

# ПРАКТИЧЕСКАЯ БИОИНФОРМАТИКА

---

Артур Залевский

ВШЭ, 2017



Факультет биотехнологии и биоинформатики  
МГУ им. М. В. Ломоносова

aozalevsky@fbb.msu.ru

<http://vsb.fbb.msu.ru>

+79268299066

## ВВЕДЕНИЕ

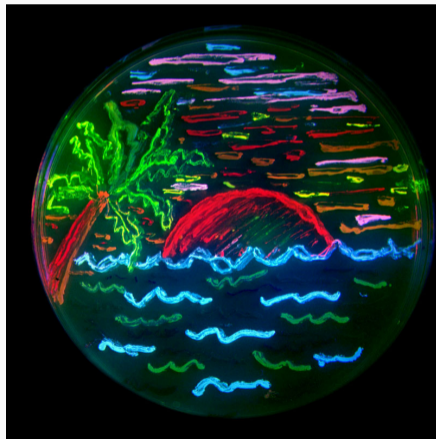
---

Это специалисты, владеющие последними достижениями фундаментальной биологической науки и способные целенаправленно изменять биологические объекты в соответствии с поставленными задачами.

Представляется...



В реальности...



- ▶ Что такое **жизнь**?





- ▶ Организация
- ▶ Метаболизм
- ▶ Рост
- ▶ Адаптация
- ▶ Реакция на раздражители
- ▶ Воспроизводство

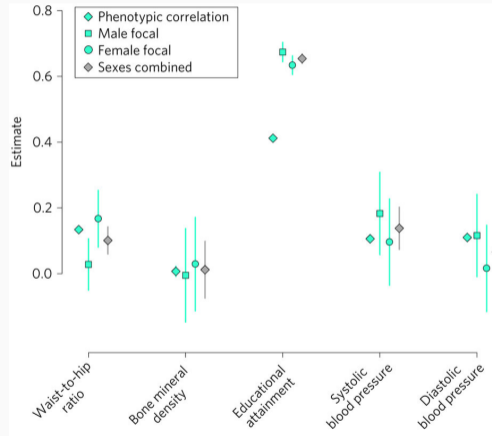


- ▶ Биогеоценоз (биосфера)
- ▶ Вид
- ▶ Популяция
- ▶ Организм (органно-тканый)
- ▶ Клеточный (субклеточный, молекулярный)

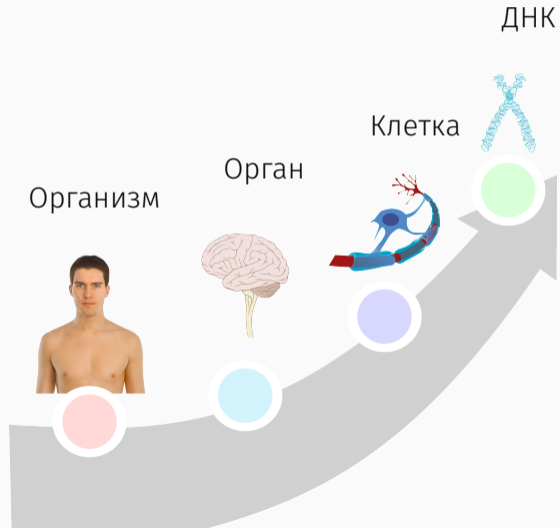
## ПРИМЕРЫ ЗАДАЧ БИОИНФОРМАТИКИ

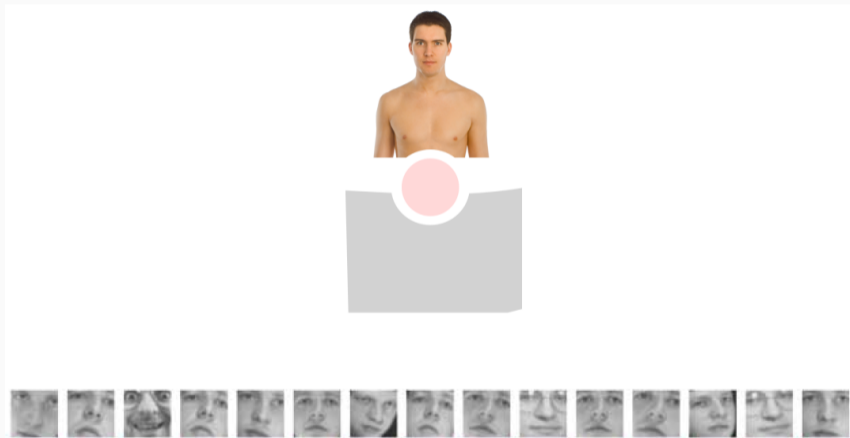
---



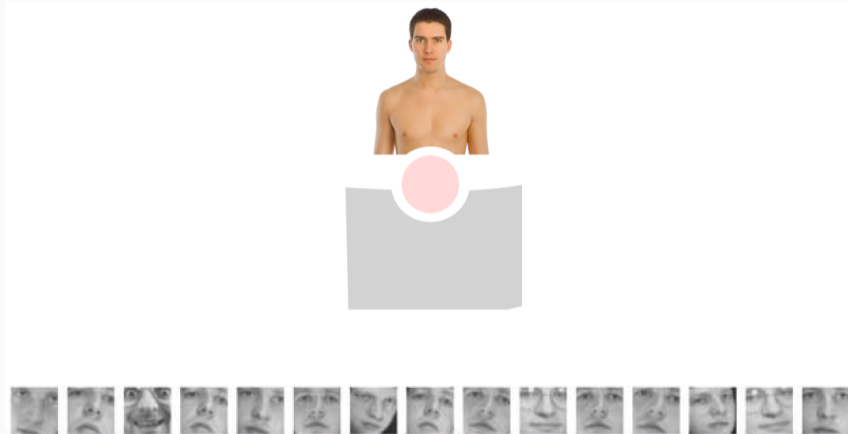


Genetic evidence of assortative mating in humans, Robinson, 2016







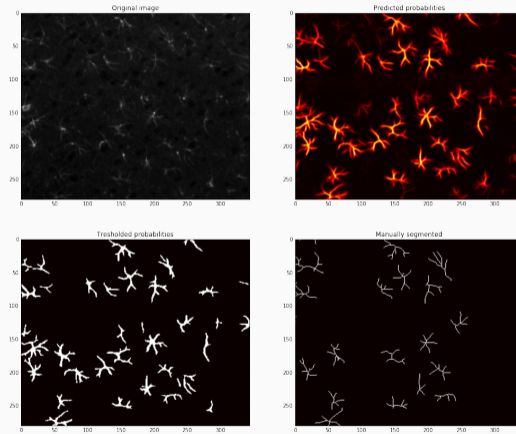


Effects of diazepam on facial emotion recognition, Coupland, 2009

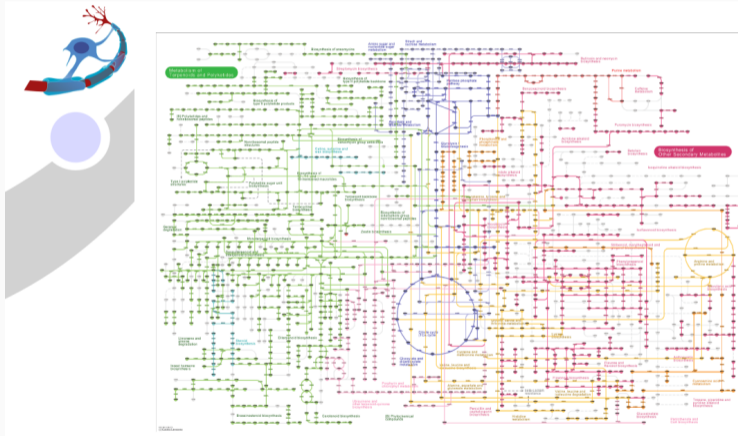


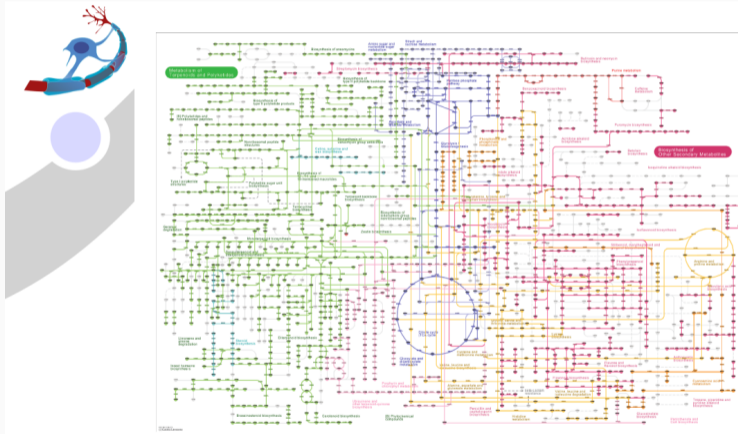


Iterative tomographic image reconstruction using Fourier-based forward and back-projectors, Matej, 2004

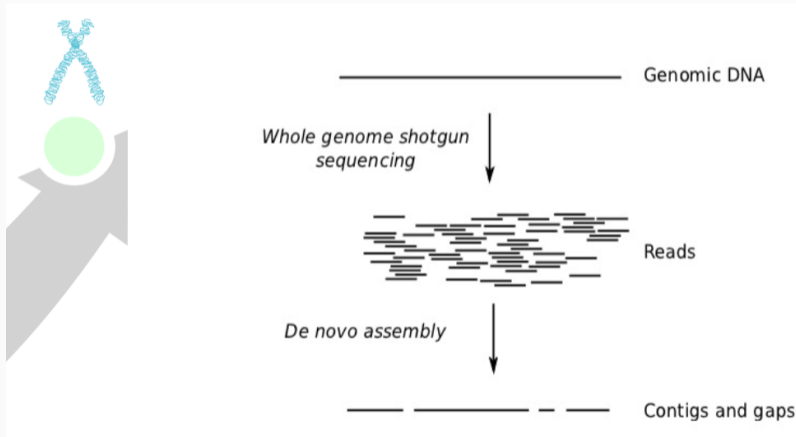


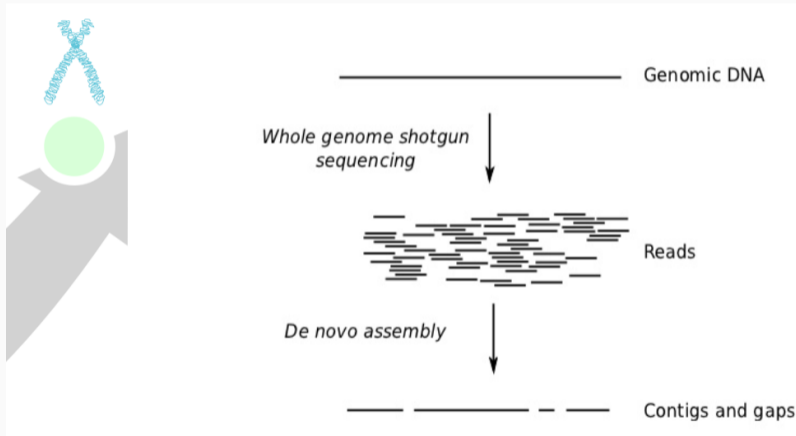
Артур Чахвадзе





Identification of regulatory properties of metabolic networks by graph theoretical modeling, Kohn, 1991

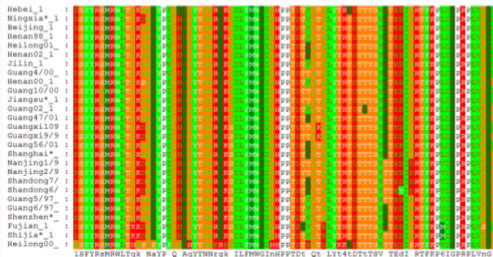




Initial sequencing and analysis of the human genome, Lander, 2001



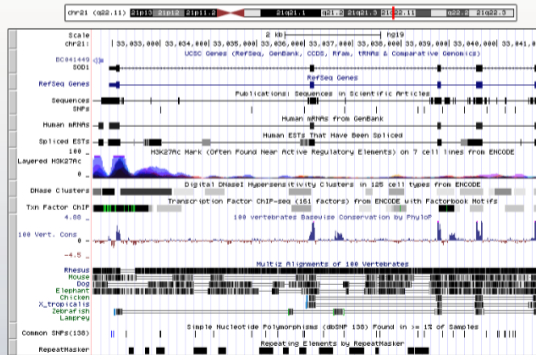
>gi|482677457|gb|KC750830.1| Homo sapiens cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial  
 CTACTCTGCTCGCATCTGTATAGTGGAGGCCGGAGCAGGAACAGGTTGAACAGTCTACCCCTCTTAG  
 CAGGGAAGTACTCCACCCTGGAGCCTCGTAGACCTAACCATCTTCTCCTTACACTAGCAGGTGCTC  
 CTCTATCTTAGGGCCATCAATTTATCACAAATTATCAATATAAAACCCCTGCCATAACCAATAC  
 CAACGCCCTCTTCTGCTGATCCGCTCAATCACAGCAGTCTACTTCTCCTATCTCTCCAGCTCTAG  
 CTGCTGGCCTACTATACTACTAACAGACGCAACTCAACACCACCTTCTTGACCCCGCCGGAGGAGG  
 GACCCCCATCTATACCAACACTATTCTGATTTT

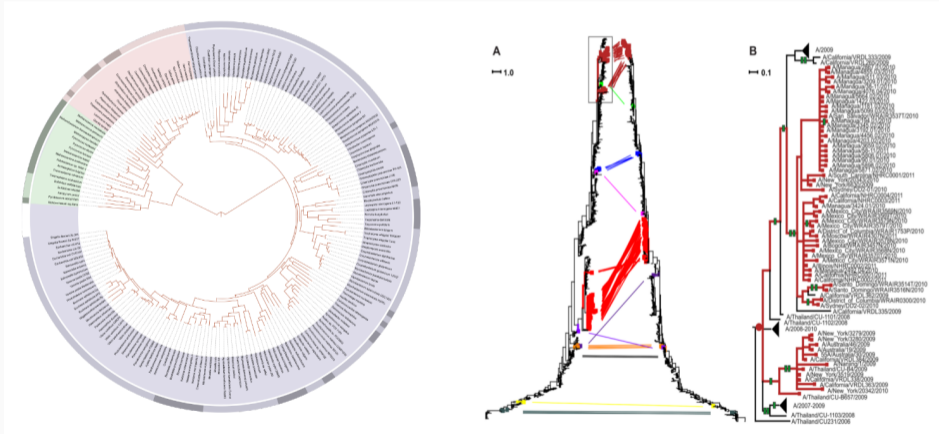


## UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Asse

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr21:33,031,597-33,041,570 9,974 bp. enter position, gene symbol or search terms





► Понимание и, даже, предсказание истории

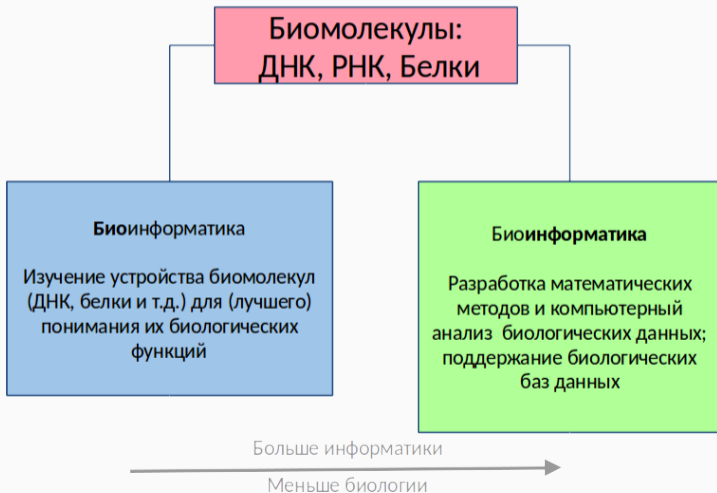
# БИОИНФОРМАТИКА

---

**Биоинформатика** - междисциплинарная область, разрабатывающая и улучшающая методы хранения, получения, организации и анализа биологической информации

Направления:

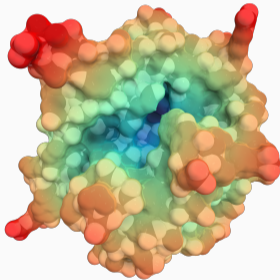
- ▶ Анализ последовательностей
- ▶ Вычислительная эволюционная биология
- ▶ Анализ изображений
- ▶ Анализ экспрессии генов
- ▶ Вычислительная системная биология
- ▶ ...
- ▶ Структурная биоинформатика



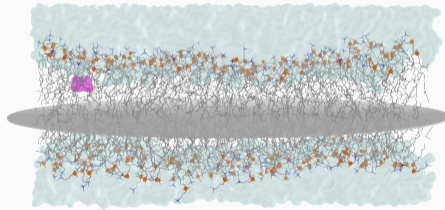
# СТРУКТУРНАЯ БИОИНФОРМАТИКА

---

## ► Белки

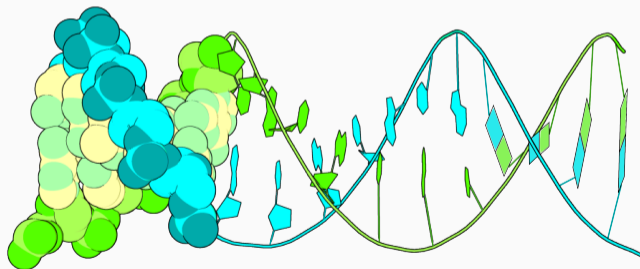


- ▶ Белки
- ▶ Липиды

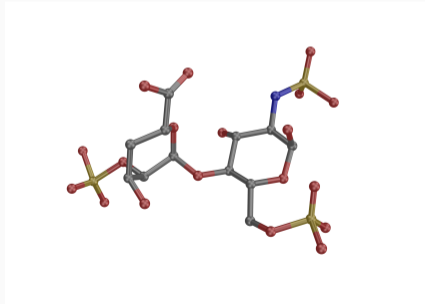




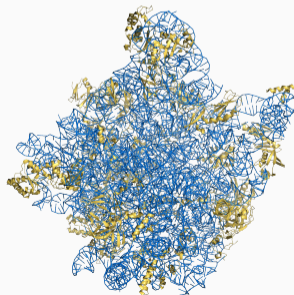
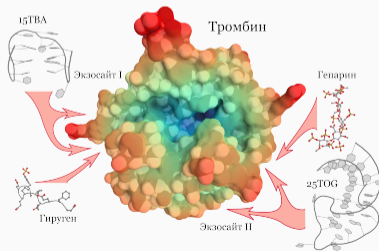
- ▶ Белки
- ▶ Липиды
- ▶ Нуклеиновые кислоты



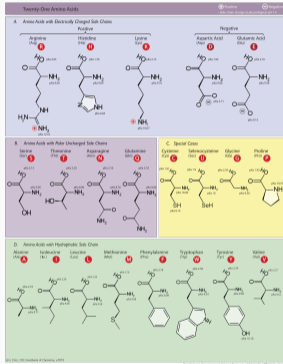
- ▶ Белки
- ▶ Липиды
- ▶ Нуклеиновые кислоты
- ▶ Низкомолекулярные соединения



- ▶ Белки
- ▶ Липиды
- ▶ Нуклеиновые кислоты
- ▶ Низкомолекулярные соединения
- ▶ Комплексы



- ▶ DNA: A, T, G, C
- ▶ RNA: A, U, G, C
- ▶ Белки: G, A, V, L, I, P, S, T, C, M, N, Q, F, Y, W, D, E, K, R, H

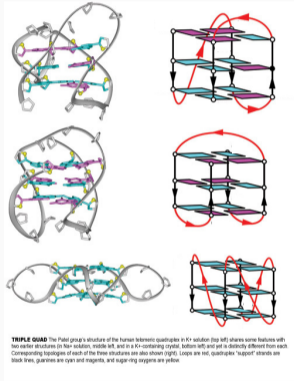


- ▶ Некоторые а.к. можно заменить без изменения структуры и свойств белка, поэтому эффективный алфавит еще меньше

en.wikipedia.org

- ▶ Одна последовательность может сворачиваться несколькими способами
- ▶ Разные последовательности могут иметь сходную трехмерную структуру

Функция определяется структурой



Patel

- ▶ Одна последовательность — 3 конформации

- ▶ Любопытство



- ▶ Любопытство
- ▶ Рациональный дизайн

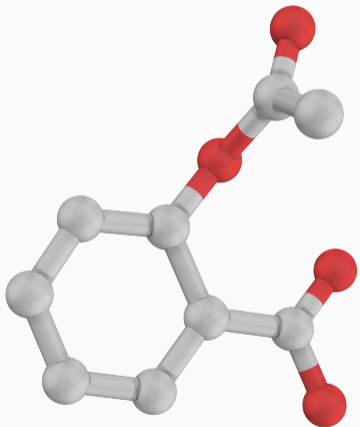
Создание новых объектов с **заданными характеристиками**, основываясь на возможности предсказания влияния структуры на свойства при помощи различных физических моделей

Применение:

- ▶ Дизайн ферментов и каскадов для биотехнологии
- ▶ Дизайн лекарств

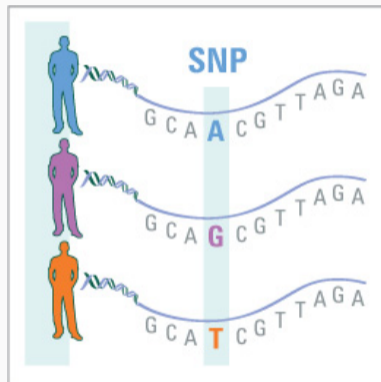
# ДИЗАЙН ЛЕКАРСТВ

---

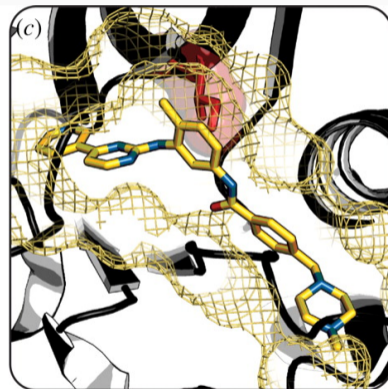
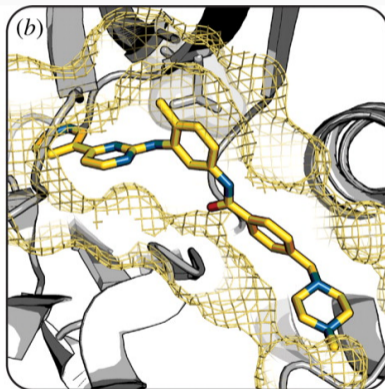


- ▶ Тошнота, снижение аппетита, гастралгия, диарея
- ▶ Аллергические реакции (кожная сыпь, ангионевротический отек, бронхоспазм)
- ▶ Нарушение функции печени и/или почек
- ▶ Тромбоцитопения, анемия, лейкопения, синдром Рейе (энцефалопатия и острая жировая дистрофия печени с быстрым развитием печеночной недостаточности)
- ▶ Формирование на основе гаптенного механизма аспириновой астмы и аспириновой триады (сочетание бронхиальной астмы, рецидивирующего полипоза носа и околоносовых пазух и непереносимости АСК и ЛС пиразолонового ряда)

# ОДНОНУКЛЕОТИДНЫЙ ПОЛИМОРФИЗМ

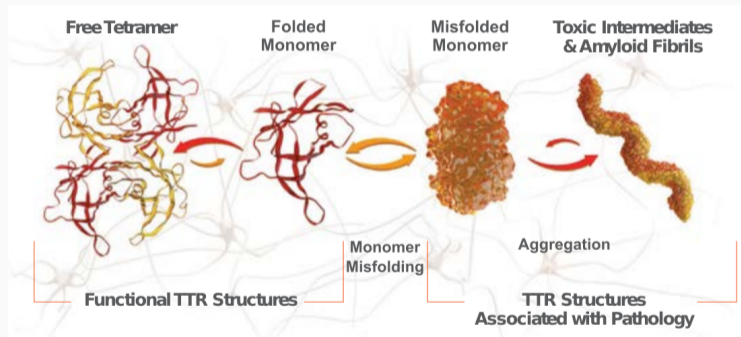


Broad institute



Lahti

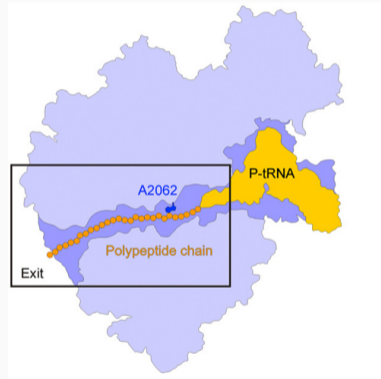
## Familial amyloid polyneuropathy – disease of protein misfolding



Pfizer

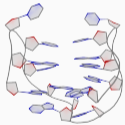
- ▶ Typical number of patients (**worldwide**) ~ 10000



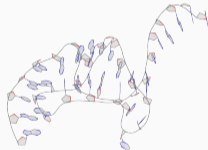


Broad institute

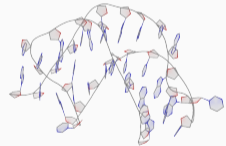
Аптамеры - синтетические олигонуклеотиды, специфично связывающиеся с мишенью



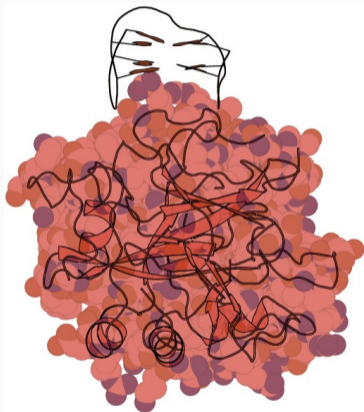
Thrombin binding aptamer



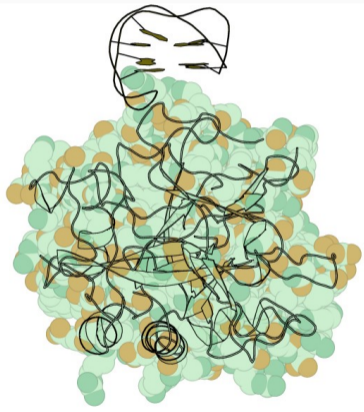
PreQ1 riboswitch aptamer



Vitamin B12 aptamer



ЯМР

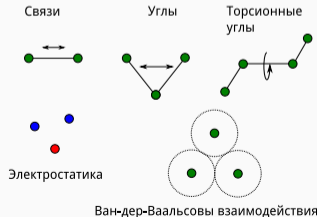


РСА

$$U = \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_0)^2 + \sum_{\text{angles}} \frac{k_i}{2} (\phi_i - \phi_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) +$$
$$+ \sum_{i=1}^N \sum_{j=i+1}^N \left( 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)$$

$$U = \sum_{bonds} \frac{k_i}{2} (l_i - l_0)^2 + \sum_{angles} \frac{k_i}{2} (\phi_i - \phi_0)^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) +$$

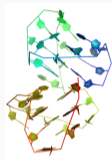
$$+ \sum_{i=1}^N \sum_{j=i+1}^N \left( 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)$$



# СУПЕРКОМПЬЮТЕР ЛОМОНОСОВ (1, 2, х)



- ▶ CPU: Intel E5-2697v3 (14 cores)
- ▶ Ядра: 42 688
- ▶ GPU: Nvidia Tesla K40M
- ▶ Сеть: Infiniband FDR
- ▶ RAM: 64Gb на узел
- ▶ HDD: 400Tb



PCT

Golovin, 2015

- ▶ G-квадруплексный аптамер к тромбину с антикоагулянтным эффектом
- ▶ Прошел доклинические испытания, готовится к 1й стадии клинических испытаний

Q&A

---





ΟΠΡΟΣ

---

<https://goo.gl/BtZ1Ex>

host: xwing.cmm.msu.ru

port: 443

user: hse

password: hse

directory: ~/challenge

question: how many CYS residues in .pdb files?

hint: proper strings start with "ATOM"