

Предсказание вторичной структуры РНК

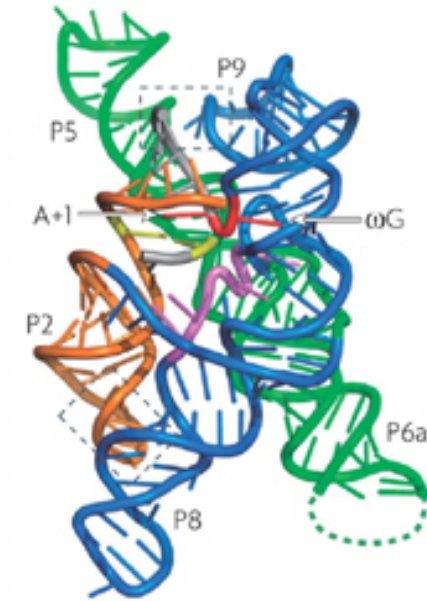
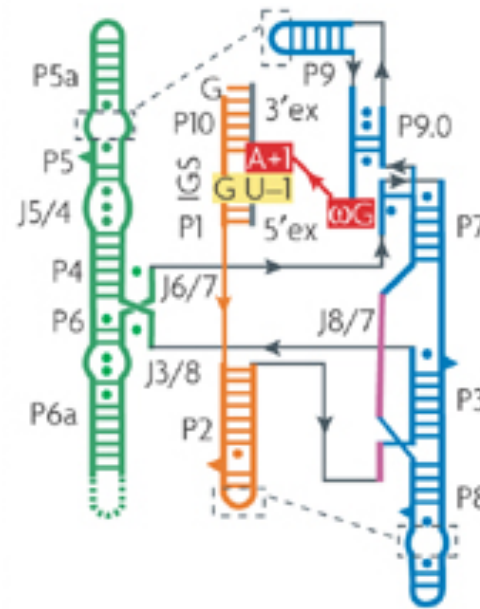
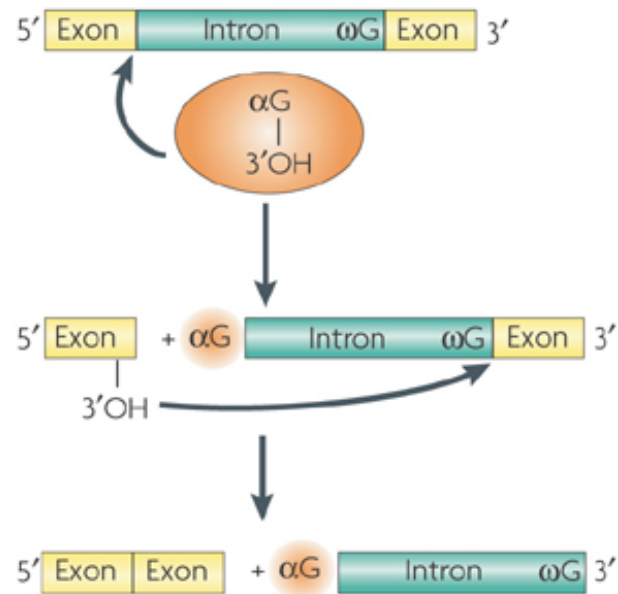
Неотъемлемые биохимические функции живых организмов

- Хранение и копирование генетической информации
- Ферментативная активность: катализ биохимических реакций
- Регуляция экспрессии генов

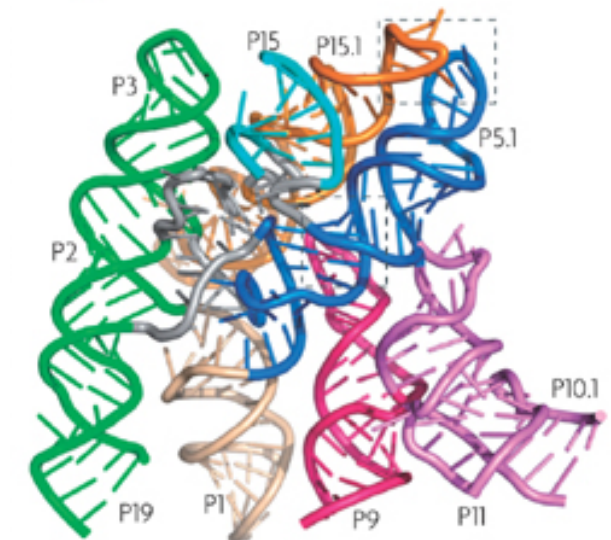
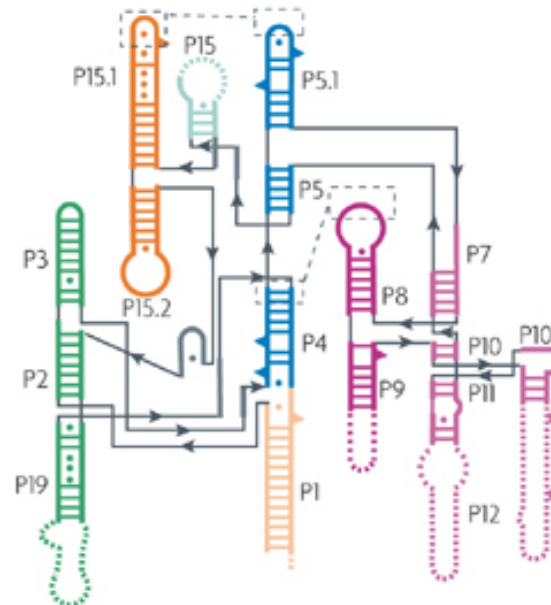
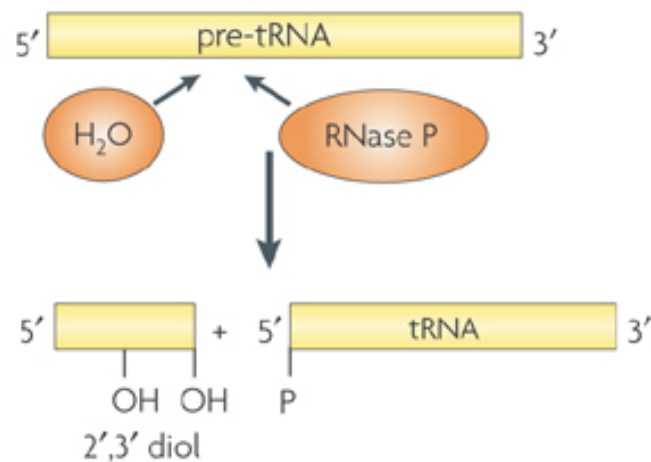
Катализ биохимических реакций: рибозимы

Самосплайсирующиеся интроны и RNase P - первые из открытых РНК, обладающих каталитической активностью

Интроны группы I

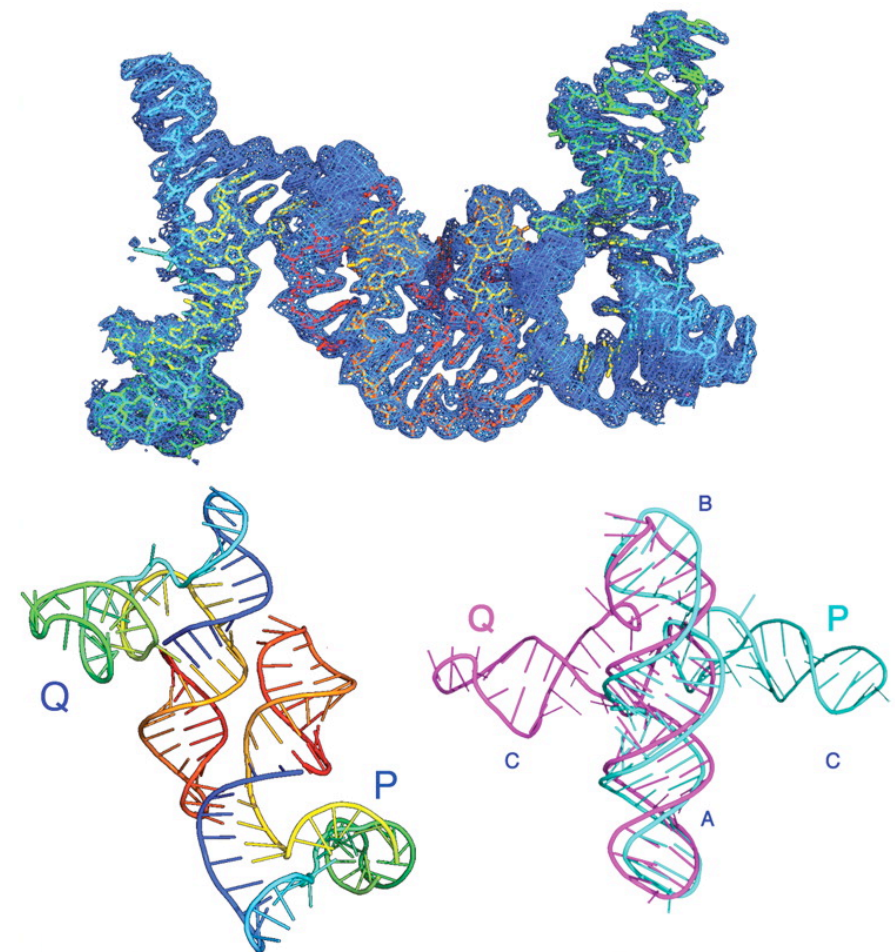
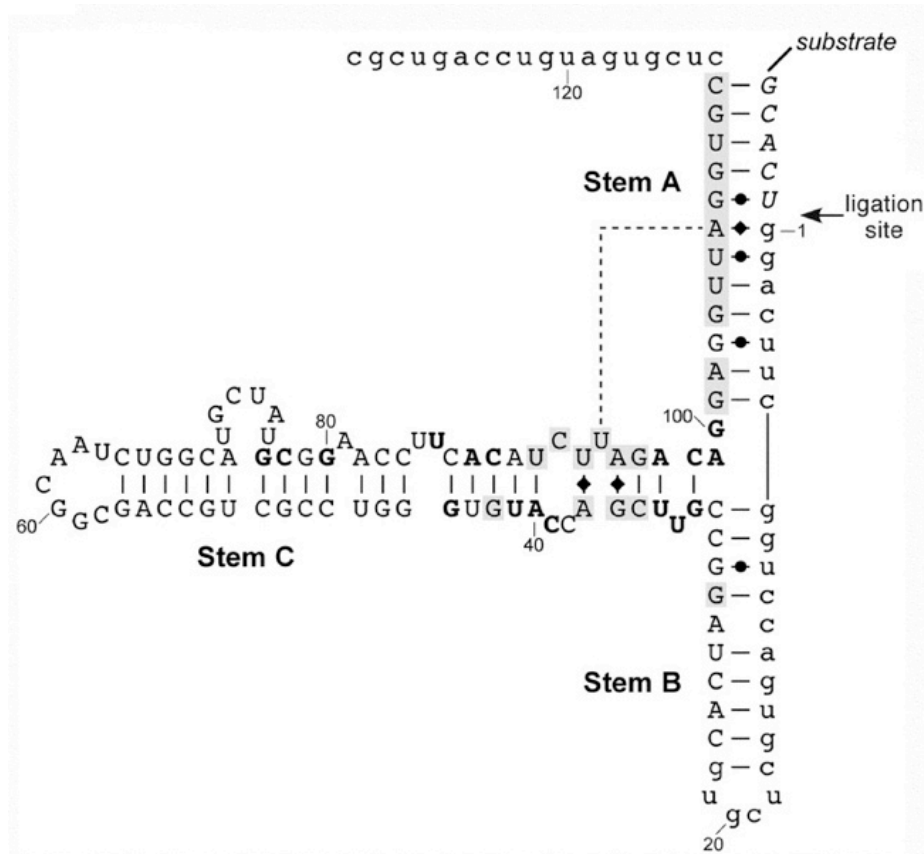
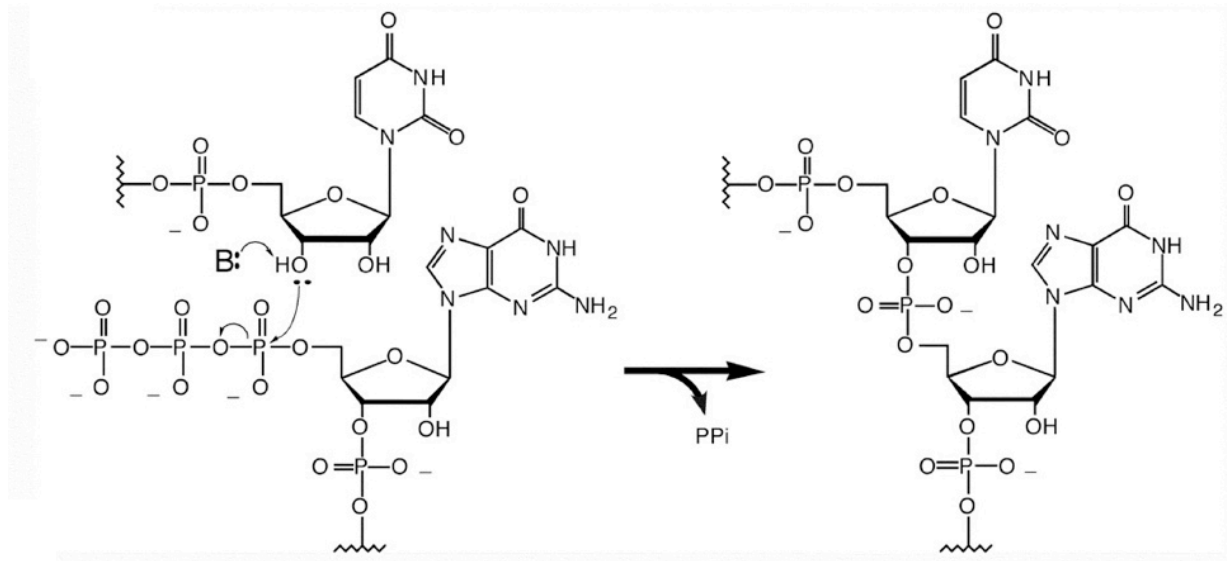


RNase P



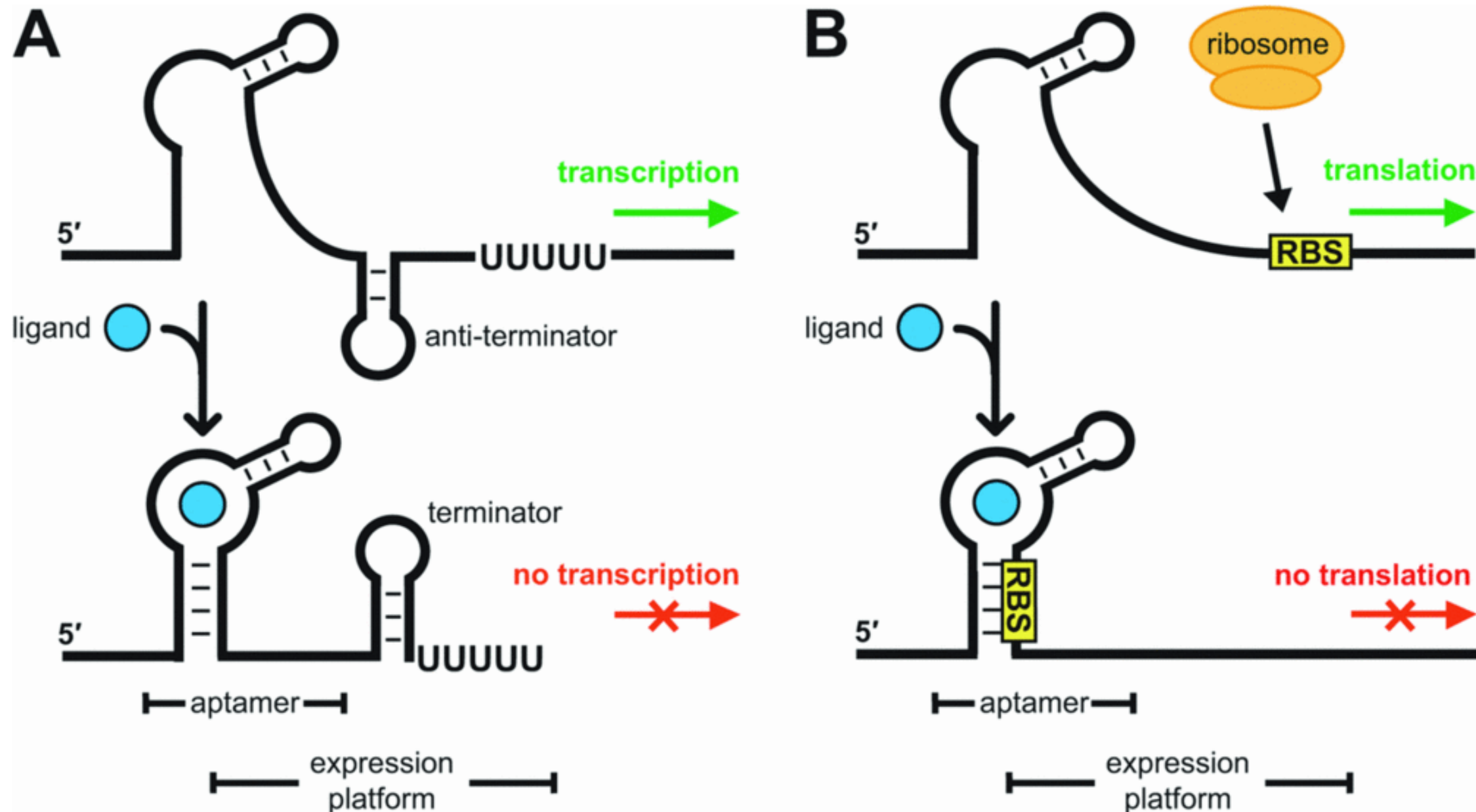
РНК репликация

Известен рибозим (молекула РНК), способный катализировать реакцию копирования (репликации) РНК-цепей



Регуляция экспрессии генов: РНК-переключатели

РНК-переключатели - структуры РНК способные связывать малые молекулы (лиганды) и в зависимости от этого включать/выключать работу генов - регулировать экспрессию генов.



Гипотеза мира РНК

РНК или РНК-подобные молекулы могли существовать до возникновения ДНК и белков - гипотеза "мира РНК"

618

NEWS AND VIEWS

NATURE VOL. 319 20 FEBRUARY 1986

Origin of life

The RNA world

from Walter Gilbert

UNTIL recently, when one thought of the varied molecular processes at the origin of life, one imagined that the first self-replicating systems consisted of both RNA and protein. RNA served to hold information, whereas protein molecules provided all the enzymic activities needed to make copies of RNA and to reproduce themselves. The cycle that developed a self-replicating system out of the primitive soup of amino acids and nucleotides had two radically different components¹.

Now it seems possible that the informational and catalytic properties of these two components may be combined in a single molecular species. Last week in these columns Frank Westheimer² described the discovery of enzymic activities in the RNA molecules of *Escherichia coli*, in which ribonuclease-P cuts phosphodiester bonds during the maturation of the transfer RNA molecule^{3,4}, and of *Tetrahymena*, whose ribosomal RNA contains a self-splicing exon⁵⁻⁷. If there are two enzymic activities associated with RNA, there may be more. And if there are activities among

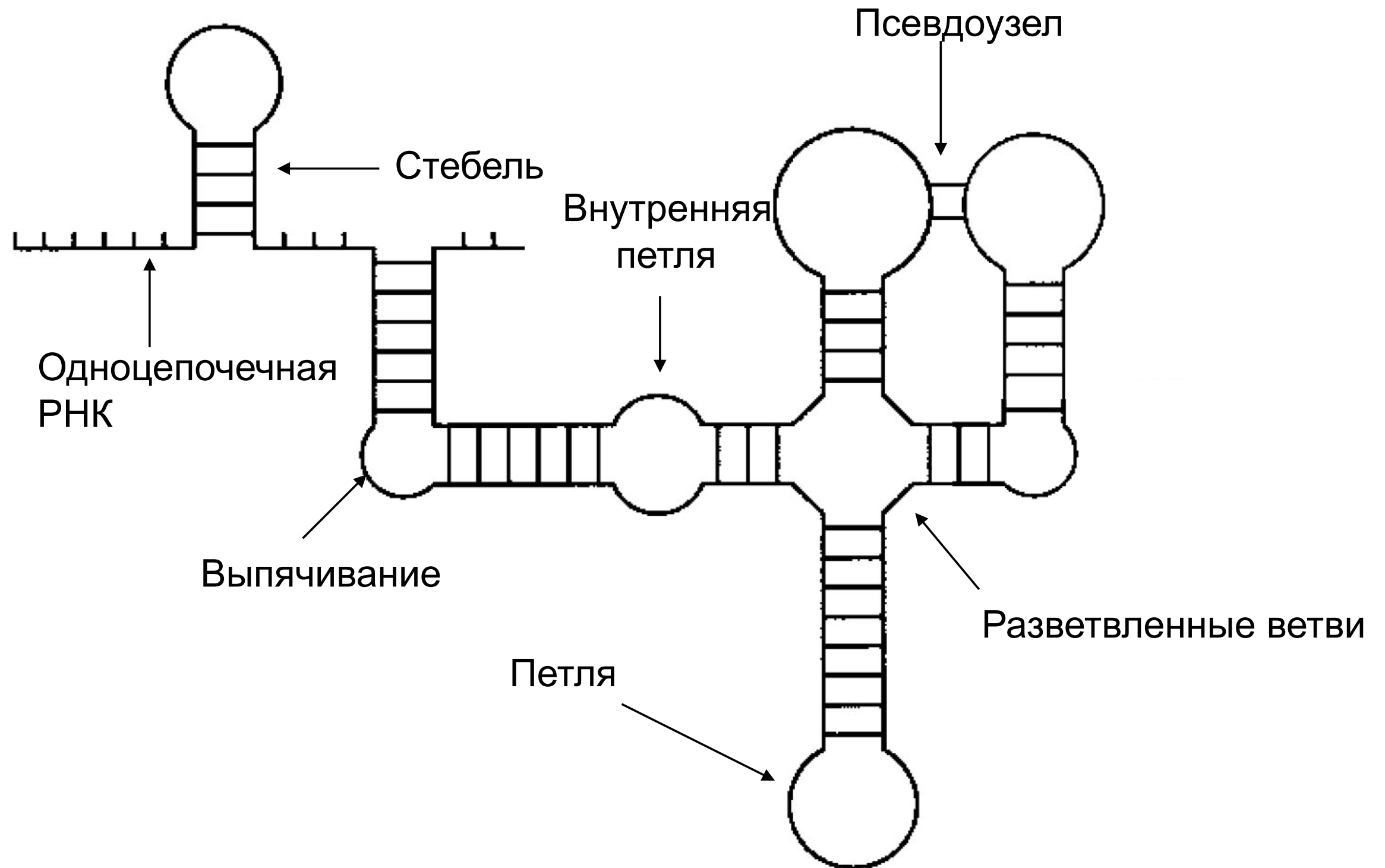
useful exon to pass from one replicating structure to an unrelated one.

This picture of the RNA world is one of replicating molecules that reassort exons by transposable elements created by introns. This process builds and remakes RNA molecules by chunks and also permits the useful distinction between information and function. Information storage needs to be one-dimensional, for ease of copying, but molecules with enzymic functions tend to be tight three-dimensional structures, whose forms are unrelated to the demands of any copying mechanism. (This dichotomy is most obvious today between the linear order along DNA and the structure of proteins.) In the RNA world, the structure that would be replicated has the full complement of introns. Some of the daughters, by splicing out all their introns, would convert to functional molecules, the ribozymes. A remnant of this process may be the structure of transfer RNA, where a compact secondary structure is broken up by the insertion of an intron.

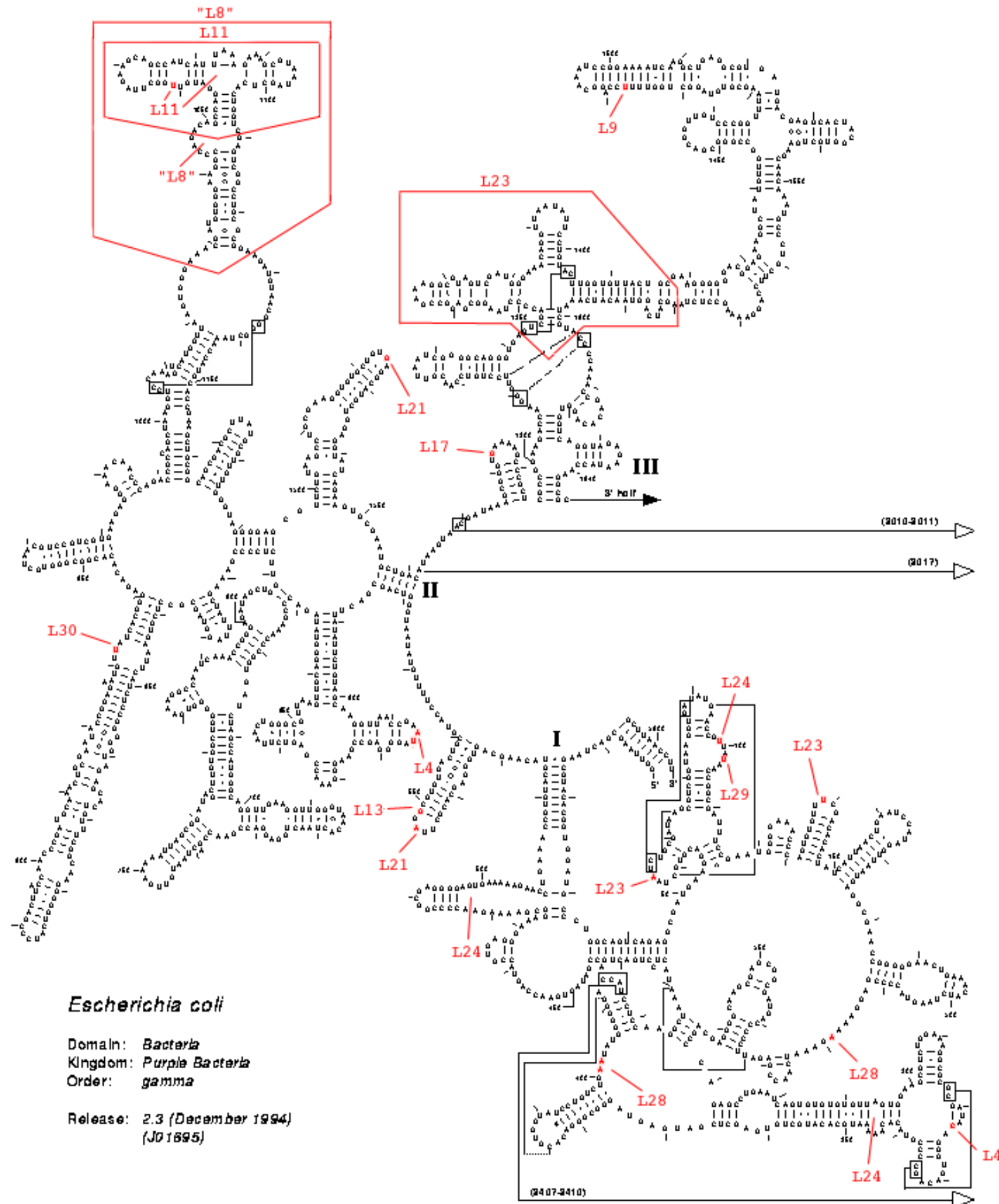
by arranging them according to an RNA template using other RNA molecules such as the RNA core of the ribosome. This process would make the first proteins, which would simply be better enzymes than their RNA counterparts. I suggest that protein molecules do not carry out enzymic reactions of a different nature from RNA molecules but are able to perform the same reactions more effectively and rapidly, and hence will eventually dominate. These protein enzymes are encoded by RNA exons, thus they, in turn, are built up of mini-elements of structure.

Finally, DNA appeared on the scene, the ultimate holder of information copied from the genetic RNA molecules by reverse transcription. After double-stranded DNA evolved there exists a stable linear information store, error-correcting because of its double-stranded structure but still capable of mutation and recombination. RNA is then relegated to the intermediate role that it has today — no longer the centre of the stage, displaced by DNA and the more effective protein enzymes. But a few RNA enzymic activities still exist, the two described recently³⁻⁷, and possibly others in the role of ribosomal RNA or in the splicing of eukaryotic messenger RNA. The relic of this process is the intron/exon structure of genes, left imprinted on DNA from the

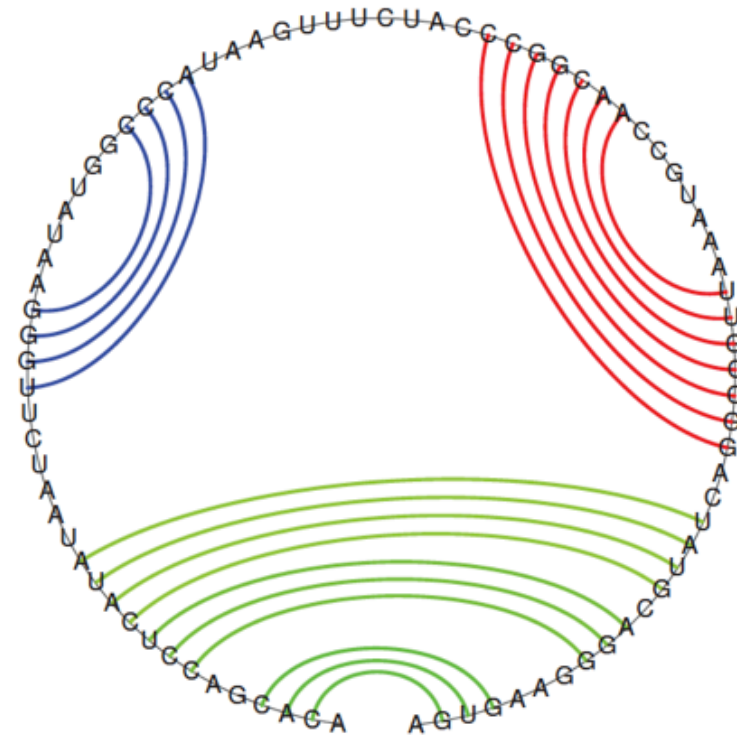
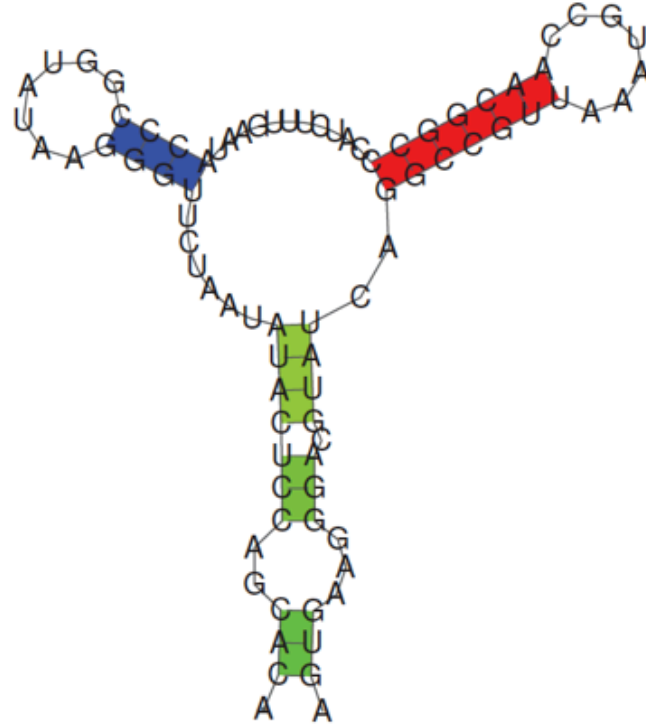
Основы вторичной структуры РНК



Secondary Structure: large subunit ribosomal RNA - 5' half

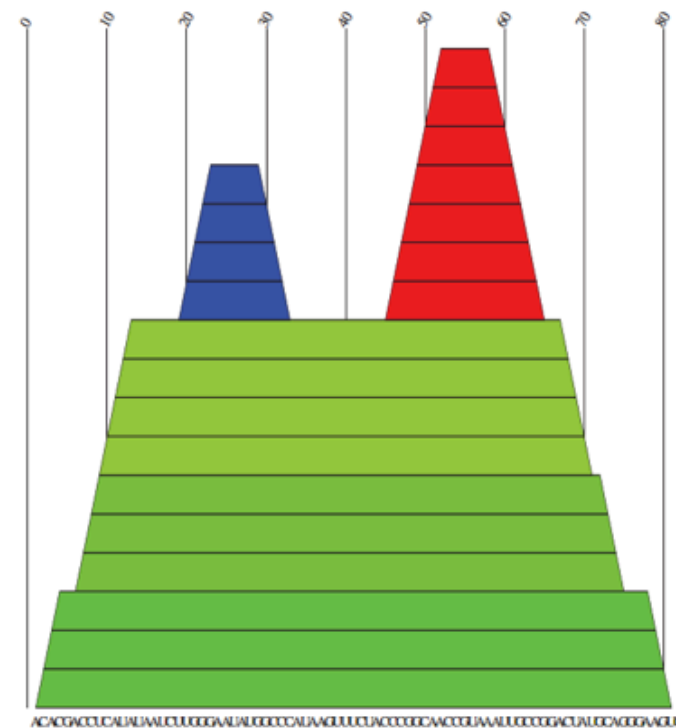
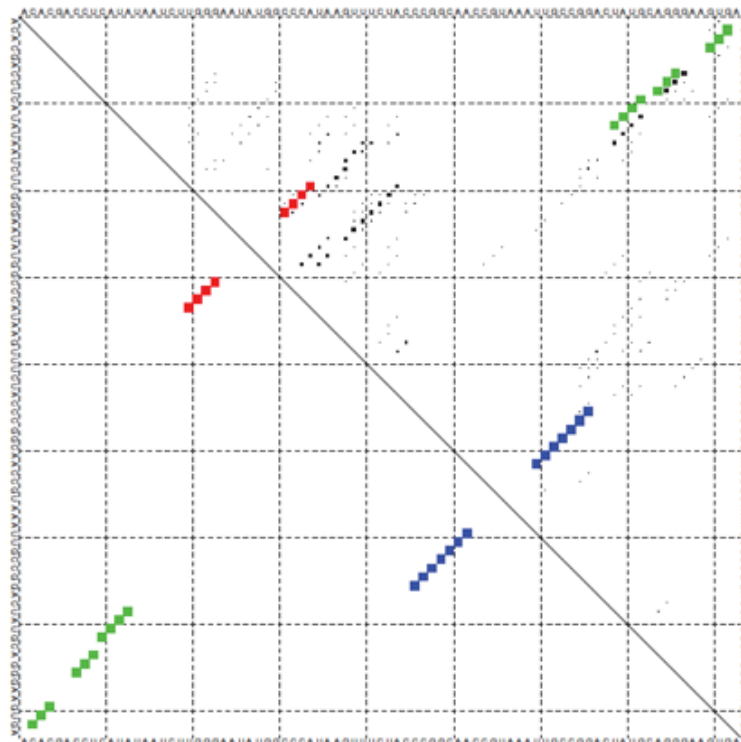


Способы представления вторичной структуры РНК



ACACGACCUCUAUUAUUCUUGGGAAUAUGGCCCAUAAGUUUCUACCCGGCAACCGUAAAUUGCCGGACUAUGCAGGGAAGUGA
.((((.(((((((.....(((.....))).....(((.....))).....))).....))).....))).....

RF00167



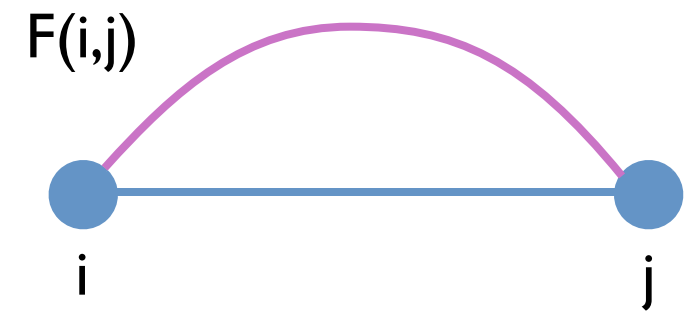
Предсказание вторичной структуры РНК: алгоритм динамического программирования Нуссинов

Основанная идея: максимизация числа спаренных оснований

Дана последовательность $X = x_1 \dots x_L$

Определим матрицу динамического программирования:

$F(i,j)$ = максимальное число спаренных оснований на отрезке (i,j)



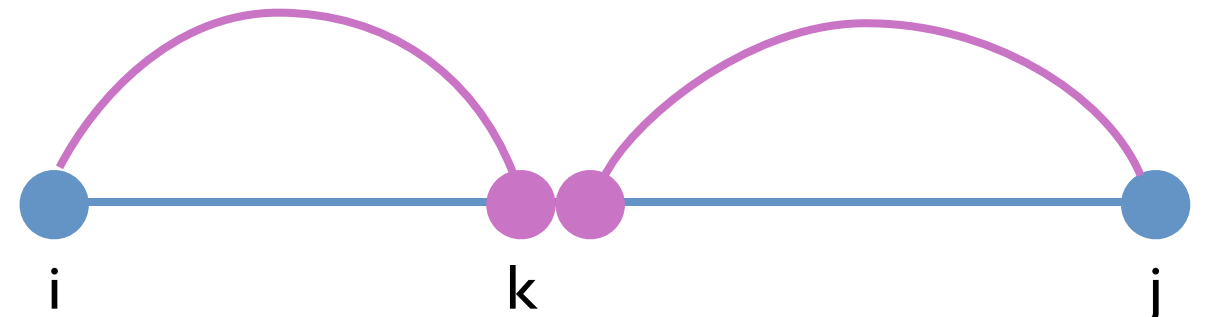
Два случая ($i < j$):

1. x_i спарено с x_j

$$F(i,j) = s(x_i, x_j) + F(i+1, j-1)$$

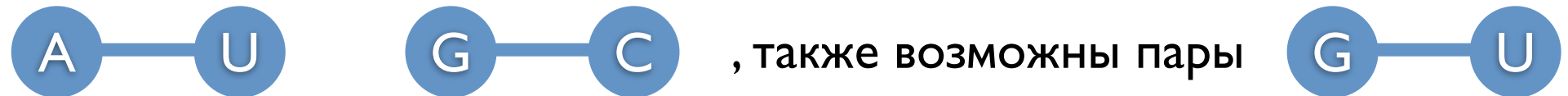
2. x_i не спарено с x_j

$$F(i,j) = \max\{ k: i < k < j-1 \} F(i,k) + F(k+1, j)$$



Алгоритм Нуссинов

- комплиментарные основания:



- способ сворачивания последовательности РНК длины L отображается в симметричной матрице $L \times L$
- бонус для комплиментарных оснований равен 1
- 0 в противном случае

Алгоритм Нуссинов

- рекурсивное нахождение наилучших структур для подпоследовательностей
- итеративное рассмотрение подпоследовательностей в направлении увеличения длины
- на каждом шаге четыре возможных случая получения наилучшей структуры

Инициализация:

$$\begin{array}{ll} F(i, i-1) = 0; & \text{от } i=2 \text{ до } L \\ F(i, i) = 0; & \text{от } i=1 \text{ до } L \end{array}$$

Итерации:

От $t=2$ до L :

От $i=1$ до $L-t$:

$$j = i + t - 1$$

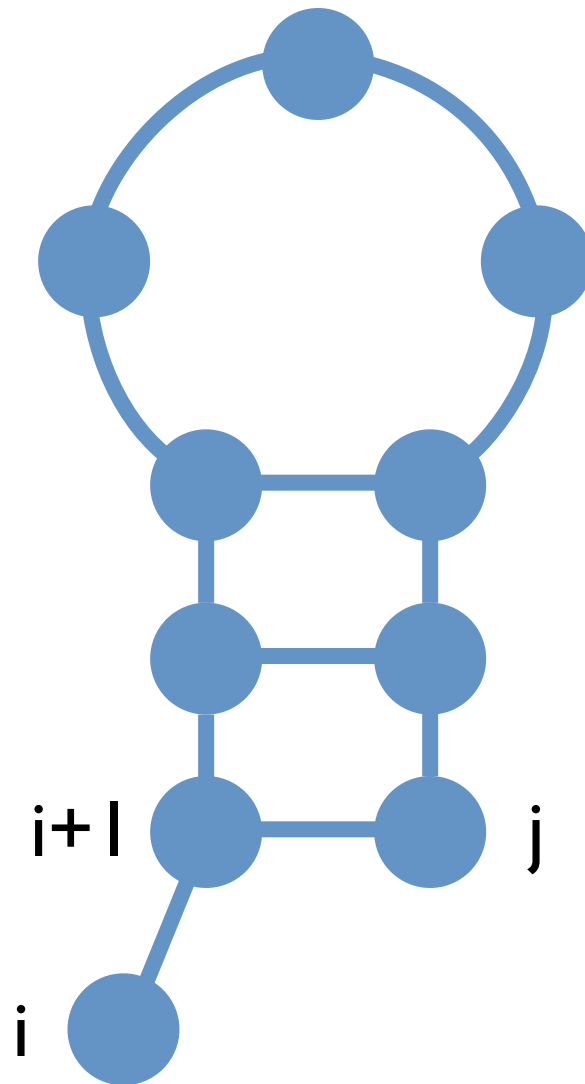
$$F(i, j) = \max \left\{ \begin{array}{l} F(i+1, j-1) + s(x_i, x_j) \\ \max_{i < k < j-1} \{ F(i, k) + F(k+1, j) \} \end{array} \right.$$

Завершение:

Наилучшая структура определяется для $F(1, L)$ процедурой обратного прохода

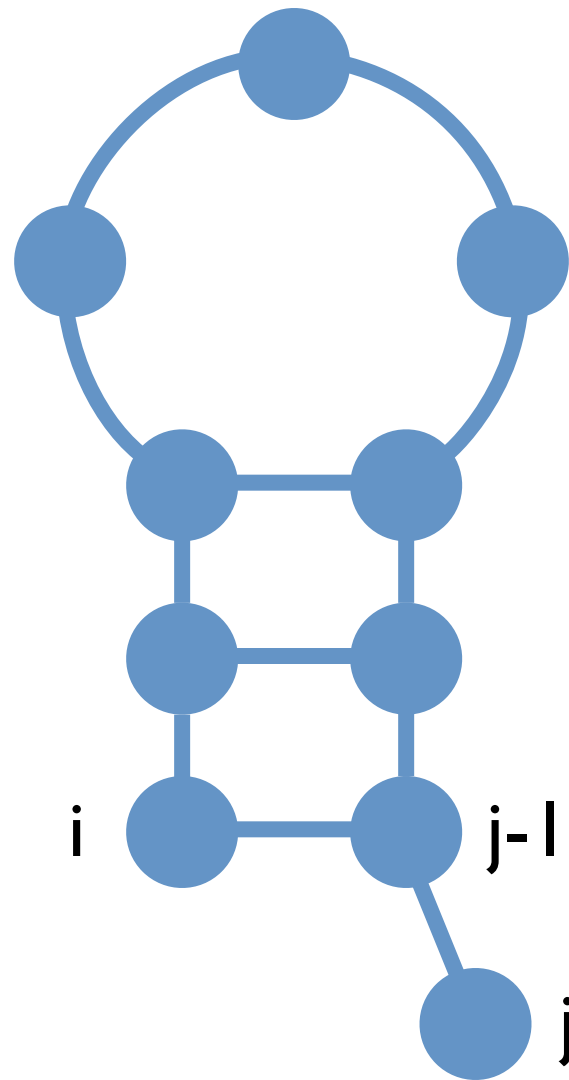
Алгоритм Нуссинов

Случай I: добавление неспаренного i -го основания для наилучшей структуры, найденной для подпоследовательности $(i+1, j)$



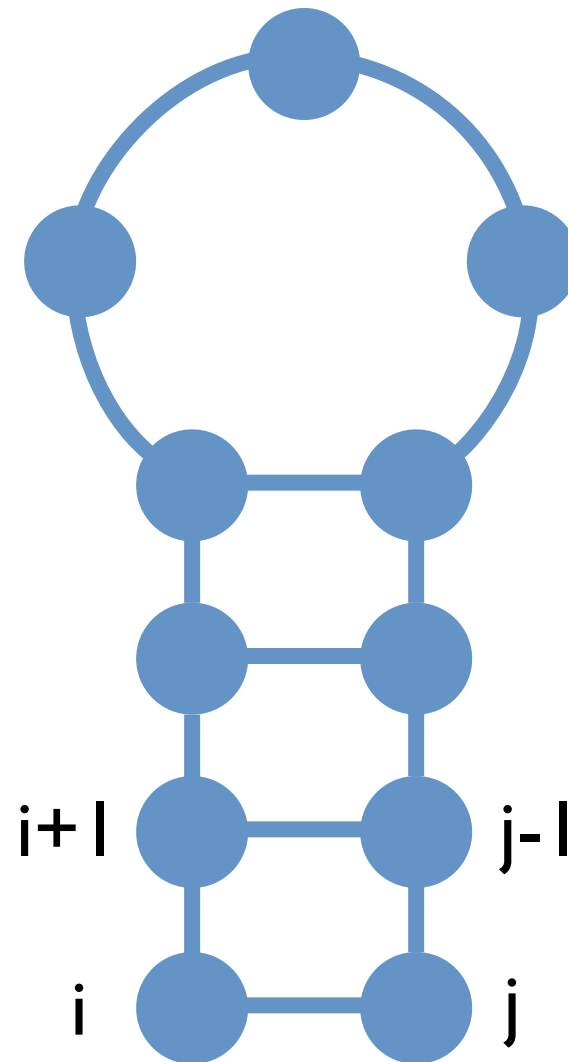
Алгоритм Нуссинов

Случай 2: добавление неспаренного j -го основания для наилучшей структуры, найденной для подпоследовательности $(i, j-1)$



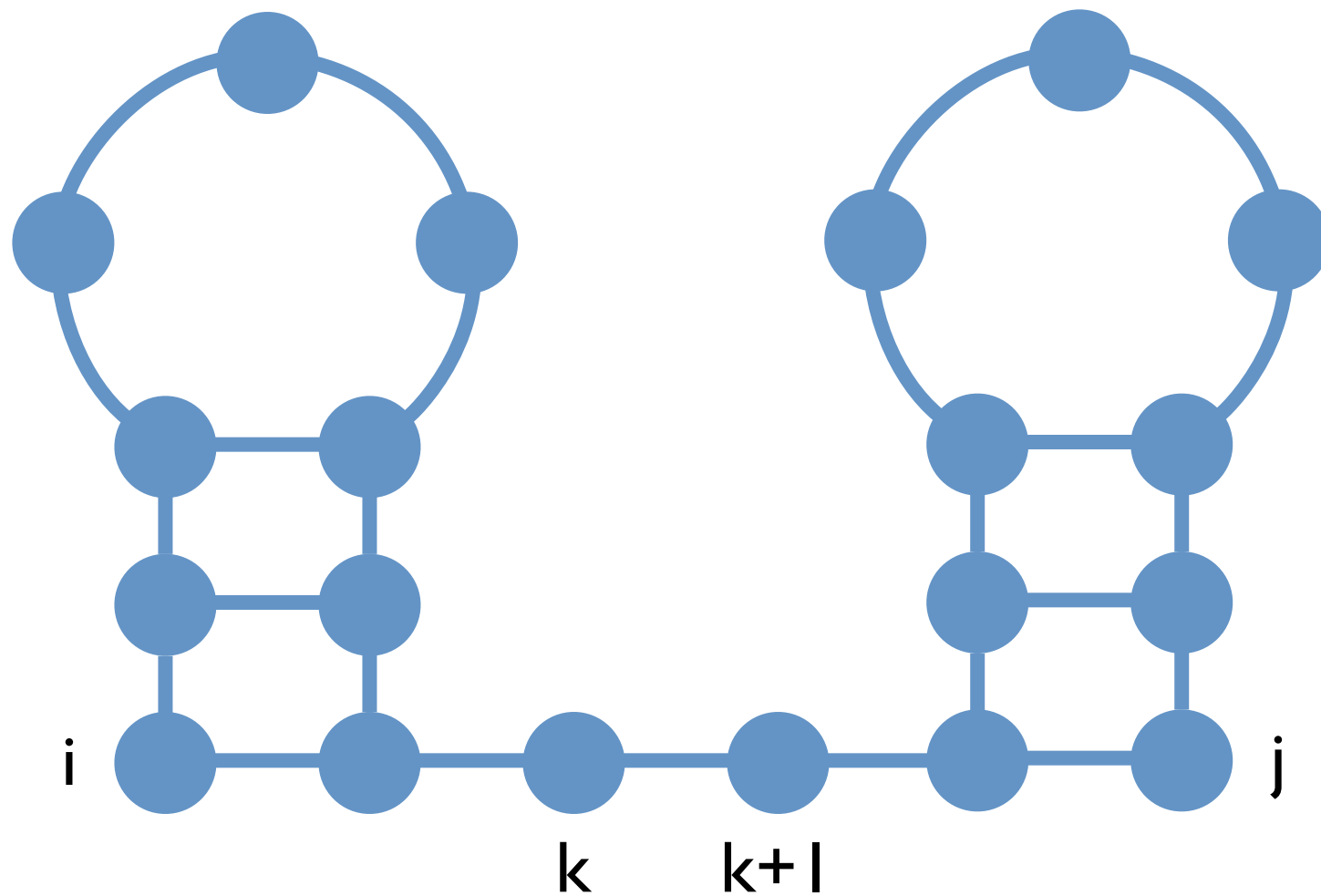
Алгоритм Нуссинов

Случай 3: добавление спаренных оснований (i,j) для наилучшей структуры, найденной для подпоследовательности $(i+1,j-1)$



Алгоритм Нуссинов

Случай 4: бифуркация - объединение двух наилучших структур полученных для подпоследовательностей (i,k) и $(k+1,j)$



Алгоритм Нуссинов

Инициализация:

$F(i, i-1) = 0;$

$F(i, i) = 0;$

от $i=2$ до L

от $i=1$ до L

	G	G	G	A	A	A	T	C	C
G	0								
G	0	0							
G		0	0						
A			0	0					
A				0	0				
A					0	0			
T						0	0		
C							0	0	
C								0	0

Алгоритм Нуссинов

$$F(i,j) = \max \left\{ \begin{array}{l} F(i+1,j) \\ F(i,j-1) \\ F(i+1,j-1) + s(i,j) \\ \max_{i < k < j-1} \{ F(i,k) + F(k+1,j) \} \end{array} \right.$$

	G	G	G	A	A	A	T	C	C
G	0								
G	0	0							
G		0	0						
A			0	0					
A				0	0				
A					0	0			
T						0	0		
C							0	0	
C								0	0

Алгоритм Нуссинов

$$F(i,j) = \max \left\{ \begin{array}{l} F(i+1,j) \\ F(i,j-1) \\ F(i+1,j-1) + s(i,j) \\ \max_{i < k < j-1} \{ F(i,k) + F(k+1,j) \} \end{array} \right.$$

	G	G	G	A	A	A	T	C	C
G	0	0							
G	0	0	0						
G		0	0	0					
A			0	0	0				
A				0	0	0			
A					0	0	1		
T						0	0	0	
C							0	0	0
C								0	0

Алгоритм Нуссинов

$$F(i,j) = \max \left\{ \begin{array}{l} F(i+1,j) \\ F(i,j-1) \\ F(i+1,j-1) + s(i,j) \\ \max_{i < k < j-1} \{ F(i,k) + F(k+1,j) \} \end{array} \right.$$

	G	G	G	A	A	A	T	C	C
G	0	0	0						
G	0	0	0	0					
G		0	0	0	0				
A			0	0	0	0			
A				0	0	0	1		
A					0	0	1	1	
T						0	0	0	0
C							0	0	0
C								0	0

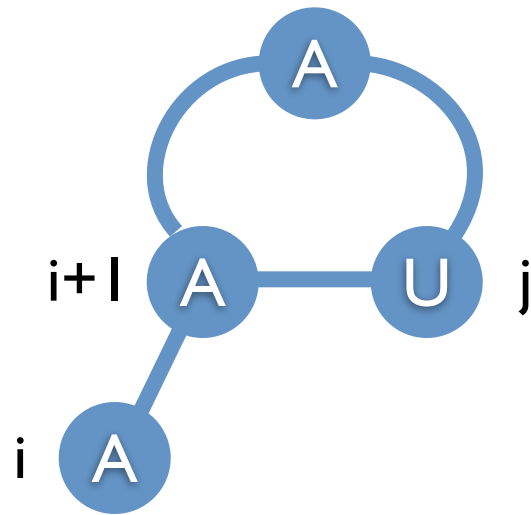
Алгоритм Нуссинов

$$F(i,j) = \max \left\{ \begin{array}{l} F(i+1,j) \\ F(i,j-1) \\ F(i+1,j-1) + s(i,j) \\ \max_{i < k < j-1} \{ F(i,k) + F(k+1,j) \} \end{array} \right.$$

	G	G	G	A	A	A	T	C	C
G	0	0	0	0					
G	0	0	0	0	0				
G		0	0	0	0	0			
A			0	0	0	0	1		
A				0	0	0	1	1	
A					0	0	1	1	1
T						0	0	0	0
C							0	0	0
C								0	0

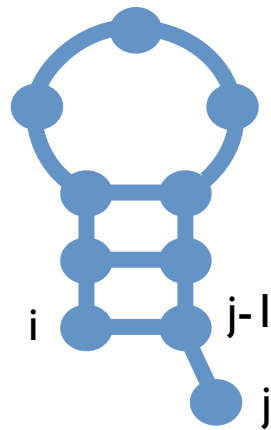
Алгоритм Нуссинов

$$F(4,7) = \max \left\{ \begin{array}{l} F(5,7) \leftarrow \\ F(4,6) \\ F(5,6) + s(4,7) \\ \max_{4 < k < 6} \{ F(4,k) + F(k+1,7) \} \end{array} \right.$$

[illegible]

Алгоритм Нуссинов

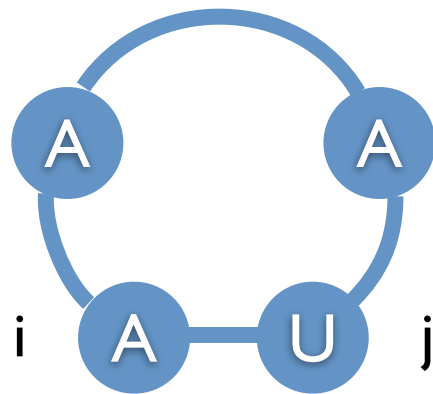
$$F(4,7) = \max \left\{ \begin{array}{l} F(5,7) \\ F(4,6) \quad \leftarrow \\ F(5,6) + s(4,7) \\ \max_{4 < k \leq 6} \{F(4,k) + F(k+1,7)\} \end{array} \right.$$



		1	2	3	4	5	6	7	8	9
1	G	0	0	0	0					
2	G	0	0	0	0	0				
3	G		0	0	0	0	0			
4	A			0	0	0	0			
5	A				0	0	0	1	1	
6	A					0	0	1	1	1
7	T						0	0	0	0
8	C							0	0	0
9	C								0	0

Алгоритм Нуссинов

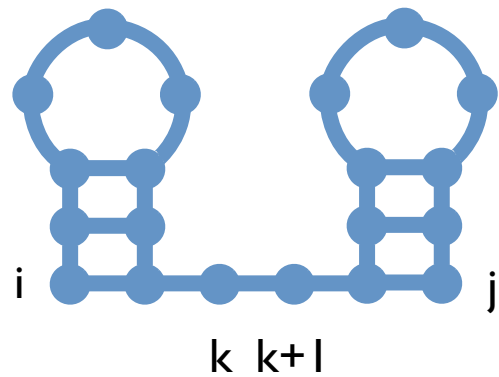
$$F(4,7) = \max \left\{ \begin{array}{l} F(5,7) \\ F(4,6) \\ F(5,6) + s(4,7) \quad \leftarrow \\ \max_{4 < k < 6} \{F(4,k) + F(k+1,7)\} \end{array} \right.$$



		1	2	3	4	5	6	7	8	9
1	G	0	0	0	0					
2	G	0	0	0	0	0				
3	G		0	0	0	0	0			
4	A			0	0	0	0			
5	A				0	0	0	1	1	
6	A					0	0	1	1	1
7	T						0	0	0	0
8	C							0	0	0
9	C								0	0

Алгоритм Нуссинов

$$F(4,7) = \max \left\{ \begin{array}{l} F(5,7) \\ F(4,6) \\ F(5,6) + s(4,7) \\ \max_{4 < k < 6} \{F(4,k) + F(k+1,7)\} \end{array} \right. \quad \leftarrow$$



		1	2	3	4	5	6	7	8	9
		G	G	G	A	A	A	T	C	C
1	G	0	0	0	0					
2	G	0	0	0	0	0				
3	G		0	0	0	0	0			
4	A			0	0	0	0			
5	A				0	0	0	1	1	
6	A					0	0	1	1	1
7	T						0	0	0	0
8	C							0	0	0
9	C								0	0

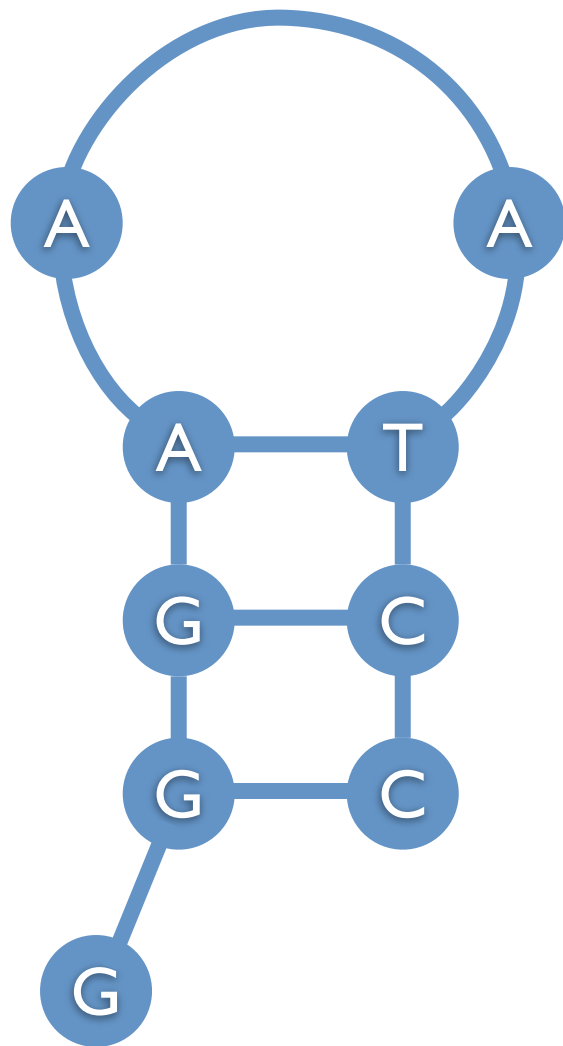
Алгоритм Нуссинов

$$F(i,j) = \max \left\{ \begin{array}{l} F(i+1,j) \\ F(i,j-1) \\ F(i+1,j-1) + s(i,j) \\ \max_{i < k < j-1} \{ F(i,k) + F(k+1,j) \} \end{array} \right.$$

		1	2	3	4	5	6	7	8	9
		G	G	G	A	A	A	T	C	C
1	G	0	0	0	0	0	0	1	2	3
2	G	0	0	0	0	0	0	1	2	3
3	G		0	0	0	0	0	1	2	2
4	A			0	0	0	0	1	1	1
5	A				0	0	0	1	1	1
6	A					0	0	1	1	1
7	T						0	0	0	0
8	C							0	0	0
9	C								0	0

Алгоритм Нуссинов

Процедура обратного прохода



		1	2	3	4	5	6	7	8	9
		G	G	G	A	A	A	T	C	C
1	G	0	0	0	0	0	0	1	2	3
2	G	0	0	0	0	0	0	1	2	3
3	G		0	0	0	0	0	1	2	2
4	A			0	0	0	0	1	1	1
5	A				0	0	0	1	1	1
6	A					0	0	1	1	1
7	T						0	0	0	0
8	C							0	0	0
9	C								0	0

Благодарности

- При подготовке слайдов использовались материалы лекций:
 - Михаила Гельфанда (ИППИ)
 - Андрея Миронова (МГУ)
 - Serafim Batzoglou (Stanford)
 - Manolis Kellis (MIT)
 - Pavel Pevzner (UCSD)