



Конформационные и физико-химические контекстно зависимые свойства ДНК и их использование для построения методов анализа и распознавания функциональных районов ДНК

Метод конформационных параметров: система B-DNA Video и SITECON

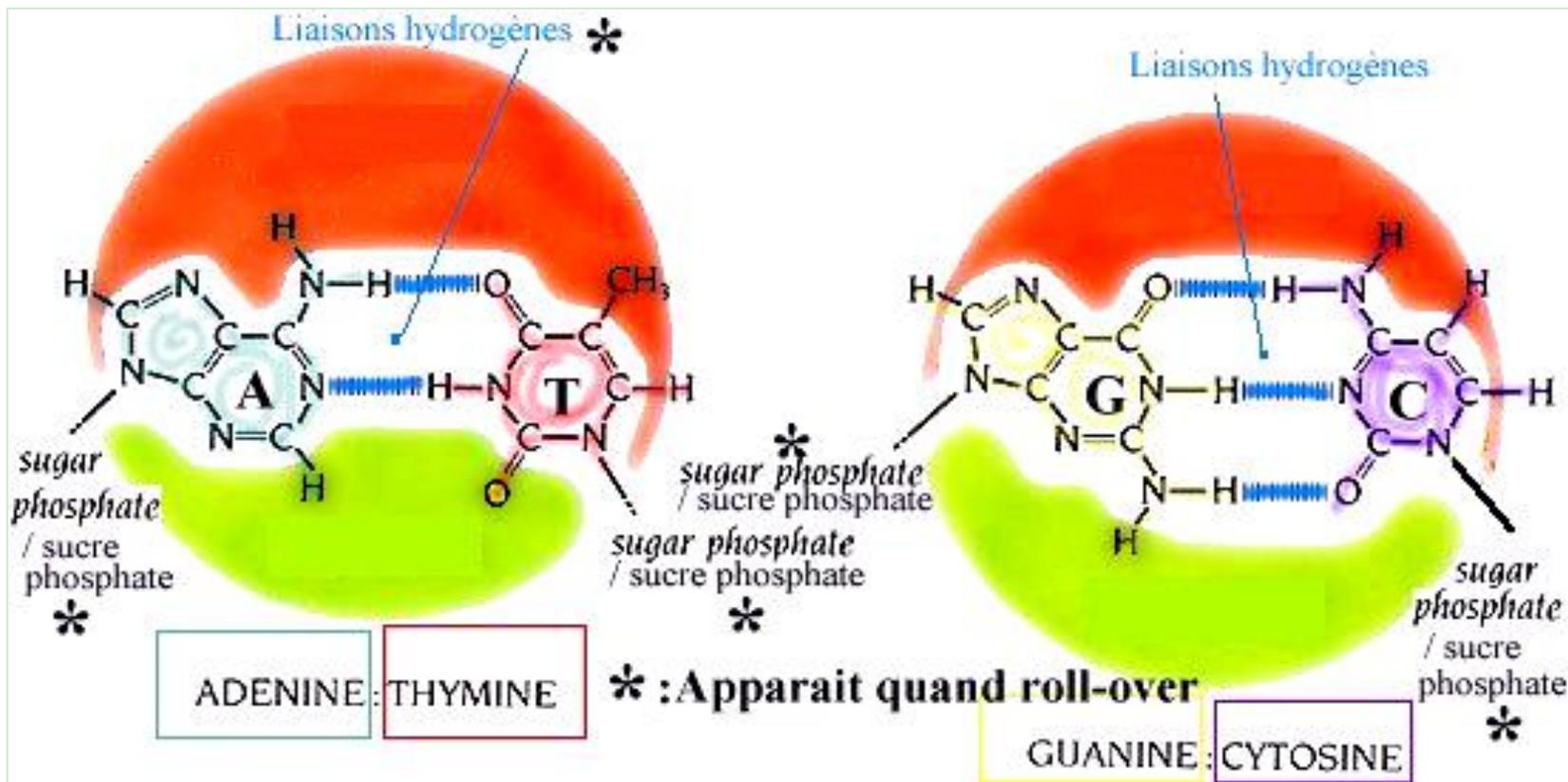
Ощепков Дмитрий Юрьевич

К.б.н., с.н.с. лаб. эволюционной биоинформатики и теоретической генетики

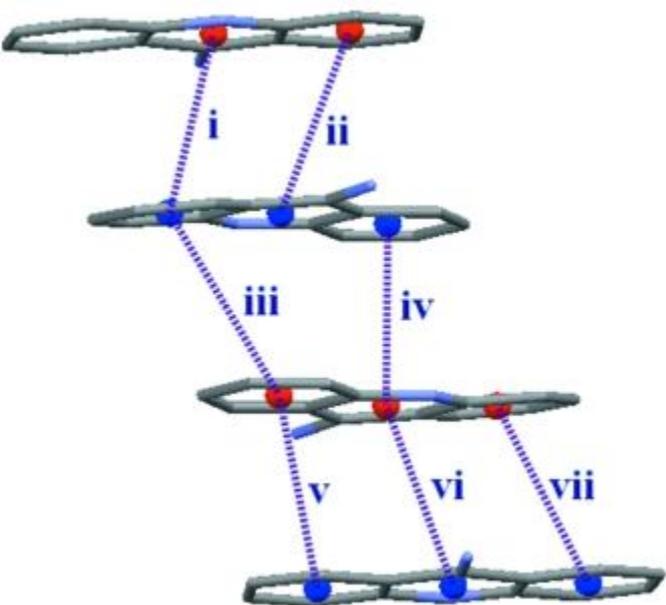
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia



Структура ДНК



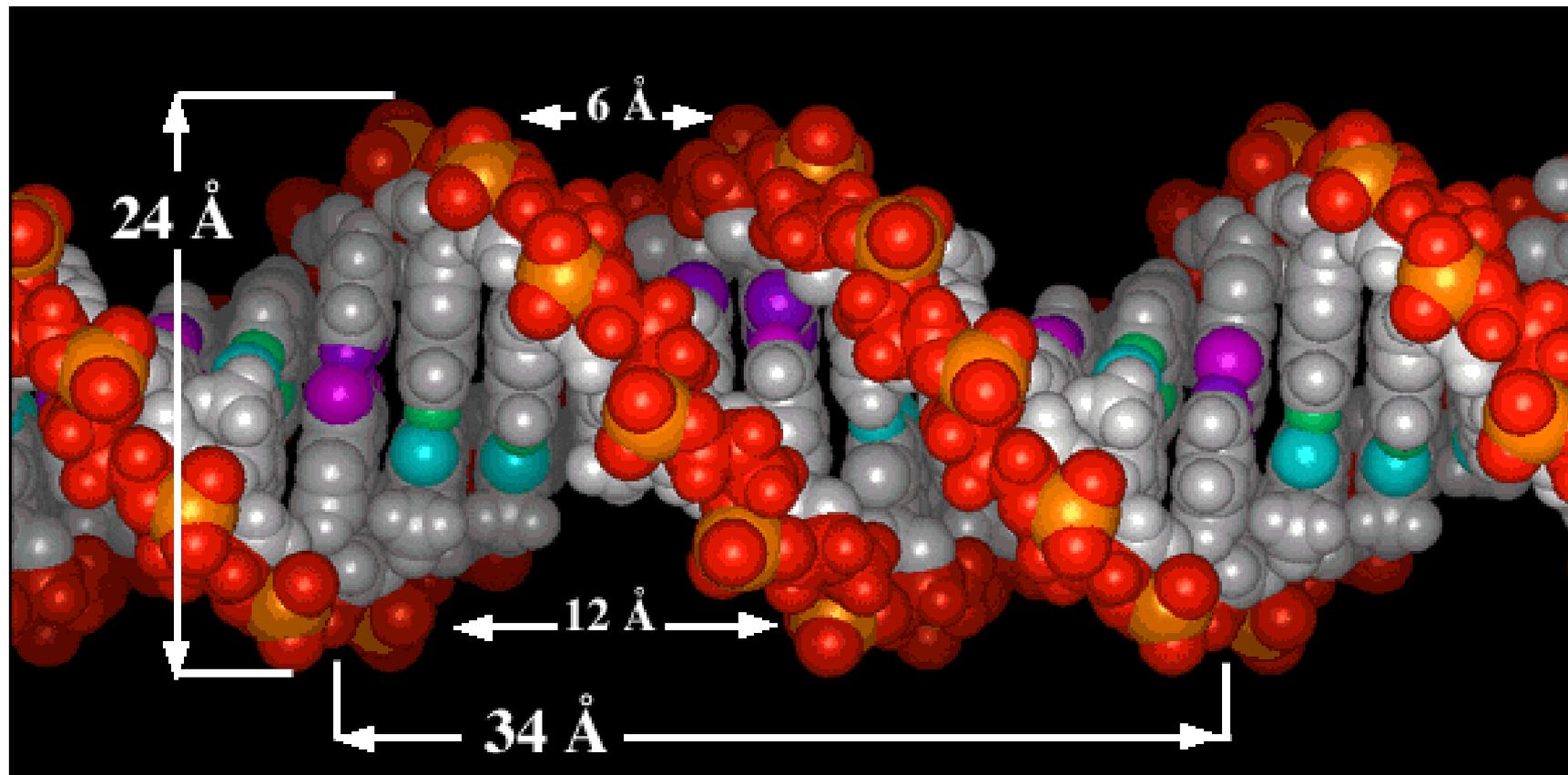
Стэкинг-взаимодействия



Термин стэкинг относится к такому расположению ароматических молекул, которое напоминает расположение монет в стопке и поддерживается ароматическими взаимодействиями. В ДНК параллельный стэкинг имеет место между соседними парами нуклеотидов и повышает стабильность молекулярной структуры. Азотистые основания нуклеотидов имеют пуриновые или пиримидиновые группы в своем составе, состоящие, в свою очередь, из ароматических колец. В молекуле ДНК ароматические кольца расположены примерно перпендикулярно оси спирали, поэтому их поверхности расположены параллельно, что способствует перекрыванию р-орбиталей этих оснований.



Напоминание: структура ДНК



Watson, J.D., Crick, F.H.C., A structure of deoxyribose nucleic acid, *Nature*, 171, 737-738 (1953)



- **Триплетный (генетический) код**
 - **Код укладки РНК**
 - **Хроматиновый код**
 - **Конформационный код**
- и т.д. (Трифонов, 1997)**

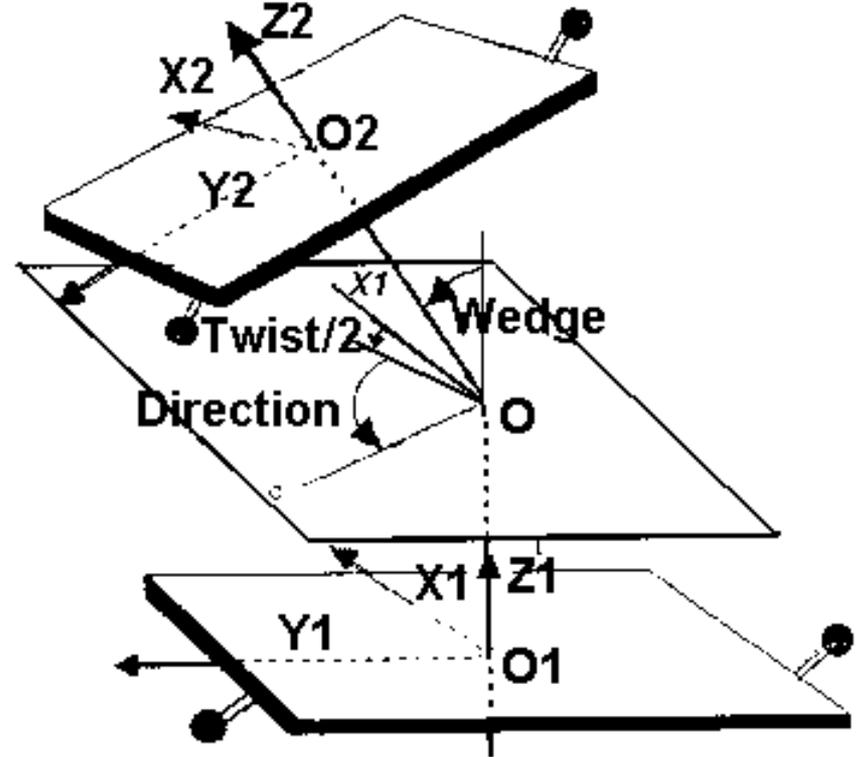
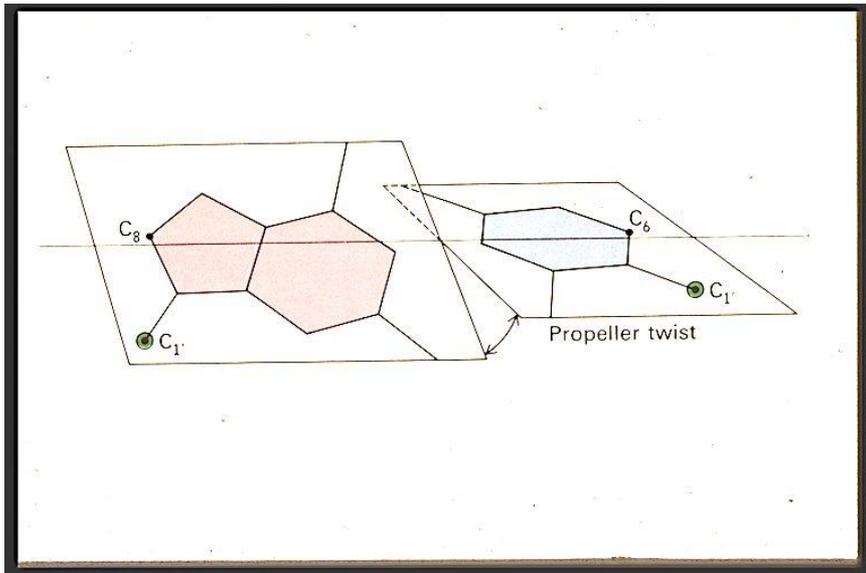
Конформационный код:

Определяемая контекстом локальная конформация молекул ДНК является одним из факторов специфичности ДНК – белкового узнавания. Нарастающий объем экспериментальных данных дает основания считать, что функционирование сайтов связывания транскрипционных факторов в значительной степени определяется конформационными и физико-химическими свойствами ДНК.



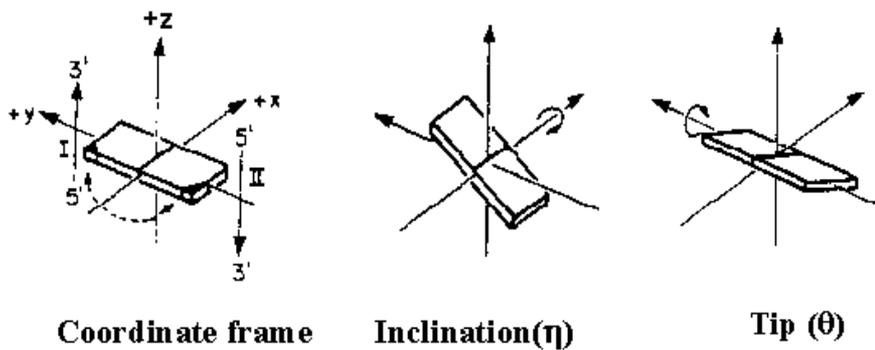
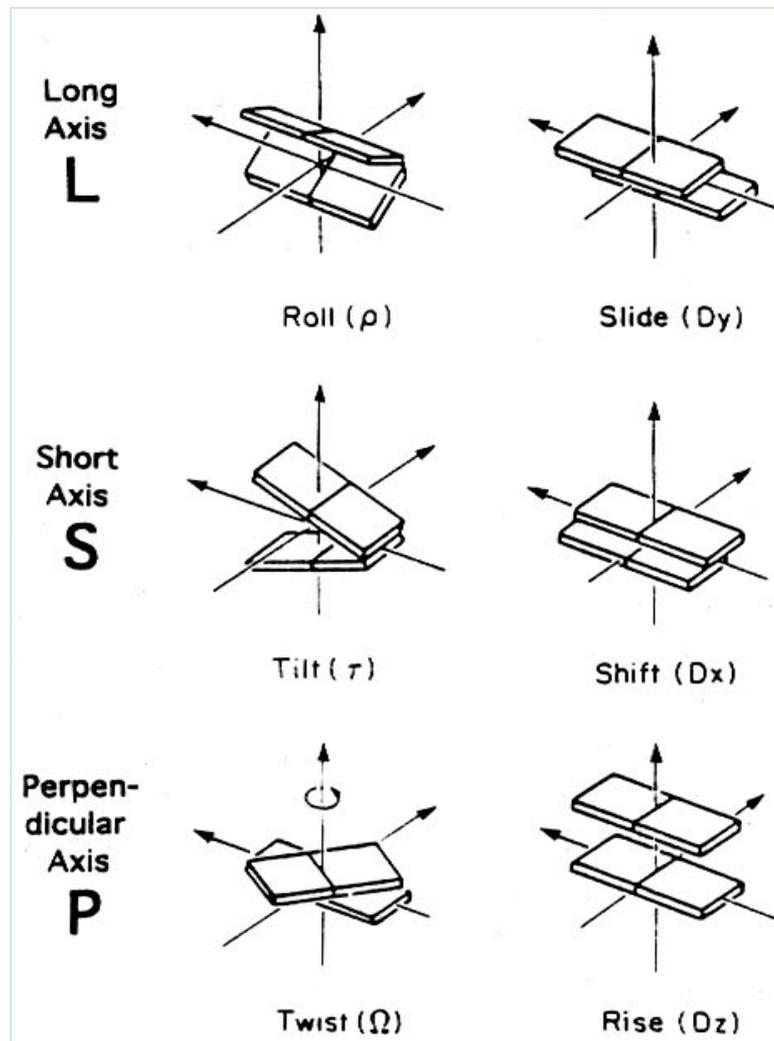
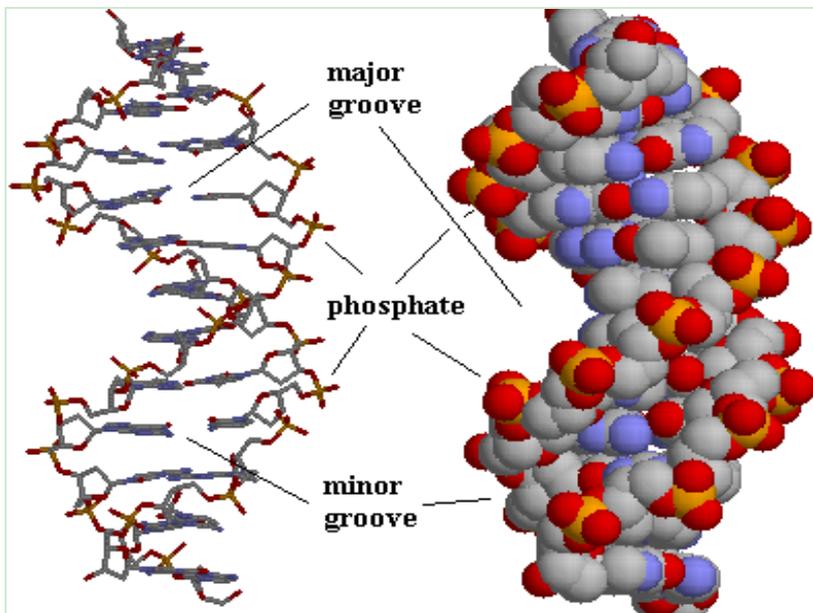
Конформационные свойства ДНК

Впервые зависимость конформации ДНК от контекста была обнаружена Dickerson and Drew (1981)





Конформационные свойства ДНК





Конформационные свойства ДНК



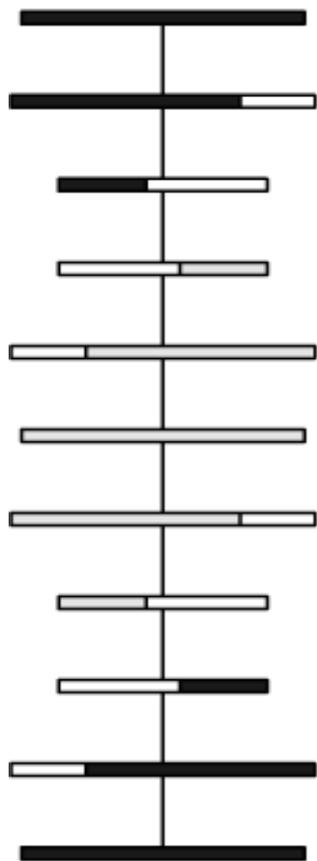
Динуклеотид	Roll °	Slide, Å	Tilt °	Shift, Å	Twist °	Rise, Å
AA	0.3	-0.1	0.5	-0.06	35.3	3.16*
AT	-0.8	-0.4*	0.0	0.12*	31.2*	3.89
AG	4.5	0.4	2.8*	0.06	31.2*	3.63
AC	0.5	-0.2	0.1	0.06	32.6	3.41
TA	2.8	0.9	0.0	-0.17*	40.5*	3.21
TT	0.3	-0.1	0.5	-0.06	35.3	3.16*
TG	0.5	1.6*	-0.7*	0.02	32.6	3.23
TC	-1.3	0.0	0.9	0	40.3	3.47
GA	-1.3	0.0	0.9	0	40.3	3.47
GT	0.5	-0.2	0.1	0.06	32.6	3.41
GG	6.0*	0.8	2.7	0.05	33.3	4.08*
GC	-6.2*	0.4	0.0	-0.3	37.3	3.81
CA	0.5	1.6*	-0.7*	0.02	39.2	3.23
CT	4.5	0.4	2.8*	0.06	31.2*	3.63
CG	-6.2*	0.7	0.0	0.06	36.6	3.60
CC	6.0*	0.8	2.7	0.05	33.3	4.08*



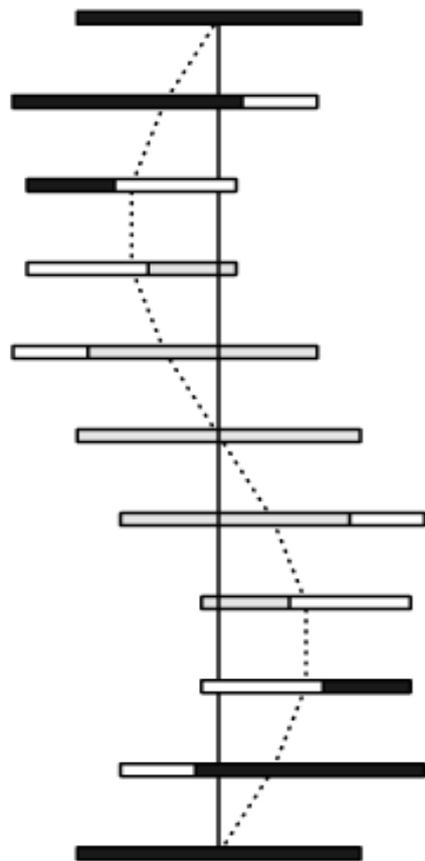
Конформационные свойства ДНК



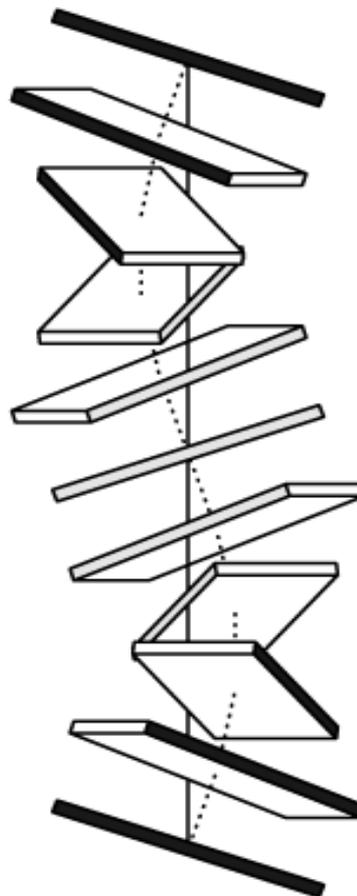
Зависимость конформации молекулы ДНК от конформационных параметров



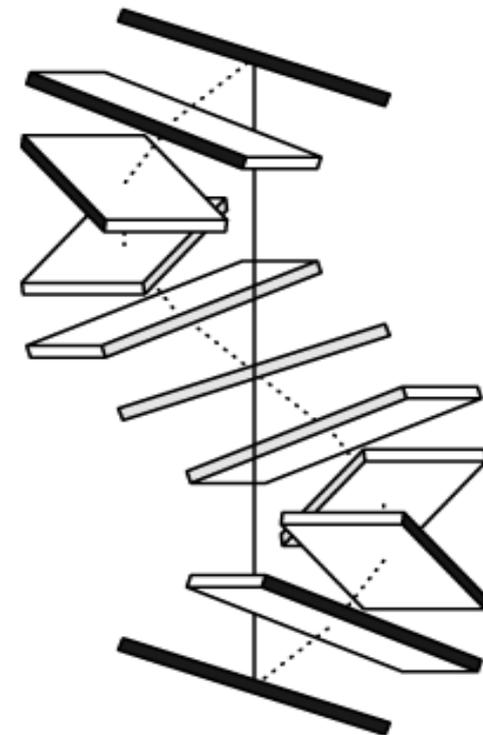
Roll = 0°, Slide = 0 Å



Roll = 0°, Slide = -2 Å



Roll = 12°, Slide = 0 Å



Roll = 12°, Slide = -2 Å



Конформационные свойства ДНК (Правила Калладина)

Зависимость конформационных параметров ДНК от стерических взаимодействий экзоциклических групп NH_2 и O в бороздках спирали

Столкновение пуринов в противоположных полинуклеотидных цепях В-ДНК

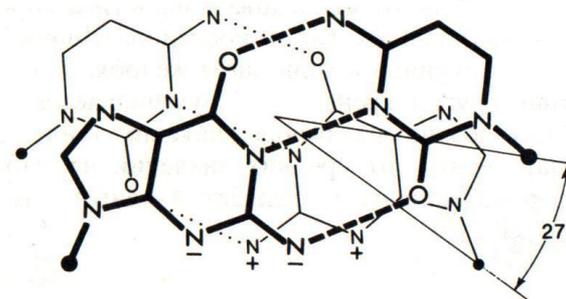
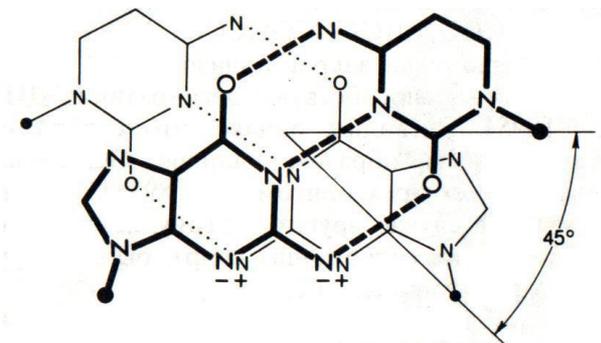
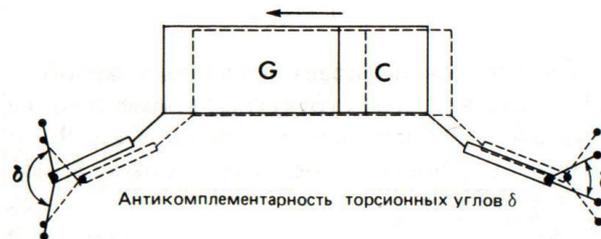
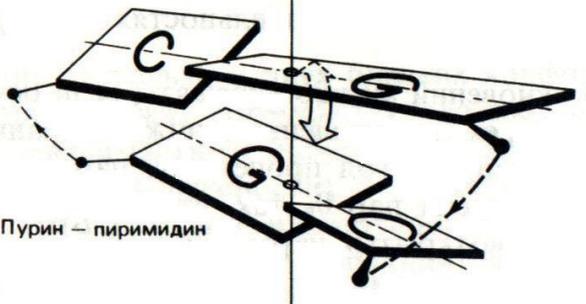
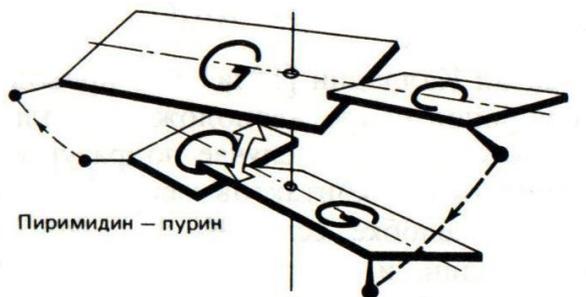
Возникновение стерически недопустимых контактов между «большими» экзоциклическими группами (NH_2 ; O) пуринов противоположных цепей: в случае последовательности пиримидин-3'-5'-пурин они контактируют в минорном желобке. В обратном случае контакты наблюдаются в главном желобке. Нежелательного эффекта можно избежать следующими способами:

- Локально уменьшить угол пропеллера оснований
- Изменить угол крена Roll таким образом чтобы уменьшить невыгодный контакт
- Сдвинуть пару оснований параллельно ее длинной оси и «вытянуть» пурин из спиральной стопки
- Уменьшить угол спирального вращения, чтобы минимизировать нежелательные взаимодействия

Calladine C.R. (1982) Mechanics of sequence-dependent stacking of bases in B-DNA. *J Mol Biol.* 161(2):343–352



Конформационные свойства ДНК (Правила Калладина)



Зенгер В. (1987) Принципы структурной организации нуклеиновых кислот. Москва, “Мир”.



Физико – химические свойства ДНК



Физико-химические свойства – это характеристика того, как молекула ДНК реагирует на внешние взаимодействия.

1. Температура плавления

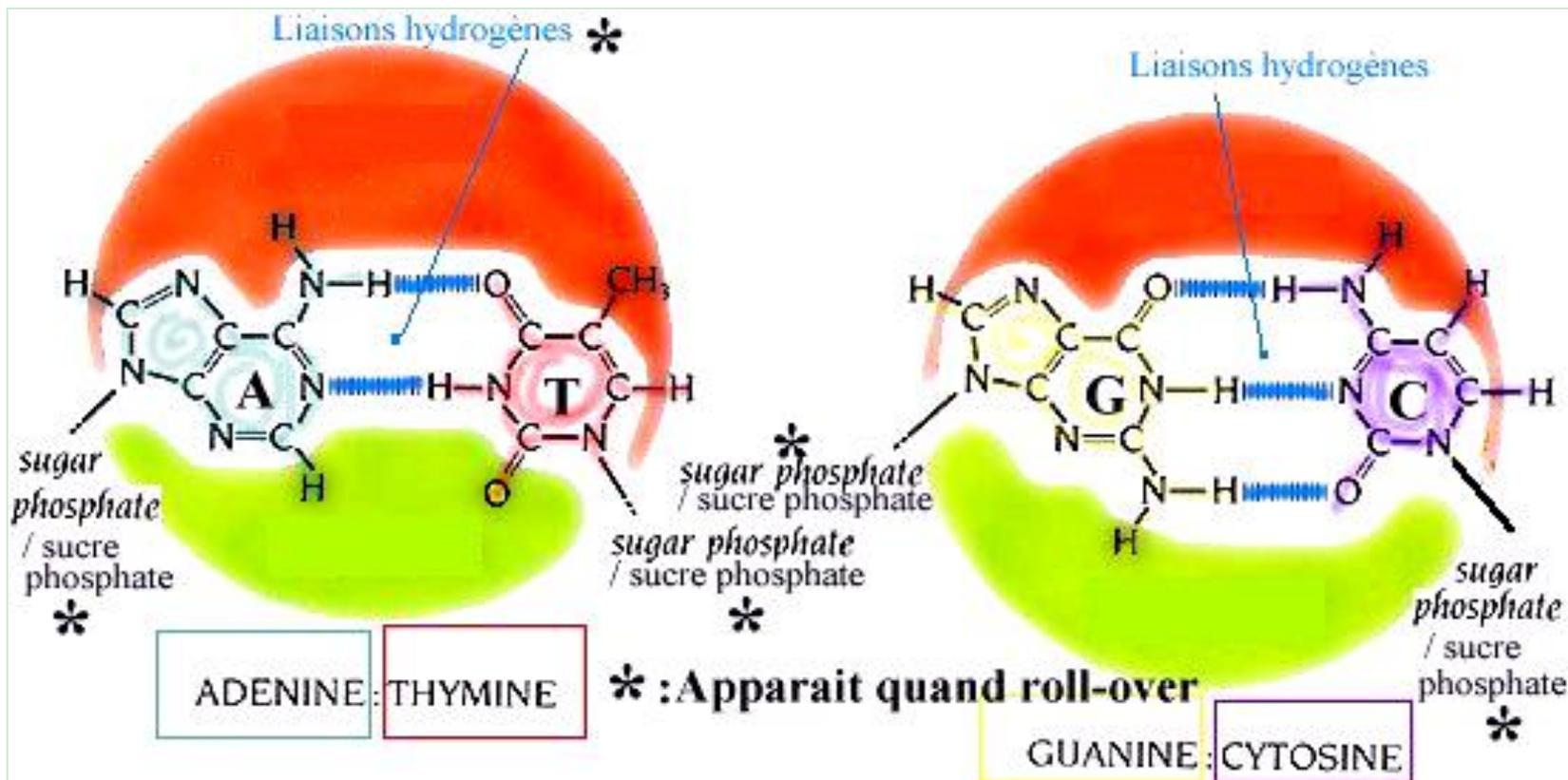
2. Изгибная жесткость

3. Заряд

...



Напоминание: структура ДНК





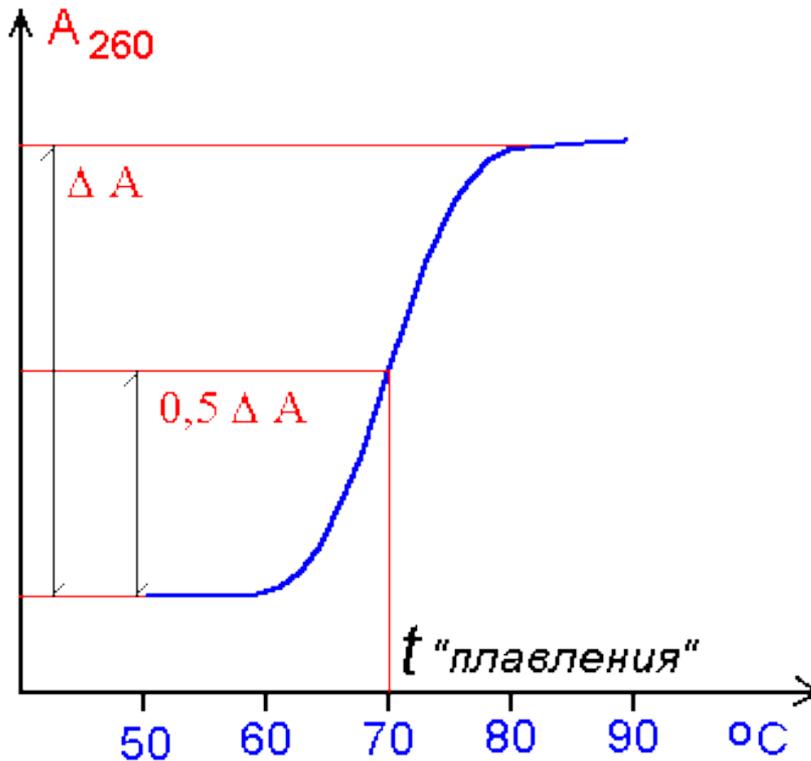
Физико – химические свойства ДНК:

температура плавления



Денатурация ДНК заключается в разрыве Н-связей и стэкинг-взаимодействий, что приводит к расплетанию и разделению цепей (без разрыва ковалентных связей) под действием температуры или рН.

О степени денатурации судят по изменению интенсивности поглощения в ультрафиолете при $\lambda = 260 \text{ nm}$, поскольку дезэкранирование азотистых оснований в результате расплетания цепей устраняется, что вызывает увеличение степени поглощения раствором ДНК ультрафиолета указанной длины волны:



Денатурацию иначе называют плавлением, а температура плавления соответствует моменту 50%-ной денатурации молекулы. Температура плавления различается у разных ДНК.

Температура плавления линейно зависит от доли GC-пар в ДНК, x_{GC} :

$$T_m = T_{AT} + (T_{GC} - T_{AT}) \cdot x_{GC}$$

Температура плавления GC выше чем AT. Для разрушения двух Н-связей AT-пар требуется меньше энергии, чем для разрыва трех Н-связей GC-пар, значения температуры и рН, при которых происходит денатурация, зависят от нуклеотидного состава ДНК.

Температура плавления, °C

AA	54.50
AT	57.02
AG	58.42
AC	97.73
TA	36.73
TT	54.50
TG	54.71
TC	86.44
GA	86.44
GT	97.73
GG	85.97
GC	136.12
CA	54.71
CT	58.42
CG	72.55
CC	85.97



изгибная жесткость



Персистентная длина измеряется как средняя длина между двумя участками двойной спирали, направления которых отличаются на 1 радиан, и отражает ее изгибную жесткость.

Изгибная жесткость ДНК характеризуется персистентной длиной или длиной Куна, $L_{\text{кун}}=0.5L_{\text{перс}}$.

Изгибная жесткость ДНК анизотропна

Персистентная длина, нМ

AA 35.0

AT 20.0

AG 60.0

AC 60.0

TA 20.0

TT 35.0

TG 60.0

TC 60.0

GA 60.0

GT 60.0

GG 130.0

GC 85.0

CA 60.0

CT 60.0

CG 85.0

CC 130.0



Транскрипционные факторы



Транскрипционные факторы (ТФ) белки, специфически связывающиеся с ДНК и регулирующие транскрипцию.

Они взаимодействуют с короткими участками ДНК (сайтами связывания), которые расположены в различных районах генов. ТФ специфично регулируют определенные группы генов ;

От них отличаются :

Базальные транскрипционные факторы (GTFs = БТФ) - особый класс вспомогательных белков, необходимых для точной посадки РНК-полимераз на участки ДНК, прилежащие к стартам транскрипции. БТФ являются общими для всех генов и вовлечены в образование прединициаторного комплекса (ПИК) ;

Некоторые БТФ, например, ТВР, имеют ДНК-связывающий домен, однако функция БТФ реализуется, в основном, через белок-белковые взаимодействия с другими БТФ, ТФ, а также коактиваторами, медиаторами и т.д.

Белки, не имеющие ДНК-связывающих доменов:

Белки-медиаторы и корегуляторные белки (коактиваторы и корепрессоры) не имеют ДНК-связывающих доменов и участвуют в регуляции транскрипции без непосредственного специфического взаимодействия с ДНК



Транскрипционные факторы



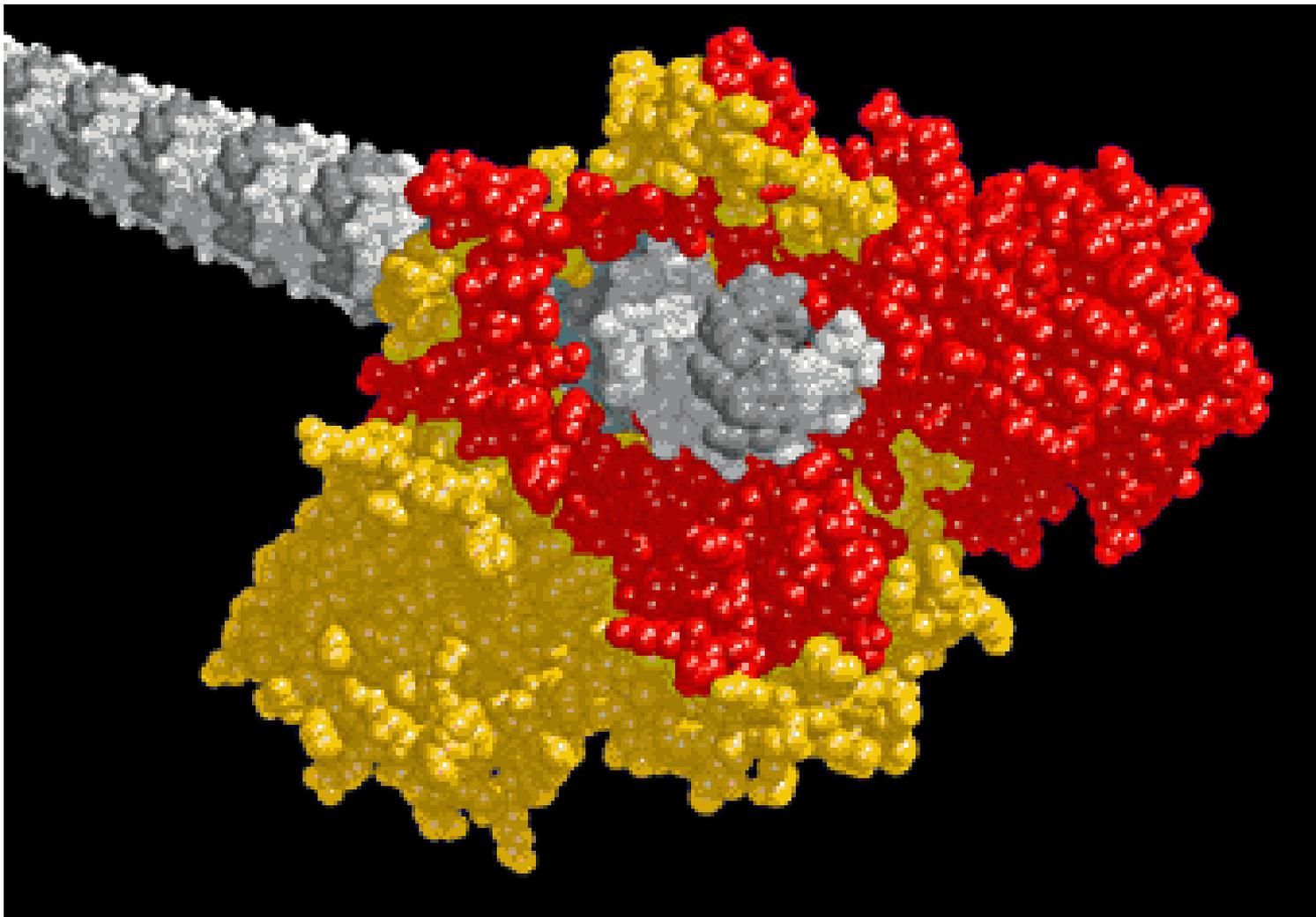
ТРАНСКРИПЦИОННЫЙ ФАКТОР – это белок, который после его перемещения в ядро клетки регулирует транскрипцию, специфически взаимодействуя с ДНК, либо стехиометрически взаимодействуя с другим белком, который может образовывать специфичный к последовательности ДНК комплекс “белок - ДНК”. (Э.Вингендер)

- В большинстве изученных к настоящему моменту случаев ТФ стимулируют формирование комплекса предынициации на ТАТА-боксе / инициаторном элементе за счет взаимодействия их транскрипционно-активирующих доменов с компонентами базального транскрипционного комплекса, либо непосредственно, либо через коактиваторы / медиаторы.
- Некоторые ТФ вызывают изменения структуры хроматина, делая его более доступным для РНК-полимераз.
- Некоторые ТФ являются вспомогательными, создавая оптимальную конформацию ДНК для действия других ТФ
- Известны ТФ, которые подавляют транскрипцию за счет непосредственного действия своих ингибирующих доменов, либо нарушая совместное функционирование комплекса ТФ внутри регуляторной области гена (промотора, энхансера)

Иногда некоторые ТФ, воздействуют на транскрипцию, не связываясь с ДНК, а объединяются в более сложные комплексы посредством белок – белковых взаимодействий. В любом случае, ТФ имеют ДНК-связывающий домен, в обратном случае такой белок называется транскрипционный регулятор



ДНК - белковые взаимодействия

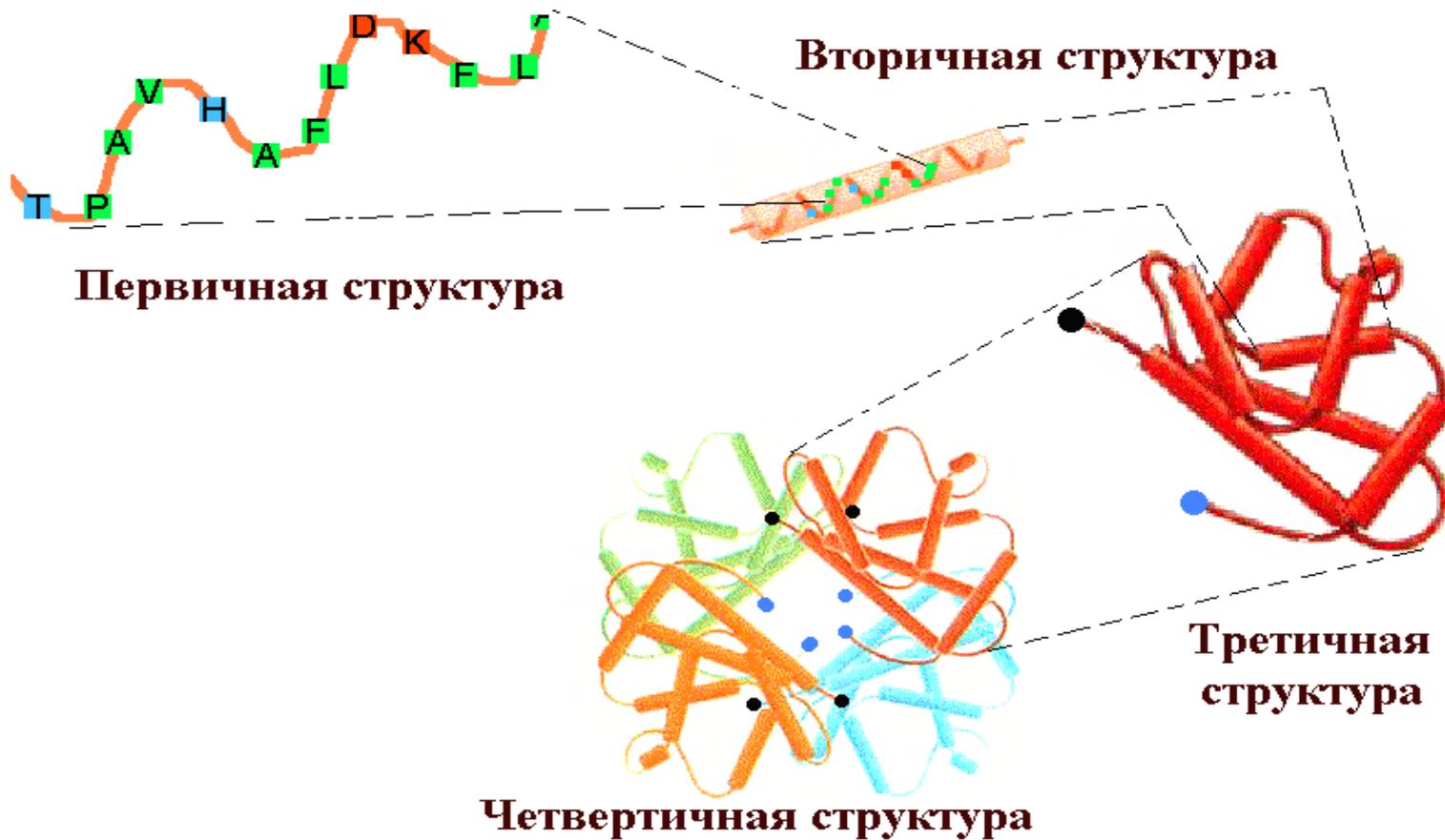




Структура белков – уровни организации

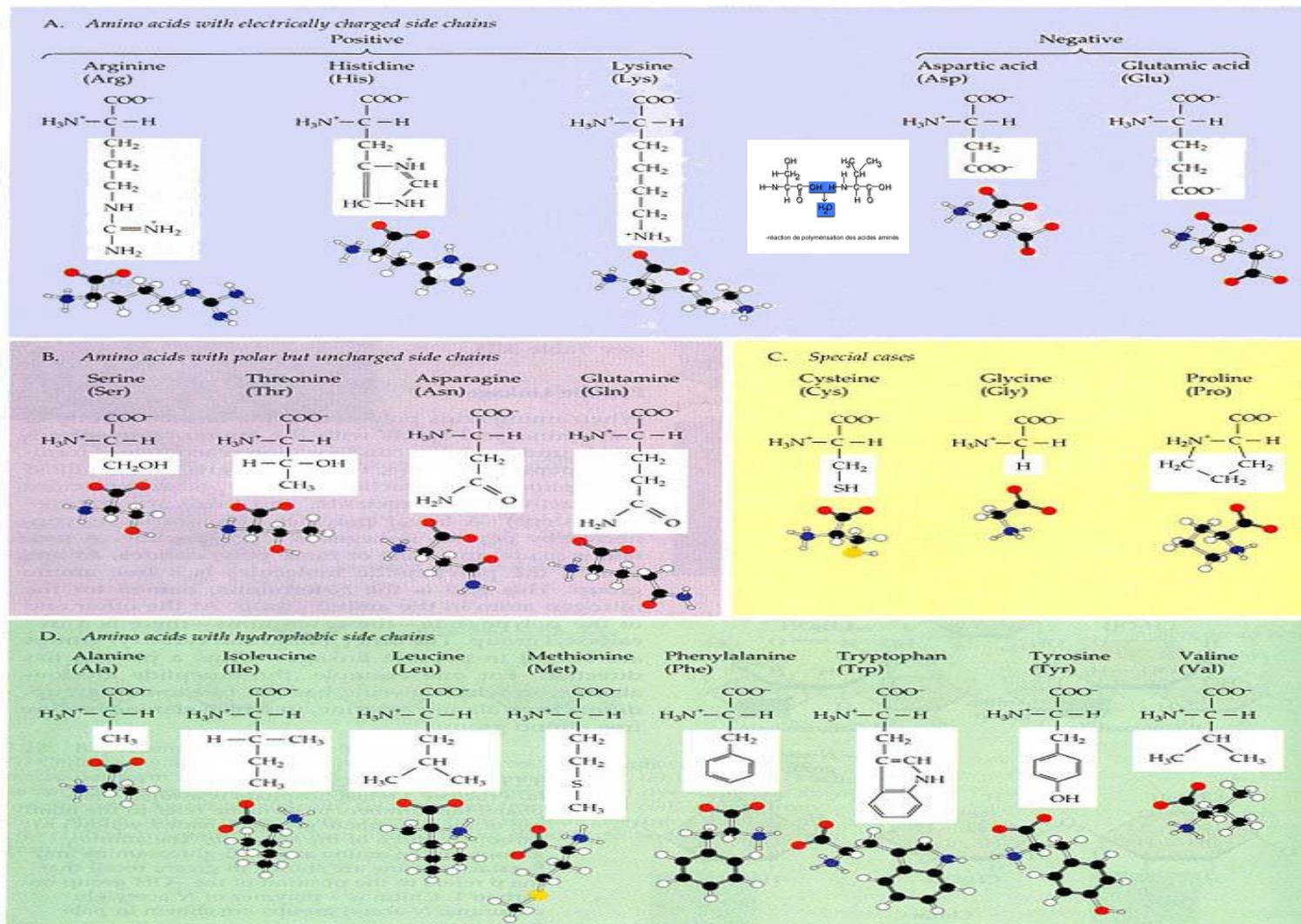


Уровни организации белков





Структура белков - аминокислоты

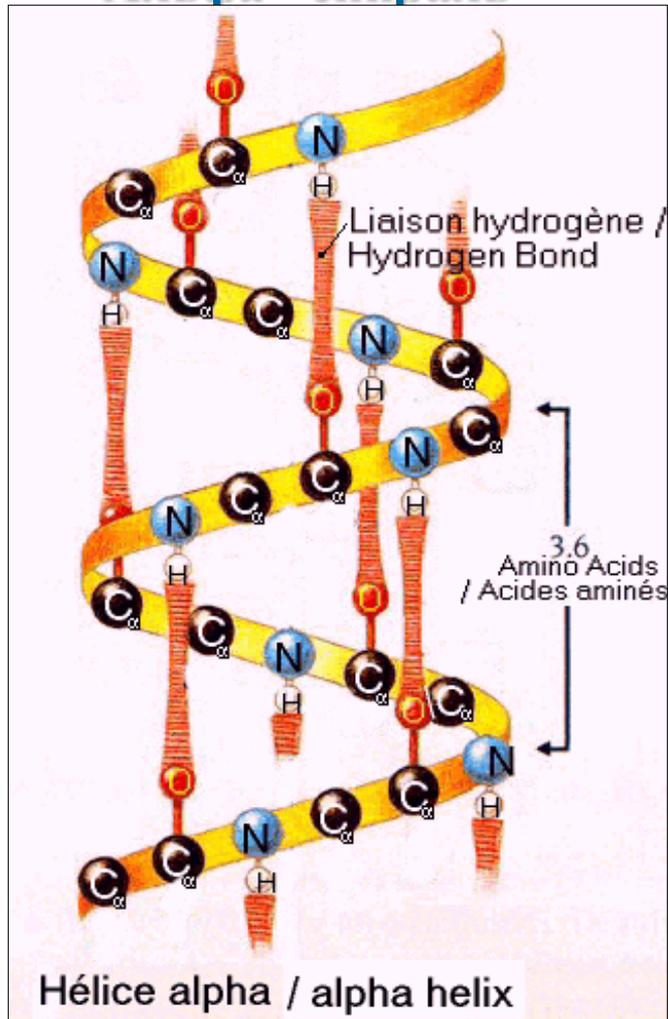




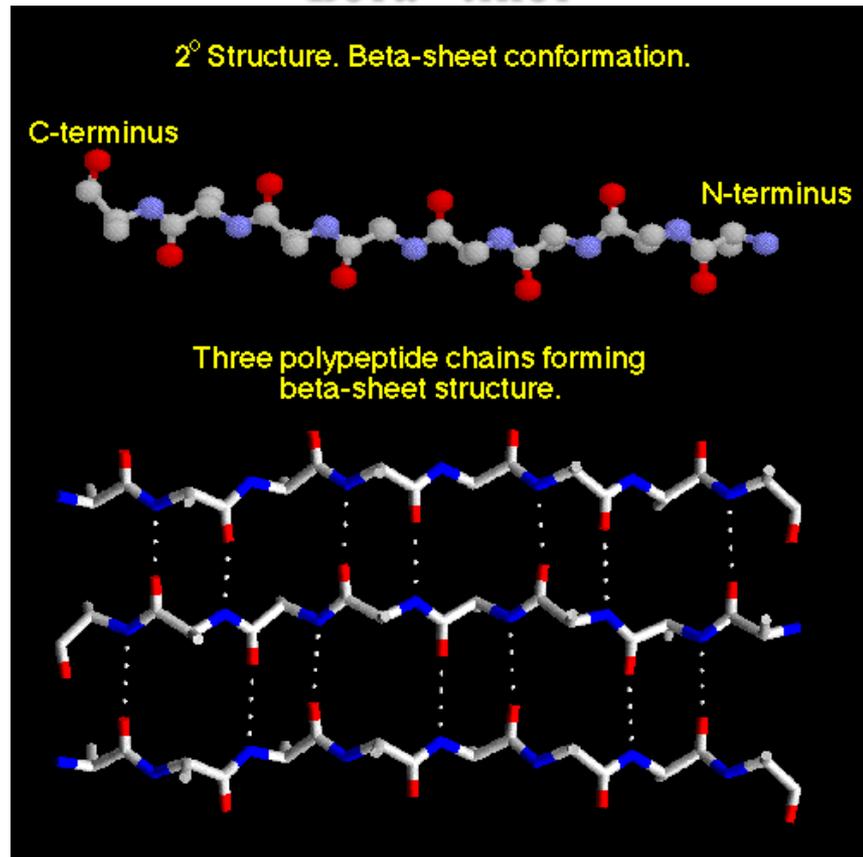
Структурные единицы белков



Альфа - спираль



Бета - лист





ДНК – связывающие домены

КЛАССИФИКАЦИЯ ДНК - СВЯЗЫВАЮЩИХ ДОМЕНОВ



<i>уровень</i>	<i>Структура нумерации</i>	<i>Наименование группы</i>	<i>Критерий</i>	<i>Пример</i>
1	N	Суперкласс	Общая топология ДСД	Домены, координированные цинком
2	N.N	Класс	Структурная характеристика ДСД	Ядерные рецепторы с “цинковыми пальцами”
3	N.N.N	Семейство	Функциональные критерии, такие, как образование ДНК - белкового комплекса	T₃R/RAR (в отличие от рецепторов стероидных гормонов)
4	N.N.N.N	Подсемейство	Как правило, в соответствии со сходством последовательностей в ДСД	RAR (рецептор ретиновой кислоты)
5	N.N.N.N.N	Род фактора	В соответствии с геном фактора	RAR-α, RAR-β
6	N.N.N.N.N.N	Вид фактора	Варианты инициации/ сплайсинга / процессинга	RAR-β1, RAR-β2

Схема классификации ТФ согласно Вингендеру и структура кода



ДНК – связывающие домены

КЛАССИФИКАЦИЯ ДНК - СВЯЗЫВАЮЩИХ ДОМЕНОВ

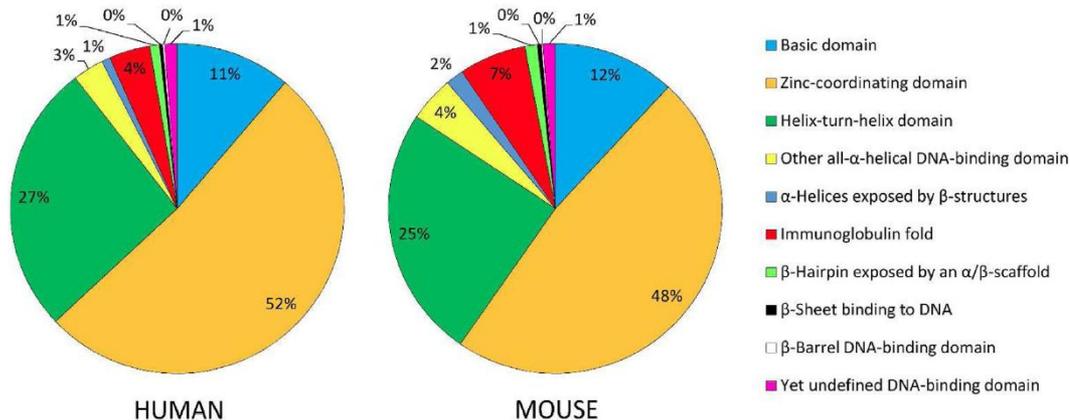
Суперкласс 1. Факторы, ДСД которых обогащен положительно заряженными аминокислотными остатками (basic domain);

Суперкласс 2. Факторы, ДСД которых формируются с участием координирующих ионов цинка (zinc-coordinating DNA-binding domain);

Суперкласс 3. Факторы, имеющие ДНК - связывающий мотив типа спираль – поворот – спираль (helix-turn-helix DNA – binding motif);

Суперкласс 4. Факторы, у которых поверхность, контактирующая с ДНК, представлена в виде сложным образом организованного скэффолда из β -нитей.

Вингендер, Э. (1997) Классификация транскрипционных факторов эукариот. Молекулярная биология. 31, 584-600.



Superclass distribution of human TF genes and their mouse orthologs.

Wingender E, et al., (2015) TFClass: a classification of human transcription factors and their rodent orthologs. Nucleic Acids Res. Jan;43(Database issue):D97-102



ДНК – СВЯЗЫВАЮЩИЕ ДОМЕНЫ



- **1** Basic domains
 - **1.1** Basic leucine zipper factors (bZIP)
 - **1.2** Basic helix-loop-helix factors (bHLH)
 - **1.3** Basic helix-span-helix factors (bHSH)
- **2** Zinc-coordinating DNA-binding domains
 - **2.1** Nuclear receptors with C4 zinc fingers
 - **2.2** Other C4 zinc finger-type factors
 - **2.3** C2H2 zinc finger factors
 - **2.5** DM-type intertwined zinc finger factors
 - **2.6** CXXC zinc finger factors
 - **2.7** C2HC
 - **2.8** C3H zinc finger factors
 - **2.9** C2CH THAP-type zinc finger factors
- **3** Helix-turn-helix domains
 - **3.1** Homeo domain factors
 - **3.2** Paired box factors
 - **3.3** Fork head / winged helix factors
 - **3.4** Heat shock factors
 - **3.5** Tryptophan cluster factors
 - **3.6** TEA domain factors
 - **3.7** ARID
- **4** Other all-alpha-helical DNA-binding domains
 - **4.1** High-mobility group (HMG) domain factors
 - **4.2** Heteromeric CCAAT-binding factors
- **5** alpha-Helices exposed by beta-structures
 - **5.1** MADS box factors
 - **5.3** SAND domain factors
- **6** Immunoglobulin fold
 - **6.1** Rel homology region (RHR) factors
 - **6.2** STAT domain factors
 - **6.3** p53 domain factors
 - **6.4** Runt domain factors
 - **6.5** T-Box factors
 - **6.6** NDT80 domain factors
 - **6.7** Grainyhead
- **7** beta-Hairpin exposed by an alpha/beta-scaffold
 - **7.1** SMAD/NF-1 DNA-binding domain factors
 - **7.2** GCM domain factors
- **8** beta-Sheet binding to DNA
 - **8.1** TATA-binding proteins
 - **8.2** A.T hook factors
- **9** beta-Barrel DNA-binding domains
 - **9.1** Cold-shock domain factors
- **0** Yet undefined DNA-binding domains
 - **0.1** AXUD/CSRNP domain factors
 - **0.2** NonO domain factors
 - **0.3** Leucine-rich repeat flightless-interacting proteins
 - **0.4** NFX1-type putative zinc finger factors
 - **0.0** Uncharacterized

<http://tfclass.bioinf.med.uni-goettingen.de/>
http://www.edgar-wingender.de/huTF_classification.html



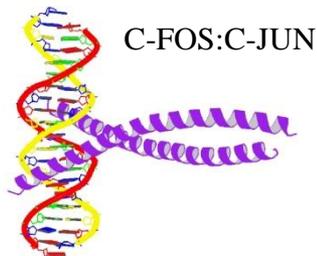
ДНК – СВЯЗЫВАЮЩИЕ ДОМЕНЫ

ОСНОВНЫЕ ТИПЫ ДНК - СВЯЗЫВАЮЩИХ ДОМЕНОВ

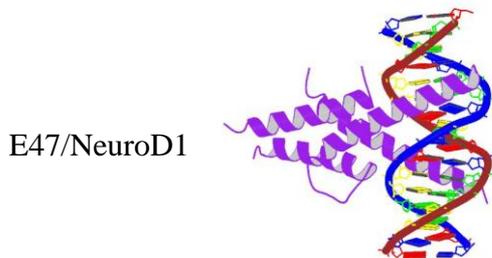


•1 Basic domains

1.1 Basic leucine zipper factors (bZIP)



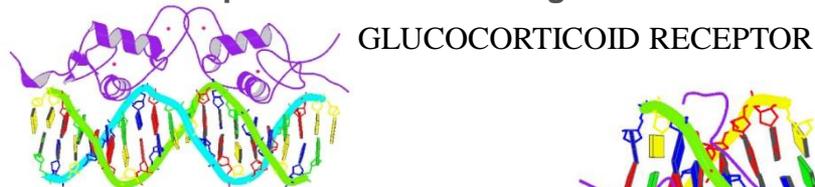
• 1.2 Basic helix-loop-helix factors (bHLH)



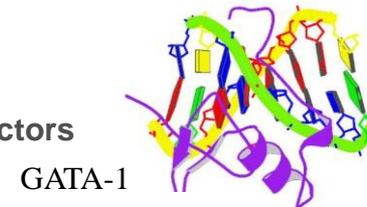
• 1.3 Basic helix-span-helix factors (bHSH)

•2 Zinc-coordinating DNA-binding domains

• 2.1 Nuclear receptors with C4 zinc fingers

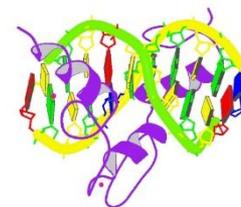


• 2.2 Other C4 zinc finger-type factors



• 2.3 C2H2 zinc finger factors

ZIF268



• 2.5 DM-type intertwined zinc finger factors

• 2.6 CXXC zinc finger factors

• 2.7 C2HC



• 2.8 C3H zinc finger factors

• 2.9 C2CH THAP-type zinc finger factors

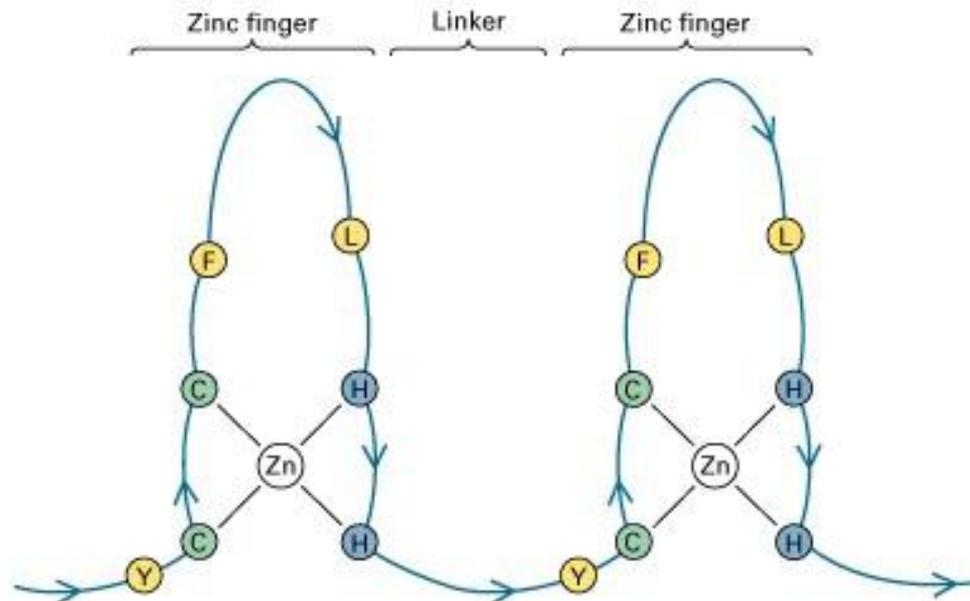




ДНК – связывающие домены

Домены, координированные атомами цинка (Zn - finger)

A



B

1	Y	I	C	S	F	A	D	C	G	A	A	Y	N	K	N	W	K	L	Q	-	A	H	L	C	-	K	H						
2	T	G	E	K	-	P	F	P	C	K	E	E	G	C	E	K	G	F	T	S	L	H	H	L	T	-	R	H	S	L	-	T	H
3	T	G	E	K	-	N	F	T	C	D	S	D	G	C	D	L	R	F	T	T	K	A	N	M	K	-	K	H	F	N	R	F	H
4	N	I	K	I	C	V	Y	V	C	H	F	E	N	C	G	K	A	F	K	K	H	N	Q	L	K	-	V	H	Q	F	-	S	H
5	T	Q	Q	L	-	P	Y	E	C	P	H	E	G	C	D	K	R	F	S	L	P	S	R	L	K	-	R	H	E	K	-	V	H
6	A	G	-	-	-	Y	P	C	K	K	D	D	S	C	S	F	V	G	K	T	W	T	L	Y	L	K	H	V	A	E	C	H	
7	Q	D	-	-	-	L	A	V	C	-	-	D	V	C	N	R	K	F	R	H	K	D	Y	L	R	-	D	H	Q	K	-	T	H
8	E	K	E	R	T	V	Y	L	C	P	R	D	G	C	D	R	S	Y	T	T	A	F	N	L	R	-	S	H	I	Q	S	F	H
9	E	E	Q	R	-	P	F	V	C	E	H	A	G	C	G	K	C	F	A	M	K	K	S	L	E	-	R	H	S	V	-	V	H



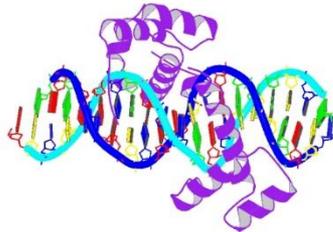
ДНК – СВЯЗЫВАЮЩИЕ ДОМЕНЫ

ОСНОВНЫЕ ТИПЫ ДНК - СВЯЗЫВАЮЩИХ ДОМЕНОВ

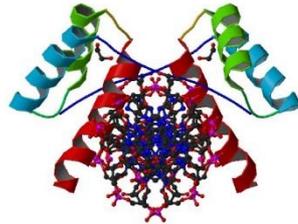
• 3 Helix-turn-helix domains

• 3.1 Homeo domain factors

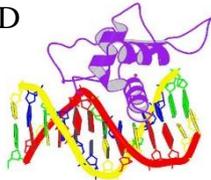
PBX1



• 3.2 Paired box factors PAX3 Homeodomain



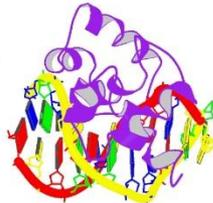
• 3.3 Fork head / winged helix factors HNF-3/FORK HEAD



• 3.4 Heat shock factors

• 3.5 Tryptophan cluster factors

MYB



• 3.6 TEA domain factors

• 3.7 ARID

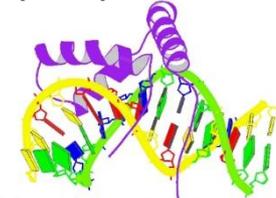
Mrf2



• 4 Other all-alpha-helical DNA-binding domains

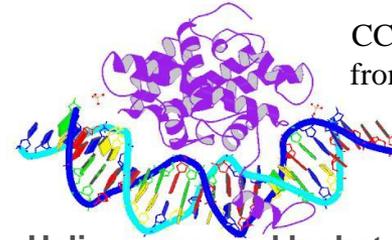
• 4.1 High-mobility group (HMG) domain factors

HMG-BOX



• 4.2 Heteromeric CCAAT-binding factors

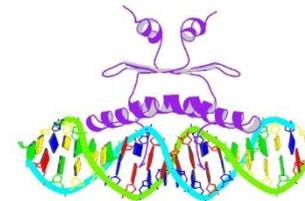
CCAAT-binding complex
from *Aspergillus nidulans*



• 5 alpha-Helices exposed by beta-structures

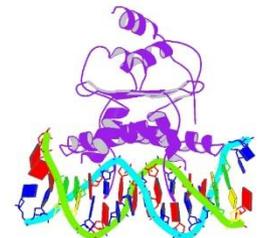
• 5.1 MADS box factors

MEF2A



• 5.3 SAND domain factors

myocyte enhancer factor-2

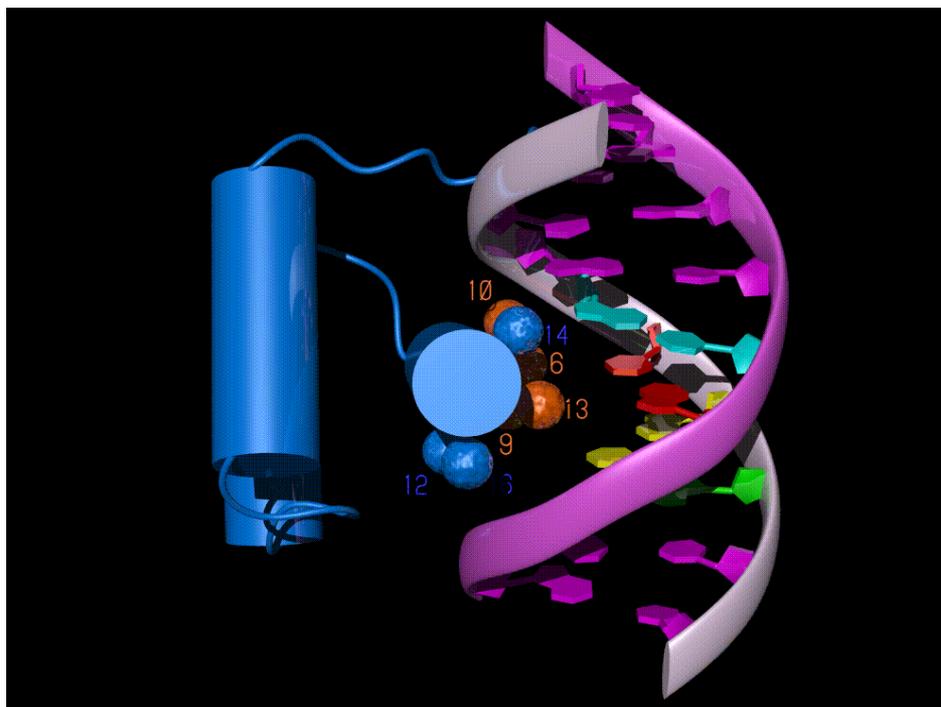




ДНК – связывающие домены



Спираль - поворот - спираль
(Helix - turn - helix)

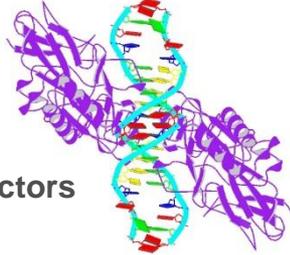




•6 Immunoglobulin fold

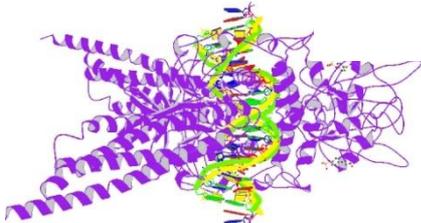
- 6.1 Rel homology region (RHR) factors

NFKB P50 HOMODIMER



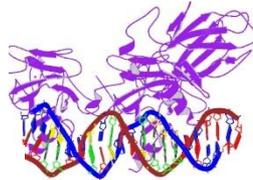
- 6.2 STAT domain factors

STAT-1



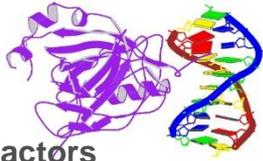
- 6.3 p53 domain factors

P53



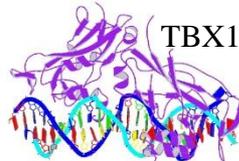
- 6.4 Runt domain factors

Aml1/cbf-beta



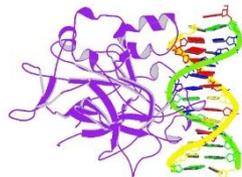
- 6.5 T-Box factors

TBX1



- 6.6 NDT80 domain factors

Ndt80

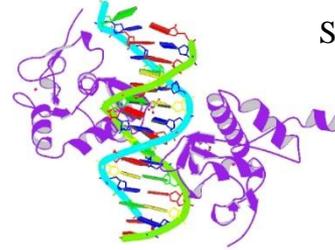


- 6.7 Grainyhead

•7 beta-Hairpin exposed by an alpha/beta-scaffold

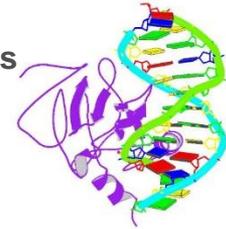
- 7.1 SMAD/NF-1 DNA-binding domain factors

SMAD1-MH1



- 7.2 GCM domain factors

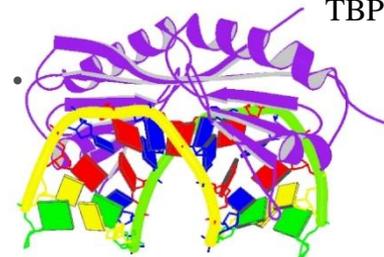
GCM



• 8 beta-Sheet binding to DNA

- 8.1 TATA-binding proteins

TBP



- 8.2 A.T hook factors

• 9 beta-Barrel DNA-binding domains

- 9.1 Cold-shock domain factors

• 0 Yet undefined DNA-binding domains



Система B-DNA Video

Интерфейс системы B-DNA Video в системе Gene Express

Address <http://www.mgs.bionet.nsc.ru/mgs/gnw/bdna/>



HOME DNA RNA PROTEIN GENENETWORKS MAP



General information

[B-DNA publications](#)
[B-DNA workgroup](#)
[B-DNA introduction](#)

User's guide

[HELPS](#)
Format description:
[FEATURES](#)
[The conformational and physico-chemical DNA properties.](#)
[How to use the Recognition Tools?](#)
[Significant conformational and physico-chemical features](#)
[Mean Recognition](#)

B-DNA-VIDEO. This database is designed to study the sets of various transcription factor binding sites, providing evidence that transcription factor binding sites are characterized by specific sets of significant conformational and physicochemical DNA properties. For a fixed site, by using the B-DNA features selected for this site recognition, the C-program recognizing this site may be generated, control tested and stored in the database B-DNA-VIDEO. Each B-DNA-VIDEO entry links to the Web-applet recognizing the site, whose significant B-DNA features are stored in this entry as the 'site recognition programs'. The pairwise linked entry-applet pairs are compiled within the B-DNA-VIDEO system, which is simultaneously the database and the program tools package applicable immediately for recognizing the sites stored in the database. Indeed, this is the novelty.

ACCESS to B-DNA

SRS ACCESS : [SAMPLES](#) [FEATURES](#) [PROFILE LIST](#) [PROPERTY](#)
Recognition: [bDNA Profiles](#) [DNA](#) [Property Plot](#)

General information

[B-DNA publications](#)
[B-DNA workgroup](#)
[B-DNA introduction](#)

User's guide

[HELPS](#)
Format description:
[FEATURES](#)
[The conformational and physico-chemical DNA properties.](#)
[How to use the Recognition Tools?](#)
[Significant conformational and physico-chemical features](#)
[Mean Recognition](#)



Система B-DNA Video

Система B-DNA Video: База данных PROPERTY



Карточка БД Property

```

MI P0000001
MN Conformational
MD B-DNA
ML dinucleotide step
YY
HN SCIO0001
YY
RN RF000012
RN RF000017
XX
PN Twist
PM Calculated by Sklenar, and averaged by Ponomarenko
PV TwistCalc
PU Degree
YY
WW http://wwwmgs.bionet.nsc.ru/Programs/acts2/images/TWIST.html
YY
WW http://wwwmgs.bionet.nsc.ru/Programs/acts2/gallery/Tw00\_bdna.html
YY

```

```

HN SCIO0001
HH Ponomarenko Mikhail Pavlovich, Ph.D.
CV http://wwwmgs.bionet.nsc.ru/mgs/info/pon\_mp/
HP +7-3832-333-119
HF +7-3832-331-278
HE pon@bionet.nsc.ru
HL Laboratory of Theoretical Molecular Genetics
HI Institute of Cytology and Genetics
HD Siberian Branch
HM Russian Academy of Science
HC Russia
HA 10 Lavrentyev Ave., Novosibirsk, 630090, Russia
WW http://wwwmgs.bionet.nsc.ru/Programs/acts2/
WW http://cl.sdsc.edu/
WW http://www.bionet.nsc.ru/SRCG/Activity/

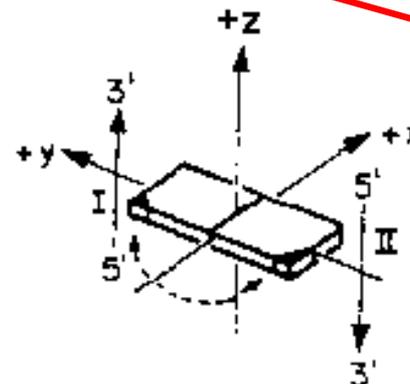
```

```

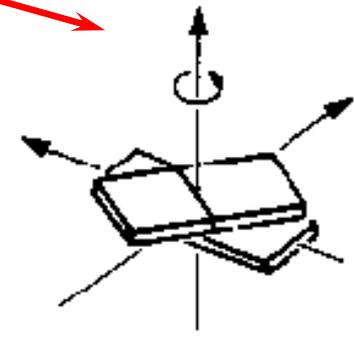
DINUCLEOTIDE
AA 38.90
AT 33.81
AG 32.15
AC 31.12
TA 33.28
TT 38.90
TG 41.41
TC 41.31
GA 41.31
GT 31.12
GG 34.96
GC 38.50
CA 41.41
CT 32.15
CG 32.91
CC 34.96

```

38 карточек, создана в 1997
 году М.П и Ю.В. Пономаренко



Coordinate frame



Twist(Ω)

//



Система B-DNA Video

Система B-DNA Video: программа Property plot



DNA Property Plot

Input DNA Sequence :

from Screen:

from DB: Bases Available: EMBL by ID from Heidelberg (SR55)

from File: Browse... [File formats](#) here.

Choose Property: Twist

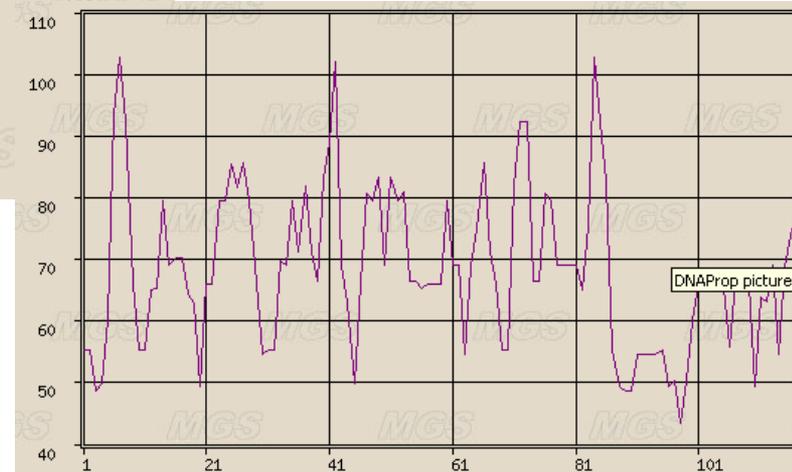
Start Position : 1 End Position: All

Window Size

[About](#)

Интерфейс Internet –
доступной программы
для построения
профилей
конформационных и
физико – химических
свойств
последовательностей
ДНК

Результат анализа участка промотора
вблизи старта транскрипции по
величине температуры плавления





Карточка БД SAMPLES

Содержит 80 выборок
сайтов связывания ТФ
и других биологически
значимых
последовательностей

```
FI AP-1
NM The sample of AP-1 (PEA1; (Jun)2; AP1; Fos/Jun; yAP1; PAR1; PDR4; SNQ3)
NM transcription factor DNA binding sites with flanks
NM (based on TRANSFAC and EMBL databases).
NM The sample contains 69 sequences.
NM The sites with the location determined experimentally
NM are taken into account. (Please, pay attention to the fields MM and BF
NM in the SITE tables of TRANSFAC database. They contain
NM the name of the method and the parameter quality of the site.)
NM The sites are aligned relative to the centers of the
NM TRANSFAC sequence regulatory elements (the field SQ
NM in the SITE table of TRANSFAC database)
NM so that the total length of the sequences was equal to 120 bp.
OR Institute of Cytology and Genetics, Novosibirsk, Russia
AU Julia Ponomarenko
DA 19-SEP-1997
LU 21-JUL-1998; Julia Ponomarenko
ST {0,0} ; AP-1 binding site; AP-1.
FV 1.0
WA Aligned: AP-1
WF Features: AP-1
WW http://wwwngs.bionet.nsc.ru/programs/acts2/images/FOS\_JUN.html
*****
ID AP_10001; DNA
AC AP_10001
OS Simian virus 40
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Polyomavirus.
DR TRANSFAC; R01385; SV$SV40_17; 3.2;
DR EMBL; V01380; SV40XX; ; join(57..176)
FT {0,0} [57;64] direct; EXP
FT {0,0} [58;66] direct; GBS
FT {0,0} [58;63] reverse; RCG
SQ ggaactgggc ggagttaggg gcgggatggg cggagttagg ggcgggacta tggttgctga
ctaattgaga tgcatgcttt gcatacttct gcctgctggg gagcctgggg actttccaca
//
ID AP_10002; DNA
AC AP_10002
OS Simian virus 40
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Polyomavirus.
DR TRANSFAC; R01394; SV$SV40_26; 3.2;
DR EMBL; V01380; SV40XX; ; join(129..248)
```



Система B-DNA Video

Система B-DNA Video: База данных SAMPLES



Описание формата БД SAMPLES

key	description
ID	identification
AC	accession number
DT	date
DE	description
KW	keyword
OS	organism species
OC	organism classification
OG	organelle
RN	reference number
RC	reference comment
RP	reference positions
RX	reference cross-reference
RA	reference author(s)
RT	reference title
RL	reference location
DR	database cross-reference
FT	site description
CC	comments or notes
SQ	sequence header
bb - (blanks)	sequence data
//	termination line



Система B-DNA Video

Система B-DNA Video: База данных FEATURES



БД B-DNA-FEATURE содержит данные по тем конформационным и физико – химическим свойствам, которые значительно отличаются от значений свойств для случайных выборок. Анализируемые выборки были взяты из БД SAMPLES

Каждый вход B-DNA-FEATURE соответствует значимым конформационным и физико – химическим особенностям (features) сайта. Эти матрицы были вычислены на основе данных БД PROPERTY. Только 50% анализируемой выборки было использовано для расчета матрицы. Каждая найденная особенность была использована для написания C-кода праспознающей программы и код сохранен во ссылке [C-CODE](#) каждого входа базы. C-код для среднего распознавания ([Mean Recognition](#)) был также включен в каждый вход.

Оставшиеся 50% выборки были использованы для КОНТРОЛЯ. Контрольные результаты тестирования каждой программы были сохранены в полях [ST](#) (Means, Standard Deviation, False Negatives for Control Sequences), [NT](#) (Means, Standard Deviation, False Positives for Random Sequences), and [FG](#) (Graphical Representation of Test Results).



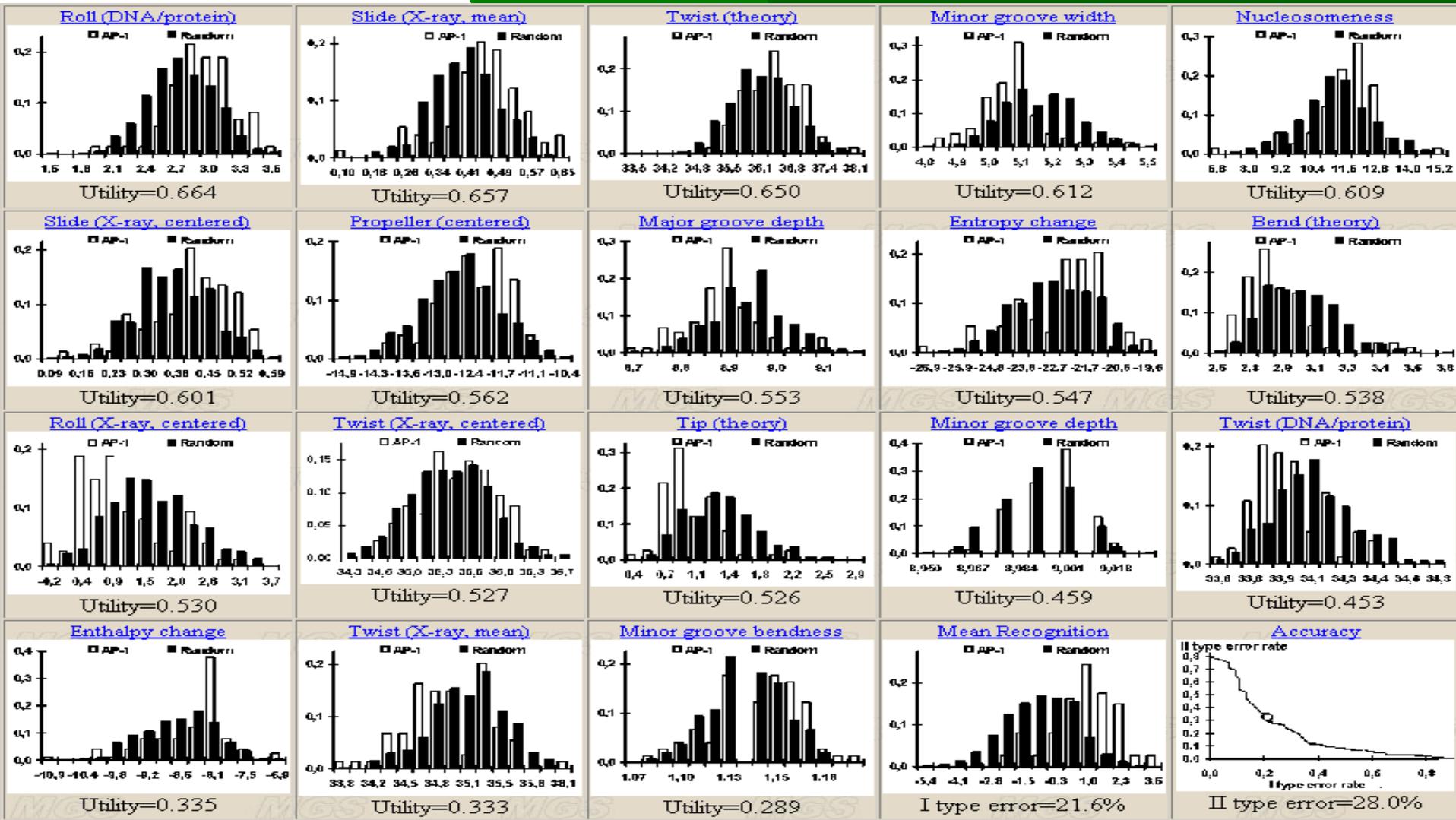
Format description

MI	Entry ID
MN	Site Name
HN	SCIENTIST
DR	SRS-links
WW	Web-link to Recognition Tools
DP	Link to PROPERTY database
PV	Property Name
HL	Feature Indicator
AB	Analyzed Region
UT	Utility
ST	Means, Standard Deviation, False Negatives for Control Sequences
NT	Means, Standard Deviation, False Positives for Random Sequences
FG	Graphical Representation of Test Results
C-CODE	



Система B-DNA Video

Система B-DNA Video: База данных FEATURES





Система B-DNA Video

Система B-DNA Video: База данных FEATURES

Интерфейс Internet – доступной программы для распознавания сайта связывания ТФ AP-1 по значимым конформационным и физико – химическим особенностям из БД FEATURES – карточка БД B-DNA PROFILES

AP-1 transcription factor binding DNA-region

Input DNA Sequence :

from Screen:

from DB: Bases Available: SRS5 from Heidelberg (EMBL) by ID

from File: Browse... [File formats here.](#)

Execute Reset form

Select one of the B-DNA Features listed below to analyse the Sequence inputted:

- AP-1 has Highest Twist (theory)
- AP-1 has Lowest Bend (theory)
- AP-1 has Lowest Tip (theory)
- AP-1 has Lowest Major groove depth (theory)
- AP-1 has Lowest Minor groove width (theory)
- AP-1 has Highest Minor groove depth (theory)
- AP-1 has Lowest Twist (X-ray, averaged)
- AP-1 has Highest Slide (X-ray, averaged)
- AP-1 has Highest Roll (DNA/protein)
- AP-1 has Lowest Twist (DNA/protein)
- AP-1 has Highest Probability to be contacting nucleosome core
- AP-1 has Highest Mobility to bend towards Minor groove
- AP-1 has Highest Twist (X-ray, centered)
- AP-1 has Lowest Roll (X-ray, centered)
- AP-1 has Highest Slide (X-ray, centered)
- AP-1 has Highest Propeller (X-ray, centered)
- AP-1 has Highest Enthalpy change
- AP-1 has Highest Entropy change
- AP-1 has the Positive Mean B-DNA-Score



Локальные конформационные и физико-химические свойства ДНК важны при взаимодействии ДНК с белком

Сайты связывания транскрипционных факторов (ССТФ) – это участки ДНК длиной 5-25 п.н., с которыми связываются транскрипционные факторы.

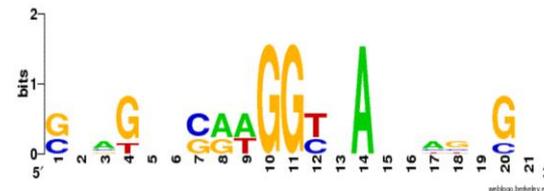
Консенсус: nBSRWGGYAn

Матрица встречаемости:

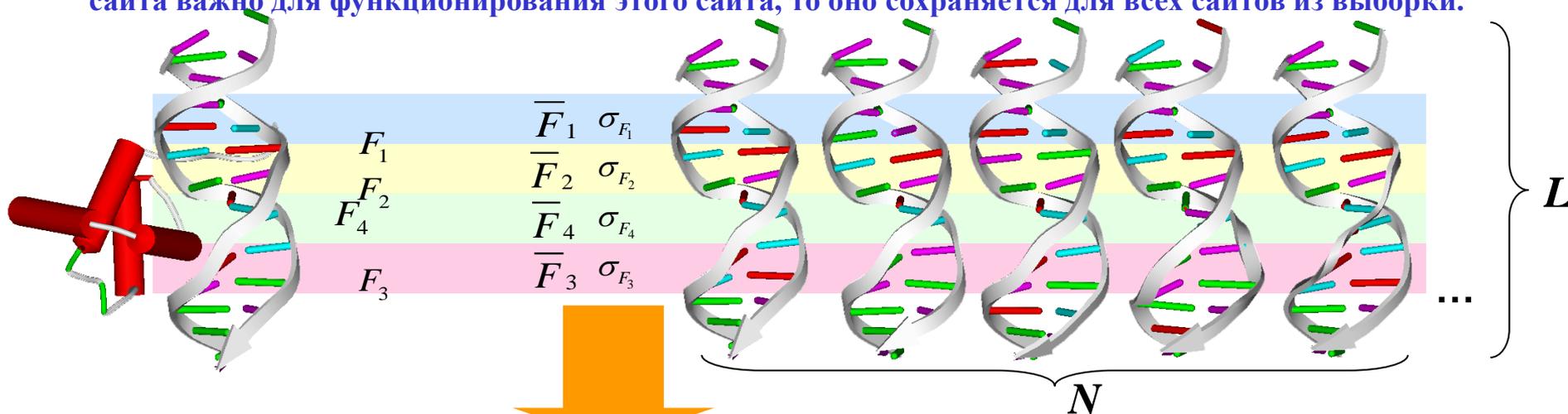
ggagtTCAAGGTAAtaagggc
ccagatCATGGCTAggtgaga
gaggggggAGGTCAacactcc
ctagcgggTGGCTAcgagagg
gtctcCCAAGGTCAtcctgtg



A	1	0	0	3	3	0	0	5	1
T	1	2	0	0	2	0	0	3	2
G	1	2	2	2	0	5	5	0	1
C	2	1	3	0	0	0	0	2	0



Предполагается, что если значение конкретного свойства ДНК в определенном участке последовательности сайта важно для функционирования этого сайта, то оно сохраняется для всех сайтов из выборки.



$$\bar{F}_{il} = \frac{1}{N} \sum_{k=1}^N F_{ikl}$$

$$\sigma^2_{F_{il}} = \frac{1}{N-1} \sum_{k=1}^N (F_{ikl} - \bar{F}_{il})^2$$



SITECON: метод выявления консервативных конформационных и физико-химических свойств ДНК

Выравнивание сайтов

```
tcaatccctg ggtttgccca ...
acagctagaa ttgtctccta ...
cttcagatt cctgagaggc ...
tgccttcta tcaactgaata ...
```

Сопоставление каждому динуклеотиду определенной величины конформационного или физико-химического свойства F_i

```
1.2, 1.4, 2.1, 3.1, 1.6, 1.8, 2.5, ...
2.3, 3.1, 3.7, 3.2, 1.8, 1.2, 2.1, ...
1.3, 3.1, 2.6, 2.4, 1.6, 2.3, 1.3, ...
2.8, 1.4, 2.6, 1.8, 3.2, 1.5, 2.1, ...
```

Оценка среднего \bar{F}_i и дисперсии σ_{F_i} для каждого свойства F_i в каждой позиции k выравнивания

$$\bar{F}_i = \frac{1}{N} \sum_k F_{ik}, \quad \sigma_{F_i} = \frac{1}{N-1} \sum_k (F_{ik} - \bar{F}_i)^2$$

Оценка с помощью критерия χ^2 консервативности σ_{F_i} каждого свойства i в каждой позиции k в сравнении со случайными последовательностями.

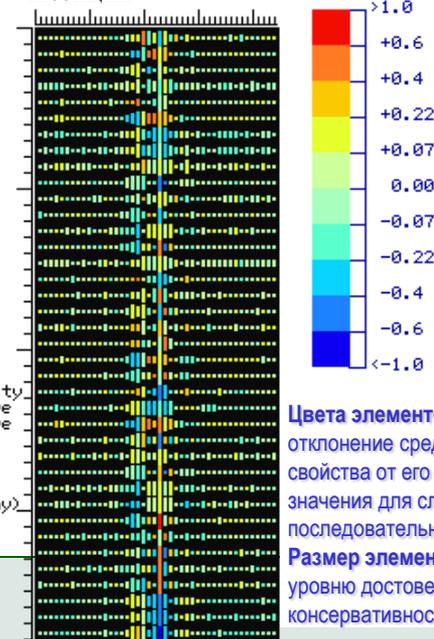
На диаграмме размер блока соответствует уровню значимости, а цвет – отклонению от среднего для случайных последовательностей

$$\bar{F}_{show} = \frac{\bar{F}_{il} - \bar{F}_{rand}}{\sigma_{rand}}$$

Свойство

- 1 Twist (averaged)
- 2 Rise (averaged)
- 3 Bend (averaged)
- 4 Tip (averaged)
- 5 Inclination (averaged)
- 6 Major groove Width
- 7 Major groove Depth
- 8 Minor groove Width
- 9 Minor groove Depth
- 10 Roll (Free DNA)
- 11 Twist (Free DNA)
- 12 Tilt (Free DNA)
- 13 Slide (Free DNA)
- 14 Roll (DNA/protein)
- 15 Twist (DNA/protein)
- 16 Tilt (DNA/protein)
- 17 Slide (DNA/protein)
- 18 Twist (Regressed X-rays)
- 19 Wedge (Regressed X-rays)
- 20 Direction (Regr. X-rays)
- 21 Persistence length
- 22 Melting Temperature
- 23 Nucleos. cont. probability
- 24 Bend towards Major groove
- 25 Bend towards Minor groove
- 26 Twist (Averaged X-ray)
- 27 Tilt (Averaged X-ray)
- 28 Roll (Averaged X-ray)
- 29 Slide (Averaged X-ray)
- 30 Propeller (Averaged X-ray)
- 31 Minor groove size
- 32 Minor groove distance
- 33 Major groove size
- 34 Major groove distance
- 35 Clash strength
- 36 Enthalpy change
- 37 Entropy change
- 38 Free Energy change

Позиция

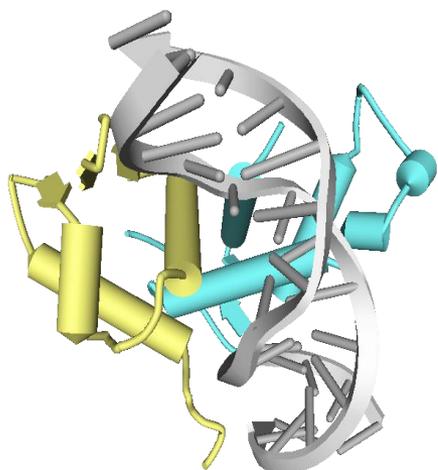


Цвета элементов отражают отклонение среднего значения свойства от его среднего значения для случайных последовательностей. Размер элемента соответствует уровню достоверности консервативности свойства



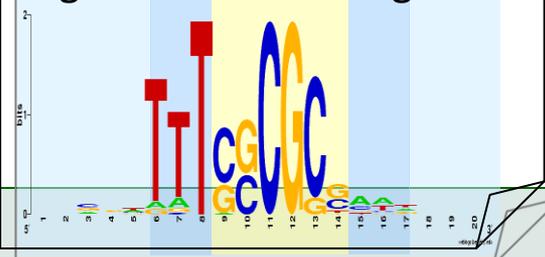
Выборка 40 сайтов связывания гетеродимера E2F/DP длиной 49 п.н. были взяты из базы данных TRRD

ТФ E2F участвуют в регуляции генов клеточного цикла млекопитающих. Связываются с ДНК в форме гетеродимеров E2F/DP.



```
ggggcgggcgggcacaatttcgCGCCAAActtgacgcgcttctgc
gctcggcggtcgtggctcTTTCGCGGCAAAaaggatctggcgcgtaaa
tctttcggcgcaaaaaggattTTCGCGCGTAAaagtggcgggactttgc
acgtcccgcagctgcagatTTTCGCGCGCTTtggcgcgggtggttgg
acgtcccgcagctgcagatTTTCGCGCGCTTtggcgcgggtggttgg
gccaatggcccagcttcttTTTCGCGCCAGcagctgtataaaatgcg
ttttacgcgcaaatcttTTTCGCGGAAAagaccagcggcccgaa
tctttcggcgcaaaaaggattTTCGCGCGTAAaagtggcgggactttgc
atcagaatcgaggcgggctTTTGGCGGGAGGtgggaaagctgtggccat
aacgctgtggcattcggattTTCGCGGAGCcggtgctgatttctgctgc
tccatccggtagcgcaagcTATCCCGCCCTtttcaagcactttctgctc
gtacacaggaagtgaatTTTCGCGCGGTTtagcgagatgtttagt
tcacttctcttattcagttTTCGCGGAAAatggcgaatcttactcg
cgcgcgaggattttaaTGTCGCGCTCTGagccggcgagggagcag
gaaaaagaacggaggaggGATCGCGCTGAGtataaaagccggttttcg
ggaccagaggagggtttTTTGGCGGAAAagaccagctggggacgcg
tcacagaaccactttcctaTCTCCCGCCAAAGtgctcggcgccggcg
ctgctcgcgcaacctgcttTTTCGCGCCAAAGtcacaaagcgggtggtg
acacagctcagggtcgcgaTTTCGCGCCAAActtgacggcgaatcctag
ttcaattatcaggtcggattTTCGCGCCAAAtatagcagctcatttaa
gtagctctgttatcgatatGTTCCCGCCATTcccgtttgatcatctc
ttatcgatatgtcccgccATTCCCGCTTTGatcatctccatccctgg
ccgcgggcggaagtgacgtTTTCCCGCGGTTggacgggcgctcagttg
cagcgccgggcaatgcaaTTTCGCGCCAAActtgggggaagcagcagcg
gctgactcttaaggactagTTTCGCGCCCTTctcaaattaagcgcga
cccctcgtgctcccggcgTTTGGCGCCCGccccctccccctgccc
gcactgcagggcgccagaTTTGGCGGGAGGgggagtgccaaagctct
gcgcgcgctaggcaaacgTTTGGCGGGTTttcctgtgctgtctcct
gccatccggtagcgcaagcTATCCCGCCCTtttcaagcactttggtt
gatccctccctcggttcttTTTCCCGCAAGcctctgagaagccctgcc
aaattttgagtttcaaatTTTGGCGCCACTatttagatattcaatat
tgccgctgtttacagttTTTGGCGCAATTccacctgcggttttcc
tgccgctgtttacagttTTTGGCGCAATTccacctgcggttttcc
ggcatgcgccggggcgtTTTGGCGGGAAGcggggcgggcgggaca
ggttttgaatcactgcgcgTTTCCCGCGCTggggtcaggggtcaggggt
caattacaatctatctaaaATTCCCGCGCTCctcgtcgccaaggaaacc
ccgctggagatgacgtagttTTCGCGCTTAAatttgagaaggcgcgga
tctgttctgagttcaagTATCCCGGACTattgaaatggacaaatga
cgcgcgaggattttaaTGTCGCGCTCTGagccggcgagggagcag
attggtgaatccggggccTTTAGCGCGGTGagtttgaactgctcga
```

```
cacaatttcgCGCCAAActt
ggctcTTTCGCGGCAAAaag
aaggattTTCGCGCGTAAaag
cagatTTTCGCGCGCTTtgg
cagatTTTCGCGCGCTTtgg
ttcttTTTCGCGCCAGcag
tccttTTTGGCGGAAAagag
aaggattTTCGCGCGTAAaag
cgggcTTTGGCGGGAGGtgg
tcggaTTTGGCGCGAGCgcg
cacgcTATCCCGCGCCTttt
acaatTTTCGCGCGGTTtta
tcagttTTTCCCGGAAAatg
ttaaaTGTCGCGCTCTGagc
ggaggGATCGCGCTGAGtat
ggtttTTTGGCGGAAAaga
tcctaTCTCCCGCCAAAGtgc
tgcttTTTCGCGCCAAAgtc
```





Анализ консервативных контекстно – зависимых свойств ДНК сайтов связывания E2F/DP

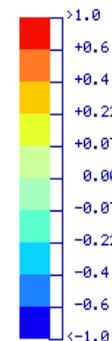
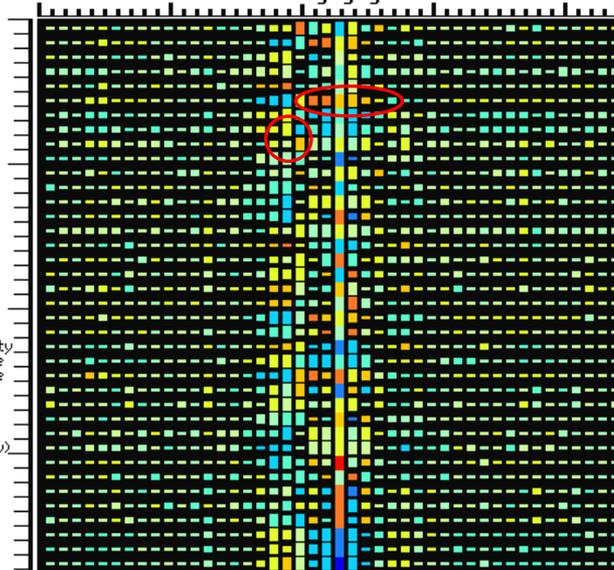
ggggcgggcgggccacaaTTTCGCGCCAAActtgaccgcgcttctgc
gctcggcggtcctgtggctcTTTCGCGGCAAAaggatttggcgcgtaaa
tctttcgggcaaaaaggaTTTCGCGCTAAaagtggccgggactttgc
acgtcccgcagctgcagatTTTCGCGCTTtggcgcaggtggttgg
acgtcccgcagctgcagatTTTCGCGGCTTtggcgcaggtggttgg
gccaatggccagcttcttTTTCGCGCCAGcagctgtataaaatgcg
ttttacgcgccaaatccttTTTCGCGGAAAgagccacgcccgcgaa
tctttcgggcaaaaaggaTTTCGCGCTAAaagtggccgggactttgc
atcagaatcaggcggggcTTTCGCGGAGGtgggaacgctgtggccat
aacgctgtggccattcgggTTTCGCGGAGGcggctggagtttctgc
tccatccggtagcgcacgcTATCCCGCCCTtttaagcactttcgttt
gtacacaggaagtgcacaaTTTCGCGGCTTtaggcggatgttgaagt
tcacttctcttattcagttTCGCGGCAAAatggcacaatcttactcg
cgcgcgagggttttaaaTTTCGCGCTCTGagccggcgaggagcag
gaaaaagaacggaggaggGATCGCGCTGAGtataaaagccggttttcg
gggaccagaggagggtttTTTCGCGGAAAgacacgtggggacgcg
tcacagaaccactttcctaTCTCCCGCAAGtgctcggcgccgcccgg
ctgctgcggaacctgcttTTTCGCGCAAAgtcacaagcgggtggg
acacagctcagggtcgcgaTTTCGCGCAAAActtgacggcaaatcctag
ttcaattatcaggtcgggTTTCGCGCAAAatatagcagtcattttaa
gtagctctgttatcgatatGTTCCCGCATTcccgctttgatcatctc
ttatcgatatgttcccgcATTCCCGCTTtGatcatctccatccctgg
ccgggcggaagtgcagttTCGCGGCTTggacgggcgctcagttg
cagcgccgggcaatgcaaTTTCGCGCAAAActtgggggaagcacagc
gctgactcttaaggactagTTTCGCGCCTTtctcaaattaagcgcga
cccctcgtctccggcgTTTCGCGCCCGcggccctcccctgcgccc
gcactcagggcgccagatTTTCGCGGAGGgggagtgctcaaaagctct
gcgcgctaggcaaacgTTTCGCGGTTTtctctgtgctgtgctcct
gccatccggtagcgcacgTATCCCGGCTtttaagcactttctggtt
gatccctcccctcgttcttTTTCGCGGCAAAatgagccctgc
aaattttgagtttcaaatTTTCGCGGCTTtcaaatgagccctgc
tgccgcgctggttacagttTTTCGCGGCAAAatgagccctgc
tgccgcgctggttacagttTTTCGCGGCAAAatgagccctgc
cgcatgcccggcggtTTTCGCGGCAAAatgagccctgc
ggtttgaatcactgTTTCGCGGCTTtcaaatgagccctgc
caattacaatctatcTTTCGCGGCTTtcaaatgagccctgc
ccgctggagatgagTTTCGCGGCTTtcaaatgagccctgc
tctgttcttgagttcagTTTCGCGGCTTtcaaatgagccctgc
cgcgcgagggttttaaaTTTCGCGCTCTGagccggcgaggagcag
attggtgaatccggggcccTTTCGCGGCTTtcaaatgagccctgc

Свойство

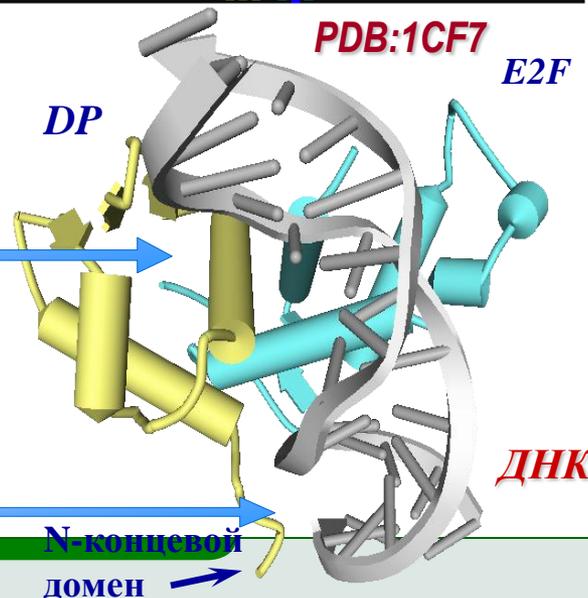
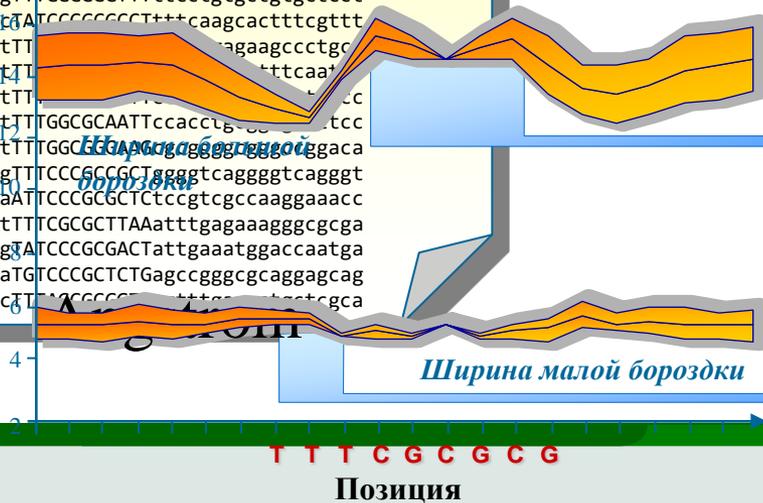
- 1 Twist (averaged)
- 2 Rise (averaged)
- 3 Bend (averaged)
- 4 Tip (averaged)
- 5 Inclination (averaged)
- 6 Major groove Width
- 7 Major groove Depth
- 8 Minor groove Width
- 9 Minor groove Depth
- 10 Roll (Free DNA)
- 11 Twist (Free DNA)
- 12 Tilt (Free DNA)
- 13 Slide (Free DNA)
- 14 Roll (DNA/protein)
- 15 Twist (DNA/protein)
- 16 Tilt (DNA/protein)
- 17 Slide (DNA/protein)
- 18 Twist (Regressed X-rays)
- 19 Wedge (Regressed X-rays)
- 20 Direction (Regr. X-rays)
- 21 Persistence length
- 22 Melting Temperature
- 23 Nucleos. cont. probability
- 24 Bend towards Major groove
- 25 Bend towards Minor groove
- 26 Twist (Averaged X-ray)
- 27 Tilt (Averaged X-ray)
- 28 Roll (Averaged X-ray)
- 29 Slide (Averaged X-ray)
- 30 Propeller (Averaged X-ray)
- 31 Minor groove size
- 32 Minor groove distance
- 33 Major groove size
- 34 Major groove distance
- 35 Clash strength
- 36 Enthalpy change
- 37 Entropy change
- 38 Free Energy change

позиция

tttcgsgsg



Цвета элементов отражают отклонение среднего значения свойства от его среднего значения для случайных последовательностей. Размер элемента соответствует уровню достоверности консервативности свойства





Исследование сайтов связывания двух представителей класса MADS белков (SRF и MCM1). Анализ сходства и различия выявленных конформационных особенностей

SRF (- 13 посл. ССТФ, 42 п.н.)

MCM1 (61 посл. ССТФ, 25 п.н.)

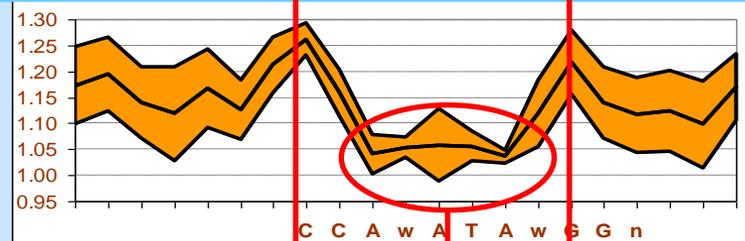
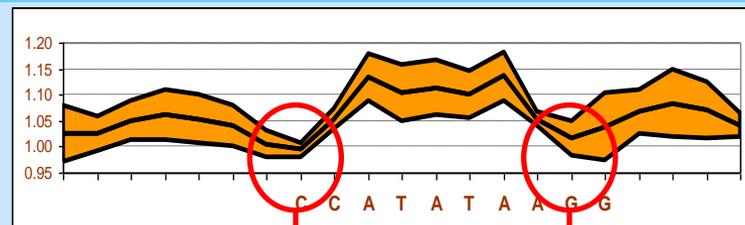
ДНК-связывающие домены этих ТФ обладают сходством – более 70%, их ДСД содержит консервативный район длиной 56 а.о., называемый MADS-бокс.

Изгибная жесткость:

■ в сторону большой бороздки

■ в сторону малой бороздки

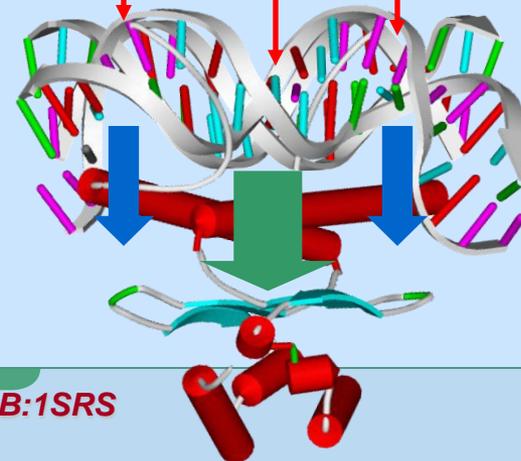
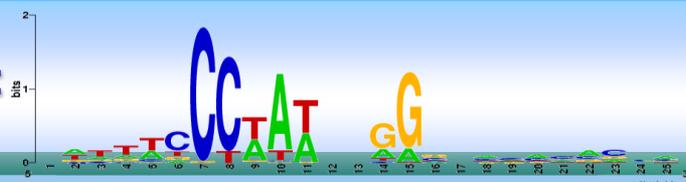
SRF



SRF:



MCM1:

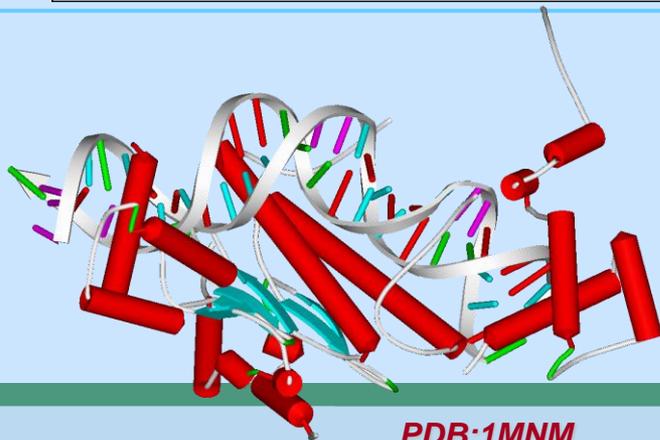
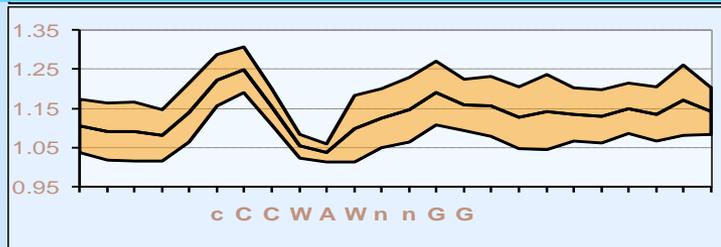
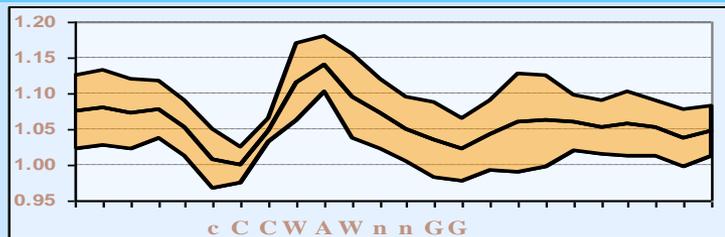
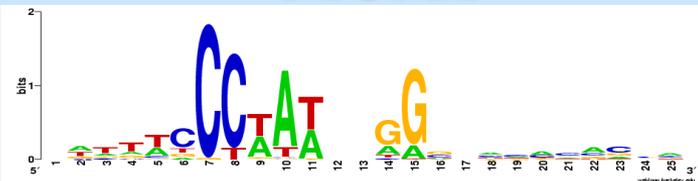


PDB:1SRS



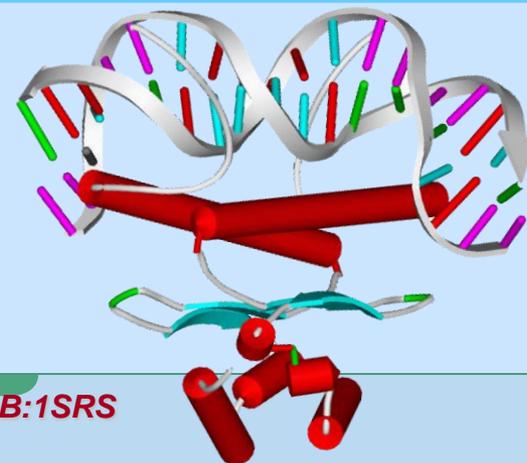
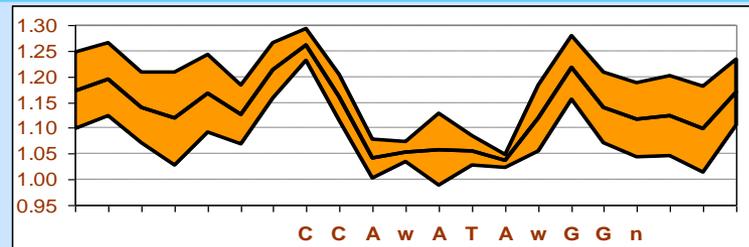
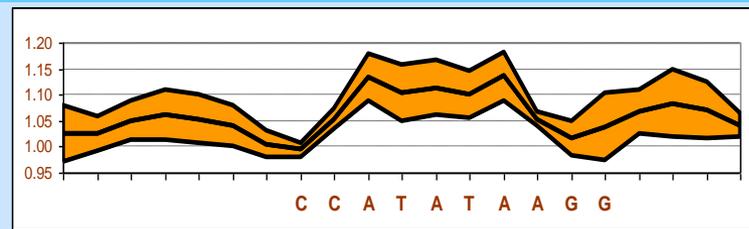
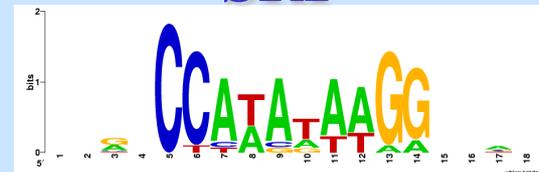
Исследование сайтов связывания двух представителей класса MADS белков (SRF и MCM1). Анализ сходства и различия выявленных конформационных особенностей

MCM1



PDB:1MNM

SRF



PDB:1SRS

Изгибная
жесткость:

■ в сторону
большой
бороздки

■ в сторону
малой
бороздки



Распознавание сайтов связывания транскрипционных факторов по их значимым конформационным и физико-химическим характеристикам

Вероятность P_{il} i -го свойства в позиции l анализируемой последовательности принять требуемое для функционирования значение F_{il} , характерное для выборки сайтов данного типа при значении F_{il} соответствует распределению Гаусса

$$P_{il} = \frac{1}{\sqrt{2\pi}(\sigma_{F_{il}} + \sigma_{err}^i)} \exp \left[-\frac{1}{2} \left[\frac{(\overline{F_{il}} - F_{il})}{(\sigma_{F_{il}} + \sigma_{err}^i)} \right]^2 \right]$$

Величина конформационного сходства P_{Σ} - это сумма всех P_{il} для всех позиций l и всех свойств i :

$$P_{\Sigma} = \frac{\sum_{i=0, l=0}^{I, L} \delta_F P_{il}}{\sum_{i=0, l=0}^{I, L} \frac{\delta_{il}}{\sqrt{2\pi}(\sigma_{F_{il}} + \sigma_{err}^i)}}$$

где $\delta_{il} = 1$, если $\sigma_{F_{il}}^2$ значимо мало, иначе $\delta_{il} = 0$, σ_{err}^i - поправочный параметр

```
cacaaTTTCGCGCCAAActt
ggctcTTTCGCGCCAAAaag
aaggaTTTGCGCGGTAAaag
cagatTTTCGCGCGCTTtgg
cagatTTTCGCGCGCTTtgg
ttcttTTTCGCGCCAGcag
tccttTTTGCCGCGAAAgag
aaggaTTTGCGCGGTAAaag
cgggcTTTGCGGGAGGtgg
tcggaTTTGCGCGAGCgcg
cacgcTATCCCGCGCTttt
acaatTTTCGCGCGTTtta
tcagtTTTCCCGCGAAAatg
ttaaTGTCCTCTGagc
```

$\overline{F_{il}}$

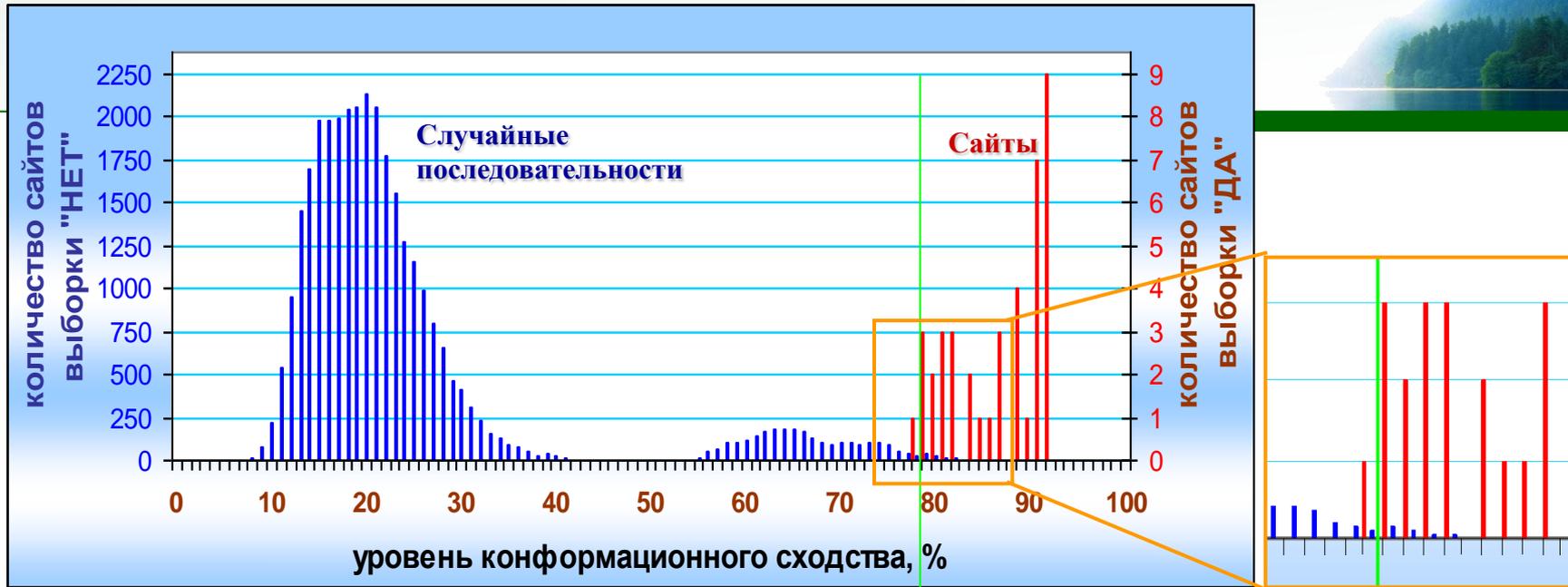
$\sigma_{F_{il}}$

?

F_{il}

анализируемая
последовательность

ggactcctgcctgtccctgcgcatgacagccctcgatgtct



Ошибка I рода рассчитывается с применением подхода Jack-knife. Процедура включает в себя последовательное распознавание одной последовательности из обучающей выборки на основе обучения по выборке, не включающей распознаваемый сайт.

Ошибки II рода рассчитывается путем распознавания сайтов в последовательности длиной 1000000 п.о., сгенерированной путем многократного случайного перемешивания нуклеотидов обучающей выборки

Таблица ошибок распознавания ССТФ E2F/DP методом SITECON

Порог конформационного сходства	77%	78%	79%
Ошибка I рода (недопредсказание)	32.5%	35%	45%
Ошибка II рода (перепредсказание)	$2.1 \cdot 10^{-4}$ (1/4760)	$1.25 \cdot 10^{-4}$ (1/8000)	$7.5 \cdot 10^{-5}$ (1/13300)



Распознавание сайтов связывания транскрипционных факторов по их значимым конформационным и физико-химическим характеристикам

$$\overline{F}_{il} = \frac{1}{N} \sum_{k=1}^N F_{ikl} \quad \sigma_{F_{il}}^2 = \frac{1}{N-1} \sum_{k=1}^N (F_{ikl} - \overline{F}_{il})^2$$

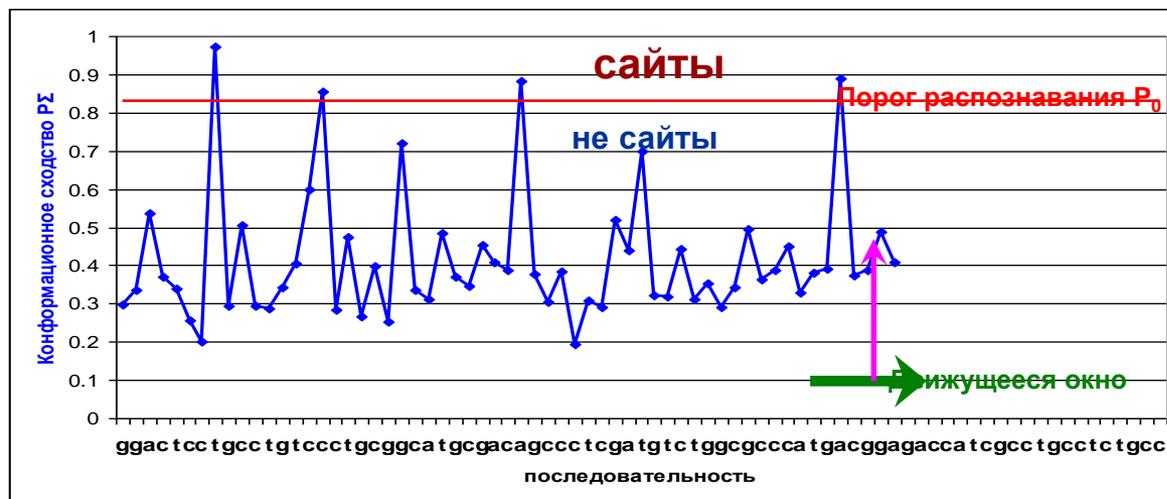
Величина конформационного сходства $P_{\Sigma} = \sum_{i=0, l=0}^{I, L} \delta_{il} \frac{1}{\sqrt{2\pi}(\sigma_{F_{il}} + \sigma_{err})} \exp(-((\overline{F}_{il} - F_{il}) / (\sigma_{F_{il}} + \sigma_{err}))^2)$

где $\delta_{il} = 1$, если $\sigma_{F_{il}}$ значимо мало, иначе $\delta_{il} = 0$

$$\sum_{i=0, l=0}^{I, L} \delta_{il}$$

```

сасааТТТСГСГССАААстт
ggctcТТТСГСГССАААаг
аагаТТТГСГСГТААаг
сгаТТТСГСГСГТТгг
сгаТТТСГСГСГТТгг
ттттТТТСГСГССАГсг
тсттТТТСГСГССААаг
аагаТТТГСГСГТААаг
сгггсТТТГСГССААаг
тсггаТТТГСГСГАГсг
сгсгТТТСГСГСГТТт
ааатТТТСГСГСГТТта
тсггТТТСГСГССААаг
ттааТГТСССГСГТГсг
ггггГТГСГСГТГсг
ггттТТТСГСГССААаг
тсстаТТТСГСГССААаг
тгттТТТСГСГССААаг
  
```



ggactcctgctgtccctgsgcatgagagccctgatgtctggcgccatgacg



Выбор порога распознавания сайтов связывания SF-1

SF-1 (Steroidogenic factor 1)

является одним из ключевых регуляторов экспрессии генов системы стероидогенеза млекопитающих

Нуклеотидные последовательности 54 сайтов связывания транскрипционного фактора SF-1 длиной 93 п.н. были взяты из базы данных TRRD

Для выбора порога распознавания ССТФ SF-1 было проведено его распознавание в промоторных районах 5-ти функциональных групп генов.

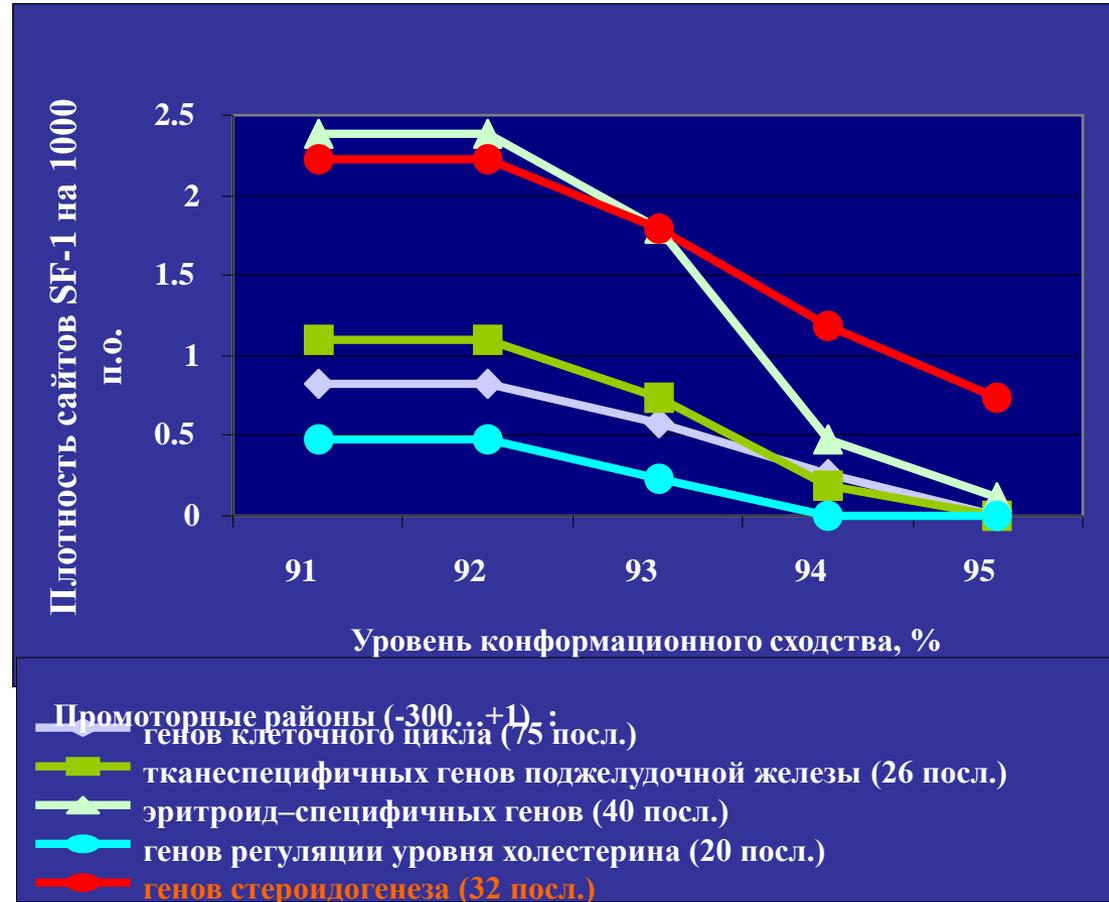


Таблица ошибок распознавания ССТФ SF-1 методом SITECON

Уровень конф. сходства	92%	93%	94%	95%
Недопредсказание	0.30	0.39	0.56	0.70
Перепредсказание	7.31E-04 (1/1368)	5.22E-04 (1/1915)	2.23E-04 (1/4484)	6.97E-05 (1/14347)



Распознавание сайтов связывания SF-1

а) Исследованы промоторные районы (-500...+1) 32-х генов стероидогенеза, где экспериментальный поиск сайтов связывания SF-1 еще не проводился. **Выявлено 15** новых сайтов связывания SF-1

б) Исследованы промоторные районах генов уже содержащих экспериментально выявленные сайты. **Выявлено** дополнительно **3** новых сайта связывания.

	Название гена (вид)	Позиция*	R**	Экспериментальное подтверждение
а				
1	<i>Cyp17</i> (Mouse)	-283	0.944	+
2	<i>Cyp17</i> (Mouse)	-49	0.949	+
3	<i>Ad</i> (Bovine)	-428	0.962	+
4	<i>Cyp11B1</i> (Guinea pig)	-126	0.945	+
5	<i>Cyp11B3</i> (Rat)	-309	0.945	+
6	<i>Cyp11B1</i> (Sheep)	-337	0.947	+
7	<i>Oxt</i> (Mouse)	-164	0.966	+
8	<i>Oxt</i> (Rat)	-167	0.962	+
9	<i>Oxt</i> (Human)	-159	0.961	+
10	<i>Cyp11B2</i> (Rat)	-324	0.951	+
11	<i>HSD3b</i> (Mouse)	-113	0.942	+
12	<i>Ad4BP/SF-1</i> (Mouse)	-224	0.952	+
13	<i>CYP17</i> (Porcine)	-51	0.946	+
14	<i>HSD17BI</i> (Rat)	-84	0.941	+
15	<i>LH beta</i> (Porcine)	-114	0.959	+
б				
16	<i>CYP17</i> (Human)	-44	0.944	+
17	<i>CYP17</i> (Rat)	-309	0.944	+
18	<i>LHbeta</i> (Ss)	-58	0.928	+

100
%

* Позиция указана относительно старта транскрипции

** Уровень конформационного сходства с известными сайтами связывания SREBP, оцененный методом SITECON

Экспериментальная проверка проводилась методом задержки комплекса ДНК/белок в геле (EMSA) с использованием экстракта ядер клеток семенников крыс и специфических антител к SF-1 в лаборатории регуляции экспрессии генов ИЦиГ СО РАН.



ТФ SREBP участвуют в регуляции экспрессии генов липидного метаболизма и биосинтеза холестерина. 38 нуклеотидных последовательностей ССТФ SREBP SRE-типа длиной 40 п.н. были взяты из БД TRRD.

Исследованы промоторные районы (-500...+1) 46 генов липидного метаболизма.

Выявлено 50 новых сайтов связывания SREBP (уровень конформационного сходства >73%)

15 из них было отобрано для экспериментальной проверки *in vitro*

Распределение значений конформационного сходства для положительной и отрицательной выборок



Таблица ошибок распознавания ССТФ SREBP методом SITECON

Уровень конф. сходства	73%	74%	75%	76%
Недопредсказание	0.53	0.58	0.61	0.63
Перепредсказание	7.55E-04 (1/1324)	5.65E-04 (1/1769)	4.15E-04 (1/2409)	3.20E-04 (1/3124)



Распознавание и результаты экспериментальной проверки сайтов связывания SREBP

Название гена (вид)	Позиция*	Ориентация	R**	Последовательность ССТФ	Экспериментальное подтверждение
А) Потенциальные SRE, имеющие уровень конформационного сходства более 73%					
1. MSR (человек)	-439	+	0.750	GGACCATGAGGTCAGGGGATCAAGACCATC	+
2. MSR (человек)	-303	+	0.778	GAGATCACACCACTGCACTCCACCSTGGTG	+
3. MSR (человек)	-164	-	0.796	AGATAACTCACTCTTCACCCCATTTAGGCC	+
4. SCAP (человек)	-311	+	0.777	TCCTCCCCCGTCACACCAGGGGCCCGCGGA	+
5. FAS (цыпленок)	-484	+	0.756	GCCTAGCAACACCCCTCACGGGCTCCGCATT	+
6. FAS (цыпленок)	-63	+	0.749	GGAGTAAACCGTCAGCCCATGTGGTGGCCG	+
7. ABCG1 (человек)	-558	-	0.732	CCCTTCCCCGCCACCCCCACCCCGACGGTGT	+
8. INSIG1 (человек)	-385	+	0.736	GGCCCGAAAAAATCACCCAAGGATGCAAGC	+
9. INSIG1 (человек)	-99	+	0.769	GCGGGGCAAGCTCAGGCCACGCCCTGGGC	+
10. LXRA (человек)	-81	-	0.780	CTGCCGAGGGGTCAGGCCACCCCTCCTTCC	+
11. CYP4A6 (кролик)	-641	-	0.783	ACAAGCATTTCCACCCCACTGCCCTCAACT	+
12. LPL (мышь)	-773	+	0.805	CCACACTCCTAGCACCCCATATCCCTCGC	+
13. CPT2 (человек)	-789	+	0.731	CAATCCGAGGTCACCTAACTAGATAACAAA	+
14. IBABP (мышь)	-67	-	0.863	AGCCCCAAAGTCACCCCACTTCTCAGTCT	+
15. IBABP (человек)	-66	-	0.881	CAGCCCCAAGTCACCCCACTTCTTCTCCC	+
В) Потенциальные SRE, имеющие уровень конформационного сходства менее 73%					
16. LXRA (мышь)	-222	+	0.729	TCCTGTCCCCCTCAGGGCATCCACACACAA	+
17. PLCG1 (человек)	-36	+	0.728	GGGGTGGGGCCTCAGGGCAGCCCCGCCCG	+
18. RXRG (мышь)	+25	+	0.727	GGAAAGAGTCGCTCACAGCAGCCCACCCAG	+
19. SRB1 (человек)	-769	+	0.725	GGGGGCCCGCCTCAGGGTGTGGCATCCAG	-
20. MSR (человек)	-323	-	0.704	AGTGGTGTGATCTCAGTCACTGCAACCTC	+
21. LXRA (человек)	-707	-	0.717	CAACTCAGGTGGCAGCCAGGATGGACCTG	+
22. LXRA (человек)	-894	-	0.717	CCAGCCCTGTTTCAGGGGTGTTGGGGGAGA	-
23. FAS (цыпленок)	-429	-	0.716	AGTGCCGGCCGGCACCCAACCGGGGGCG	-
24. CAV1 (мышь)	-239	-	0.715	GACCCACAACTCACCCTGCTCCGTGGTCC	-
25. SIP (человек)	-101	+	0.713	GAACACGCCTGGGCACTCCATTCGGGGCTG	-
26. ABCG1 (человек)	-390	+	0.712	CCCCGAGCCTCACCCGTGCTGGCCAGCC	+
27. INSIG1 (человек)	-47	-	0.712	GCGGGGCCCGCATCACGTGAGCCTGCCCGG	-
28. FAS (цыпленок)	-129	+	0.705	CAGGGGCTGGCATCACCCGGGGCCGGCTC	+
29. CAV1 (мышь)	-501	-	0.699	TGCCAAGACCTGTCAACCCCAACCCCATGC	-
30. SCAP (человек)	-60	+	0.694	CAGGCGCCCTGAGCACCCCGGTGACGGTAC	+

100
%

* Позиция указана относительно старта транскрипции

** Уровень конформационного сходства с известными сайтами связывания SREBP, оцененный методом SITECON

Экспериментальная проверка проводилась методом задержки комплекса ДНК/белок в геле (EMSA) с использованием рекомбинантного белка SREBP1-а человека в лаборатории регуляции экспрессии генов ИЦиГ СО РАН.



Спасибо за внимание!



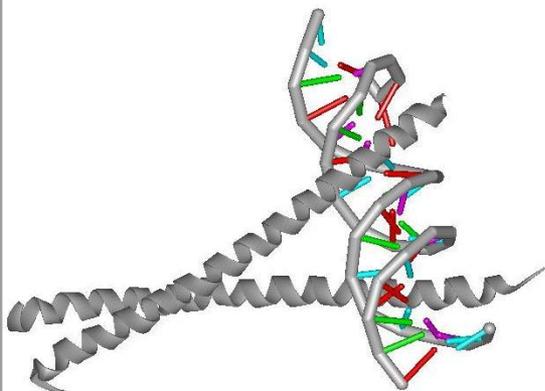


ДНК – связывающие домены

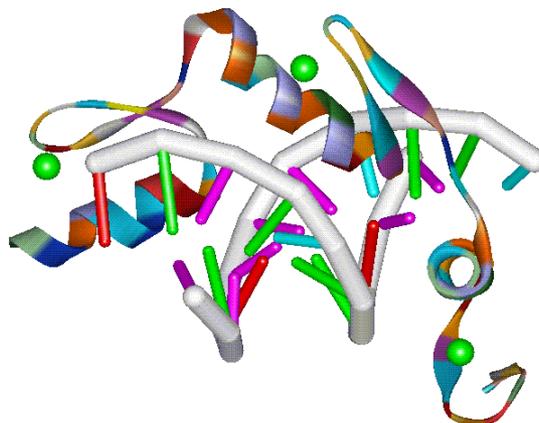
ОСНОВНЫЕ ТИПЫ ДНК - СВЯЗЫВАЮЩИХ ДОМЕНОВ



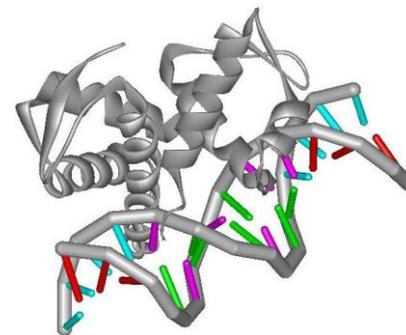
1. bZIP



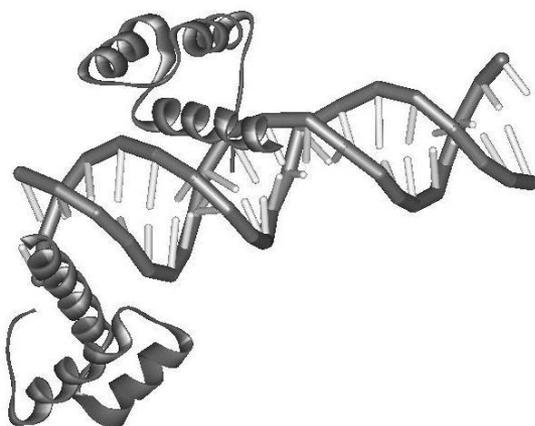
2. Zinc coordinated



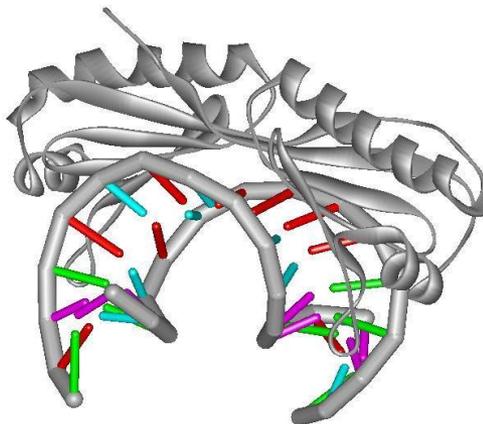
Другие -E2



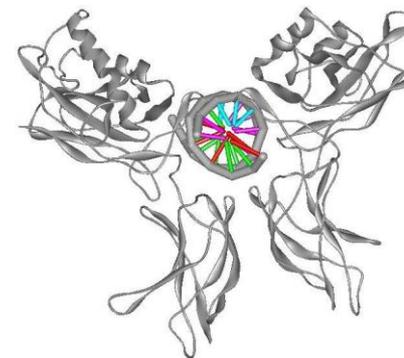
3. Helix-Turn-Helix



4. β -scaffold

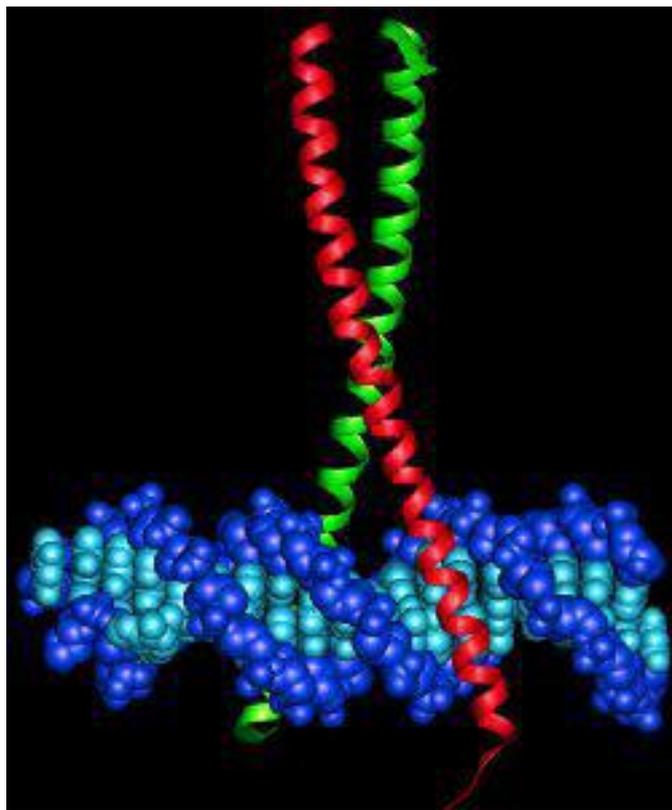


Другие – REL

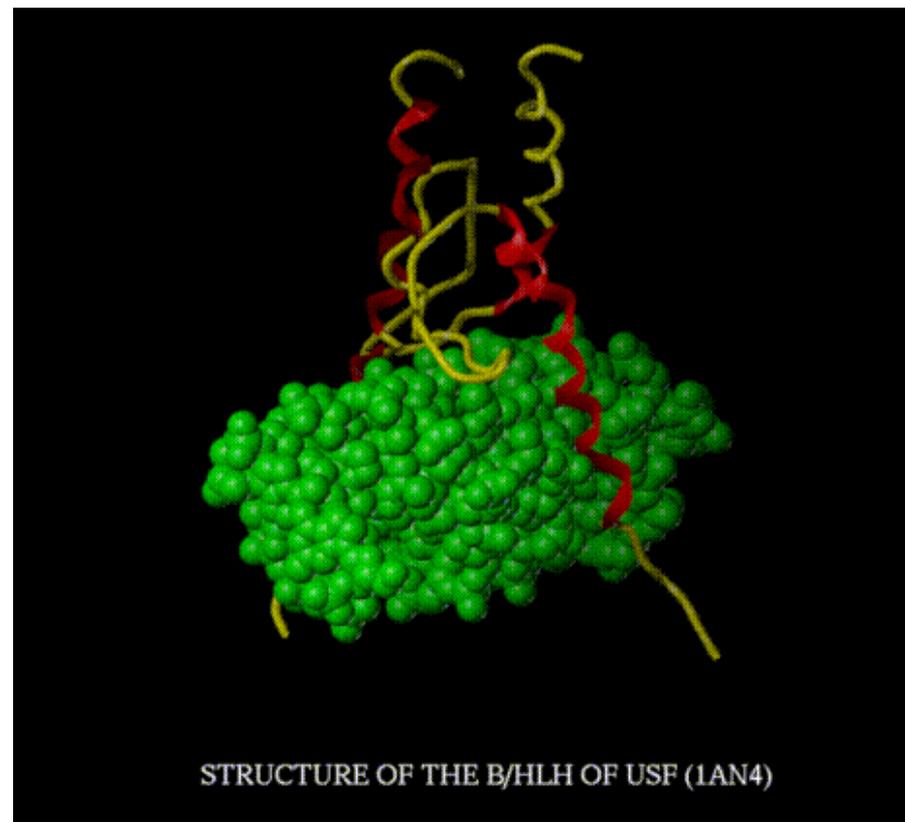




Лейциновая застежка (b-zipper)



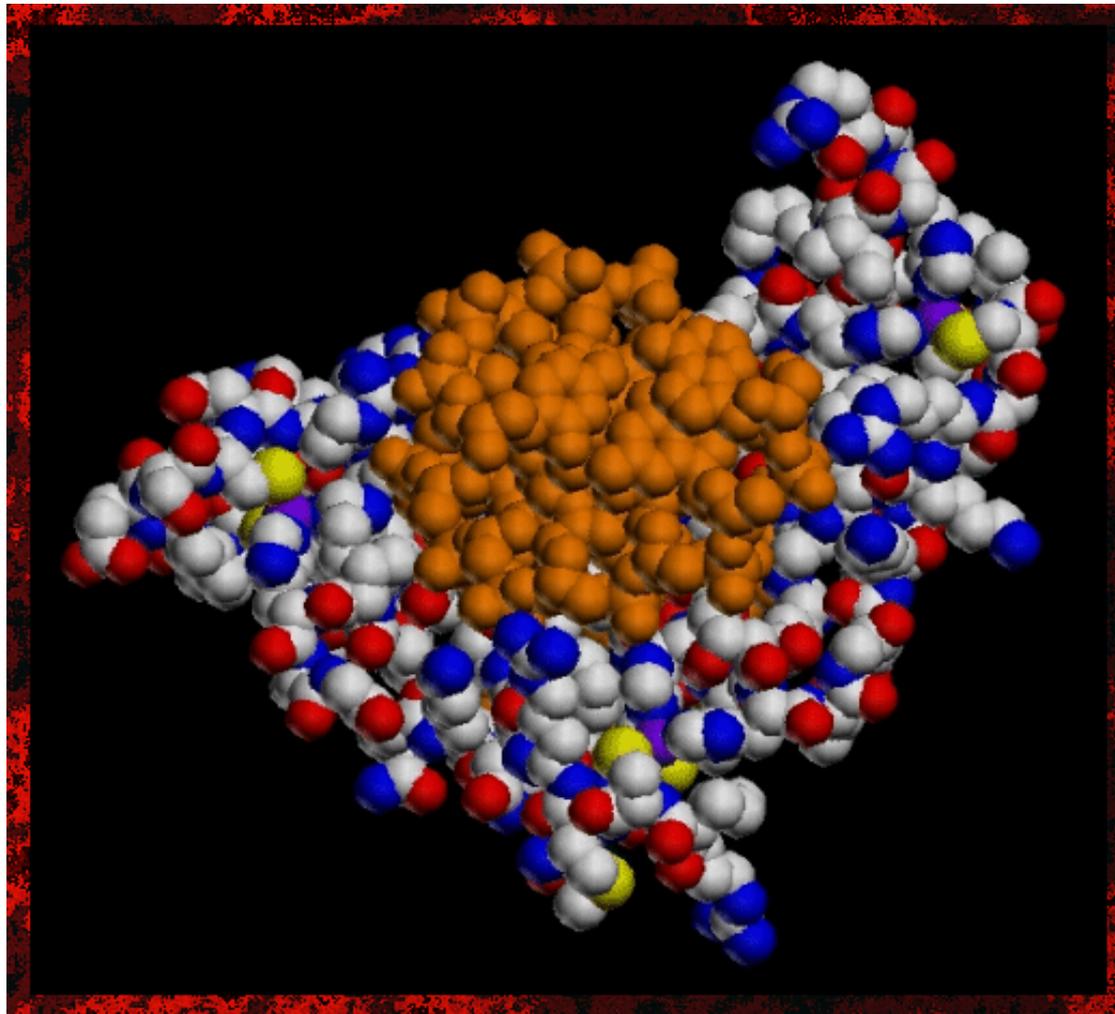
Спираль – петля – спираль (Helix-Loop-Helix)





ДНК – связывающие домены

Домены, координированные атомами цинка (Zn - finger)

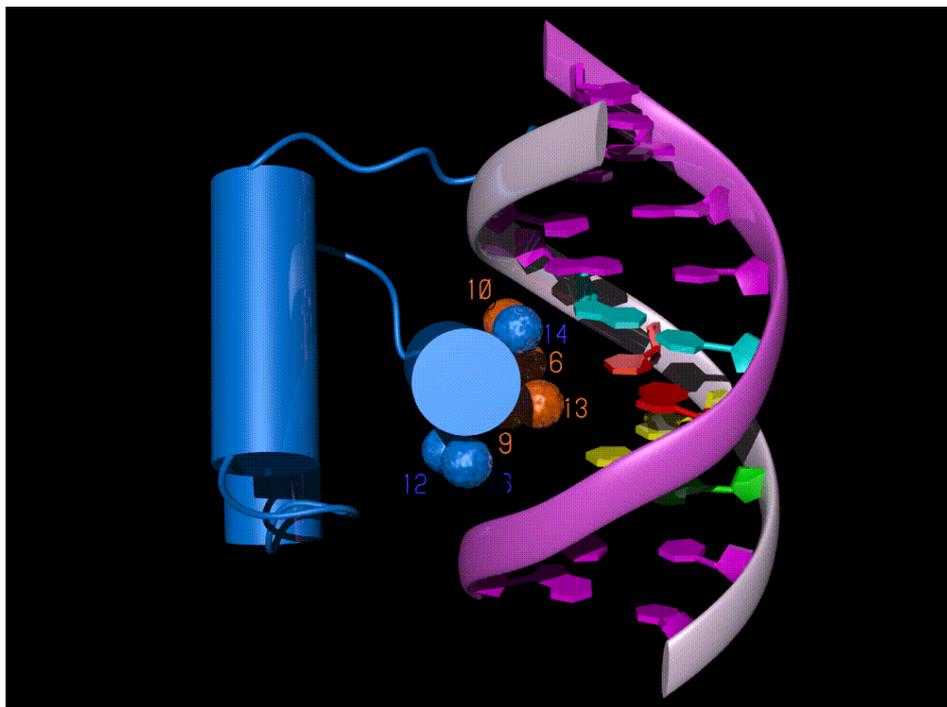




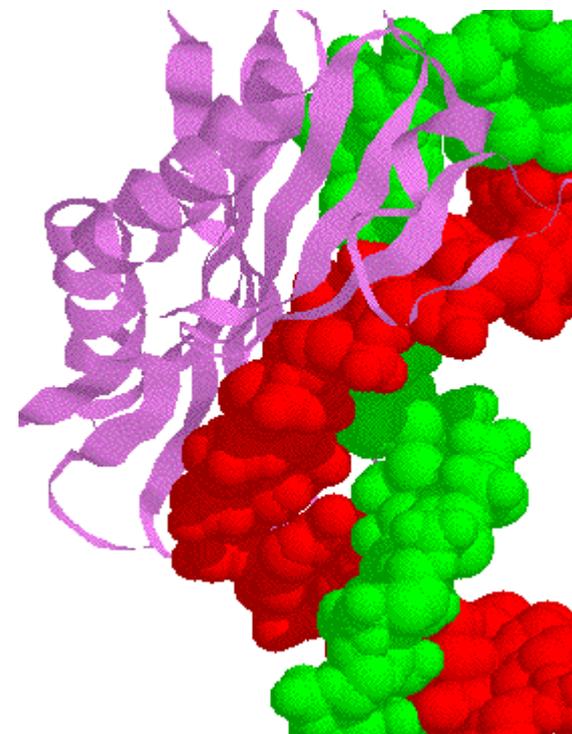
ДНК – связывающие домены



Спираль - поворот - спираль
(Helix - turn - helix)



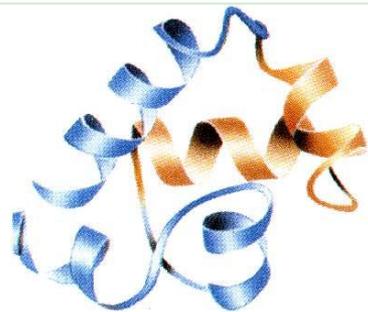
β -scaffold
(TATA - box +
DNA - polymerase)



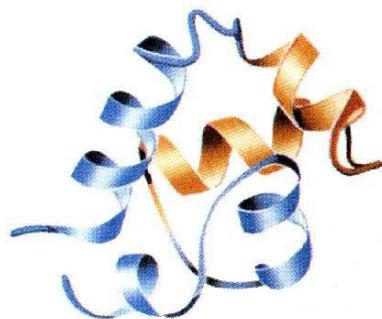


ДНК – СВЯЗЫВАЮЩИЕ ДОМЕНЫ

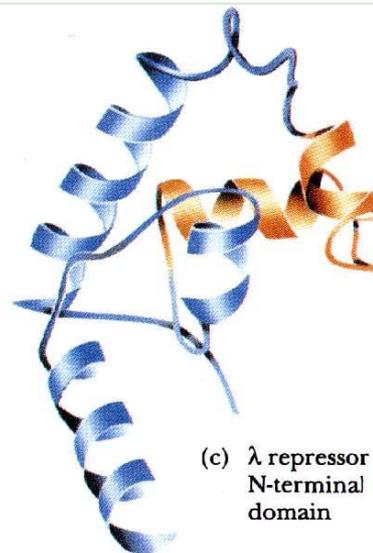
Домены с мотивом Спираль – поворот – спираль (helix – turn - helix)



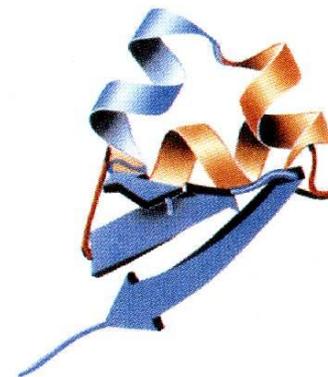
(a) 434 repressor
N-terminal
domain



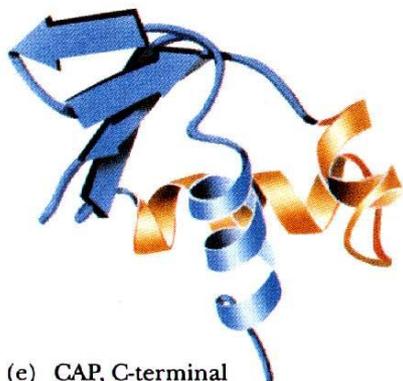
(b) 434 Cro



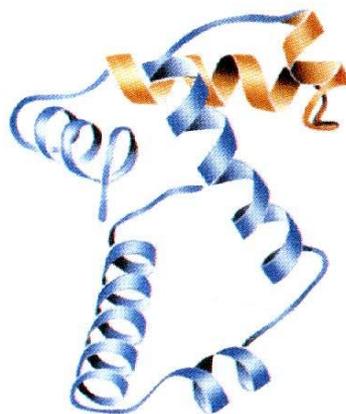
(c) λ repressor
N-terminal
domain



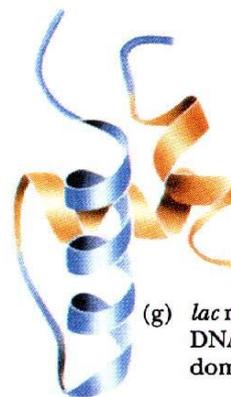
(d) λ Cro



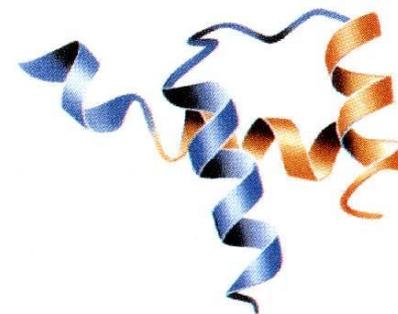
(e) CAP, C-terminal
domain



(f) *trp* repressor
subunit



(g) *lac* repressor
DNA-binding
domain



(h) *Antp* homeodomain