

# BIOINFORMATIKA V PRAXI – CVIČENÍ 4

## DESIGN PRIMERŮ

### PRIMERY A JEJICH VÝZNAM V MOLEKULÁRNÍ BIOLOGII

Krátké oligonukleotidové řetězce, které slouží pro provedení polymerázové řetězové reakce (PCR) označujeme jako primery. PCR má široké využití pro namnožení vybraných úseků DNA, detekci genů a dalších sekvencí, úpravu sekvencí (mutace) a řadu dalších aplikací. Návrh primerů tak musí zohledňovat účel, ke kterému budou použity.

### SEZNÁMENÍ SE S PROGRAMEM PRO ANALÝZU PRIMERŮ

Krátké oligonukleotidy pro amplifikaci DNA in vitro (primery) lze v zásadě navrhovat ručně, s výhodou však můžeme používat počítačové programy. Ty mají svou nezastupitelnou roli zejména při analýze navržené sekvence primeru. Existuje řada programů pro tento účel a to včetně volně použitelných. V tomto cvičení budete používat program **OligoAnalyzer** na serveru firmy IDT: (<http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/>)

The screenshot displays the OligoAnalyzer 3.1 web application. At the top, there is a navigation menu with options like Home, Products, Order, Support, Services, and SciTools. The main content area is titled 'OligoAnalyzer 3.1' and contains a 'Sequence' input field with a '# Bases' counter. To the right of the sequence field are several dropdown menus and input fields for 'Target Type' (set to DNA), 'Oligo Conc' (0.25 μM), 'Na+ Conc' (50 mM), 'Mg++ Conc' (0 mM), and 'dNTPs Conc' (0 mM). Below these are buttons for 'Analyze', 'Hairpin', 'Self-Dimer', 'Hetero-Dimer', 'NCBI Blast', and 'TM Mismatch'. At the bottom, there are tabs for 'Results', '5' mods', 'Internal Mods', '3' mods', and 'Mixed Bases'. The 'Mixed Bases' tab is active, showing 'Standard Mixed Base Instructions' and 'Custom Mixed Base Instructions' with a table of IUPAC symbols (R, Y, M, K, S) and their corresponding base pairs.

### ÚKOL 1

Seznamte se s aplikací **OligoAnalyzer**.

Se kterými typy sekvencí umí aplikace pracovat? DNA RNA PNA protein

Jaká je základní (default) uvažovaná koncentrace  $Mg^{2+}$  iontů a s jakými hodnotami umí program zacházet?

Jaké písmeno je užito pro označení následujících bazí (mixed base)?

A/T

A/C/G

A/C/G/T

### VÝZNAM DÉLKY PRIMERŮ A ZASTOUPENÍ BAZÍ (A+T/G+C) – $T_m$

Jednou z charakteristik primerů je jejich délka, která v kombinaci s konkrétním zastoupení bazí určuje tzv. teplotu tání ( $T_m$ , melting temperature). Ta se zvyšuje s délkou primeru a se zvyšujícím se zastoupením G+C bazí. Vyšší  $T_m$  zvyšuje specifitu primeru, pokud je však hodnota  $T_m$  příliš vysoká, nebude PCR probíhat správně. Syntéza dlouhých primerů je rovněž technicky a tudíž i finančně náročnější.

Pro typickou PCR, kde míra shodnosti sekvence primeru a cílové DNA sekvence je vysoká (100 – 90%) se jako optimální délka primerů uvádí cca 17 – 28 bází. Optimální  $T_m$  se obvykle pohybuje v rozmezí 50 – 65°C.  $T_m$  dvojice primerů by se neměla lišit o více než  $\pm 2^\circ\text{C}$ .

### OTÁZKA

Jaký základní vzorec je použit programem OligoAnalyzer pro výpočet  $T_m$  (melting temperature, teplota tání) primeru?

### ÚKOL 2

Zjistěte  $T_m$  následujících primerů (nastavení: default). Uvědomte si změnu  $T_m$  danou prodloužením sekvence a změnu danou složením (zastoupením G+C resp. A+T bazí).

#	Sekvence	Délka	G+C [%]	$T_m$
1	GAATGCTCACTCAAGGATTAC			
2	GAATGCTCACTCAAGGATTACAAT			
3	CACGGAATGCTCACTCAAGGATTAC			
4	GAGGGCTCACTCAAGGAGGAC			
5	GCCTGCTCACTCAAGGAG			

### ÚKOL 3

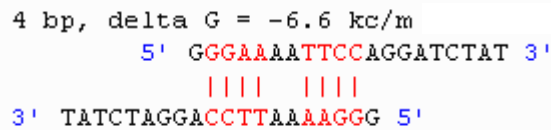
Z následujících primerů vyberte vhodnou dvojici na základě jejich  $T_m$ .

#	Sekvence	$T_m$	Vhodná dvojice s #
1	TCCAGTAATGACCTCAGAACAA		
2	AAACGACTTACTTTACTTTG		
3	TTCATCATGTGCTACGCTTACGCGTC		
4	ATGAATGCTCATCCGGAATT		
5	GCTGAATTCGGCAACGGCA		

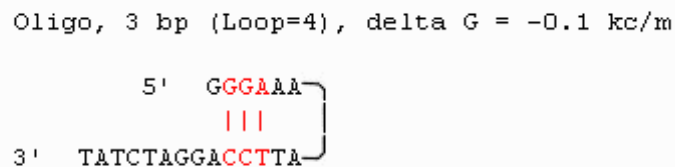
### ANALÝZA SEKVENCE PRIMERŮ – DIMERY, VLÁSENKY

I primery s vhodně zvolenou  $T_m$  mohou být nevhodné pro PCR, pokud by vytvářely nevhodné sekundární struktury. Pro správné nasednutí na cílovou DNA je třeba, aby primery existovaly jako monomerní molekuly a žádná z bazí nebyla blokována. K tomu však nedochází, pokud

primery tvoří dimery (ať již samy se sebou – homodimery, nebo s druhým primerem ve dvojici – heterodimery):



Podobně není žádoucí ani interakce uvnitř primeru, která vede ke tvorbě vlásenky:



V obou případech je důležité sledovat i konkrétní místo, kde by k těmto nežádoucím interakcím mohlo docházet. Obzvláště nevhodná je tvorba sekundárních struktur na 3' konci primeru.

#### ÚKOL 4

Odhalte možnou tvorbu dimerů u následujících dvojic primerů a označte, zda se jedná o homodimer nebo heterodimer. Zároveň uveďte sílu dané interakce, tj. volnou Gibbsovu energii dané vazby ( $\Delta G$ ). Pokud může vznikat více různých dimerů, uvádějte jen nejstabilnější z nich.

Primer 1a GATTCCACCTAAAAGCTC  
Primer 1b TAGATGGTTCACCTAACAGG

Dimer			
Typ			
$\Delta G$			

Primer 2a ACCTTTATTTTCATCGCTCTG  
Primer 2b GTGGACTAATCATGTTTCACGC

Dimer			
Typ			
$\Delta G$			

#### ÚKOL 5

U následujících primerů detekujte tvorbu vlásenek a seřadte tyto primery podle tendence k tvorbě vlásenek od nejvyšší k nejnižší.

Primer	Vlásenka (ANO – NE)	$\Delta G$	Pořadí
--------	---------------------	------------	--------

TCCAGTAATGACCTCAGAAC			
TTACCTTCCCCTCTCTTCACTT			
ATCATGTGCTACGCTTACGCGTC			
GATATTACGGCCTTTTTAAAG			
ACACTTAGAACGGTAGCAACT			

## ANALÝZA MOŽNÉHO ŠPATNÉHO NASEDNUTÍ PRIMERŮ – FALSE PRIMING SITES (FPS)

I primery, které splňují všechny výše probírané parametry, se mohou ukázat jako nevhodné pro konkrétní účel, pokud nenesedají dostatečně specificky. Přítomnost více míst, která jsou rozpoznávána primery (false priming sites) vede k tvorbě více různých PCR produktů. Význam této analýzy se liší v závislosti na účelu PCR. Pro následné klonování a tvorbu rekombinantního proteinu je žádoucí tvorba jednoho PCR produktu, akceptovatelná je tvorba několika málo PCR produktů o různé délce. Na rozdíl od předchozích úkolů je v tomto případě klíčová nejen znalost sekvence primerů a samotného cílového úseku ( genu) ale kompletní DNA přítomné v PCR směsi.

Komerční programy běžně nabízejí možnost detekce míst špatného nasednutí. U volně dostupných programů je tato funkce bohužel často opomíjena. Program **Primer-BLAST** na serveru NCBI (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) využívá propojení s databázemi a umožňuje tuto detekci provést.

### ÚKOL 7

Pro následující geny navrhnete primery tak, aby ve výsledku byl amplifikován celý gen. Ověřte možnost tvorby nežádoucích produktů při použití DNA mateřského organismu jako templátu v PCR směsi. Ke každému genu doplňte navrženou dvojici primerů, délku PCR produktu a délku dvou velikostně nejbližších možných nežádoucích produktů.

Gen 1

Organismus: Homo sapiens

Sekvence:

```
TTTACCTTCACGCTGGAGCCAAGATCGCTGCGGGGAGTCCCGTGAAGCACCCTGCCCCTAAGACCTTGAAGG
GGAAACACCAGAAGGTGTGGGTGCTGAGCTCCGCTGCGTCAGACTGCCAGGACCTGAGTGGAACCTCAGTGCTGAA
ACCTGGGTTCTCACTGCAGCTGGATAGCAGGTGGTTACAAAATTATATGTATTTTTAGGGTTATTTCTAATTTTT
CTTTAATAGTTATTGGCAATTTTCAAAGTCCTTTAAGAAACACTGAAAACCTGCTCAAGAAAATGGCAGTTCTGCA
GGGCTCAGGTAGCAGGGCAGGCTGTTGGGACAGGTTTGACAGACTGTCCATAACAAGGAGCGCCTTTTCCCAGA
TCAGCAGCTCTAGAGGGTGGCGGCCTTGCCCTGGAACCTCCGCAACCTCCGCGCCACCACACAAGGGCTGAGAAC
AGTAAGGTATGGGCTGTGCTGGGCTCAGGAGCAGGGGTAGTTGCTCACCTGGACAGGGTGTGTCTGGGAGGGCCA
TACTAGGGCCTGGACAGGGCGTCTTCTGGCCTTGAGCTGCTACGATGGGGGTAGGAGTTGGGAAGGGTGGAAAGGC
TGGTGCCCTCAGGGAGTCGGGGAGGCTGCGCAGTCAGCAGGTGAGGACTGTGTTCTCAGAGAAGGGTGGCCTGT
GCAGGCGAGGCAGGGCTGCCCCCTGCAGGGCTGAGGAACATTAGTCCATCCAGGTGGTGTGGGGAGGCTGAGGGCA
CGGTCCTCCTGCCAGAACGCAGGTGTCCTGTCTGCTACAGACACATCAAGGAAGTCACCAATGTGTGCCCTTT
CTCTAGCCTAGGGACTGGCAAACCCAAAAACACCATGCCACAGGTTAGAGCCGACCTGGTGCCCGTTTGTGAAC
AGCCACAAGCTAGGAATGGCTTTTACATGTTTAAATGGCCGAAAAGTCAAAAATATCTCATGATGTGTGAAAACCTG
TAGGTCATTCTCATCTGCGTCGGTAAGTAACGCCTTATCTGAGCACAGCCAGGCGCATTCAGGTGGTGCAGGGCCC
AGTGGAGGCTCTGTGGGCCGGGAGCCACGTCCCAGGTGACAGACAGAGCTCCCCTGCGGGCTGACCTGTGACT
TCCCCTCTCCCTGCAGCTCTGCCCGGATGGCCCTAGTCTTCGTGTACGGCACCCCTGAAGCGGGGTGAGCCCAACC
ACAGGGTCTGCGGGGACGGCGCCACGGCTCCGCAGCCTTTTCGGGCGCGGGCCGACGCTGGAGCCCTACCCGT
TGGTGATCGCGGGGGAGCACAAACATCCCCTGGCTGCTGCACCTGCCCGGCTCGGGGCGCCTCGTGGAGGGCGAGG
TCTACGCGGTAGACGAGCGGATGCTGCGCTTTCTGGATGACTTCGAGAGTTGCCCGGCCCTGTACCAGCGCACGG
TGCTGCGGGTACAGCTGCTGGAGGACCGGGCCCCGGGCGCAGAGGAGCCGCCAGCGCCACCAGCGGTGACGTGCT
TCGTGTACAGCAGGGCCACCTTCCCAGCGGAGTGGGCCCCAGCTCCCGCACCATGACAGCTACGACTCCGAGGGGC
CGCACGGGCTGCGCTACAACCCCGGGAGAACAGATAAGGGGGACGGGCAGGGTGGGCCTAGGTTTGTAGAGCCCT
```

GGGGCTCCAAGATGCGCCAGCCCATGCTGGGTGAAGGCGGAAGCCGAACAGGGCCCTTTCCAATGAATCTGCCG  
GAAAGGAACCAATCTTTTCAGTGGCAGCTGATTTTACAAATAATGTTGAGATACGAATAGCAAGGTGCTTCCCTCC  
CATCTTTCTACCTGGTAAGAAAAATTTAGGATTTTAACTCCCCTAAATGACATTTAGAGAACTCGTGTTATGCC  
AATTCCTTCTCCTCCTCGTGTTGTTTCTGCTGTTGGCTCTGCTTTGAGCTCAAGATAATAATAAATATTTAGGAT  
CAGTGTAAGACTTGGTGTGGCCGCTAGATTTTAGCAGCCCTACTATACTGATTCTGGCCTGTAACCCCTGAGAA  
AGCCGATTTTACACGGCTGGGTAGAATTTGTAGAAAAGATCCACAGGGCAAGCATGCTGTATATCAGAGTGCGTA  
TAGCACCATTCTTCTAATTTTCAGATCAAGCTTCACAGCAAATATTAAGATTATTTAAATTTGAAGTCGATGT  
TTTGGGAAATCAG

## Gen 2

Organismus: *Saccharomyces cerevisiae*

### Sekvence:

ATGATGAATAACAACGGCAACCAAGTGTGCAATCTCTCCAATGCGCTCCGTCAAGTAAACATAGGAAAACAGGAAC  
AGTAATACAACCACCGATCAAAGTAATATAAATTTTGAATTTTCAACAGGTGTAAATAATAATAATAACAAT  
AGCAGTAGTAATAACAATAATGTTCAAAAACAATAACAGCGCCGCAATGGTAGCCAAAATAATGATAACGAGAA  
AATATCAAGAATACCTTAGAACAACATCGACAACAACAACAGGCATTTTCCGATATGAGTCACGTGGAGTATTC  
AGAATTACAAAATTTTTTCAAGAACAACCCTGGAGGGATATACCCTTTTCTCTCACAGGTCTGCGCTAATGGA  
TTCAAAGTTGCTATAGTACTAAGTGAACCTGGATTTTATTATAACACAATCTTCCCTAGATTTCAATCTTGGCGAA  
CATAGGGCCCCCGAATTTGTGTCTGTGAACCCTAATGCAAGAGTTCAGCTTTAATCGATCATGGTATGGACAAC  
TTGTCTATTTGGGAATCAGGGGCGATTTTATTACATTTGGTAAATAAATATTACAAAGAGACTGGTAATCCATTA  
CTCTGGTCCGATGATTTAGCTGACCAATCACAAATCAACGCATGGTTGTTCTTCCAAACGTCAGGGCATGCGCCA  
ATGATTGGACAAGCTTTACATTTAGATACTTCCATTACAAAAGATAGCAAGTGTGTAGAAAAGATATACGGAT  
GAGGTTAGAAGAGTTTACGGTGTAGTGGAGATGGCCTTGGCTGAACGTAGAGAAGCGCTGGTGATGGAATTAGAC  
ACGGAAAATGCGGCTGCATACTCAGCTGGTACAACACCAATGTCACAAAGTCGTTTCTTTGATTATCCCGTATGG  
CTTGTAGGAGATAAATTAATATAGCAGATTTGGCCTTTGTCCCATGGAATAATGTCGTGGATAGAATTGGCATT  
AATATCAAAAATTTGAATTTCCAGAAGTTTACAAATGGACGAAGCATATGATGAGAAGACCCGCGGTATCAAGGCA  
TTGCGTGGTGAATGA

## Gen 3

Organismus: *Escherichia coli*

### Sekvence:

ATGAGTGAAGATTGTTTGAAAATGTTTACAGGTGTTGTTCTGTTAATATTTGTCATTATTGCCGGTTATTTCTTT  
TCTGAGCGTAATGACAGGAAAATGTTTCTCCTGAGCTCACTGGTTTTCTTGTATTAAATATCGCGTGTATATAT  
GTGCTGACTGCCAGATTCTGGTTTTCTGTGTGGTGCAATTATGAATCAGGGCGCAGCAATGGTTGTTCCAATGGTT  
ATTGGCTCATTACCGAACGTTACGAGCTTCGACGGGTTCAGAAGAATATTTATCTGTATTATGTTGTCATCAGTA  
TGGTCCGGAGTGATGTGGTTTTTTATAAGGGGGCTTATGACAGGCTAA

## VLASTNÍ DESIGN PRIMERŮ

### ÚKOL 8

Primer navrhujeme vždy s ohledem na smysl jeho použití. Je-li cílem detekce přítomnosti/nepřítomnosti genu, nemusíme primery navrhovat tak, aby PCR produkt obsahoval celý gen, nebo naopak může zahrnovat i část okolní DNA. Pro tuto aplikaci lze s výhodou využít on-line programy, které v rámci zadané sekvence navrhnou optimalizované primery. Úkolem vědce je mezi nabízenými alternativami vybrat tu nejvhodnější.

Pomocí programu **Primer3** (<http://frodo.wi.mit.edu/primer3>) navrhnete pro následující sekvenci vhodné primery tak, aby výsledný PCR produkt měl délku mezi 350 a 500 bp. Pro nastavení ostatních parametrů využijte znalostí z předchozích úkolů.

### Sekvence:

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAAT  
TGTGAGCGGATAACAATTTACACAGAATTCATTAAGAGGAGAAATTAATCTATGAGAGGAT

CGCATCACCATCACCATCACGGATCCCACGTGATATCCTCAATCGCTTCTAGAAGCGATTGA  
 GGAGATCTGAGCTCGGTACCCGGGTCGACAGGCCTCTGCAGCCAAGCTTAATTAGCTGAGCT  
 TGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTTCAGAACGC  
 TCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGG  
 AGCTAAGGAAGCTAAAATGGAGAAAAAATCACTGGATATAACCACCGTTGATATATCCCAAT  
 GGCATCGTAAAGAACATTTTGAGGCATTTTCAGTCAGTTGCTCAATGTACCTATAACCAGACC  
 GTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAAGTTTTATCC  
 GGCCTTTATTACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTTCGTATGGCAATGA  
 AAGACGGTGAGCTGGTGATATGGGATAGTGTTACCCTTGTTACACCGTTTTCCATGAGCAA  
 ACTGAAACGTTTTTCATCGCTCTGGAGTGAATACCACGA

Primer	Sekvence	Délka	Pozice	T <sub>m</sub>	GC [%]	Dimer	3' dimer	Vlásenka
Left								
Right								

Délka PCR produktu:

## ÚKOL 9

Na základě výše získaných znalostí navrhnete vhodnou dvojici primerů ke genu, s nímž budete pracovat v následujícím cvičení:

Genová sekvence:

GGCCAATAACGAGCCATGGGCGGCCGC

ATGGCTGATTCTCAAACGTCATCCAACCGCGCCGGCGAATTCTCGATTCCGCCGAATACCGATTTCGGCGCGATT  
 TTCTTCGCGAATGCCGCCGAGCAACAGCACATCAAATTGTTTCATCGGCGACAGCCAGGAACCCGCCGCTATCAC  
 AAGCTGACGACGCGCGACGGCCCGCGAAGCCACGCTGAATTCCGGCAACGGCAAGATCCGTTTCGAGGTGTCG  
 GTGAACGGCAAGCCGTCGGCGACCGACGCGCTCTCGCGCCGATCAACGGCAAGAAGTCGGACGGCTCGCCGTTTC  
 ACGGTCAACTTCGGGATCGTCGTGTCGGAAGACGGCCACGACAGCGACTACAACGACGGCATCGTCGTGCTCCAG  
 TGGCCGATCGGCTGA

CTCGAGGGCTCTTCCACCCGAATTTTCG

Primery by měly zahrnovat počátek a konec genu tak, aby výsledný PCR produkt obsahoval celý gen. Zkontrolujte T<sub>m</sub> obou primerů, možnou tvorbu dimerů a vlásenek, primery by měly mít délku mezi 20 a 28 bp a obsah G+C bazí by měl být mezi 40 a 60 %.

Primer	Sekvence	Délka	Pozice	T <sub>m</sub>	GC [%]	Dimer	3' dimer	Vlásenka
Left								
Right								

## SAMOSTATNÝ PROJEKT

Vaším úkolem je navrhnout vhodnou dvojici primerů ke genu, který jste identifikovali v sekvenci zadané v předchozím cvičení. Opět platí, že primery by měly zahrnovat počátek a konec genu tak, aby výsledný PCR produkt obsahoval celý gen. Zkontrolujte T<sub>m</sub> obou primerů, možnou tvorbu dimerů a vlásenek, primery by měly mít délku mezi 20 a 28 bp a obsah G+C bazí by měl být mezi 40 a 60 %.