

# Биоинформатика

Андрей Александрович Миронов

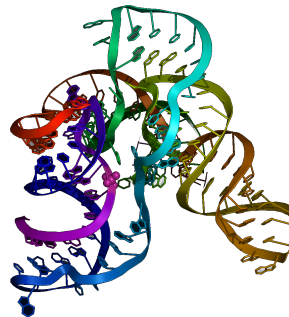
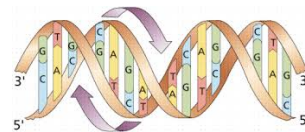
Факультет биоинженерии и биоинформатики

29 ноября 2015 г.

# 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 (самосплайсирующие интроны)
  - Рибонуклеаза Р — фермент процессинга тРНК
  - Другие рибозимы
- Гипотеза РНК-мира

# Направляющие функции РНК

## Направляющие функции

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

# Регуляторные функции РНК

## Регуляторные функции

- терминация транскрипции прокариот
- аттенюаторы
- секвесторы трансляции
- рибоперключатели
- Т-боксы
- микро РНК
- структуры, регулирующие сплайсинг
- PIWI - РНК
- компенсация дозы гена X-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

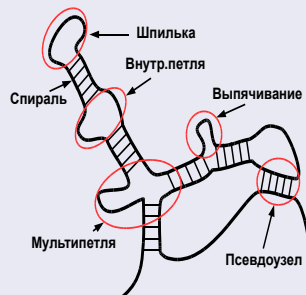


# Вторичная структура РНК

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 → вероятность спаривания

# Золотой стандарт

## Золотой стандарт

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

# Алгоритм Нуссинофф

## Оптимальная структура на сегменте $[i, j]$

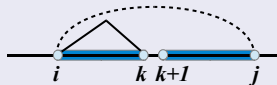
$i$  не спарено



$i$  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  
 $\Delta G \rightarrow \min$
- вычисление вероятностей спаривания  
$$Z = \sum_{structures} \exp\left(-\frac{\Delta G}{kT}\right); p(ij) = \frac{Z(ij)}{Z}$$
- поиск субоптимальных структур
- учет экспериментальных данных

# 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



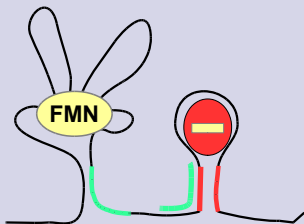
## Рибопереклюатели

	1	2	2'	3	Add.	3'	Variable	4	4'	5	5'	1'
	=====>	=====>	=====>	=====>	=====>	=====>	=====>	=====>	=====>	=====>	=====>	=====>
BS	TGATCTCTGGGG-CAGTGGAAATGACCCTCGT	21	AGC	GGAC	--	8	4	8	----	TGATTCAGTTTAA-CGTGAGCCCGAATGAA-AGT	TTGAT-GGGAGAGATGAT	
BQ	AGCATCTCTGGGG-TCGAATATGACCCTCGT	19	AGC	GGAC	--	8	5	8	----	TGATTCAGTTTAA-CGTGAGCCCGAATGAA-AGT	TTGAT-GGGAGAGATGAT	
BS	TGATCTCTGGGG-CAGTGGAAATGACCCTCGT	20	AGC	GGAC	--	8	5	8	----	TGATTCAGTTTAA-CGTGAGCCCGAATGAA-AGT	TTGAT-GGGAGAGATGAT	
BT	AGCATCTCTGGGG-TCGAATATGACCCTCGT	19	AGC	GGAC	--	10	4	10	----	TGATTCAGTTTAA-CGTGAGCCCGAATGAA-AGT	TTGAT-GGGAGAGATGAT	
CA	TGATCTCTGGGG-CAGTGGAAATGACCCTCGT	23	AGC	GGAC	--	8	4	8	----	TGATTCAGTTTAA-CGTGAGCCCGAATGAA-AGT	TTGAT-GGGAGAGATGAT	
Can	GATCTCTCTGGGG-TCGAATATGACCCTCGT	2	AGC	GGAA	--	3	4	3	----	AGATTCGCTTAAACCTGGGCCCCAATTA-AGT	TTGAT-GAAAGAGATATG	
CF	GTCATCTCTGGGG-TAGTAAATGACCCTCGT	2	AGC	GGAA	--	3	4	3	----	AGATTCGCTTAAACCTGGGCCCCAATTA-AGT	TTGAT-GAAAGAGATATG	
CH	GTCATCTCTGGGG-TAGTAAATGACCCTCGT	2	AGC	GGAA	--	11	5	11	----	CTGCTGATGAGATTTAGCCCGCAATTA-AGT	TTGAT-GGGAGAGATGAT	
LLX	ATAAATCTAGGG-CAGTGGAAATGACCCTCGT	2	AGC	GGAA	--	4	4	4	----	ATGATTCGCTTAAACCTGGGCCCCAATTA-AGT	TTGAT-GAAAGAGATATG	
FN	AACATCTCTAGGG-CAGTGGAAATGACCCTCGT	2	AGC	AGCA	--	3	4	3	----	ATGATTCGCTTAAACCTGGGCCCCAATTA-AGT	TTGAT-GAAAGAGATATG	
FN	AAGGCTCTAGGG-CAGTGGAAATGACCCTCGT	3	AGC	GGAA	--	5	4	5	----	TTGCTCCGCTGGCAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATGAT	
BR	GTCATCTCTGGGG-TAGTAAATGACCCTCGT	15	AGC	GGAA	--	8	12	9	----	CCGTCGCTGGCAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
TO	CATCTCTCTGGGG-CAGTGGAAATGACCCTCGT	3	AGC	GGAA	--	5	4	5	----	CCGCTCCGCTGGCAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATGAT	
AO	AATAATCTAGGG-CAGTGGAAATGACCCTCGT	2	AGC	GGAA	--	7	7	7	----	AGGACCGCTGGATTCGGTACCGCAATTA-AGT	TTGAT-GAAAGAGATATG	
DU	TTTAAATCTAGGG-CAGTGGAAATGACCCTCGT	2	AGC	GGAA	--	13	4	12	----	AGGACCGCTGGATTCGGTACCGCAATTA-AGT	TTGAT-GAAAGAGATATG	
CAU	GAAATCTCTAGGG-CAGTGGAAATGACCCTCGT	20	AGC	GGAA	--	3	4	3	----	AGGACCGCTGGATTCGGTACCGCAATTA-AGT	TTGAT-GGGAGAGATGAT	
FN	TAAATCTCTAGGG-CAGTGGAAATGACCCTCGT	2	AGC	AGCA	--	5	4	5	----	GATTCGCTTAAACCTGGGCCCCAATTA-AGT	TTGAT-GGGAGAGATATG	
TFU	ACCGCTCTAGGG-CAGTGGAAATGACCCTCGT	3	AGC	GGAC	--	8	5	8	----	TGATTCAGTTTAA-CGTGAGCCCGAATGAA-AGT	TTGAT-GGGAGAGATGAT	
SP	AAGGCTCTAGGG-CAGTGGAAATGACCCTCGT	3	AGC	GGAC	--	8	5	8	----	TTGCTCCGCTGGCAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATGAT	
BS	TTGCTCTCTAGGG-CAGTGGAAATGACCCTCGT	21	AGC	GGAGG		8	4	8		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
REU	TTAGCTCTCTAGGG-CAGTGGAAATGACCCTCGT	31	AGC	GGAGG		7	5	7		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
BSO	TTAGCTCTCTAGGG-CAGTGGAAATGACCCTCGT	21	AGC	GGAGG		11	3	11		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
KCO	GCTATCTCTAGGG-CAGTGGAAATGACCCTCGT	17	AGC	GGAGG		8	4	8		GACAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
FX	GCTATCTCTAGGG-CAGTGGAAATGACCCTCGT	67	AGC	GGAGG		3	8	8		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
BP	GCTATCTCTAGGG-CAGTGGAAATGACCCTCGT	20	AGC	GGAGG		8	4	8		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
HL	TCCATCTCTAGGG-CAGTGGAAATGACCCTCGT	2	AGC	AGAGC		26	9	30		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
VK	GCTATCTCTAGGG-CAGTGGAAATGACCCTCGT	14	AGC	AGAGC		11	9	11		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
VC	CAATCTCTAGGG-CAGTGGAAATGACCCTCGT	13	AGC	AGAGG		5	4	5		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
YB	GCTATCTCTAGGG-CAGTGGAAATGACCCTCGT	40	AGC	AGAGG		16	14	16		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
AB	GCCATCTCTAGGG-CAGTGGAAATGACCCTCGT	25	AGC	AGAGG		16	4	27		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
BP	TTAGCTCTCTAGGG-CAGTGGAAATGACCCTCGT	18	AGC	GGAGG		10	4	10		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
AC	ACATCTCTAGGG-CAGTGGAAATGACCCTCGT	16	AGC	AGAGC		10	3	11		CCGATTCGCTTAAACCTGGGCCCCAATTA-AGT	TTGAT-GAAAGAGATATG	
Spu	AACATCTCTAGGG-CAGTGGAAATGACCCTCGT	34	AGC	GGAGG		6	6	6		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
PP	GTCATCTCTAGGG-TAGTAAATGACCCTCGT	13	AGC	GGAGG		7	3	7		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
AU	GCCTCTCTAGGG-CAGTGGAAATGACCCTCGT	17	AGC	GGAGG		7	9	7		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
FU	AAAGCTCTAGGG-CAGTGGAAATGACCCTCGT	19	AGC	GGAGG		19	4	18		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
MY	TAAATCTCTAGGG-CAGTGGAAATGACCCTCGT	19	AGC	GGAGG		15	4	16		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
HA	TAAATCTCTAGGG-CAGTGGAAATGACCCTCGT	19	AGC	GGAGG		14	4	13		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
MLO	TAAATCTCTAGGG-CAGTGGAAATGACCCTCGT	16	AGC	GGAGG		8	5	8		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
SM	AGGCTCTCTAGGG-CAGTGGAAATGACCCTCGT	34	AGC	GGAGG		8	3	8		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
INE	GCCTCTCTAGGG-CAGTGGAAATGACCCTCGT	17	AGC	GGAGG		10	15	10		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
BS	AACATCTCTAGGG-CAGTGGAAATGACCCTCGT	18	AGC	GGAA	--	5	4	5	----	AGGACCGCTGGATTCGGTACCGCAATTA-AGT	TTGAT-GGGAGAGATATG	
BQ	GTCATCTCTGGGG-TAGTAAATGACCCTCGT	27	AGC	GGAA	--	3	3	3	----	AGGACCGCTGGATTCGGTACCGCAATTA-AGT	TTGAT-GGGAGAGATATG	
BE	ATTCACTCTGGGG-CAGTGGAAATGACCCTCGT	20	AGC	GGAA	--	3	4	3	----	AGGATTCGCTTAAACCTGGGCCCCAATTA-AGT	TTGAT-GGGAGAGATATG	
CA	AATCACTCTGGGG-CAGTGGAAATGACCCTCGT	2	AGC	GGAG	--	3	4	3	----	TATGATCGCTTTGATCGGAGCCCGAATTA-AGT	TTGAT-GAAAGAGATATG	
DF	GAAATCTCTGGGG-CAGTGGAAATGACCCTCGT	2	AGC	GGGG	--	6	4	6	----	CTTTGCTGGATTCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
EF	GTCATCTCTGGGG-TAGTAAATGACCCTCGT	2	AGC	GGAA	--	5	3	5	----	ATGATTCGCTTAAACCTGGGCCCCAATTA-AGT	TTGAT-GGGAGAGATATG	
LLX	AAATATCTCTAGGG-CAGTGGAAATGACCCTCGT	21	AGC	AGAT	--	4	4	4	----	TTGATTCAGTTTAA-CGTGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
LXO	GCTATCTCTGGGG-CAGTGGAAATGACCCTCGT	3	AGC	AGAT	--	3	10	3	----	TTGACTCTCTGATAATCCGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
FN	AGAGCTCTCTAGGG-CAGTGGAAATGACCCTCGT	125	AGC	GTC	--	3	4	3	----	GATTCGCTTAAACCTGGGCCCCAATTA-AGT	TTGAT-GGGAGAGATATG	
ST	AGGCTCTCTAGGG-CAGTGGAAATGACCCTCGT	14	AGC	GGAG	--	3	4	3	----	CTTTGCTGGATTCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
HN	AGGCTCTCTAGGG-CAGTGGAAATGACCCTCGT	14	AGC	GGAG	--	3	4	3	----	CTTTGCTGGATTCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
SA	ATTCACTCTGGGG-CAGTGGAAATGACCCTCGT	6	AGC	TGGAC	--	11	3	11	----	CTGCTGATGAGATTTAGCCCGCAATTA-AGT	TTGAT-GGGAGAGATATG	
AM	TCCAGCTTCTAGGG-CAGTGGAAATGACCCTCGT	4	AGC	GCCC	--	5	5	5	----	TGATCTCTGGCAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
DLA	ACGATCTCTAGGG-TAGTAAATGACCCTCGT	20	AGC	GGAA	--	11	4	11	----	CGACTCTCTGGCAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
FN	AATATCTCTAGGG-CAGTGGAAATGACCCTCGT	2	AGC	AGCA	--	5	4	5	----	TTGCTCCGCTGGCAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	
GBU	CTTTCTCTAGGG-CAGTGGAAATGACCCTCGT	28	AGC	GGAGG		10	4	10		GTCAGGAGCTGCTGCGAATCGGAGCCCGAATTA-AGT	TTGAT-GGGAGAGATATG	

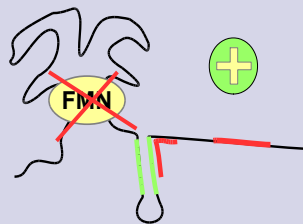
# Рибопереключатели

## 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

