

# Ćwiczenie C06B

## Transkrypcja i translacja

### Transkrypcja DNA

### Translacja

### Kod genetyczny

### Struktura białek

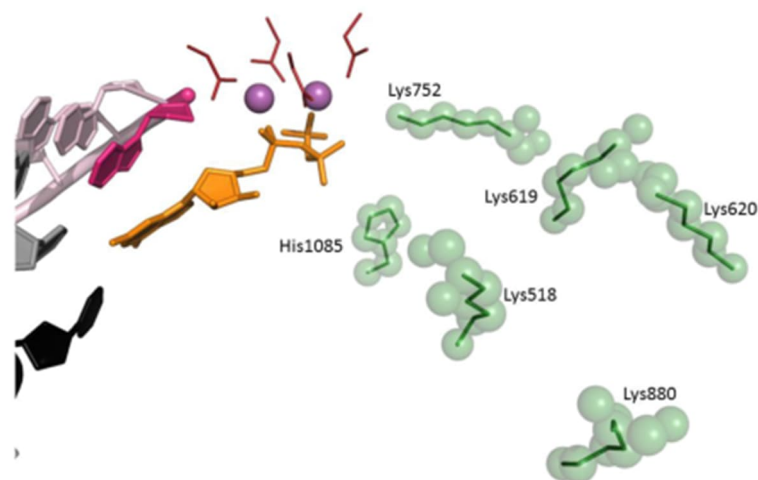
Prof. dr hab. Roman Zieliński

## 1. Transkrypcja DNA

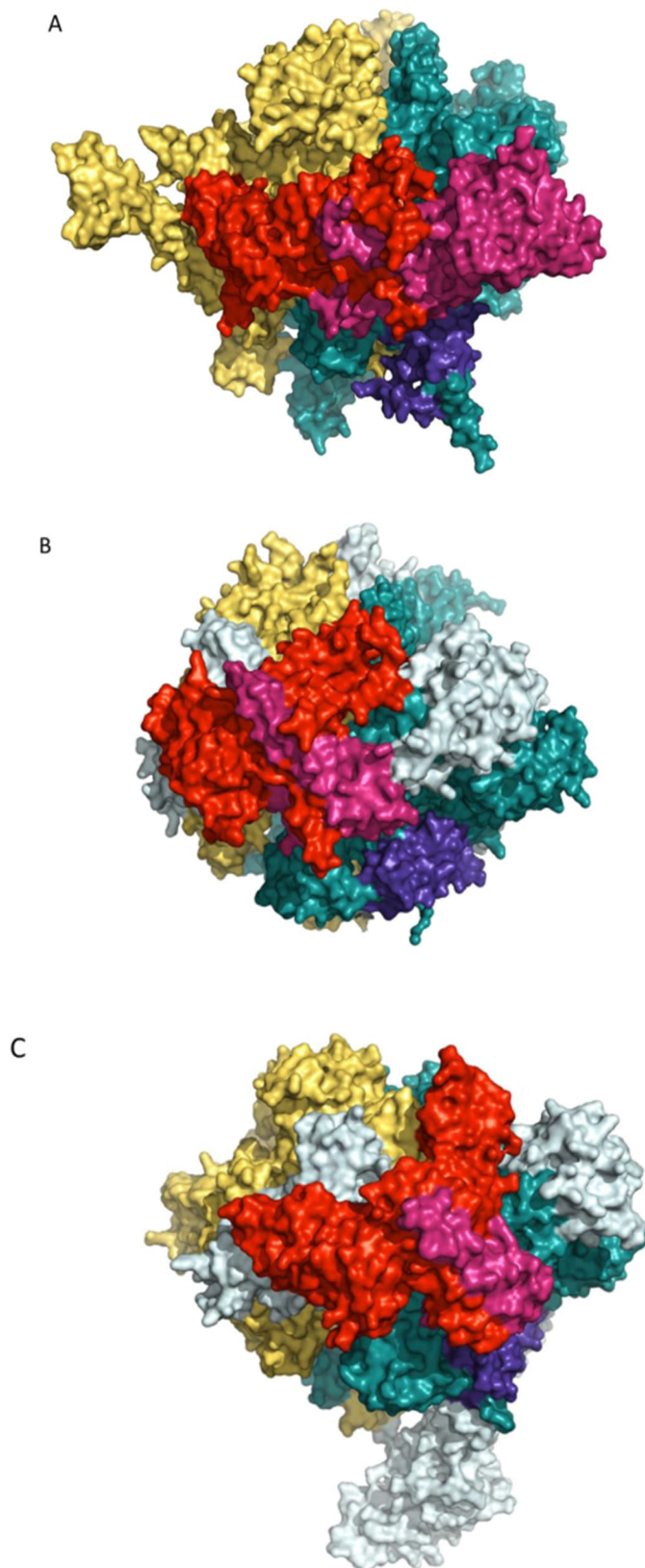
**Ekspresja genów: wykorzystywanie informacji genetycznej do syntezy funkcjonalnego produktu (białko, RNA), który ostatecznie wpływa na fenotyp. Proces ekspresji genów jest opisany przez Centralny Dogmat Biologii Molekularnej (Crick 1958, 1970), potwierdzony także odkryciem odwrotnej transkryptazy i replikacji RNA. Typowy przepływ informacji genetycznej obejmuje transkrypcję i translację.**

### ➔ 1.1. Przebieg transkrypcji

**Transkrypcja to przepisanie DNA na RNA.** Przeprowadzana jest ona przy pomocy enzymu polimerazy RNA DNA-zależnej. Synteza RNA zachodzi w kierunku 5' do 3' na matrycy o orientacji przeciwnej. Rozpoczyna się ona atakiem nukleofilowym końca 3'OH pierwszego rybonukleotydu na atom fosforu znajdujący się w reszcie trójfosforanowej w pozycji 5' rybozy drugiego rybonukleotydu (Rys. 1.1a).



**Rys. 1.1a.** Uwalnianie dwóch reszt fosforanowych podczas syntezy RNA. Aminokwasy polimerazy RNA II zabarwione na zielono ułatwiają uwolnienie reszt fosforanowych.



**Rys. 1.1b.** Polimeraza RNA u Prokariota (A), polimeraza RNA II u Eukariota (B) oraz polimeraza RNA u Archaea (C).

Transkrypcja jest inicjowana rozpoznaniem specyficznego miejsca – promotora. Promotory różnią się u Prokariota i Eukariota. Najczęściej rozpoznawaną sekwencją jest TATA. U Prokariota występuje jedna polimeraza RNA, u Eukariota występują trzy polimerazy RNA, przy czym polimeraza RNA II syntetyzuje mRNA. Polimerazy RNA są wielodomenowymi białkami (Rys. 1.1b). **Wszystkie polimerazy RNA II zawierają stałą triadę obejmującą asparaginę w pozycjach 481, 483, 485.** Polimerazy eukariotyczne i bakteryjne są ortologami.

U Prokariota polimeraza RNA może samodzielnie rozpoczynać transkrypcję, u Eukariota i Archaea niezbędne są czynniki transkrypcyjne. **Polimerazy RNA, w przeciwieństwie do polimeraz DNA, nie wymagają startera do rozpoczęcia syntezy RNA.**

Transkrypcja obejmuje inicjację, elongację i terminację. Najdłuższy jest etap elongacji. Początkowo RNA łączy się z matrycowym DNA na obszarze 8-9 bp tworząc hybrydę DNA-RNA, która odpowiada za stabilność kompleksu. Długość hybrydy DNA-RNA jest utrzymywana na stałym poziomie przez cały okres trwania elongacji. Odczytanie matrycy przez pojawiające się NTPs wymaga zerwania wiązań wodorowych między niciami DNA i utworzenia hybrydy DNA-RNA. Lokalnie zerwane wiązania wodorowe między niciami DNA tworzą segment określany jako „bąbel transkrypcyjny”.

U Prokariota geny są ciągłe, dlatego otrzymany transkrypt od razu może być matrycą do translacji. U Eukariota pierwotny transkrypt zawiera introny, które są wycinane w procesie splicingu. Dojrzały mRNA jest transportowany do cytoplazmy gdzie odbywa się translacja.



### 1.1.1. Przebieg transkrypcji

Proszę obejrzeć animację transkrypcji: <https://www.youtube.com/watch?v=WsofH466lqk>. Na podstawie przedstawionej animacji proszę odpowiedzieć na pytania.

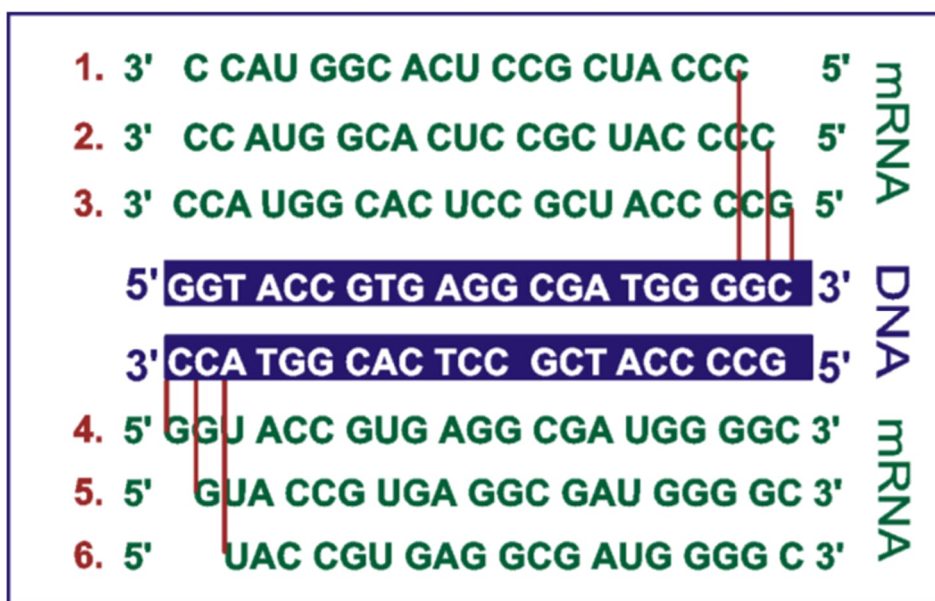
- Wymień wszystkie czynniki zaangażowane w proces transkrypcji.
- Gdzie zlokalizowana jest sekwencja TATA?
- Który element łączy się z sekwencją TATA? I jaką pełni funkcję?
- Skąd pochodzi energia niezbędna do transkrypcji?
- Proszę podać grupy organizmów, u których transkrypcja przebiega zgodnie ze schematem przedstawionym na filmie. Proszę uzasadnić.



### 1.2. Otwarta ramka odczytu, ORF

**ORF (ang. Open Reading Frame): segment DNA kodujący polipeptyd. Na poziomie mRNA, ORF rozpoczyna się kodonem start, najczęściej AUG, a kończy kodonem STOP.**

ORF identyfikuje się na podstawie sekwencjonowania genomów, gdy nie są znane białka kodowane przez dany fragment DNA. Wówczas sekwencję mRNA i następnie białek przewiduje się na podstawie sekwencji DNA. Każdy region DNA może być matrycą dla co najmniej 6 typów mRNA



Rys. 1.2. Możliwe ORF dla danego odcinka DNA.

różniących się początkiem odczytu. Sekwencje DNA w bazach danych podawane są dla orientacji 5'→3'. Aby podać wszystkie ramki odczytu należy „dopisać” drugą nić.

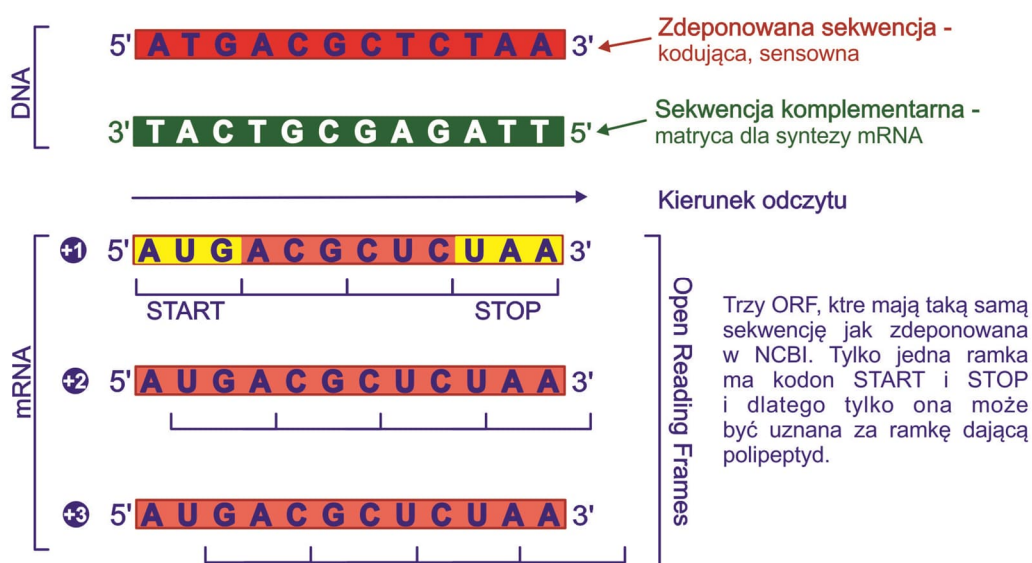


1.2.1. Dana jest sekwencja DNA: TGG TTA CAA ATT GAG zdeponowana w NCBI. Proszę podać wszystkie możliwe ORF dla tej sekwencji.

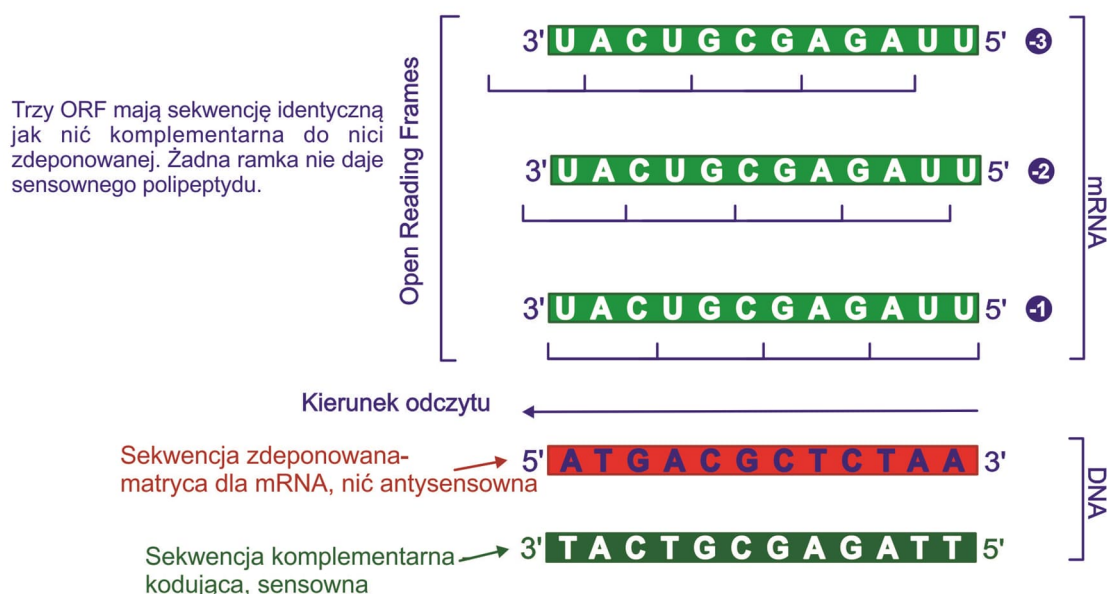
## 1.2.2. Wykorzystanie programu ORFfinder

Odczytując ORF wyznaczone przy pomocy programu istotna jest prawidłowa interpretacja nici sensownej i antysensownej oraz odpowiadających im ramek odczytu (Rys. 1.2.2a). Ponieważ kodony są trójnukleotydowe, każda nić DNA może dać trzy potencjalne ramki odczytu (ORF) w zależności od miejsca, w którym rozpoczynany jest odczyt. Przykładowo, jeżeli sekwencja zdeponowana w NCBI zaczyna się od ATG (AUG) odczyt może rozpoczynać się od A, od T lub od G. Odpowiednie ramki numeruje się jako +1, +2, +3 jeżeli odczyt dotyczy nici sensownej, oraz -1, -2 i -3 jeżeli odczyt dotyczy nici komplementarnej do nici zdeponowanej w NCBI (antysensownej).

## A. Sekwencja zdeponowana w NCBI jest sekwencją kodującą (sensowną)



## B. Sekwencja zdeponowana jest nicią antysensowną (niekodującą)



**Rys. 1.2.2.a.** Konwencja przyjęta dla odczytu i numerowania ORF w analizach bioinformatycznych i bazach danych.



W praktyce ORF odczytuje się za pomocą programów dostępnych on line, np. ORFfinder (Rys. 1.2.2b) na stronach NCBI (<https://www.ncbi.nlm.nih.gov/orffinder/>). Program pozwala przewidywać ramki odczytu dla sekwencji do 50 kbp. Na ogół za właściwą ramkę odczytu przyjmuje się tę ramkę, która wyznacza najdłuższy polipeptyd.

Wśród parametrów można wybrać:

- minimalną długość ORF (zaleca się min. 100 bp);
- kod genetyczny, jest on identyczny u większości organizmów, czasami występują niewielkie odstępstwa dotyczące częstości wykorzystania określonych kodonów;
- kodon START. Na ogół kodonem START jest AUG (ATG dla DNA), który odpowiada metioninie u Eukariota.


**Open Reading Frame Finder**

ORF finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF finder to search newly sequenced DNA for potential protein encoding segments, verify predicted protein using newly developed SMART BLAST or regular BLASTP.

This web version of the ORF finder is limited to the subrange of the query sequence up to 50 kb long. Stand-alone version, which doesn't have query sequence length limitation, is available for [Linux x64](#).

**Examples** (click to set values, then click Submit button) :

- NC\_011604 Salmonella enterica plasmid pWES-1; genetic code: 11; 'ATG' and alternative initiation codons; minimal ORF length: 300 nt
- NM\_000059; genetic code: 1; start codon: 'ATG only'; minimal ORF length: 150 nt




Enter Query Sequence

Enter accession number, gi, or nucleotide sequence in FASTA format:

From:  To:

Miejsce wstawienia sekwencji lub numeru akcesyjnego.



Rys. 1.2.2b. Zrzut ekranu dla strony startowej ORFfinder.



### 1.2.3. Wyznaczanie ORF dla wybranej sekwencji przy pomocy ORFfinder

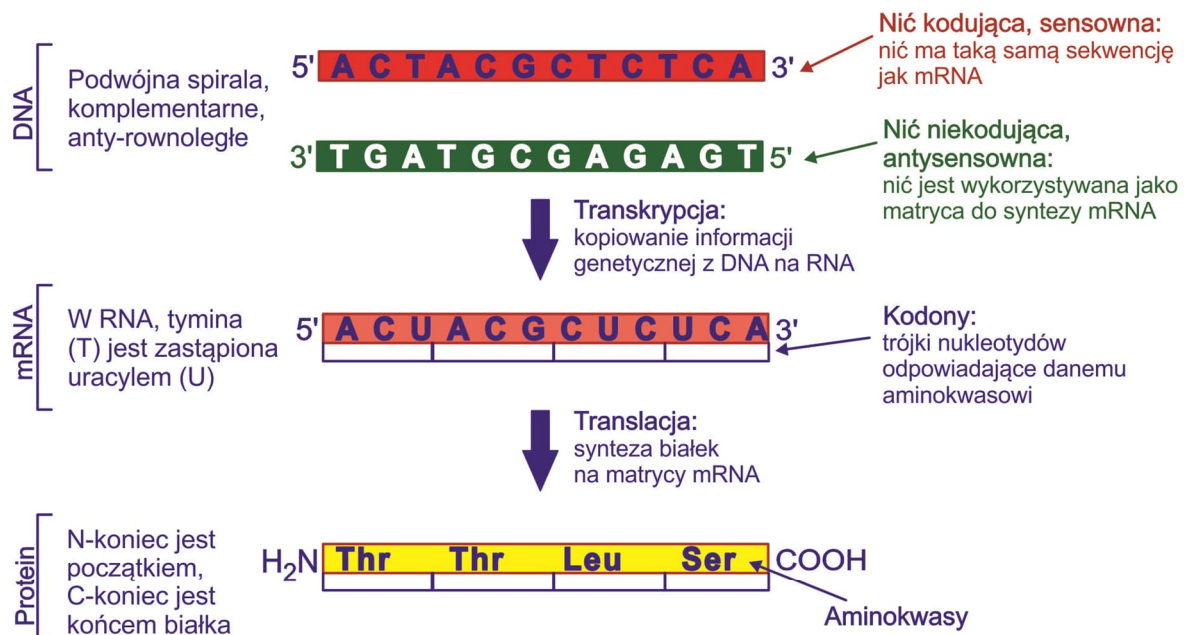
Numerem akcesyjnym X60703.1 oznaczono sekwencję genu peroksydazy aktywnej w tarczycy myszy (*Mus musculus*). Proszę dla tej sekwencji wyznaczyć wszystkie możliwe ORF przy założeniu, że najkrótszy ORF wynosi 150 bp.

- Ile możliwych ORF wyznaczył program, ile wyznaczył dla nici sensownej, a ile dla nici antysensownej?
- Który ORF jest najprawdopodobniej właściwym polipeptydem? Której nici dotyczy ten ORF?
- Proszę podać pozycję kodonu START i STOP dla sekwencji z punktu B oraz długość białka otrzymanego na bazie tej sekwencji.
- Seqwencję polipeptydu otrzymaną na bazie najdłuższego ORF proszę skopiować do notatnika i zapisać ten plik jako „*Mus musculus ORFX*”, gdzie X oznacza numer ORF z punktu B.

## ➔ 2. Translacja

W procesie zwanym translacją informacja zawarta w mRNA zostaje przepisana na sekwencję aminokwasów w białku (Rys. 2). Translacja obejmuje inicjację, elongację i terminację.

- Inicjacja prowadzi do złożenia jednostki rybosomowej 80S i jej pozycjonowania w pobliżu kodonu START mRNA. Kluczowym elementem translacji jest utworzenie kompleksu preinicjacyjnego 43S (PIC) oraz jego połączenie z końcem 5' mRNA. Inicjacja jest etapem limitującym efektywność translacji. Jest ona uwarunkowana około 25 czynnikami, które są istotne w przebiegu procesu. Pozostałe etapy wymagają zaledwie kilka czynników. Warunkiem rozpoczęcia translacji jest obecność wolnych podjednostek rybosomów.
- Elongacja prowadzi do wydłużania łańcucha polipeptydowego.
- Terminacja to uwolnienie utworzonego polipeptydu, gdy rybosom dojdzie do kodonu STOP.



```

1 actacgctct catcccaaac cc
61 tctettctec acaatcetta cc
121 ggtgctcttc gaaatcaaga ca
181 cgacgccgac tgetgtaccg ac
241 caactccctc accatettca cc
301 agaacttgccg tacctagaaa cc

```

Fragment sekwencji nukleotydowej w banku genów (NCBI). Domyślnie jest to sekwencja napisana od 5' do 3'. Każda linia zawiera 60 nukleotydów, po sześć dziesiątek w linii.

```

1 mmdfklfslt llfstiltpa l
61 cvecdptthr insltiftdn n
121 lkmlrlswng legsvpdfis q
181 gqipessfgkf vgtvpalfle h
241 tqivdlsernm lefdlsvvfv s
301 kipvggkqls ldttisyfhnr c

```

Fragment sekwencji aminokwasowej w formacie GenBank (NCBI). Domyślnie sekwencja jest zapisana od N-końca do C-końca. Każda linia zawiera 60 aminokwasów, po sześć dziesiątek w linii.

**Rys. 2.** Transkrypcja i translacja oraz sposób odczytu sekwencji zdeponowanych w bazach danych. Białka w bazach danych zawsze podawane są od końca N, czyli aminokwasu z wolną grupą NH<sub>2</sub> do końca C, czyli aminokwasu z wolną grupą COOH.



## 2.1. Przebieg translacji

Proszę obejrzeć animację translacji (<https://www.youtube.com/watch?v=5bLEDd-PSTQ>) oraz odpowiedzieć na następujące pytania.

- A. Proszę wymienić składniki niezbędne do translacji.
- B. Kiedy rozpoczyna się translacja?
- C. Jaki kodon najczęściej pełni funkcję kodonu START?
- D. Kiedy zostają złożone dwie podjednostki rybosomu?
- E. Jak nazywają się miejsca, które powstają po złożeniu rybosomu, i jakie cząsteczki się w nich znajdują?
- F. Co dzieje się z tworzonym łańcuchem polipeptydowym w trakcie elongacji?
- G. Co się dzieje, gdy pojawi się kodon STOP?

## 3. Kod genetyczny



**Kod genetyczny to sposób przepisania informacji zawartej w materiale genetycznym (DNA, RNA) na sekwencję aminokwasów w białku. Kod genetyczny jest trójkowy, uniwersalny, niezachodzący, zdegenerowany, bezprzecinkowy. Kodon to trójki nukleotydów w mRNA, antykodon to komplementarne trójki nukleotydów w tRNA.**

Degeneracja kodu genetycznego polega na kodowaniu jednego aminokwasu przez więcej kodonów. Cztery nukleotydy dają 64 kombinacje trójek, czyli są 64 kodony. Kodują one 20 aminokwasów, co sprawia, że jednemu aminokwasowi może być przyporządkowane więcej kodonów. Trzy kodony, UAA, UAG, UGA oznaczają sygnał terminacji translacji.

Kod genetyczny to reguła, szyfr. Jest to sposób przetłumaczenia języka nukleotydów na język aminokwasów. **Kodem genetycznym nie jest sekwencja nukleotydowa w DNA.**



### 3.1. Analiza twierdzeń dotyczących kodu genetycznego

Poniżej przedstawiono kilka cytatów opisujących kod genetyczny. Proszę ustosunkować się do podanych informacji pod kątem ich poprawności.

- **Cytat 1:** *Kod genetyczny człowieka określany jest mianem zdegenerowanego, ponieważ większość aminokwasów może być określana przez więcej niż jeden kodon (Medicover: Genetyka, podstawowe pojęcia).*
  - **Cytat 2:** *Kod genetyczny: współzależność między sekwencją zasad w DNA (lub w mRNA stanowiącym jego transkrypt) a sekwencją aminokwasów w białku (Encyklopedia PWN).*
  - **Cytat 3:** *Kod genetyczny wirusa AH1N1 został złamany. Niedługo szczepionka (Rzeczpospolita 2015).*
- A. Który z powyższych cytatów przedstawia prawidłową definicję kodu genetycznego?
  - B. Proszę przeprowadzić analizę twierdzeń zawartych w każdym cytacie pod kątem ich poprawności.

Table 3.1 Standard genetic code									
1 <sup>st</sup> base	2 <sup>nd</sup> base								3 <sup>rd</sup> base
	U		C		A		G		
U	UUU	Phenylalanine, Phe/F <sup>1</sup>	UCU	Serine, Ser/S	UAU	Tyrosine, Tyr/Y	UGU	Cysteine, Cys/C	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine, Leu/L	UCA		UAA	STOP - Ochre	UGA	STOP - Opal	A
	UUG		UCG		UAG	STOP - Amber	UGG	Tryptophan, Trp/W	G
C	CUU	Leucine Leu/L	CCU	Proline, Pro/P	CAU	Histidine, His/H	CGU	Arginine, Arg/R	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine, Gln/Q	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine Ile/I,	ACU	Threonine, Thr/T	AAU	Asparagine, Asn/N	AGU	Serine, Ser/S	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine, Lys/K	AGA	Arginine, Arg/R	A
	AUG <sup>2</sup>	Methionine, Met/M	ACG		AAG		AGG		G
G	GUU	Valine, Val/V	GCU	Alanine, Ala/A	GAU	Aspartic acid, Asp/D	GGU	Glycine, Gly/G	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamic acid, Glu/E	GGA		A
	GUG		GCG		GAG		GGG		G

<sup>1</sup>PHE/F - abbreviations.

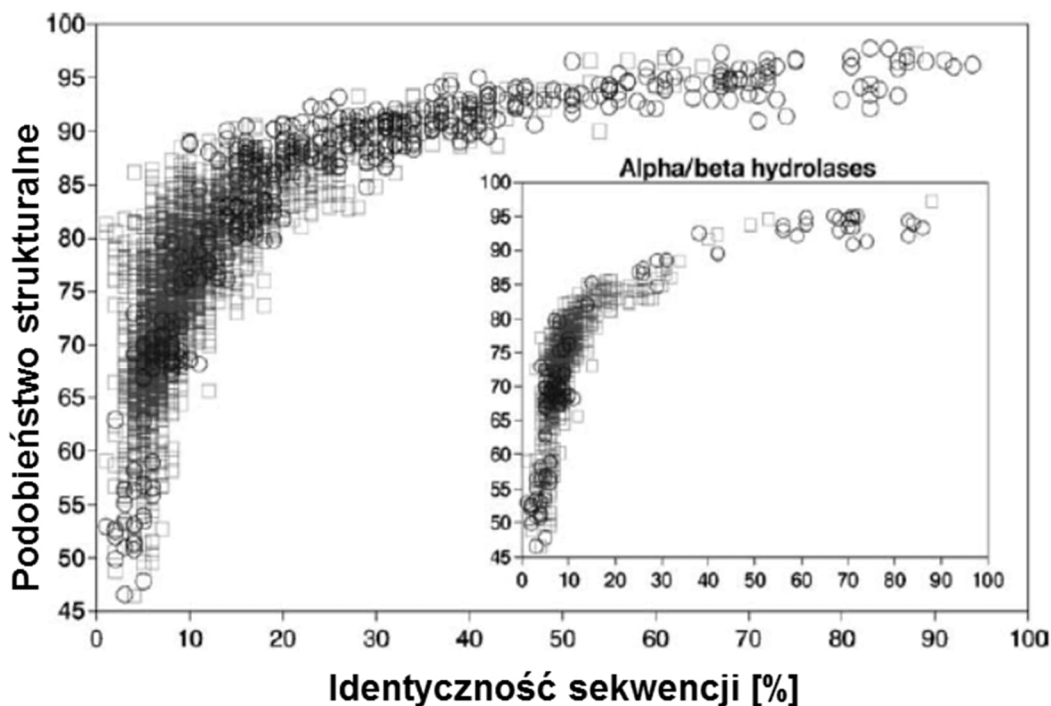
<sup>2</sup>The codon AUG has two functions: it codes for both methionine and an initiation site.

Nonpolar Polar Basic Acidic

## ➔ 4. Struktura białek

Nazwa „protein” wywodzi się z greckiego słowa „*proteios*” oznaczającego składnik występujący we wszystkich żywych komórkach. W komórkach bakterii, *Escherichia coli* zidentyfikowano około 3 000 różnych białek. U człowieka w 1 ml krwi występuje średnio 108 różnych białek. Białka klasyfikowane są w rodziny na podstawie podobieństwa sekwencji aminokwasowej (Rys. 4). Białka pochodzące od wspólnego przodka to homologi, które dzielimy na ortologi i paralogi (powstałe w wyniku duplikacji). Sekwencja białek klasyfikowanych w obrębie jednej rodziny jest identyczna w >30%, przy czym wartości >40% oznaczają bliskie podobieństwo. Białka są składnikami budulcowymi komórek, katalizują reakcje chemiczne, uczestniczą w przekazywaniu sygnału oraz mogą być źródłem energii. Najogólniej wyróżnia się białka fibrylarne, membranowe i globularne. Białka stanowią złożone makromolekuły, często zbudowane z kilku łańcuchów polipeptydowych. Struktura białek obejmuje poziom pierwszorzędowy, drugorzędowy, trzeciorzędowy i czwartorzędowy.





**Rys. 4.** Korelacja między podobieństwem sekwencji aminokwasowej a podobieństwem strukturalnym dla białek w bazie CATH. Im wyższe podobieństwo sekwencji tym bardziej białka są podobne strukturalnie. Białka o podobieństwie sekwencji 30% wykazują 90% podobieństwo strukturalne.

#### ➔ 4.1. Struktura pierwszorzędowa

**Struktura pierwszorzędowa białka to sekwencja aminokwasów. Jest ona określona sekwencją nukleotydów w genie.**

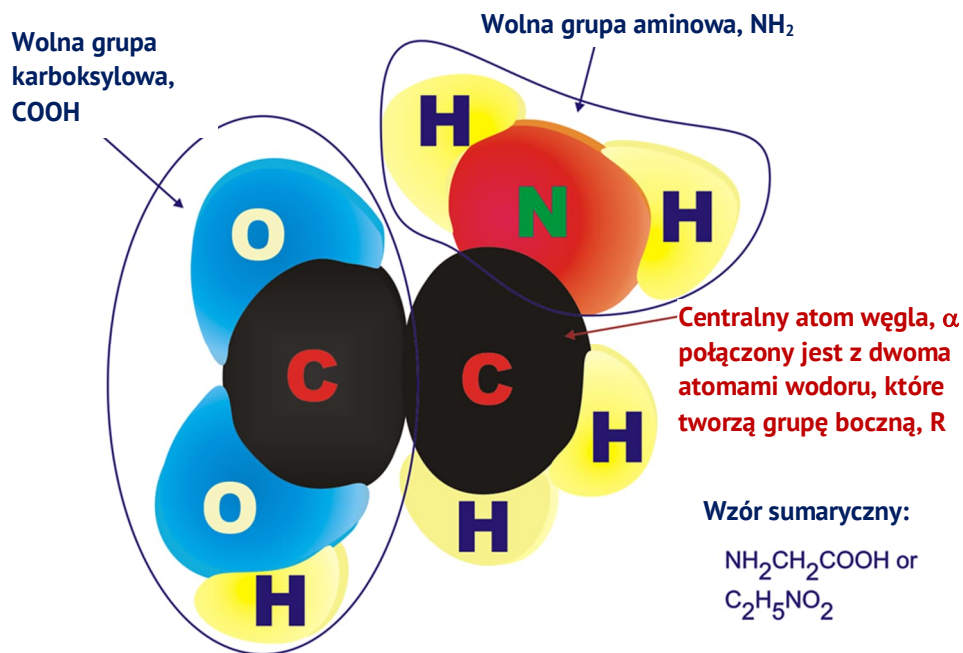
Aminokwasy stanowią element budulcowy białek. Każdy aminokwas jest zbudowany z grupy aminowej,  $\text{NH}_2$ , grupy karboksylowej,  $\text{COOH}$ , centralnego atomu węgla ( $\alpha\text{-C}$ ) oraz grup bocznych (R), które decydują o różnorodności i właściwościach aminokwasów (Rys. 4.1). Przykładowo, seryna i treonina zawierają grupę hydroksylową (OH), natomiast kwas glutaminowy i asparaginowy dodatkową grupę karboksylową ( $\text{COOH}$ ). Kwas glutaminowy i asparaginowy są często mylone z glutaminą i asparaginą, które stanowią amidy. Aminokwasy łączą się ze sobą tworząc wiązanie peptydowe.



##### 4.1.1. Struktura pierwszorzędowa wybranej sekwencji białkowej

Dla sekwencji z zadania 1.2.3 punkt D (*Mus musculus ORFX*) proszę podać trzy pierwsze i trzy ostatnie aminokwasy.

- A. Korzystając z tabeli kodu genetycznego, proszę podać pełną nazwę oraz skróty jednoliterowe i trzyliterowe dla każdego aminokwasu.
- B. Proszę podać wszystkie kodony dla aminokwasów z punktu A.



**Rys. 4.1.** Glicyna, najmniejszy ze wszystkich aminokwasów. Grupę boczną stanowią dwa atomy wodoru.

## ➔ 4.2. Struktura drugorzędowa

**Struktura drugorzędowa dotyczy przestrzennej organizacji aminokwasów w motywy takie jak: α-helisy, β-kartki i zwoje (coils).**

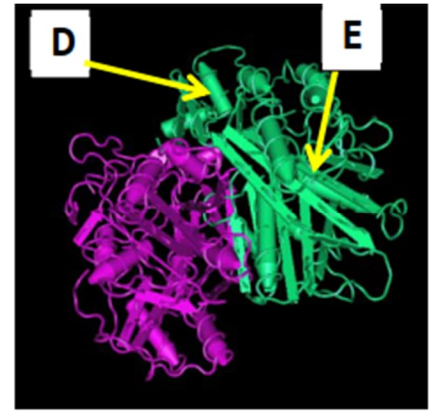
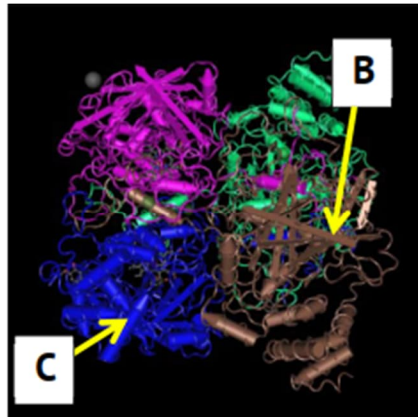
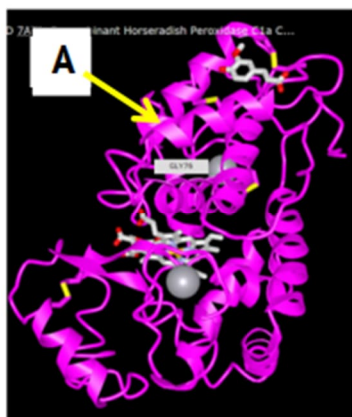
Tworzenie α-helis i β-kartek jest powszechne, ponieważ maksymalizują one powstawanie stabilizujących wiązań wodorowych.

- α-helisy są utworzone z 4-10 aminokwasów, tworzą cylindry, są prawoskrętne z 3,6 aminokwasów na skręt.
- β-kartki są złożone z 5-10 aminokwasów, które tworzą charakterystyczny „zig-zag”.



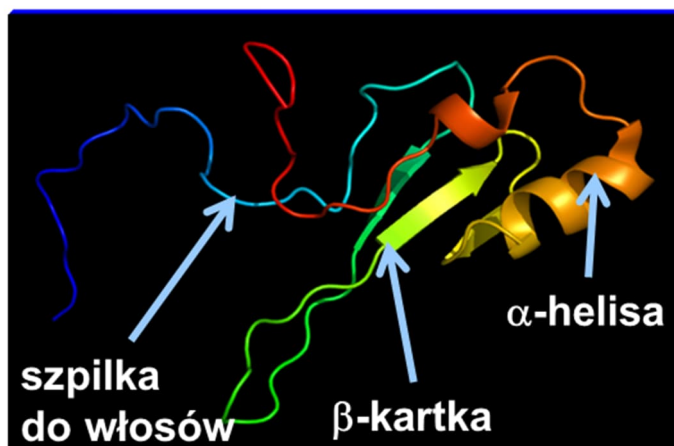
### 4.2.1. Identyfikacja struktur drugorzędowych w modelach przestrzennych

Na rysunkach poniżej przedstawiono strukturę przestrzenną trzech białek. Które litery oznaczają α-helisy, a które β-kartki?



### ➔ 4.3. Struktura trzeciorzędowa i czwartorzędowa

Struktury trzecio- i czwartorzędowa odnoszą się do zwijania się białka w przestrzeni trójwymiarowej. **Struktura trzeciorzędowa opisuje jak poszczególne motywy łączą się ze sobą.** Utrzymywana jest ona przez szereg słabych wiązań niekowalencyjnych, za wyjątkiem kowalencyjnego wiązania dwusiarczkowego (S-S). Białka składają się z kilku segmentów  $\alpha$ -helis i  $\beta$ -kartek oddzielonych pętlami. Pętle są wrażliwe na trawienie enzymami proteolitycznymi. Białka mogą być klasyfikowane na



**Rys. 4.3.** Domena MBD wiążąca się z sekwencjami CpG (zmetylowanymi).

podstawie przeważających domen. Domeny są podstawową jednostką strukturalną polipeptydów, zawierają miejsce aktywne i pełnią specyficzną funkcję (Rys. 4.3). Białko może składać się z domeny jednego typu lub z różnych domen. Domena może być powtórzona kilka razy. **Struktura czwartorzędowa opisuje łączenie dwóch lub więcej polipeptydów w multimeryczne białko.**

### ➔ 4.4. Właściwości fizyko-chemiczne białek

Generalnie, właściwości fizyko-chemiczne białek są podobne – wszystkie białka są bezbarwne, bezsmakowe, mają duży ciężar właściwy oraz tworzą koloidy. Jednakże skład aminokwasowy wpływa na ładunek białka, rozpuszczalność, zdolność do tworzenia soli i tym warunkuje różnice między białkami. Rozkład aminokwasów hydrofobowych i hydrofilnych na powierzchni białka oraz ładunek decydują o jego rozpuszczalności. Białka klasyfikowane są do poszczególnych grup, rodzin na podstawie właściwości fizyko-chemicznych. Wysoka zawartość lizyny i argininy oraz niski ciężar właściwy wskazują na histony. Na podstawie właściwości fizyko-chemicznych białka dzieli się na globularne i fibrylarne. Właściwości białka można określać przy pomocy programów dostępnych on line, jak np. ProtParam, PepCaI, ComputePI/MW (tabela 7.1).

- **Masa cząsteczkowa (MW):** określana jest w Daltonach (Da). Masa 24 500 Da oznacza, że jeden mol substancji waży 24 500 g. Masę cząsteczkową można obliczyć na podstawie wzoru chemicznego i mas atomowych.
- **Ładunek: sumaryczny:** np. 16 dodatnich ładunków i 15 ujemnych daje ładunek +1.
- **Hydrofobowość/ hydrofilność:** powinowactwo do wody. Cząsteczki hydrofilne są polarne, a hydrofobowe są niepolarne. Większość aminokwasów jest hydrofilna, jedynie siedem aminokwasów jest hydrofobowych: alanina, cysteina, leucyna, izoleucyna, metionina, fenyloalanina i walina. Aminokwasy hydrofobowe najczęściej występują w regionach transmembranowych lub w rdzeniu białka. Znając rozkład aminokwasów hydrofobowych można przewidzieć regiony transmembranowe i aktywne.
- **Czas półtrwania:** okres, w którym połowa ilości białka w komórce ulega degradacji.
- **Punkt izoelektryczny:** jest to pH, w którym cząsteczka nie ma ładunku. Ładunki pozytywne i negatywne danej cząstki równoważą się. W  $\text{pH} < \text{pI}$  cząsteczka pobiera protony ( $\text{H}^+$ ), ma ładunek dodatni i zachowuje się jak zasada, w  $\text{pH} > \text{pI}$  cząsteczka oddaje protony ( $\text{H}^+$ ), ma ładunek ujemny i zachowuje się jak kwas.

Table 7.1 Physico-chemical parameters of rPGIP1 and AJ620336.1 proteins estimated by different prediction programs			
Parameter	Program	rPGIP1 protein	AJ620336.1 protein
Chemical formula	IsotopIdent ProtParam	C <sub>1115</sub> H <sub>1782</sub> N <sub>298</sub> O <sub>326</sub> S <sub>4</sub>	C <sub>1652</sub> H <sub>2615</sub> N <sub>431</sub> O <sub>484</sub> S <sub>14</sub>
Number of amino acids [n]	Compute pI/MW ProtParam PepCalc	224	331
Molecular weight, MW [Da]: a - average/ m - monoisotopic	Compute pI/MW GPMW lite IsotopIdent ProtParam PepCalc KP <sup>1</sup>	24 706.45a/ 24 691.09m 24 705.47a/ 24 690.08m 24 691.09a/ 24 691.09m 24 706.40a 24 706.52a 24 706.16a	36 707.37a/ 36 683.93m 36 925.65a/ 36 902.00m 36 706.95a/ 36 683.94m 36 707.30a 36 707.52a 36 706.96a
Isoelectric point, [pI]	Compute pI/MW GPMW lite ProtParam PepCalc	9.24 9.64 9.24 9.65	8.04 9.12 8.04 7.80
Net charge at pH 7	ProtParam <sup>2</sup> PepCalc	16 negatively charged 20 positively charged 4.5	27 negatively charged 29 positively charged 2.2
Molar extinction (attenuation) coefficient, $\epsilon^3$ [M <sup>-1</sup> cm <sup>-1</sup> ]	GPMW lite ProtParam <sup>4</sup> PepCalc	9 530 9 970/9 970 9 530	23 950 24 450/23 350 23 470
Molar absorbance, A <sup>5</sup>	GPMW lite ProtParam <sup>4</sup>	0.39 0.404/0.404	0.65 0.666/0.652
Aliphatic index	ProtParam	104.42	98.37
Grand average hydropathicity, GRAVY <sup>6</sup>	ProtParam	-0.044	-0.012
Hydrophobicity index <sup>6</sup>	GPMW lite	-0.04	-0.02
Estimated water solubility	PepCalc PROSO SPpred	Poor water solubility	Poor water solubility
Estimated half-life	ProtParam: Reticulocutes, <i>in vitro</i> Yeast, <i>in vivo</i> , E.coli, <i>in vivo</i>	5.5 hours 3 minutes 2 minutes	30 hours >20 hours >10 hours
Instability index	ProtParam	26.46 (stable)	28.59 (stable)

<sup>1</sup>Molecular weight calculated manually by K. Pollock.

<sup>2</sup>ProtParam does not calculate the net charge. It only gives the number of positively and negatively charged residues.

<sup>3</sup>The term "molar attenuation coefficient" has been recommended since 1960s due to wrong uses of the term "extinction coefficient".

<sup>4</sup>The first value assumes that all pairs of cysteine are linked by disulfide bonds thus forming cystines. The second value considers no disulfide bonds between cysteine pairs, i.e., cysteine is reduced.

<sup>5</sup>The actual absorbance (A) is estimated for the 0.1% solution (1 g/l).

<sup>6</sup>GRAVY and Hydrophobicity index represent the same parameter.



4.4.1. Proszę wejść na stronę programu ProtParam, <https://web.expasy.org/protparam/> i przeprowadzić za pomocą tego narzędzia analizę właściwości białka z punktu 1.2.3D (*Mus musculus ORFX*). Proszę wkopiować sekwencję w ramkę (bez nazwy sekwencji). Proszę podać następujące właściwości fizyko-chemiczne:

- Liczba aminokwasów i masa cząsteczkowa.
- Wzór chemiczny.
- Punkt izoelektryczny. Czy w pH 7,0 białko jest kwasem czy zasadą?
- Który aminokwas występuje najczęściej, a który najrzadziej? Proszę uzasadnić.
- Proszę podać ładunek białka. Proszę uzasadnić.



## Odpowiedzi

### 1. Transkrypcja DNA

#### 1.1. Przebieg transkrypcji

1.1.1. Proszę obejrzeć animację transkrypcji: <https://www.youtube.com/watch?v=WsofH466lqk>. Na podstawie przedstawionej animacji proszę odpowiedzieć na pytania.

**A.** Wymień wszystkie czynniki zaangażowane w proces transkrypcji.

- DNA
- Czynniki transkrypcyjne
- Polimeraza RNA
- ATP

**B.** Gdzie zlokalizowana jest sekwencja TATA?

- W górę (upstream) od miejsca początku transkrypcji (jednostki transkrypcyjnej).

**C.** Który element łączy się z sekwencją TATA? I jaką pełni funkcję?

- TBP (transcription binding protein), element największego czynnika transkrypcyjnego, TFIID.
- Do TBP przyłączają się kolejne czynniki transkrypcyjne TFIIA i TFIIB.
- Przygotowują one DNA do przyłączenia się polimerazy RNA i tworzą dojrzały kompleks transkrypcyjny.

**D.** Skąd pochodzi energia niezbędna do transkrypcji?

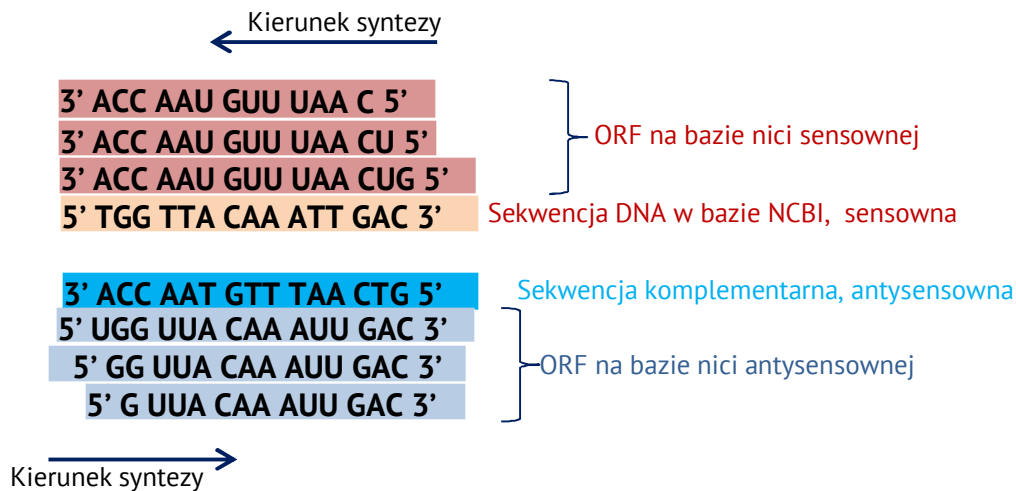
- Z redukcji ATP do ADP.

**E.** Proszę podać grupy organizmów, u których transkrypcja przebiega zgodnie ze schematem przedstawionym na filmie. Proszę uzasadnić.

- Film dotyczy transkrypcji u Eukariota i ewentualnie Archaea, gdyż w procesie uczestniczą czynniki transkrypcyjne. U Bacteria transkrypcja przebiega bez udziału czynników transkrypcyjnych.

## 1.2. Otwarta ramka odczytu, ORF

1.2.1. Dana jest sekwencja DNA: TGG TTA CAA ATT GAG zdeponowana w NCBI. Proszę podać wszystkie możliwe ORF dla tej sekwencji.



1.2.3. Wyznaczanie ORF dla wybranej sekwencji przy pomocy ORFfinder

Numerem akcesyjnym X60703.1 oznaczono sekwencję genu peroksydazy aktywnej w tarczycy myszy (*Mus musculus*). Proszę dla tej sekwencji wyznaczyć wszystkie możliwe ORF przy założeniu, że najkrótszy ORF wynosi 150 bp.

A. Ile możliwych ORF wyznaczył program, ile wyznaczył dla nici sensownej, a ile dla nici antysensownej?

- 14 wszystkich ORF
- 5 ORF dla nici sensownej
- 9 ORF dla nici antysensownej

B. Który ORF jest najprawdopodobniej właściwą ramką odczytu? Której nici dotyczy ten ORF?

- ORF2
- Dotyczy nici sensownej, czyli mRNA ma taką sekwencję jak nić zdeponowana w NCBI.

C. Proszę podać pozycję kodonu START i STOP dla sekwencji z punktu B oraz długość białka otrzymanego na bazie tej sekwencji.

- START: pozycja 74
- STOP: pozycja 2818
- 914 aminokwasów

Label	Strand	Frame	Start	Stop	Length (nt   aa)
ORF2	+	2	74	2818	2745   914
ORF13	-	3	1531	1118	414   137
ORF5	+	3	2157	2447	291   96

D. Sekwencję polipeptydu otrzymaną na bazie najdłuższego ORF proszę skopiować do notatnika i zapisać ten plik jako „*Mus musculus ORFX*”, gdzie X oznacza numer ORF z punktu B.

- Nazwa pliku „*Mus musculus ORF2*”

## 2. Translacja

### 2.1. Przebieg translacji

Proszę obejrzyć animację translacji (<https://www.youtube.com/watch?v=5bLEDd-PSTQ>) oraz odpowiedzieć na następujące pytania.

**A.** Proszę wymienić składniki niezbędne do translacji.

- mRNA
- rybosom złożony z małej i dużej podjednostki
- tRNA
- czynnik uwalniający

**B.** Kiedy rozpoczyna się translacja?

- Translacja rozpoczyna się gdy mała podjednostka rybosomu przyłącza się do czapeczki (cap) na końcu 5' i przemieszcza się wzdłuż mRNA do miejsca inicjacji.

**C.** Jaki kodon najczęściej pełni funkcję kodonu START?

- AUG dla metioniny

**D.** Kiedy zostają złożone dwie podjednostki rybosomu?

- Podczas translacji, mała podjednostka przyłącza się na początku inicjacji, duża podjednostka dopiero po przyłączeniu tRNA dla metioniny do kodonu AUG (START).

**E.** Jak nazywają się miejsca, które powstają po złożeniu rybosomu, i jakie cząsteczki się w nich znajdują?

- P: miejsce peptydylowe, zajmowane przez pierwszy tRNA.
- A: miejsce aminoacylowe, zajmowane przez kolejny tRNA.

**F.** Co dzieje się z tworzonym łańcuchem polipeptydowym w trakcie elongacji?

- Wydłużający się łańcuch polipeptydowy jest sukcesywnie przenoszony do miejsca A.

**G.** Co się dzieje, gdy pojawi się kodon STOP?

- Czynnik uwalniający łączy się z miejscem A i translacja się kończy.

### 3. Kod genetyczny

#### 3.1. Analiza twierdzeń dotyczących kodu genetycznego

Poniżej przedstawiono kilka cytatów opisujących kod genetyczny. Proszę ustosunkować się do podanych informacji pod kątem ich poprawności.

- **Cytat 1:** *Kod genetyczny człowieka określany jest mianem zdegenerowanego, ponieważ większość aminokwasów może być określana przez więcej niż jeden kodon* (Medicover: Genetyka, podstawowe pojęcia).
- **Cytat 2:** *Kod genetyczny: współzależność między sekwencją zasad w DNA (lub w mRNA stanowiącym jego transkrypt) a sekwencją aminokwasów w białku* (Encyklopedia PWN).
- **Cytat 3:** *Kod genetyczny wirusa AH1N1 został złamany. Niedługo szczepionka* (Rzeczpospolita 2015).

**A.** Który z powyższych cytatów przedstawia prawidłową definicję kodu genetycznego?

- Żadna definicja nie opisuje prawidłowo kodu genetycznego ponieważ kod genetyczny to sposób przepisania informacji zawartej w materiale genetycznym (DNA lub RNA) na sekwencję aminokwasów w białku. Kod genetyczny jest uniwersalny. Żadna z definicji nie podaje kodu w takim rozumieniu.
- Najbliższa poprawnej definicji jest definicja PWN (Cytat 2), ale odnosi się ona tylko do DNA, tymczasem kod genetyczny mówi o informacji genetycznej a ta jest zawarta w DNA lub RNA (wirusy RNA).

**B.** Proszę przeprowadzić analizę twierdzeń zawartych w każdym cytacie pod kątem ich poprawności

- Cytat 1 nie jest poprawny.
  - ▶ Błędnie opisuje zjawisko degeneracji kodu genetycznego jako cechę charakterystyczną dla człowieka (cyt. „określany jest mianem zdegenerowanego”).
  - ▶ Degeneracja jest cechą charakterystyczną kodu genetycznego wszystkich organizmów i człowiek nie jest wyjątkiem.
- Cytat 2: nie jest poprawny
  - ▶ Odnosi się tylko do sytuacji, gdy materiałem genetycznym jest DNA i informacja jest przepisywana na białko za pośrednictwem mRNA.
  - ▶ Kod genetyczny opisuje zależność między informacją zawartą w materiale genetycznym jakim jest DNA lub RNA a sekwencją aminokwasów. Retrowirusy, których materiałem genetycznym jest RNA przepisują informację genetyczną na DNA w procesie „odwrotnej transkrypcji” i dopiero później na białko za pomocą mRNA. Wirusy RNA mają informację zawartą w RNA, która natychmiast jest przepisywana na białko.
  - ▶ Nie jest prawidłowe pojęcie współzależności, gdyż oznacza ono obustronną relację (np. informacja przepływa w obie strony), tymczasem kod genetyczny przekazuje informację z kwasu nukleinowego do białka, gdy znajdzie się ona w białku nie może być powtórnie przekazana do kwasu nukleinowego (dogmat biologii molekularnej w wersji Cricka).



● Cytat 3: nie jest poprawny

- ▶ Kod genetyczny jest uniwersalny i wirus AH1N1 ma taki sam kod genetyczny jak inne wirusy i organizmy żywe.
- ▶ Kod genetyczny został złamany w latach 60-tych ubiegłego wieku, 1968 Nobel dla Nirenberga, Khorany i Holleya za złamanie kodu genetycznego;
- ▶ W artykule pomyłono pojęcie kodu genetycznego i informacji genetycznej.

## 4. Struktura białek

### 4.1. Struktura pierwszorzędowa

#### 4.1.1. Struktura pierwszorzędowa wybranej sekwencji białkowej

Dla sekwencji z zadania 1.2.3 punkt D (*Mus musculus ORF2*) proszę podać trzy pierwsze i trzy ostatnie aminokwasy.

A. Korzystając z tabeli kodu genetycznego, proszę podać pełną nazwę oraz skróty jednoliterowe i trzyliterowe dla każdego aminokwasu.

B. Proszę podać wszystkie kodony dla aminokwasów z punktu A.

● Trzy pierwsze aminokwasy

- ▶ **M**: metionina, Met, kodony: AUG.
- ▶ **R**: arginina, Arg, kodony: CGU, CGC, CGA, CGG, AGA, AGG.
- ▶ **T**: treonina, Thr, kodony: ACU, ACC, ACA, ACG.

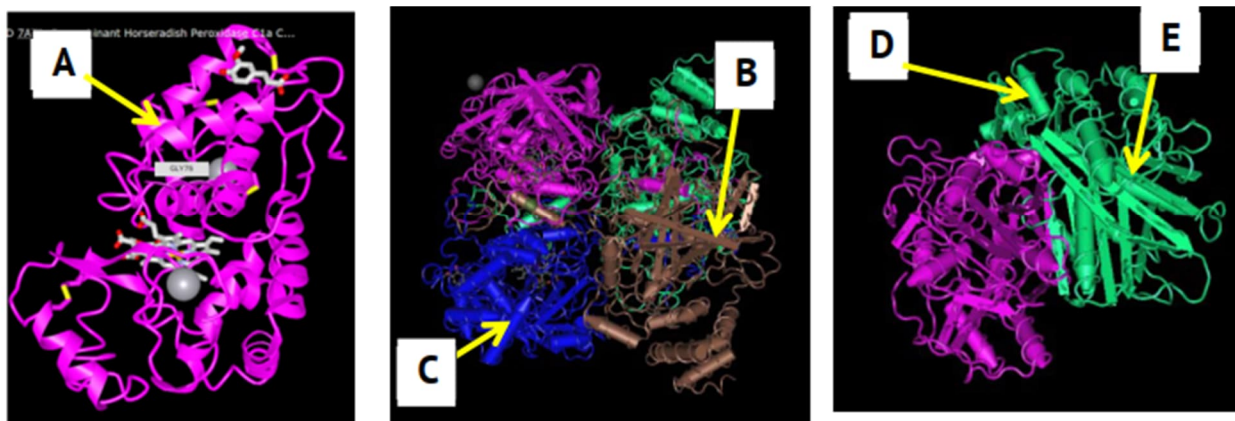
● Trzy ostatnie aminokwasy.

- ▶ **L**: leucyna, leu, kodony: UUA, UUG, CUU, CUC, CUA, CUG.
- ▶ **C**: cysteina, Cys, kodony: UGU, UGC.
- ▶ **E**: kwas glutaminowy, Glu, kodony: GAA, GAG.

### 4.2. Struktura drugorzędowa

#### 4.2.1. Identyfikacja struktur drugorzędowych w modelach przestrzennych

Na rysunkach poniżej przedstawiono strukturę przestrzenną trzech białek. Które litery oznaczają  $\alpha$ -helisy, a które  $\beta$ -kartki?



- $\alpha$ -helisa: A, C, D
- $\beta$ -kartka: B, E

#### 4.4. Właściwości fizyko-chemiczne białek

4.4.1. Proszę wejść na stronę programu ProtParam, <https://web.expasy.org/protparam/> i przeprowadzić za pomocą tego narzędzia analizę właściwości białka z punktu 1.2.3D (*Mus musculus ORF2*). Proszę wkopiować sekwencję w ramkę (bez nazwy sekwencji). Proszę podać następujące właściwości fizyko-chemiczne:

**A.** Liczba aminokwasów i masa cząsteczkowa.

- 914 aminokwasów
- Masa cząsteczkowa: 101 342,45

**B.** Wzór chemiczny.

- $C_{4483}H_{7005}N_{1267}O_{1324}S_{47}$

**C.** Punkt izoelektryczny. Czy w pH 7,0 białko jest kwasem czy zasadą?

- $pI = 6,9$
- $pI$  prawie jest równe pH 7, zatem w pH 7,0 białko nie będzie ani kwasem, ani zasadą.

**D.** Który aminokwas występuje najczęściej, a który najrzadziej? Proszę uzasadnić.

- Najczęściej występuje leucyna, 10%
- Najrzadziej występuje metionina, 1,9%.

**E.** Proszę podać ładunek białka. Proszę uzasadnić.

- Występują 94 aminokwasy z ładunkiem ujemnym i 92 z ładunkiem dodatnim.
- Łączny ładunek białka wynosi -2.