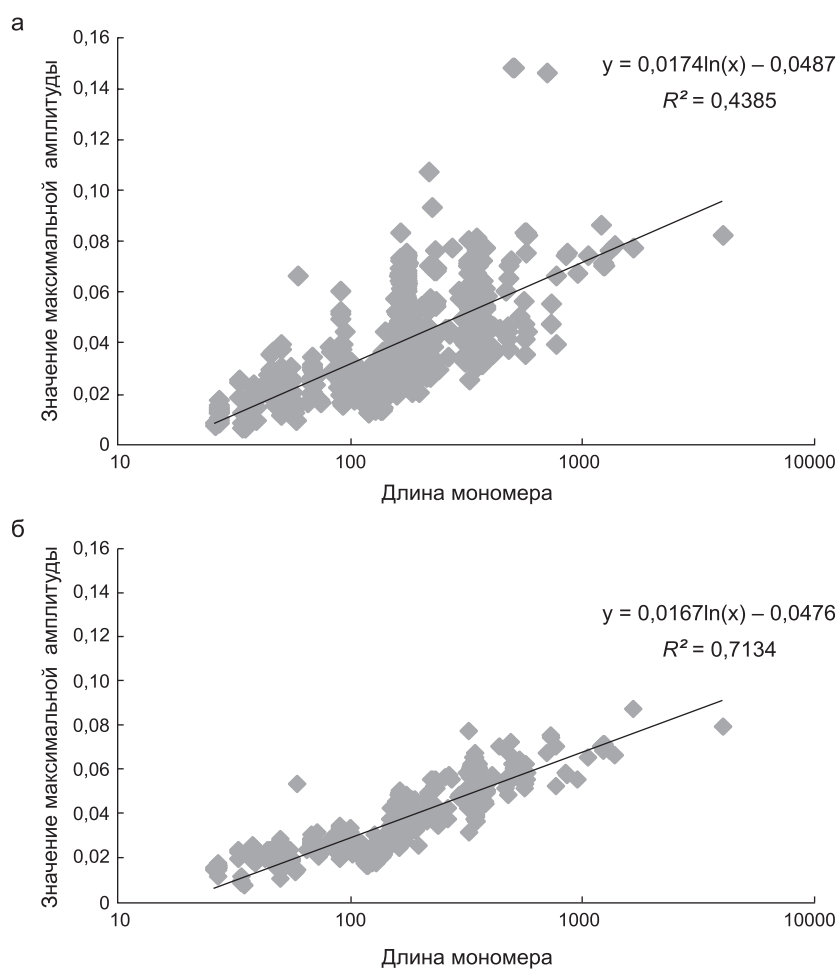
$$I_k = \frac{n}{2} \sum_{i=1}^n (a_k^2 + b_k^2).$$

## Результаты

### Максимальная амплитуда профиля изгиба ДНК как неслучайная характеристика позиционирования нуклеосом на сателлитах

Мы вычислили профили ДНК с помощью программы Phase для 901 мономера (74 мономера имели вырожденные последовательности, поэтому были исключены из анализа). После изучения профилей изгиба ДНК мы заключили, что большая часть мономеров имеет специфический профиль с высокими пиками и глубокими падениями. Основываясь на опубликованных ранее данных о выраженных амплитудах профиля изгиба ДНК для ряда функциональных элементов, таких, как старты

транскрипции, экзоны, поли-А-сигналы (Spies *et al.*, 2009; Ringrose, 2010), свидетельствующих о степени предпочтения нуклеосомного позиционирования в районах этих элементов, мы предположили, что высокая амплитуда является специфическим признаком для позиционирования нуклеосом в tandemных повторах. Для более детального исследования этого эффекта был построен график зависимости амплитуды от длины мономера для 901 элемента базы PlantSat (рис. 2, а). Параллельно была проведена симуляция для 901 мономера путем случайного перемешивания последовательности с теми же весами динуклеотидов, что и в исходной последовательности (рис. 2, б). Как видно, значение максимальной амплитуды экспоненциально зависит от длины, что вполне согласуется с теорией временных рядов и экстре-



**Рис. 2.** Зависимость значения амплитуды профиля изгиба ДНК от длины мономера для: а – 901 мономера из PlantSat; б – 901 симулированного значения амплитуды, полученного путем усреднения 100 испытаний для каждого мономера путем перемешивания последовательности с идентичным динуклеотидным составом.

мальных значений в приложении к ДНК (Karlin, Altschul, 1990).

При одинаковых параметрах (тангенс угла наклона и смещение) регрессионных линий (рис. 2, а, б) значения остаточной вариации ( $R^2$ ) резко отличаются. Дисперсионный критерий Фишера статистически подтвердил это, отвергнув гипотезу о гомогенности реальных и смоделированных данных с уровнем значимости  $P < 10^{-5}$ . Для того чтобы выявить факторы, обусловившие эту разницу, мы предприняли следующие шаги.

### Значение максимальной амплитуды и распределение длины сателлитных мономеров

Стоит заметить, что значение максимальной амплитуды, кроме корреляции с длиной (рис. 2, а, б), зависит также от профиля изгиба ДНК. Была поставлена цель – оценить черты этой статистики на основе нескольких характерных примеров.

Мы выбрали последовательности со значением амплитуды профиля изгиба ДНК, превышающую случайное значение с определенным порогом. Для этого мы провели симуляцию, аналогичную упоминавшейся выше, путем перемешивания исходной последовательности с фиксированным динуклеотидным составом. Получив 100 точек для случайных последовательностей, мы оценили среднее и дисперсию для каждого мономера. Реальное значение

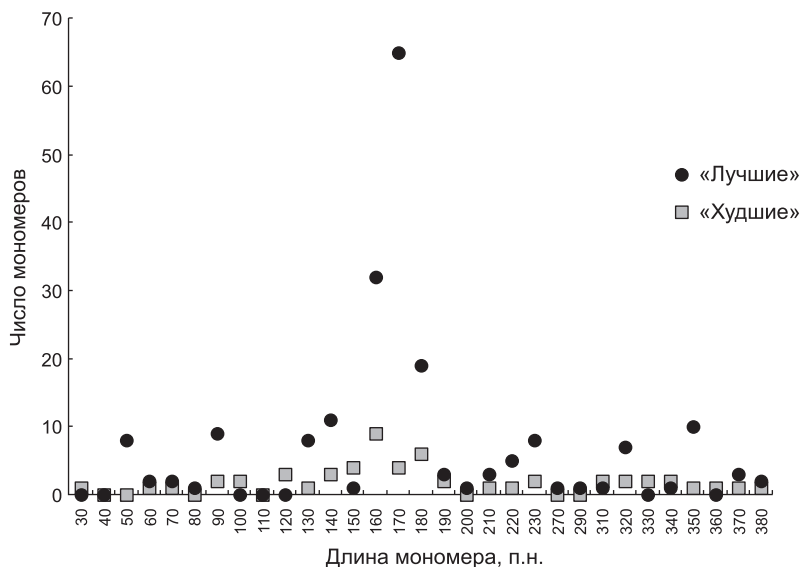
амплитуды было выше случайного на два стандартных отклонения в 103 случаях, на одно стандартное отклонение – для 218 мономеров. Мы назвали последние мономеры (218) «лучшими» (рис. 3). В то же время мономеры со значениями амплитуды меньше случайного среднего на одно стандартное отклонение и более мы назвали «худшими» (рис. 3).

Как можно видеть на рис. 3, популяция «лучших» мономеров имеет 2 пика: один – в районе 170 п.н., другой – в районе 350 п.н. Число «худших» мономеров оказалось чрезвычайно малым – 14 случаев.

Используя информацию о локализации мономеров, мы построили диаграмму (рис. 4), из которой можно сделать вывод о том, что высокие значения амплитуды профиля изгиба ДНК характеризуют последовательности центромерных и парацентромерных повторов, в то время как субтеломерные повторы практически не позиционируют нуклеосомы в целом. Эти выводы хорошо согласуются с различными моделями процесса сборки нуклеосом в зависимости от локализации (Kiyama, Trifonov, 2002; Valouev *et al.*, 2008).

### Симметрия профиля изгиба в ДНК мономеров благоприятна для гладкой стыковки

Для мономеров с высоким значением амплитуды мы наблюдали в основном профили в виде буквы «U». Поскольку во многих случаях



**Рис. 3.** «Лучшие» мономеры: их амплитуда превышает среднее на 1 станд. откл. «Худшие» мономеры: их значения амплитуды меньше случайного, по крайней мере, на 1 станд. откл.

период изгиба профиля ДНК равен длине мономера (рис. 2, 4), гладкая стыковка мономеров, возможно, обуславливает равное значение профиля на начале и конце мономера, что еще раз подтверждает неслучайный паттерн профиля изгиба ДНК на мономерах.

Наблюдалась также форма профиля, обратного к «U», но не так часто. Количество несимметричных мономеров с монотонно возрастающим или убывающим профилем было крайне мало. Мономеры с нерегулярным трендом профиля изгиба имели низкое значение амплитуды, согласно теории стохастического движения. Вероятно, такие мономеры не предполагают строгого позиционирования нуклеосом.

Стоит отметить, что максимальное значение профиля изгиба мономеров, показанное программой Phase, не превосходит в большинстве случаев случайное среднее значение, что свидетельствует, с одной стороны, об относительной природе способа позиционирования нуклеосом, а с другой – об ограничении максимальной энергии связывания ДНК с нуклеосомой для функциональности хроматина, как было указано выше.

### Спектральный анализ профиля изгиба ДНК мономеров

Мы провели анализ Фурье профилей изгиба ДНК для 901 мономера. Периодограммы, по-

лученные в результате анализа, изучали двумя способами: а) анализировалось распределение периодов, соответствующих наивысшей гармонике (гармоники с максимальным вкладом) (рис. 5, а); б) анализировалось распределение, соответствующее двум высшим гармоникам (рис. 5, б). Во втором случае выявили, что 50 % мономеров (436 из 901) имеют период, близкий к 170 п.н. Если добавить мономеры, высшая гармоника которых лежит в интервале периодов 20–90 п.н. (рис. 5, б), являющихся кратным значений 160–180 п.н., тогда можно считать, что большинство мономеров (763 из 901) имеют изгиб, подходящий для оптимального плотного позиционирования нуклеосом. Когда рассматривалась только наивысшая гармоника (рис. 5, а), в интервал периодов 150–190 п.н. укладывалось более 30 % мономеров (317 из 901). В то же время период высшей гармоники был примерно равен длине мономера в 60 % последовательностей (587 из 901). Остальные 40 % имели период, значительно отличающийся от длины мономера (рис. 6).

Даже при рассмотрении двух высших гармоник наблюдается повышение числа последовательностей с периодом в районе 320–340 п.н. (рис. 5, а, б). Возможно, это подтверждает гипотезу двустадийной сборки нуклеосом в процессе формирования нитей хроматина: на первом шаге нуклеосомы формируют разряженную регулярную сеть на ДНК, на втором шаге

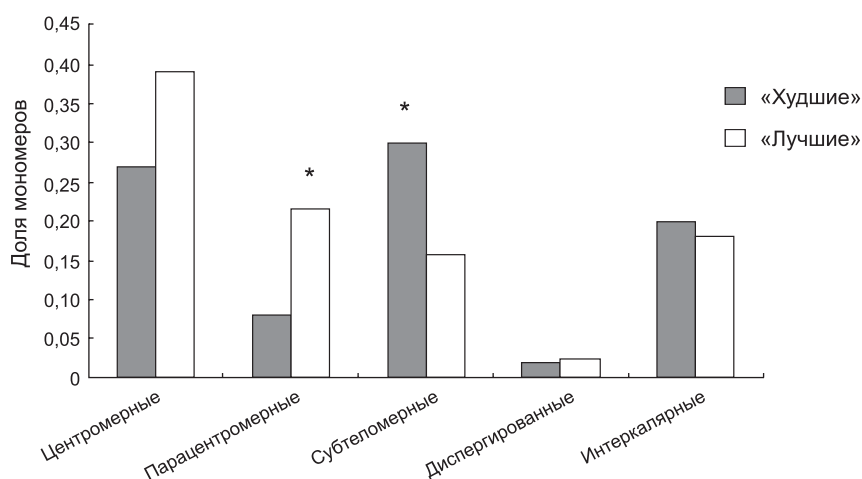
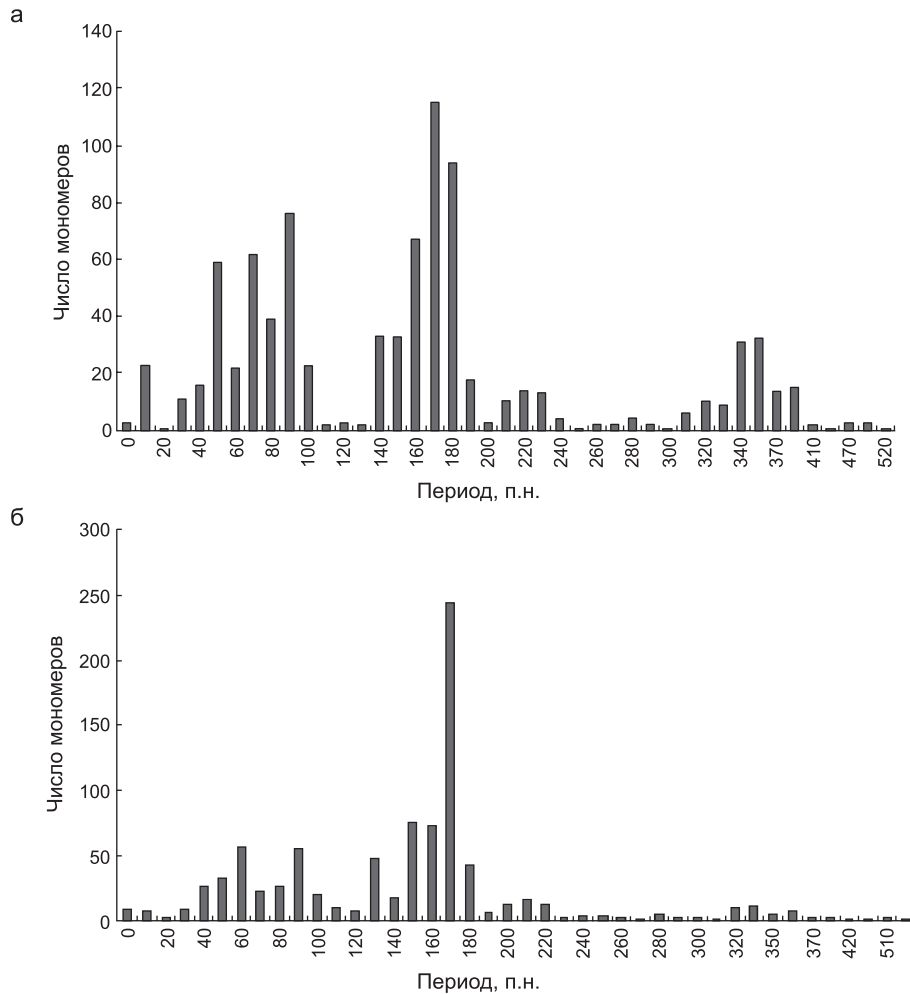


Рис. 4. «Лучшие» и «худшие» мономеры в различных локализациях.

\* Статистически значимая разница ( $P < 0,05$ ) между «лучшими» и «худшими» внутри класса локализации.



**Рис. 5.** Распределение мономеров в зависимости от периода, соответствующего высшей гармонике (а); двум высшим гармоникам (б).

нуклеосомы неспецифически заполняют пустые места для формирования плотно упакованного хроматина (Kiyama, Trifonov, 2002; Valouev *et al.*, 2008).

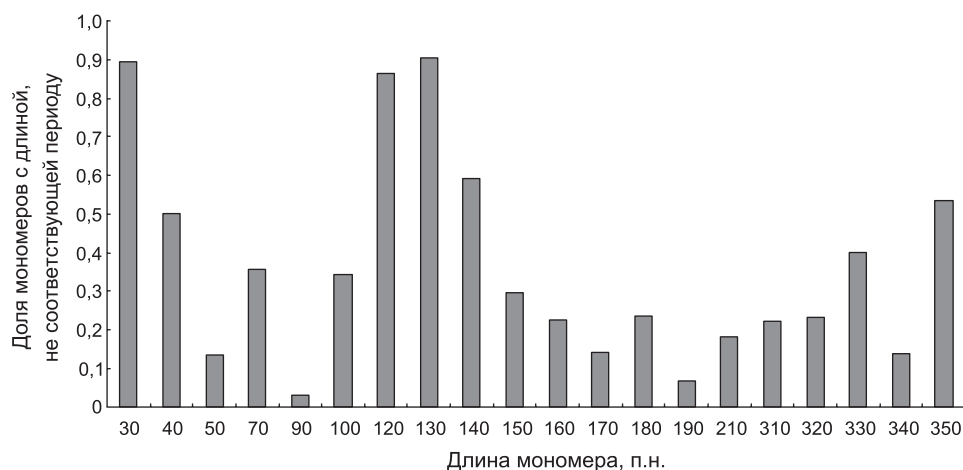
### Обсуждение

Мы дополнили начальное наблюдение неслучайного распределения длин мономеров (рис. 1) двумя относительно независимыми типами анализа на основе профиля изгиба ДНК: 1) путем введения статистики значения амплитуды; 2) дискретным преобразованием Фурье (рис. 5). На основании этих исследований можно заключить, что предпочтительная длина мономеров в 170 п.н. основана на специфике формирования изгиба ДНК, имеющее отношение к позиционированию нуклеосом. Более

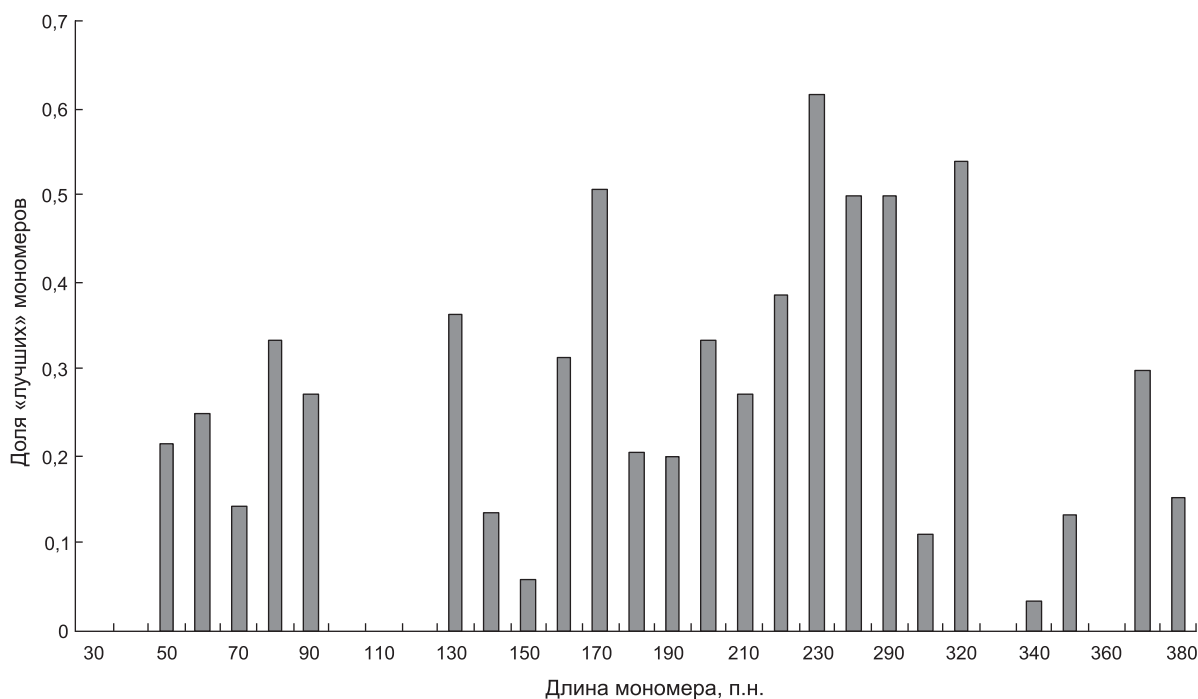
того, анализ Фурье подтвердил, что большое число мономеров имеет период, значительно отличающийся от длины (рис. 6).

В свою очередь, нуклеосомное позиционирование с периодом в 170 п.н. не является характеристикой всех сателлитных мономеров. В таких случаях возможно, что ДНК мономера не нуждается в позиционировании нуклеосом по ряду причин (в частности, от влияния положения сателлита на хромосоме – центромерный или теломерный).

На основании анализа значения амплитуды, можно констатировать, что максимальная амплитуда характерна для мономеров с длиной 170 п.н., но при этом даже в этом случае только 50 % таких мономеров превышают случайное значение на одно стандартное отклонение (рис. 7). Таким образом, «строгость» нуклео-



**Рис. 6.** Доля мономеров с отличием длины более чем на 30 п.н. от периода, соответствующего высшей гармонике.



**Рис. 7.** Доля «лучших» мономеров в каждой категории длины.

сомного позиционирования варьирует для разных длин мономеров. Возможно, что это обусловлено также как контекстом, так и локализацией мономеров.

В целом приведенные результаты подтверждают заключение, сделанное в работе E. Segal, J. Widom (2009) о том, что нуклеосомы позиционируются в основном путем исключения их посадки в определенных сайтах ДНК. Таким образом, можно предположить, что глубокие

«ямы» профиля изгиба ДНК – это основная черта, определяющая строгость нуклеосомного позиционирования на примере сателлитных повторов.

### Благодарности

Авторы выражают благодарности В.Г. Левицкому и А.В. Вершинину за критические замечания. Работа была частично поддержана

Госконтрактом № 07.514.11.4003. Вычисления произвелись в суперкомпьютерном центре ИВМиМГ СО РАН ([www.sccc.ru](http://www.sccc.ru)).

### Литература

- Deal R.B., Henikoff J.G., Henikoff S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones // *Science*. 2010. V. 328. P. 1161–1164.
- Cohanim A.B., Haran T.E. The coexistence of the nucleosome positioning code with the genetic code on eukaryotic genomes // *Nucl. Acids Res.* 2009. V. 37. No 19. P. 6466–6476.
- Feller W. *An Introduction to Probability Theory and its Applications*. V. 2. N.Y.: Wiley, 1971.
- Gupta S., Dennis J., Thurman R.E. *et al.* Predicting human nucleosome occupancy from primary sequence // *PLoS Comput. Biol.* 2008. 14:e1000134.
- Kaplan N., Moore I.K., Fondufe-Mittendorf Y. *et al.* The DNA-Encoded nucleosome organization of an eukaryotic genome // *Nature*. 2009. V. 458. P. 362–366.
- Kaplan N., Moore I., Fondufe-Mittendorf Y. *et al.* Nucleosome sequence preferences influence in vivo nucleosome organization // *Nat. Struct. Mol. Biol.* 2010. V. 17. No 8. P. 918–920.
- Karlin S., Altschul S.F. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes // *Proc. Natl Acad. Sci. USA*. 1990. V. 87. No 6. P. 2264–2268.
- Kiyama R., Trifonov E.N. What positions nucleosomes? – A model // *FEBS Lett.* 2002. V. 523. No 1/3. P. 7–11.
- Locke G., Tolkunov D., Moqtaderi Z. *et al.* High-throughput sequencing reveals a simple model of nucleosome energetic // *Proc. Natl Acad. Sci. USA*. 2010. V. 107. No 49. P. 20998–21003.
- Macas J., Meszaros T., Nouzova M. PlantSat: a specialized database for plant satellite repeats // *Bioinformatics*. 2002. V. 18. P. 28–35.
- Ringrose L. How do RNA sequence, DNA sequence, and chromatin properties regulate splicing? // *F1000 Biol. Rep.* 2010. 2. P. 74.
- Segal E., Widom J. What controls nucleosome positions? // *Trends Genet.* 2009. V. 25. P. 335–343.
- Schones D.E., Cui K., Cuddapah S. *et al.* Dynamic regulation of nucleosome positioning in the human genome // *Cell*. 2008. V. 132. No 5. P. 887–898.
- Spies N., Nielsen C.B., Padgett R.A., Burge C.B. Biased chromatin signatures around polyadenylation sites and exons // *Mol. Cell*. 2009. V. 36. No 2. P. 245–254.
- Sussman J.L., Trifonov E.N. Possibility of nonkinked packing of DNA in chromatin // *Proc. Natl Acad. Sci. USA*. 1978. V. 75. No 1. P. 103–107.
- Tanaka Y., Nakai K. An assessment of prediction algorithms for nucleosome positioning // *Genome Inform.* 2009. V. 23. No 1. P. 169–178.
- Trifonov E.N. Sequence-dependent deformational anisotropy of chromatin DNA // *Nucl. Acids Res.* 1980. V. 8. No 17. P. 165–171.
- Trifonov E.N. Cracking the chromatin code: Precise rule of nucleosome positioning // *Phys. Life Rev.* 2011. V. 8. No 1. P. 39–50.
- Valouev A., Ichikawa J., Tonthat T. *et al.* A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning // *Genome Res.* 2008. V. 18. No 7. P. 1051–1063.
- Valouev A., Johnson S.M., Boyd S.D. *et al.* Determinants of nucleosome organization in primary human cells // *Nature*. 2011. V. 474. No 7352. P. 516–522.
- Xi L., Fondufe-Mittendorf Y., Xia L. *et al.* Predicting nucleosome positioning using a duration Hidden Markov Model // *BMC Bioinformatics*. 2010. doi:10.1186/1471-2105-11-346.

## NUCLEOSOME ORGANIZATION IN PLANT DNA SATELLITES

V.N. Babenko<sup>1</sup>, K.O. Kutashov<sup>1</sup>, V.F. Matvienko<sup>2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: bob@bionet.nsc.ru;

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

### Summary

Satellite DNA, consisting of tandem repeats, forms heterochromatic regions packed into nucleosomes. Sequence and conformational DNA features responsible for the high efficiency of this package have been sought with a Fourier transform. The curvature profiles in tandem repeats differ significantly from a random model. As many as 50 % of the monomers have a periodicity of DNA bends in the vicinity of 170 bp, fitting the wrapping length of a single nucleosome within two highest-order harmonics in the DNA curvature profile. The nucleosome positioning scheme in subtelomeric satellites in plants differs from that in centromeric satellites.

**Key words:** nucleosome, positioning, Fourier transform, satellite repeats.