

Задания:

1. Транскрипция. Sequence 1 перевести код ДНК в код РНК
2. Трансляция. Sequence 1 перевести в код белка
3. Посчитать встречаемость
 - а) каждой аминокислоты,
 - б) GC (почему GC - они дают три водородные связи и дают термоустойчивость ДНК, это важно для отжига праймеров для ПЦР)
4. Сравнить в двух белках по Sequence2 и Sequence3
 - а) GC состав
 - б) встречаемость кодонов для глицина (G) и аланина (A)

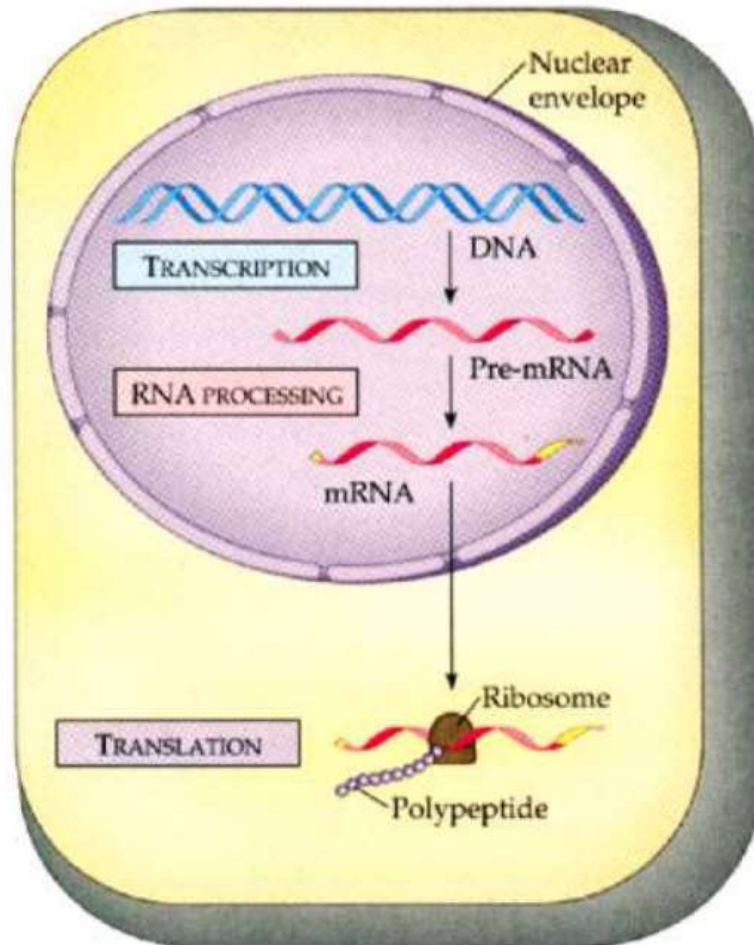
Задание со *

5. В ДНК последовательностях Sequence1_1, Sequence1_2, Sequence1_3 обнаружить неоднозначные нуклеотиды и определить варианты изменения белкового кода и вывести результирующие белковые последовательности.

Задание с **

6. В последовательности Sequence4 найти старт и стоп кодоны, определить рамку считывания, количество аминокислот в белке и вывести последовательность белка

От ДНК к белку



CCTGAGCCAACTATTGATGAA



CCUGAGCCAAACUAUUGAUGAA



белок

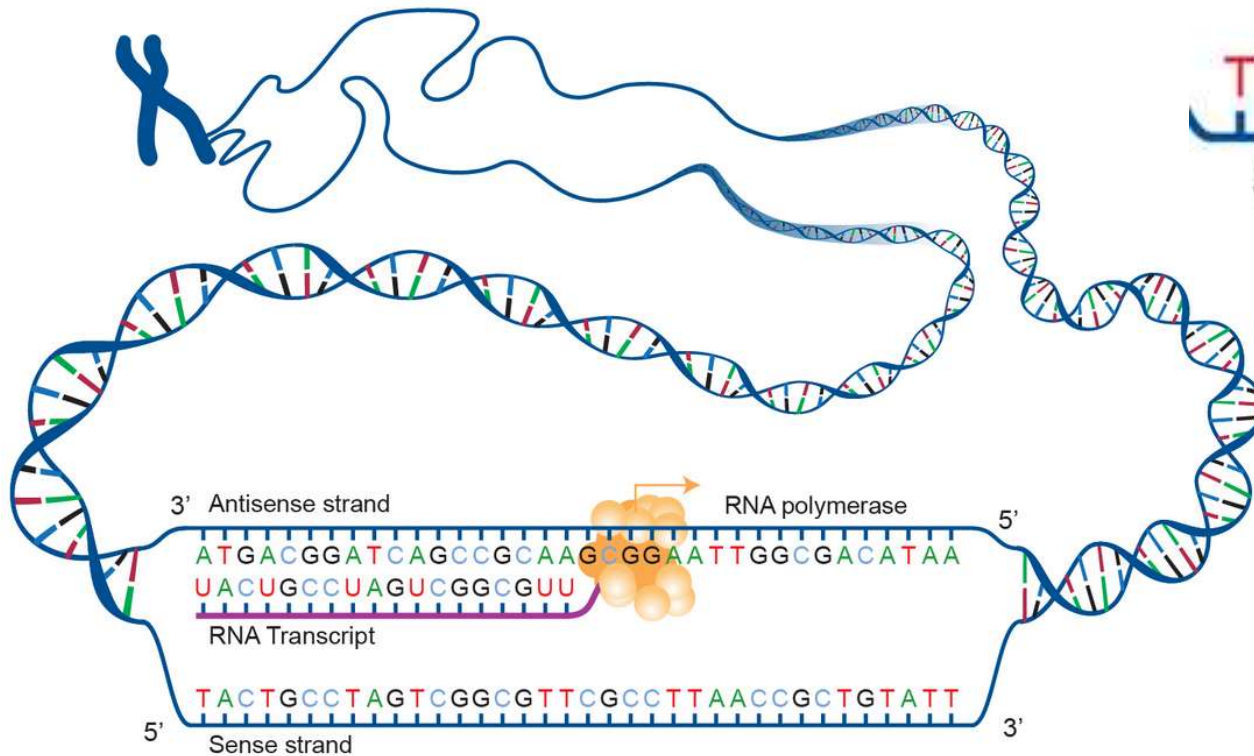
Транскрипция

U A C U G C C U A G U C G G C G U U

RNA Transcript

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

Sense strand



Библиотека

A – A

T – U

C – C

G – G

Синонимичные кодоны

TTT	F	TCT	S	TAT	Y	TGT	C
TTC	F	TCC	S	TAC	Y	TGC	C
TTA	L	TCA	S	TAA	stop	TGA	stop
TTG	L	TCG	S	TAG	stop	TGG	W
CTT	L	CCT	P	CAT	H	CGT	R
CTC	L	CCC	P	CAC	H	CGC	R
CTA	L	CCA	P	CAA	Q	CGA	R
CTG	L	CCG	P	CAG	Q	CGG	R
ATT	I	ACT	T	AAT	N	AGT	S
ATC	I	ACC	T	AAC	N	AGC	S
ATA	I	ACA	T	AAA	K	AGA	R
ATG	M/ start	ACG	T	AAG	K	AGG	R
GTT	V	GCT	A	GAT	D	GGT	G
GTC	V	GCC	A	GAC	D	GGC	G
GTA	V	GCA	A	GAA	E	GGA	G
GTG	V	GCG	A	GAG	E	GGG	G

Библиотека ДНК кодонов

```
dna_codons = {"GCT": "A", "GCC": "A", "GCA": "A", "GCG": "A",  
              "TGT": "C", "TGC": "C",  
              "GAT": "D", "GAC": "D",  
              "GAA": "E", "GAG": "E",  
              "TTT": "F", "TTC": "F",  
              "GGT": "G", "GGC": "G", "GGA": "G", "GGG": "G",  
              "CAT": "H", "CAC": "H",  
              "ATA": "I", "ATT": "I", "ATC": "I",  
              "AAA": "K", "AAG": "K",  
              "TTA": "L", "TTG": "L", "CTT": "L", "CTC": "L", "CTA": "L",  
              "CTG": "L",  
              "ATG": "M", "AAT": "N", "AAC": "N",  
              "CCT": "P", "CCC": "P", "CCA": "P", "CCG": "P",  
              "CAA": "Q", "CAG": "Q",  
              "CGT": "R", "CGC": "R", "CGA": "R", "CGG": "R", "AGA": "R",  
              "AGG": "R",  
              "TCT": "S", "TCC": "S", "TCA": "S", "TCG": "S", "AGT": "S",  
              "AGC": "S",  
              "ACT": "T", "ACC": "T", "ACA": "T", "ACG": "T",  
              "GTT": "V", "GTC": "V", "GTA": "V", "GTG": "V",  
              "TGG": "W",  
              "TAT": "Y", "TAC": "Y",  
              "TAA": "_", "TAG": "_", "TGA": "_"}}
```


От РНК к белку - трансляция

		Second base				
		U	C	A	G	
First base	U	UUU } Phenyl- alanine F UUC UUA } Leucine L UUG	UCU } UCC } Serine S UCA UCG	UAU } Tyrosine Y UAC UAA Stop codon UAG Stop codon	UGU } Cysteine C UGC UGA Stop codon UGG Tryptophan W	Third base U C A G U C A G U C A G U C A G
	C	CUU } CUC } Leucine L CUA CUG	CCU } CCC } Proline P CCA CCG	CAU } Histidine H CAC CAA } Glutamine Q CAG	CGU } CGC } Arginine R CGA CGG	
	A	AUU } Isoleucine I AUC AUA AUG Methionine start codon M	ACU } ACC } Threonine T ACA ACG	AAU } Asparagine N AAC AAA } Lysine K AAG	AGU } Serine S AGC AGA } Arginine R AGG	
	G	GUU } GUC } Valine V GUA GUG	GCU } GCC } Alanine A GCA GCG	GAU } Aspartic acid D GAC GAA } Glutamic acid E GAG	GGU } GGC } Glycine G GGA GGG	

Библиотека РНК кодонов

```
rna_codons = {"GCU":"A", "GCC":"A", "GCA":"A", "GCG":"A",  
              "UGU":"C", "UGC":"C",  
              "GAU":"D", "GAC":"D",  
              "GAA":"E", "GAG":"E",  
              "UUU":"F", "UUC":"F",  
              "GGU":"G", "GGC":"G", "GGA":"G", "GGG":"G",  
              "CAU":"H", "CAC":"H",  
              "AUA":"I", "AUU":"I", "AUC":"I",  
              "AAA":"K", "AAG":"K",  
              "UUA":"L", "UUG":"L", "CUU":"L", "CUC":"L", "CUA":"L",  
              "CUG":"L",  
              "AUG":"M", "AAU":"N", "AAC":"N",  
              "CCU":"P", "CCC":"P", "CCA":"P", "CCG":"P",  
              "CAA":"Q", "CAG":"Q",  
              "CGU":"R", "CGC":"R", "CGA":"R", "CGG":"R", "AGA":"R",  
              "AGG":"R",  
              "UCU":"S", "UCC":"S", "UCA":"S", "UCG":"S", "AGU":"S",  
              "AGC":"S",  
              "ACU":"U", "ACC":"U", "ACA":"U", "ACG":"U",  
              "GUU":"V", "GUC":"V", "GUA":"V", "GUG":"V",  
              "UGG":"W",  
              "UAU":"Y", "UAC":"Y",  
              "UAA":"_", "UAG":"_", "UGA":"_"}
```


Частота использования кодонов в экзонах

Codon Usage: The Standard Code (transl_table=1)

CCA(P) 1.77%	CGA(R) 0.64%	CAA(Q) 1.26%	CTA(L) 0.79%
CCG(P) 0.61%	CGG(R) 0.98%	CAG(Q) 3.24%	CTG(L) 3.62%
CCT(P) 1.85%	CGT(R) 0.43%	CAT(H) 1.09%	CTT(L) 1.37%
CCC(P) 1.76%	CGC(R) 0.90%	CAC(H) 1.45%	CTC(L) 1.95%
GCA(A) 1.67%	GGA(G) 1.68%	GAA(E) 2.87%	GTA(V) 0.81%
GCG(A) 0.62%	GGG(G) 1.42%	GAG(E) 3.83%	GTG(V) 2.71%
GCT(A) 2.02%	GGT(G) 1.13%	GAT(D) 2.16%	GTT(V) 1.18%
GCC(A) 2.52%	GGC(G) 2.02%	GAC(D) 2.55%	GTC(V) 1.56%
ACA(T) 1.74%	AGA(R) 1.28%	AAA(K) 2.49%	ATA(I) 0.85%
ACG(T) 0.52%	AGG(R) 1.23%	AAG(K) 3.35%	ATG(M) 2.30%
ACT(T) 1.46%	AGT(S) 1.32%	AAT(N) 1.62%	ATT(I) 1.69%
ACC(T) 1.85%	AGC(S) 1.94%	AAC(N) 2.00%	ATC(I) 2.22%
TCA(S) 1.23%	TGA(*) 0.10%	TAA(*) 0.06%	TTA(L) 0.72%
TCG(S) 0.40%	TGG(W) 1.19%	TAG(*) 0.05%	TTG(L) 1.36%
TCT(S) 1.68%	TGT(C) 1.20%	TAT(Y) 1.28%	TTT(F) 1.81%
TCC(S) 1.75%	TGC(C) 1.17%	TAC(Y) 1.55%	TTC(F) 2.09%

- частоты синонимичных кодонов
 - специфичны для генома
 - коррелируют с концентрациями тРНК

IUPAC symbols for nucleotides.

Symbol	Name	Nucleotides represented
A	Adenine	A
C	Cytosine	C
G	Guanine	G
T	Thymine	T
U	Uracil	U
K	Keto	G, T
M	Amino	A, C
R	Purine	A, G
S	Strong	C, G
W	Weak	A, T
Y	Pyrimidine	C, T
B	Not A	C, G, T
D	Not C	A, G, T
H	Not G	A, C, T
V	Not T	A, C, G
N	Any base	A, C, G, T