



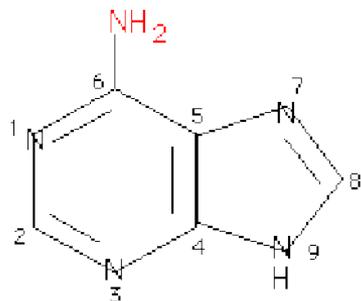
Вторичная структура РНК: компьютерные методы анализа (1-2)

К.ф.-м.н. И.И. Тимов

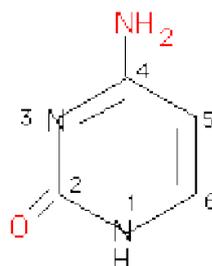
Кафедра информационной биологии ФЕННГУ



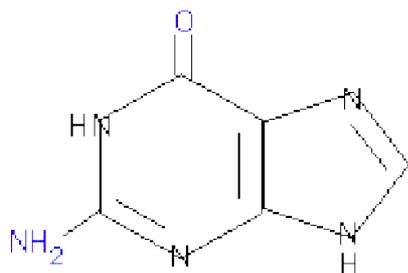
Химическое строение РНК



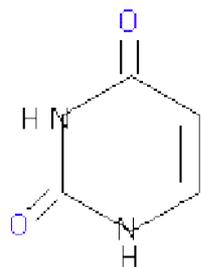
Adenine



Cytosine



Guanine



Uracil

РНК (рибонуклеиновая кислота) – биологический гетерополимер, состоящий из рибонуклеотидов 4х основных типов (А, U, G и С).

Нуклеотид:
фосфатная группа
рибоза
остаток

РНК - односторонняя молекула с выделенным направлением вдоль цепи (5' → 3').



Типы взаимодействий между остатками



- Уотсон-криковские пары: AU, GC
- GU "wobble" пары
- «неканонические» пары, напр. AA, GG
- Взаимодействия троек и другие пространственные взаимодействия

Взаимодействия пар влечет образование непрерывных двунитиевых участков (спиралей)

Стабилизация пар в спиральных:

- водородные связи
- стэкинг



Что такое и зачем нужна вторичная структура РНК

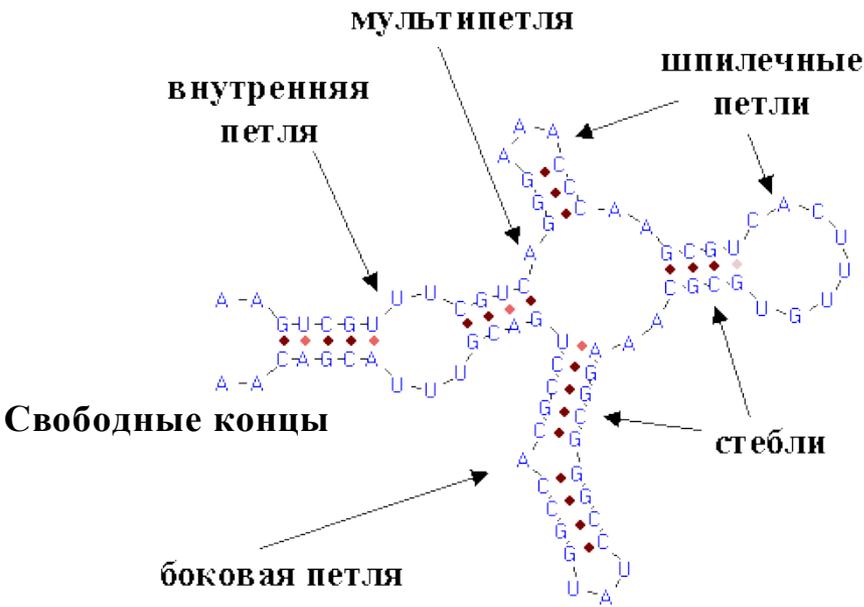


ВС РНК – плоская модель пространственной структуры РНК, состоящее из чередования спиральных и однонитиевых участков

ВС - разумная аппроксимация 3D структуры:

- поддерживается эволюцией
- доминирующий вклад в энергию 3D структуры
- начальный этап образования 3D

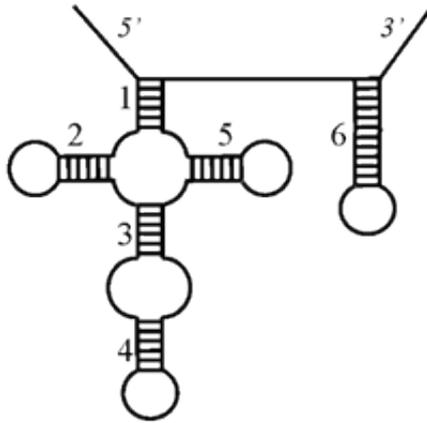
- Способы определения



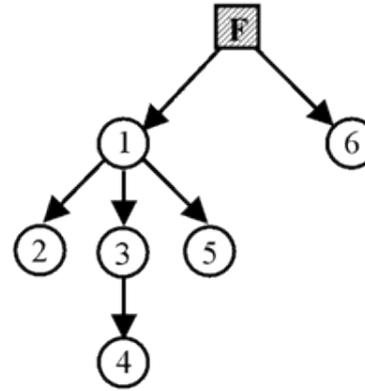
Элементы ВС



Графические представления ВС РНК



1. Традиционное



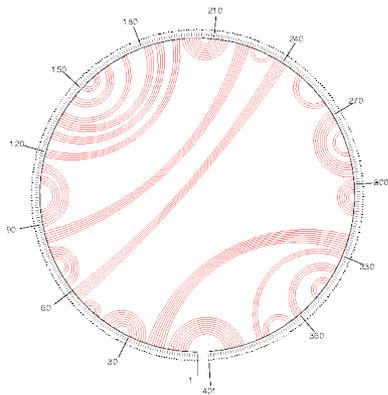
2. Дерево



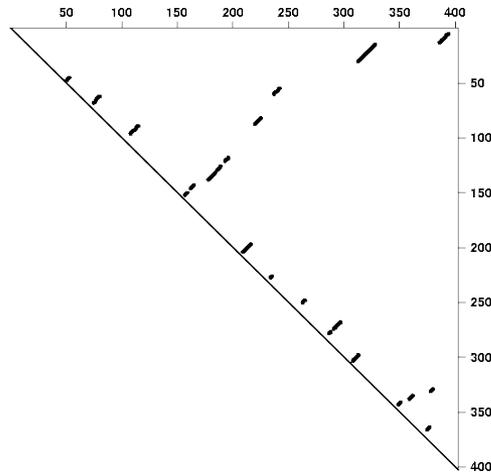
Графические представления ВС РНК



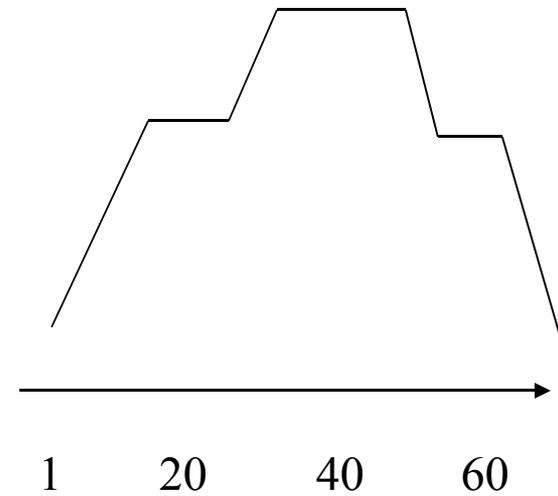
© 1997 by S. Chakravarti, Sage
© 2001, Washington University



ENERGY = -85.7 Bacillus subtilis RNase P RNA



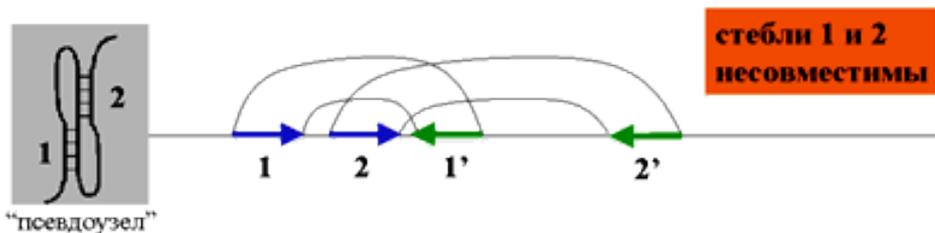
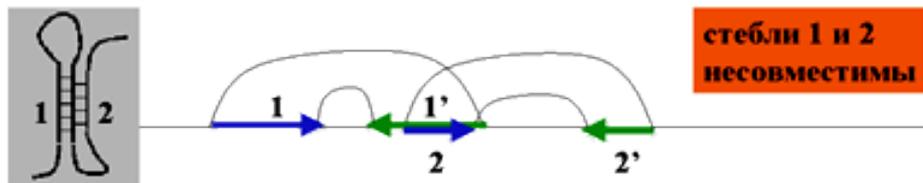
Шпилька с боковой петлей



- 3. Круговое
- 4. Матрица контактов
- 5. Горы и долины



Сtereoхимические правила ВС



Запрещение на

Одновременное вовлечение
одного нуклеотида в 2х и более
спирали

Псевдоузлы (входят в третичные
взаимодействия)



Термодинамика вторичной структуры РНК



??

$$\Delta G = \Delta U - T\Delta S$$



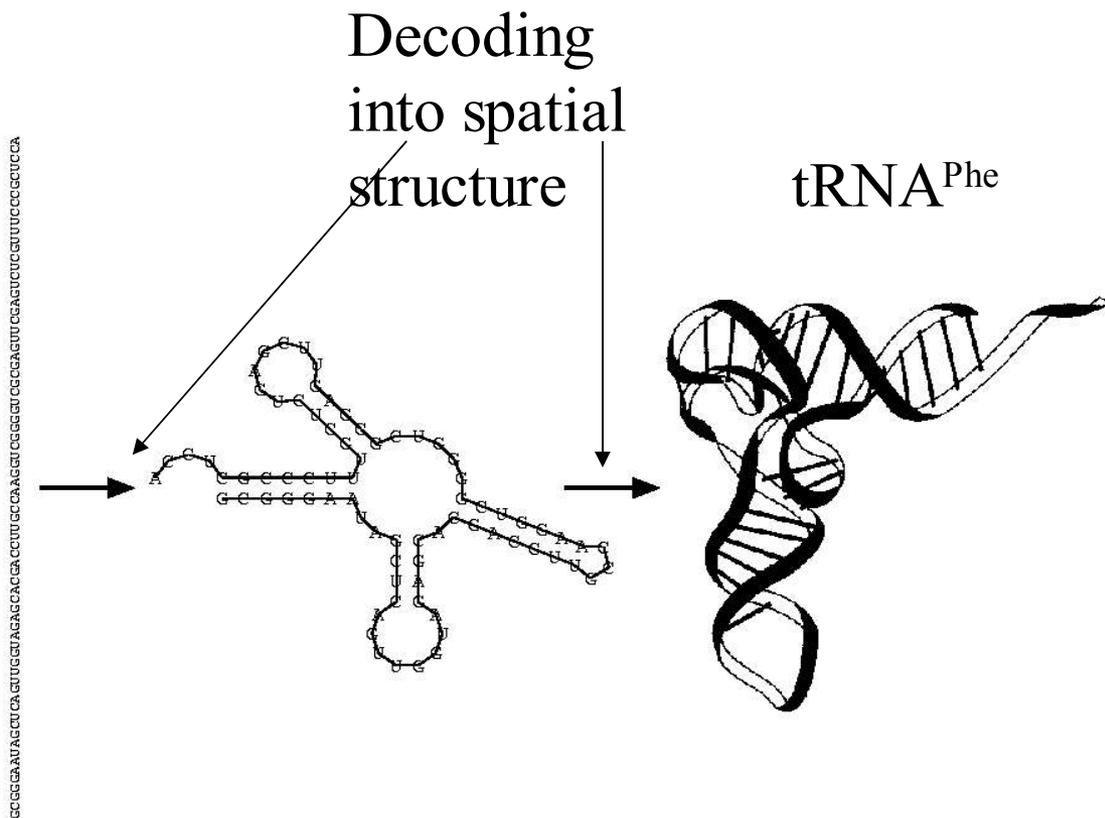
Энтропийные факторы

- Резкая угловая зависимость энергии водородной связи
- Ограничение вращения мономеров
- Сближенность взаимодействующих участков

$$\Delta G = \Delta U - T\Delta S$$



Прямая и обратная задачи фолдинга РНК





Обратная задача: Application Areas



Bionanotechnology

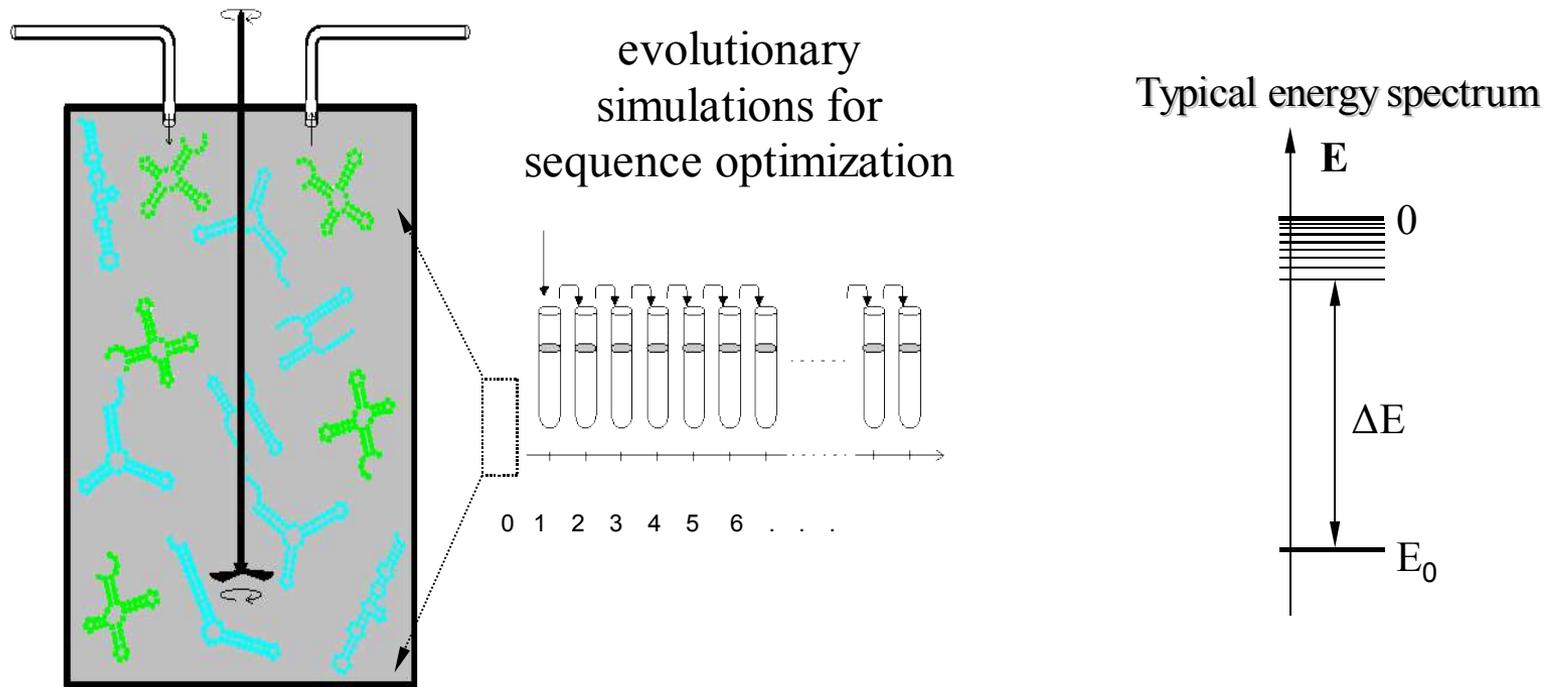
Pharmaceutics

RNA structure research

Optimization algorithms

Technology Features

- The energy gap between a target and the misfolded structures will be explicitly optimized by evolutionary simulations thus maximizing the thermodynamical probability of a target structure

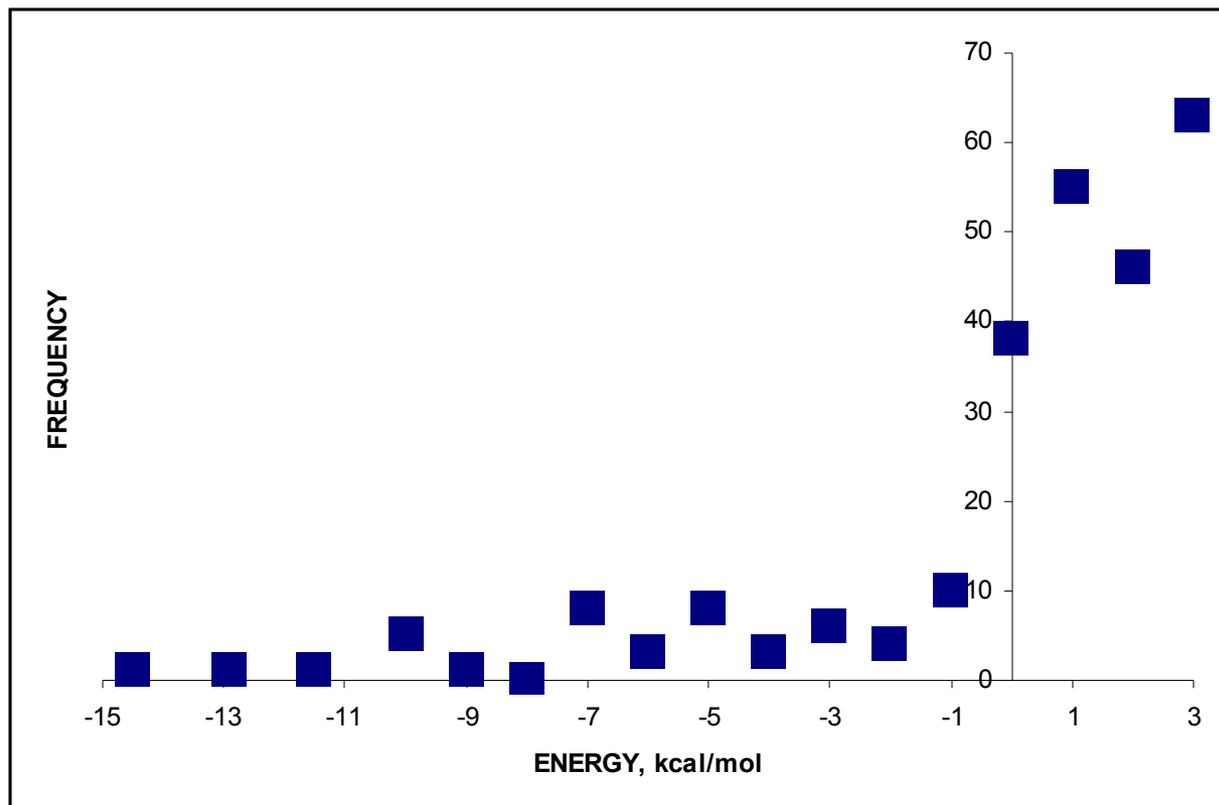




Энергетический спектр ВС РНК

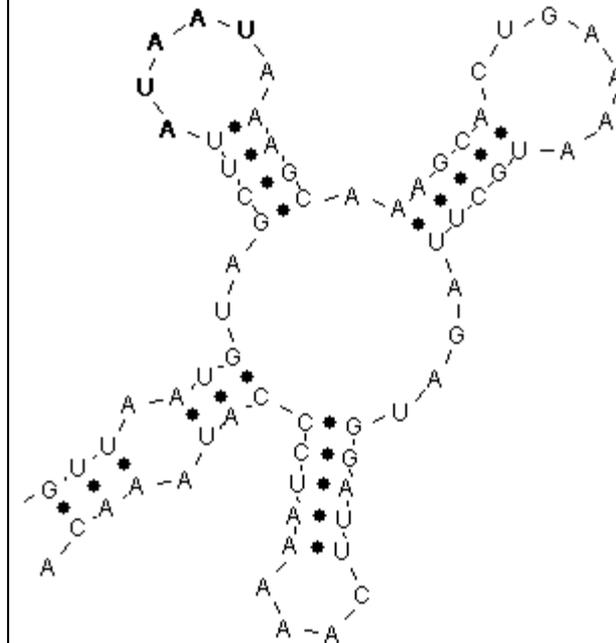
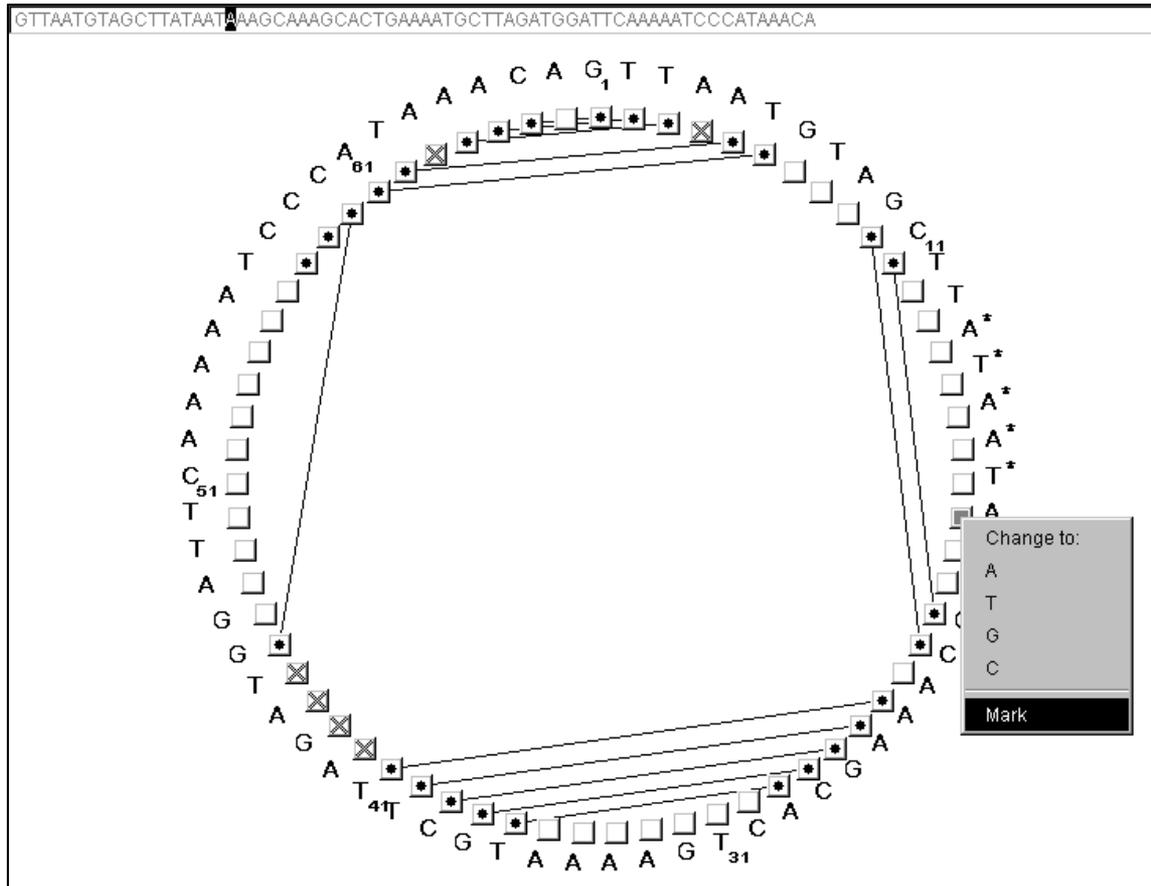


LOW-ENERGY TAIL OF THE DENSITY OF STATES (leader region of *thrS* mRNA *E.Coli*)

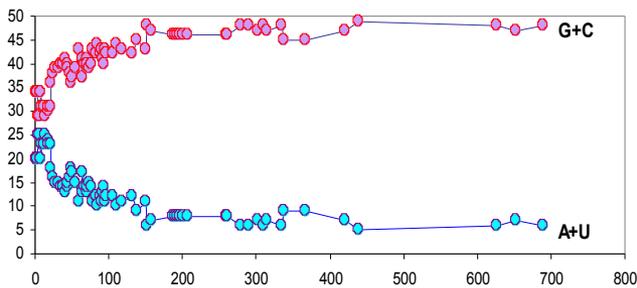
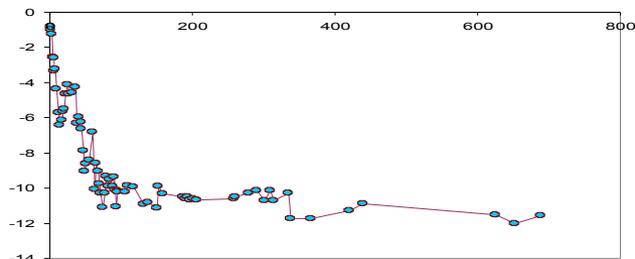
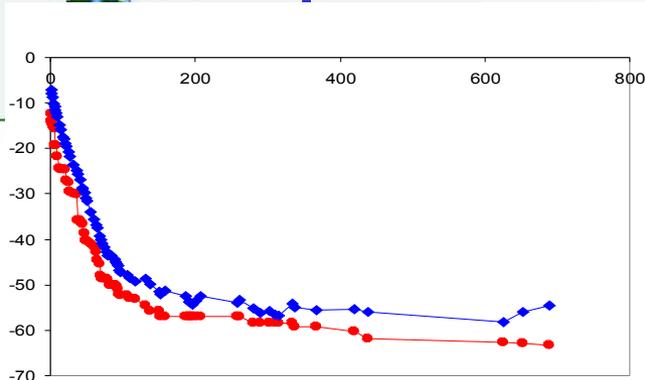




- The programs are implementing for PC with Windows and Linux OS
- Unique easy-to-use Java interface which will help accounting for experimental information of base-pairing



Моделирование эволюции к стабильной вторичной структуре



Конечный результат

Z-score = -13.0



Эволюция энергии вторичной структуры
(лучшая особь и среднее значение по популяции)

Динамика относительного (от случайной последовательности того же состава) отклонения энергии самой стабильной вторичной структуры

Абсолютная величина относительного отклонения характеризует упорядоченность следования нуклеотидов в последовательности

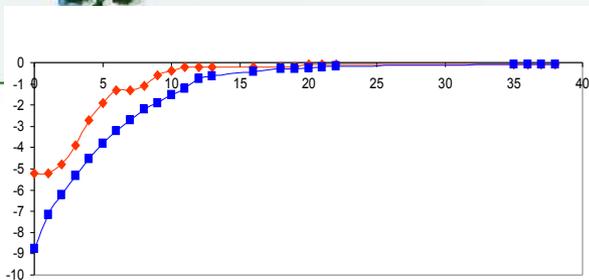
Временная эволюция нуклеотидного состава

Два основных механизма оптимизации последовательности:

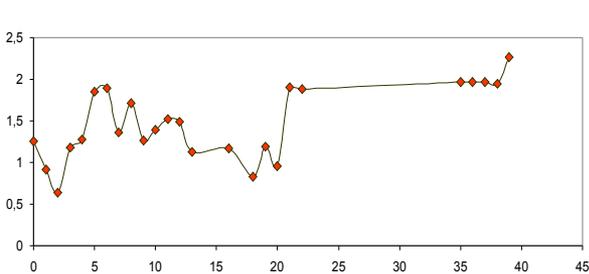
- а) увеличение содержания GC
- б) упорядочения нуклеотидов посредством коадаптивных замен

действуют **одновременно**

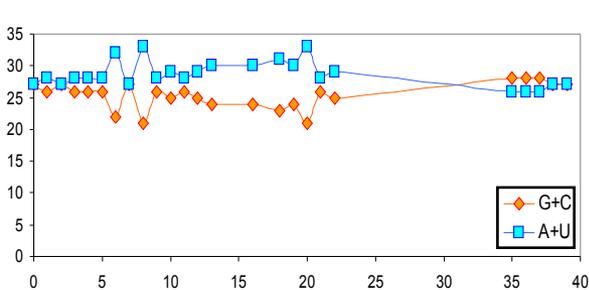
Моделирование эволюции к нестабильной вторичной структуре



Эволюция энергии вторичной структуры
(лучшая особь и среднее значение по популяции)



Динамика относительного (от случайной последовательности того же состава) отклонения энергии самой нестабильной вторичной структуры. Абсолютная величина относительного отклонения характеризует упорядоченность следования нуклеотидов в последовательности

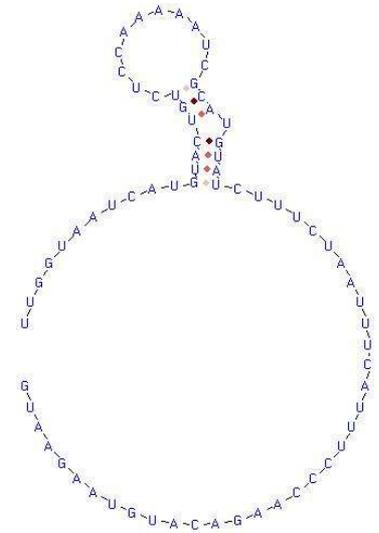


Временная эволюция нуклеотидного состава

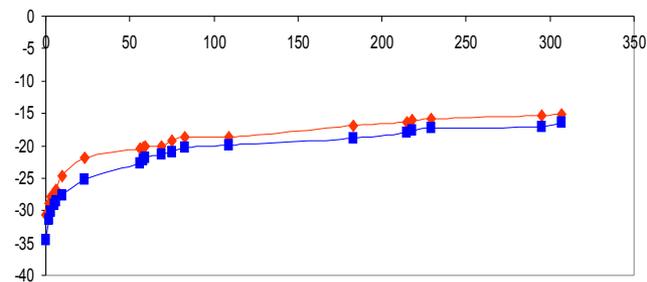
$$Z = \frac{Observed(x) - Expected(x)}{\sqrt{disp(x)}}$$

Легко получить последовательность с нестабильной вторичной структурой, поскольку таких последовательностей подавляющее большинство

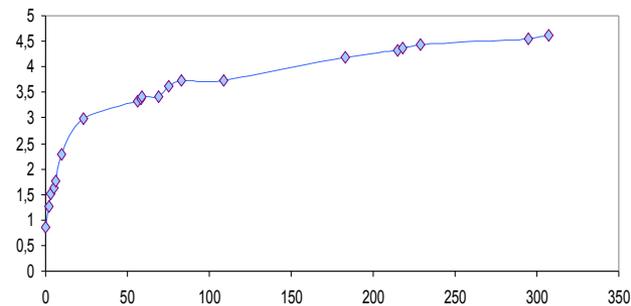
Результат
Z-score = 2.06



Моделирование эволюции к нестабильной вторичной структуре при фиксированном содержании нуклеотидов (50%G, 50%C, A,U - 0%)



Эволюция энергии вторичной структуры (лучшая особь и среднее значение по популяции)



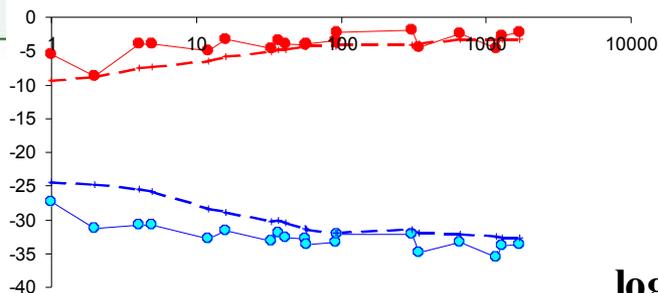
Динамика относительного (от случайной последовательности того же состава) отклонения энергии самой нестабильной вторичной структуры. Абсолютная величина относительного отклонения характеризует упорядоченность следования нуклеотидов в последовательности

Результат эволюции:

CCGCCCGCCGGGGGGGGGGCCGCCGCCGCCGCCGCCGGGGCCGCCGCCG

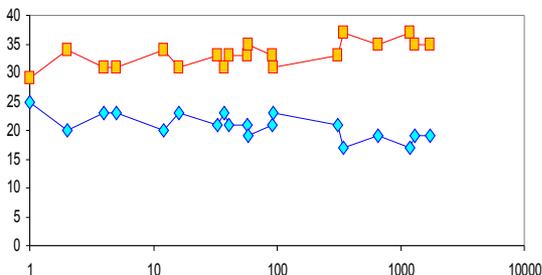
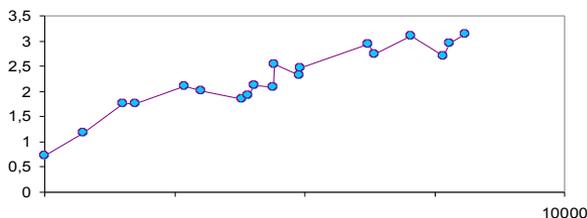
содержит триплет CCG и поли-G

Моделирование эволюции к РНК-кубу



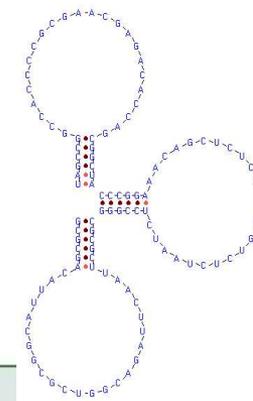
Эволюция энергии **целевой структуры** и **альтернативной вторичной структуры** Сплошная линия - лучшая особь, пунктир - среднее

log t Динамика относительного (от случайной последовательности того же состава) отклонения энергии самой нестабильной вторичной структуры. Абсолютная величина относительного отклонения характеризует упорядоченность следования нуклеотидов в последовательности



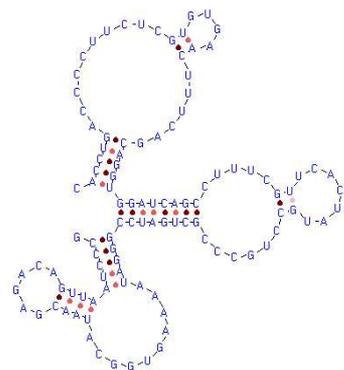
log t Временная эволюция нуклеотидного состава

Структура итоговой последовательности



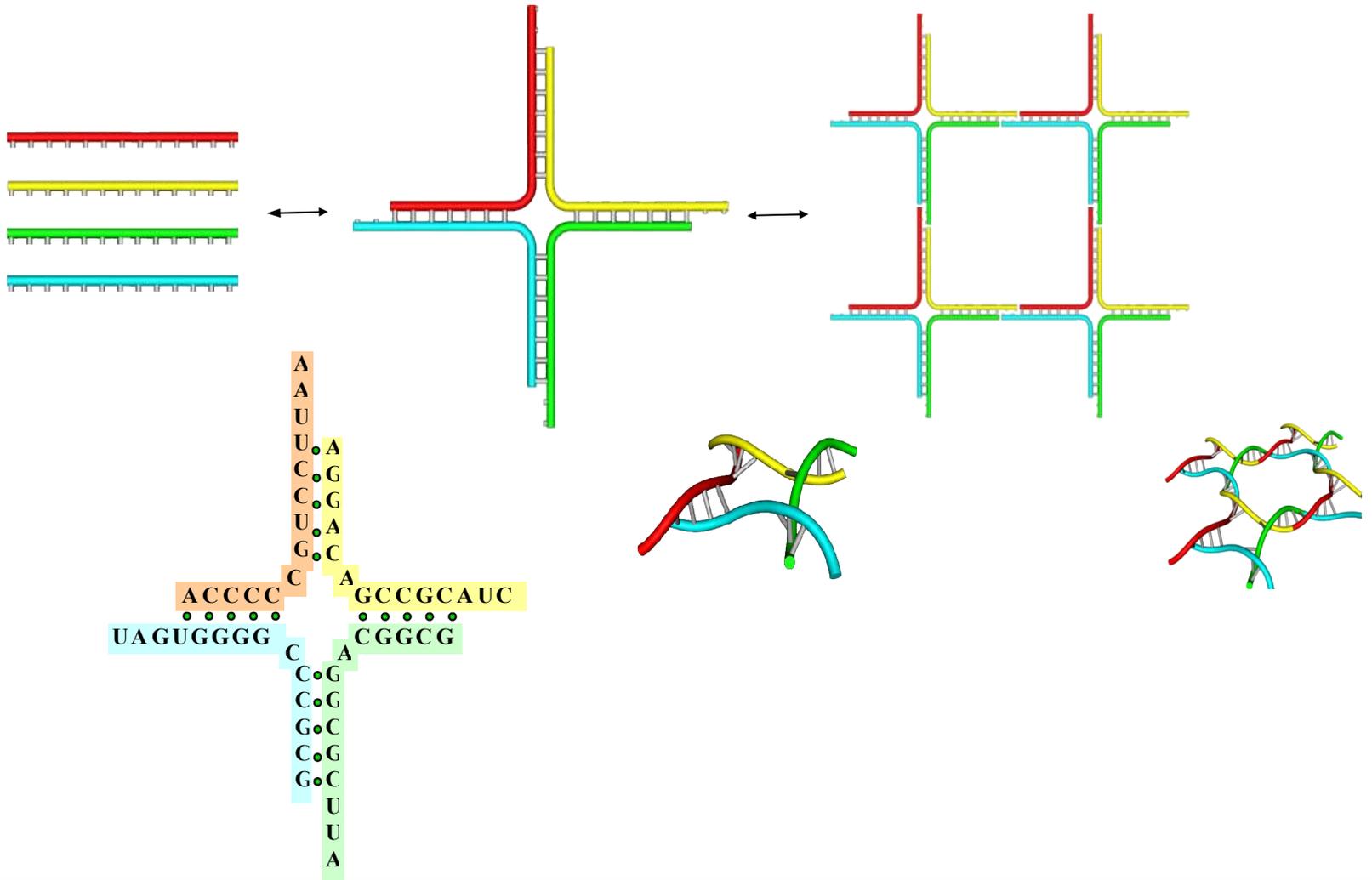
Длина последовательности 108 нукл (длина ребра нанокуба 6 нукл) Z-score = 2.76

Структура исходной последовательности

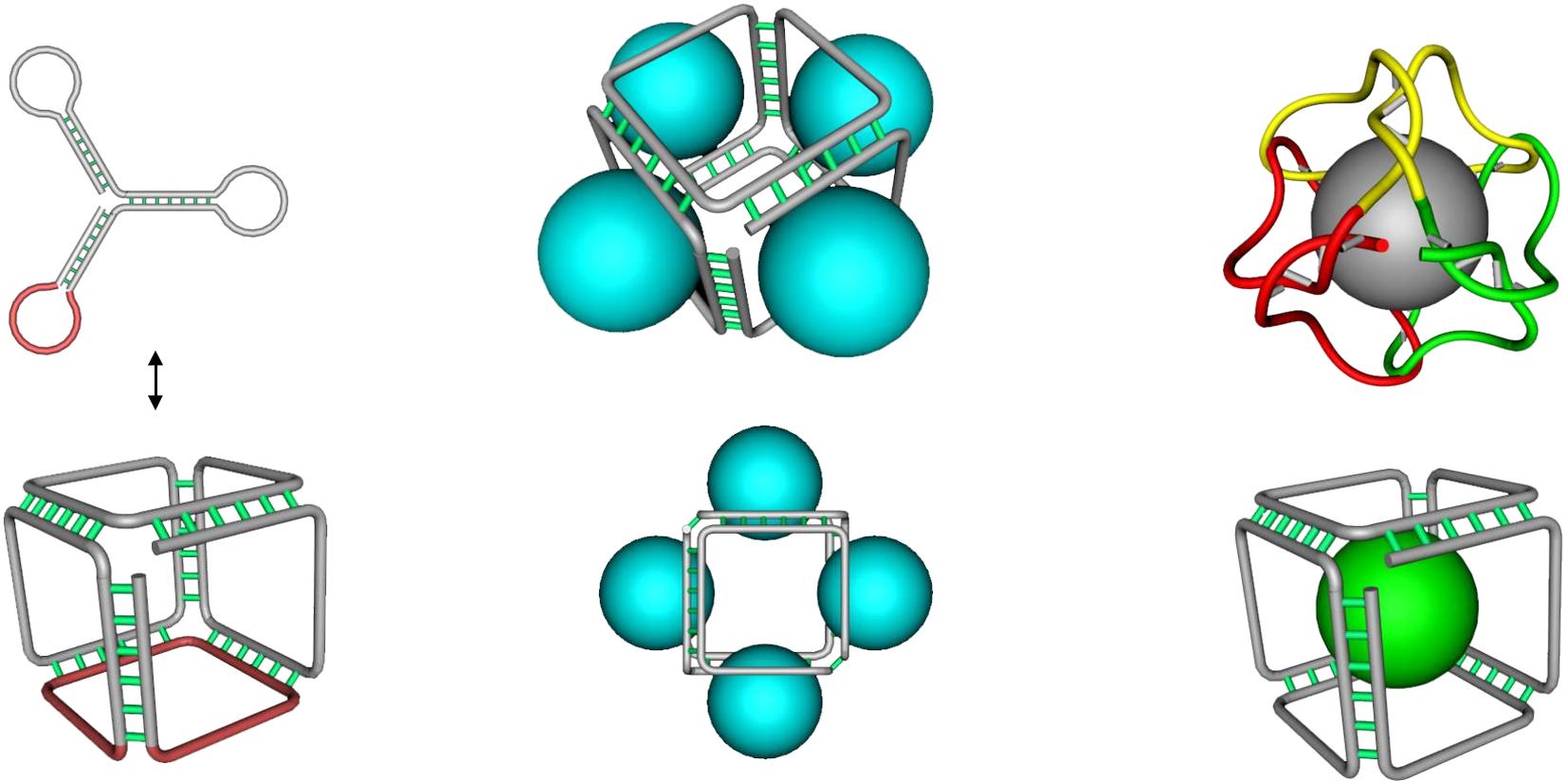




Calculation of the oligonucleotide sequences forming a square lattice



Possible applications of a cube spontaneously formed from a single-stranded RNA



Such a cube may be used as a matrix for protein assembly, as a ligand cage, as a "multiribozyme", etc.

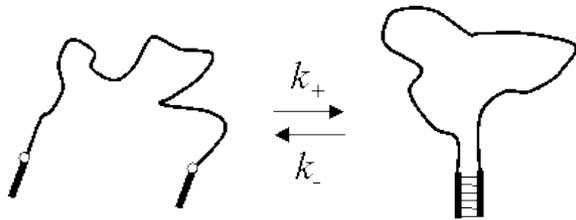


Кинетика вторичной структуры РНК: начальная стадия



Образование комплементарных взаимодействий – кинетически контролируемый процесс

(Wetmur, Davidson, 1968)



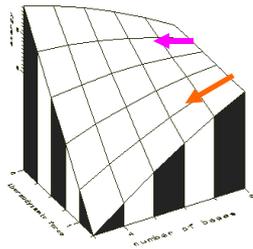
Скорость реакции описывается формулой Стокмайера (1947)



Кинетика вторичной структуры РНК: рост спирали



free energy surface
of a helix



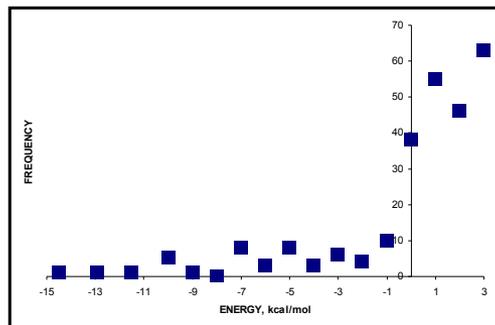
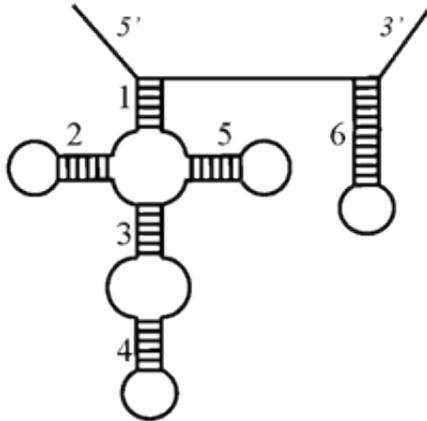
Helix length

Спираль растет после образования критического ядра (Porschke, Eigen, 1969)

Модель кооперативного застегивания спирали (Титов, 1998)



Кинетика вторичной структуры РНК: зачем?



Пути фолдинга РНК могут ветвиться и заканчиваться в метастабильных состояниях

Скорость образования спирали определяется энергией петли

Скорость распада спирали определяется энергией спирали (Миронов, Дьяконова и Кистер, 1983)

Альтернативные состояния РНК могут нести биологическую функцию



Компьютерные методы *ab initio* расчета ВС РНК через Интернет



Mfold (Zuker)

Vienna Package (Hofacker et al.)

Garna (Titov et al.)

минимизация?



Лингвистический анализ (поиск тРНК-генов, сигналов и проч)



Пример

Сайт связывания оболочечного белка R17 (R=A/G, Y=U/C)

N Y

A A

N-N'

N-N'

R

N-N'

N-N'

N-N'

N-N'

N-N'

$$I = \sum_x p_x \log_2 \frac{p_x}{f_x}$$

Информация в наборе x

P – наблюдаемая частота

F – ожидаемая частота



Лингвистический анализ (пример)



F=0.25

X=A и только A

P=1

I=2 bit (Binary digit)

$$I = \sum_x p_x \log_2 \frac{p_x}{f_x}$$

I ?

X=Y

X=N

Информация в наборе x

P – наблюдаемая частота

F – ожидаемая частота

I консенсуса, встречаемость



Лингвистический анализ (пример)



Консенсус длины L

$$I=2^L$$

По случайным причинам

в среднем однажды в
последовательности длиной
 2^l

В 1Mb – 20 bit паттерн

Геном человека 4Gb → 32bit

Дрожжи 24 bit

Вирусы 16 bit

$$I = \sum_x p_x \log_2 \frac{p_x}{f_x}$$

Информация в наборе x

P – наблюдаемая частота

F – ожидаемая частота



$$M = \sum_{xy} p_{xy} \log_2 \frac{p_{xy}}{f_x f_y}$$

Взаимная информация в наборе x :

«если знаю A , что могу сказать о B ?»

Курица или яйцо?



Спецкурс «информационная биология»

Лекция 5

Лекция 7

Структурная биология в:

Кантор и Шиммель «Биофизическая химия-3»

Зенгер «Принципы структурной организации нуклеиновых
КИСЛОТ»

Рассчитать энергию и время