Биоинформатика

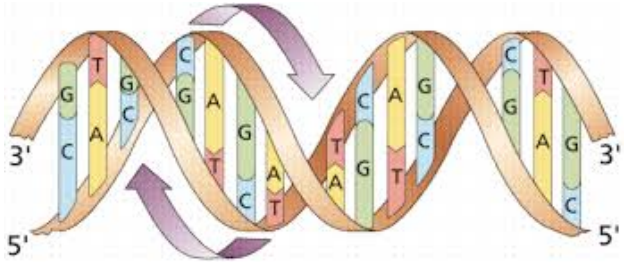
Андрей Александрович Миронов

Факультет боиинженерии и биоинформатики

РНК - одноцепочечная полинуклеотидная цепь

ДНК vs РНК

- в РНК используется рибоза вместо дезокси-рибозы
- в РНК обычно U вместо T
- ДНК – обычно двунитевая полинуклеотидная цепь
- РНК – обычно однонитевая полинуклеотидная цепь
- нуклеотиды в РНК образуют комплементарные пары.
- РНК может образовывать сложную структуру



Информационные функции РНК

Информационные функции

- перенос информации от ДНК к белку
- т-РНК — соответствие между кодоном и аминокислотой
- геномы вирусов
  - двунитевая РНК: ротавирусы, вирусы без симптомов
  - РНК+ полиомиелит, менго, риновирусы (насморк), ящур
  - РНК- грипп,

Каталитические функции РНК

Каталитические функции

- Синтез белка — катализируется РНК
- Процессинг РНК
  - сплайсинг — катализируется РНК
  - авто-сплайсинг (самосплайсирующиеся интроны)
  - Рибонуклеаза P — фермент процессинга тРНК
  - Другие рибозимы
- Гипотеза РНК-мира

Направляющие функции РНК

Направляющие функции

- редактирование РНК
- модификация рРНК, тРНК
- микро-нуклеус $\rightarrow$ макро-нуклеус
- внутренняя инициация трансляции — IRES
- репарация ДНК

Регуляторные функции РНК

Регуляторные функции

- терминация транскрипции прокариот
- аттенюаторы
- секвесторы трансляции
- рибопереключатели
- Т-боксы
- микро РНК
- структуры, регулирующие сплайсинг
- PIWI - РНК
- компенсация дозы гена Х-хромосомы
- контроль репликации
- ...

Прочие функции РНК

Прочие функции

- теломеразная РНК
- тмРНК
- затравки для репликации

Неизвестные функции РНК

Неизвестные функции

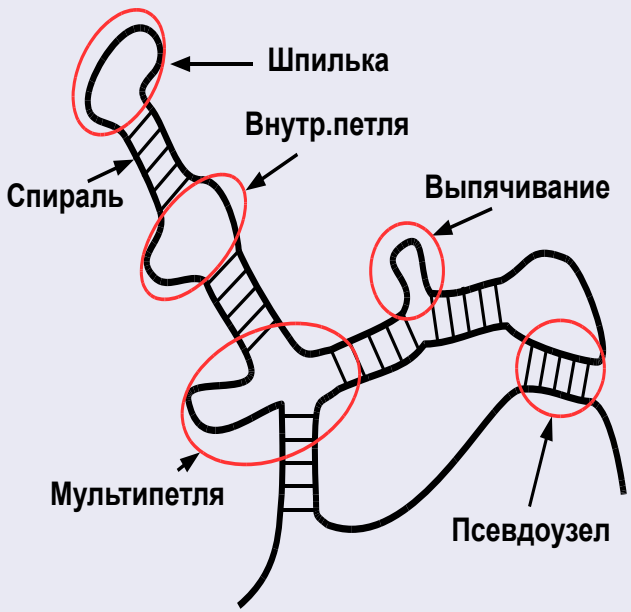
- длинные некодирующие РНК
- очень длинные некодирующие РНК
- короткие некодирующие РНК
- энхансерные и промоторные РНК

Вторичная структура РНК

Вторичная структура - совокупность спаренных оснований

Элементы вторичной структуры

- **спираль** = посл. спаренных оснований
- **шпилька** = петля, 1 спираль
- **внутр.петля** = 2 спирали
- **мультипетля** >2 спиралей
- **выпячивание** =2 спирали, по одной нити пары подряд



Экспериментальные данные

Эксперименты

- специфическая модификация спаренных или неспаренных оснований
- обратная транскрипция
  - на модифицированных основаниях срывается — смотрим где чаще срывы
  - на модифицированных основаниях делает ошибки — смотрим частоту ошибок
- массовое секвенирование и анализ результатов $\rightarrow$ вероятность спаривания

Золотой стандарт

Золотой стандарт

- Экспериментально определенные структуры
- Консервативные структуры (например, тРНК)

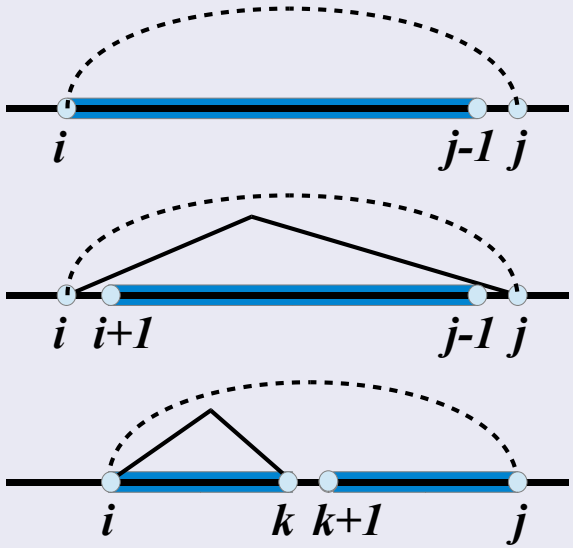
Алгоритм Нуссинофф

Оптимальная структура на сегменте $[i, j]$

i не спарено

i спарено с j

i спарено с $i < k < j$



рекурсия

$$S(i,j) = \max \begin{cases} S(i, j-1) \\ S(i+1, j-1) + 1 \\ S(i+1, k) + 1 + S(k+1, j) \end{cases}$$

Оптимизация энергии

Энергия вторичной структуры

- Энергия водородных связей — связи между основаниями на разных цепях
- Энергия стэкинга — взаимодействие оснований в стопке
- Энергия петель — энтропийный вклад

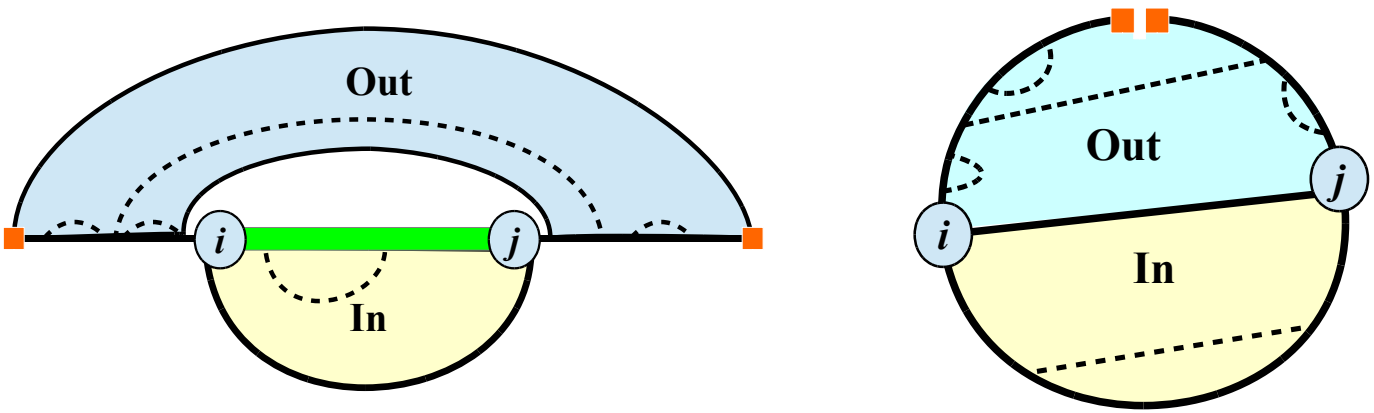
Задачи и алгоритмы

- минимизация полной энергии вторичной структуры $\Delta G \to \min$
- вычисление вероятностей спаривания $Z = \sum_{structures} \exp(-\frac{\Delta G}{kT})$; $p(ij) = \frac{Z(ij)}{Z}$
- поиск субоптимальных структур
- учет экспериментальных данных

Статистическая сумма

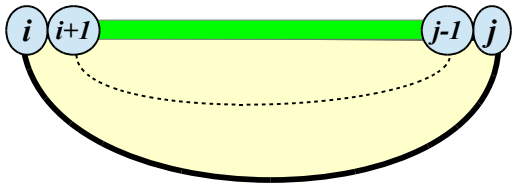
Условная стат.сумма: сумма по всем структурам, где $i \diamond j$:

$$Z(i,j) = \sum_{struct|i\diamond j} \exp\left(-\frac{\Delta G}{kT}\right) = \sum_{struct|i\diamond j} \exp\left(-\frac{(\Delta G_{in} + \Delta G_{out}}{kT}\right)$$

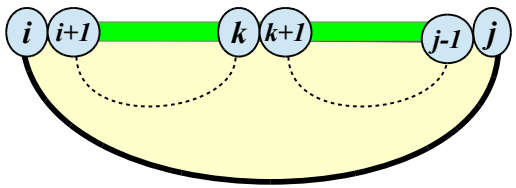


$$Z(i,j) = \sum_{In\ struct} \exp\left(-\frac{\Delta G_{In}}{kT}\right) \cdot \sum_{Out\ struct} \left(-\frac{\Delta G_{Out}}{kT}\right)$$

Статистическая сумма



$$In(i+1, j-1) \cdot exp\left(-\frac{\Delta G(i,j)}{kT}\right)$$



$$(In(i+1, k) + In(k+1, j-1)) \cdot exp\left(-\frac{\Delta G(i,j)}{kT}\right)$$

Внутренняя стат. сумма

$$In(i,j) = In(i+1, j-1) \cdot exp\left(-\frac{\Delta G(i,j)}{kT}\right) + \sum_k (In(i+1, k) + In(k+1, j-1)) \cdot exp\left(-\frac{\Delta G(i,j)}{kT}\right)$$

Статистическая сумма

Вероятность спаривания:

$$P(i \diamond j) = \frac{In(i,j) \cdot Out(i,j)}{In(0,L)}$$

Сравнительный анализ

Сравнительный анализ

Если есть много последовательностей, для которых мы ожидаем схожую структуру, то надо искать общую структуру для них

- Строим выравнивание, потом ищем общую структуру.
- Ищем общую структуру вместе с построением выравнивания.

Сравнительный анализ

Предсказания

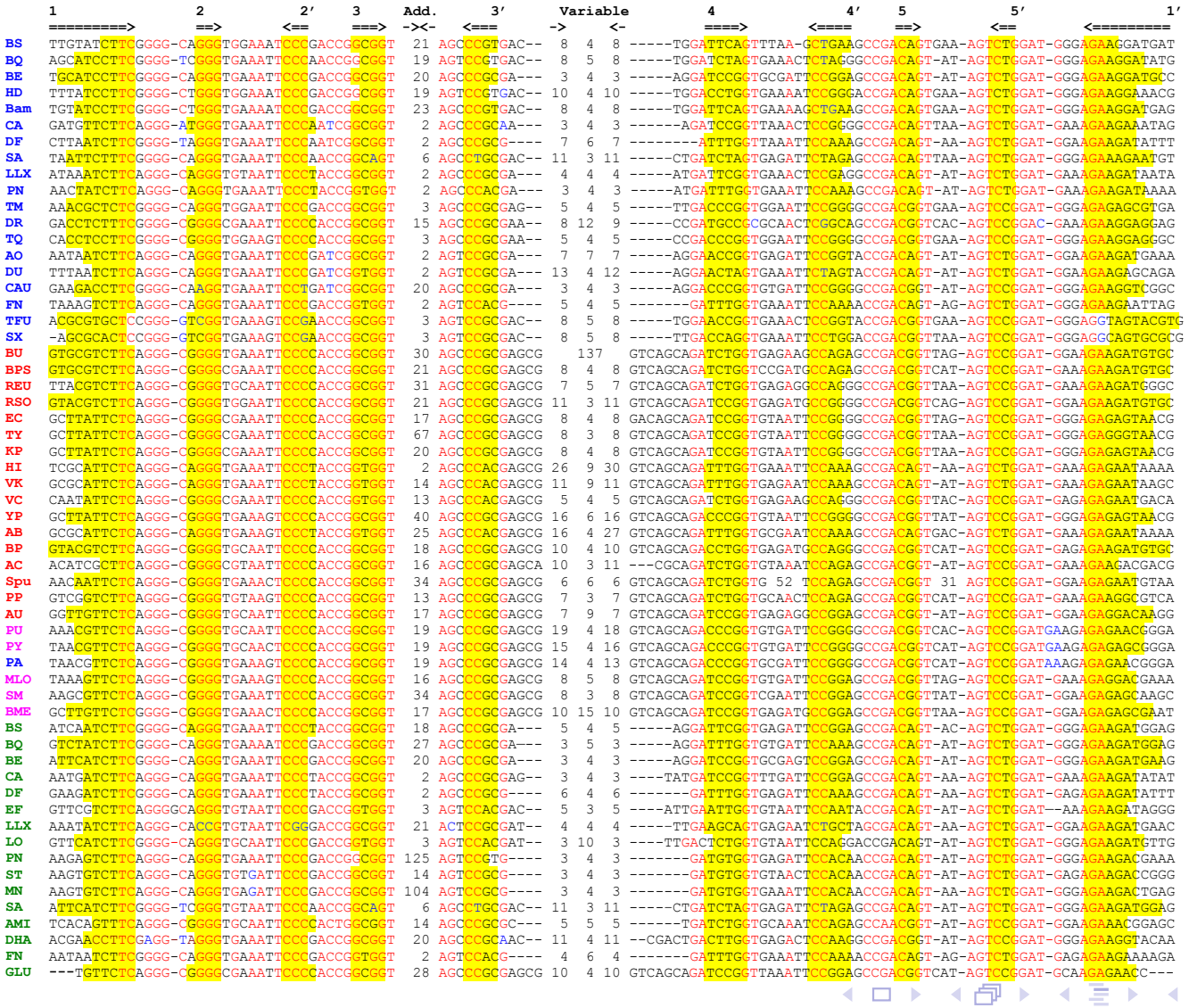
- транспортные РНК (Robert W. Holley, Hans Zachau)
- рибосомные РНК
- рибопереключатели
- рибозимы

Рибопереключатели

Исходные данные

- Есть перед одним из генов биосинтеза флавинов область ок. 150 нукл.:
  - некоторые мутации снимают регуляцию
  - некоторые мутации не меняют регуляцию
- Есть похожие области перед генами синтеза флавинов почти во всех геномах

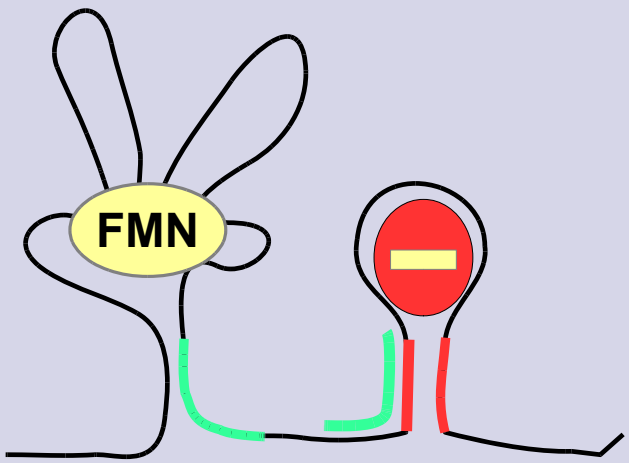
Рибопереключатели



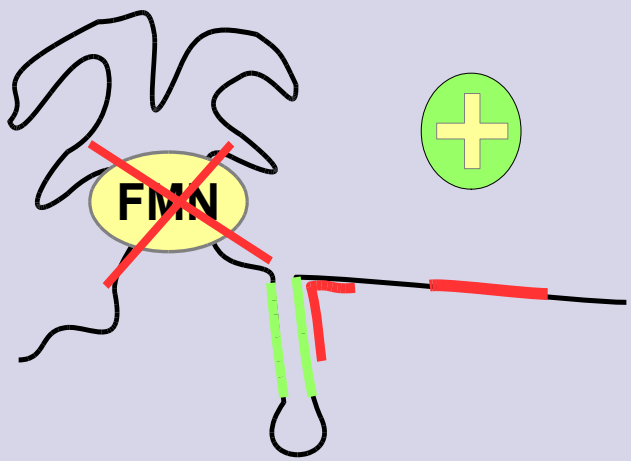
Рибопереключатели

Механизм работы

Есть флавин → запрет



Нет флавина → разрешение



Экспериментальные подтверждения

- разрушение структуры разрушает регуляцию
- рентгеноструктурный анализ

Рибопереключатели

Рибопереключатели

- Cobalamin
- FMN – flavin
- Glutamine
- Glycine
- Lysine
- PreQ1 – pre-queuosine1
- Purine riboswitches
- SAH – S-adenosyl homocysteine
- SAM – S-adenosyl methionine
- TPP – thiamin pyrophosphate
- ...

рибопереключатель Lis