

ВВЕДЕНИЕ В ИНФОРМАЦИОННУЮ БИОЛОГИЮ

Н.А. Колчанов, С.А. Лашин

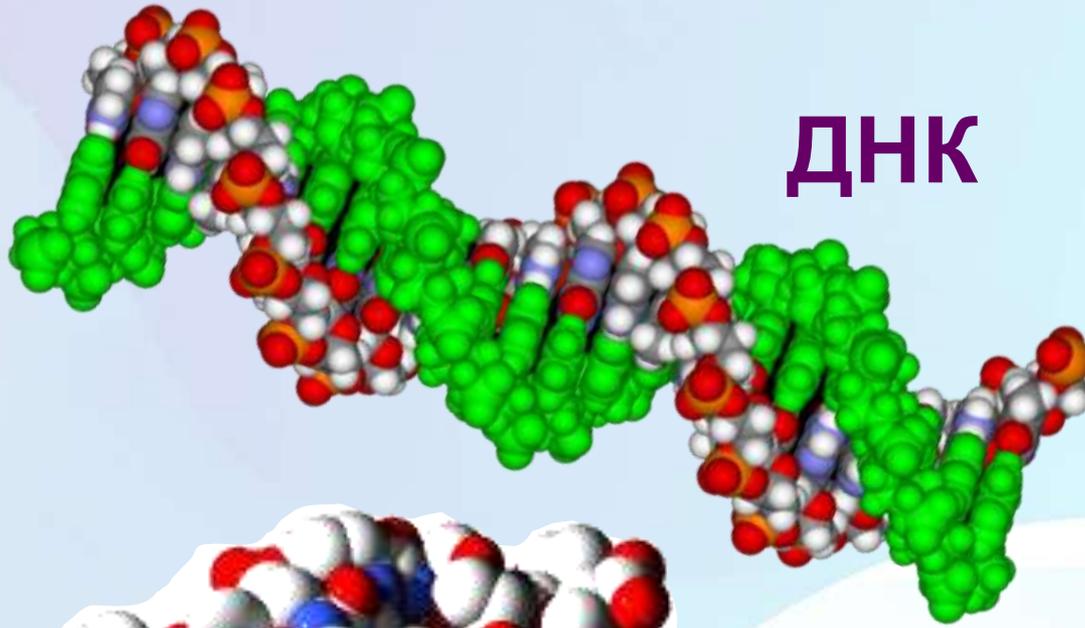
Электронно-лекционный курс
разработан в рамках реализации Программы развития НИУ-НГУ
2012 год

- Курс подготовлен с использованием лекционных материалов, прочитанных в 2008-2012 годах сотрудниками кафедры информационной биологии ФЕН НГУ и Института цитологии и генетики СО РАН:

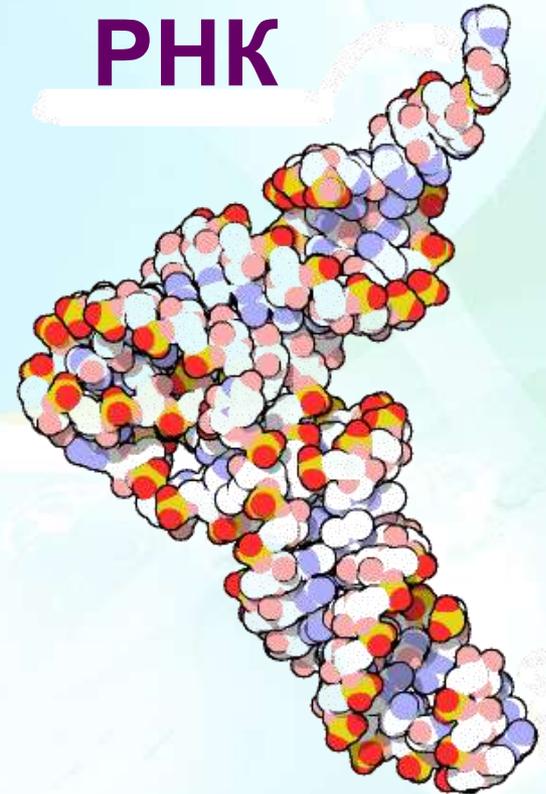
- *Абердин И.Р.*
- *Алемасов Н.А.*
- *Афонников Д.А.*
- *Дорошков А.В.*
- *Иванисенко В.А.*
- *Иляскин А.В.*
- *Колчанов Н.А.*
- *Лашин С.А.*
- *Миронова В.В.*
- *Сормачева И.Д.*
- *Сорокина К.Н.*

Лекция 1
Объекты информационной
биологии –
биологические
самовоспроизводящиеся
системы и их молекулярные
и информационные
компоненты

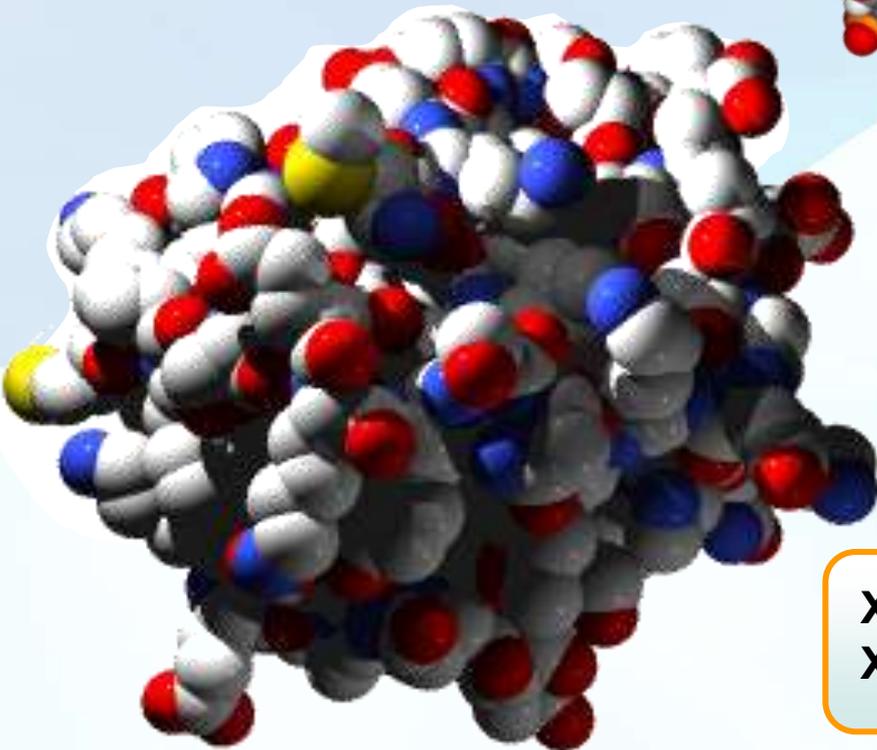
Биологические макромолекулы



ДНК



РНК



Белки

Характерный размер: 0.1 - 3.0 нм;
Характерное время процесса: 0.1 - 100 нс;

РИБСОМА - супрамолекулярный РНК-белковый комплекс



**Характерный размер : 1.0 - 30 нм
Характерное время процесса 10 мс**

Геном - супрамолекулярная ДНК-белковая информационная структура

Хромосомы

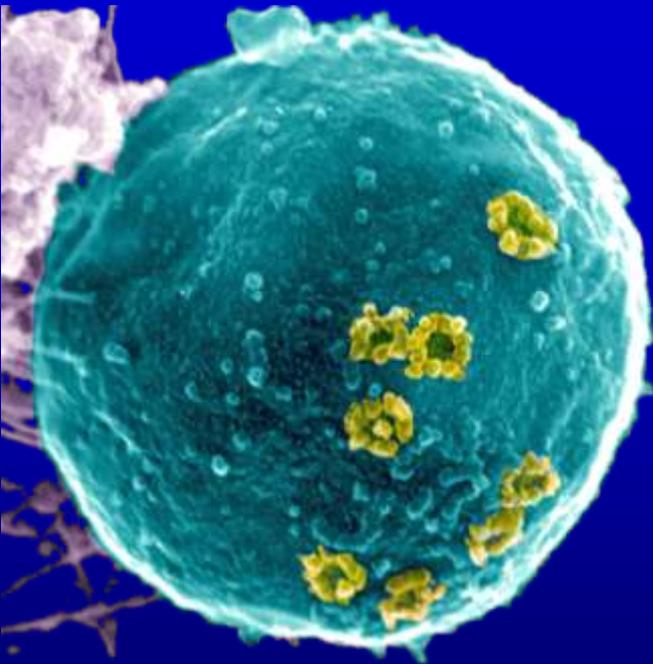


Упаковка ДНК в
хроматине

Характерный размер: 2 – 20 мкм

Ядро

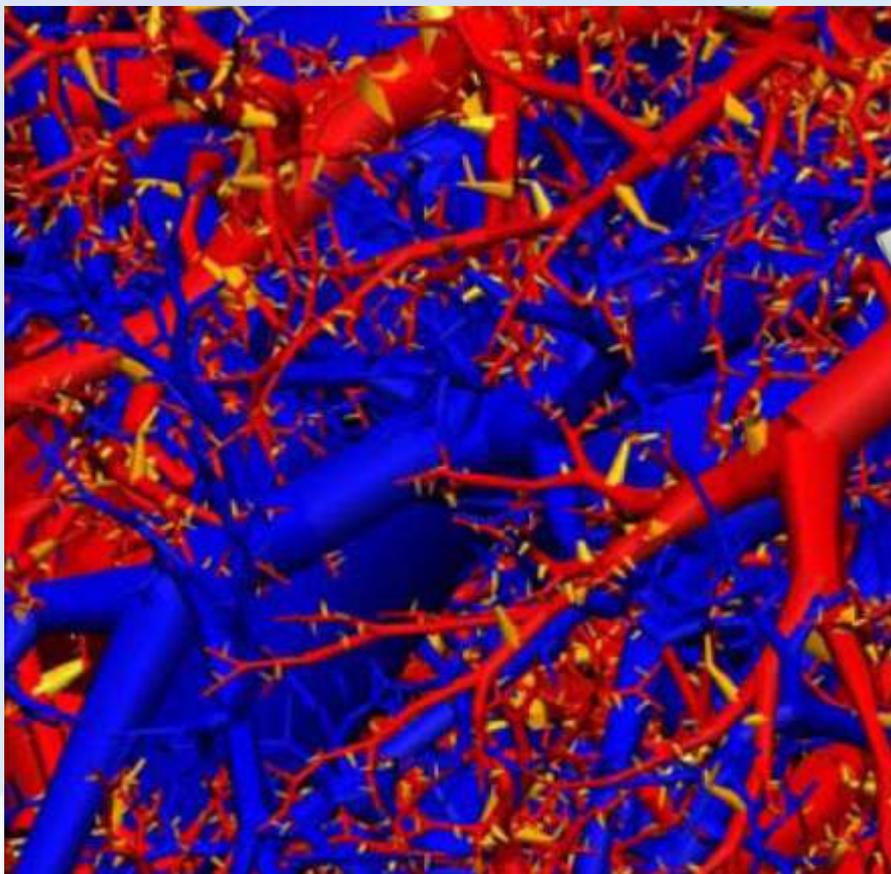
Клетка



Характерный размер 50 - 100 мк

Характерное время процесса 10 мс - 1000 с

Ткань



Характерный размер: 0.001м – 0.01 м
Характерное время процесса: 1 с – 1 ч.

Орган



Характерный размер: 0.001м – 0.01 м
Характерное время процесса: 1 с – 1 ч.

Различия между иерархическими уровнями в живых системах по характерным размерам объектов – 10 порядков

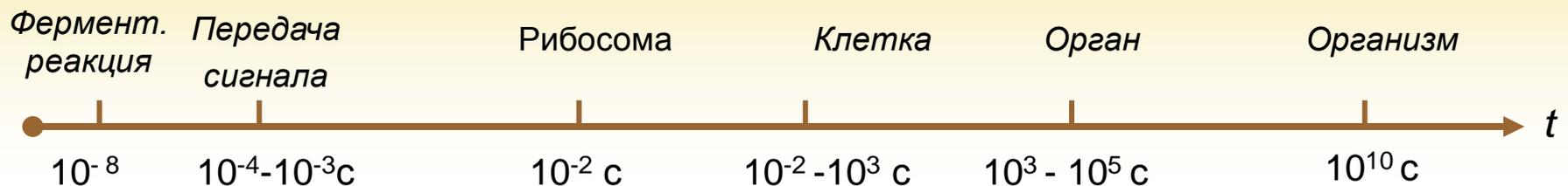


Каждый иерархический уровень организации живых систем контролируется своими регуляторными системами с характерными временами протекания процессов

Сети регуляторных процессов интегрируют выше и ниже расположенные иерархические уровни, обеспечивая целостность организации живых систем

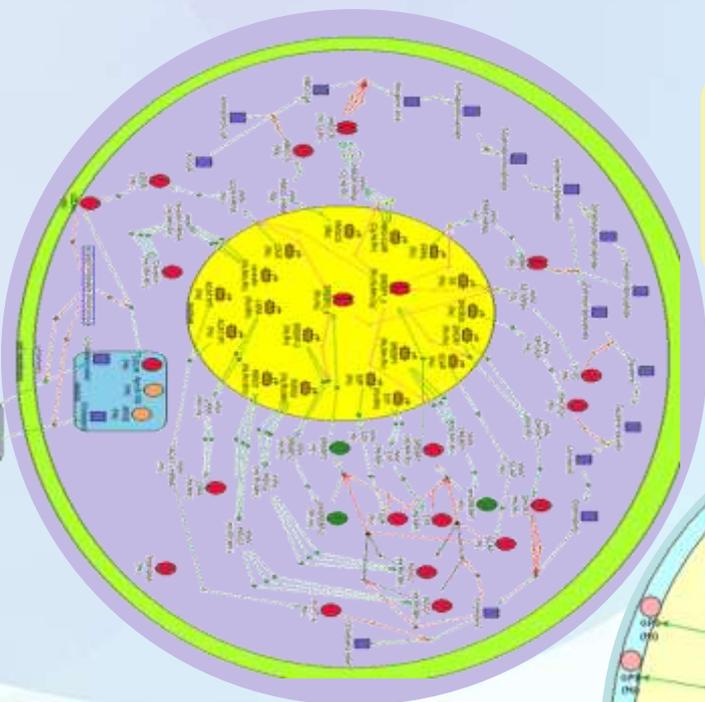


Различия между иерархическими уровнями в живых системах по характерным временам процессов – 18 порядков



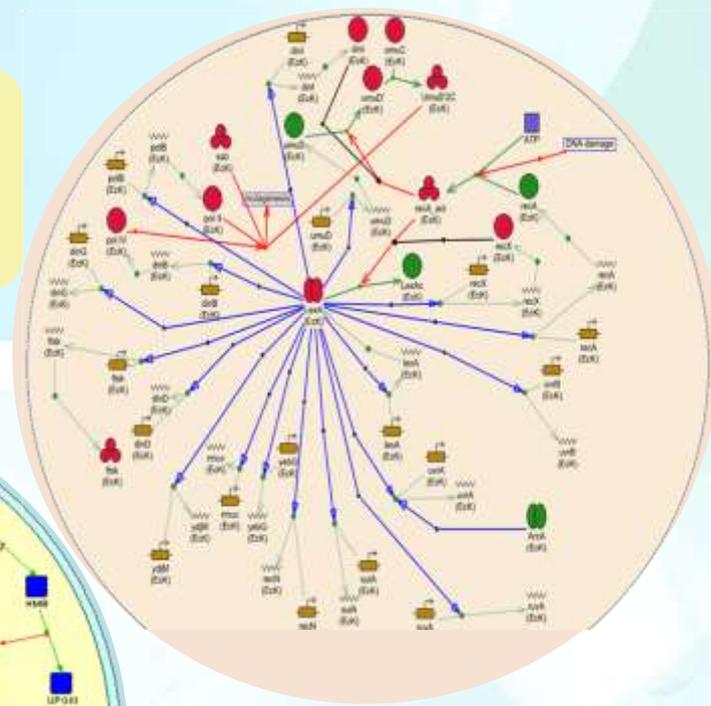
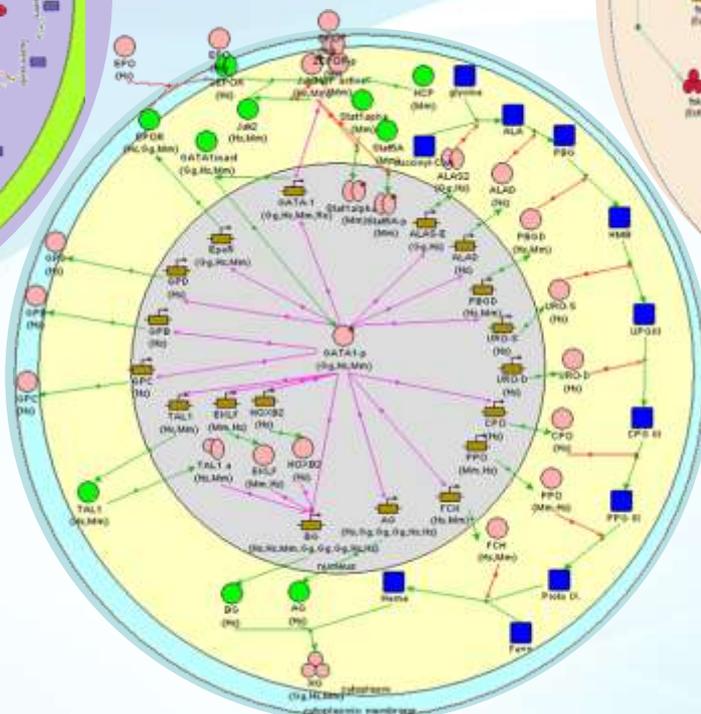
Центральный объект системной биологии - ГЕННЫЕ СЕТИ

группы координировано функционирующих генов, обеспечивающих формирование фенотипических признаков организмов (молекулярных, биохимических, физиологических, морфологических, поведенческих и т.д.)



Генная сеть биосинтеза холестерина

Генная сеть дифференцировки эритроцита



Генная сеть теплового шока

Все фенотипические характеристики организмов формируются в ходе их индивидуального развития на основе информации, закодированной в геномах, и в результате взаимодействия с факторами окружающей среды

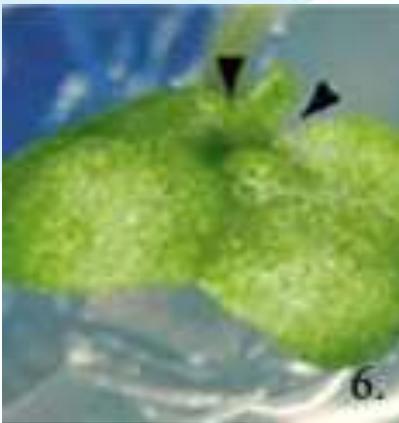
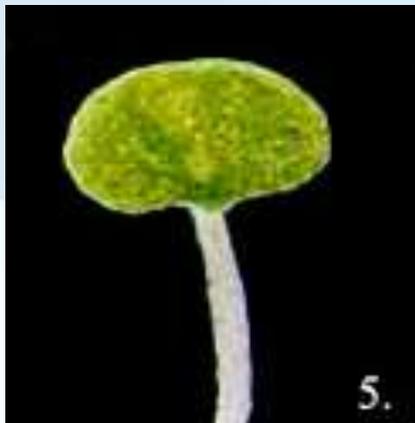
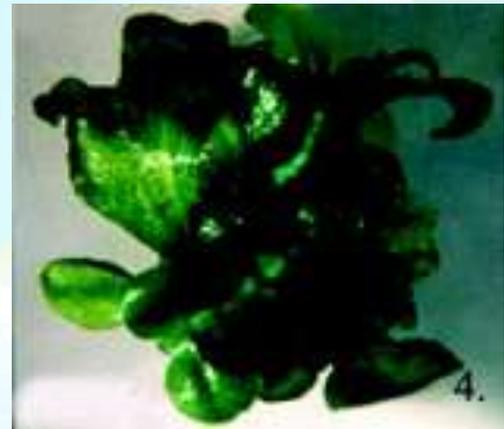
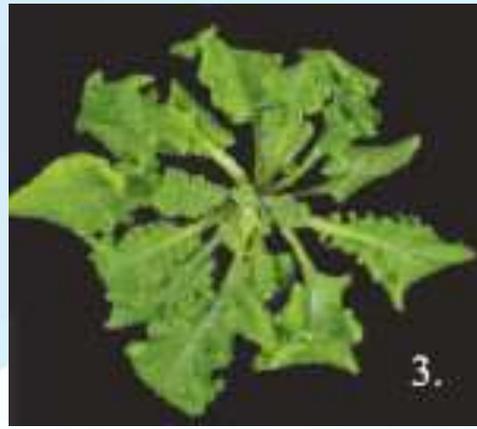
Наследование в ряду поколений



Не наследуется!!!!!!

Уникальная особенность живых систем - минимальные изменения на нижележащих иерархических уровнях могут вызывать глобальные изменения на более высоких уровнях организации

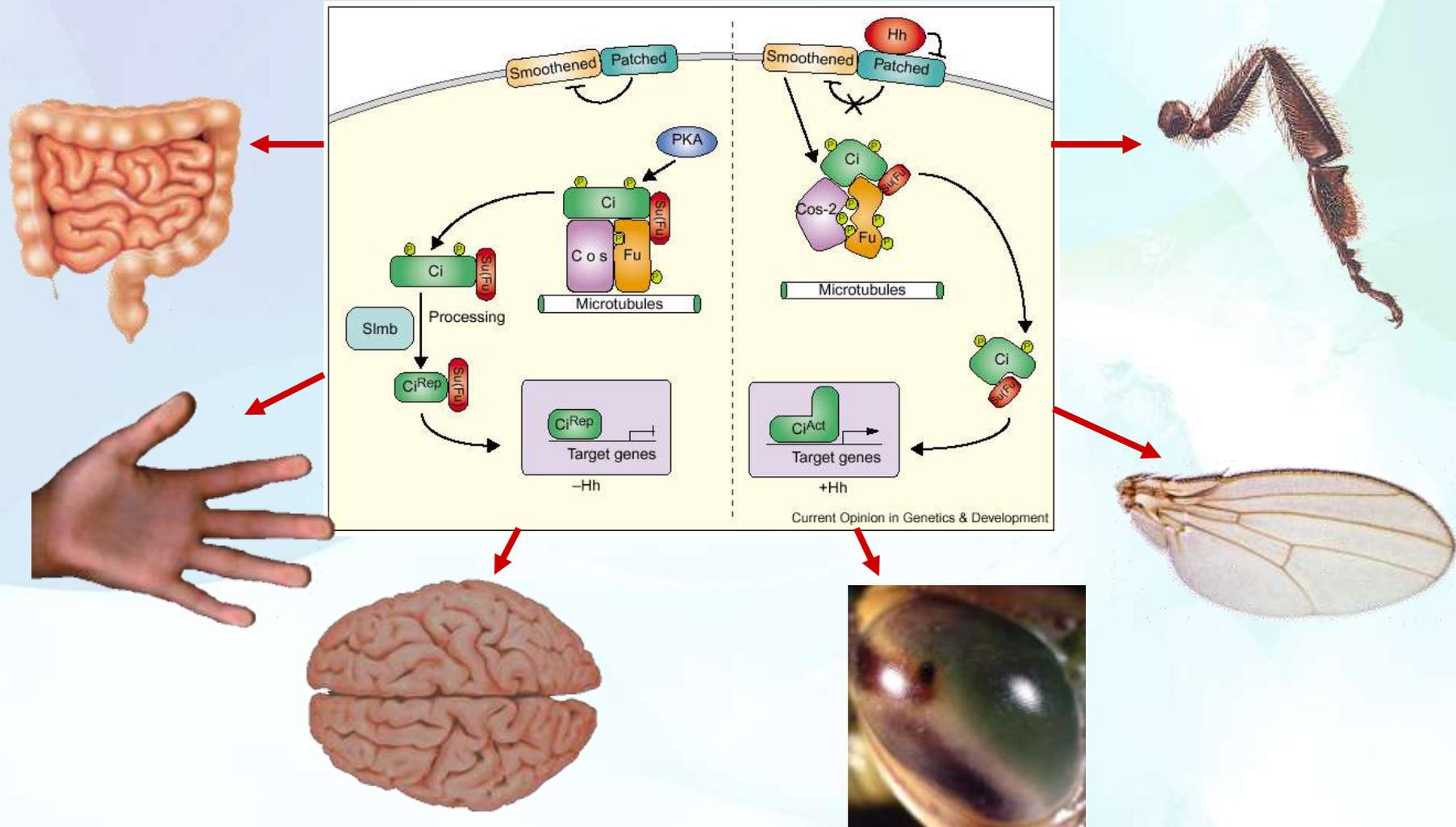
Норма



Влияние мутаций в геноме Arabidopsis thaliana на фенотип взрослого растения



Биологические самовоспроизводящиеся системы и их компоненты формируются в ходе эволюционного процесса.
Hh-каскад передачи сигналов - универсальный молекулярный механизм, вовлеченный в контроль морфогенеза многоклеточных организмов



Экспериментально-вычислительная база системной биологии

Кластер «Системная биология» Новосибирского научного центра СО РАН)

Геномика: автоматический секвенатор



Протеомика: массовый анализ белков и метаболитов



ТРАНСКРИПТОМИКА: производство биочипов высокой плотности и анализ профилей экспрессии генов



Клеточная биология: лазерный сканирующий микроскоп LSM510 META

Высокопроизводительные вычислительные

Прижизненная томография экспериментальных животных

Современная биология стала источником беспрецедентно огромных объемов экспериментальных данных и знаний. Их осмысление и применение невозможны без привлечения современных информационных технологий, эффективных методов компьютерного анализа данных и математического моделирования биологических систем и процессов на различных уровнях организации живой материи: от молекулярно-генетического, включая организменный и заканчивая экосистемным и биосферным.

В ответ на эти вызовы возникла *НОВАЯ наука, называемая *информационной биологией* (*биоинформатикой*).*

Информационная биология

Цель - изучение организации и механизмов развития и функционирования живых систем на основе информации, закодированной в их геномах, в ходе их взаимодействия с окружающей средой.

Описание в базах данных и интеграция огромных объемов гетерогенной экспериментальной информации, характеризующей живые системы на различных уровнях их структурно-функциональной организации

Крупномасштабный анализ экспериментальных данных

Построение математических моделей организации и функционирования живых систем

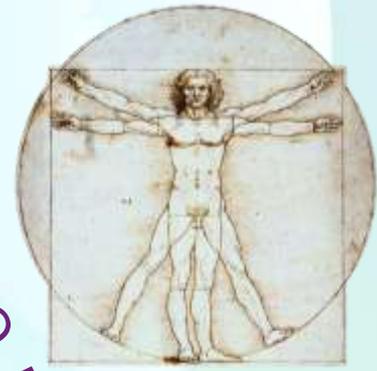
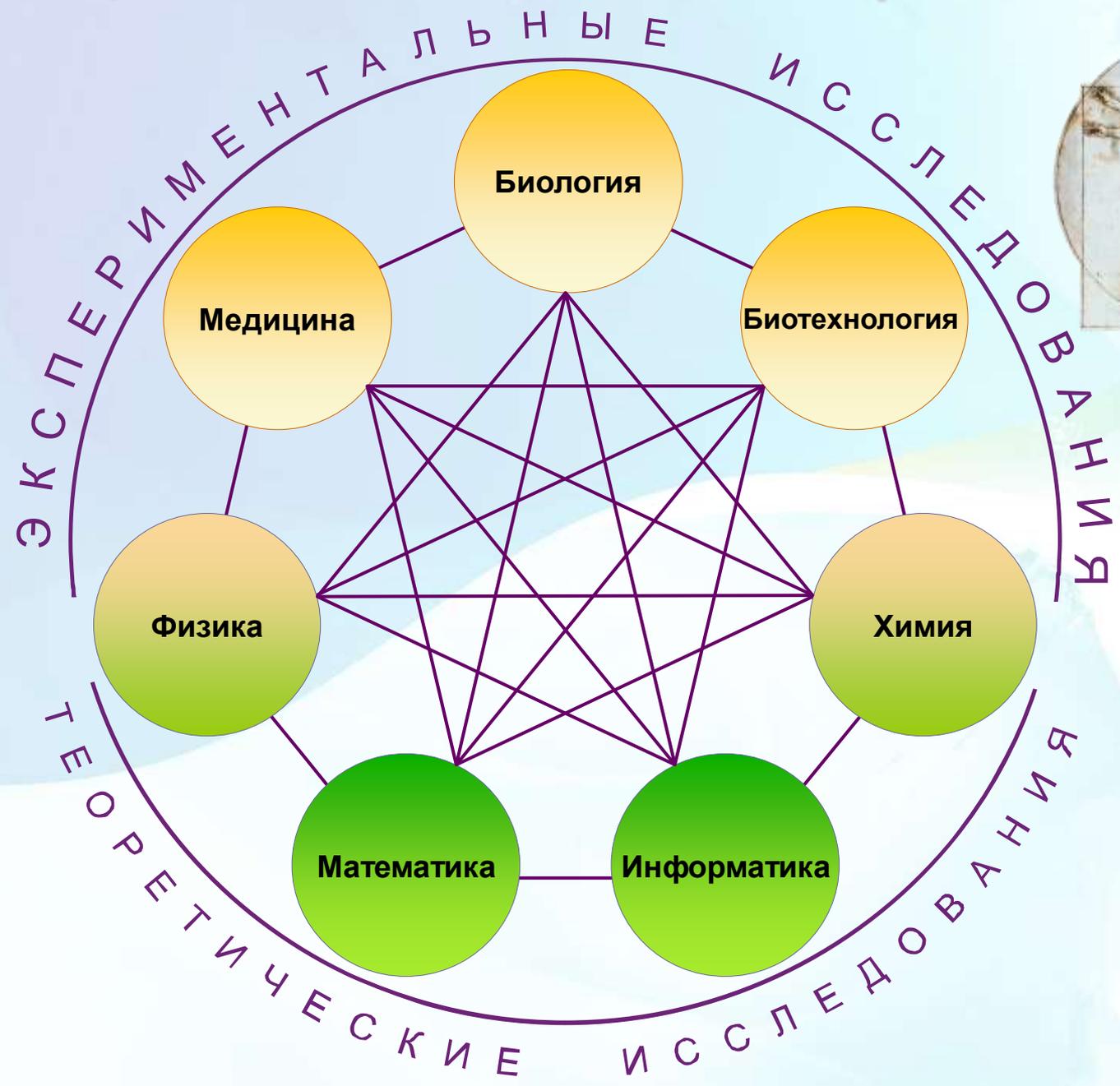
Предсказание новых особенностей организации и функционирования живых систем

Планирование экспериментов по проверке результатов предсказания

Проведение экспериментов и получение новых данных и знаний

ИНФОРМАЦИОННАЯ БИОЛОГИЯ ВОЗНИКЛА, КОГДА ОНА СТАЛА ПРЕДСКАЗАТЕЛЬНОЙ НАУКОЙ

Информационная биология – интегративная наука



ГЕНОМИКА

Экспериментально-компьютерная геномика

(Лаборатория молекулярной и клеточной биологии ИЦиГ СО РАН):

1. крупномасштабное и полное секвенирование геномов,
2. изучение регуляторных районов генов,
3. выявление и анализ полиморфизмов.

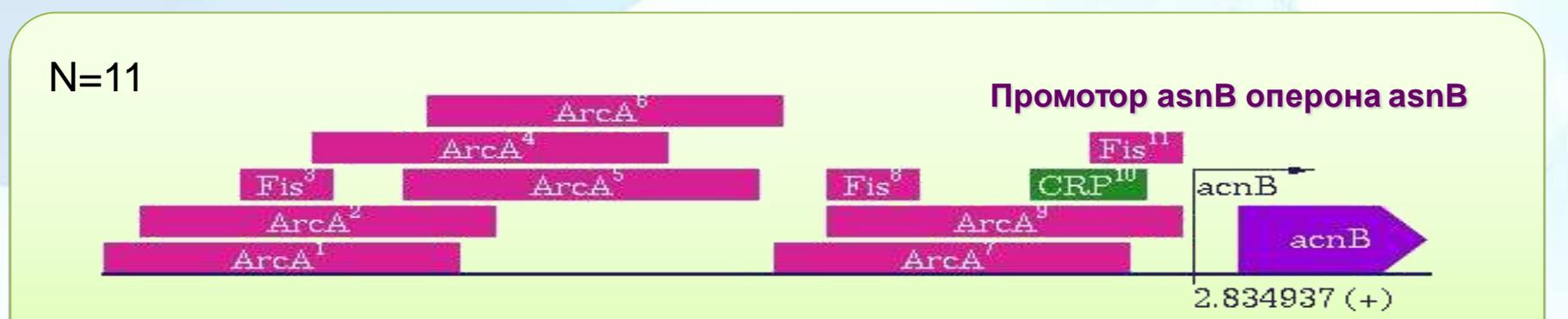
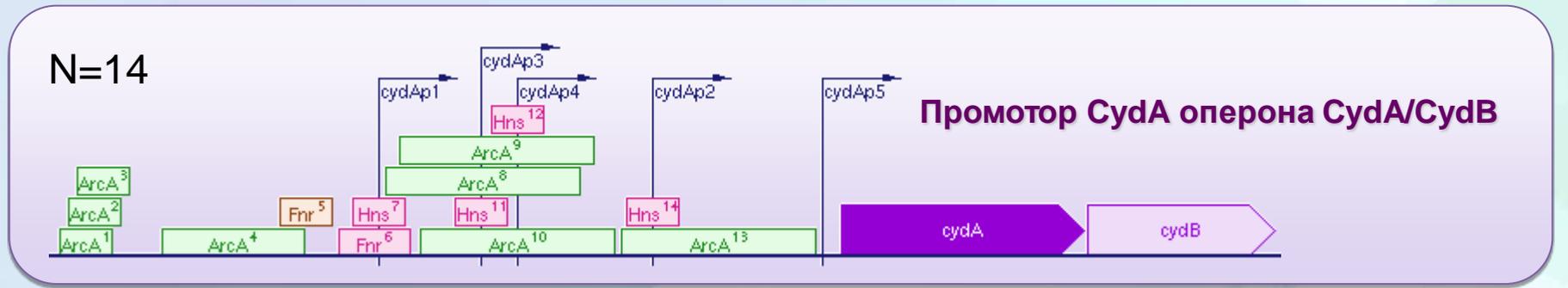
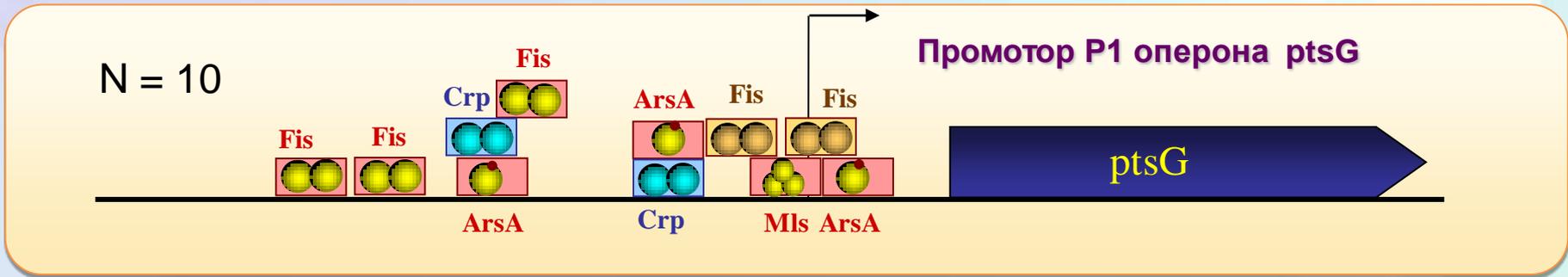
Компьютерная геномика

(Группа компьютерной геномики Отдела системной биологии и Лаборатория молекулярных механизмов экспрессии генов ИЦиГ СО РАН):

1. компьютерная поддержка полногеномных проектов,
2. экспериментано-компьютерное исследование регуляторных районов генов про- и эукариот.



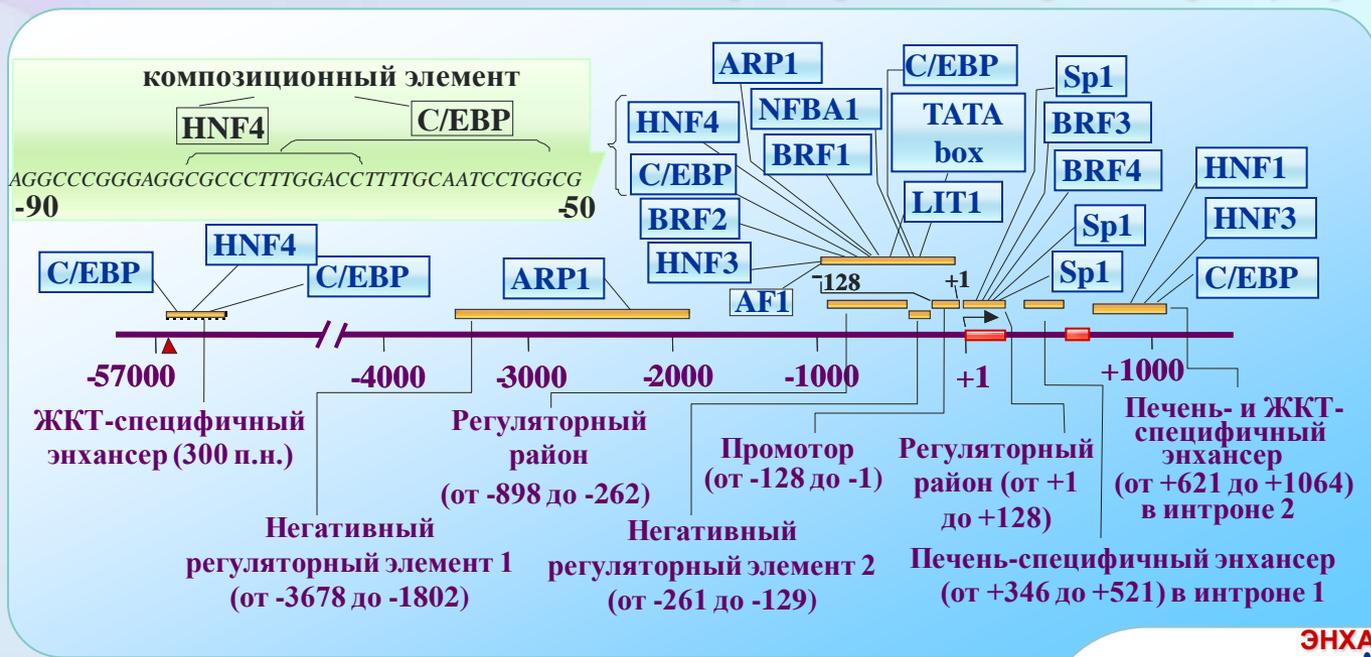
«Молекулярная бюрократия» на геномном уровне. Большое количество сайтов связывания транскрипционных факторов, контролируют один процесс – транскрипцию: промоторы бактерий.



С учетом взаимного перекрытия регуляторных сайтов, для данного промотора возможно 70 различных состояний, каждое из которых соответствует транскрипционному комплексу, сформированному определенным набором регуляторных белков. Разнообразие вариантов регуляции этого промотора - 140 ПРОЦЕССОВ.



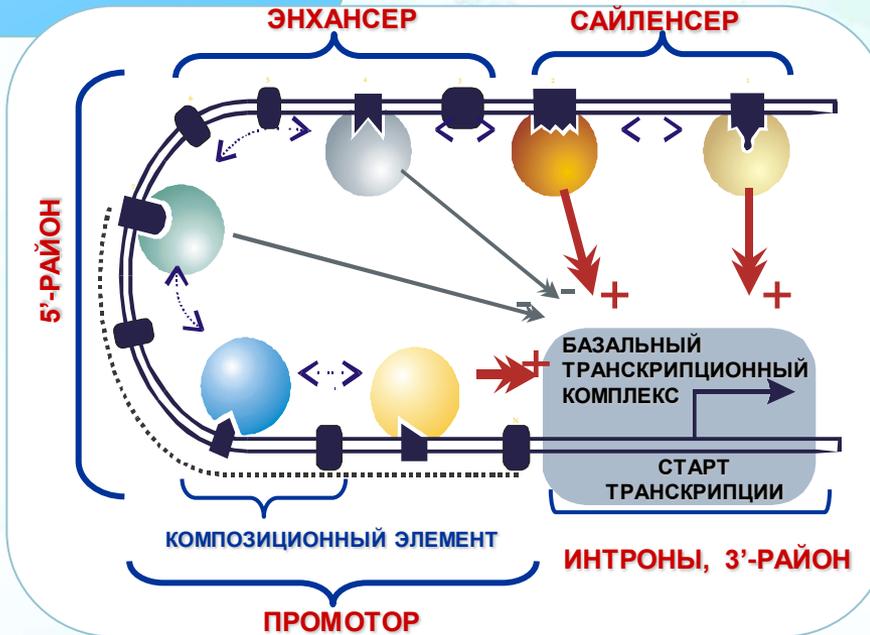
«Молекулярная бюрократия» на геном уровне. Большое количество сайтов связывания транскрипционных факторов, контролируют один процесс – транскрипцию: промоторы эукариот.



$W \sim 2^N$, где N - кол-во сайтов связывания транскрипционных факторов.
 При N = 20,
 $W \sim 1\ 000\ 000$

Организация регуляторных районов, контролирующих транскрипцию гена аполипопротеина В человека.

Регуляторные районы генов эукариот содержат большое количество сайтов связывания транскрипционных факторов, что обеспечивает большое разнообразие паттернов экспрессии



Регуляторный портрет генома человека: некоторые количественные оценки

Геном человека содержит
~ 20000 -25000 генов :

Среднее количество сайтов связывания
транскрипционных факторов:

40 на один ген

Ожидается, что общее количество функционально активных
сайтов связывания транскрипционных факторов:

40 x 20000 ~ 800 000 сайтов на геном человека

Для столь масштабного исследования нужны эффективные
технологии, основанные на компьютерно-экспериментальных
подходах.

Пример описания сайта связывания транскрипционного фактора SF1 в базе TRRD. Ген MIS (Mullerian inhibiting substance) человека

ОПИСАНИЕ САЙТА СВЯЗЫВАНИЯ SF1

[SiteAC](#) S4799

[GeneID](#) Hs:MIS

[RegUnitAC](#) P01298

[SiteName](#) SF-1;

[PreferredName](#) SF1

[SiteNameSynonym](#) NR5A1

[SiteIndex](#) 1

[FactorName](#) SF-1; steroidogenic factor 1

[FactorInfluence](#) increase

[Sequence](#) ggcactgtcccCCAAGGTCGcg

[SequencePosition](#) -103 to -82

[FootprintSequencePosition](#) -97 to -82

[DNA_BankLink](#) K03474:98

[ExperimentCodes](#) 1.1.5, 3.5, 3.2.2, 3.3

[De Santa Barbara P. et al., 1998]

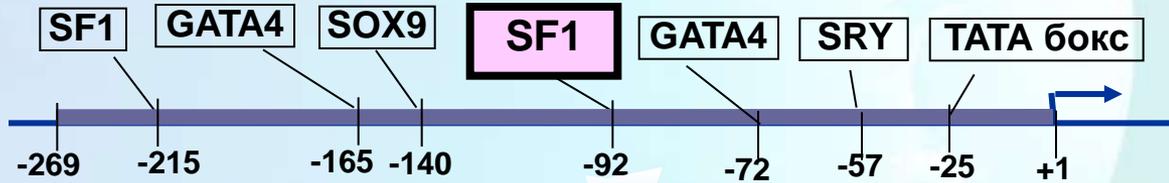
NT2/D1: 6.1.1, 3.6 [De Santa Barbara P. et al., 1998]

P2: 1.1.1, 3.1, 3.2.2, 3.3, 6.2 [Watanabe K. et al., 2000]

COS-7: 6.6.2.1(SF-1) [De Santa Barbara P. et al., 1998]

7.1, 7.2 [Watanabe K. et al., 2000]

//



ЭКСПЕРИМЕНТЫ, НА ОСНОВЕ КОТОРЫХ ИДЕНТИФИЦИРОВАН САЙТ SF-1

ДНКАЗА I ФУТПРИНТИНГ С ЯДЕРНЫМ ЭКСТРАКТОМ

ДНКАЗА I ФУТПРИНТИНГ С ОЧИЩЕННЫМ БЕЛКОМ

МЕТОД ЗАДЕРЖКИ В ГЕЛЕ С ИСПОЛЬЗОВАНИЕМ ОЧИЩЕННЫХ БЕЛКОВ

МЕТОД ЗАДЕРЖКИ В ГЕЛЕ С ИСПОЛЬЗОВАНИЕМ ЯДЕРНОГО ЭКСТРАКТА И СПЕЦИФИЧЕСКИХ АНТИТЕЛ

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭКСПРЕССИИ ПЛАЗМИДНЫХ КОНСТРУКЦИЙ С ТОЧКОВЫМИ МУТАЦИЯМИ В САЙТЕ

КОТРАНСФЕКЦИЯ ПЛАЗМИДОЙ, ЭКСПРЕССИРУЮЩЕЙ ТРАНСКРИПЦИОННЫЙ ФАКТОР SF-1

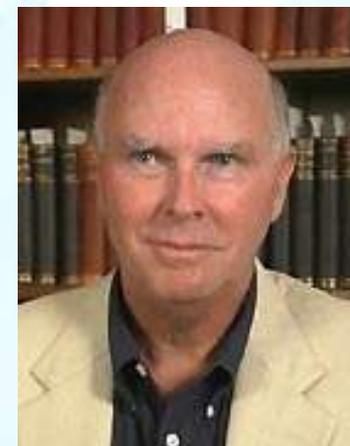




Полногеномное секвенирование человека: Дж. Уотсон, Кр. Вентер и анонимный китаец-УН. Показаны однонуклеотидные различия (SNP) между геномами

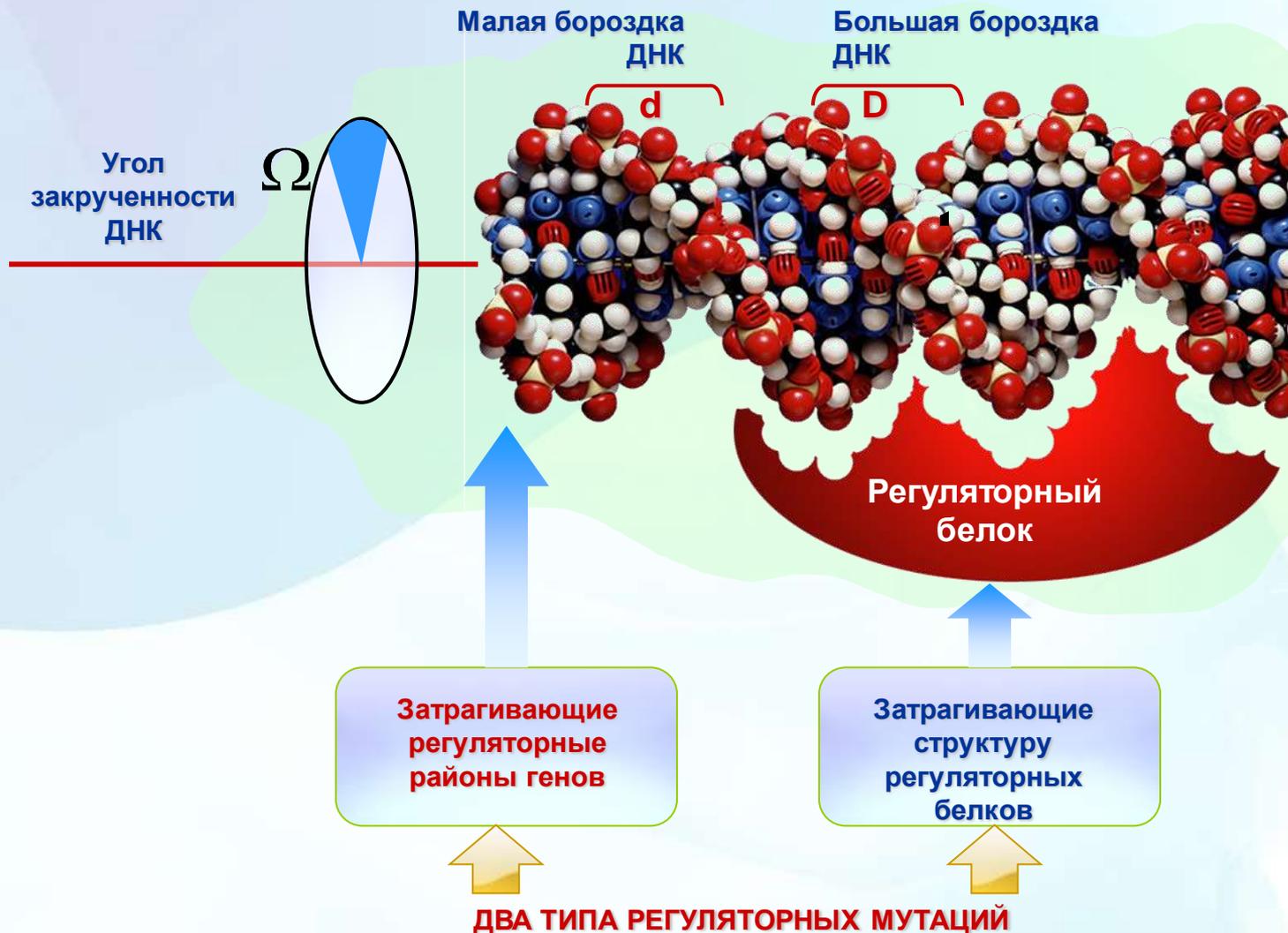
Всего SNP по сравнению с референсным геномом		Из них			Литература
		Ранее известные в dbSNP	Новые	Уникальные	
Дж. Уотсон	3 322 093	2 715 296	606 797	~1 096 873	Wheeler et al., 2008
Кр. Вентер	3 470 669	2 822 902	647 767	~647 767	
Китаец (УН)	~3 285 600	~2 260 000	~820 000	~978 370	Jun Wang et al., 2008

Число уникальных SNP (отличия от референсного генома), выявленных в геномах Дж. Уотсона, Кр. Вентера и одного из представителей монголоидной расы ~ 2 723 000

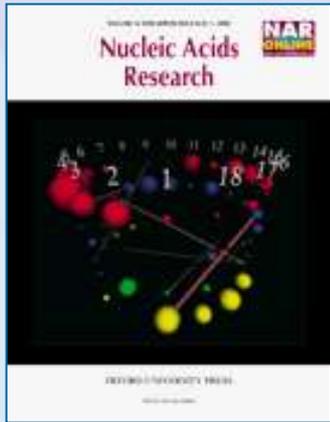




Какие типы мутаций могут приводить к выраженному изменению регуляторных функций?



Метод SITECON позволяет проводить анализ консервативных контекстно-зависимых конформационных и физико-химических свойств ДНК

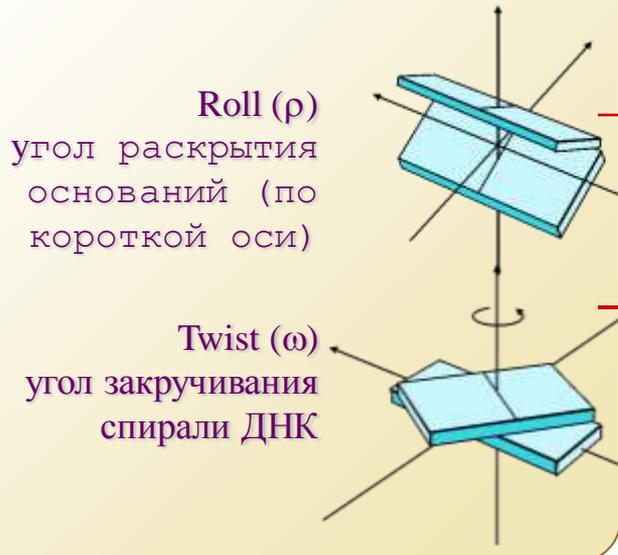


Nucleic Acids Research 2004, V.32, W208-W212.

Oshchepkov D.Y., Vityaev E.E., Grigorovich D.A., Ignatieva E.V.,
Khlebodarova T.M.

SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition.

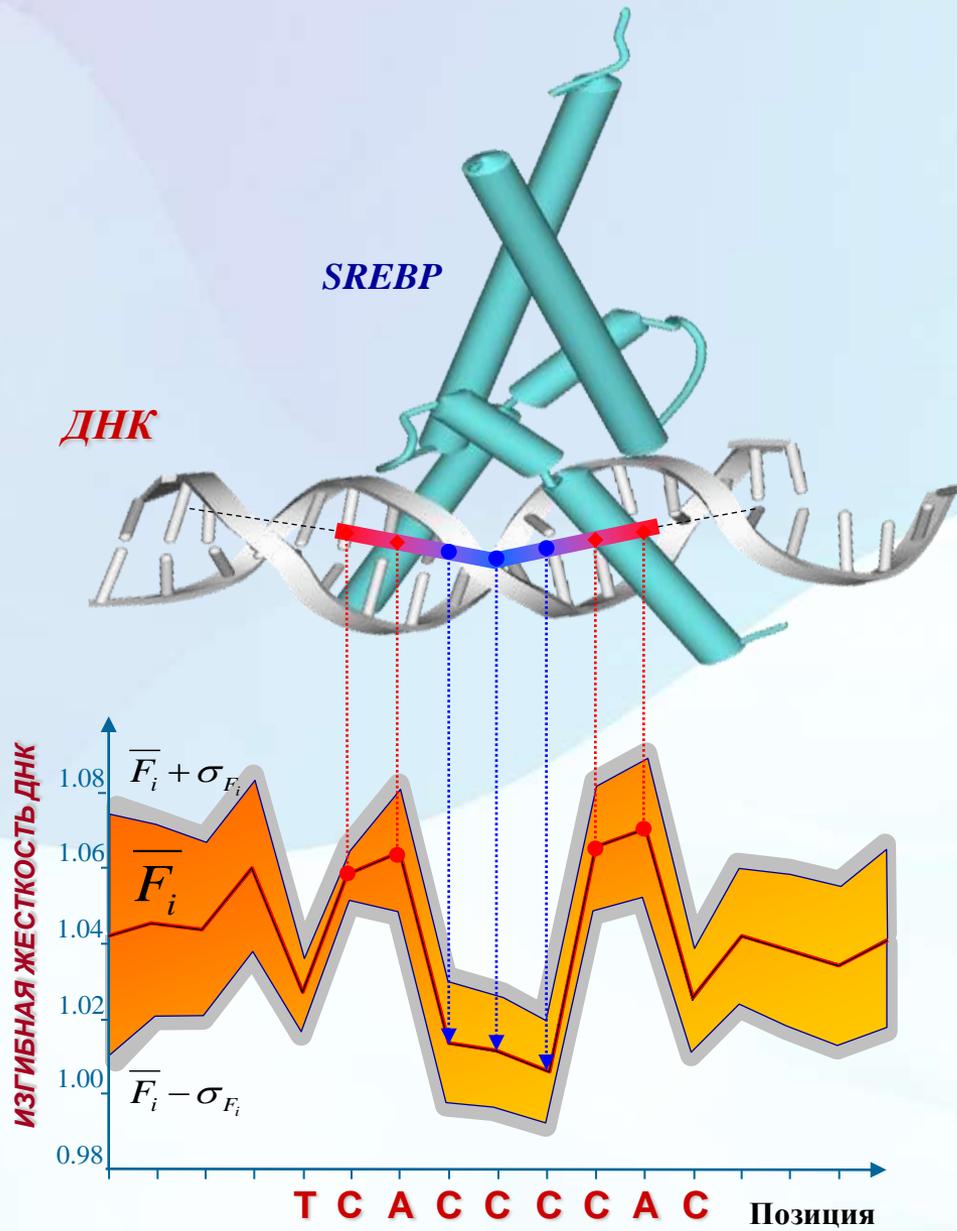
SITECON использует информацию из базы данных конформационных и физико-химических характеристик ДНК (<http://www.mgs.bionet.nsc.ru/mgs/gnw/bdna/>)



Динуклеотид	Roll, (градусы)	Twist, (градусы)
AA	0.3	35.3
AT	-0.8	31.2 *
AG	4.5	31.2 *
AC	0.5	32.6
TA	2.8	<u>40.5 **</u>
TT	0.3	35.3
TG	0.5	32.6
TC	-1.3	40.3
GA	-1.3	40.3
GT	0.5	32.6
GG	<u>6.0 **</u>	33.3
GC	<u>-6.2 *</u>	37.3
CA	0.5	39.2
CT	4.5	<u>31.2 *</u>
CG	-6.2 *	36.6
CC	6.0 **	33.3

SITECON: анализ консервативных конформационных и физико-химических свойств сайта связывания белка SREBP с ДНК

*Пространственная структура комплекса «SREBP/ДНК»
(Párraga A. et al., 1998)*



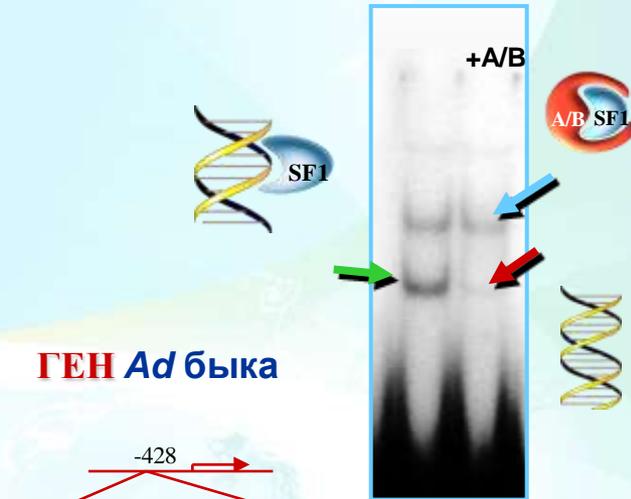
- Особенности сайта, выявленные методом SITECON**
- 1. пониженная жесткость ДНК в центре*
 - 2. повышенная жесткость ДНК по краям*
 - 3. Наличие консервативного контекста в центре*

Экспериментальная проверка потенциальных сайтов связывания SF-1, распознанных в промоторных районах генов стероидогенеза методом SITECON (Т.И.Меркулова, Г.В.Васильев, Н.В.Климова, лаб. регуляции экспрессии генов)

Ген	Позиция сайта SF1	Ориентация сайта	Экспериментальное подтверждение
1 Cyp17(Mouse)	-283	<-	+
2 Cyp17 (Mouse)	-49	->	+
3 Ad (Bovine)	-428	<-	+
4 Cyp11B1 (Guinea pig)	-126	<-	+
5 Cyp11B3 (Rat)	-309	->	+
6 Cyp11B1 (Sheep)	-337	->	+
7 Oxt (Mouse)	-164	<-	+
8 Oxt (Rat)	-167	<-	+
9 Oxt (Human)	-159	<-	+
10 Cyp11B2 (Rat)	-324	->	+
11 HSD3b (Mouse)	-113	<-	+
12 Ad4BP/SF-1 (Mouse)	-224	->	+
13 CYP17 (Porcine)	-51	->	+
14 HSD17BI (Rat)	-84	<-	+
15 LH beta (Porcine)	-114	<-	+

15
100%

ГЕН Ad быка

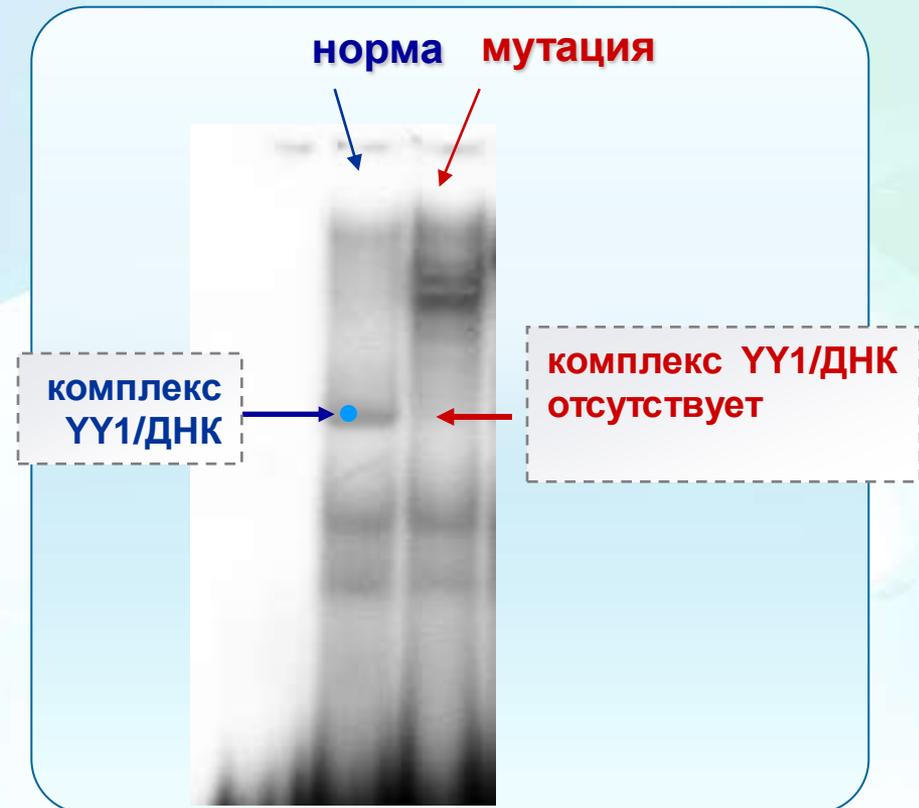
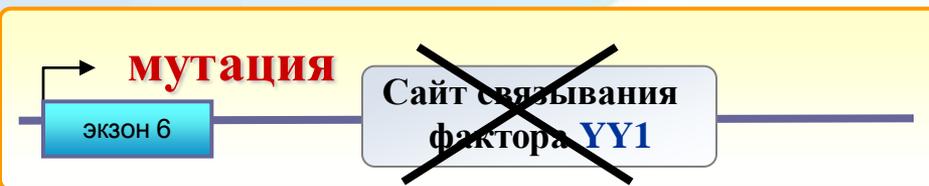
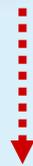
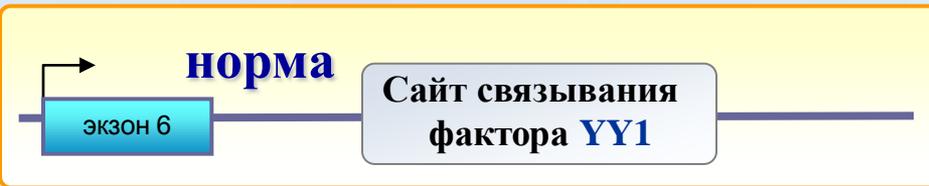


- комплекс SF1/ДНК
- ← исчезновение или ослабление комплексов SF1/ДНК при добавлении специфических антител (A/B)
- ← комплексы, образуемые другими белками



Мутация **G** --> **A** в 6-м интроне гена триптофан 2,3 диоксигеназы сопровождается комплексом психических патологий, таких как рассеянное внимание, высокий риск наркотической зависимости и др.

Экспериментально-компьютерный анализ доказал, что мутация повреждает сайт связывания транскрипционного фактора YY1.



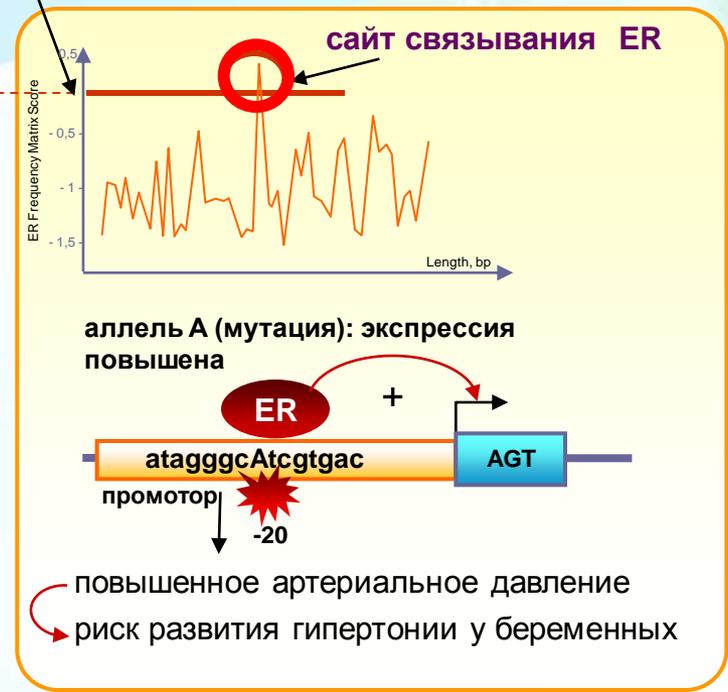
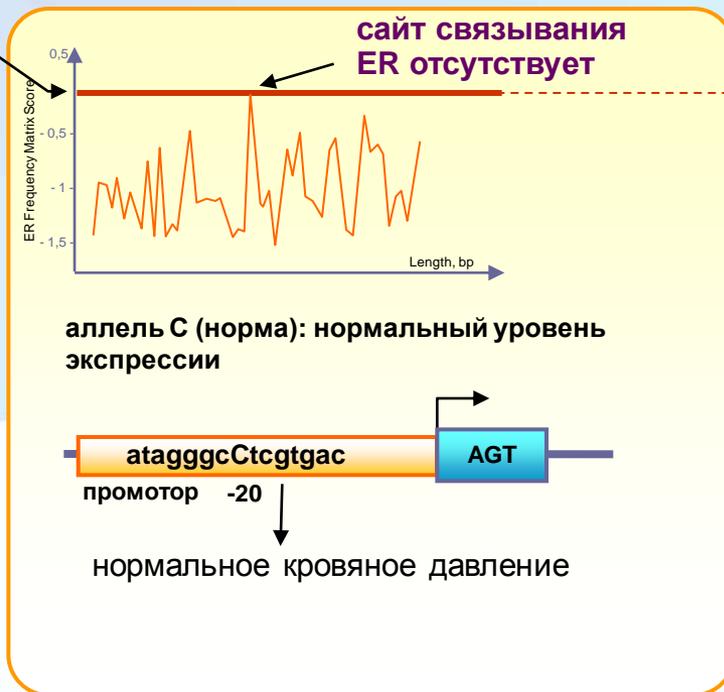
Норма : 5' -cagtTGCCAAATAATG**G**CAGATAAGAATAGGGAG-3'

Мутация : 5' -cagtTGCCAAATAATG**A**CAGATAAGAATAGGGAG-3'

rSNP-tools: Функциональная интерпретация полиморфизма -20(C -> A) в регуляторном районе гена ангиотензиногена человека (Zhao Y. et al., Hypertension, 1999), ассоциированного с риском развития гипертонии беременных

Лаборатория теоретической генетики (Пономаренко М.П., 2005).

Gene, region	SNP's and/or mutagenesis data		DNA/TF-binding		TF-core		Prediction		
	DNA, 5'-3'	Allele	Observed	X	(+)	(-)	TF site	α	$d_{TF,X}$
<i>hANG</i> promoter	<i>atagggcCtcgtgac</i>	Normal	Absent	0.5	0,1	0,6	ER	0,005	0,04
	<i>atagggcAtcgtgac</i>	Risky	Present	1	0,4	0,8			

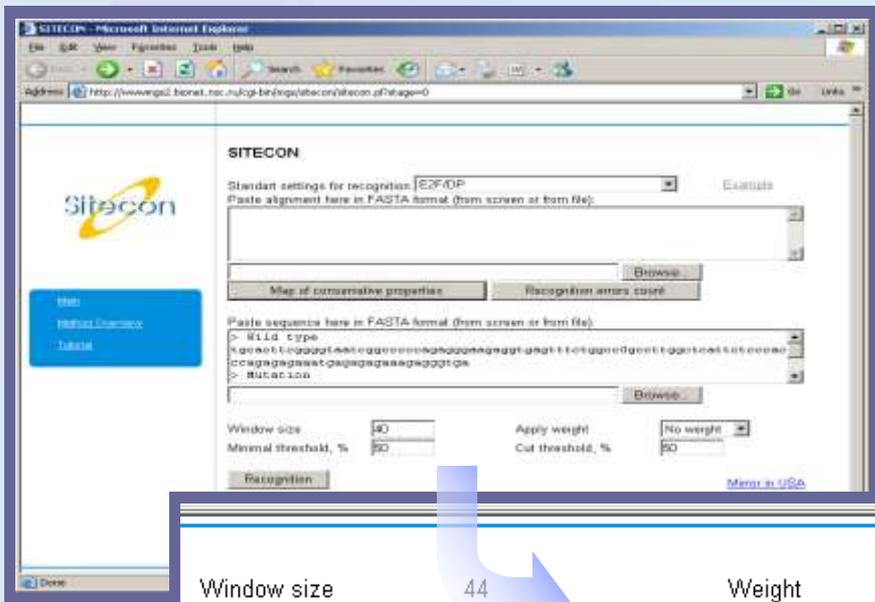


В результате мутации возникает новый сайт связывания эстрадиолового рецептора



Метод SITECON: анализ полиморфизма (G->A) в первом интроне гена TNF-alpha быка, ассоциированного с лейкозом. В результате мутации разрушается сайт связывания транскрипционного фактора NF-κB.

LABORATORY OF THEORETICAL GENETICS
INSTITUTE OF CYTOLOGY AND GENETICS
Russian Academy of Sciences, Siberian Branch



Норма - аллель G (устойчивость к лейкозу):
сайт NF-κB присутствует

ЭКЗОН 1

```
gtgagtttctggccGgccttggctcatt  
cactcaaagaccggCcggaaccgagtaa
```

Мутация - аллель A (чувствительность к лейкозу):
сайт NF-κB исчезает

ЭКЗОН 1

```
gtgagtttctggccAgccttggctcatt  
cactcaaagaccggTcggaaccgagtaa
```

Window size 44 Weight Algorithm 2
Minimal threshold, % 50 Cut threshold, % 60

Calculate other threshold >= 50

[\[Map of conservative properties\]](#) [\[Table of conservative properties\]](#) [\[Full Result Table\]](#)

Cutted result table
>1: Wild type
Pos 49, 0.603, indirect, GGGTGGGAGAAATGAGCCAAGGCCGGCCAGAAACTCACCTCTCC
>2: Mutation
Summarized length of all tested sequences = 198
Sum = 1

ТРАНСКРИПТОМИКА

Экспериментально-компьютерная транскриптомика
(лаборатории ИЦиГ СО РАН, работающие в области генетики животных, молекулярной генетики и клеточной биологии):

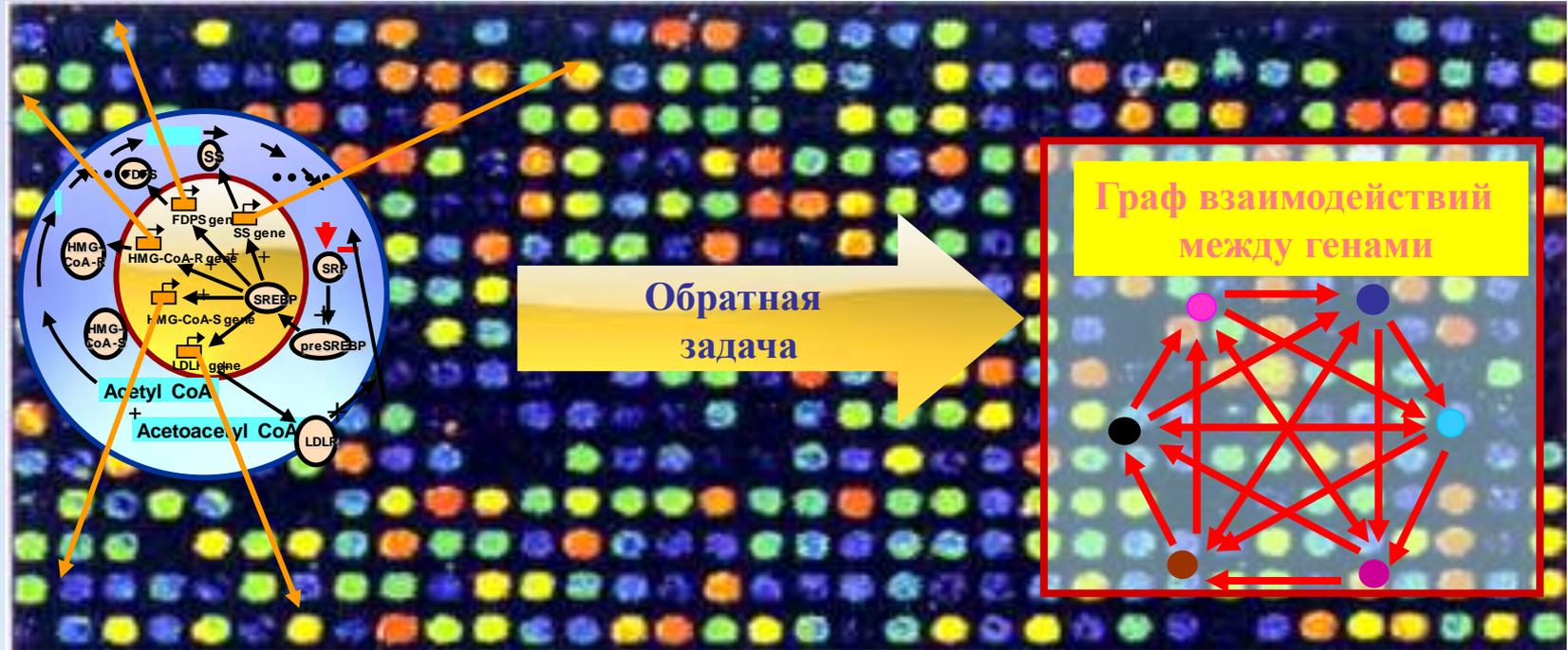
- 1. Крупномасштабные эксперименты по изучению профилей экспрессии больших групп генов на основе биочиповых технологий**

Компьютерная транскриптомика (Лаборатория молекулярно-генетических систем и Лаборатория теоретической генетики Отдела системной биологии ИЦиГ СО РАН):

- 1. Компьютерная поддержка биочиповых технологий;**
- 2. Реконструкция генных сетей на основе биочиповых данных**



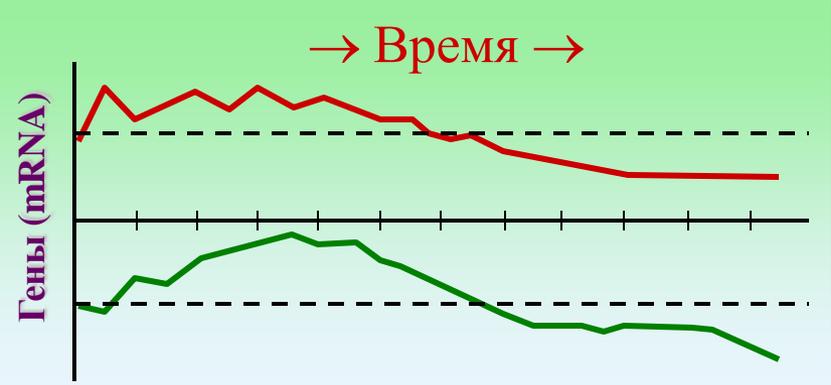
Компьютерные технологии реконструкции графов генных сетей с использованием данных Microarray Analysis



Исходные данные: χ_{ij}^j \Leftrightarrow концентрация i -го гена
 \Leftrightarrow в j -й временной точке

→ Время →

$$X_{N \times M} := \begin{pmatrix} \chi_1^1 & \chi_1^2 & \dots & \chi_1^M \\ \chi_2^1 & \chi_2^2 & \dots & \chi_2^M \\ \vdots & \vdots & \ddots & \vdots \\ \chi_N^1 & \chi_N^2 & \dots & \chi_N^M \end{pmatrix}$$





ВОССТАНОВЛЕНИЕ ГРАФА ВЗАИМОДЕЙСТВИЙ МЕЖДУ ГЕНАМИ

$$\dot{x}_i(t) = -\lambda_i x_i(t) + \sum_{j=1}^N W_{ij} x_j(t) + b_i(t) + \xi_i(t) \quad (i=1,2,\dots, N).$$

$x_i(t)$ - концентрация i -й mRNA в момент времени t
(отражает уровень экспрессии i -го гена);

λ_i - параметр самодеградации i -й mRNA;

$b_i(t)$ - внешняя стимуляция (добавка);

$\xi_i(t)$ - стохастический шум.

$\mathbf{W} = (W_{ij})$ - матрица влияний (connectivity) размерности $\mathbf{N} \times \mathbf{N}$

W_{ij} - определяет тип и силу влияния j -го гена на i -й ген;

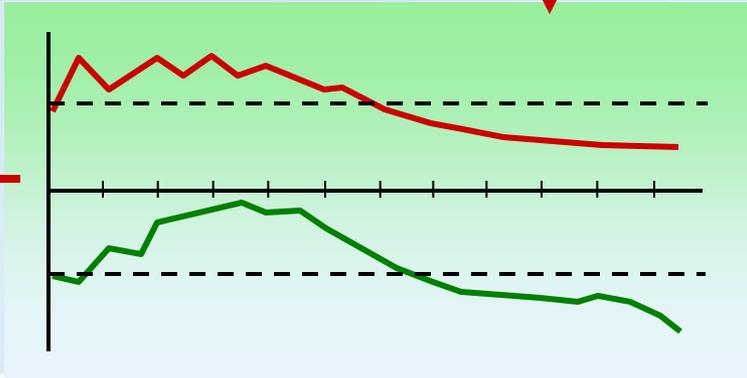
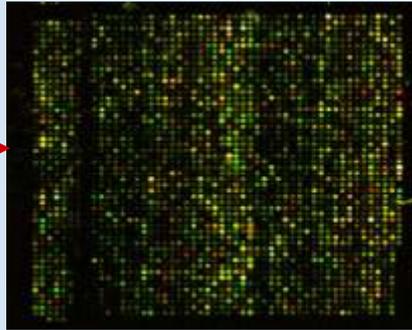
при $W_{ij} > 0$ i -й ген оказывает активирующее влияние на j -й ген;

при $W_{ij} < 0$ i -й ген оказывает репрессирующее влияние на j -й ген;

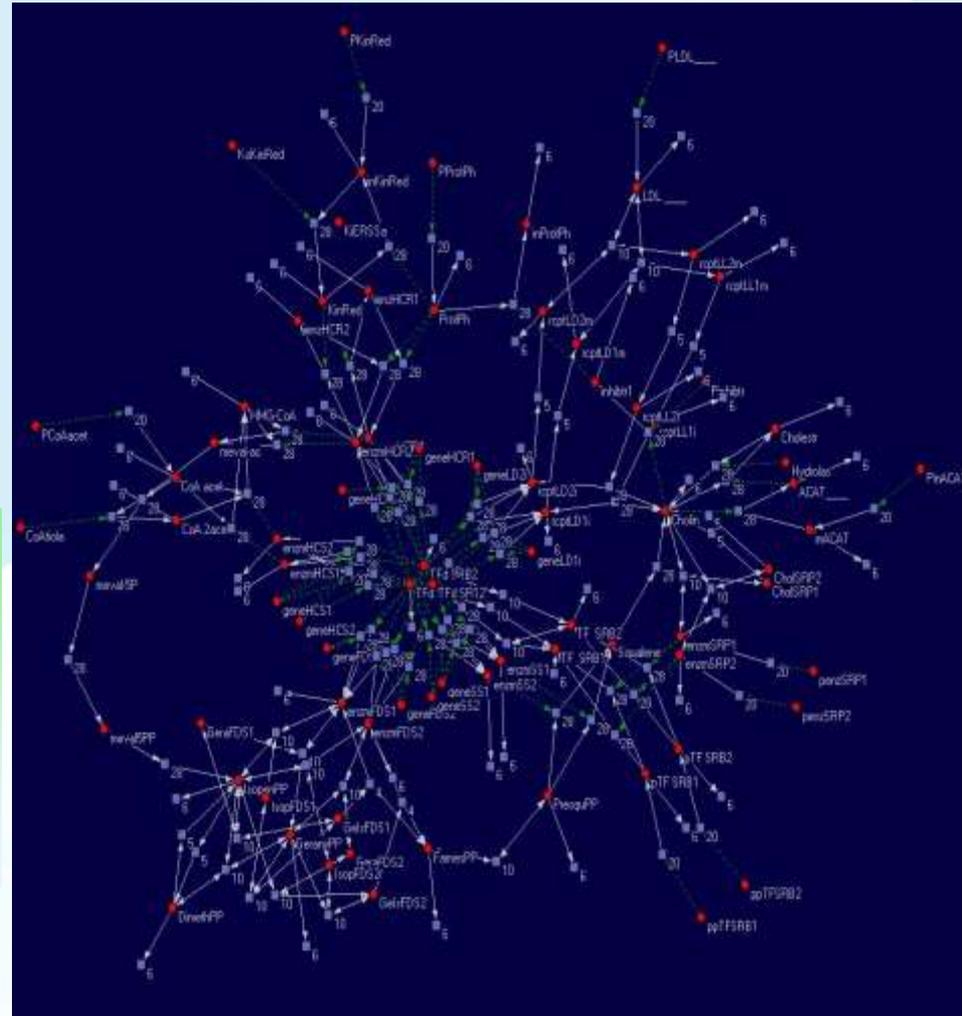
при $W_{ij} = 0$ взаимодействием между i -м и j -м геном отсутствует.



Биочиповые технологии и генные сети



$$\dot{x}_i(t) = -\lambda_i x_i(t) + \sum_{j=1}^N W_{ij} x_j(t) + b_i(t) + \xi_i(t) \text{ for } i = 1, 2, \dots, N.$$



$$W = X V D^{-1} U^T \quad (\text{сингулярное разложение})$$

ПРОТЕОМИКА

Экспериментально-компьютерная протеомика
(Лаборатория молекулярных биотехнологий ИЦиГ
СО РАН):

1. масспектрометрические методы протеомного анализа;
2. конструирование генетически модифицированных белков.

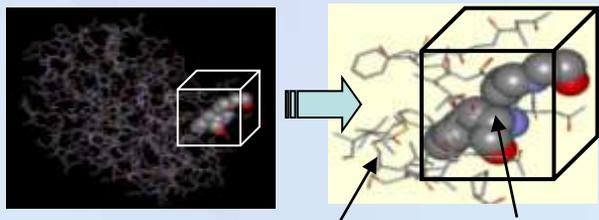
Компьютерная протеомика

(Сектор компьютерной протеомики Отдела
системной биологии ИЦиГ СО РАН):

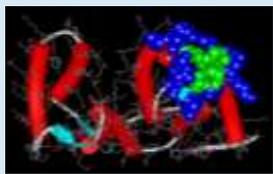
1. Компьютерная поддержка полногеномных проектов – реконструкция структурно-функциональной организация белков, кодируемых расшифрованными геномами;
2. Компьютерный анализ и предсказание функциональных сайтов в пространственных структурах белков;
3. Компьютерный дизайн белков с заданными свойствами.
4. Поиск новых мишеней в пространственной структуре белков

PDBSITE: база данных функциональных сайтов в 3-D структурах белков,

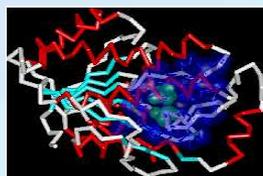
Иванисенко и др. 2008, ИЦиГ СО РАН



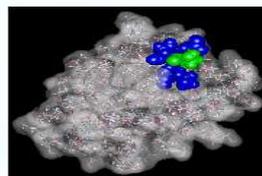
Сайты связывания лекарств



Сайты белок-белковых взаимодействий
Сайт взаимодействия белок-РНК



Каталитические центры

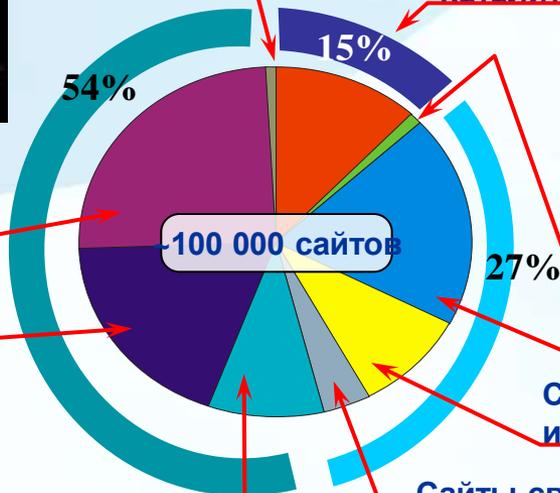


Сайты посттрансляционной модификации

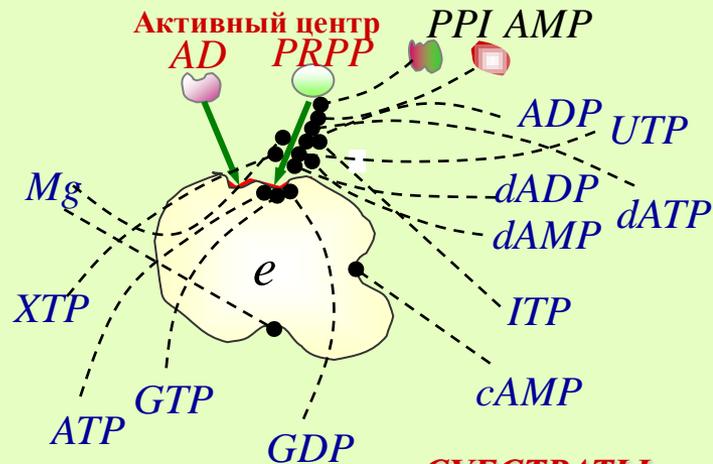
Сайты связывания органических лигандов

Сайты связывания ионов металлов

Сайты связывания неорганических лигандов



Аденин-фосфорибозил_трансфераза



СУБСТРАТЫ:

AD - adenine

PRPP - phosphoribosylpyro-phosphate

ПРОДУКТЫ:

PPI - pyrophosphate

AMP - adenosine-5'-phosphate

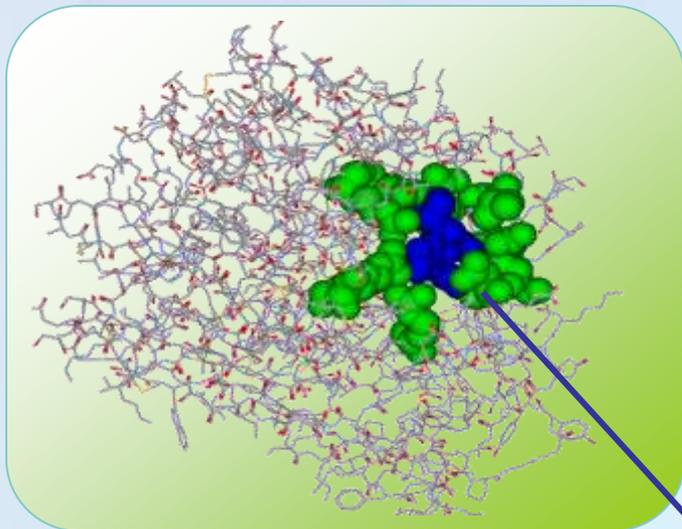
РЕГУЛЯТОРЫ

ADP; dADP; ATP; dATP;

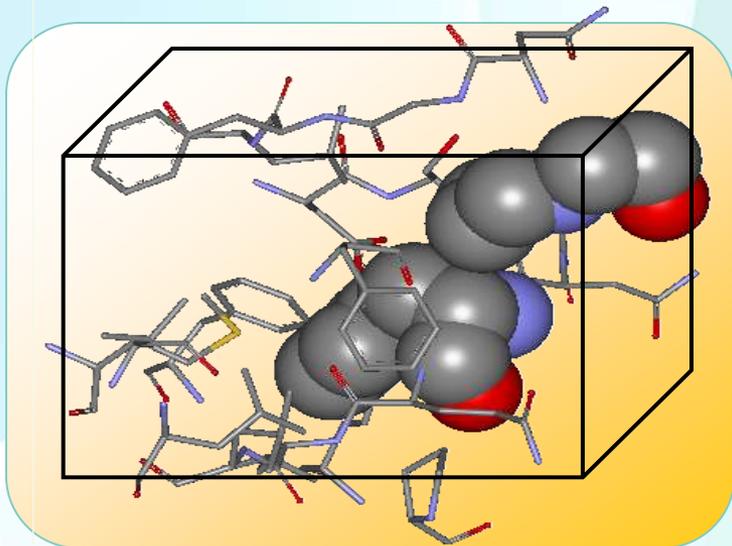
dAMP; GTP; ITP; XTP;

UTP; GDP; Mg; cAMP

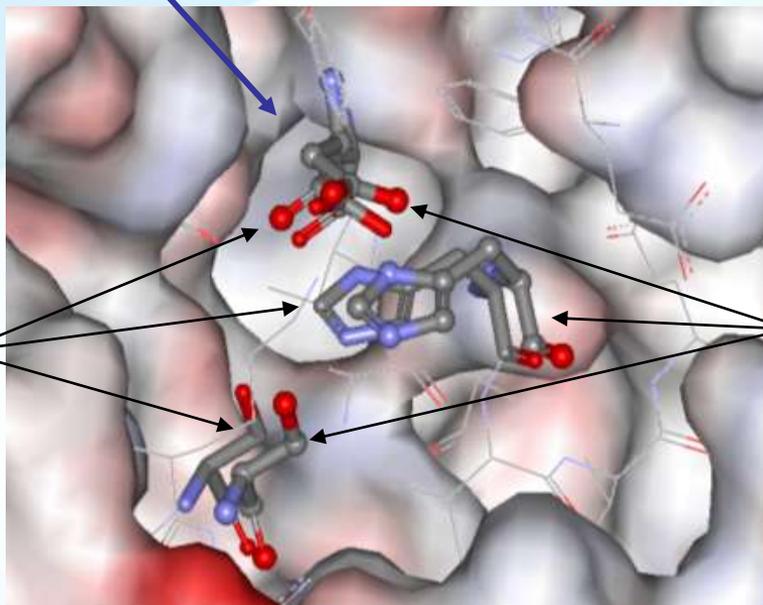
PDBSite: База данных функциональных сайтов в пространственных структурах белков: 100 000 сайтов



**Поиск сайтов
в 3-D структуре
белков**



**Аминокислотные остатки
распознанного
каталитического центра
в сериновой протеазе
человека (1ELV)**



**Аминокислотные остатки
каталитического центра
сериновой протеазы
(1BQY) из базы PDBSite.**

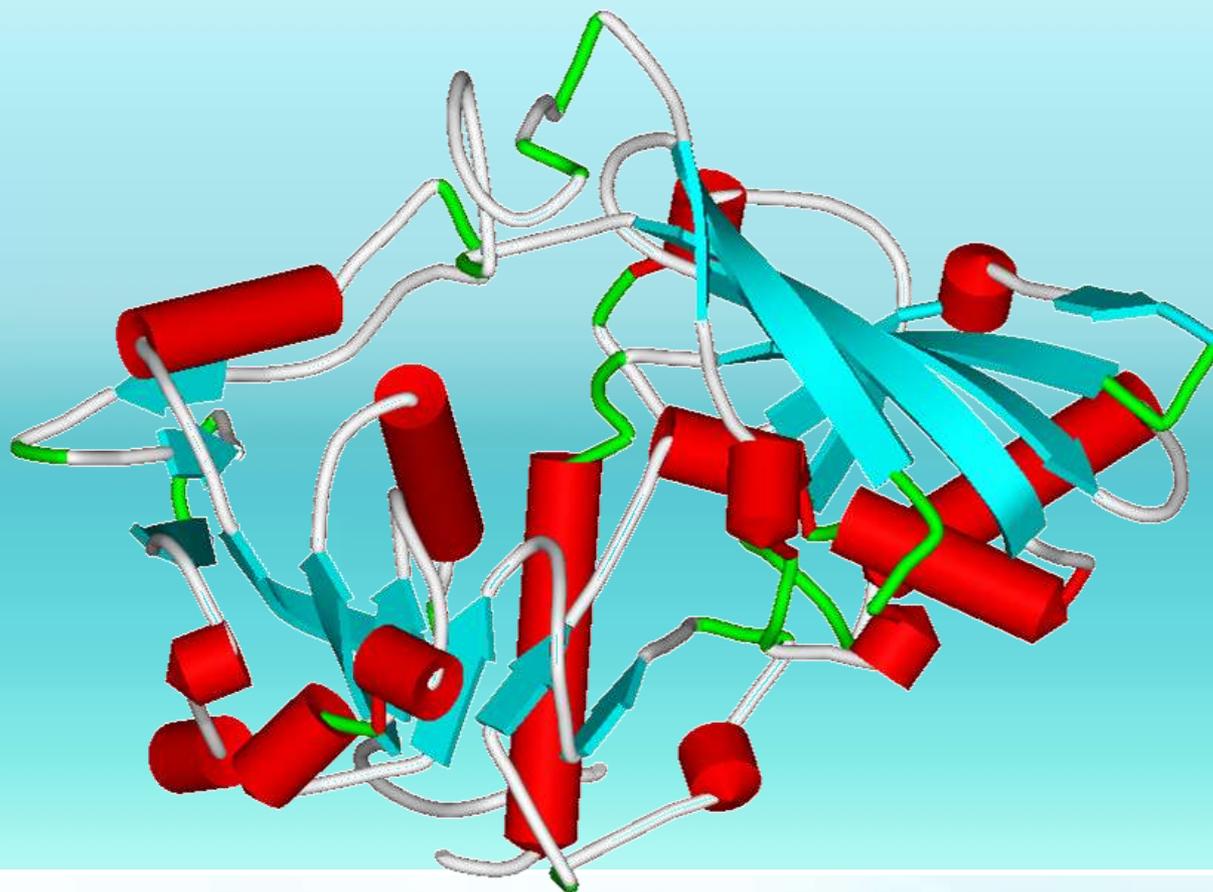
ИВАНИСЕНКО и соавт., 2005

Моделирование по гомологии пространственной структуры белка Oxidoreductase, кодируемого геном бактериального генома

(лаборатория теоретической генетики)

Исходная информация:

- (1) аминокислотная последовательность реконструируемого белка;
- (2) пространственная структура гомологичного белка с уровнем гомологии не ниже 50%.

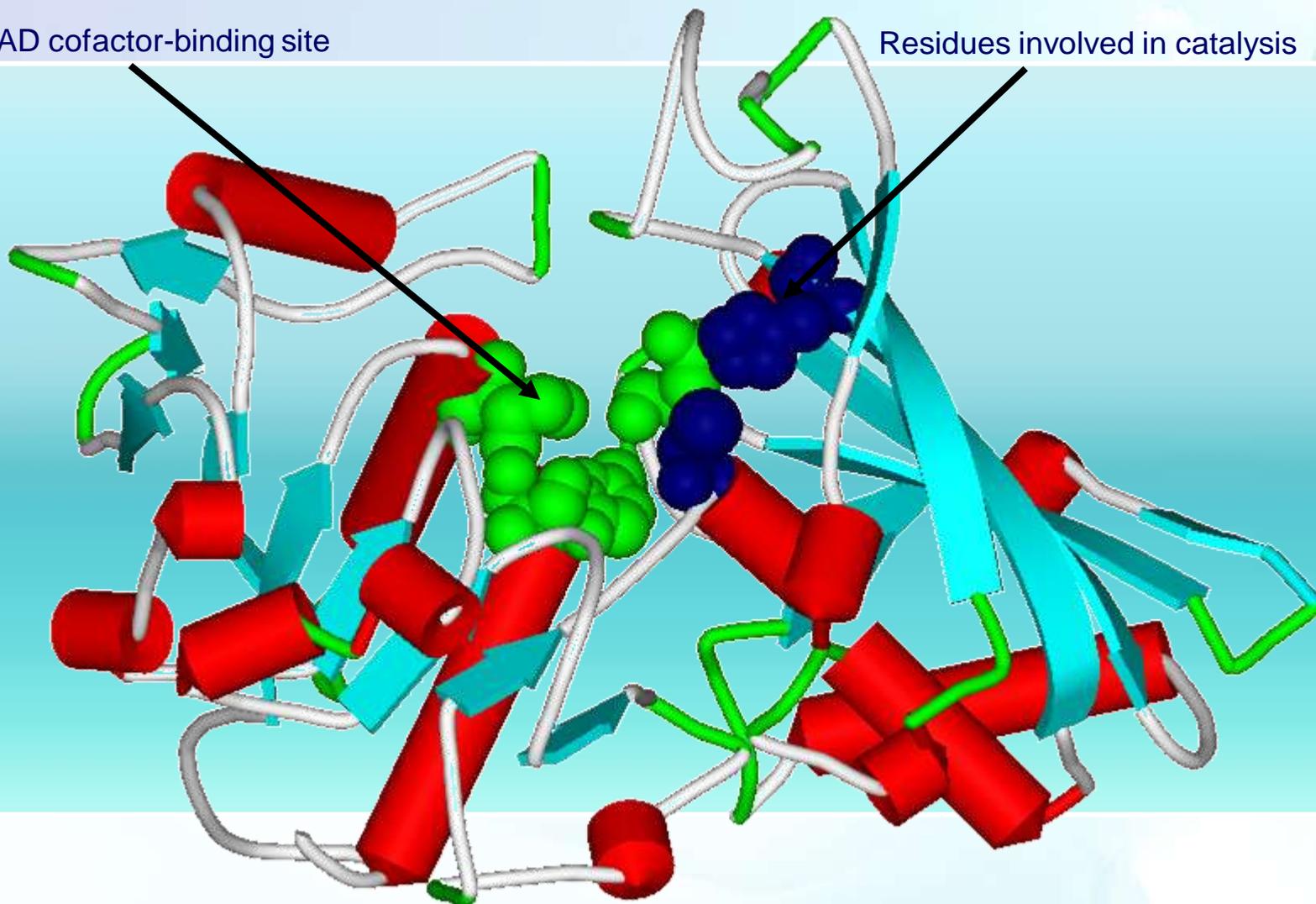


Поиск активных центров и сайтов связывания лигандов в 3-D структуре белка **Oxidoreductase**, кодируемого в исследуемом геноме

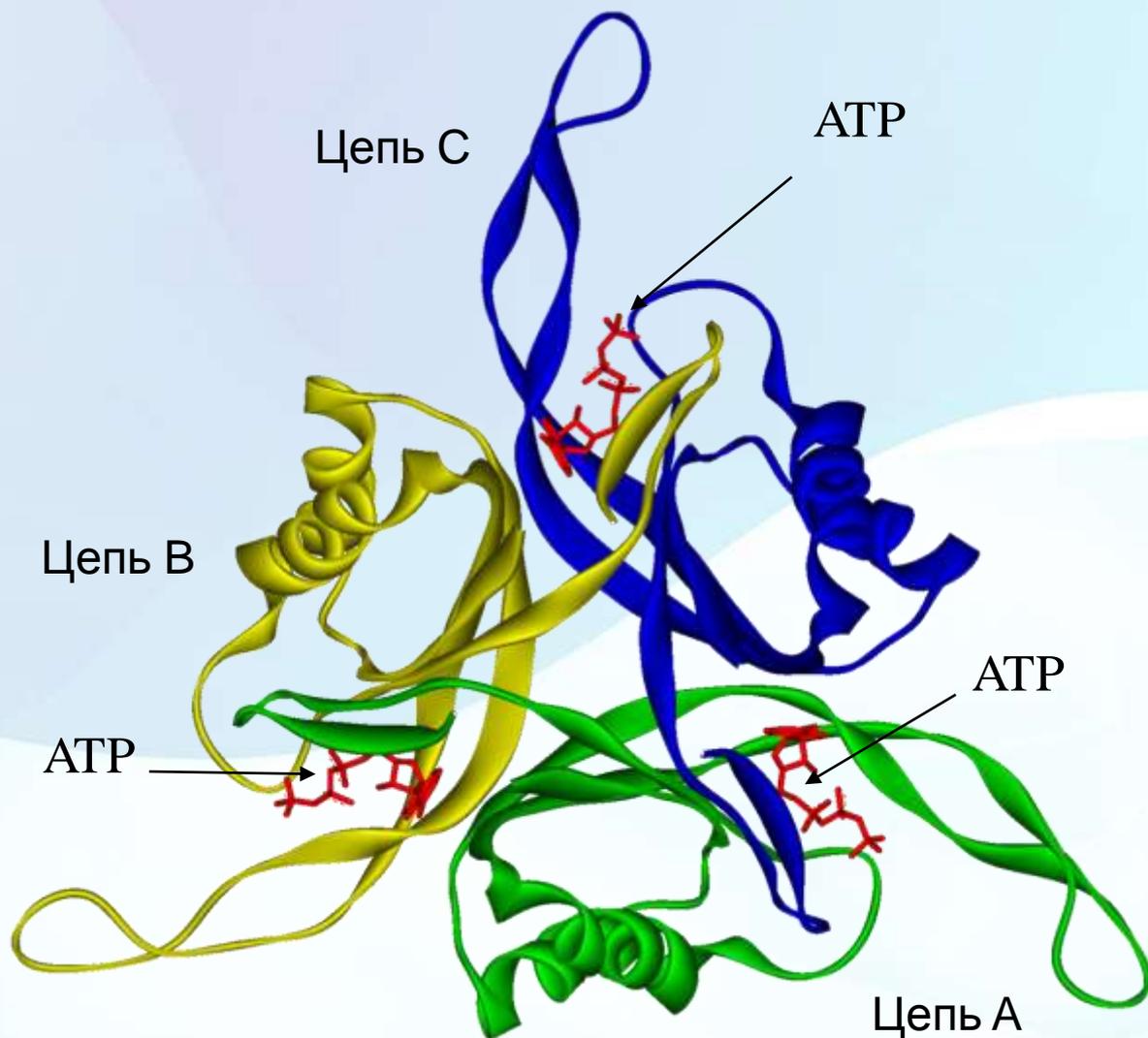
Центром «Биоинженерия» РАН

NAD cofactor-binding site

Residues involved in catalysis



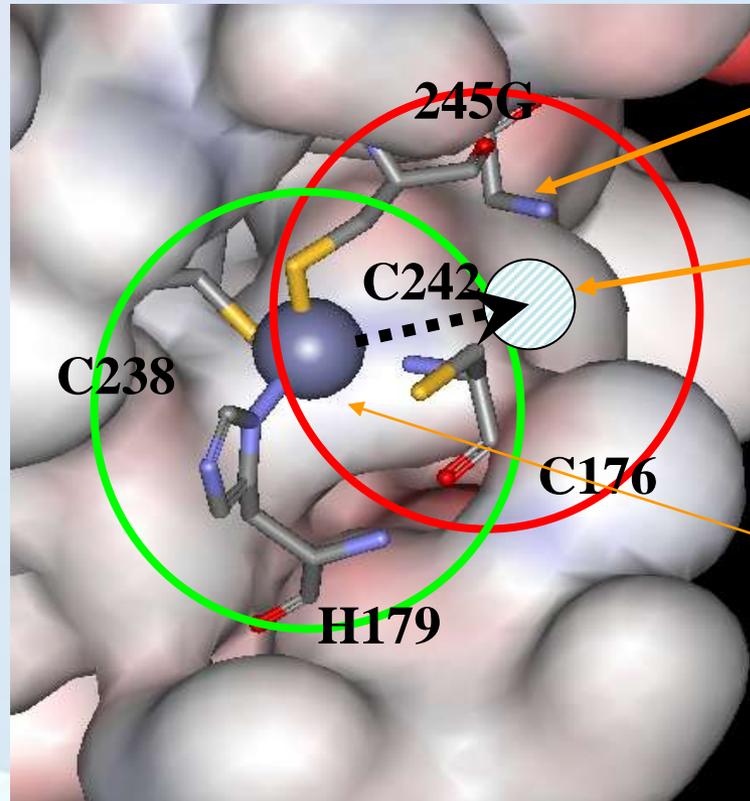
**Белок PII - важнейший регулятор метаболизма азота,
сигнальной трансдукции и мембранного транспорта
*Mycobacterium tuberculosis***



**PII – тример,
формируемых тремя
идентичными
субъединицами.**

**Три молекулы АТФ
взаимодействуют с
боковыми
карманами между
соседними
субъединицами
белка PII.**

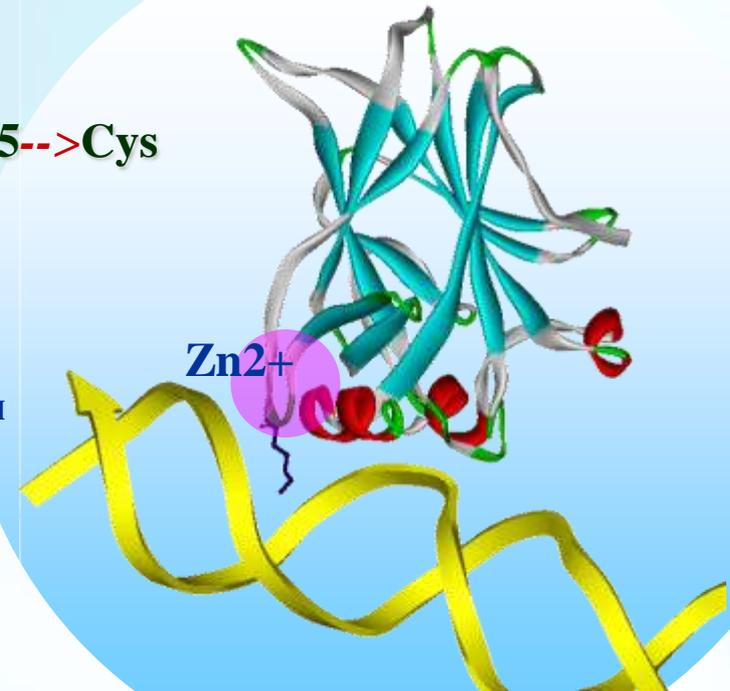
При аминокислотной замене Gly245->Cys в ДНК-связывающем домене белка p53, ассоциированной с риском канцерогенеза, возникает новый сайт связывания цинка.



Замена Gly245-->Cys

Предсказанный сдвиг места связывания иона цинка

Нативное место связывания иона цинка



С использованием метода молекулярной механики MOLKERN показано, что новый сайт связывания формирует с цинком более стабильный комплекс, чем исходный сайт связывания (Fomin et al., 2006, ICG SB RAS)

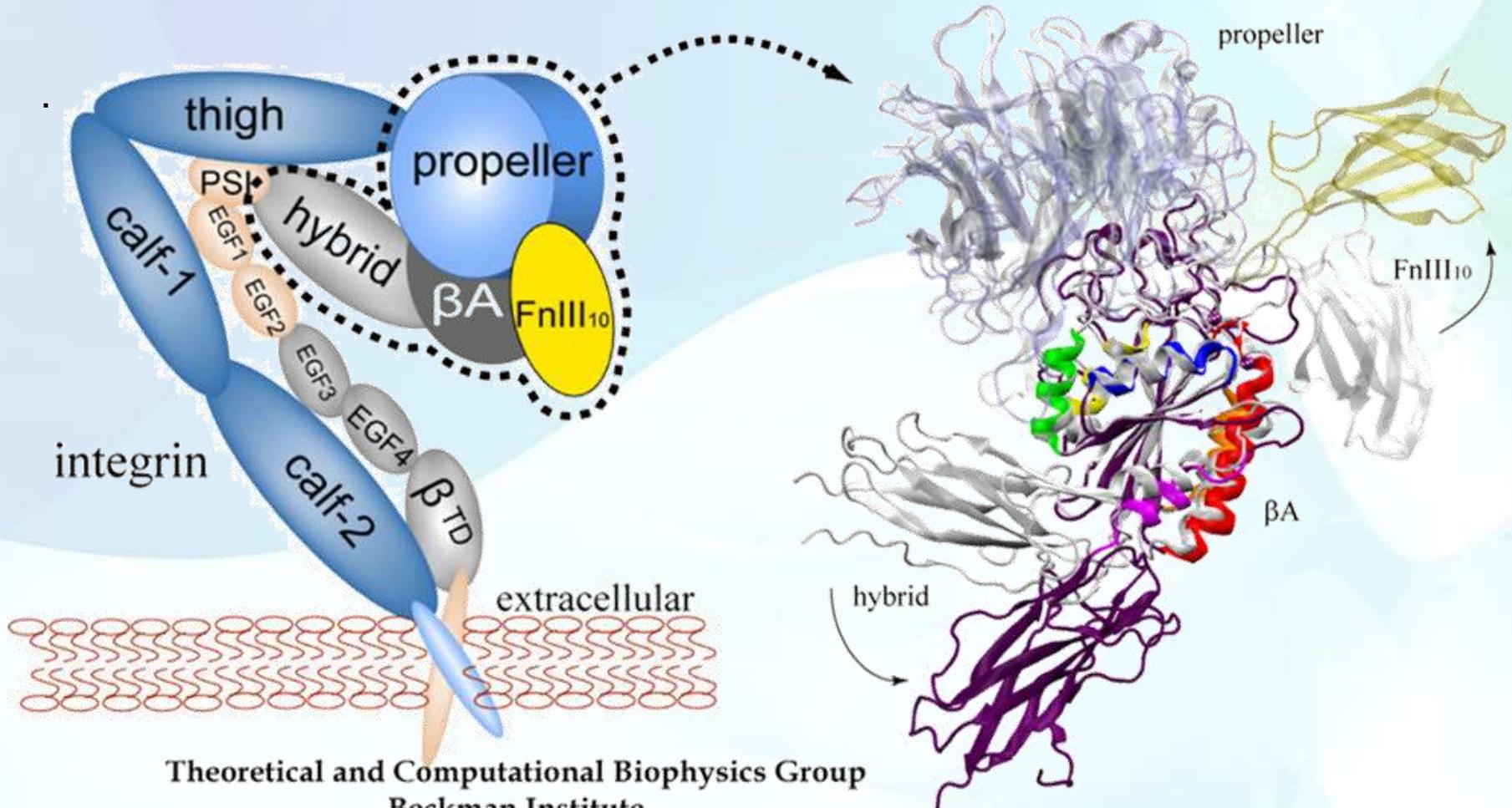
Сайт связывания Zn^{2+} - норма

Потенциальный сайт связывания Zn^{2+}

Иванисенко
и др., 2006, ИЦиГ
СО РАН

Молекулярное моделирование структуры интегринов.

Интегрины – супрамолекулярные машины, обеспечивающие контакты между клетками. Конформация интегринов зависит от поступающих к ним молекулярных сигналов из внеклеточного или внутриклеточного матрикса.



Системная биология

- 1. Экспериментальный и компьютерный анализ генных сетей, метаболических путей, путей передачи сигналов,**
- 2. Компьютерный дизайн и экспериментальное конструирование генных сетей и метаболических путей с заданными свойствами.**

(ИЦиГ СО РАН:

Отдл системной биологии

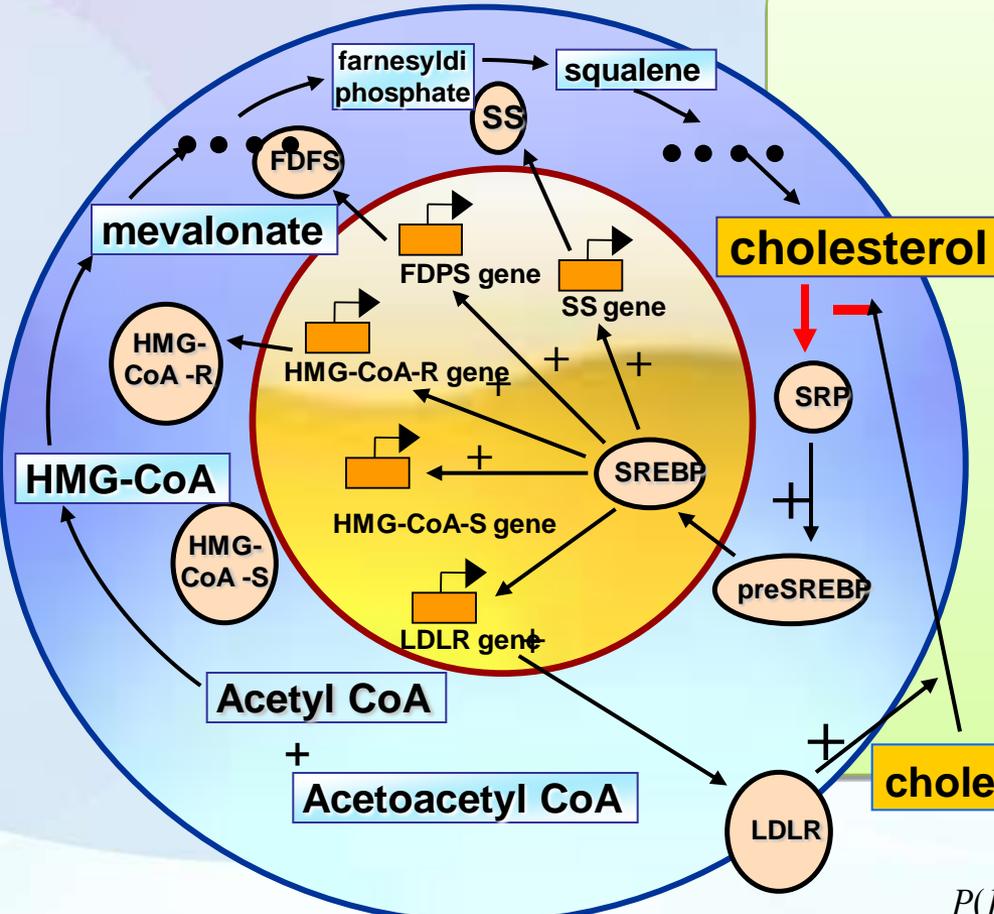
Лаборатория генной инженерии,

Лаборатория молекулярной и клеточной биологии,

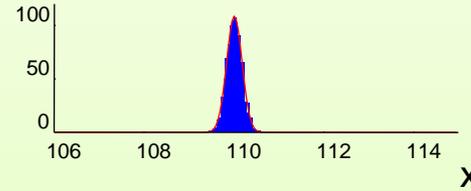
Лаборатория молекулярных биотехнологий)



Центральный фрагмент генной сети биосинтеза холестерина в клетке с регуляцией по механизму отрицательной обратной связи

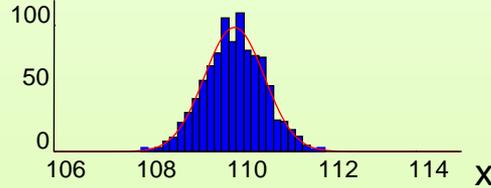


«НЕЙТРАЛИЗАЦИЯ» мутационного спектра под действием отрицательной обратной связи



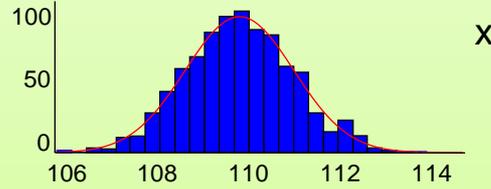
$kf=1$

Очень сильная отрицательная обратная связь



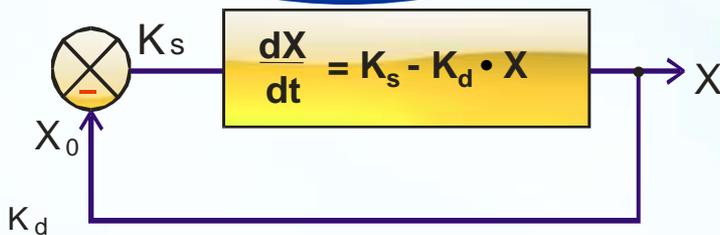
$kf=0.1$

Отрицательная обратная связь умеренной силы



$kf=0.03$

Очень слабая отрицательная обратная связь



Отрицательная обратная связь

$$P(E_i) = \sqrt{1/2\pi} e^{-\frac{1}{2} \left(\frac{E_i - E_0}{\sigma_E} \right)^2}$$

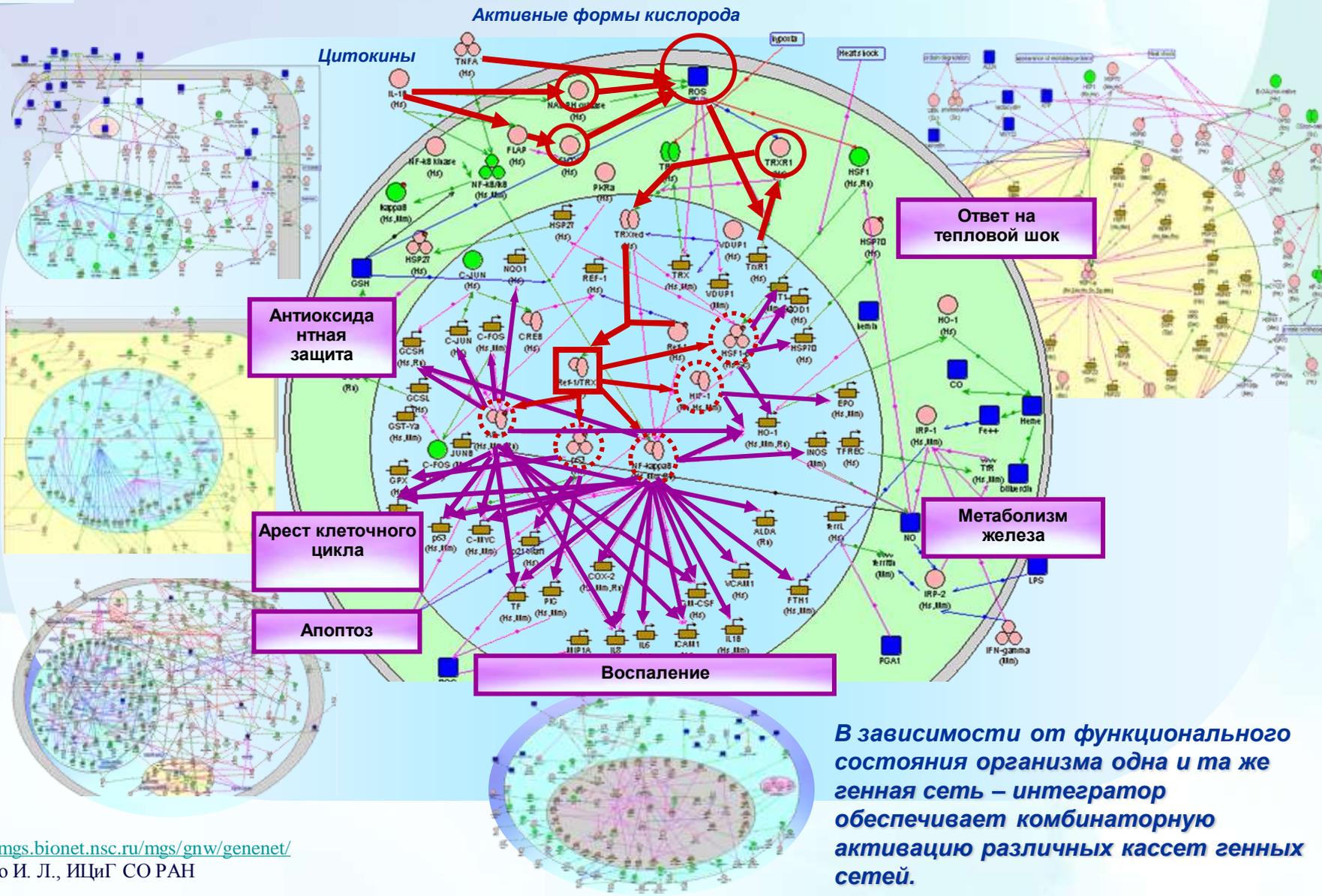
$$X_i = -\frac{C_2}{2} + \sqrt{\frac{C_2^2}{4} + \frac{C_5 C_2}{C_3} (E_i + C_4)}$$

$$kf = 1/C_2$$

$$W(X_i) = \sqrt{1/2\pi} e^{-\frac{1}{2} \left(\frac{X_i - X_0}{\sigma_X} \right)^2}$$

Генные сети – интеграторы:

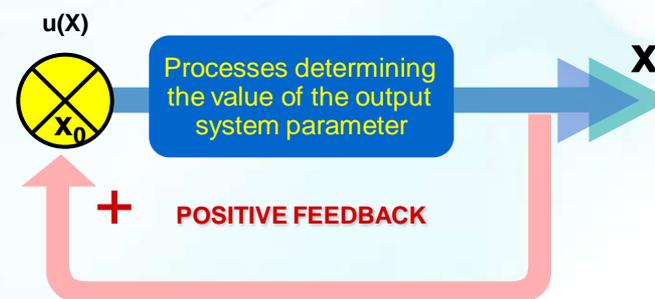
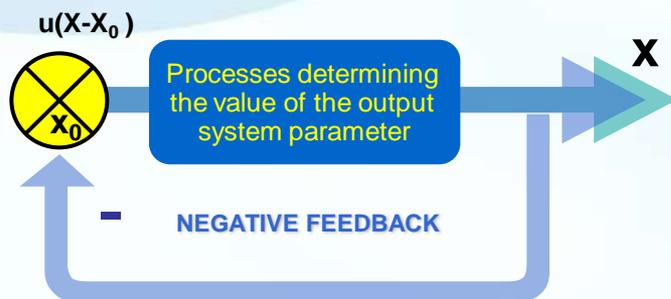
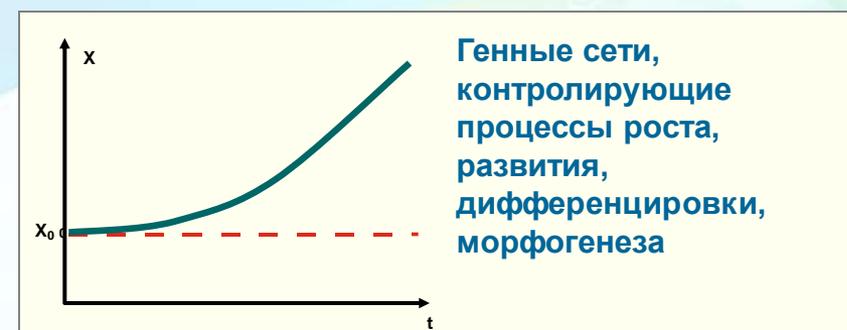
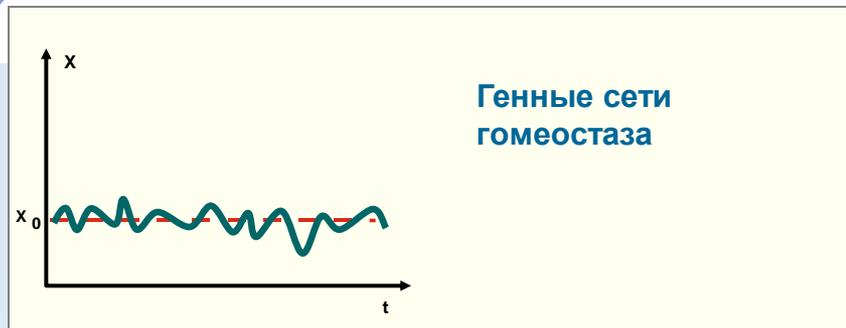
генная сеть регуляции уровня свободных радикалов и активация связанных с нею генных сетей в ходе противовоспалительного ответа организма



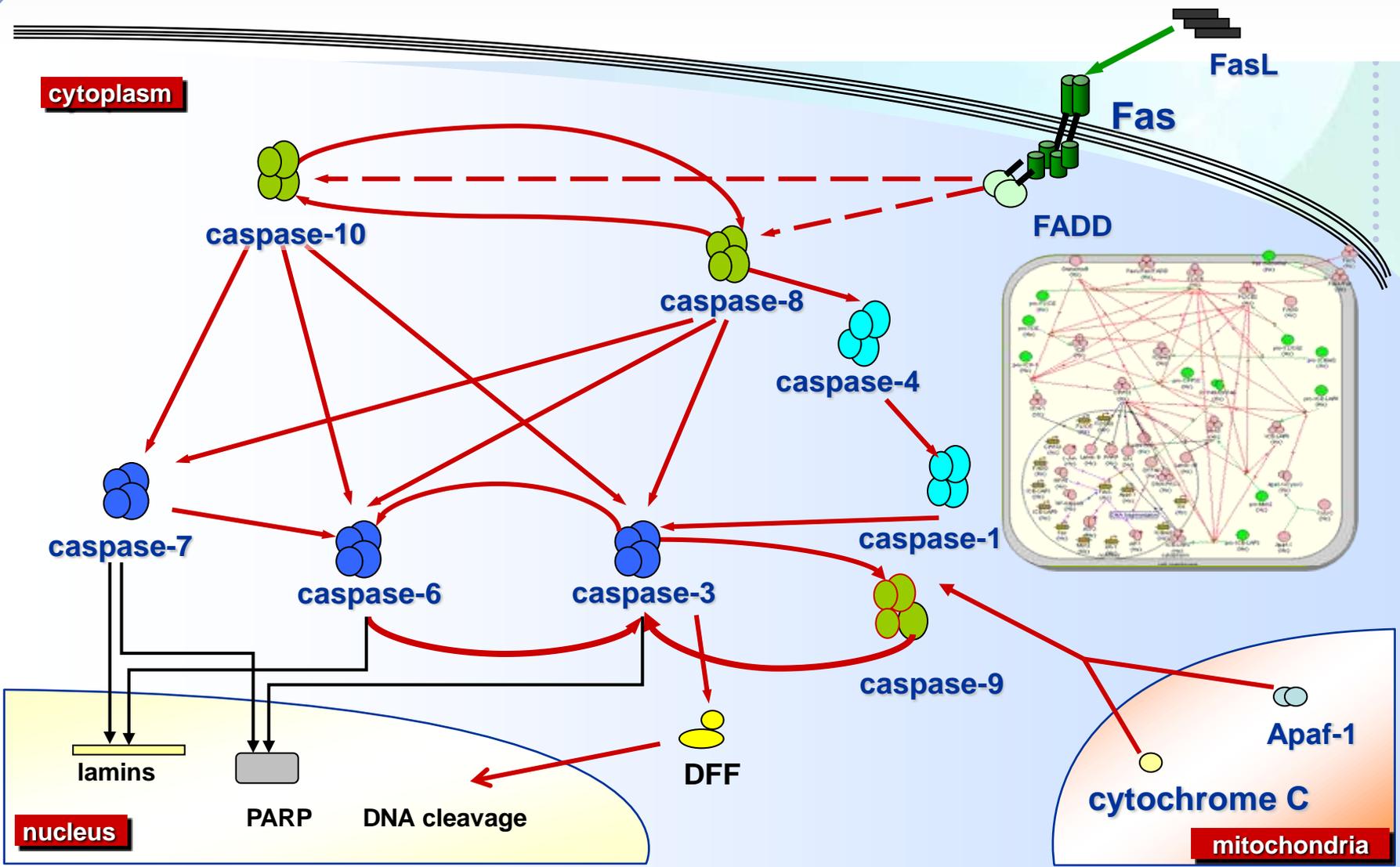
В зависимости от функционального состояния организма одна и та же генная сеть – интегратор обеспечивает комбинаторную активацию различных каскадов генных сетей.

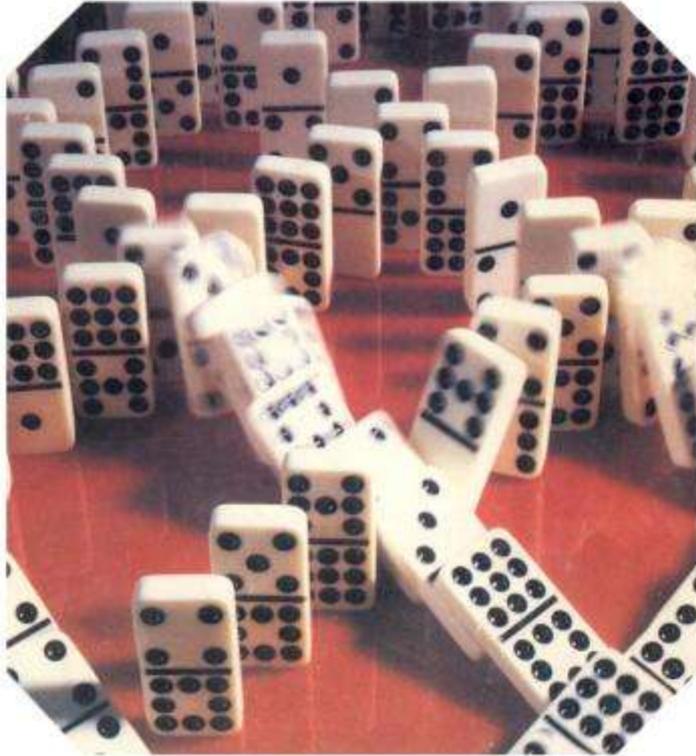


Основные типы генных сетей, описанных в базе данных GeneNet



Усиление сверхслабого сигнала в генной сети активации апоптоза на основе каскада положительных обратных связей

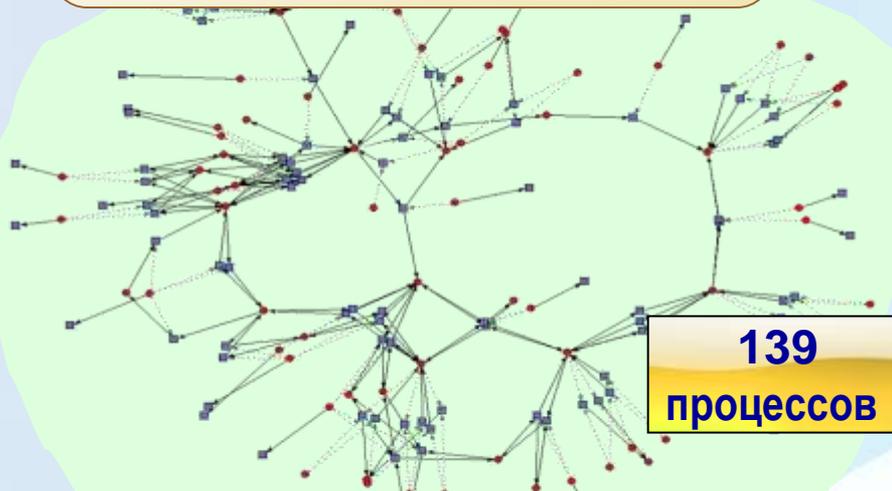




**САМООРГАНИЗОВАННАЯ СВЕРХКРИТИЧНОСТЬ: СИСТЕМА
ВОЗБУЖДАЕТСЯ ПОД ДЕЙСТВИЕМ СВЕРХСЛАБОГО
ВОЗДЕЙСТВИЯ**

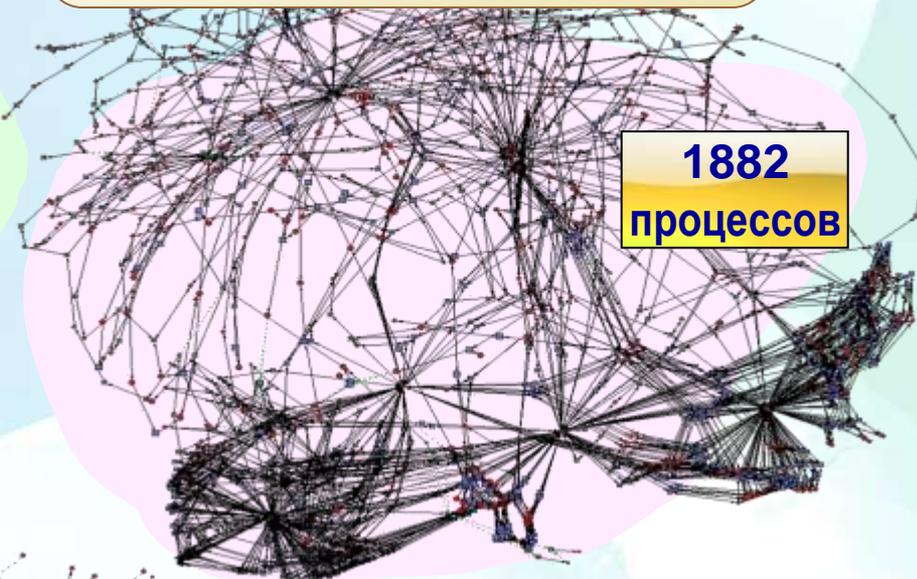
Соотношение метаболической и регуляторной компонент цикла трикарбоновых кислот *E. Coli* K-12

ИСПОЛНЯЮЩАЯ (РАБОЧАЯ)
КОМПОНЕНТА (МЕТАБОЛИЗМ)

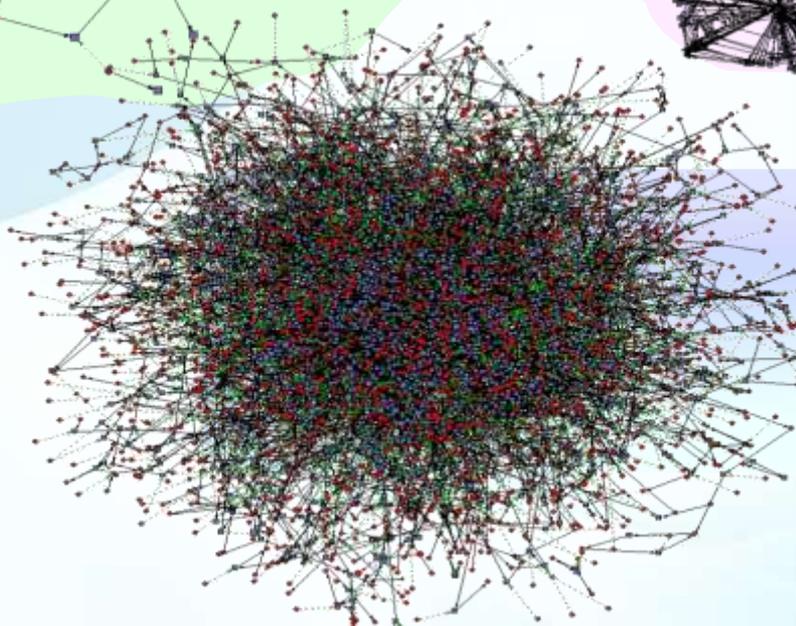


139
процессов

РЕГУЛЯТОРНАЯ КОМПОНЕНТА
(УПРАВЛЕНИЕ МЕТАБОЛИЗМОМ)

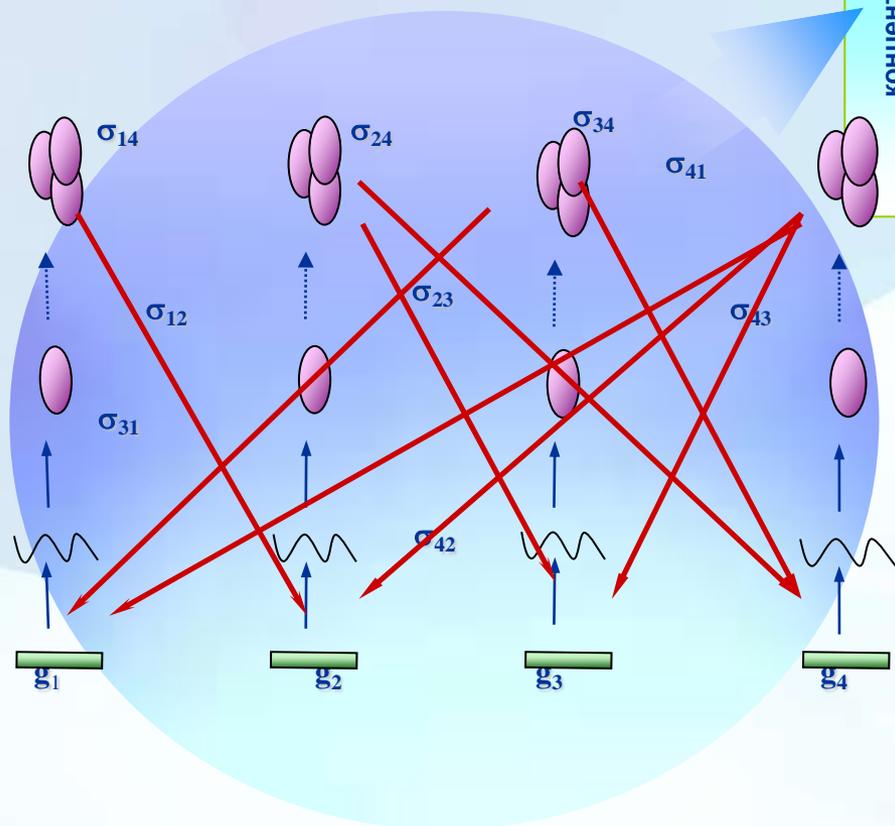


1882
процессов

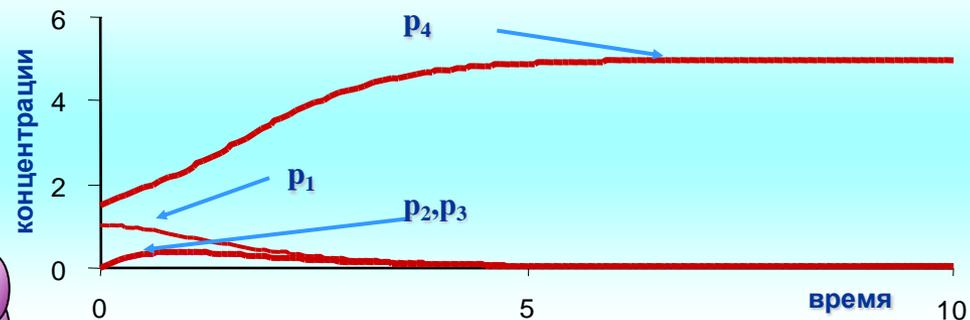


Полный граф
метаболическо
й компоненты
E. COLI K-12:
3973 процесса

Теория искусственных генных сетей (В.А. Лихошвай): граф генной задающей сеть регуляторных взаимодействий

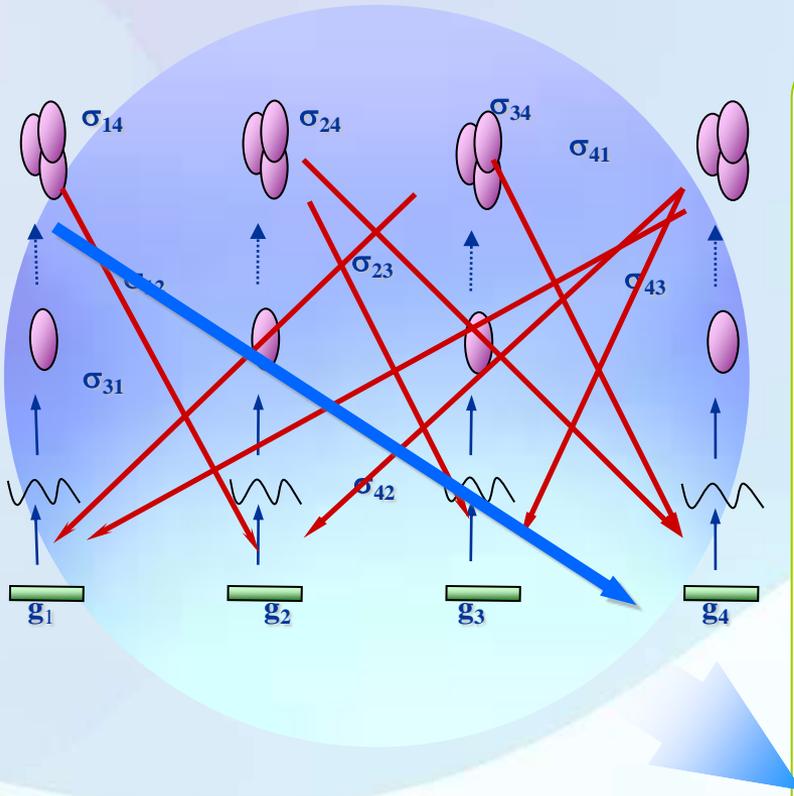


Устойчивое стационарное состояние

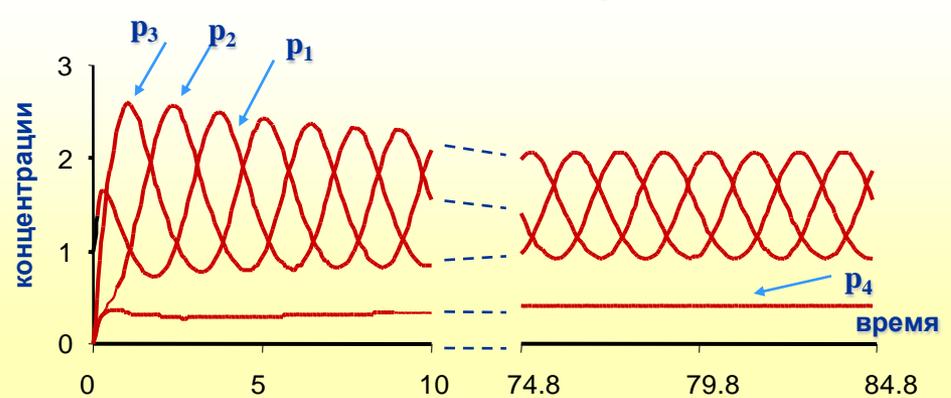


Режим функционирования
генной сети -
единственное устойчивое
стационарное состояние

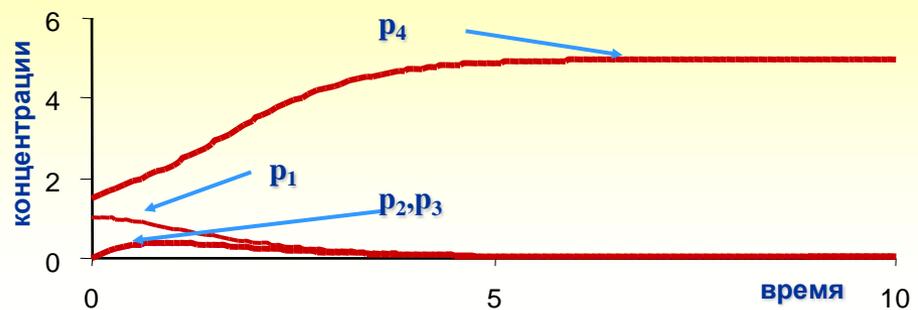
Теория искусственных генных сетей: появление новой регуляторной связи приводит к качественному изменению динамики генной сети



Устойчивый циклический режим



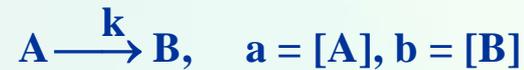
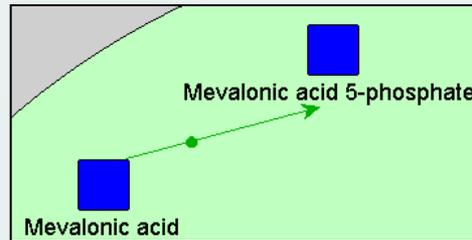
Устойчивое стационарное состояние



Два качественно различных режима функционирования:
(1) устойчивый цикл;
(2) устойчивое стационарное состояние.

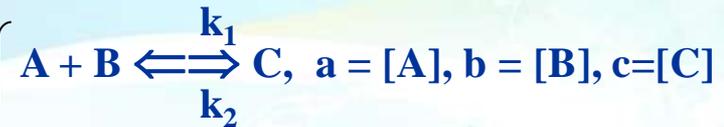
Examples of description of elementary events in the gene network of cholesterol biosynthesis in a cell within the framework of a chemical-kinetic approach

Monomolecular reaction



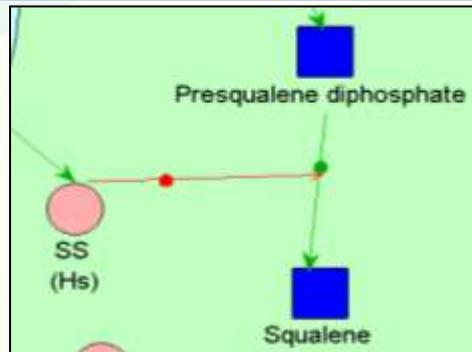
$$-da/db = db/dt = ka$$

Bimolecular reaction



$$da/dt = db/dt = -dc/dt = -k_1ab + k_2c$$

Reaction of enzymatic synthesis (Michaelis-Menten equation)



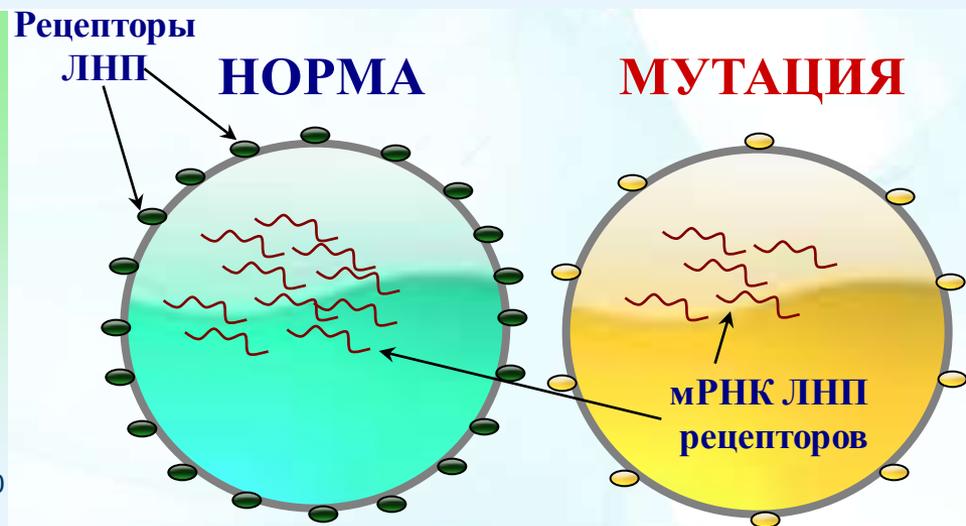
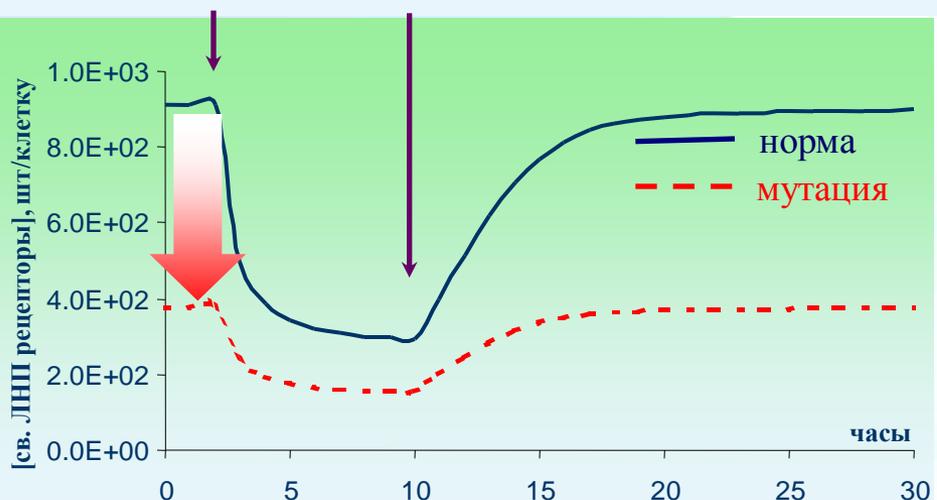
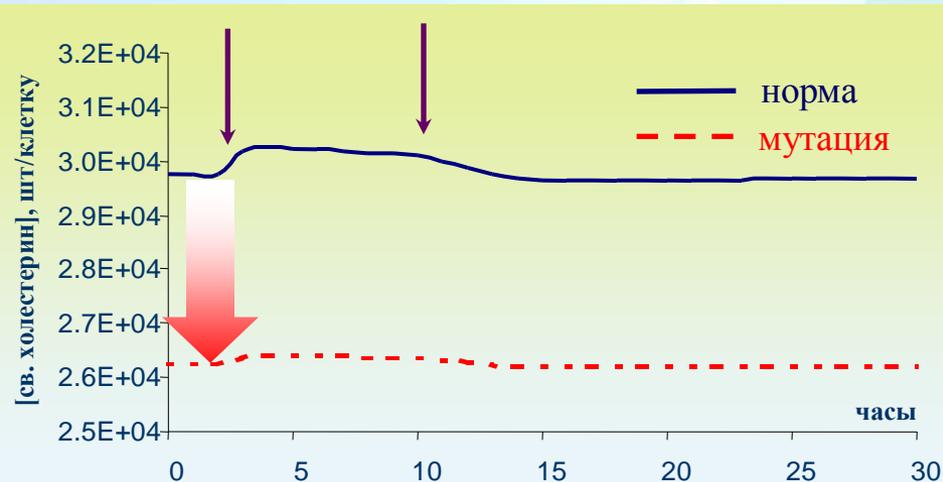
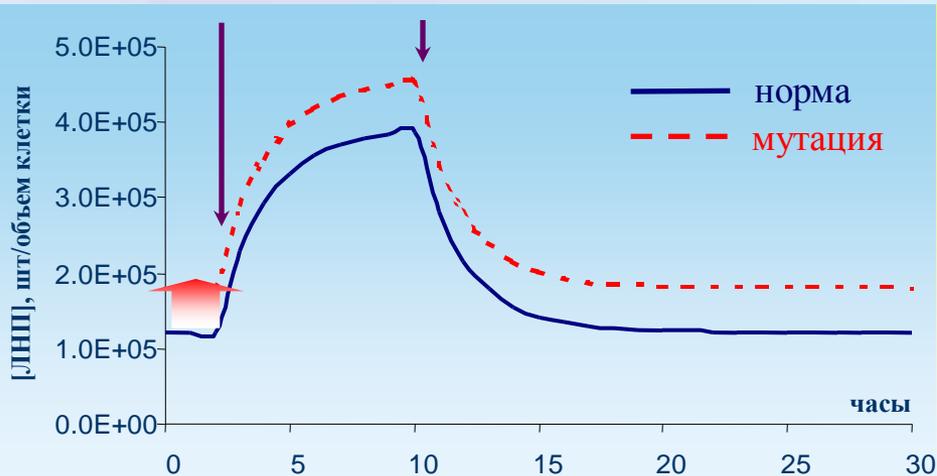
$$a = [A], e = [E], e_a = [AE], b = [B]$$

$$da/dt = -db/dt = -k_3e_0a/(K_m + a), \quad e_0 = e + e_a$$

$$K_m = (k_2 + k_3)/k_1$$

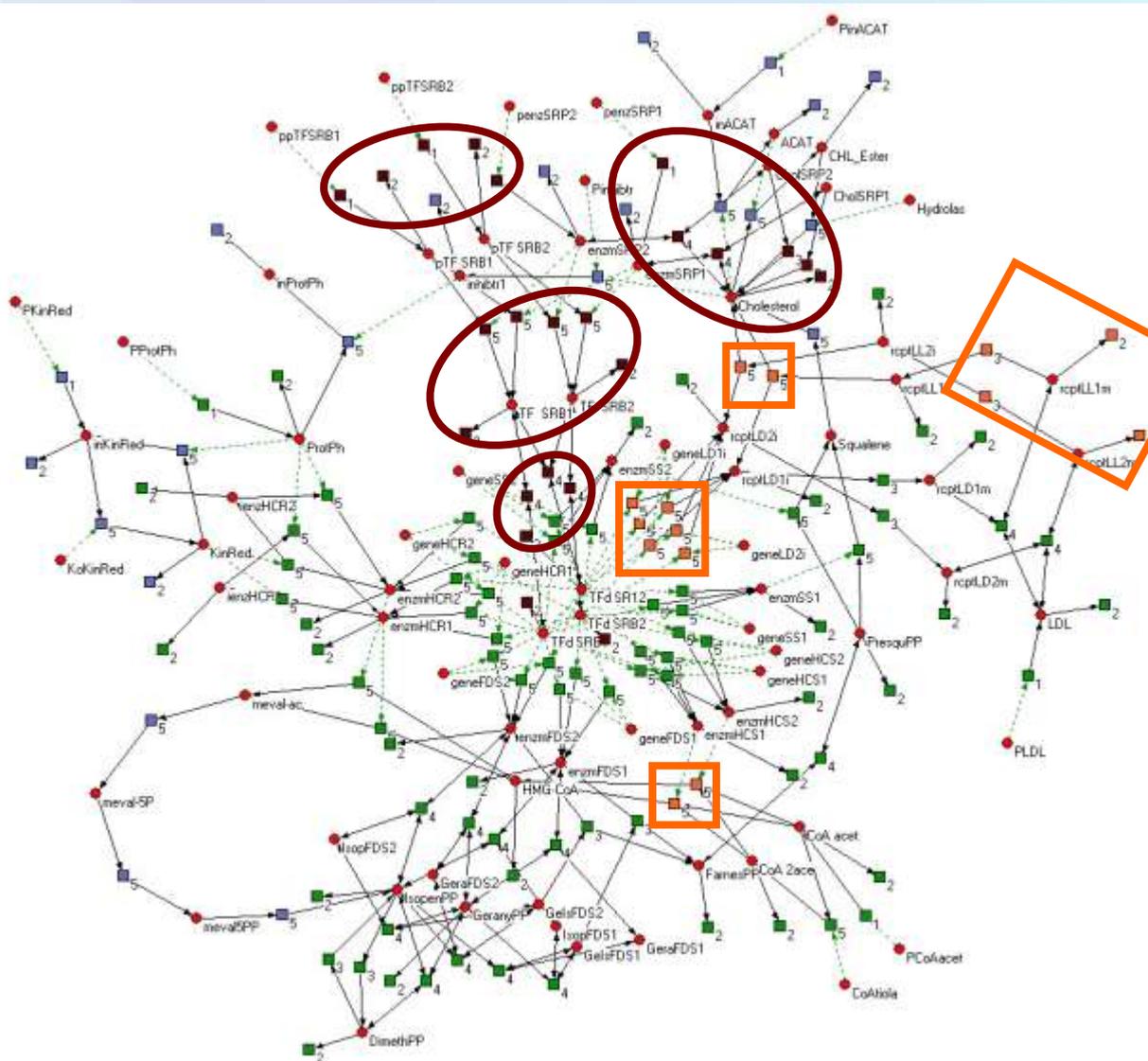
Моделирование влияния мутаций на функцию генной сети регуляции биосинтеза холестерина в клетке

Моделирование отклика генной сети на увеличенное в 2 раза поступления ЛНП в плазму крови при мутации, уменьшающей скорость экспрессии гена ЛНП рецепторов в 2 раза





Мутационный портрет генной сети регуляции биосинтеза холестерина в клетке

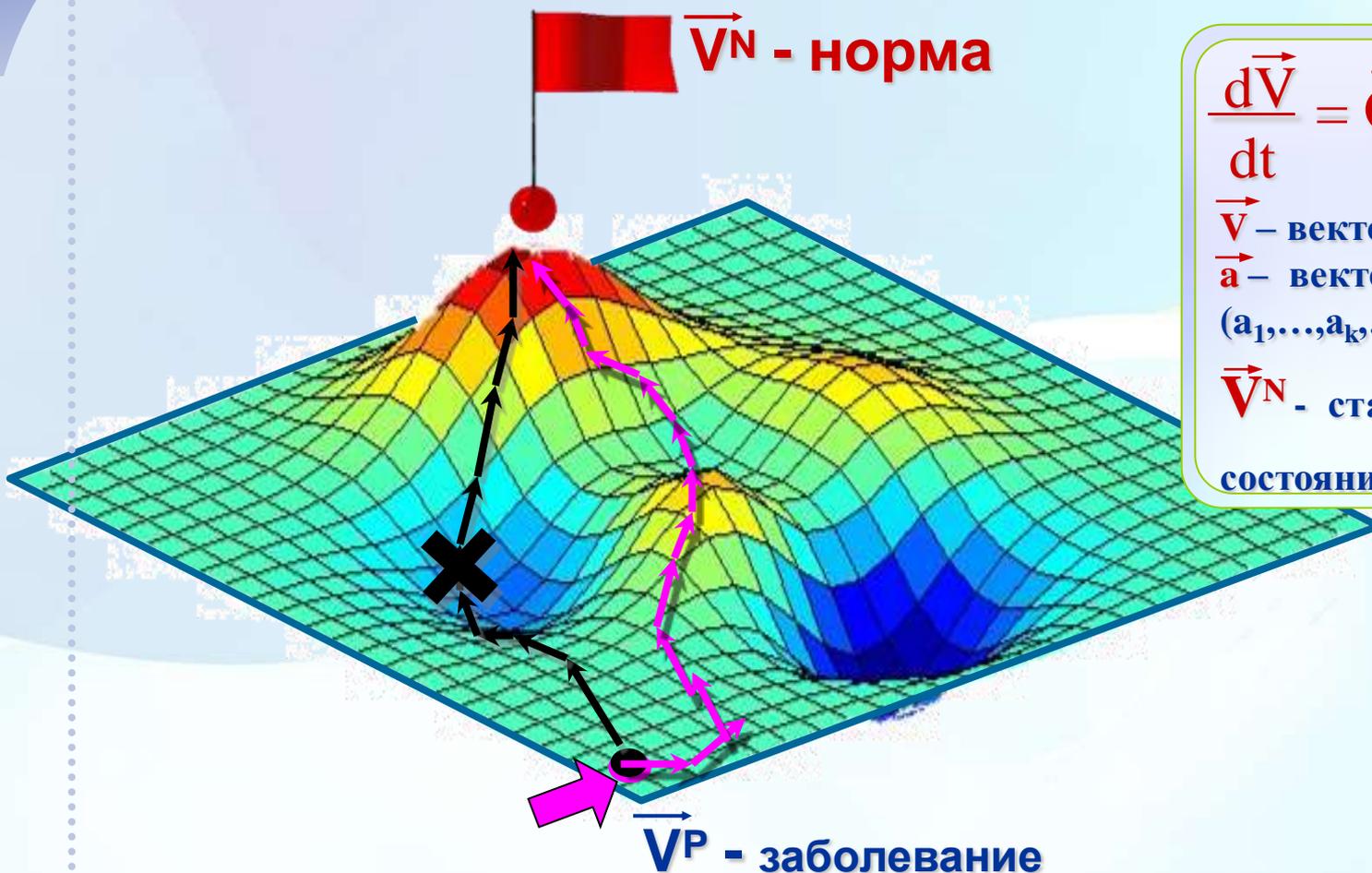


■ Changes in the rates of these processes affect considerably the stationary cholesterol concentration, changing it from 0 to over 200% of the norm;

■ Changes in the stationary concentration of free cholesterol are below 35% of the norm;



Управление функцией генных сетей



$$\frac{d\vec{V}}{dt} = \vec{\varphi}(\vec{V}, \vec{a})$$

\vec{V} - вектор переменных

\vec{a} - вектор параметров
($a_1, \dots, a_k, \dots, a_1, \dots, a_r$)

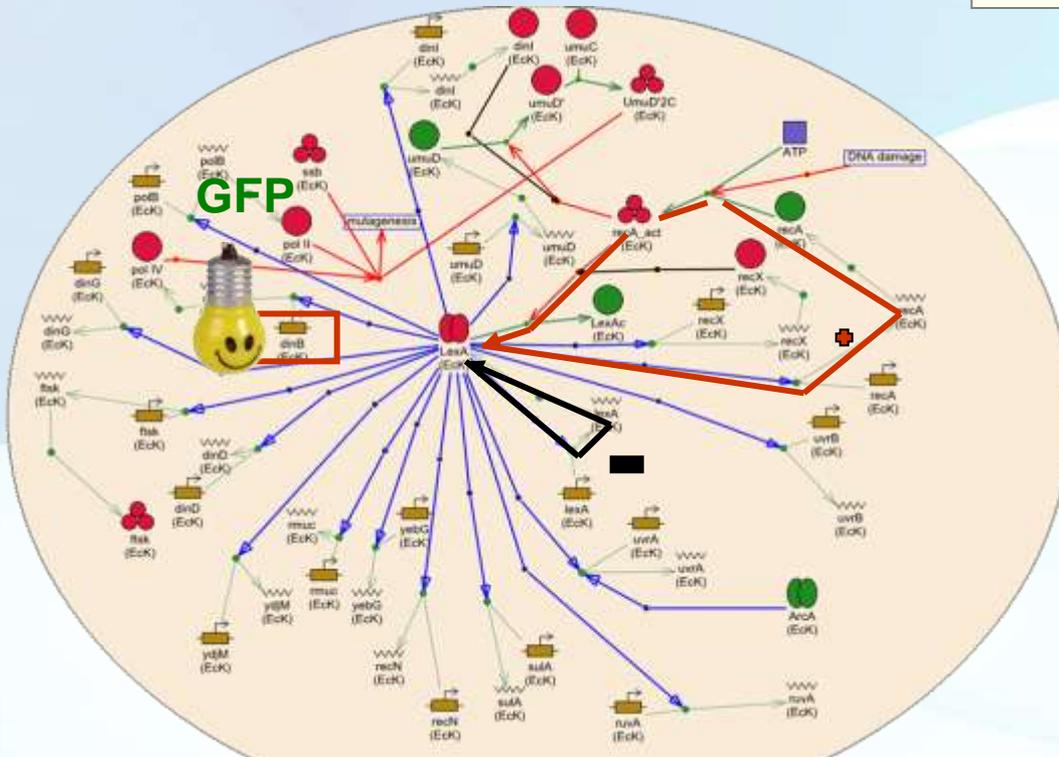
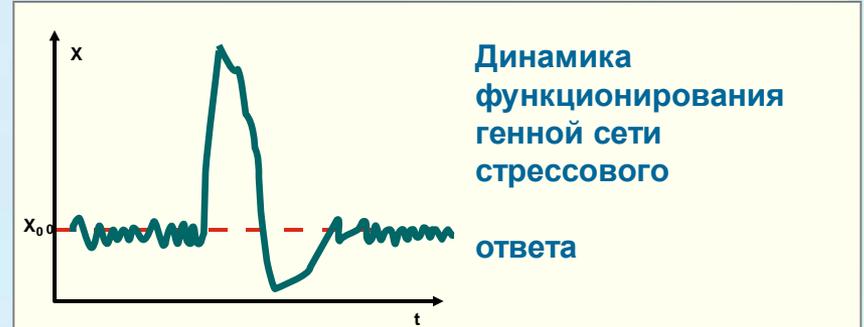
\vec{V}^N - стационарное

состояние нормы

Генная сеть стрессового ответа, как природная платформа для создания геносенсора

(лаборатория теоретической генетики)

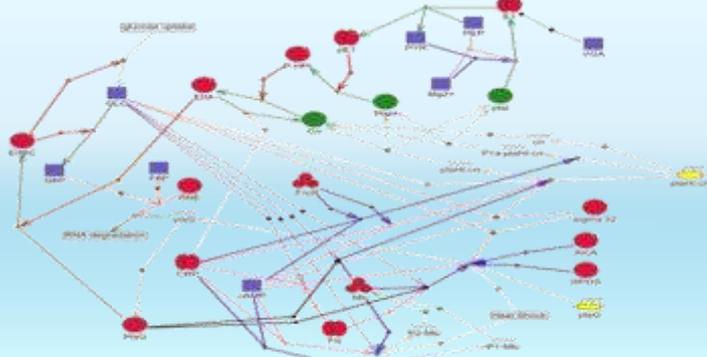
Генная сеть SOS ответа на повреждение ДНК у *E.coli*. Геносенсором может служить *dinB*, ген ДНК полимеразы IV. Динамику функционирования генной сети определяют контуры с положительной обратной связью (белок LexA ингибирует транскрипцию своего репрессора *recA*) и отрицательной (ингибирует транскрипцию своего гена *lexA*)



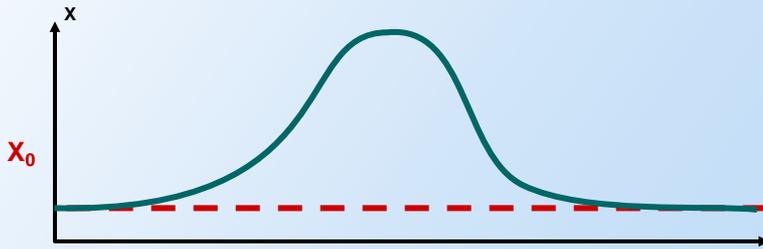
Компьютерно-экспериментальный дизайн СТРЕССЕЛЯТОРА – искусственной генной сети, активируемой стрессовыми факторами внешней среды (ИЦиГ СО РАН)

Компьютерный дизайн

База данных генных сетей

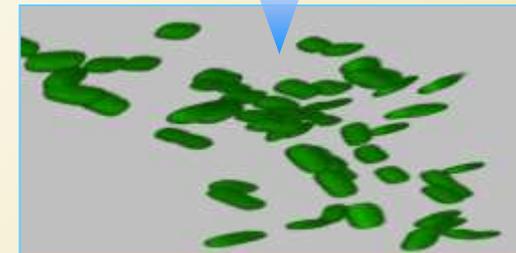


$$\frac{dp_i}{dt} = \alpha_i / (1 + \sum_{j \in D_i} \beta_{i,j} p_j^{\gamma_{i,j}}) - \beta_i p_i$$



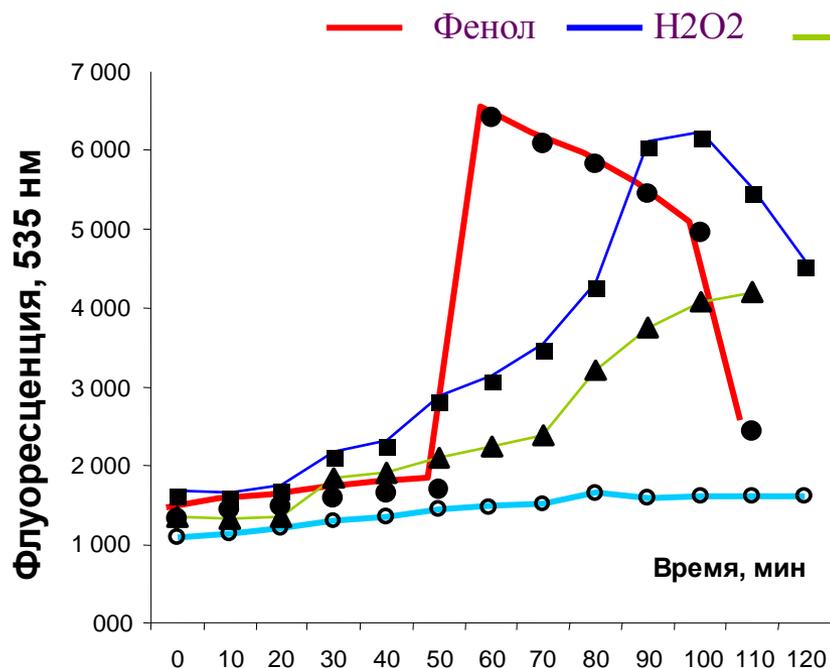
Прогнозируемая динамика
стресселятора

Экспериментальное конструирование

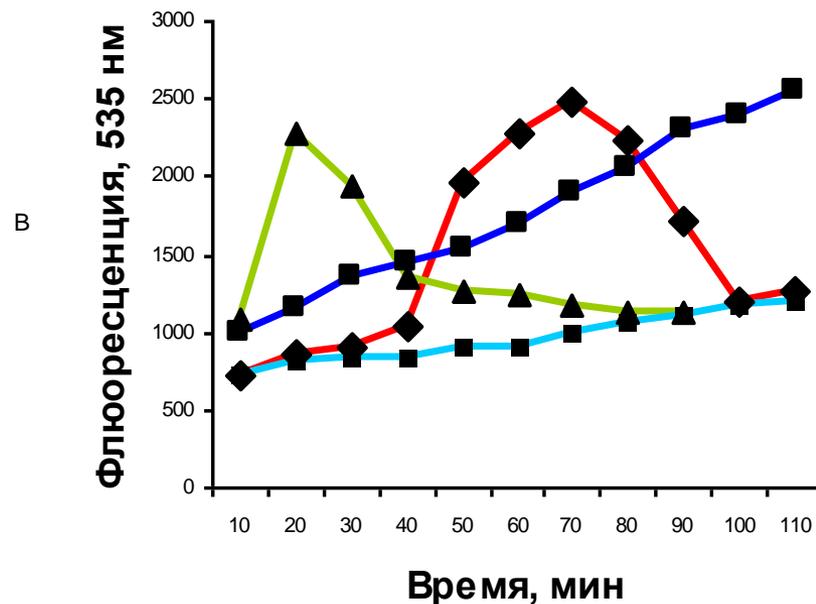


Экспериментальная проверка ответа
стресселятора по уровню GFP белка.

Реакция стресселлятора на внешние факторы



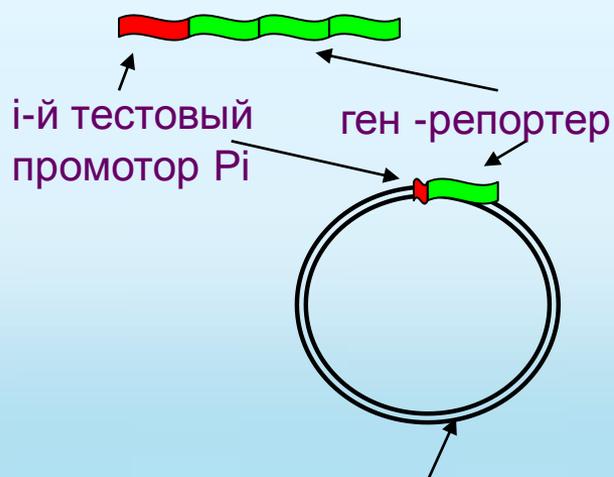
Геносенсор на основе промотора гена *ufiA* *E. coli*



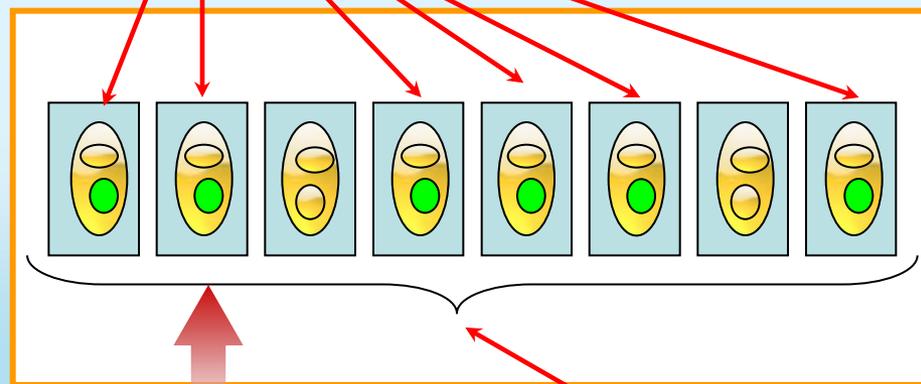
Геносенсор на основе промотора гена *dps* *E. coli*

Многопозиционный биосенсор на основе микро/нанофлюидной системы: принципиальная схема использования стресселятора

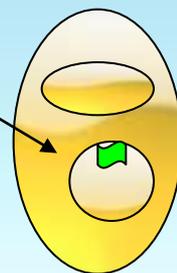
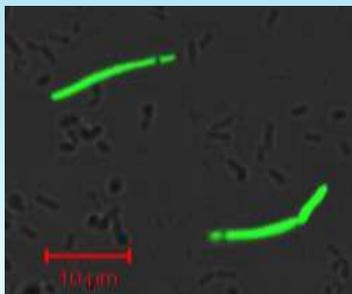
Активация геносенсора в i -м микрофлюидном канале свидетельствует о присутствии соответствующего повреждающего вещества в анализируемой жидкости



i -я гибридная плаزمида

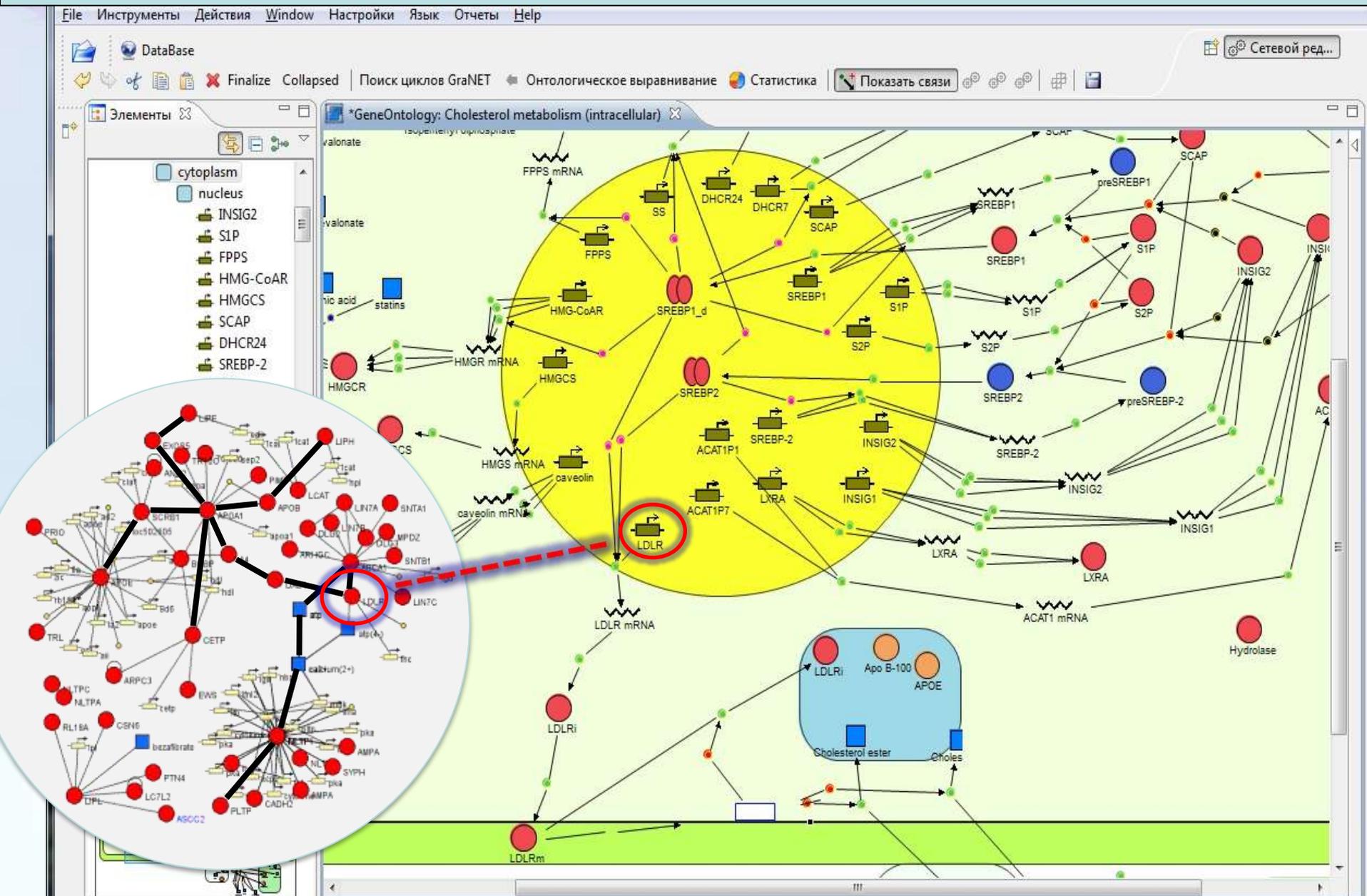


Каждый микрофлюидный канал содержит определенный геносенсор



i -я бактериальная клетка

Реконструкции генной сети биосинтеза холестерина: компьютерная система ANDCell и редактор генных сетей GeneNetStudio



ЭВОЛЮЦИОННАЯ БИОЛОГИЯ

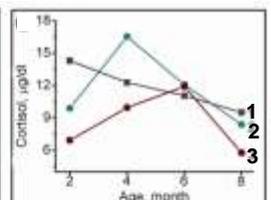
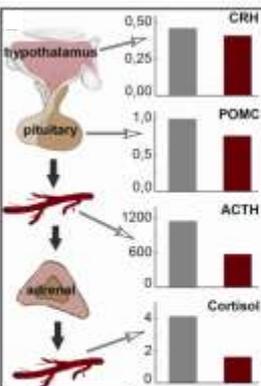


Д.К. Беляев: дестабилизирующий отбор при доместикации лис (отбор на дружественное поведение по отношению к человеку)

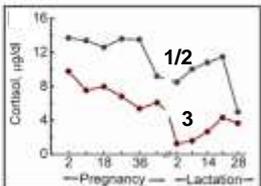
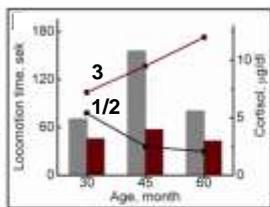
исследовательская
активность (график) на
фоне уровня кортизола
в плазме крови
(столбцы)

возраст 1–2 мес.

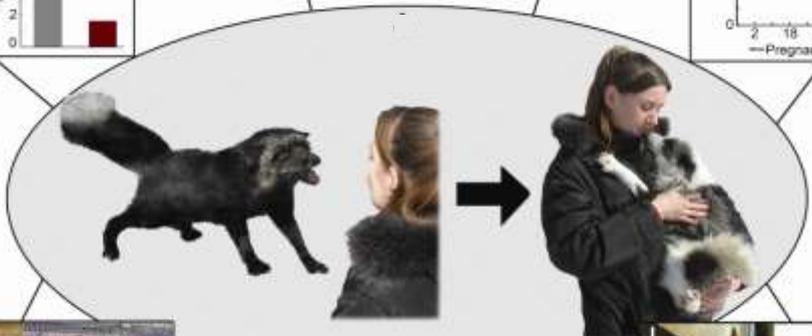
уровень
кортизола в
плазме крови
при
беременности
и лактации



возрастная динамика
кортизола в плазме
крови



показатели
активности
ГНС



1- агрессивные } дикие
2 – не агрессивные }
3 – ручные

по О.В.Трапезов,
2006, ИЦиГ СО РАН

Изменения пигментации и архитектоники у
доместичированных лисицы, соболя и норки
демонстрируют хорошие гомологические
ряды с давно одомашненными
видами.

по Л.Н.Трут, 2008, ИЦиГ СО РАН





Дестабилизирующий отбор (Д.К. Беляев)

Дикие лисицы
характеризуются
агрессивным по
отношению к человеку
поведением





Дестабилизирующий отбор: дружественное поведение одомашненных лисиц по отношению к человеку



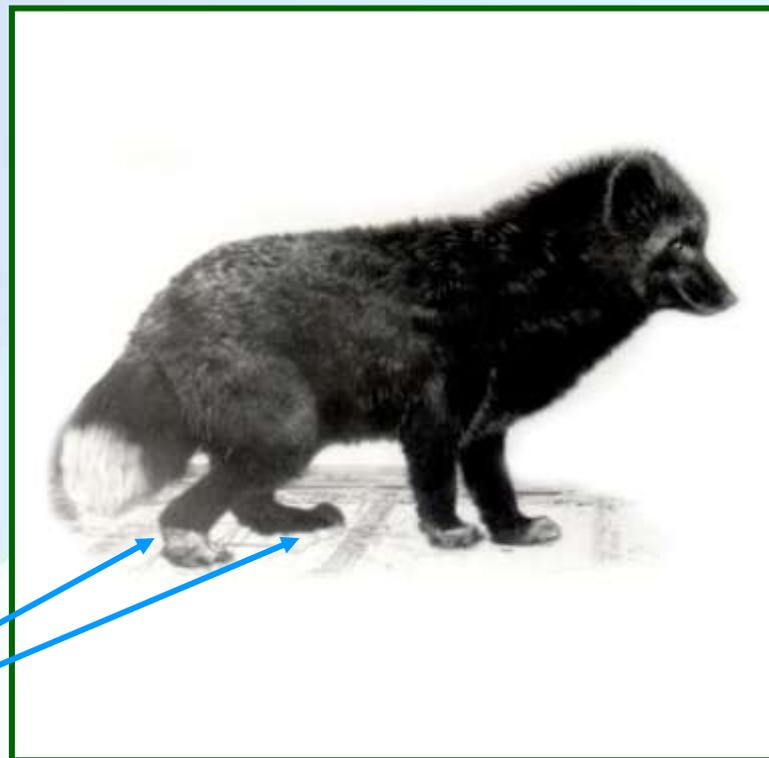


Дестабилизирующий отбор (Д.К. Беляев)

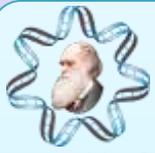
Фенотипические
характеристики
лисиц с
дружественным по
отношению к
человеку поведением



**ВЫРАЖЕННЫЕ МОРФОЛОГИЧЕСКИЕ
ОТКЛОНЕНИЯ ОТ НОРМЫ, ВОЗНИКШИЕ У
ЛИСИЦ В РЕЗУЛЬТАТЕ ДЕСТАБИЛИЗИРУЮЩЕГО
ОТБОРА**



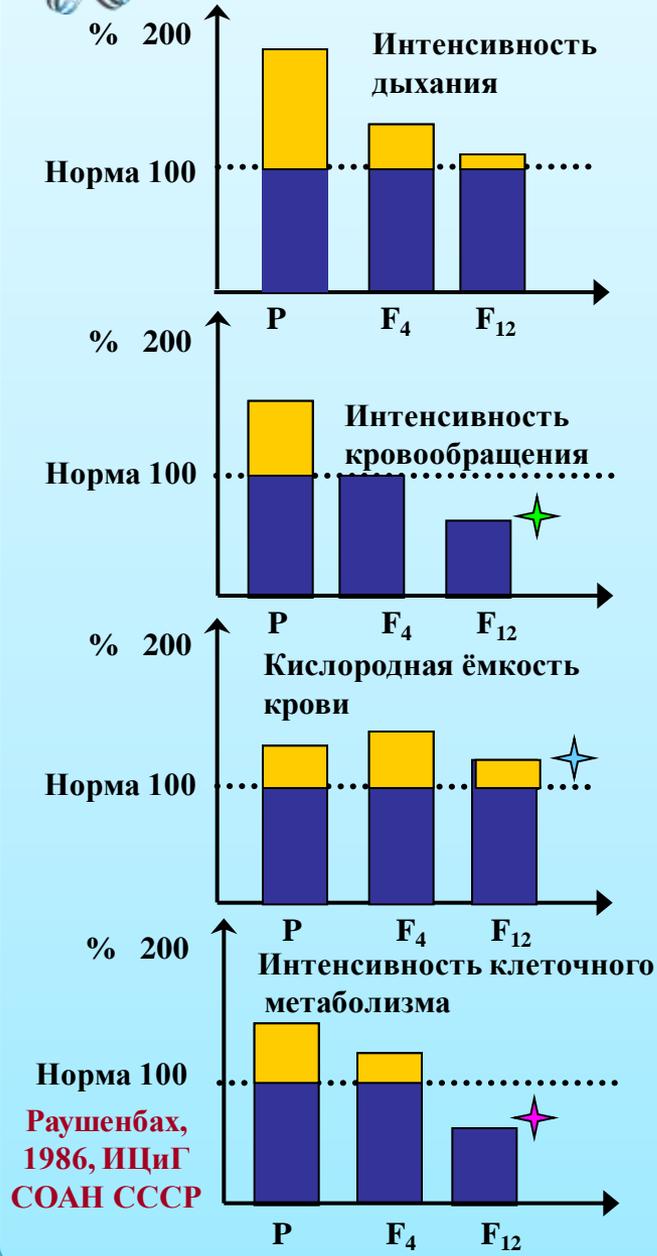
**Удлиненные
задние лапы**



Эволюционная адаптация часто копирует физиологический ответ таксонов, приспособляющихся к неблагоприятным факторам внешней среды (эффект Болдуина).



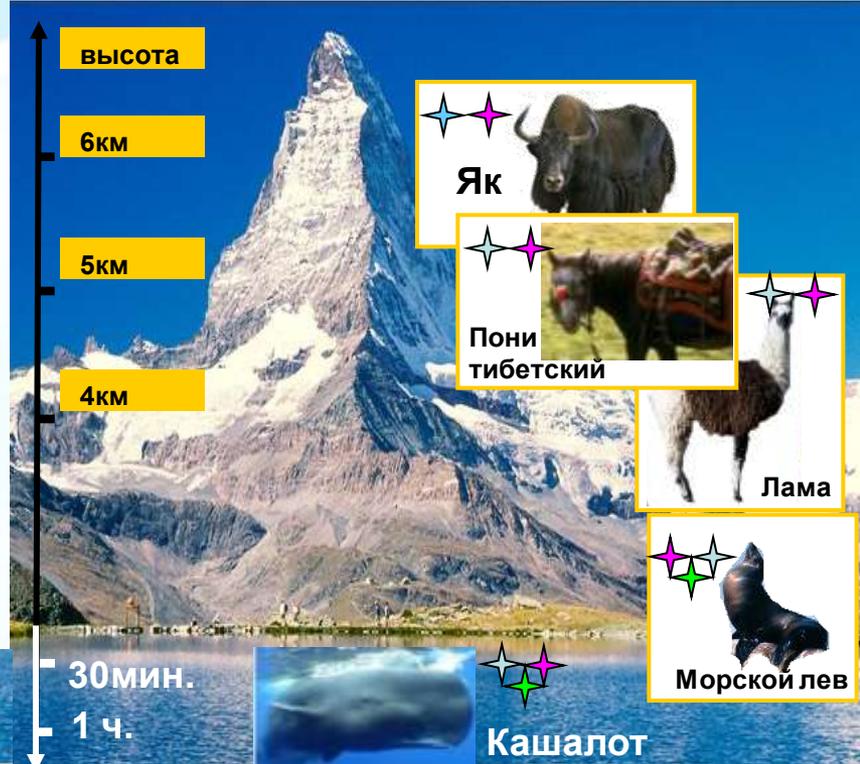
P - родители,
F₄ – потомки 4 поколения;
F₁₂ – потомки 12 поколения.



Раушенбах, 1986, ИЦиГ СОАН СССР

Адаптивная физиологическая реакция в ответ на гипоксию у копытных, попавших в горы

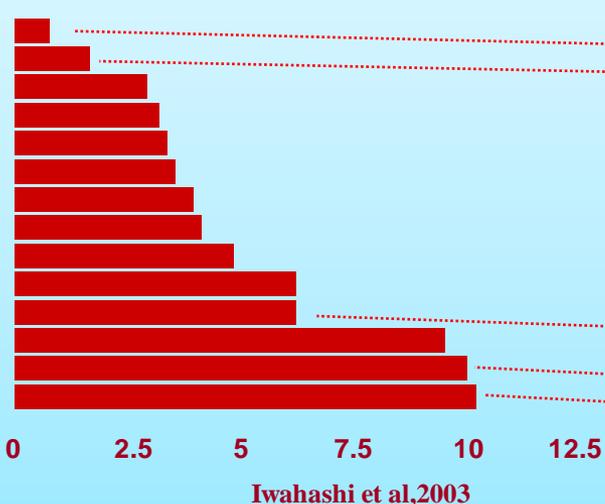
Звездочкой отмечены совпадения с видами и породами, живущими в условиях гипоксии





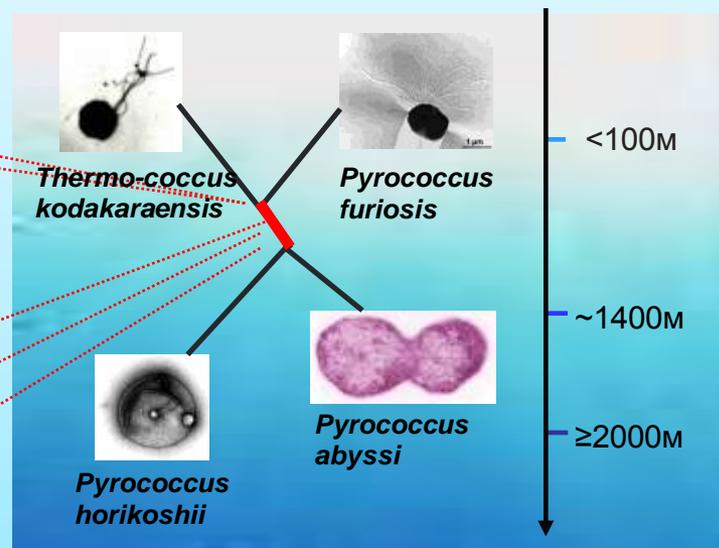
Генные сети – мишени эволюционной адаптации к неблагоприятным условиям внешней среды: адаптация к высокому давлению

S. cerevisiae: семейства генов, отвечающие на увеличение давления

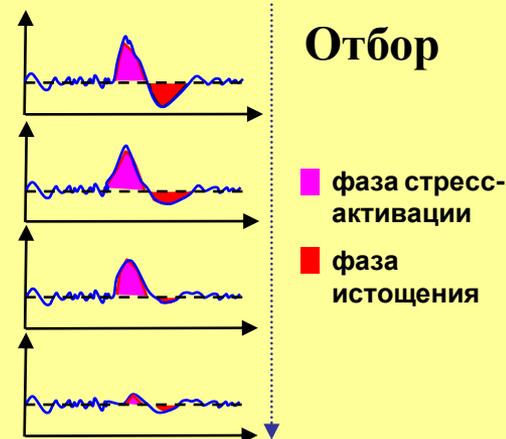


транспозоны
транспорт
синтез белков
 транскрипция
 клеточный биогенез
 ионный гомеостаз
 клеточный транспорт
 клеточный рост
 неклассифицированы
 архитектоника клетки
 межклеточные коммуникации
метаболизм
 распределение белков
стресс-системы
энергетические процессы

Адаптивная эволюция археобактерий к жизни на различных глубинах



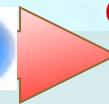
Афонников, Гунбин., 2009, ИЦИГ СО РАН





Эволюционные качели: чередование стабилизирующего и движущего отбора

СТАБИЛИЗИРУЮЩИЙ ОТБОР



ОСЛАБЛЕНИЕ или полное
РАЗРУШЕНИЕ отрицательных
обратных связей



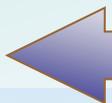
ВОЗНИКНОВЕНИЕ (или УСИЛЕНИЕ ранее существовавших) регуляторных контуров с отрицательными обратными связями.

Регуляторные механизмы

ВСКРЫТИЕ ранее замаскированной фенотипической изменчивости.

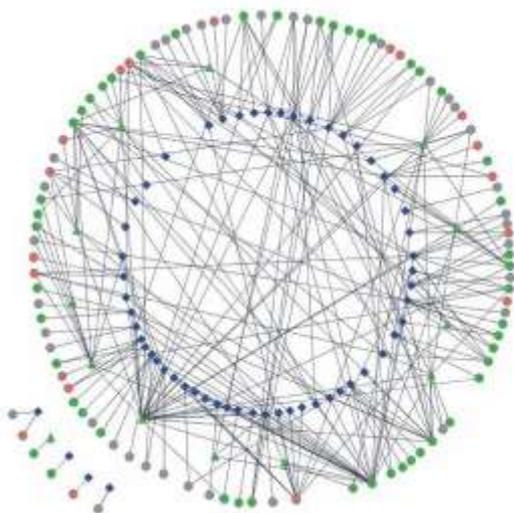


НЕЙТРАЛИЗАЦИЯ мутационных спектров и минимизация размаха фенотипической изменчивости

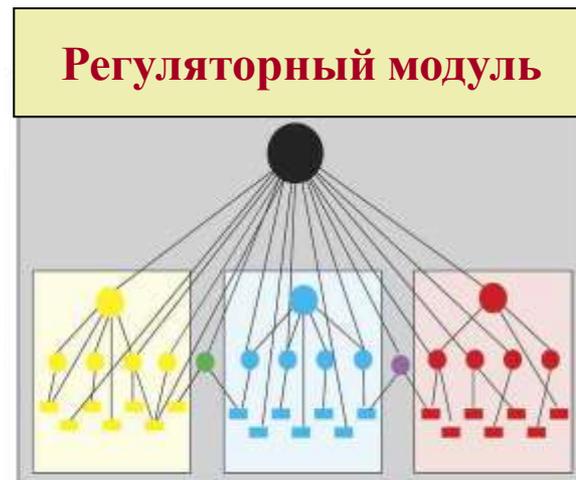


ДВИЖУЩИЙ И ДЕСТАБИЛИЗИРУЮЩИЙ ОТБОР

Суслов,
Колчанов,
Сергеев, 2006, ИЦиГ
СО РАН, НГУ



Иерархическая слоистая структура графа сети регуляции транскрипции генов кишечного тракта *S.elegans*



по Walhout, 2006

The background features a light blue and green color palette. On the right side, there is a faint, semi-transparent profile of a human head facing right. Overlaid on the head is a circular area containing a brain scan or MRI image, showing internal structures in shades of white and grey. The overall aesthetic is clean and scientific.

БИОЛОГИЯ РАЗВИТИЯ

AGNS: онтологическое описание стадий развития, органов и клеточных типов *Arabidopsis thaliana*

Wild type seedling



Словарь стадии развития:

Словарь органов

Seedling stage starts after germination. The end of seedling stage in some vocabularies is determined as the time when first true leaves emerge above the cotyledons but in many papers this stage is prolonged further.

True leaves, derivatives of the SAM, are apparent at the apex [Long J.A. and Barton M.K., 1998]. It takes on average 7.8 and 8 days for primordia of the leaf 1 and leaf 2 to reach 0.5 mm in length (Ler, LD) [Hamada S. et al., 2000]

The overall number of leaves is 9-17, on average 12.5 [Brown J.A. et al., 1964] [Schultz E.A. and Haughn G.W., 1991]

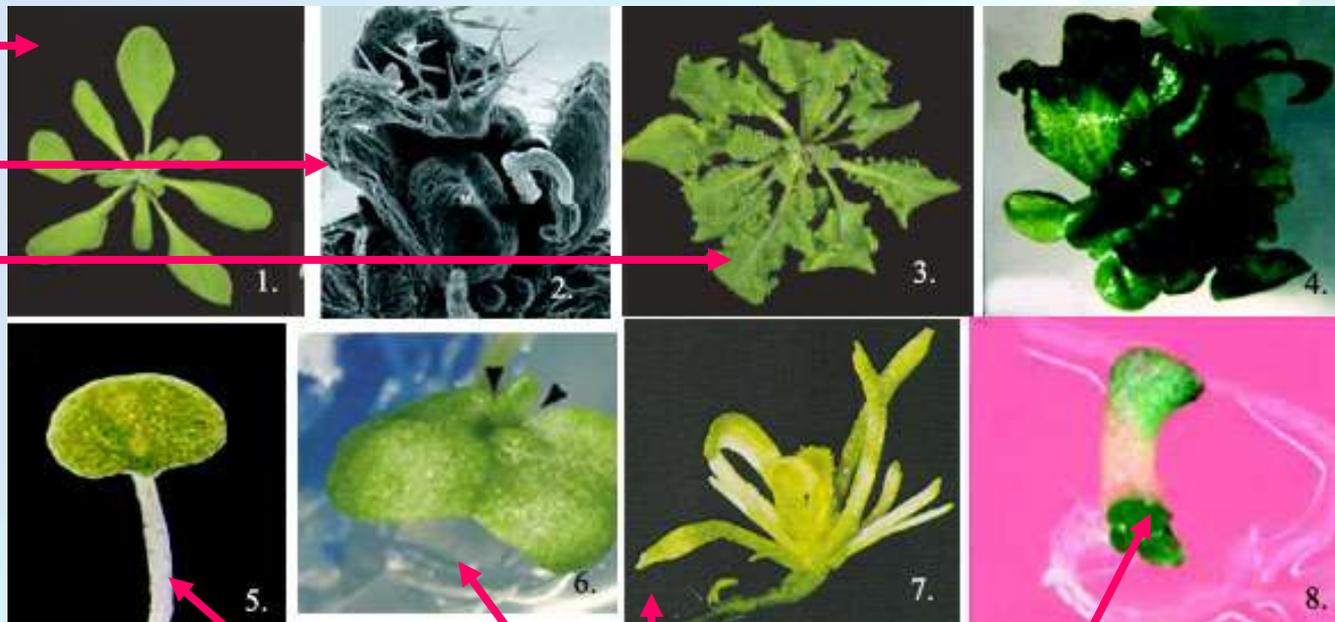
In shoot apex leaf primordia and young leaves located in spiral phyllotaxis order [Brown J.A. et al., 1964] [Schultz E.A. and Haughn G.W., 1991]. The first and second rosette leaves had round leaf blades and long petioles, in later leaves leaf blades were longer and petioles shorter. The vascular system of later leaves is more reticulated [Hamada S. et al., 2000]

AGNS: стандартизованное описание фенотипических аномалий *Arabidopsis thaliana* при мутациях (примеры для стадии проростка)

Wild type

Дезорганизованный проросток 35S::AS AtCAP1

Зазубренные розеточные листья проростка, jaw-D



Узкие розеточные листья, ago1-2

Терминация развития, tpl-1, stm1, atml1-1 pdf2-8

MA atml1-2 pdf2-8 [Abe M. et al., 2002]

MA stm-11 [Long J.A. and Barton M.K., 1998]

MA tpl-1 [Long J.A. et al., 2002]
RD mature embryo

RD seedling

RO primary SAM

FL absent or very strongly reduced and not restored at germination

AGNS: данные по экспрессии генов, контролирующих развитие арабидопсиса (пример - экспрессия мутантного гена *pin1* и фенотипические аномалии при мутации)

ID At:PIN1

XX

MA wild type [Steinmann T. et al., 1999]

MA *mp* [Steinmann T. et al., 1999]

RT protein, AB

RD flowering

RO SAM

RO flower

RL present

XX

...

XX

MA wild type [Steinmann T. et al., 1999]

RT protein, ab

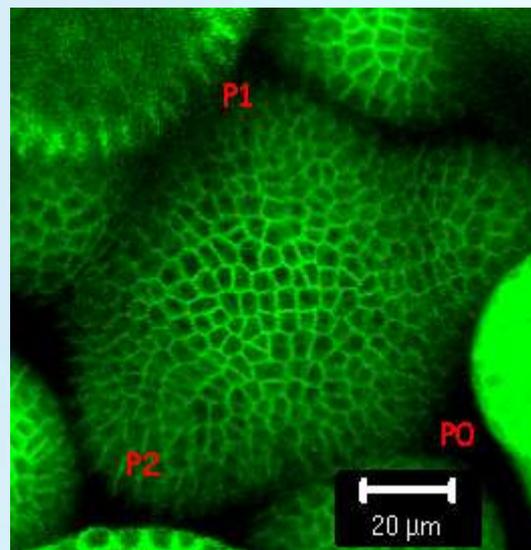
RD adult phenotype

RO siliques

RO stems

RO leaves

RL present



Arabidopsis SAM at flowering stage produces floral primordia at its flank in regular, spiral, clockwise or counterclockwise patterns [Smyth D.R. et al., 1990].



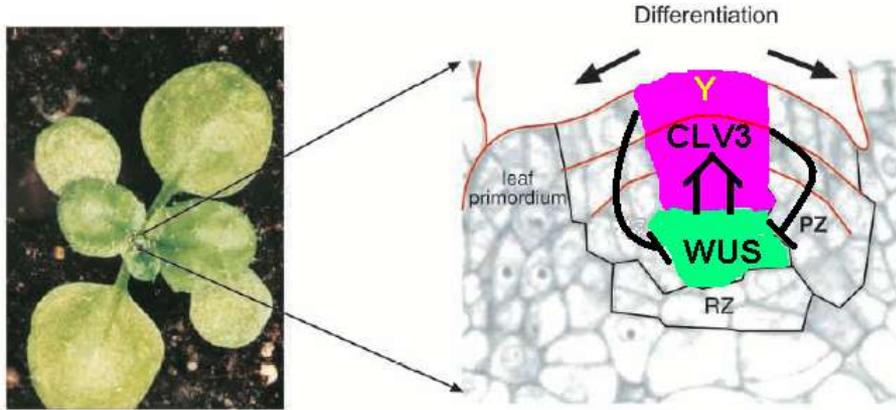
MA *pin1* [Reinhardt D. et al., 2000]
RD flowering
RO SAM
FL pin-like structure

Моделирование регуляции структуры ниши стволовых клеток в

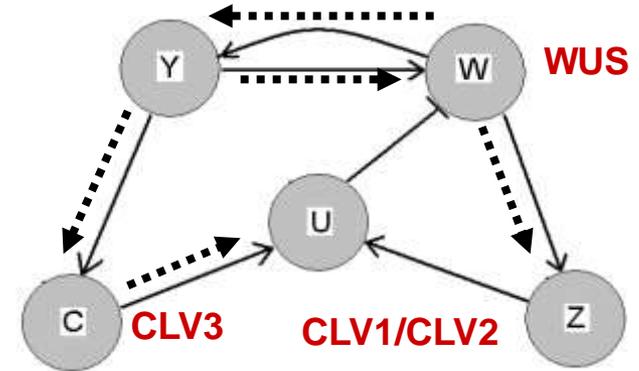
апикальной меристеме побега. Лаборатория теоретической генетики (С.В. Николаев, С.И.Фадеев)

Ниша стволовых клеток в апикальной меристеме побега соответствует зоне экспрессии гена **CLV3**. Согласно «классической» модели, гены **CLV3** и **WUS**, участвующие в регуляции структуры меристемы, экспрессируются в разных зонах, при этом **CLV3** подавляет экспрессию **WUS**, а последний, в свою очередь, активирует экспрессию **CLV3**.

адаптировано из *Groß-Hardt R. and Laux T., 2003*



Феноменологические представления «классической» модели были дополнены предположением о существовании фактора **Y**, играющего важную роль в регуляции генной сети поддержания структуры меристемы побега, и предположениями о диффузии **Y**, **W**, **C**:



Качественные представления были формализованы в виде математической модели. Показано, что стационарное решение модели качественно согласуется с наблюдаемой локализацией синтеза **CLV3** и **WUS**:

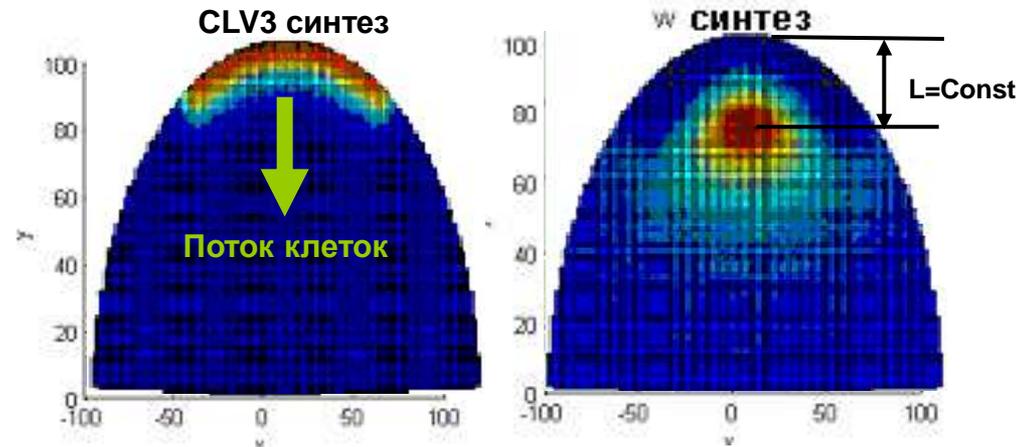
$$\frac{\partial Y}{\partial t} = D_Y \cdot \Delta Y - d_Y \cdot Y, \quad \frac{\partial Y}{\partial n_{r_1}} = \frac{1}{\tau_Y} g(h_Y + T_{YW}W),$$

$$\frac{\partial W}{\partial t} = \frac{1}{\tau_W} \cdot g(h_W + T_{WY} \cdot Y + T_{WW} \cdot W + T_{WU} \cdot C \cdot Z) + D_W \cdot \Delta W - d_W \cdot W$$

$$\frac{\partial C}{\partial t} = \frac{1}{\tau_C} \cdot g(h_C + T_{CY} \cdot Y) + D_C \cdot \Delta C - d_C \cdot C - \alpha \cdot C \cdot Z$$

$$\frac{\partial Z}{\partial t} = \frac{1}{\tau_Z} \cdot g(h_Z + T_{ZW} \cdot W) - d_Z \cdot Z - \alpha \cdot C \cdot Z$$

$$\frac{\partial Y}{\partial n_{r_0}} = 0; \frac{\partial C}{\partial n_{r_1}} = 0; \frac{\partial W}{\partial n_{r_1}} = 0; \frac{\partial Z}{\partial n_{r_1}} = 0$$



**БИОИНФОРМАТИКА:
АВТОМАТИЧЕСКИЙ
АНАЛИЗ БАЗ ДАННЫХ И
ТЕКСТОВ НАУЧНЫХ
ПУБЛИКАЦИЙ**

Информационный взрыв в современной науке (биологический домен)

Геномика



Протеомика



Метаболомика



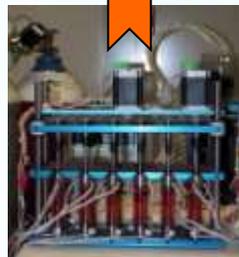
Более 50 миллионов биологических публикаций
Тысячи биологических баз данных



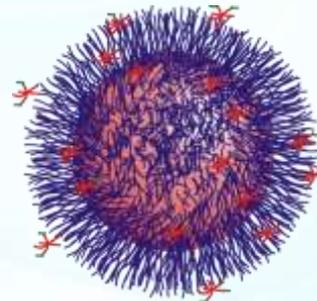
Более 2 миллионов патентов



Фармакология



Биотехнология



Нанотехнология

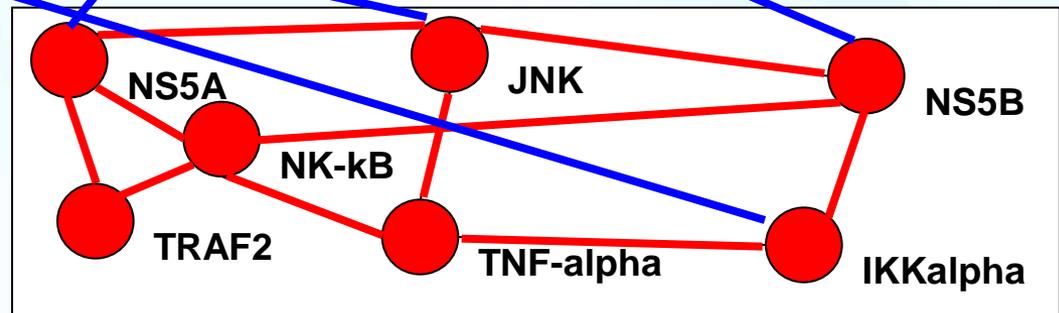


Агробиология

Реконструкция ассоциативных сетей: графическое представление функциональных взаимодействий белков вируса гепатита С (NS5A и NS5B) с белками человека

Hepatitis C virus (HCV) NS5B protein is a membrane-associated phosphoprotein that possesses an RNA-dependent RNA polymerase activity. We recently reported that **NS5A protein** interacts with **TRAF2** and **modulates tumor necrosis factor alpha (TNF-alpha)-induced NF-kappaB** and **Jun N-terminal protein kinase (JNK)**. Since **NS5A** and **NS5B** are the essential components of the **HCV** replication complex, we examined whether **NS5B** could **modulate** TNF-alpha-induced **NF-kappaB** and **JNK** activation. In this study, we have demonstrated that TNF-alpha-induced **NF-kappaB** activation is **inhibited** by **NS5B protein** in HEK293 and hepatic cells. Furthermore, **NS5B protein** **inhibited** both **TRAF2-** and **IKK-induced NF-kappaB** activation. Using coimmunoprecipitation assays, we show that **NS5B** **interacts** with **IKKalpha**. Most importantly, **NS5B protein** in **HCV** subgenomic replicon cells **interacted** with endogenous **IKKalpha**, and then **TNF-alpha-mediated IKKalpha** kinase activation was significantly **decreased** by **NS5B**. Using in vitro kinase assay, we have further found that **NS5B protein** synergistically **activated** TNF-alpha-mediated **JNK** activity in HEK293 and hepatic cells. These data suggest that **NS5B protein** **modulates** **TNF-alpha** signaling pathways and may contribute to HCV pathogenesis.

Красный – названия объектов
Синий – слова-связки
Зеленый - организмы
Розовый – детерминанты
типа объекта

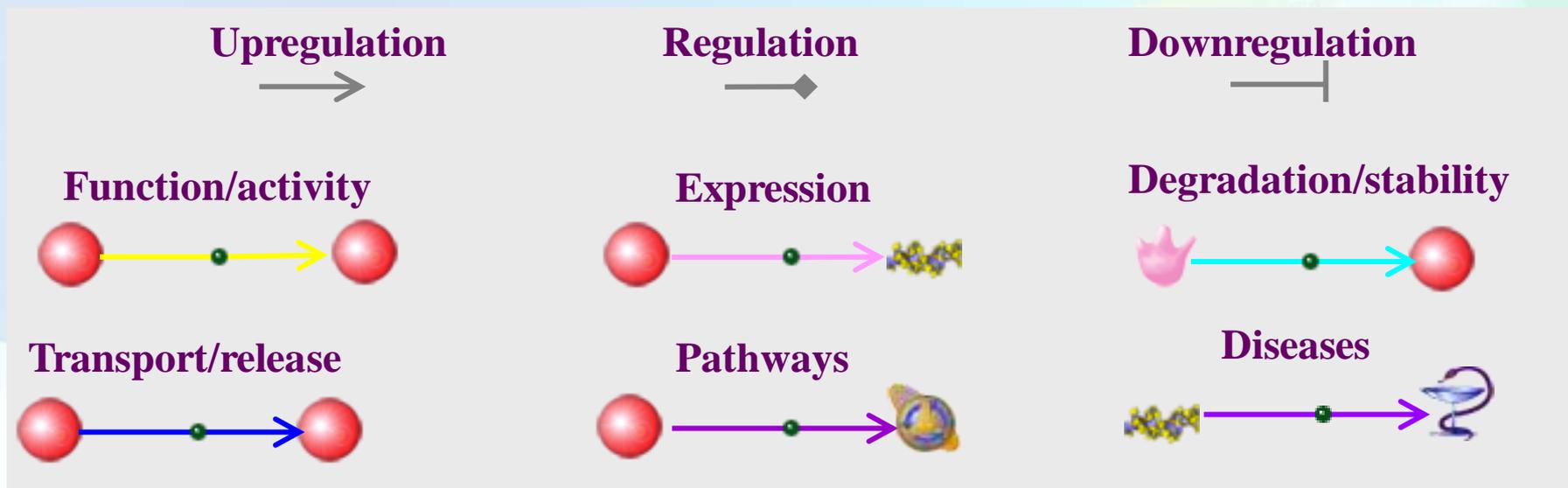
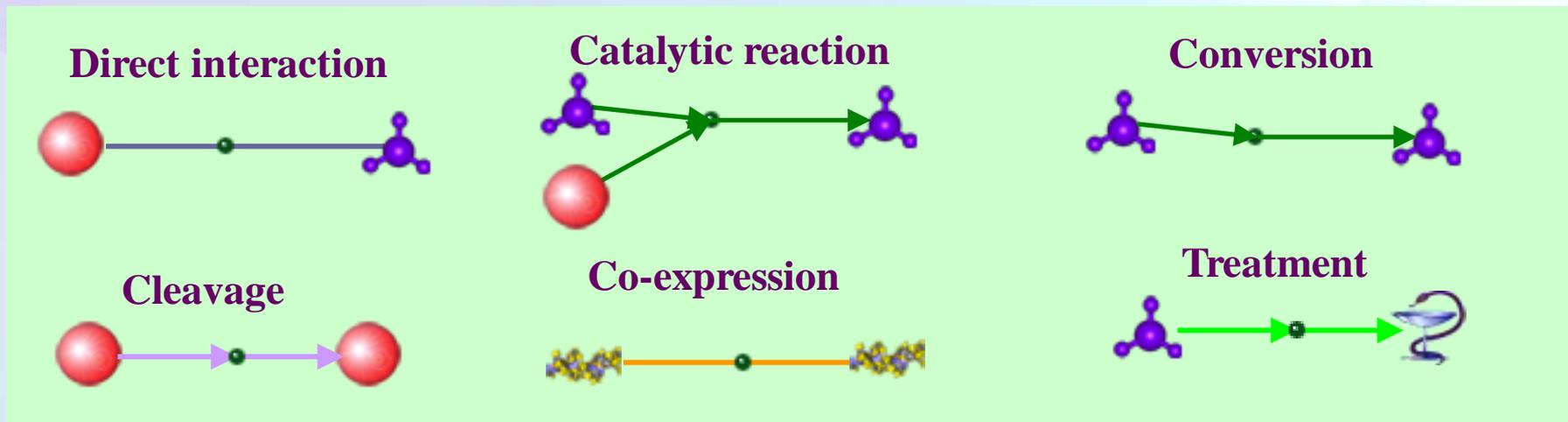


Dictionary of molecular-genetic object names

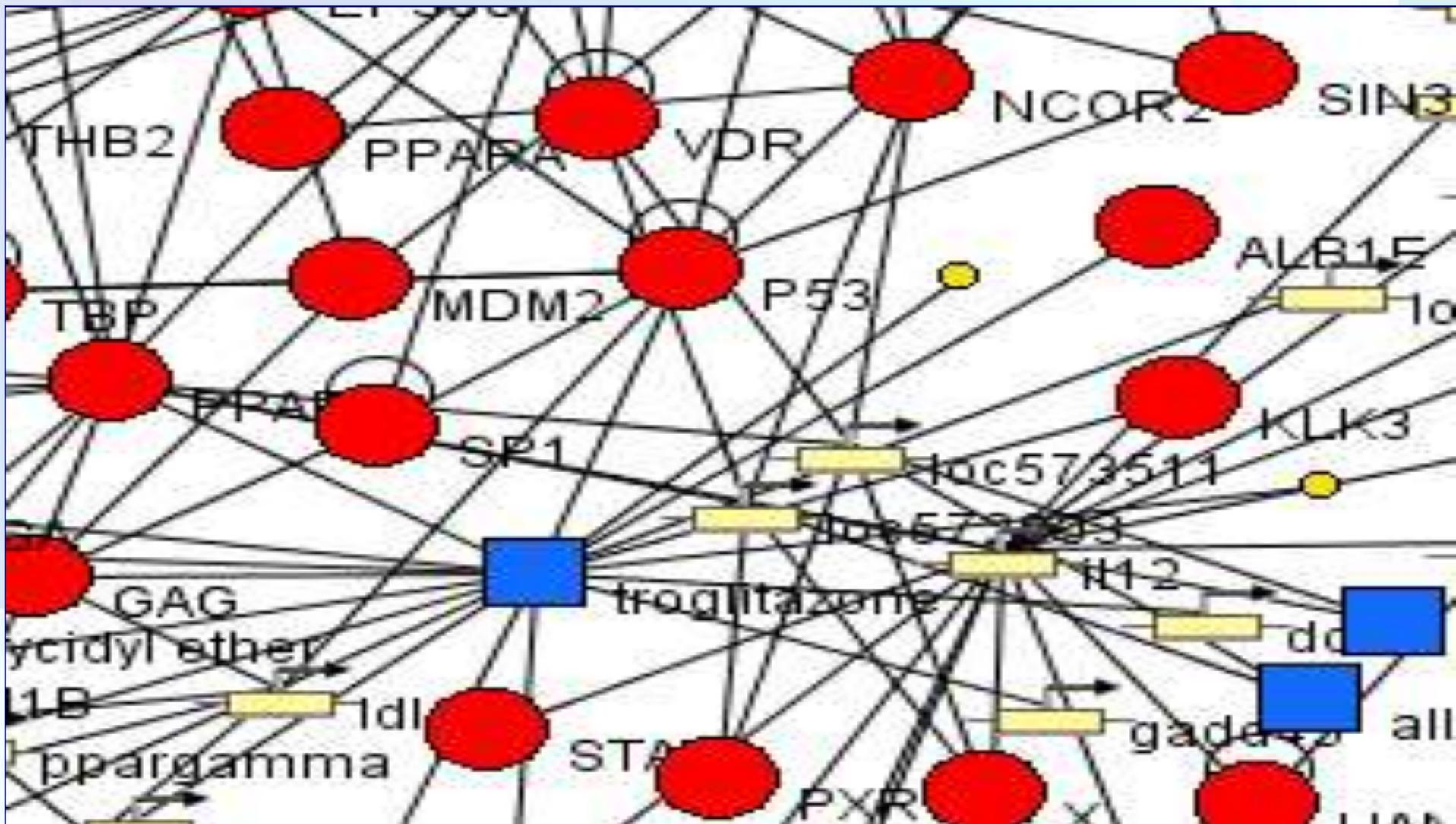


Protein names		370097
Gene names		2474201
Disease names		46109
Metabolite names		79914
Organism names		428439
MicroRNA names		4515
Pathway names		90 000 (42864 – GO; 47136 – PubMed)
Cell component names		5249
Cell names		396841-PubMed

Онтология типов взаимодействий



ANDCell: глобальная ассоциативная семантическая сеть фактов и знаний, извлеченных из различных источников информации



5 миллионов ассоциаций, извлеченных из баз данных и баз знаний

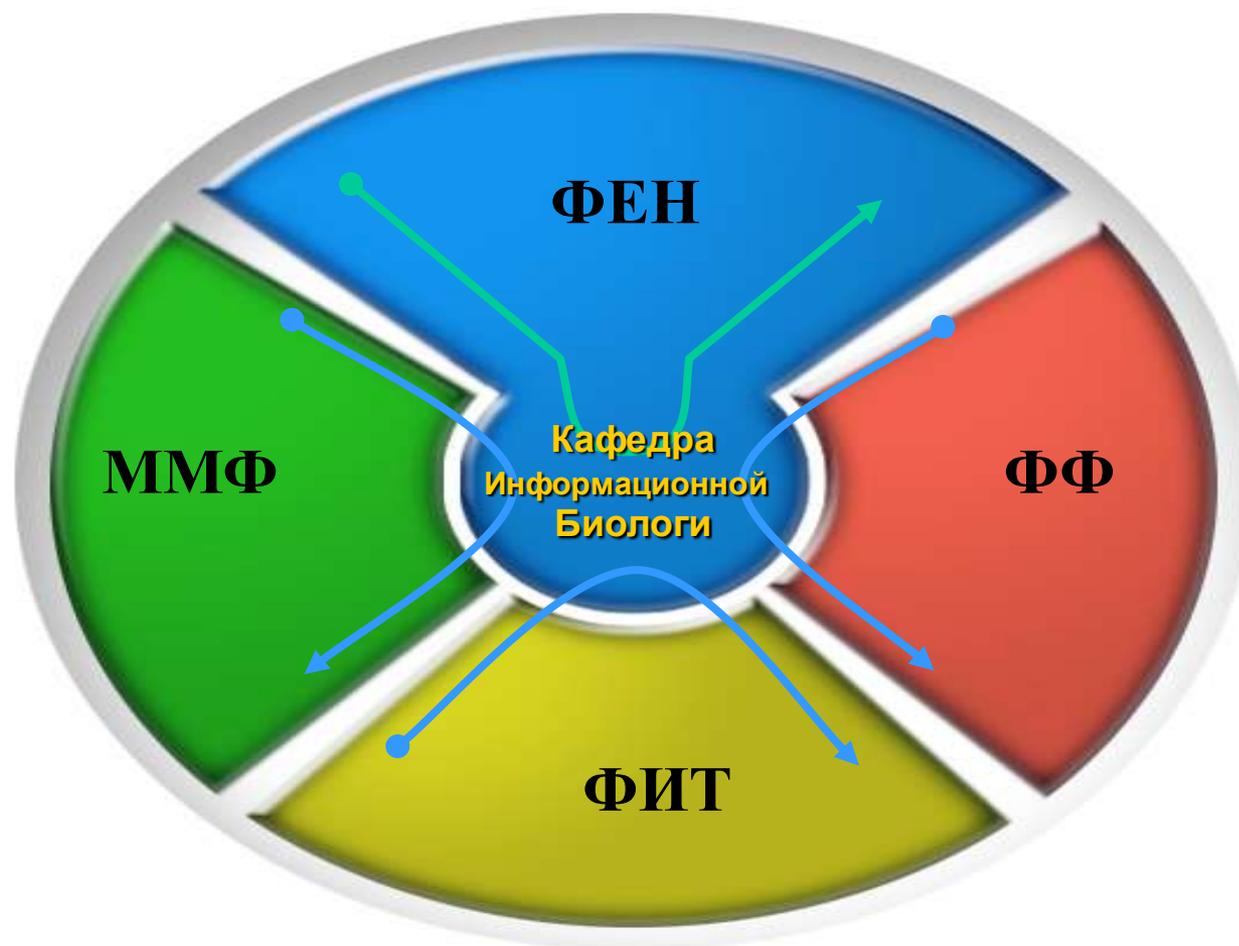
**ИЦИГ СО РАН – базовый
институт Кафедры
информационной биологии.
Лаборатория теоретической
генетики ИЦИГ СО РАН –
базовая лаборатория
кафедры информационной
биологии**

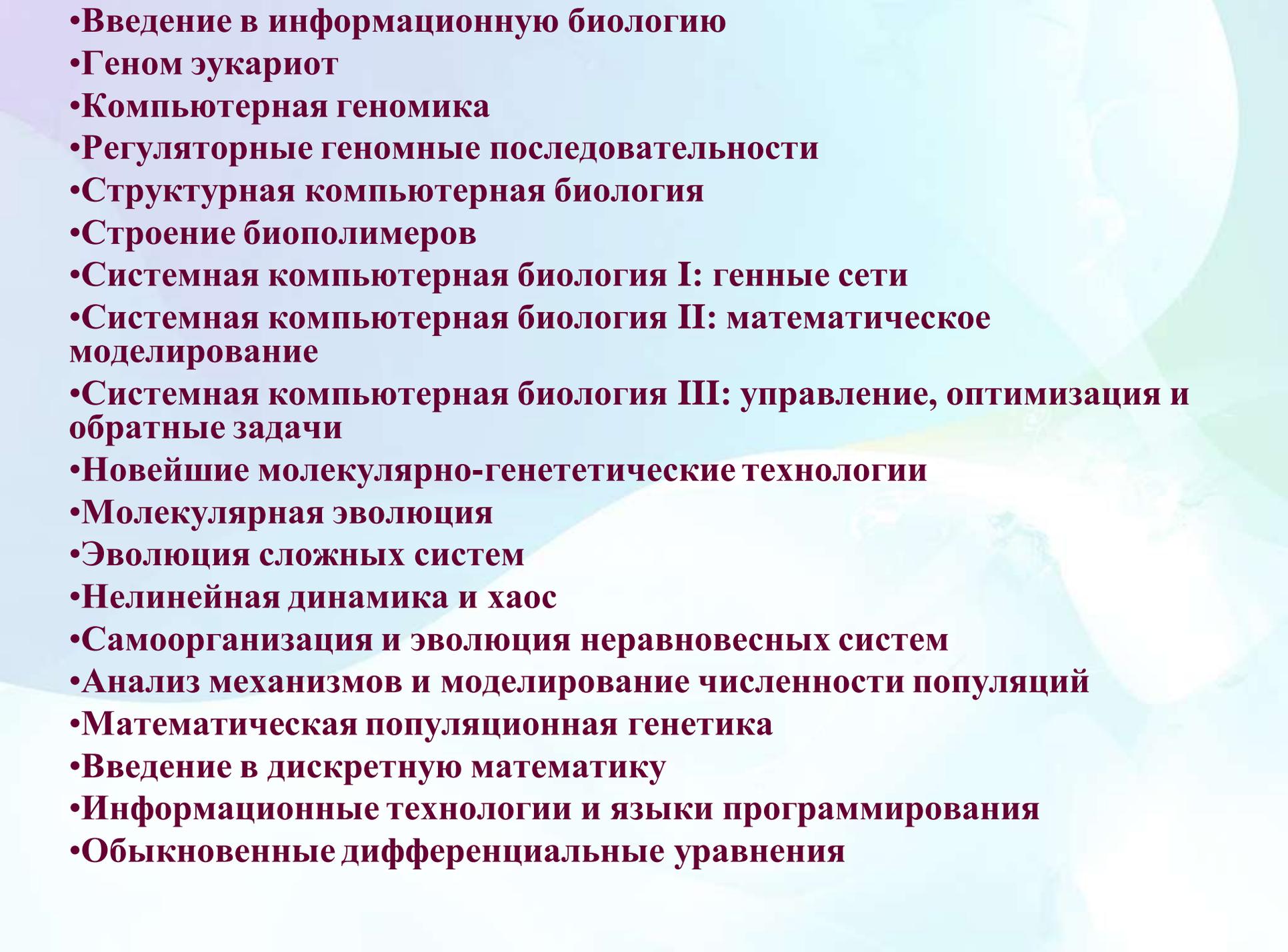
КАФЕДРА ИНФОРМАЦИОННОЙ БИОЛОГИИ (ФЕН НГУ)

Задача - подготовка специалистов–теоретиков и специалистов-экспериментаторов, владеющих методами биоинформатики, теоретического и компьютерного анализа и моделирования, необходимыми для решения широкого круга фундаментальных и прикладных проблем молекулярной биологии, молекулярной генетики, клеточной биологии, физиологии, биофизики, биомедицины, фармакологии, общей биологии, экологии, возникающих на стыке с математикой, информатикой и физикой.



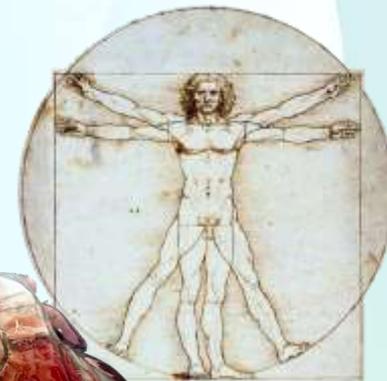
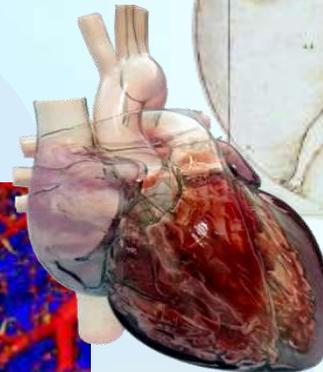
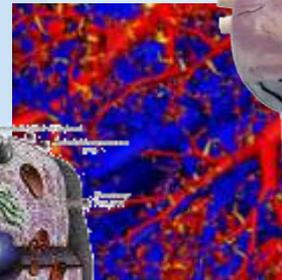
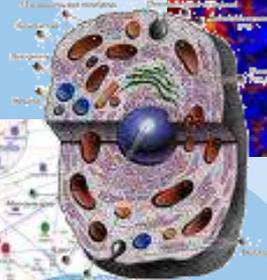
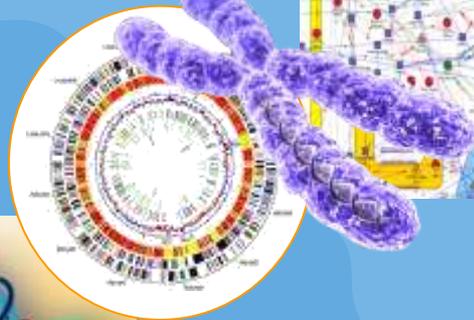
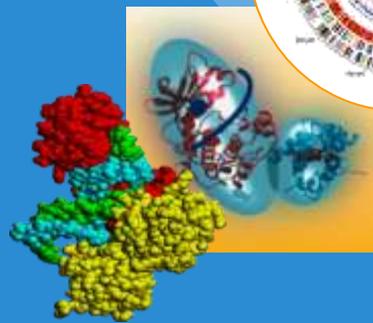
КАФЕДРА ИНФОРМАЦИОННОЙ БИОЛОГИИ: ВЗАИМОДЕЙСТВИЯ С ФАКУЛЬТЕТАМИ



- 
- **Введение в информационную биологию**
 - **Геном эукариот**
 - **Компьютерная геномика**
 - **Регуляторные геномные последовательности**
 - **Структурная компьютерная биология**
 - **Строение биополимеров**
 - **Системная компьютерная биология I: генные сети**
 - **Системная компьютерная биология II: математическое моделирование**
 - **Системная компьютерная биология III: управление, оптимизация и обратные задачи**
 - **Новейшие молекулярно-генетические технологии**
 - **Молекулярная эволюция**
 - **Эволюция сложных систем**
 - **Нелинейная динамика и хаос**
 - **Самоорганизация и эволюция неравновесных систем**
 - **Анализ механизмов и моделирование численности популяций**
 - **Математическая популяционная генетика**
 - **Введение в дискретную математику**
 - **Информационные технологии и языки программирования**
 - **Обыкновенные дифференциальные уравнения**

«Биоинформатическое цунами»

Экспериментально-компьютерные ресурсы (Log)



Биологическая сложность

Лекция 2

Исследование механизмов адаптации организмов к различным условиям среды на молекулярном уровне