



# **Структурно-функциональная организация генов и геномов и коды нуклеосомной организации геномной ДНК**

*к.б.н. В.Г.Левецкий*

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia



# СОДЕРЖАНИЕ



1. Структурно-функциональная организация геномов эукариот
2. Роль нуклеосомной организации ДНК в регуляции транскрипции генов
3. Общие характеристики нуклеосомного кода укладки хроматина
4. Подходы методы компьютерного анализа нуклеосомной организации ДНК
5. Анализ особенностей нуклеосомной организации различных структурно-функциональных районов геномной ДНК



# **Структурно-функциональная организация геномов эукариот**



# Размеры геномов в зависимости от таксономической принадлежности организмов



¶ Таксон□	¶ Вид□	Размер генома, *10 <sup>6</sup> п.о.□	
Прокариоты, эубактерии□	<i>Bacillus subtilis</i> □	4.2□	
	<i>Mycoplasma genitalium</i> □	0.6□	
	<i>Escherichia coli</i> □	4.6□	
Архебактерии□	<i>Archaeoglobus fulgidus</i> □	2.2□	
	<i>Methanococcus jannaschii</i> □	1.7□	
	<i>Pyrococcus horikoshii</i> □	1.7□	
Эукариоты.*□	Грибы□	<i>Aspergillus nidulans</i> □	13□
		<i>Saccharomyces cerevisiae</i> (дрожжи)□	4□
	Беспозвоночные□	<i>Caenorhabditis elegans</i> ¶ (нематода)□	100□
		<i>Drosophila melanogaster</i> ¶ (плодовая мушка)□	165□
	Растения□	<i>Arabidopsis thaliana</i> ¶ (кресс-салат)□	130.**□
		<i>Zea mays</i> (кукуруза)□	2290□
		<i>Secale cereale</i> (рожь)□	9000□
	Позвоночные□	<i>Homo sapiens</i> (человек)□	3400□
		<i>Mus musculus</i> (мышь)□	3454□
		<i>Gallus gallus</i> (цыплёнок)□	1200□
		<i>Rattus norvegicus</i> (крыса)□	2900□

\* – для эукариот приведён гаплоидный размер генома;

\*\* – один из самых малых среди геномов растений: ~ в 200 раз меньше, чем большинство других геномов растений.



Гены эукариот отделены друг от друга районами нетранскрибируемой ДНК (межгенными спейсерами).

У прокариот гены также отделены друг от друга спейсерами, однако длина спейсеров эукариот существенно больше. Низкая плотность кодирующих районов – общее свойство геномов эукариот, например, для человека она составляет около 2%



### Кластер глобиновых генов человека.

Общая длина ДНК кластера бета-глобиновых генов человека составляет 73308 п.о. Из них:

2.8% приходится на экзоны,

6.1% – на интроны,

8.6% – на различные типы повторяющихся последовательностей.



# Повторяющиеся последовательности геномов эукариот



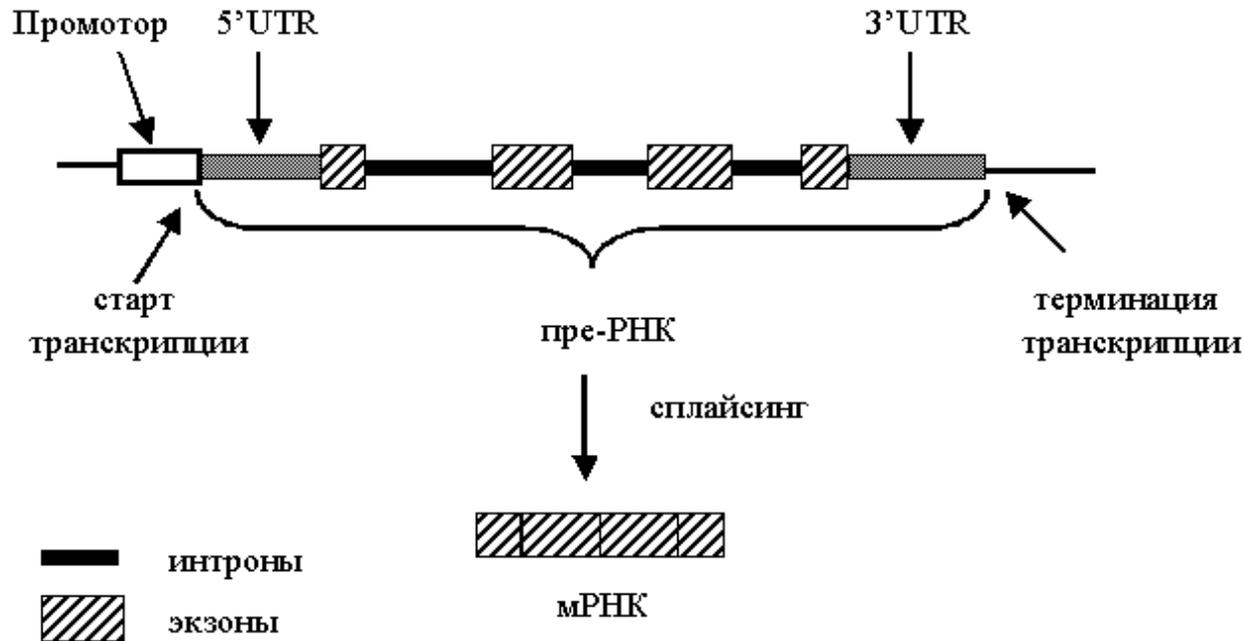
Все повторы в геноме можно разделить на два класса:

1. Тандемные повторы, к которым относятся разные виды сателлитной ДНК, гены рРНК.
2. Диспергированные повторы, распределённые в геноме по принципу чередования с уникальными последовательностями. К этому классу относятся, в частности, различные типы перемещающихся (мобильных) элементов

Для приматов, например, характерны повторы семейства Alu, их размер составляет около 300 п.о. На долю семейства повторов Alu приходится около 13% геномной ДНК человека.



# Общая структура гена эукариот



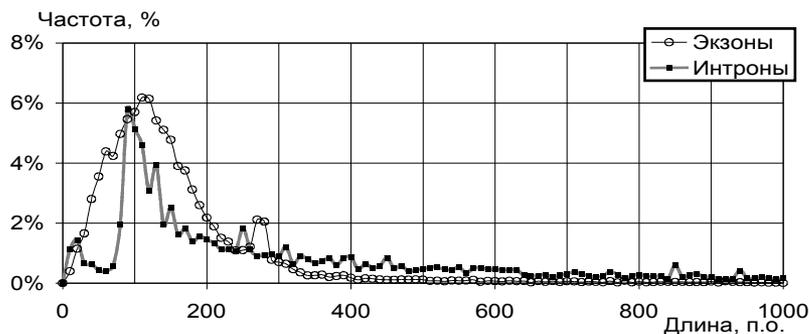
Строение генов эукариот, в отличие от прокариот, характеризуется наличием экзон-интронной структуры. В состав первичного транскрипта – пре-мРНК входят как экзоны, так и интроны (некодирующие районы). В процессе сплайсинга интроны вырезаются из пре-мРНК. Оставшиеся же части – экзоны – объединяются в зрелую матричную РНК (мРНК), которая может транслироваться в белок.



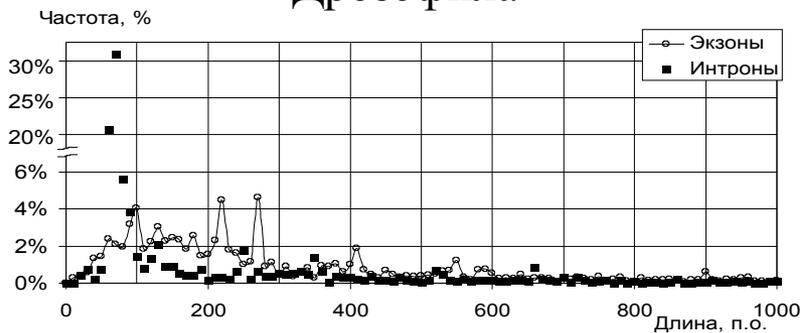
В среднем на 1000 п.о. кодирующей ДНК у позвоночных приходится 5.6 интронов. Число экзонов в одном гене может быть даже больше сотни: например, ген титина человека имеет не менее 157 экзонов, так что их общая длина превышает 28347 п.о.

### Распределения по длинам интронов и экзонов

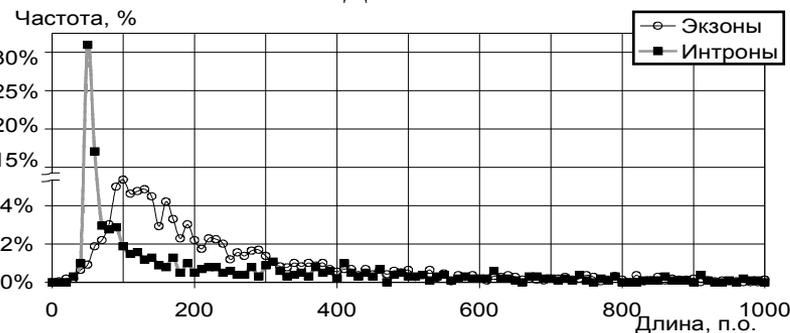
Человек



Дрозофила

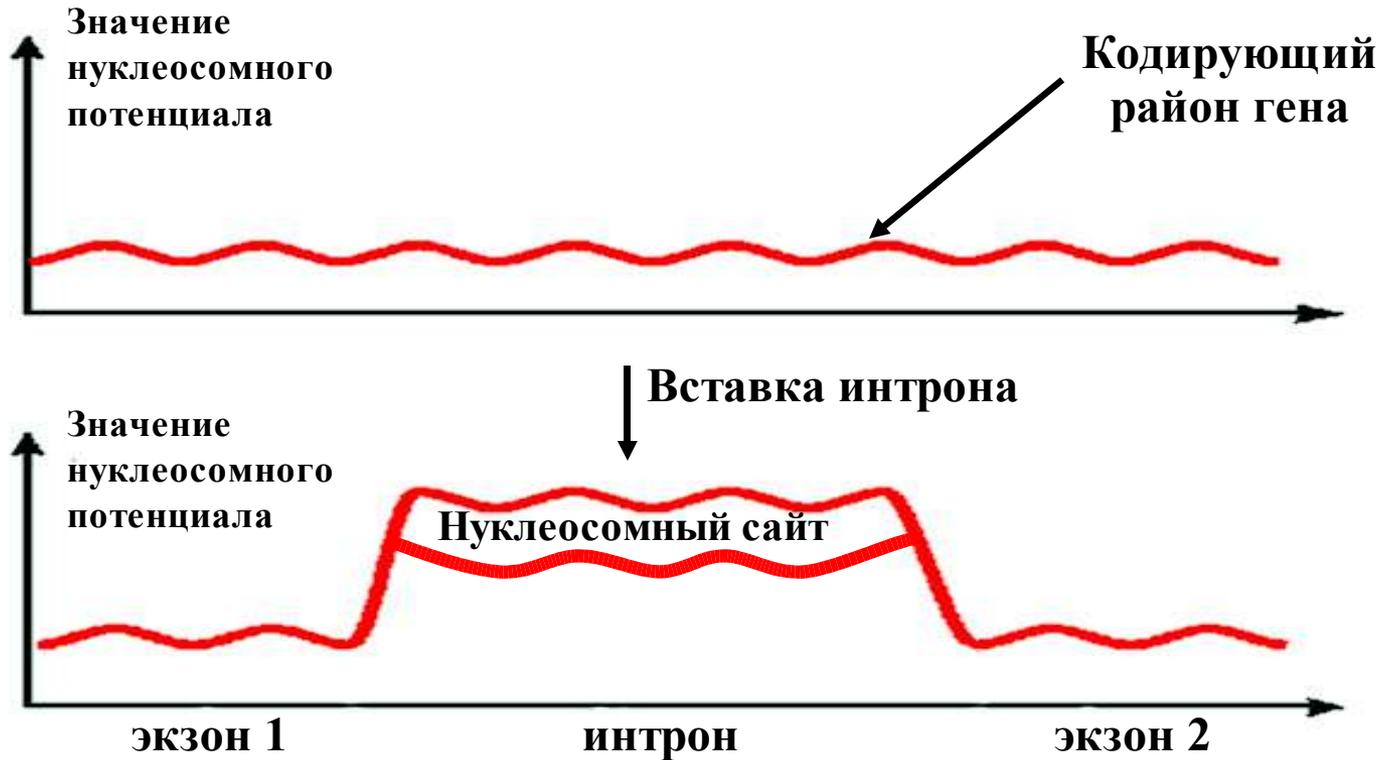


Нематода





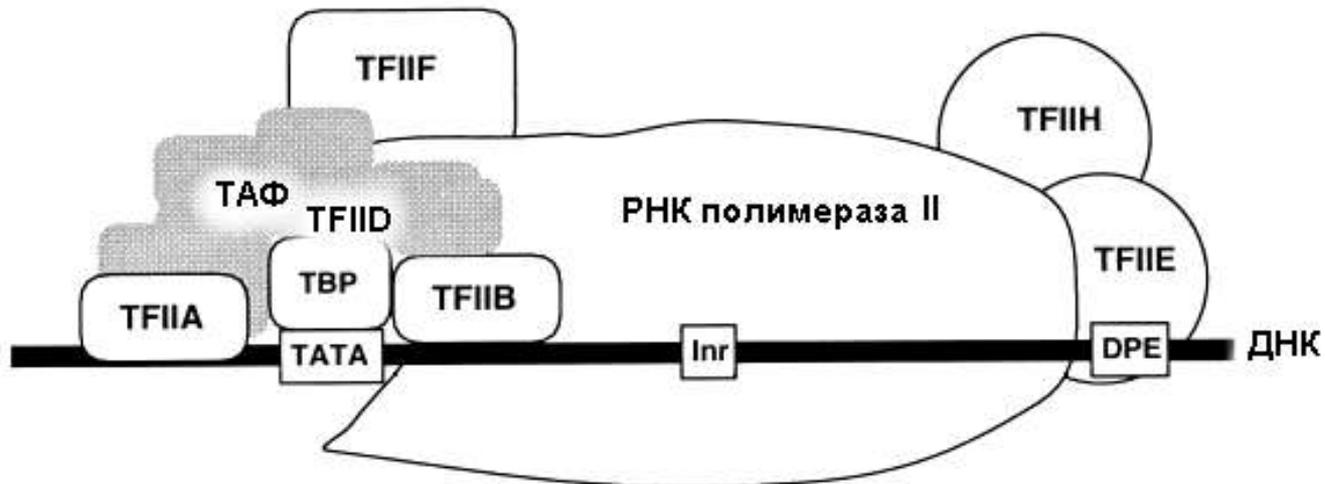
# Вставка интрона в кодирующий район гена позволяет обеспечить присутствие нуклеосомного сайта



Экзон-интронная структура генов эукариот может быть обусловлена нуклеосомной организацией хроматина и связанными с ней особенностями регуляции экспрессии генов.



# **Роль нуклеосомной организации ДНК в регуляции транскрипции генов**



Структура корового промотора генов эукариот, транскрибируемых РНК полимеразой II. TFIIA, TFIIB, TFIID = TBP + ТАФ (ТАТА-бокс связывающий белок и ассоциированные факторы), TFIIЕ, TFIIF, TFIIH - основные белки базового транскрипционного комплекса.

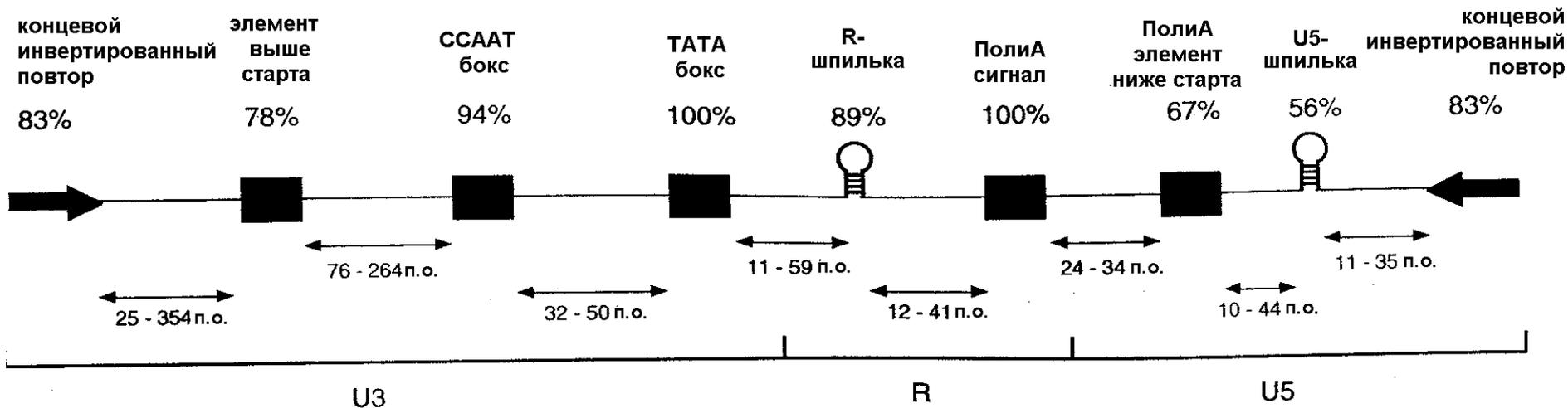
Элементы корового промотора:

- (1) ТАТА-бокс (около  $-30$  по отношению к старту транскрипции), инициатор (Inr-элемент, перекрывает точку старта),
- (2) DPE (downstream promoter element) – позиция около  $+30$ .

Присутствие каждого из перечисленных элементов в конкретном промоторе не является обязательным.

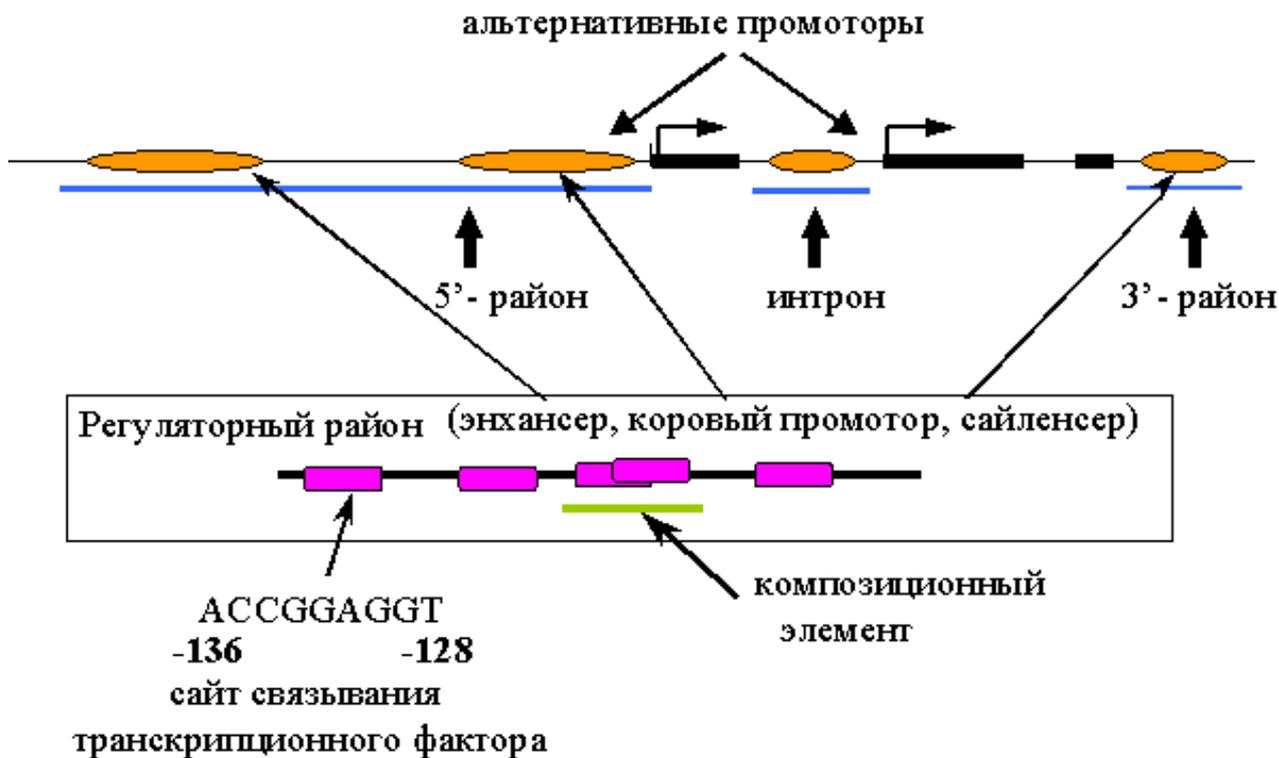


Общая длина - от 330 до 644 п.о.



Модель промоторного района, построенная пакетом программ ModelGenerator/ModelInspector для распознавания промоторов, расположенных в длинных концевых повторах (LTR)

На схеме приведены расстояния между отдельными элементами и частоты элементов (в % от числа содержащих их последовательностей).



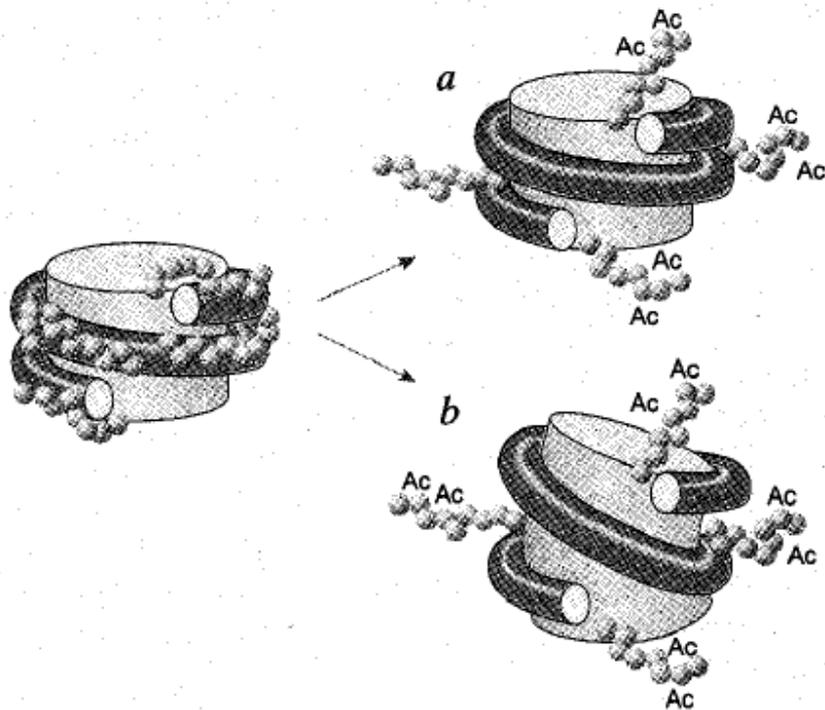
Тканеспецифическая экспрессия эукариот достигается за счёт существенно более сложного по сравнению с прокариотами строения генов: (1) наличия экзон-интронной структуры, (2) более сложного блочно-иерархического устройства регуляторных районов генов.



## Модификации аминокислотных остатков гистонов, влияющие на конформацию нуклеосом



Существуют посттрансляционные модификации аминокислотных остатков гистонов, оказывающие сильное воздействие на транскрипционную активность генов. Один из возможных механизмов их действия состоит в изменении величины положительного заряда гистонов, что влияет на прочность связи гистонов с отрицательно заряженным остовом ДНК.

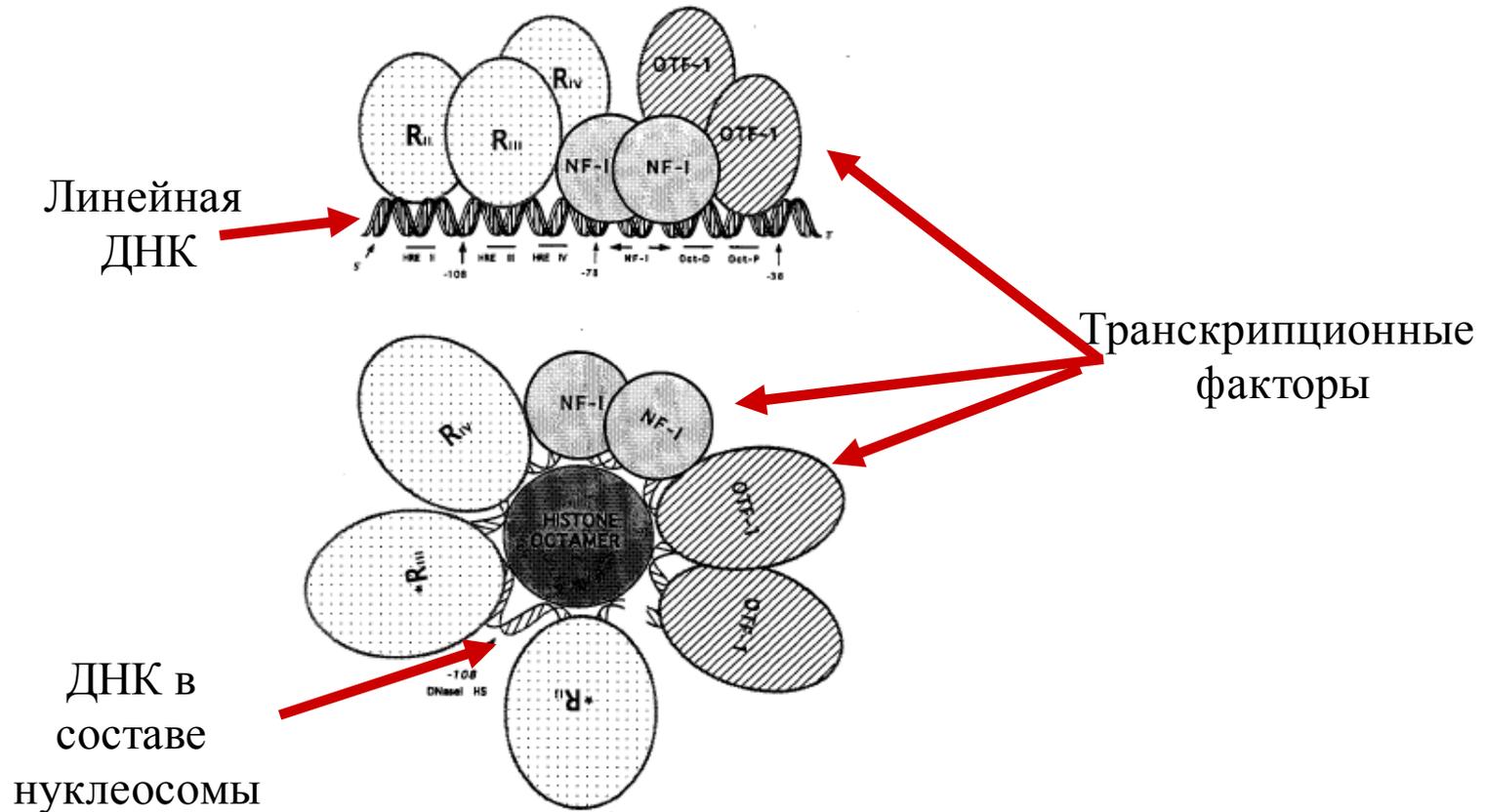


Таким образом можно объяснять эффекты ацетилирования и фосфорилирования. При этом меняется распределение заряда в концевых доменах гистонов: **ацетилирование** нейтрализует положительный заряд лизина, а **фосфорилирование** вносит отрицательный заряд в серин.



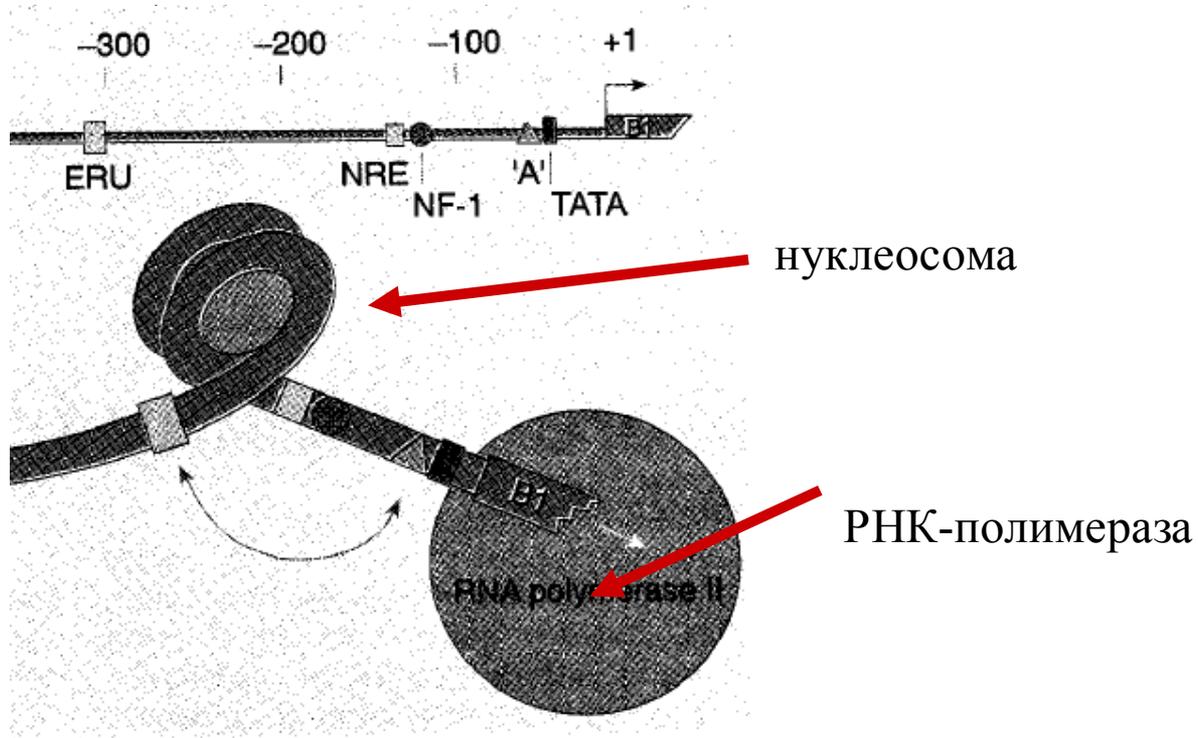
# Нуклеосомная упаковка ДНК в промоторном районе гена

Формирование нуклеосомы в промоторной ДНК может способствовать правильному стерическому взаимодействию транскрипционных факторов.





# Нуклеосомная упаковка ДНК в промоторном районе гена



Упаковка промоторной ДНК в нуклеосому может приводить к пространственному сближению удалённых сайтов связывания транскрипционных факторов, способствующему формированию транскрипционного комплекса



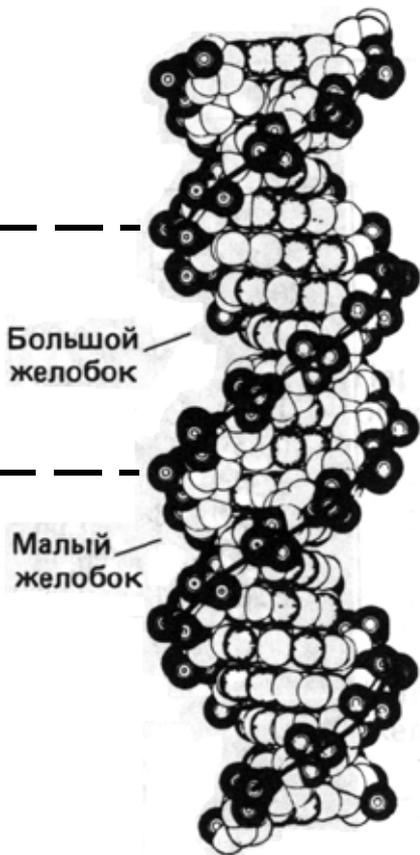
# **Общие характеристики нуклеосомного кода укладки хроматина**



# Классификация форм ДНК: В-, А- и Z-формы.



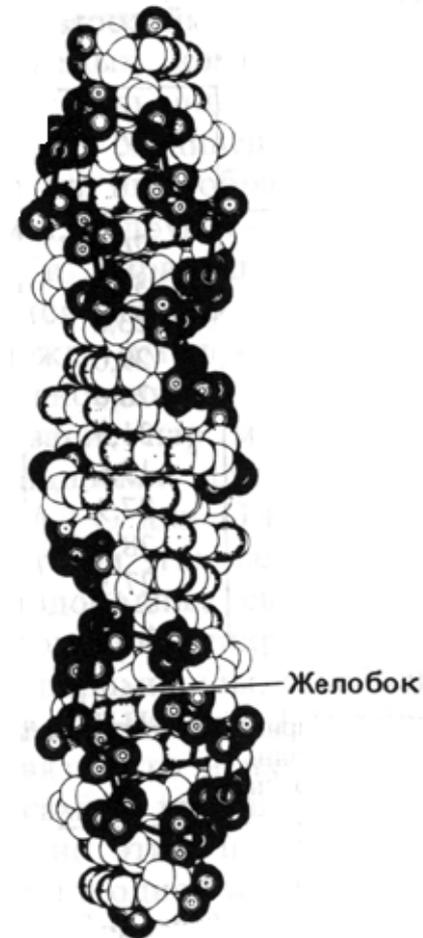
10.3÷10.6  
пар  
оснований  
на виток



В-ДНК



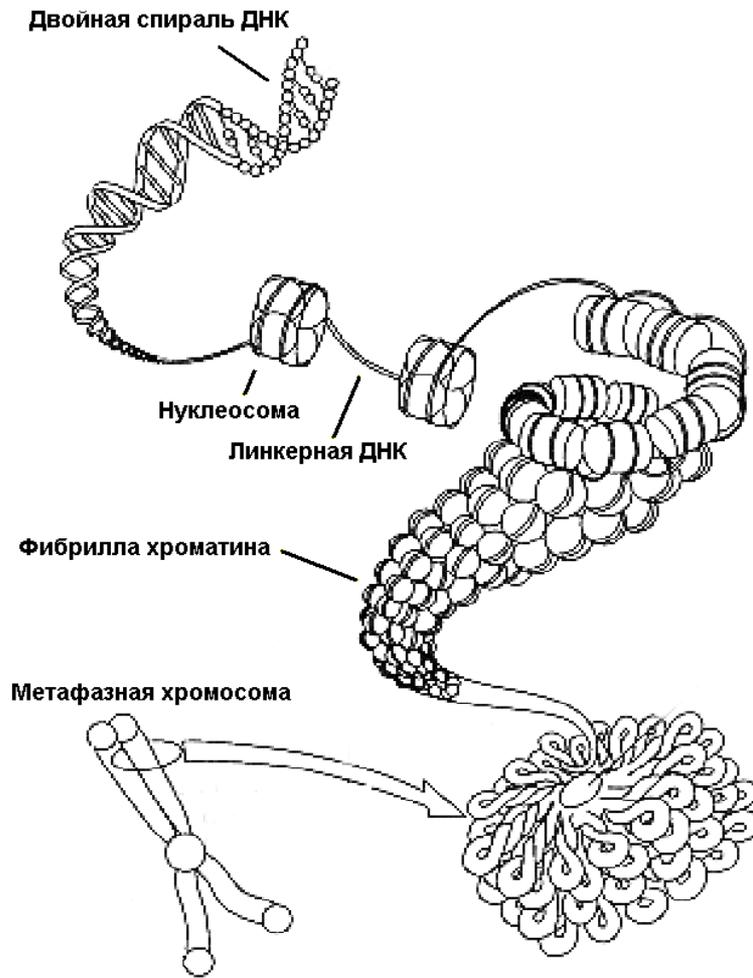
А-ДНК



Z-ДНК



# Иерархия уровней упаковки ДНК в хромосомы



Упаковка ДНК в ядре дала эукариотическим организмам возможность существенного увеличения размера генома. Это обеспечило огромную информационную емкость геномов эукариот и возможность кодирования сложных паттернов регуляции экспрессии генов. Последнее особенно необходимо для обеспечения тканеспецифичной экспрессии генов в многоклеточных организмах эукариот

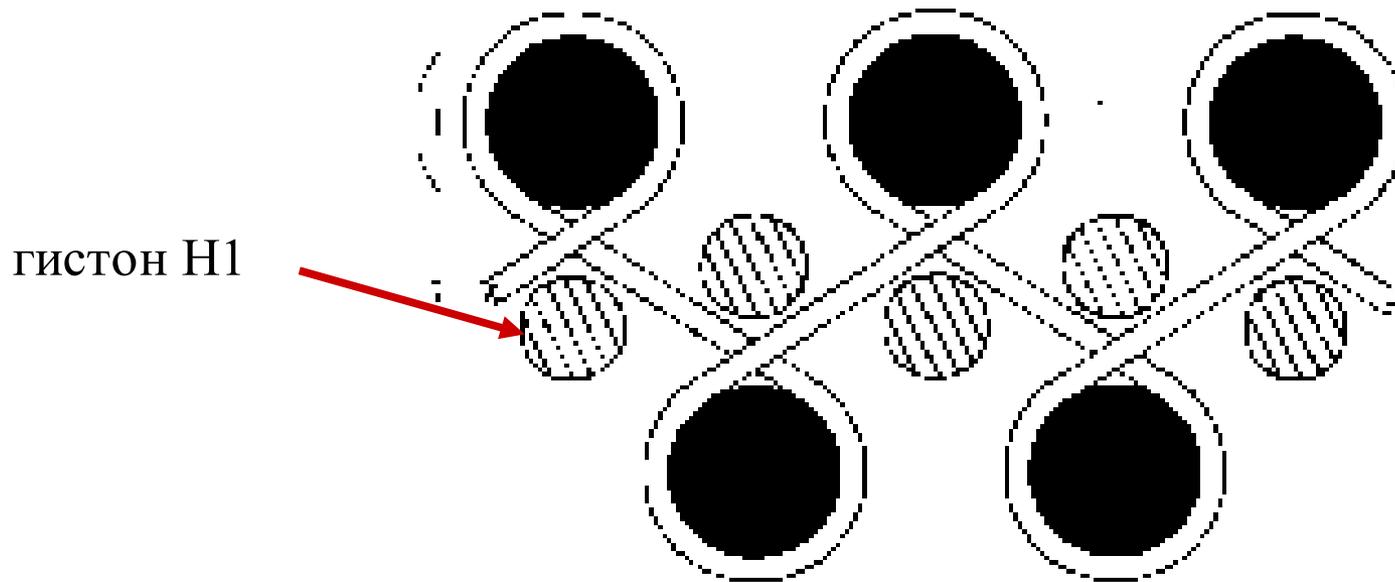




# Структурная схема нитей нуклеосом (“бусин на нити”) – нуклеопротеиновых структур хроматина:



гистон H1 отсутствует – прямолинейная форма.

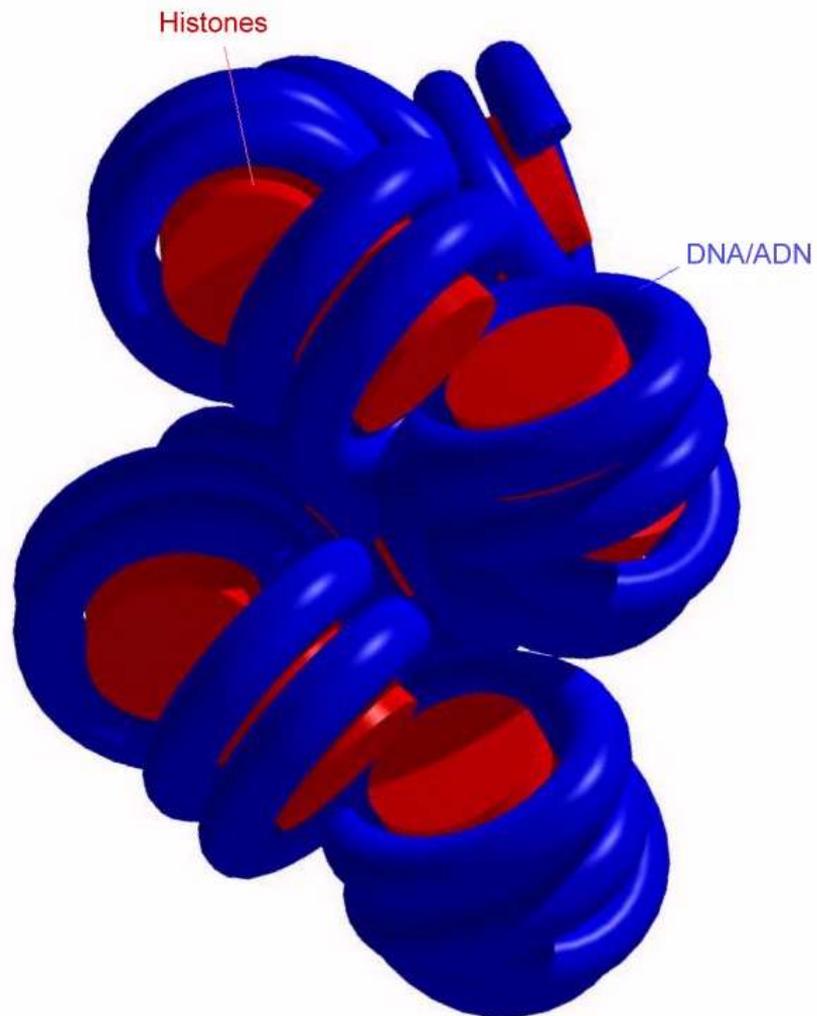


гистон H1

гистон H1 присутствует – зигзагообразная форма



# Схема упаковки нуклеосом в хроматиновые нити





## Контекстный код упаковки ДНК в хроматин



1. Отсутствие специфических последовательностей в фиксированных местах.
2. Сигналы распределены по множеству альтернативных функционально значимых позиций.

Отмечаются фазированное расположение некоторых динуклеотидов (АА, ТТ... ) с периодом  $\sim 10.5$  пар оснований.

Нуклеотидный контекст нуклеосомной ДНК способствует изгибам, необходимым для её накручивания ДНК на гистоновый октамер.



# Классификация типов неслучайного расположения нуклеосом на ДНК



Структура нуклеосомной ДНК является периодичной, так что перемещение двойной спирали ДНК на расстояние, кратное одному её витку, может не приводить к существенному изменению состояния локальных контактов ДНК с гистонами. По этой причине разделяют трансляционно- и ротационно-определенное позиционирование нуклеосом.

**Трансляционно-определённое** позиционирование относится к точному расположению нуклеосомы в каком-либо участке ДНК.

**При ротационно-определённом** позиционировании нуклеосома может находиться в альтернативных позициях, отстоящих друг от друга на один виток спирали ДНК (около 10 п.о.).



# **Механизмы определяющие расположение нуклеосом в геномной ДНК**



- 1. Позиционирование, определяемое положительными контекстными характеристиками ДНК. Экспериментально показана определяющая роль контекста ДНК для предпочтительного формирования нуклеосом в определённых позициях нуклеотидных последовательностей.**
- 2. Позиционирование, определяемое отрицательными контекстными характеристиками ДНК. Неслучайное расположение нуклеосомы на ДНК может также направляться фрагментами ДНК, которые по своим конформационным свойствам не могут быть объединены в нуклеосому. Присутствие таких фрагментов в геноме может обеспечивать его контекстно-конформационную разметку.**
- 3. Позиционирование, определяемое взаимодействием с белками. При анализе структуры хроматина иногда оказывается, что позиционирование нуклеосом направляется негистоновыми белками, сайты связывания которых расположены рядом с сайтами посадки нуклеосом**
- 4. Фазирование нуклеосом – формирование упорядоченных массивов нуклеосом, определяемое контекстными особенностями ДНК**



# **Подходы методы компьютерного анализа нуклеосомной организации ДНК**



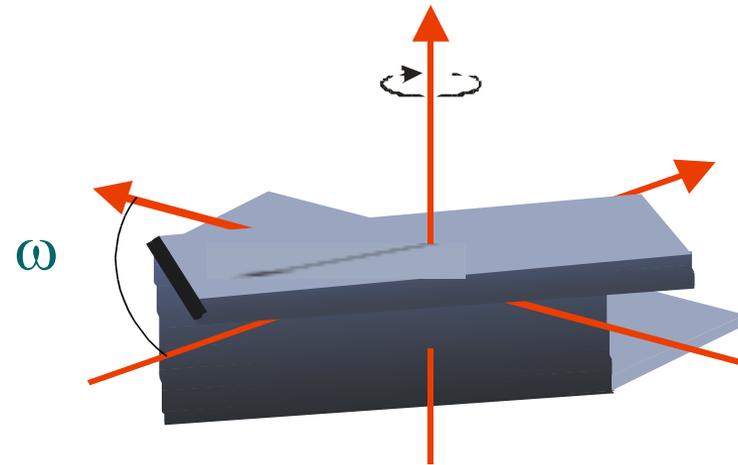
# Метод конформационных параметров ДНК



Динуклеотидный код угла спирального вращения ДНК (угол твист)

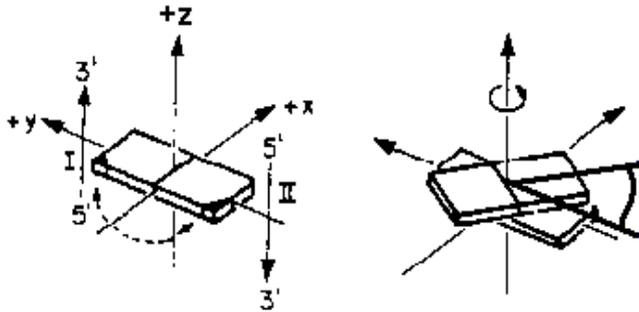
Угол спирального вращения твист ( $\omega$ )

Динуклеотид	Угол твист (градусы)
AA	38.90
AT	33.81
AG	32.15
AC	31.12
TA	33.28
TT	38.90
TG	41.41
TC	41.31
GA	41.31
GT	31.12
GG	34.96
GC	38.50
CA	41.41
CT	32.15
CG	32.91
CC	34.96

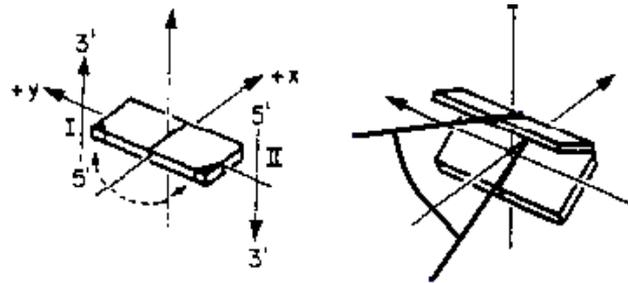




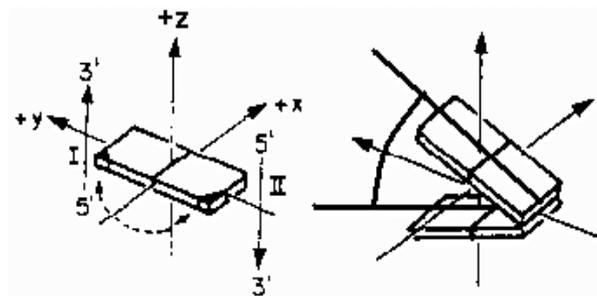
# Примеры конформационных свойств ДНК



А) Конформационное свойство ДНК - **угол твист** - угол спирального вращения между длинными осями симметрии соседних пар оснований.



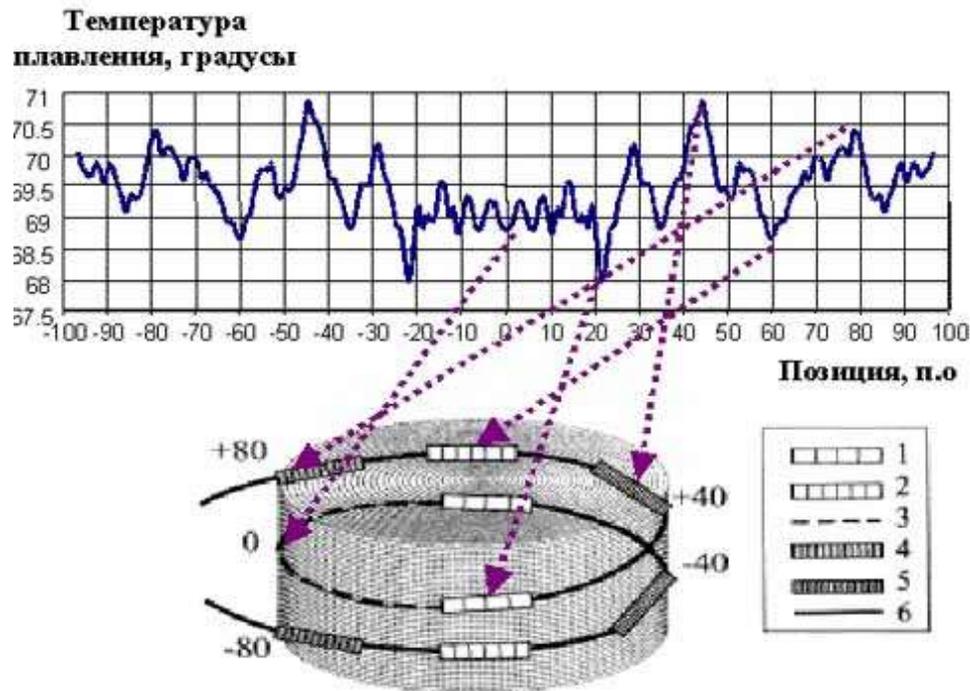
Б) Конформационное свойство ДНК - **угол ролл** - угол раскрытия плоскостей соседних пар оснований (ось двугранного угла параллельна длинной оси симметрии).



В) Конформационное свойство ДНК - **угол тилт** - угол раскрытия плоскостей соседних пар оснований (ось двугранного угла параллельна короткой оси симметрии).



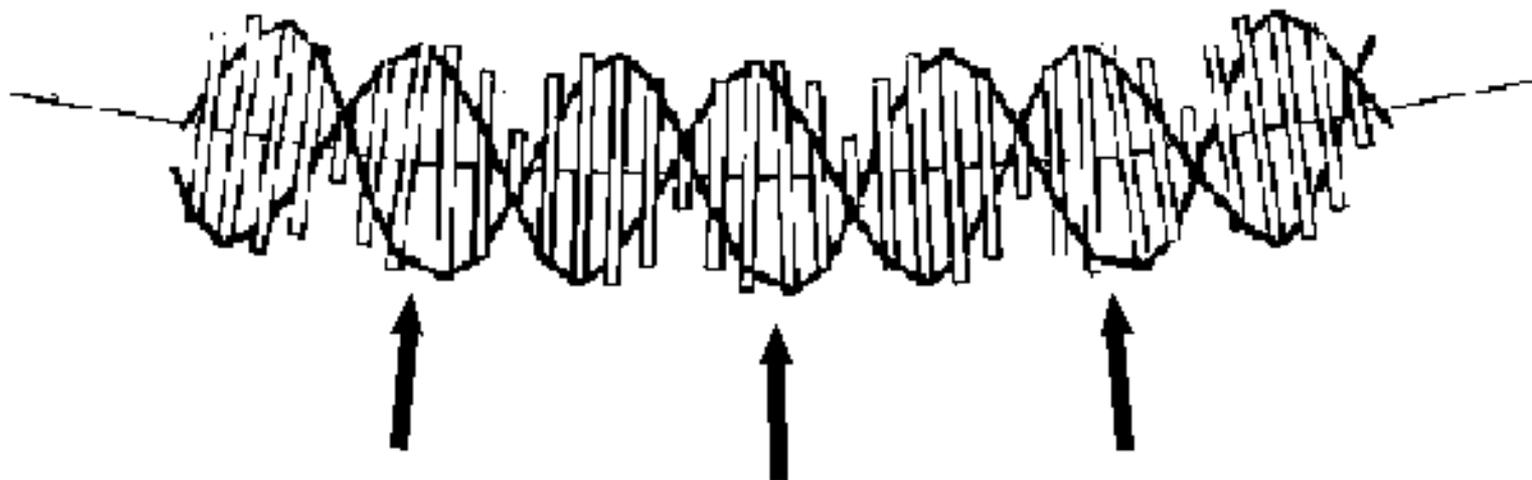
# ПРОФИЛЬ ТЕМПЕРАТУРЫ ПЛАВЛЕНИЯ НУКЛЕОСОМНОЙ ДНК(РАЗМЕР ОКНА 7 П.О.) СХЕМА ПРОСТРАНСТВЕННОГО РАСПОЛОЖЕНИЯ НА ПОВЕРХНОСТИ ГИСТОНООВОГО КТАМЕРА РАЙОНОВ, РАЗЛИЧАЮЩИХСЯ ПО ЭНЕРГИИ ЖЁСТКОСТИ ИЗГИБА ДНК



- |                                  |                          |
|----------------------------------|--------------------------|
| 1 - наиболее гибкие районы       | [-25; -17] и [+17; +25]; |
| 2 - гибкие районы                | [-64; -56] и [+56; +64]; |
| 3 - область повышенной гибкости  | [-25; +25];              |
| 4 - наиболее жёсткие районы      | [-48; -40] и [+40; +48]; |
| 5 - жёсткие районы               | [-80; -72] и [+72; +80]; |
| 6 - область повышенной жёсткости | [-96; -25] и [+25; +96]. |



Схематическая иллюстрация однонаправленного изгиба оси спирали ДНК под влиянием периодических вставок двух пар оснований, плоскости которых непараллельны друг другу (вставки отмечены стрелками)



Фазирование подобных изгибов с периодом, равным одному витку спирали ДНК, может облегчать формирование нуклеосом и упаковку ДНК в хроматин.



# Анализ автокорреляционных функций динуклеотидов для последовательностей эукариот (А), прокариот (В) и случайных (С).

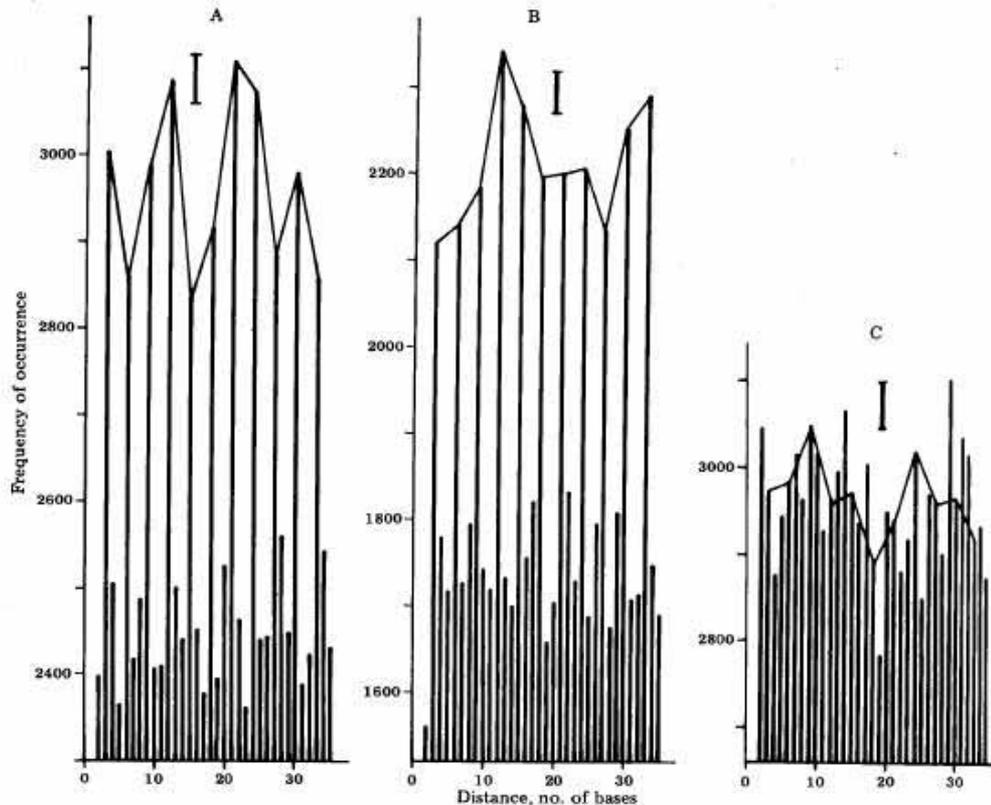
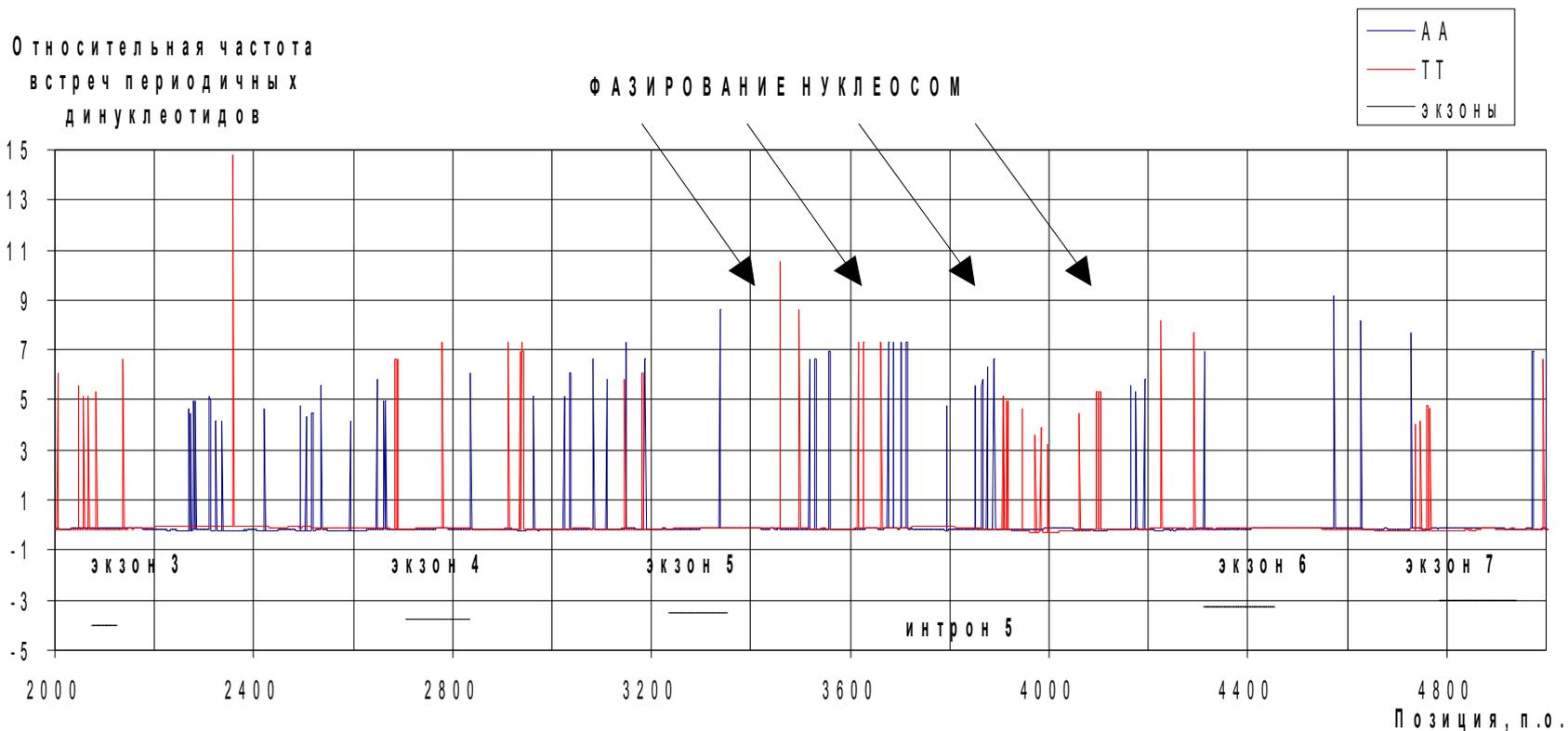


FIG. 3. Sum of the autocorrelation functions for all 16 dinucleotides, calculated for chromatin sequences (A), nonchromatin prokary sequences (B), and a random sequence (C). Envelope curves (as in Fig. 2) pass through the points that correspond to multiples of three-b distances. The Y axis origins are omitted for clarity. Because the autocorrelation functions for dinucleotides of the chromatin sequences has a slowly decaying component (e.g., as in Fig. 2B), this component was subtracted to make the oscillating part clearer. This was done by subtracting the deviation of each ordinate of the original distribution from the "running average" of seven adjacent points. The subtraction also was made for the cases of nonchromatin and random sequences for purposes of uniformity. The subtraction also was made for the cases of nonchromatin and random sequences for purposes of uniformity. The vertical bar in each graph corresponds to estimated standard deviation. The chromatin sequences used for the analysis are listed in Table 1; their total length is 36,000 bases. The nonchromatin prokary sequences are listed in the text; their total length is 30,000 bases. The computer-generated random sequence has the same dinucleotide composition as SV40 DNA; the total length is 42,000 bases.

По оси Y - суммы автокорреляционных функций  $[f_{XY}(i) \cdot f_{XY}(i+T)]$  по всем 16 динуклеотидам XY, по позициям (i) последовательностей. По оси X - период T.  $f_{XY}(i) = 0$  или 1, вероятность обнаружения в i-той позиции динуклеотида XY



# Фазирование нуклеосом в 5 интроне гена овальбумина цыплёнка



Lauderdale JD, Stein A.

Introns of the chicken ovalbumin gene promote nucleosome alignment in vitro.

Nucleic Acids Res 1992 Dec 25;20(24):6589-96



# 15-буквенный нуклеотидный код



Обозначение буквенного кода	15-	Сопоставляемая группа нуклеотидов	Пояснение обозначения кода
A		A	
T		T	
G		G	
C		C	
W		A,T	<b>WEAK</b>
R		A,G	<b>PURINE</b>
M		A,C	<b>AMINO-</b> (положительный заряд)
K		T,G	<b>KETO-</b> (отрицательный заряд)
Y		T,C	<b>PYRIMIDINE</b>
S		G,C	<b>STRONG</b>
B		T,G,C	не <b>A</b>
V		A,G,C	не <b>T</b>
H		A,T,C	не <b>G</b>
D		A,T,G	не <b>C</b>
N		A,T,G,C	Любой



## Поиск сайтов формирования нуклеосом с помощью анализа периодичности расположения тринуклеотидов VWG



Метод распознаёт сайты предпочтительного формирования нуклеосом по повышенной частоте встреч (период около 10 п.о.) тринуклеотидов вида VWG (неТ, А/Т, G). Выбор этого вырожденного тринуклеотида вызван тем, что фазированное его расположение было обнаружено в экзонах и интронах человека.

В ходе расчёта фазированности для каждого триплета VWG сначала рассматривается окно длины 100 п.о. ( $\pm 50$  п.о. от позиции центра окна). Число  $P$  фазированно расположенных триплетов VWG в окне вычислялась по формуле:  $P=M/N$ .  
Здесь  $M$  – расстояние между двумя триплетами VWG в п.о.,  
 $N$  – целое число от 1 до 5, соответствующее числу витков двойной спирали ДНК между этими триплетами.  
Затем выбраны допустимые нижний  $P_L$  и верхний  $P_H$  пределы изменения числа  $P$  фазированных триплетов VWG. Например,  $P_L = 10.0$ ,  $P_H = 10.33$ .  
Поэтому фазированно относительно центра окна могут быть расположены триплеты только в позициях  $\pm 10$ ,  $\pm 20$ ,  $\pm 30$ ,  $\pm 31$ ,  $\pm 40$ ,  $\pm 41$ ,  $\pm 50$ ,  $\pm 51$ .



# Метод множественного выравнивания



выравнивание:

```
agacaatGACTCAGCagaaat  
tccccgTACTCAGCgctttg  
caaccaTGACACAGCataaaa
```

1. Наиболее весомый вклад в контекстный код позиционирования нуклеосом вносит фазированное расположение динуклеотидов АА и ТТ. Для этих динуклеотидов выявлена периодичность  $10.3 \pm 0.2$  п.о.
2. В центральной области нуклеосомного сайта упорядоченного расположения динуклеотидов АА и ТТ не обнаружено.
3. Рассмотрение распределений по другим динуклеотидам не является столь информативным, как по динуклеотидам АА и ТТ, из оставшихся динуклеотидов, наиболее значимый вклад в неслучайное расположение нуклеосом вносят динуклеотиды СС и GG.



# Анализ лингвистической сложности ДНК



Сложность генетического текста является мерой числа слов определённой длины (олигонуклеотидов), необходимых для чтения этого текста. Количественно сложность определяется как отношение реального размера словаря, требуемого для описания данной последовательности к максимальному размеру словаря, необходимого для описания случайных последовательностей такой же длины. В качестве примера рассмотрим слова длины 2 (динуклеотиды) в 2 следующих последовательностях ДНК: (1) ACGGTAAGCTGATTCCA, (2) ACACACACACACACA. Для (1) сложность равна  $16/16 = 1$ , а для (2) -  $2/16 = 0.125$ .

Общая сложность последовательности определяется как произведение сложностей, рассчитанных при определённых длинах олигонуклеотидов.

Показано, что последовательности нуклеосомной ДНК обладают пониженными значениями сложности.

Общим свойством последовательностей с низкой сложностью является более выраженная периодичность динуклеотидов, а также возможность более легкого множественного выравнивания



# Нуклеосомный потенциал $\phi(X)$ , рассчитанный с помощью дискриминантного анализа динуклеотидным частотам $\{f(X)\}$

Две выборки последовательностей ДНК:

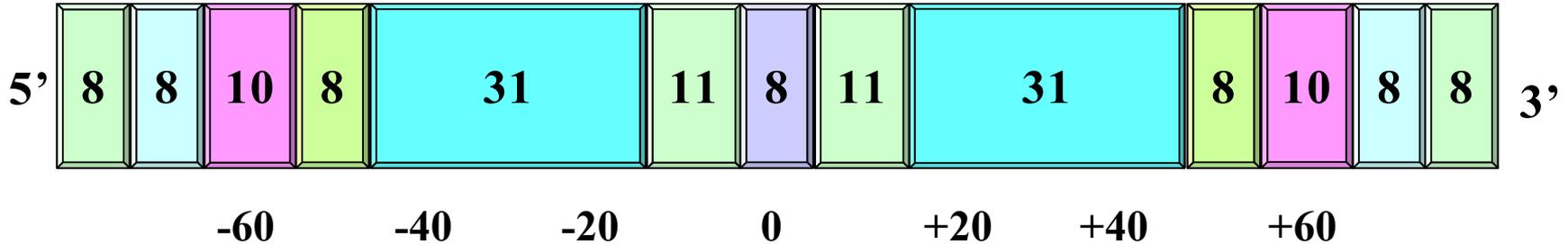
1. Нуклеосомные сайты

2. Случайные последовательности

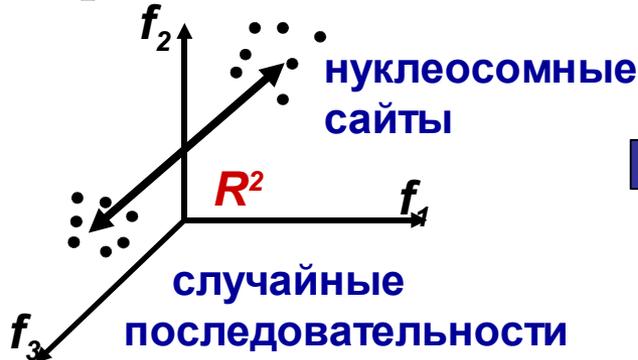
1) С А А С Т Г С С А С {  $f_{1,j}^{(1)}$  }  
 .....  
 М) С А Г Т Г Г Т Т А А {  $f_{M,j}^{(1)}$  }

1) Т Г С А С А Г С С С {  $f_{1,j}^{(2)}$  }  
 .....  
 М) Г Т Г Г С С Т С А А {  $f_{M,j}^{(2)}$  }

## Разбиение нуклеосомного сайта на части



## Дискриминантный анализ



Дискриминирующая функция – нуклеосомный потенциал  $\phi(X)$

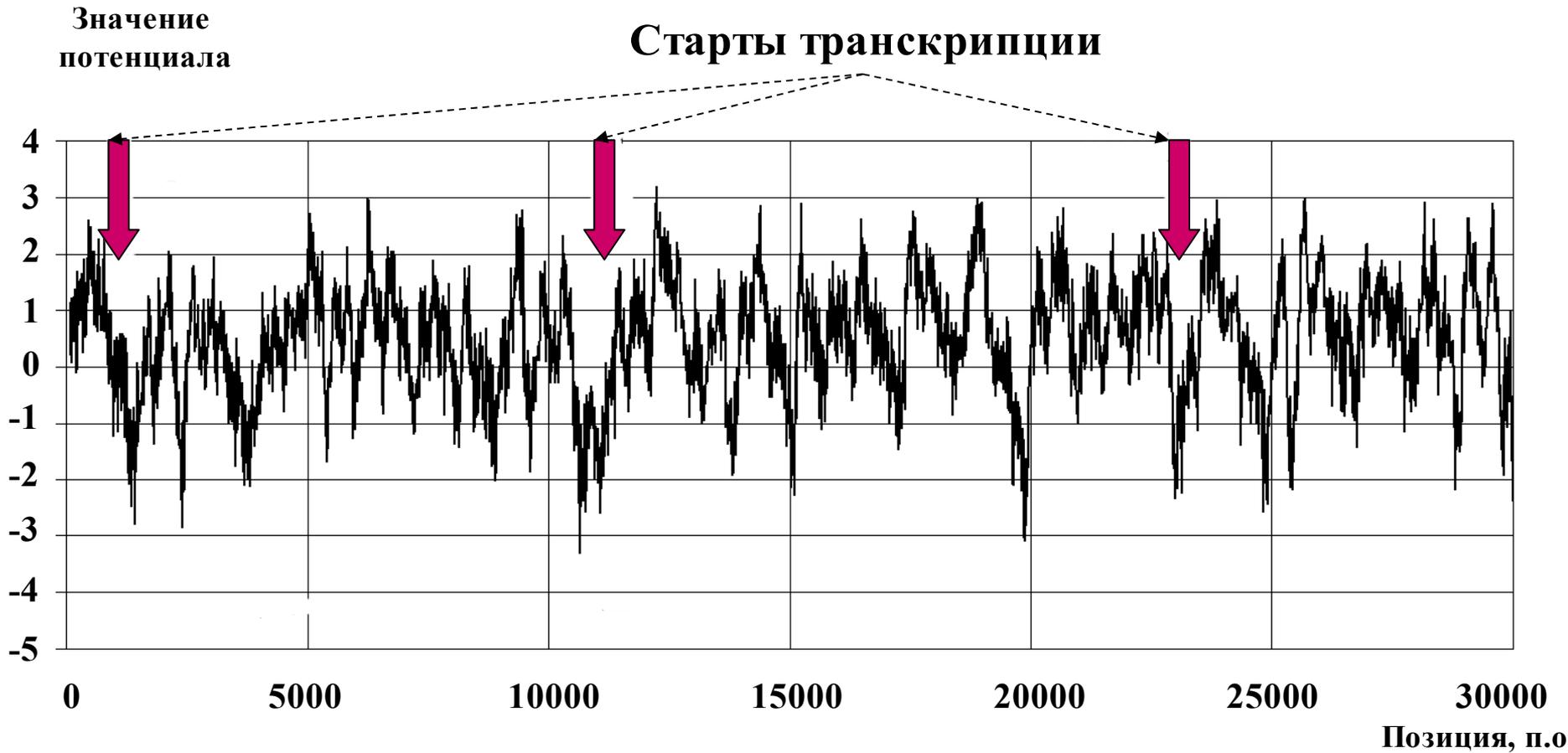
$$\phi(X) = \begin{cases} +1 & \text{нуклеосомные сайты} \\ -1 & \text{случайные последовательности} \end{cases}$$



# **Анализ особенностей нуклеосомной организации различных структурно- функциональных районов геномной ДНК**

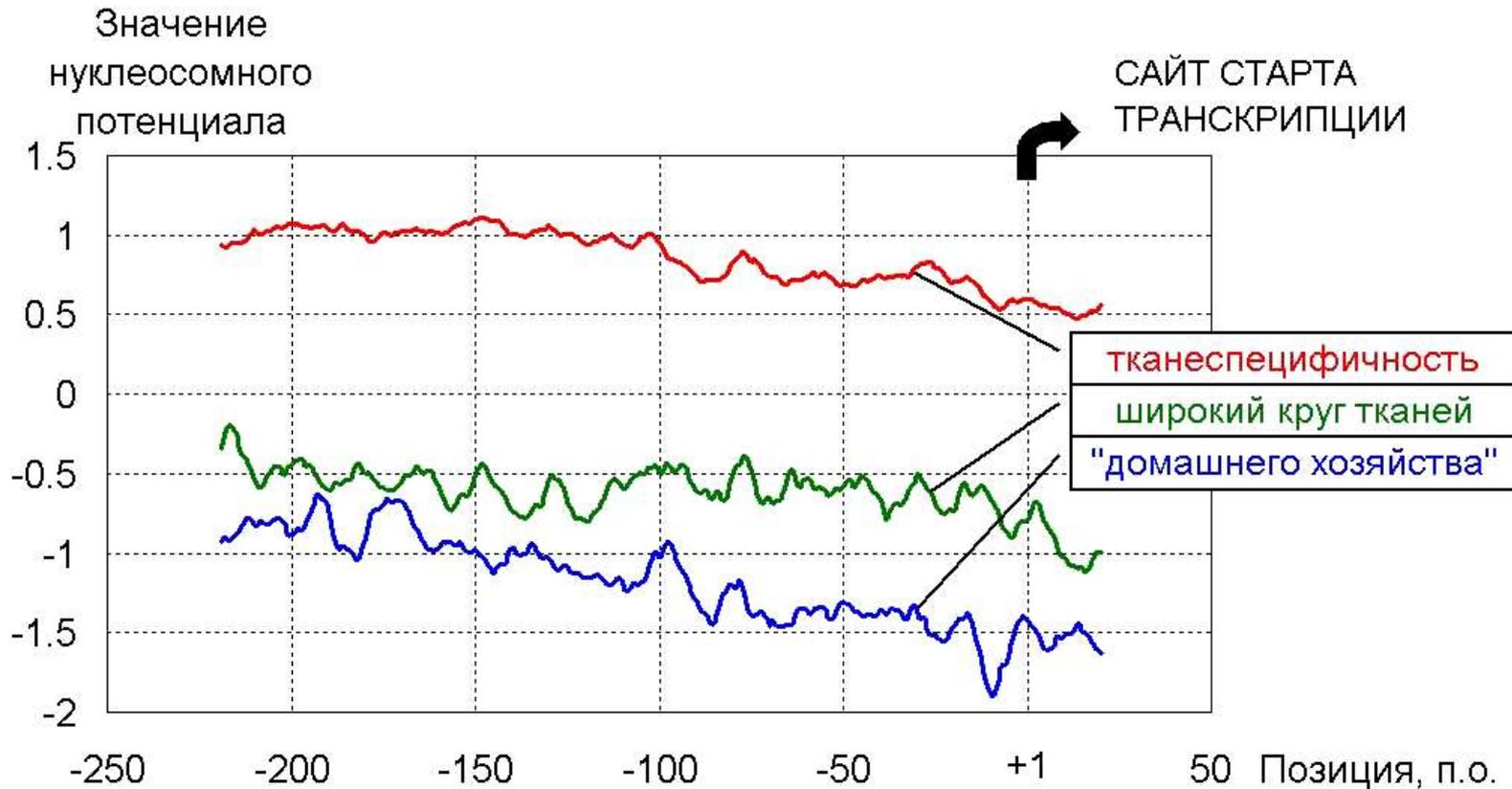


# Профиль нуклеосомного потенциала $\varphi(X)$ для фрагмента *bithorax* комплекса, *Drosophila melanogaster*





# Нуклеосомная организация промоторов и характер экспрессии генов



**Наиболее высокие значения нуклеосомного потенциала наблюдаются для промоторов ткане-специфичных генов, более низкие значения – для промоторов генов “домашнего хозяйства” и экспрессирующихся в широком круге тканей генов**



# Профили нуклеосомного потенциала $\phi(X)$ для донорных и акцепторных сайтов сплайсинга человека

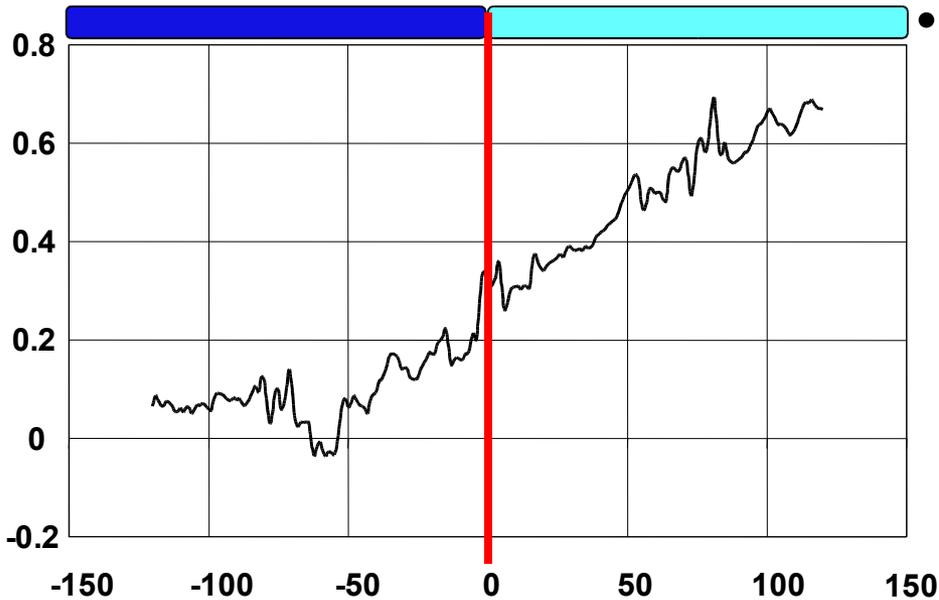


## Донорные сайты

Значение  
потенциала

ЭКЗОН

ИНТРОН



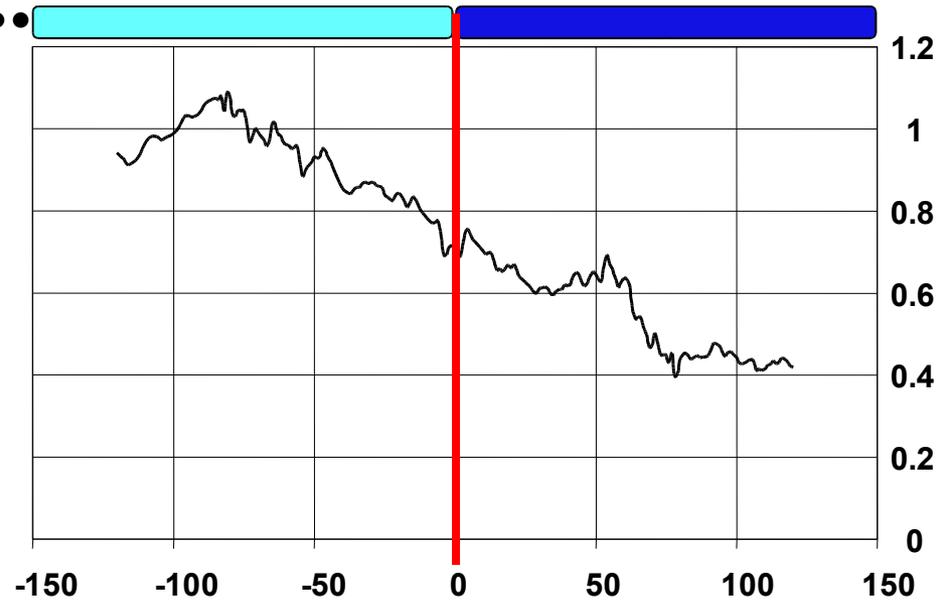
Позиция относительно  
центра сайта сплайсинга, п.о.

## Акцепторные сайты

Значение  
потенциала

ИНТРОН

ЭКЗОН



Позиция относительно  
центра сайта сплайсинга, п.о.



Профиль нуклеосомного потенциала  $\varphi(X)$  для области, расположенной в 5' спейсера  $\rho$ -глобинового гена цыплёнка. отмечены позиции гипечувствительных сайтов (HSS, стрелки вниз), и старта транскрипции (стрелка вверх).

Значение  
потенциала

Старт транскрипции

