

Palindromy, restriční enzymy, primery

C2131 Úvod do bioinformatiky

Jaro 2023

Mgr. Josef Houser, Ph.D.

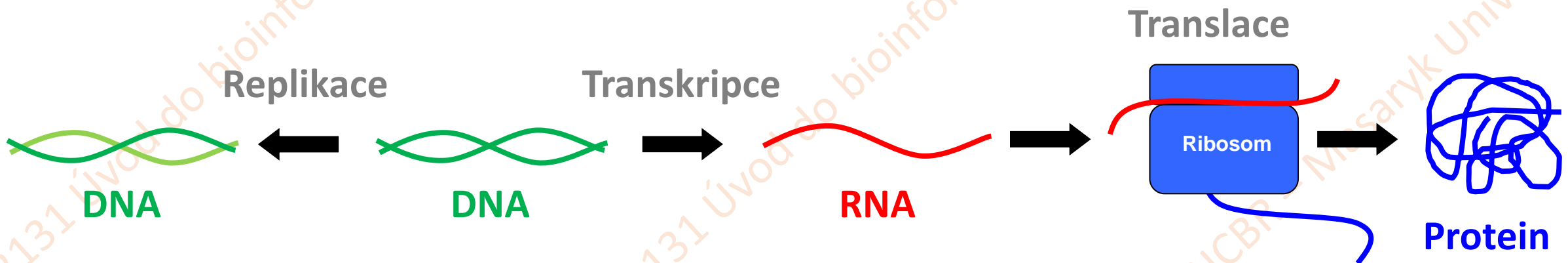
Nukleové kyseliny

- Dlouhý, nerozvětvený polymer tvořený čtyřmi typy monomerů. Informace je uložena v **pořadí (sekvenci)** monomerů.
- **DNA** – deoxyribonukleová kyselina – vlastní nosič genetické informace (skoro vždy)
- **RNA** – ribonukleová kyselina
- **Sekvence** je vždy zapisována od 5' ke 3' konci

Interakce nukleová kyselina – protein

Přenos genetické informace je vícekrokový proces

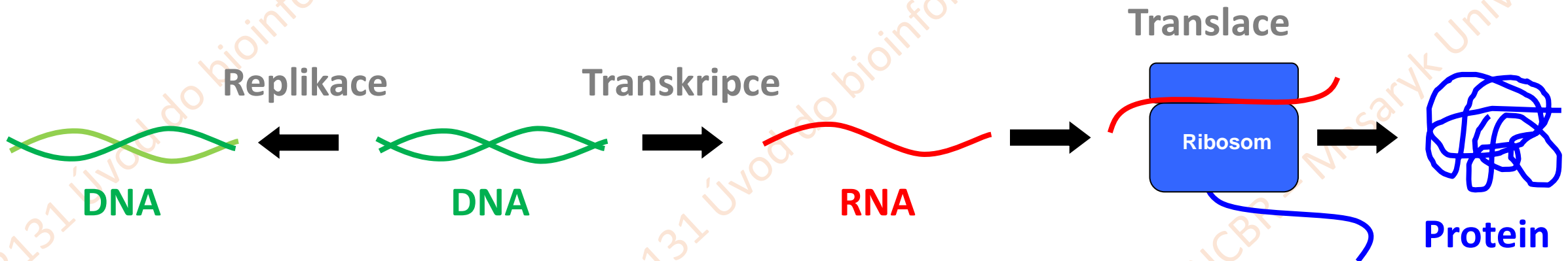
Všechny kroky probíhají za účasti proteinů



Interakce nukleová kyselina – protein

Proteiny se podílí i na dalších interakcích s nukleovými kyselinami

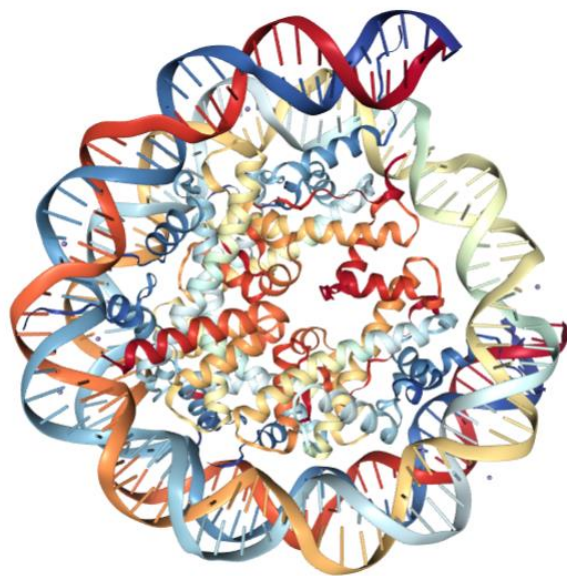
- Modifikace NK
- Regulace exprese
- Transport
- Degradace



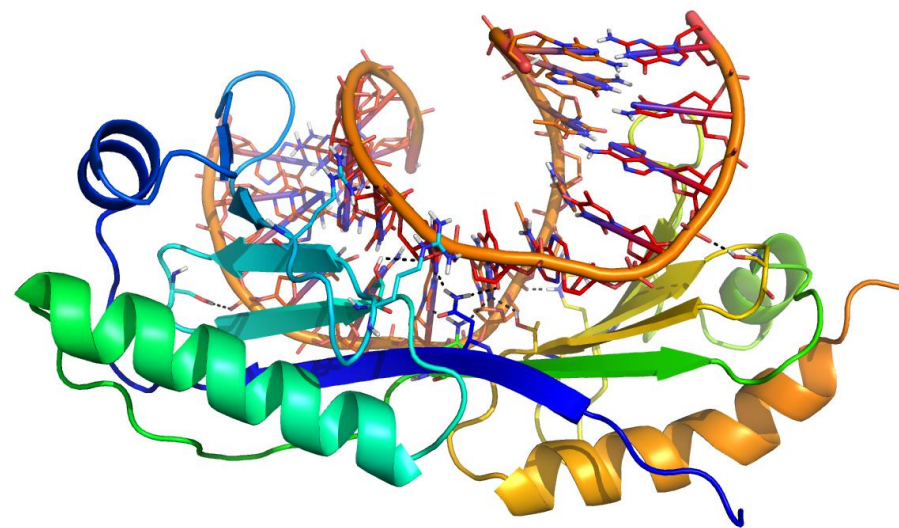
Interakce nukleová kyselina – protein

Interakce protein – nukleová kyselina může být:

- **Nespecifická** – vazba na cukr-fosfátovou kostru
- **Specifická** – zahrnující rozpoznávání bází = sekvence



Nukleosom – nespecifická interakce histonu s DNA



TATA-vázající protein – specifická interakce s DNA

NCBR – Masaryk University

C2131 Úvod do bioinformatiky

C2131 Úvod do bioinformatiky

Palindromy

C2131 Úvod do bioinformatiky

C2131 Úvod do bioinformatiky

NCBR – Masaryk University

Palindrom

- Specifický případ sekvence, která je z obou stran stejná

Matematika

0
11
121
12321
123321
95135753159

Jazyk

OKO
KAJAK
RADAR
Kobyla má malý bok.
Jelenovi pivo nelej.

Hudba



Joseph Haydn
Symfonie 47 v G dur



Palindrom


- V biologii (biochemii) se palindromy vyskytují v **sekvencích DNA**
- Aby mohla být sekvence palindrom, **MUSÍ** se číst zepředu a zezadu stejně – **POZOR**, sekvence biopolymerů (proteiny, DNA, RNA) jsou orientované!!! A proto:

Protein: MELEM 

N-Met-Glu-Leu-Glu-Met-C

≠

C-Met-Glu-Leu-Glu-Met-N

RNA/ssDNA: AGCCGA 

5'-A-G-C-C-G-A-3'

≠

3'-A-G-C-C-G-A-5'

DNA: TCATGA 

5'-T-C-A-T-G-A-3'
3'-A-G-T-A-C-T-5'

=

3'-A-G-T-A-C-T-5'

5'-T-C-A-T-G-A-3'

=

5'-T-C-A-T-G-A-3'

3'-A-G-T-A-C-T-5'

Palindromatické sekvence v DNA

- Sekvence, která je totožná s KOMPLEMENTÁRNÍ sekvencí (vždy sudý počet nukleotidů)

- Dvoupísmenné (celkem 4, bez zvláštního významu) AT TA CG GC

- **Čtyřpísmenné** (celkem 16)

AATT	ATAT	AGCT	ACGT
TATA	TTAA	TGCA	TCGA
GATC	GTAC	GGCC	GCGC
CATG	CTAG	CGCG	CCGG

- **Šestipísmenné** (celkem 64)

AAATTT	AATATT	ATATAT	ATTAAT	TAATTA	TATATA	TTATAA	TTTAAA
AAGCTT	...						
...							

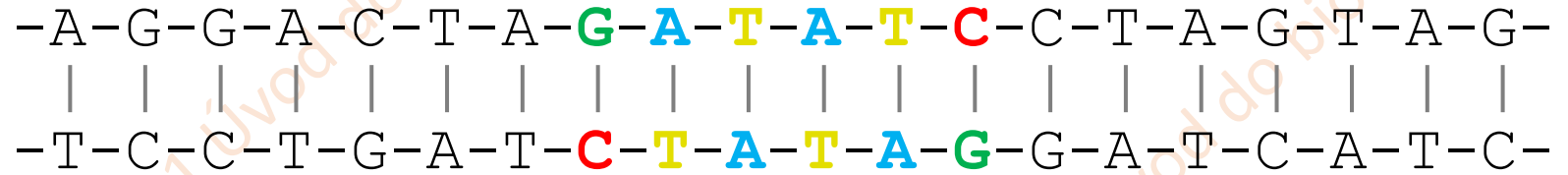
- Delší

Výskyt palindromů v DNA

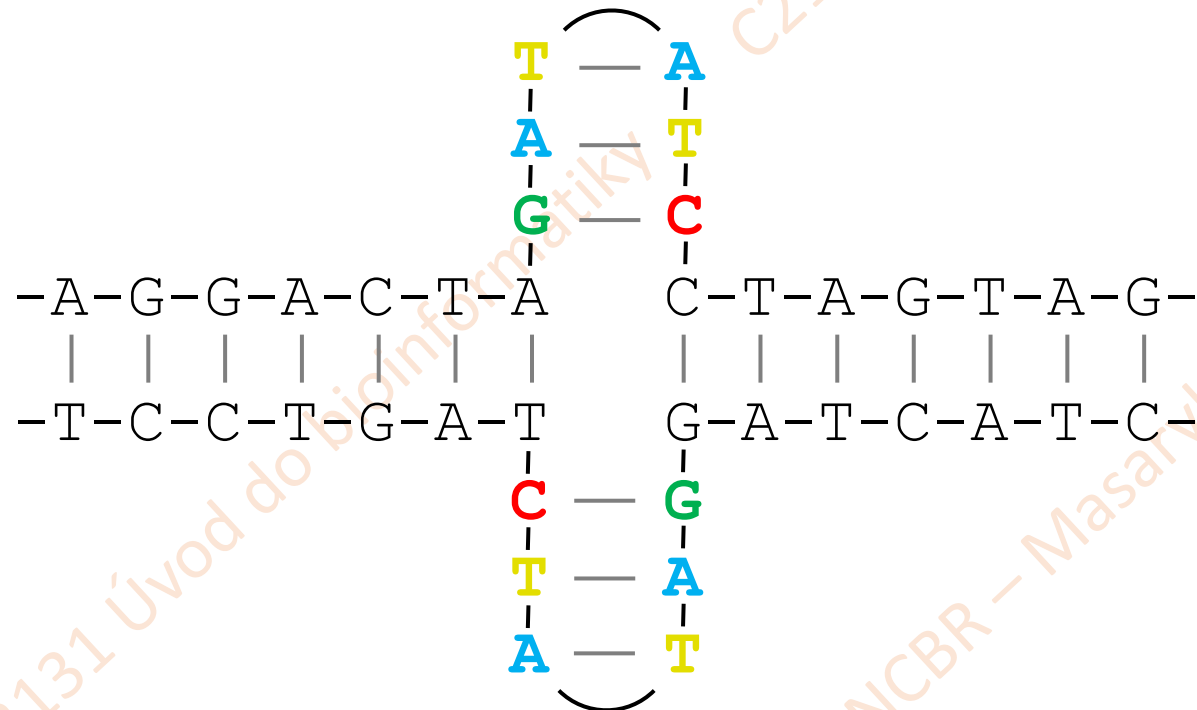
- Náhodný – čistě statisticky se mohou palindromy vyskytovat často:
 - Dvoupísmenný – průměrně každá 4. dvojice
 - Čtyřpísmenný – průměrně každá 16. čtveřice
 - Šestipísmenný – průměrně každá 64. šestice
- **Cílený** – využívaný buňkou pro rozpoznávání/manipulaci DNA
- V reálné DNA je **náhodný výskyt palindromů potlačen**
(je jich méně, než by se dalo čekat na základě náhodného výskytu)

Struktura DNA palindromů

- **Standardní párování**

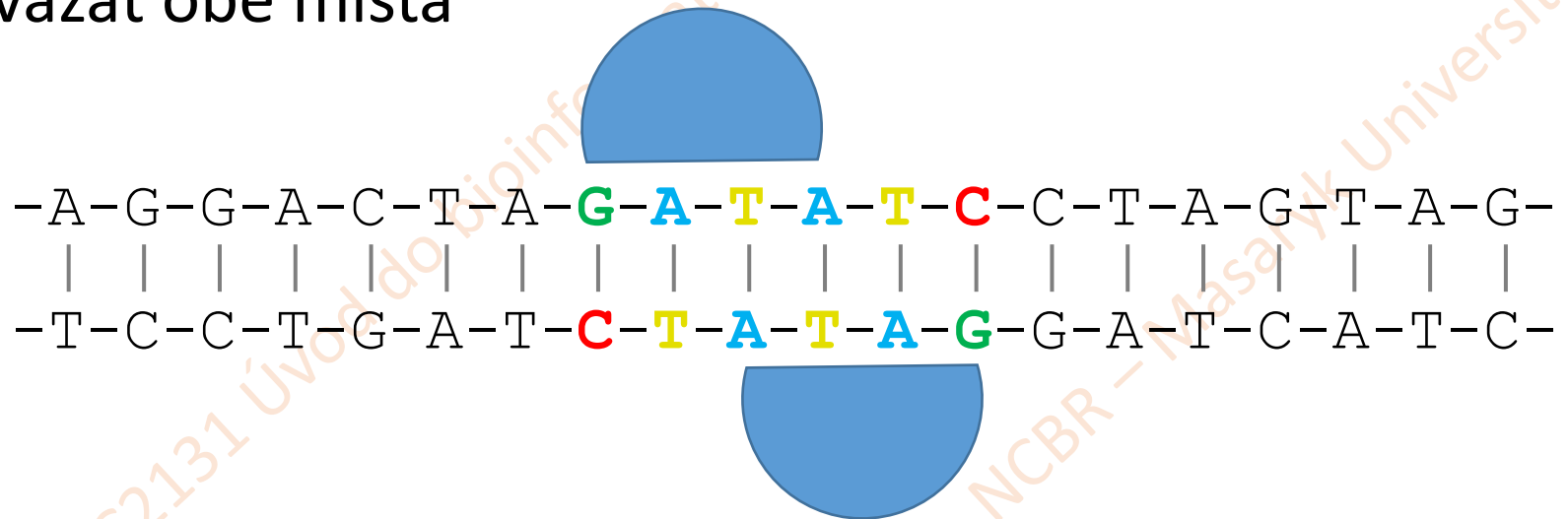


- **Tvorba vlásenek**

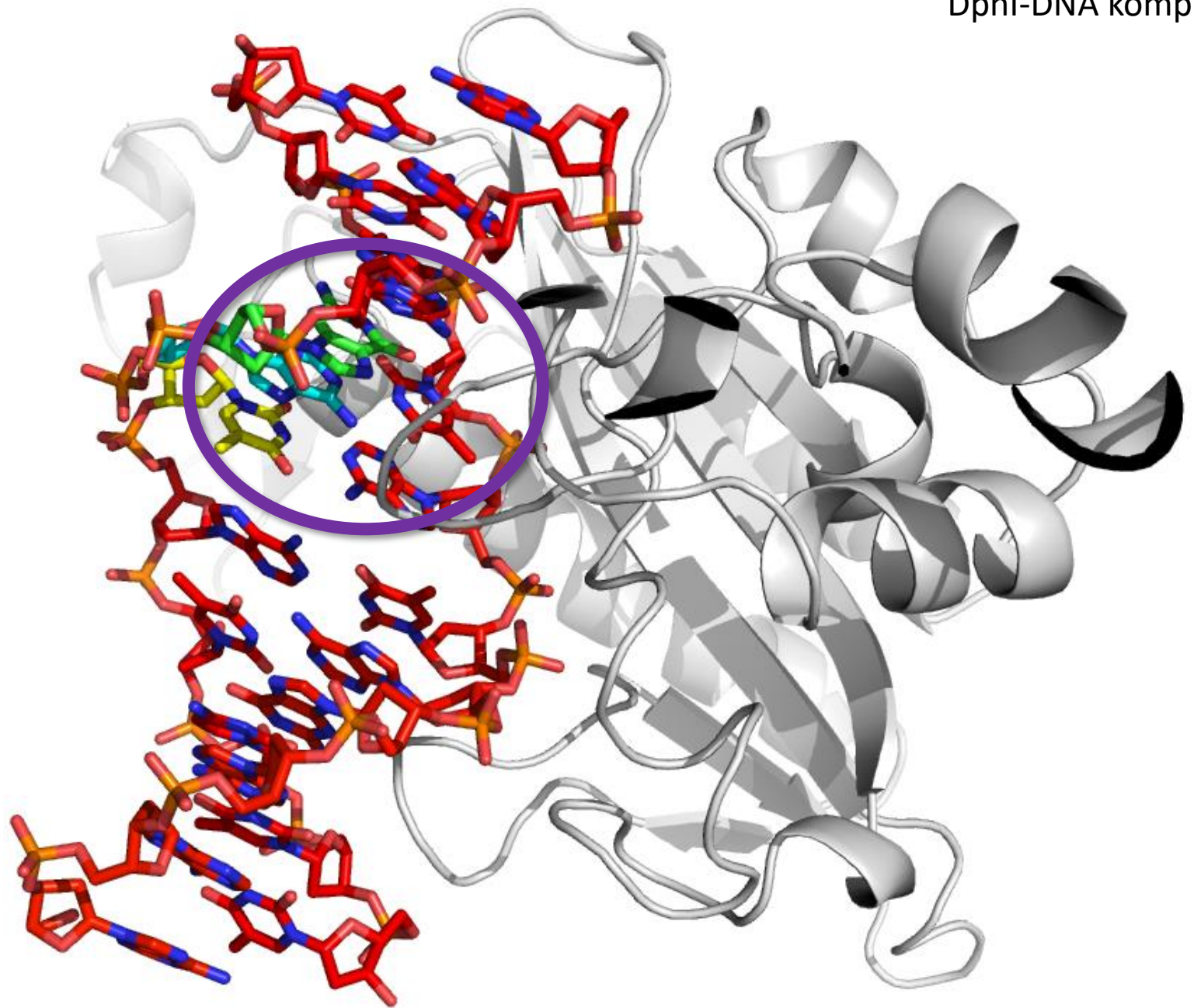


Palindromy a interakce s proteiny

- Díky palindromu se opakovaná sekvence vyskytuje dvakrát naproti sobě
- Protein může rozpoznat DNA na dvou místech
- **Dimer** proteinu může vázat obě místa
 - Výrazně silnější vazba
 - Vyšší specifita



DpnI-DNA komplex



NCr

'sity

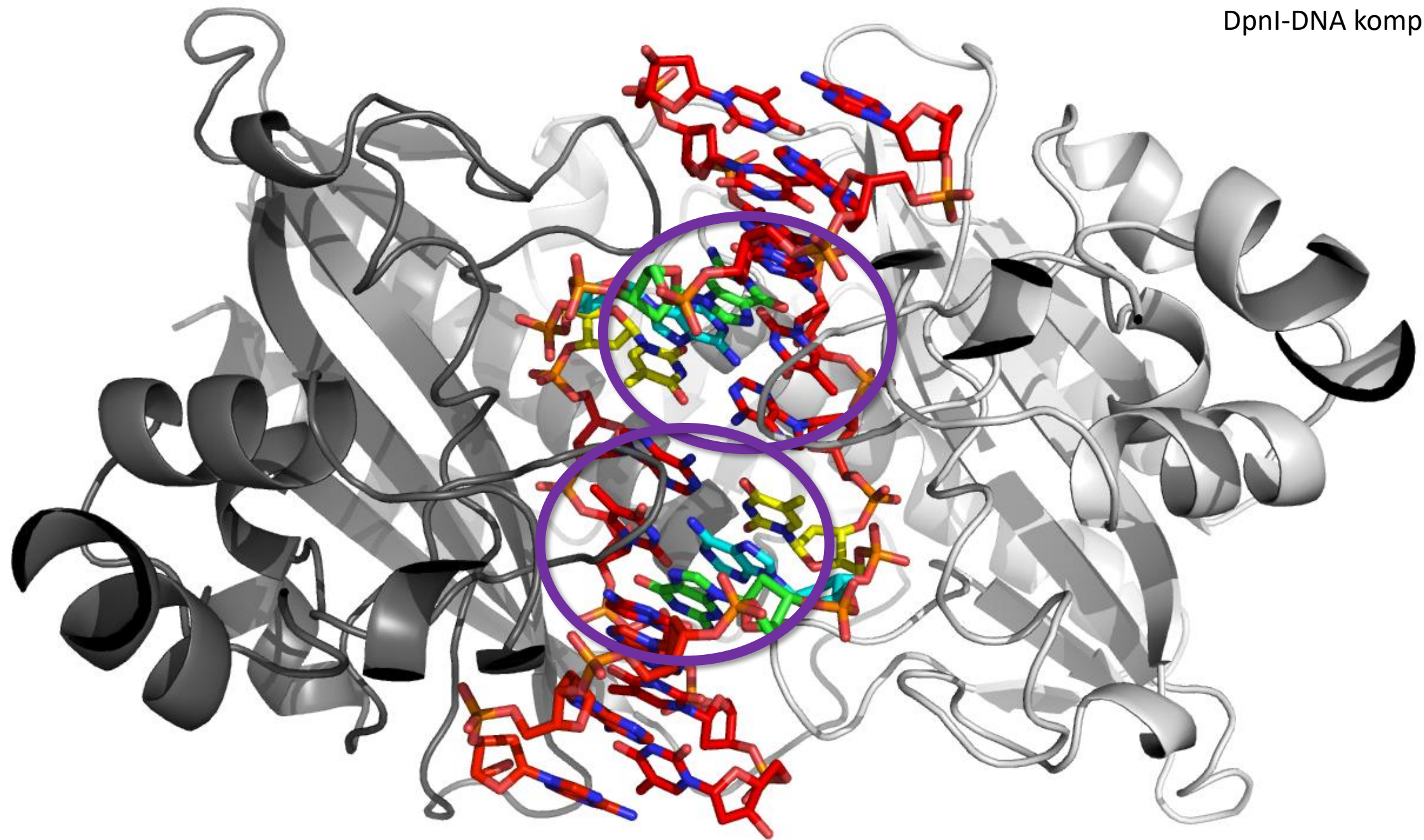
C27

DpnI-DNA komplex

Ncr

'sity

C27



NCBR – Masaryk University

C2131 Úvod do bioinformatiky

Restrikční endonukleasy

C2131 Úvod do bioinformatiky

C2131 Úvod do bioinformatiky

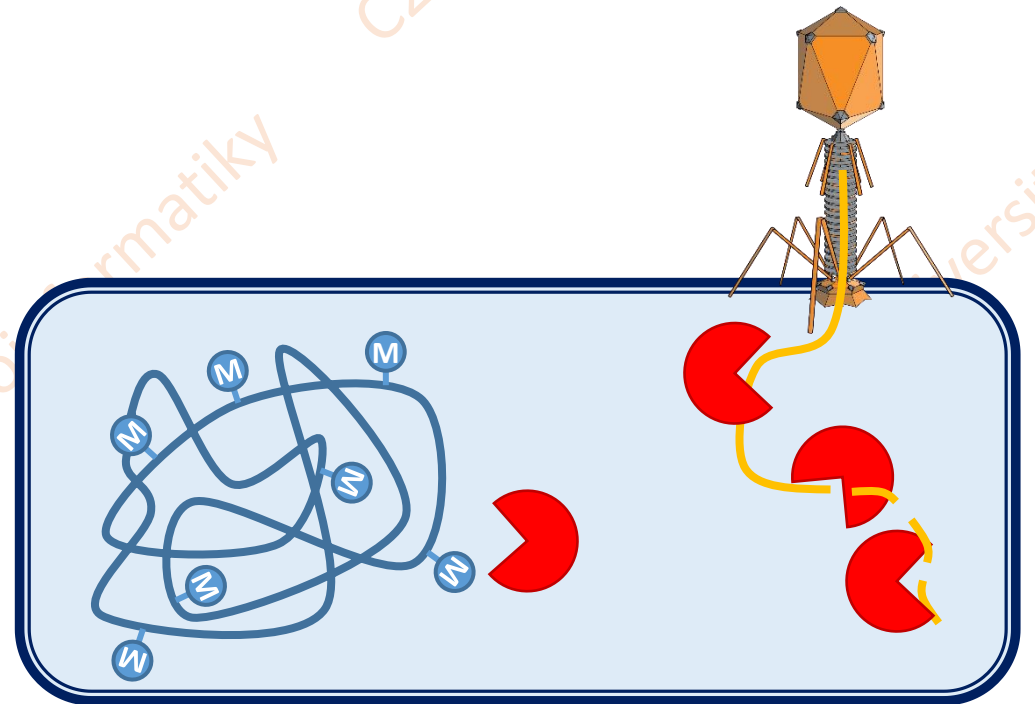
NCBR – Masaryk University

Restrikční endonukleasy

- Nukleasy – proteiny štěpící DNA
- Exonukleasy – štěpí DNA od konce
- Endonukleasy – štěpí DNA uvnitř řetězce
- **Restrikční endonukleasy** – štěpí DNA po vazbě na určité místo sekvence (často jde právě o palindrom)

Restrikční endonukleasy (RE)

- RE jsou enzymy **bakteriálního původu**
- Typicky existují v komplexu s methylasou, která methyluje vlastní DNA
- Slouží pro **ochranu** před bakteriofágy – štěpí fágovou DNA (nemethylovanou)



Název restrikčních endonukleas

- První písmeno rodového jména
- První dvě písmena druhového jména
- Další rozlišení (kmen, typ...)
- Pořadové číslo v daném organismu

Např.: EcoRI – *Escherichia coli* RY13
BamHI – *Bacillus amyloliquefaciens* H
SacI – *Streptomyces achromogenes*
HindIII – *Haemophilus influenzae* Rd
Sall – *Streptomyces albus* G

EcoRI

Escherichia coli kmen RY13 První popsaná restriktasa z této *E. coli*

Klasifikace RE

- **Klasifikace** podle místa štěpení a kofaktorů
 - Typ I – proteinový komplex, náhodné místo štěpení
 - **Typ II** – štěpí v definovaném místě v blízkosti rozpoznávané sekvence, Mg²⁺ ionty jako kofaktor
 - Typ III – hydrolyzují ATP, specifické místo štěpení
 - Typ IV – štěpí modifikovanou DNA (např. methylovanou)

- Existuje větší množství **podtypů**

Table 1. Subtypes of Type II REases

Subtype ^a	Defining feature	Examples	Recognition sequence
A	Asymmetric recognition sequence	FokI AclI	GGATG (9/13) CCGC (-3/-1)
B	Cleaves both sides of target on both strands	BglI	(10/12) CGANNNNNTGTC (12/10)
C	Symmetric or asymmetric target. R and M functions in one polypeptide	GsuI HaeIV BclI	CTGGAG (16/14) (7/13) GAYNNNNNRTC (14/9) (10/12) CGANNNNNTGTC (12/10)
E	Two targets; one cleaved, one an effector	EcoRII NaeI	↓CCWGG GCC↓GGC
F	Two targets, both cleaved coordinately	SfiI SgrAI	GGCCNNNN↓NGGCC CR↓CCGGYG
G	Symmetric or asymmetric target. Affected by AdoMet	BsgI Eco57I	GTGCAG (16/14) CTGAAG (16/14)
H	Symmetric or asymmetric target. Similar to Type I gene structure	BclI AhdI	(10/12) CGANNNNNTGTC (12/10) GACNNN↓NNGTC
M	Subtype IIP or IIA. Require methylated target	DpnI	Gm6 A↓TC
P	Symmetric target and cleavage sites	EcoRI PpuMI BslI	G↓AATTC RG↓GWCCY CCNNNNN↓NNGG
S	Asymmetric target and cleavage sites	FokI MmeI	GGATG (9/13) TCCRAC (20/18)
T	Symmetric or asymmetric target. R genes are heterodimers	Bpu10I BslI	CCTNAGC (-5/-2) ^b CCNNNNN↓NNGG

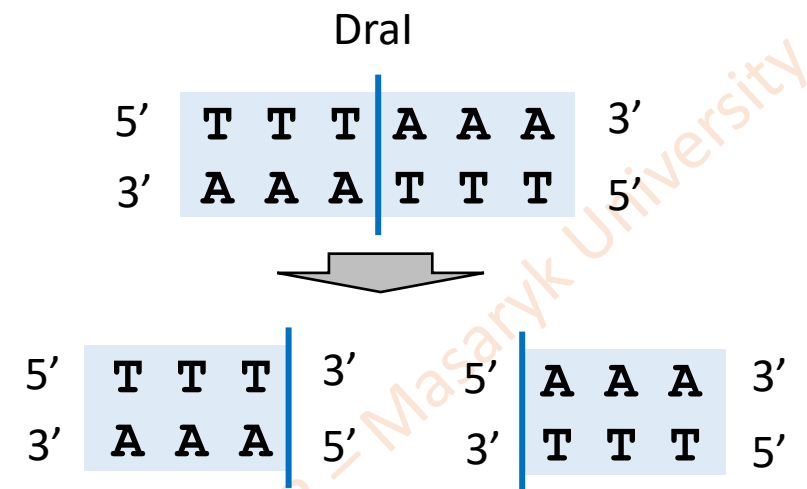
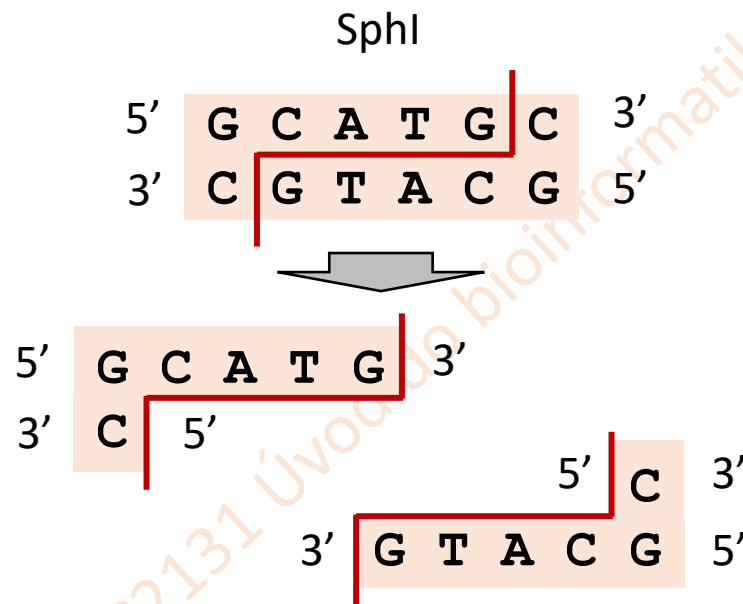
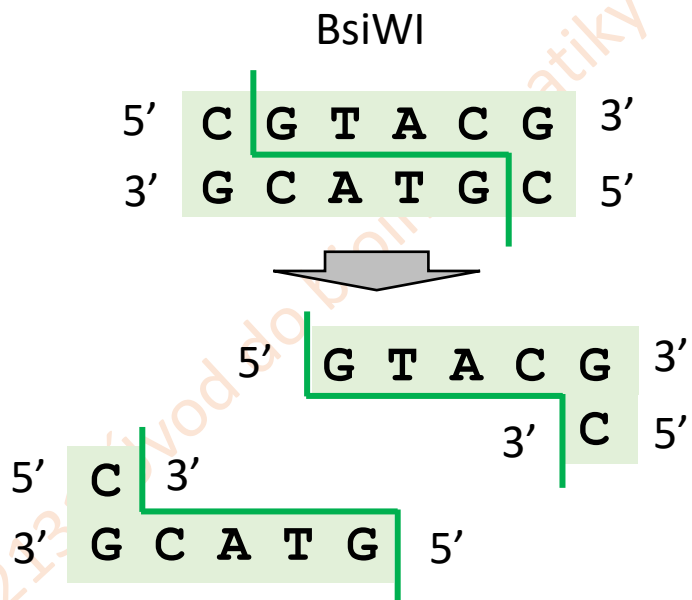
^aNote that not all subtypes are mutually exclusive. E.g. BslI is of subtype P and T.

^bThe abbreviation indicates double strand cleavage as shown below:

5' C C ↓ T N A G C
3' G G A N T ↑ C G

Štěpení – tvorba konců na DNA

- RE typicky štěpí obě vlákna DNA ve stejném místě sekvence
- V závislosti na pozici štěpení mohou vznikat:
 - 5' převislé konce (5' overhangs)
 - 3' převislé konce (3' overhangs)
 - Tupé konce (blunt ends)



Schizomery

- **Isoschizomery** – dvě RE, které mají stejnou specifitu (rozpoznávají stejnou sekvenci a štěpí ji stejně)

HpaII: C↓CGG

MspI: C↓CGG

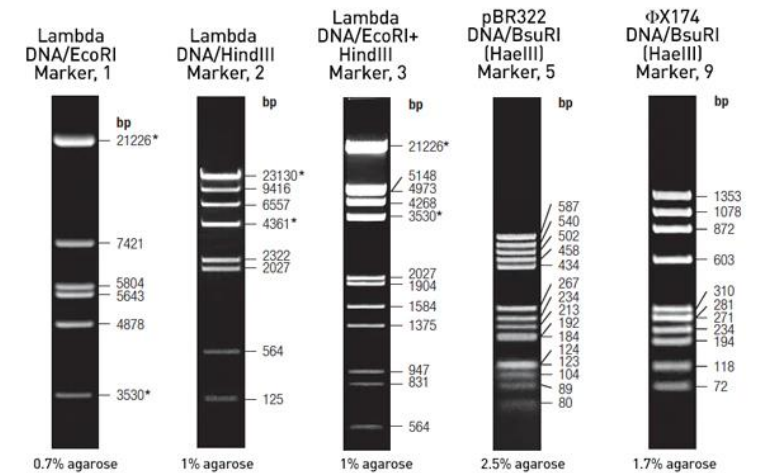
- **Neoschizomery** – rozpoznávají stejnou sekvenci, avšak štěpí ji různě

AatII: GACGT↓C

ZraI: GAC↓GTC

Využití restrikčních endonukleas

- Charakterizace **fágové DNA**
- Tvorba **velikostních standardů DNA**
- **Analýza DNA** – mutace, množství kopií genů
- Příprava DNA **pro sekvenaci** a další analýzy
- Modifikace DNA (**klonování**) – tvorba GMO



*Cohesive ends (the 12 nt cos site of lambda DNA) may anneal and form additional bands.

ThermoFisher Scientific

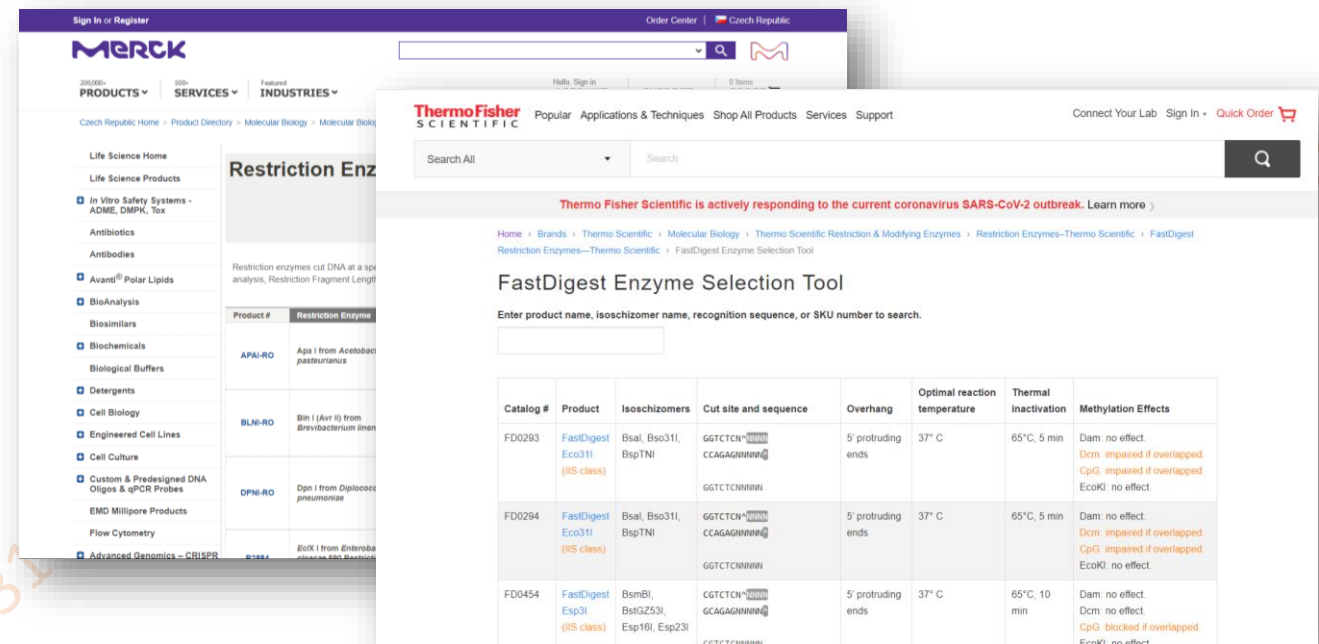
Naštěpená fágová DNA sloužící jako velikostní standardy pro DNA elektroforézu

Databáze restrikčních endonukleas

- Typicky spravované výrobci restrikčních endonukleas
 - Pozor: Mohou obsahovat jen enzymy daného výrobce !
- Obsahují název, rozpoznávané místo, způsob štěpení a další informace
- Často včetně **nástrojů** pro analýzu a vyhledávání

Příklady:

- New England Biolabs (NEB)
- **REBASE** (spravováno NEB)
- ThermoFisher Scientific
- a další






ThermoFisher Scientific is actively responding to the current coronavirus SARS-CoV-2 outbreak. Learn more

Home » Brands » Thermo Scientific » Molecular Biology » Thermo Scientific Restriction & Modifying Enzymes » Restriction Enzymes—Thermo Scientific » FastDigest Restriction Enzymes—Thermo Scientific » FastDigest Enzyme Selection Tool

FastDigest Enzyme Selection Tool

Enter product name, isoschizomer name, recognition sequence, or SKU number to search.

Catalog #	Product	Isoschizomers	Cut site and sequence	Overhang	Optimal reaction temperature	Thermal inactivation	Methylation Effects
FD0293	FastDigest Eco311 (HS class)	Bsal, Bso311, BspTNI	GGTCTCH-  CCAGAGNNNN	5' protruding ends	37° C	65°C, 5 min	Dam: no effect. Dcm: impaired if overlapped CpG: impaired if overlapped EcoKI: no effect.
FD0294	FastDigest Eco311 (HS class)	Bsal, Bso311, BspTNI	GGTCTCH-  CCAGAGNNNN	5' protruding ends	37° C	65°C, 5 min	Dam: no effect. Dcm: impaired if overlapped CpG: impaired if overlapped EcoKI: no effect.
FD0454	FastDigest Esp31 (HS class)	BsmBI, BstGZ531, Esp16I, Esp23I	CGTCTCH-  GCAGAGNNNN	5' protruding ends	37° C	65°C, 10 min	Dam: no effect. Dcm: no effect. CpG: blocked if overlapped EcoKI: no effect.

REBASE

- Nejobsáhlejší („kompletní“) databáze restrikčních enzymů

<http://rebase.neb.com/>







Use this tool as a guide to the ever-changing landscape of restriction enzymes.
REBASE is a dynamic, curated database of restriction enzymes and related proteins.

RESOURCES ?

Data files
Sequence data
PacBio data
Crystal Data
Methylation Sensitivity

Lists and compilations
Tools
Enzymes
Genomes
Suppliers

Publications
Related websites
FTP

-  WHAT'S NEW?
-  SUBMIT DATA
-  SUBSCRIBE
-  CONTACT US

SEARCH ?

use quotes around phrases

by

Search

Clear

[Advanced Search](#)



ABOUT REBASE



CITING REBASE



HELP



Classic

REBASE

- Enzymy podle typů, podle rozpoznávaných sekvencí, ...

REBASE Lists

Citing REBASE...

Enzyme Navigation Tables:

- All Acronyms...
- Enzymes Grouped by Type
- Prototypes/neoschizomers...
- Newest Enzymes

REBASE Genomes:

- Genome List...
- Genomes with shotgun...
- Archaea summary
- Bacteria summary

Sequence Collections:

- Download DNA/Protein sequences...
- Collect customized FastA files...
- Type II specificities...
- Sequenced Type II R genes...
- Methyltransferases...
- Methyltransferases with protein seqs...
- Methyltransferases with refs...
- Cloned/Sequenced enzymes...
- COG table...
- PID table...
- RM gene coordinates in sequenced genomes...
- Genbank Cross Reference...
- Protein Id Cross Reference...
- UniProtKB Cross Reference...
- GI Cross Reference...
- Jumbo Cross Reference...

Specialized Information:

- PacBio...
- Restriction enzyme properties...
- Split Type II rec seqs...
- Methylase recognition sequences...
- Methylase recognition sequences (plain text)...
- Type II m4C methylases (plain text)
- Type II m5C methylases (plain text)
- Type II 6mA methylases (plain text)
- Type II m4C, m5C, and 6mA methylases (plain text)
- FDA Salmonella strain summary...
- Gold Standard...
- Prophages...
- Pt components...
- Type I recognition sequences...
- Type I S subunit TRD sequences...
- Gold Standards Type I S subunit TRD sequences...
- Fused enzymes...
- Crystal data...
- N-term data...
- Star Activity...
- Methylation Sensitivity...
- Kinetics data...
- CG list...
- Cleavage of RNA/DNA Hybrids...
- Methyltransferase protein sequence lengths...
- Restriction Enzyme Molecular Weights...
- Methyltransferase Molecular Weights...
- Molecular Weights (all enzymes)...
- Enzyme sub types...
- Orphan methyltransferases...
- Methyltransferase list...
- Methyltransferase list (with subtypes)...
- Tetranucleotide palindrome methylases...
- Enzymes with alternative names...
- Growth Temperatures...
- Organism Types...
- Journal abbreviations...



REBASE Enzymes 03/31/2021

Type II restriction enzymes
producing 5' overhangs



Length	Overhang	Enzyme	Recognition Sequence	Molecular Weight	Suppliers	Isoschizomers
4	NNNN	AarI	CACCTGC (4/8)	-	B	PacCI
4	TCGA	AbaI	CC↓TCGAGG	-	I	-
2	MK	AccI	GT↓MKAC	42493	BJKMNOX	FbiI XmiI
4	NNNN	AceII	CAGCTC (7/11)	-	-	-
2	CG	AciI	CCGC (-3/-1)	60993	N	BspACI SsiI
2	CG	AclI	AA↓CGTT	37882	INV	Psp1406I
2	CG	AcyI	GR↓CGYC	-	J	AhaII AosII AstWI AsuIII BbiII BsaHI BssNI BstACI HgiI HgiDI HgiGI HgiHII HinII Hsp92I Msp17I PamII
4	TTAA	AflII	C↓TTAAG	36774	JKN	BfrI BspTI Bst98I BstAFI BstPZ740I Esp4I MspCI Vha464I
4	CRYG	AflIII	A↓CRYGT	26954	MNS	-
4	CCGG	AgeI	A↓CCGGT	30536	JNR	AsiAI AsiGI BshTI CsiAI CspAI PinAI
4	TGCA	ApaLI	G↓TGAC	-	CKN	Alw44I SnoI VneI
4	AATT	ApoI	R↓AATY	-	N	AcsI Cfal FsiI XapI
4	CGCG	AscI	GG↓CGGCC	50498	N	PalAI SgsI
2	AT	Asi256I	G↓ATC	-	-	-
4	GTAC	Asp718I	G↓GTACC	-	MS	Acc65I AhaB8I SthI
3	GNC	AsoI	G↓GNCC	-	-	AspS9I AvcI Bac36I Bal228I BavAII BavBII Bce22I BmgT120I BshKI BsiZI Bsp1894I BspBII BspF4I Bsu54I CcuI Cfr13I MaeK81II NspIV Nsp712II Pde12I PspPI Sau96I
2	CG	AsuII	TT↓CGAA	-	C	AcpI Asp10HI BimI Bim19I Bpu14I BsiCI Bsp119I BspLAI BspT104I BstBI CbiI Csp45I Csp68KII EspII LspI MlaI NspV PlaII PpaAI SfuI Ssp1I SspRFI SviI
4	YCGR	AvaI	C↓YCGRG	35676	JNQX	Ama87I ApuI BcoI BmeT110I Bse15I BsiHKCI BsoBI BspLU4I BstSI Eco88I Eco27kI NspIII NspSAI OfoI PlaAI PunAI
3	GWC	AvaII	G↓GWCC	26127	JNX	AflI Asp745I BamNII BcuAI Bme18I Bme216I BmpI BsrAI BthAI Caul Csp68KI DsaIV EagMI Eco47I ErpI FdiI EspMSI FsslI HgiBI HgiCII HgiEI HgiHIII HgiJ Kzo49I MspNI SinI SnuEI VpaK11BI
4	CTAG	AvtII	C↓CTAGG	-	N	Acp2I AvtBII BinI BspA2I XmaJI
4	GATC	BamHI	G↓GATCC	24569	BCIJKMNQRSVXY	AccEBI AliI ApaCI AsiI Bce75I BnaI Bsp98I Bsp4009I BspAAIII BstI Cell GstI Mlu23I Nsp29132II NspSAIV OkraI Pfi8I RspLkII Soll Surl Uba4009I
4	NNNN	BbvI	GCAGC (8/12)	61734	N	AlwXI BseKI BseXI Bsp423I Bst12I Bst7I BstV1I GeoICI Lsp1109I
4	NNNN	BbvII	GAAGAC (2/6)	-	-	Bbr7I BbsI Bbv16II BpiI BpuAI Bsc9I BspBS3I BspIS4I BspTS514I BstBS32I BstTS5I BstV2I
3	TCA	BbvCI	CCTCAGC (-5/-2)	32615	N	AbeI

REBASE – příklad záznamu



GOLD STANDARD

EcoRI

Type II restriction enzyme
subtype: P

Different enzyme:

Go

Recognition Sequence: [help?](#)

G[^]AATTC



REBASE Enz Num 993 entered Jan 1 1972 ... modified Jul 3 2014

Acronym: [EcoR](#)

Prototype: EcoRI

Org #: [1394](#)

Organism: [Escherichia coli RY13](#)

Organism type: [plasmid](#)

Organism source: [R.N. Yoshimori](#)

Growth Temperature: 37 °

Experimental Evidence: [biochemistry](#)

Exhibits star activity

Single-stranded cleavage: y

Enzyme gene cloned

Enzyme gene sequenced

Crystal data present

[Kinetics data present](#)

Molecular Weight: 31057

sites on

Adeno2: 5

Lambda: 5

pBR322: 1

PhiX174: 0

SV40: 1

[Status of methylation sensitivity testing...](#)

[Site frequency in sequenced genomes...](#)

[DNA RNA hybrids...](#)

[RNA duplexes...](#)

[Single stranded DNA...](#)

[Combined report...](#)

Related Enzymes:

[M.EcoRI](#)

Related References

[sorted by date](#) [in new window](#)

[sorted by authors](#) [in new window](#)

[Commercially Available...](#)

[NEB EcoRI-HF](#)

[Similar enzymes...](#)

Sequence Data:

[EcoRI](#)

[Crystal Data...](#) (8 structures)

[Kinetic Data...](#) (5 records)

[Methylation Sensitivity...](#)

New England Biolabs – Restrikční enzymy

- Dodavatel řady RE
- Užitečné **nástroje** na webových stránkách

NEW ENGLAND **BioLabs** Inc. be INSPIRED drive DISCOVERY stay GENUINE

Applications & Products Tools & Resources Support About

Home > Tools & Resources > Selection Charts > Alphabetized List of Recognition Specificities

Alphabetized List of Recognition Specificities

All restriction endonuclease recognition specificities available from New England Biolabs are listed below. For enzymes that recognize non-palindromic sequences, the complementary sequence of each strand is listed. For example, CCTC(7/6) and (6/7)GAGG both represent an MnlI (NEB #R0163) site.

All recognition sequences are written 5' to 3' using the **single letter code** nomenclature with the point of cleavage indicated by a "/>. Numbers in parentheses indicate point of cleavage for non-palindromic enzymes.

For example, GGTCTC(1/5) indicates cleavage at:
5' ...GGTCTCN/...3'
3' ...CCAGAGNNNN/...5'

Recognition Sequence
AA/CGTT
A/AGCTT
AAT/ATT
/AATT
A/CATGT
A/CCGGT
ACCTGC(4/8)
A/CCWGGT

NEW ENGLAND **BioLabs** Inc. be INSPIRED drive DISCOVERY stay GENUINE









Applications & Products Tools & Resources Support About

Search NEB

Home > Tools & Resources > Interactive Tools

Interactive Tools

NEB Tools

-  **Competitor Cross-Reference Tools**
Use this tool to select another company's product and find out which NEB product is compatible. Choose either the competitor's product name or catalog number from the available selections, and this tool will identify the recommended NEB product.
-  **Double Digest Finder**
Use this tool to guide your reaction buffer selection when setting up double-digests, a common timesaving procedure. Choosing the right buffers will help you to avoid star activity and loss of product.
-  **Exo Selector**
Use this tool to simplify the process of selecting the appropriate exonucleases for use in your nucleic acid digestion workflows. The tool guides you to product recommendations based on your answers to a few simple questions.
-  **NEB Golden Gate Assembly Tool**
Use this tool to assist with in silico DNA construct design for Golden Gate DNA assembly. It enables the accurate design of primers with appropriate type IIS restriction sites and overlaps, quick import of sequences in many formats and export of the final assembly, primers and settings.
-  **DNA Sequences and Maps Tool**
Use this tool to find the nucleotide sequence files for commonly used molecular biology tools, including plasmid, viral and bacteriophage vectors.
-  **Enzyme Finder**
Use this tool to select restriction enzymes by name, sequence, overhang or type. Enter your sequence using single letter code nomenclature, and Enzyme Finder will identify the right enzyme for the job.
-  **Glycan Analyzer**
Use this tool to interpret Ultra or High Pressure liquid chromatography (UPLC/HPLC) N-glycan profiles following exoglycosidase digestions.
-  **NEBaseChanger**
NEBaseChanger can be used to design primers specific to the mutagenesis experiment you are performing using the Q5 Site-Directed Mutagenesis Kit. This tool will also calculate a recommended custom annealing temperature based on the sequence of the primers by taking into account any mismatches.

NCBR – Masaryk University

C2131 Úvod do bioinformatiky

C2131 Úvod do bioinformatiky

Primery

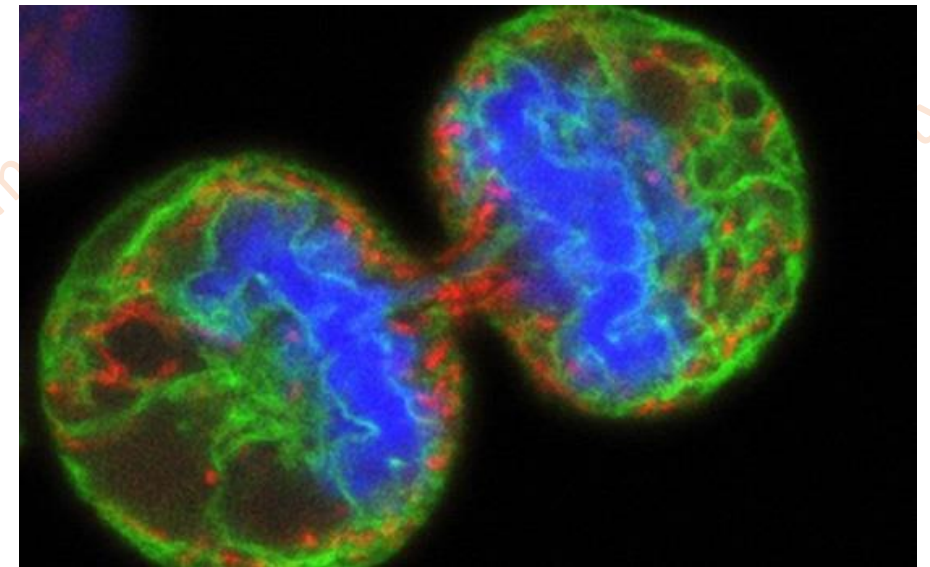
C2131 Úvod do bioinformatiky

C2131 Úvod do bioinformatiky

NCBR – Masaryk University

Replikace DNA

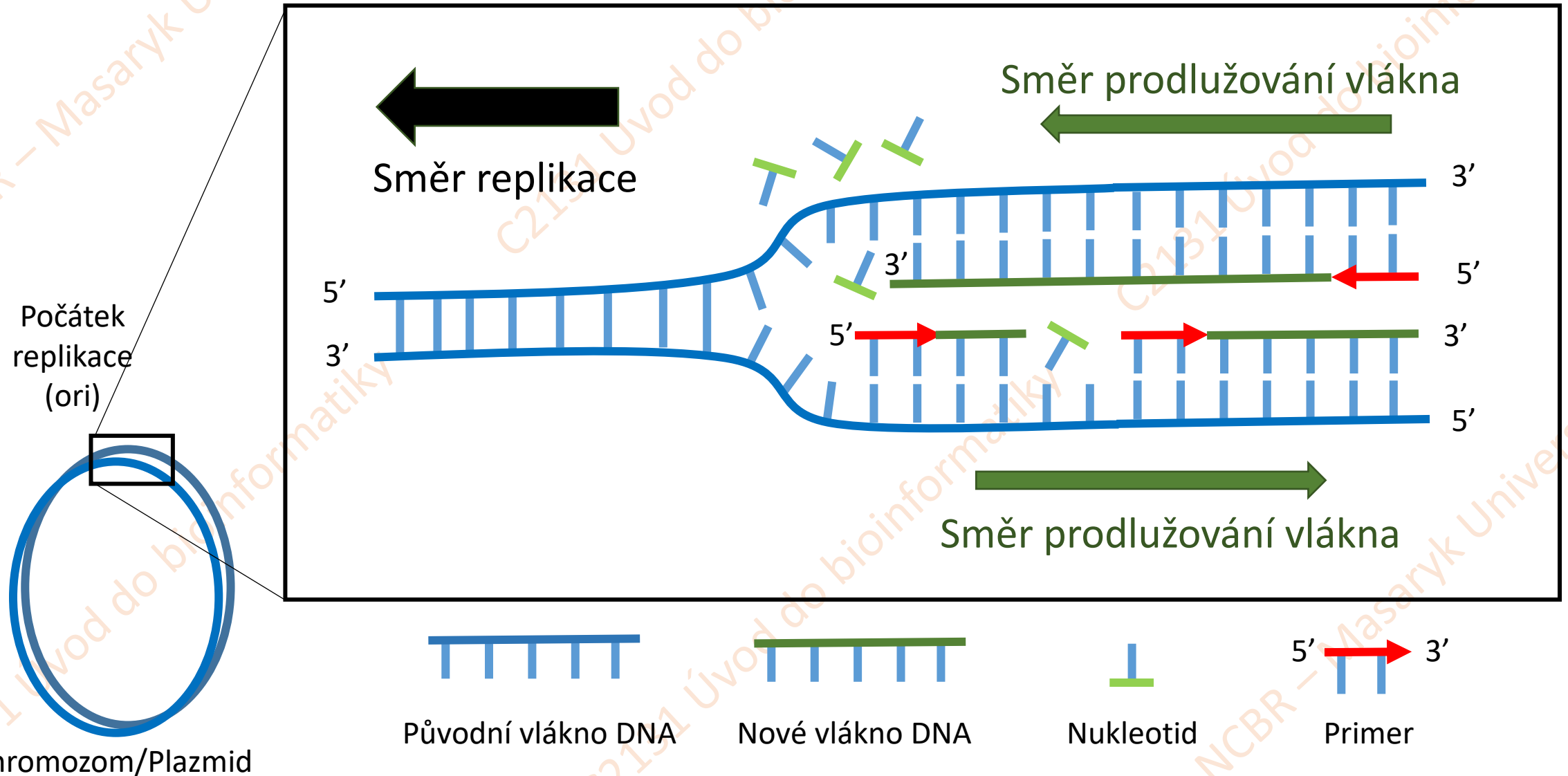
- Aby se mohla buňka dělit, je nutné vytvořit kopii její DNA
- Podobně je nutné zkopírovat i DNA případných organel (mitochondrie, plastidy)
- Celý proces je komplexní, vyžaduje přísun stavebního materiálu, energie a podílí se na něm řada enzymů



Replikace DNA

- DNA je dvouvláknová – je nutno ji „rozplést“
- Na základě párování bází je ke každému vlákně doplněno vlákno komplementární
- Syntéza nového vlákna probíhá VŽDY ve směru **5'→3'**. To znamená, že:
 - Jedno vlákno je syntetizováno nepřetržitě
 - Druhé vlákno je syntetizováno „v protisměru“ po krátkých úsecích (tzv. Okazakiho fragmenty)
- Syntéza každého úseku DNA začíná krátkým oligonukleotidem zvaným **PRIMER**

Replikace DNA – Schéma

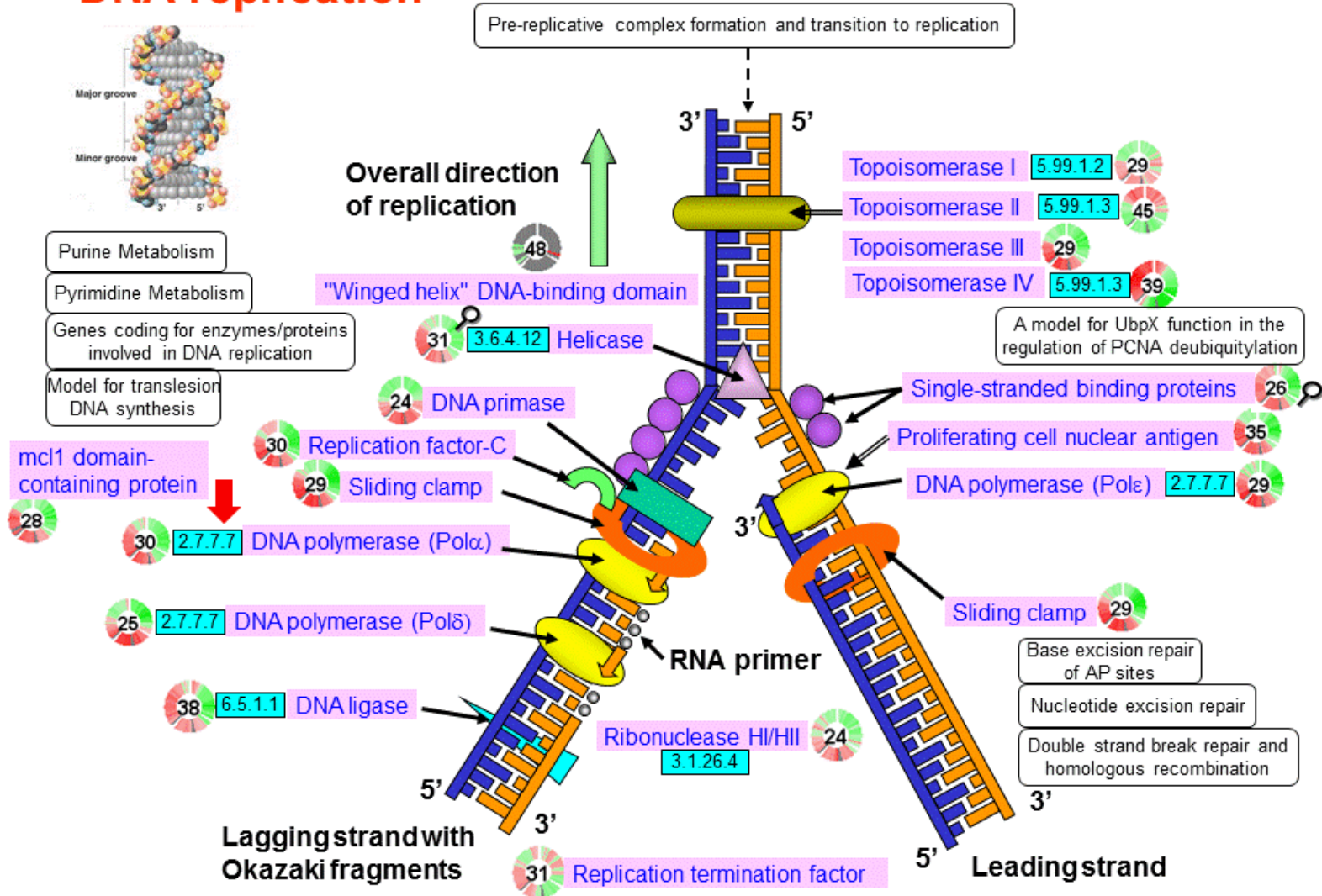


Replikace DNA

• Na procesu se podílí řada proteinů:

- DNA polymerasa
- Primasa
- Ligasa
- Topoisomerasa
- „Svorky“
- Replikační faktory
- ...

DNA replication



Polymerázová řetězová reakce (PCR)

- Replikace DNA *in vitro* (ve zkumavce)
- Využívá **DNA polymerasu z termofilních organismů** – je stabilní a aktivní při vysoké teplotě
- Je založena na vícenásobném opakování cyklu:
 - Rozvolnění DNA (*denaturing*)
 - Nasednutí primerů (*annealing*)
 - Syntéza komplementárního vlákna (*extending*)

Polymerázová řetězová reakce



Vstupní (templátová) DNA

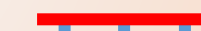
Tepelná
denaturace

95°C



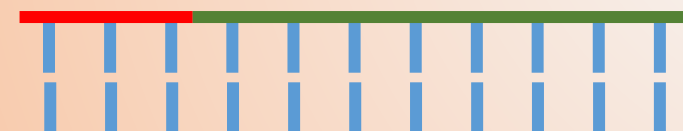
Nasednutí
primerů
(annealing)

45-60°C



72°C

Syntéza DNA
(extending)



C2131 Úvo

NCBR - Masaryk University

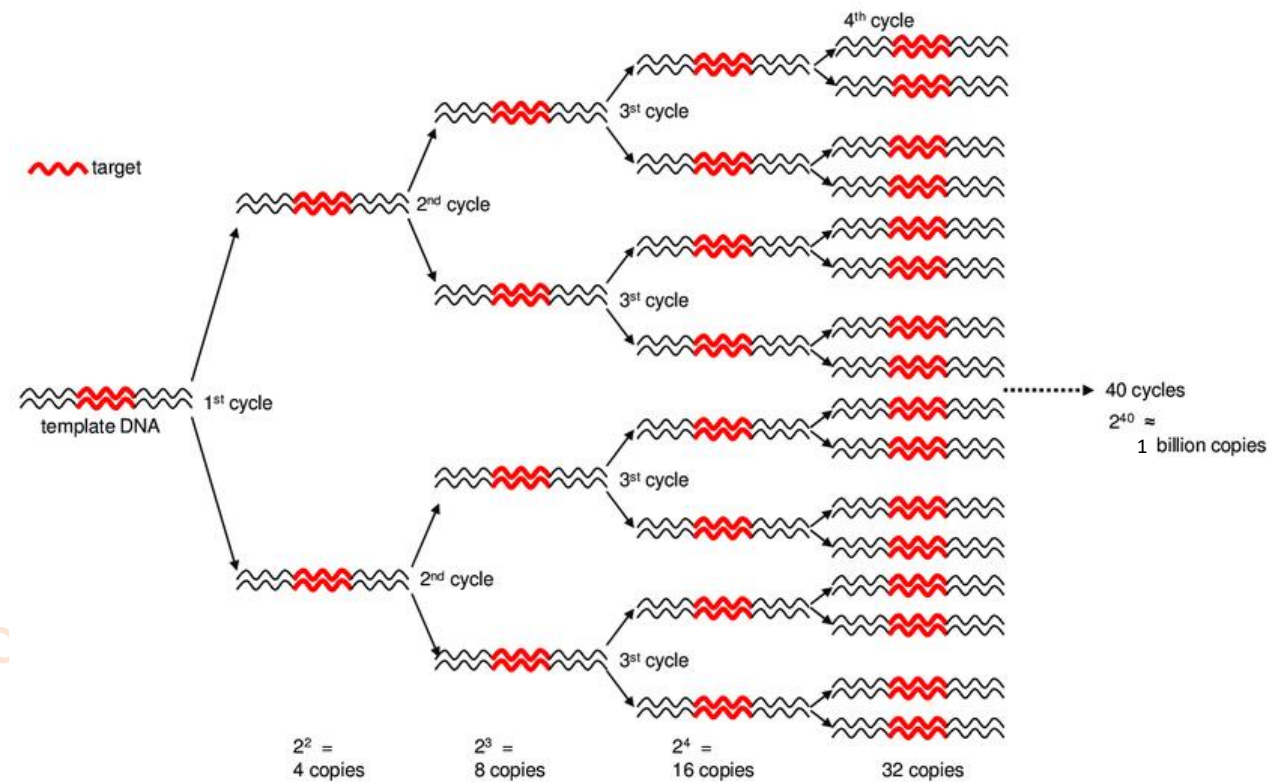
od do bioinformatiky

od do bioinformatiky

NCBR - Masaryk University

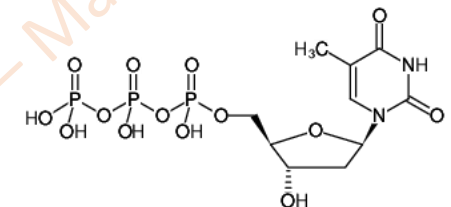
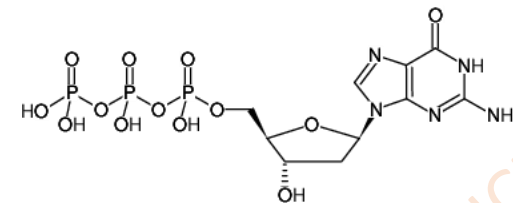
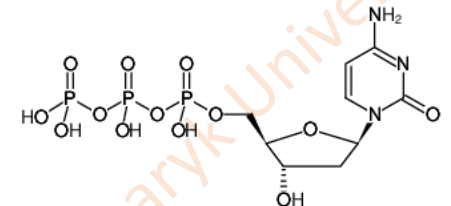
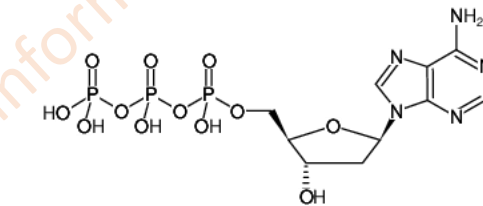
Polymerázová řetězová reakce

- Počet molekul DNA vzrůstá (teoreticky) s druhou mocninou
 - Po prvním cyklu – 2 molekuly
 - Po druhém cyklu – 4 molekuly
 - Po třetím cyklu – 8 molekul
 - ...
 - Po 25. cyklu – 33,5 mil. molekul
 - Po n-tém cyklu – 2^n molekul



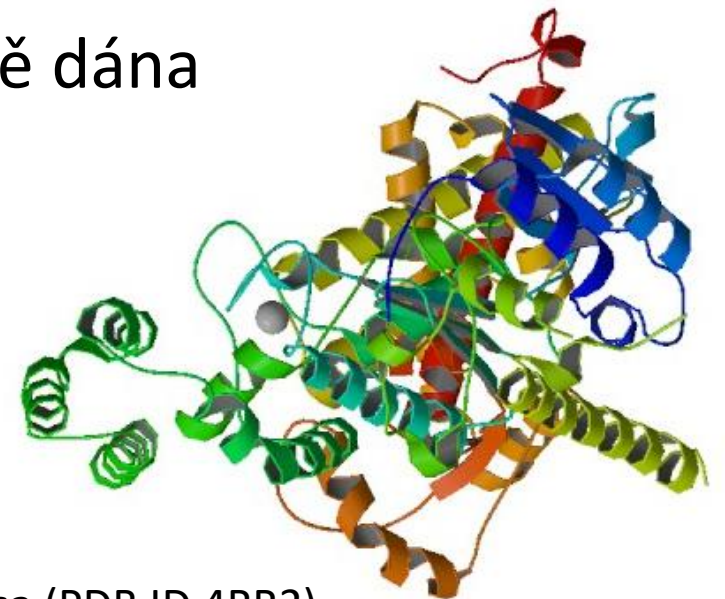
Komponenty PCR

- **Templátová DNA**
- **Termostabilní DNA polymerasa**
- **Primery** – krátké DNA nebo RNA oligonukleotidy
- **Deoxynukleotidtrifosfáty (dNTP)** – stavební materiál a zároveň zdroj energie
- **Mg²⁺ ionty** – kofaktory pro enzymy
- **Vhodné prostředí** – pH, iontová síla



Tvorba primerů v organismu

- Organismy používají **RNA primery**
- Syntetizovány DNA-dependentní-RNA-polymerasou (primasou)
- Pozice primerů (a tedy jejich sekvence) není striktně dána
- **Během replikace** jsou následně **odbourány** a nahrazeny DNA



Lidská primasa (PDB ID 4RR2)

Umělá syntéza primerů

- **Syntetické primery** je možno vytvořit jako RNA i DNA primery
- **DNA primery** jsou stabilnější a používají se častěji
- Syntetické primery lze použít k:
 - Replikaci DNA *in vitro*
 - Vložení mutace do DNA – substituce, delece, inserce
 - Vložení nestandardní báze – např. značené (fluorescenčně, izotopově)

Design primeru

- Syntetické primery mají **definovanou sekvenci**
- Konkrétní sekvence primeru je závislá na účelu použití
- Nejjednodušší **aplikace** – tvorba kopií žádaného úseku DNA
 - Primery jsou zcela komplementární k začátku a konci sekvence
 - Tzn. primery mají stejnou sekvenci jako vlákno v orientaci 5'→3'
- Pro speciální aplikace se může sekvence primeru a DNA lišit !

Parametry primeru

- U každého primeru (resp. každého oligonukleotidu) lze definovat několik parametrů
- **Parametry jsou dané zejména sekvencí primeru** (a částečně podmínkami)
 - Délka primeru (počet bazí/nukleotidů)
 - Obsah guaninu a cytosinu (G+C) v sekvenci primeru
 - Teplota „tání“ (T_m)
 - Tendence k tvorbě sekundárních struktur
 - Specifita nasednutí

Délka primeru

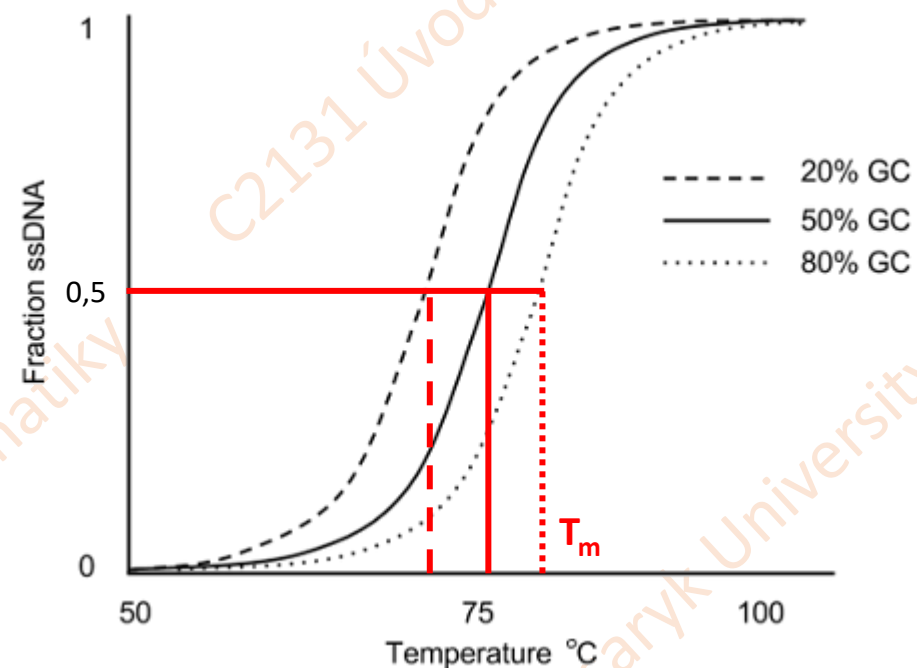
- Běžné primery mají délku **17-28 bp** (*base pairs* = párů bazí)
- Délka primeru má vliv na ostatní parametry (T_m , specifita)
- **Krátké** primery mají nižší specifitu (viz dále), jsou „univerzálnější“
- **Dlouhé** primery (až 40+ bp) se používají zejména pro cílené modifikace DNA

Obsah G+C bází v primeru

- Typické primery jsou tvořeny čtyřmi základními bázemi (A, C, G, T pro DNA, příp. A, C, G, U pro RNA)
- Vzhledem k tomu, že při kanonickém párování se guanin vždy páruje s cytosinem, obsah G+C v primeru (jednom vláknu DNA) odpovídá i obsahu G+C ve dvouvláknové DNA
- Logicky pak platí, že: $(\text{obsah G+C [\%]}) = 100 - (\text{obsah A+T [\%]})$
- **Vyšší obsah G+C** znamená vyšší **stabilitu** při párování bází

Teplota tání primeru

- Každý oligonukleotid (tedy i primer) je schopen se párovat se svým komplementárním vláknem (tvoří tzv. duplex)
- **Stabilita** duplexu je závislá na délce primeru, sekvenci (obsahu G+C), **teplotě** a iontové síle prostředí
- Stabilita klesá s rostoucí teplotou – duplex se snáze rozpadá
- **Teplota tání primeru (T_m)** je teplota, při níž je právě polovina molekul primeru volná a polovina vázaná v duplexu



Upraveno z www.khanacademy.org

Teplota tání primeru

- T_m lze stanovit experimentálně, častěji je však odhadována **výpočtem**
- Pro určení T_m se používají různé vzorce. Např.

$$T_m [^{\circ}\text{C}] = \frac{\Delta H^0}{\Delta S^0 + R \times \ln c_p} - 273,15$$

kde ΔH^0 je entalpie tání, ΔS^0 je entropie tání, R je molární plynová konstanta, c_p je koncentrace primeru

Výpočet může zahrnovat také **efekt iontové síly** roztoku. Např.

$$T_m \text{ (corrected)} = \frac{1}{\frac{1}{T_m} + \left[(4.29 \cdot f_{gc} - 3.95) \cdot \ln(m) + 0.94 \cdot (\ln(m))^2 \right] \cdot 10^{-5}}$$

kde T_m je teplota tání dle předchozího vztahu, f_{gc} je zastoupení G+C bází, m je koncentrace monovalentních iontů

Teplota tání primeru

- V rámci PCR je nutné aby primery opakovaně nasedaly na templátovou DNA
- **Typická T_m** používaných primerů se pohybuje v rozmezí 50 – 65°C
- Zároveň je potřeba, aby se oba použité primery chovaly podobně – fungují zároveň v jedné reakční směsi
- **Rozdíl T_m obou primerů** by měl být ideálně menší než 2°C

Sekundární struktury primeru

- Jako každá nukleová kyselina, i primery mohou tvořit sekundární struktury
- V případě primerů jsme schopni rozlišit:
 - **Vlásenky**
 - **Dimery**
- Pro účely PCR je tvorba sekundárních struktur u primeru **nežádoucí**
- Při analýze sekundárních struktur nás obvykle zajímá ta nejstabilnější

Sekundární struktury – dimerizace

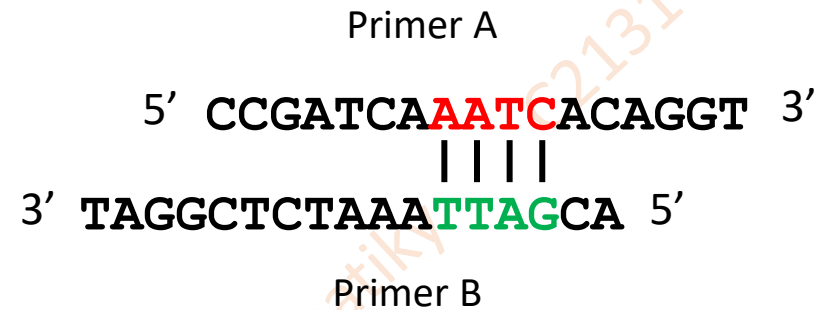
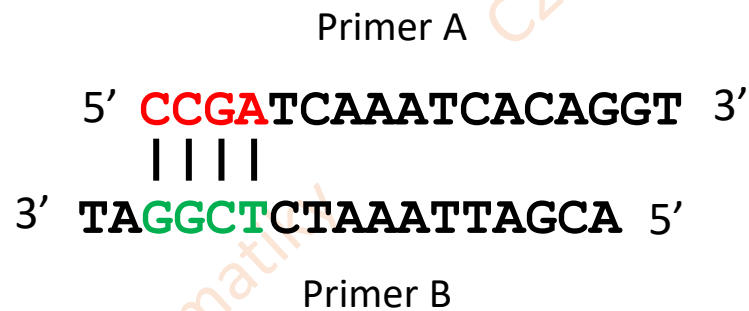
- Pokud se v rámci primeru nachází vzájemně komplementární sekvence (v podstatě obsahují palindrom), mohou molekuly tvořit tzv. **homodimer**



- **Stabilita dimeru** je dána teplotou tání komplementárního úseku
- V extrémním případě může dimer tvořit celý primer
- Úseky kratší než 4 nukleotidy jsou málo stabilní

Sekundární struktury – dimerizace

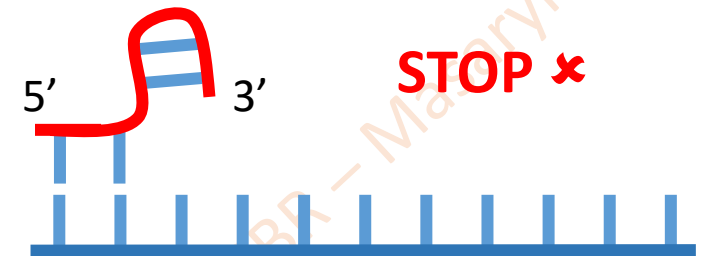
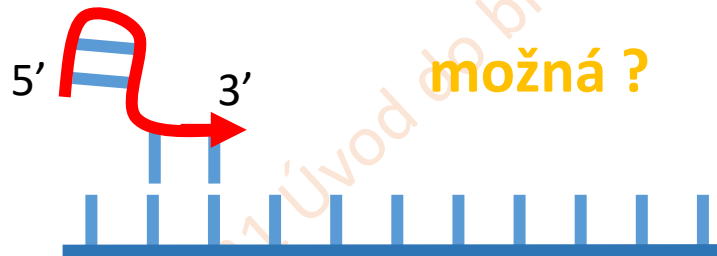
- Při PCR používáme typicky dvojici primerů. Pokud jsou navzájem komplementární části v jednom a druhém z nich, mohou tvořit tzv. **heterodimer**



- Pro stabilitu heterodimeru platí totéž, co pro homodimer.
- Někdy může vznikat několik různých dimerů

Sekundární struktury

- Tvorba sekundárních struktur DNA je vratná – systém se nachází v rovnováze, **zastoupení sekundární struktury závisí na podmínkách**
- Sekundární struktury brání nasedání primeru na templátovou DNA
- Obzvláště **nevhodná** je tvorba sekundárních struktur na **3' konci** primeru – blokují prodlužování primeru = syntéza DNA neprobíhá

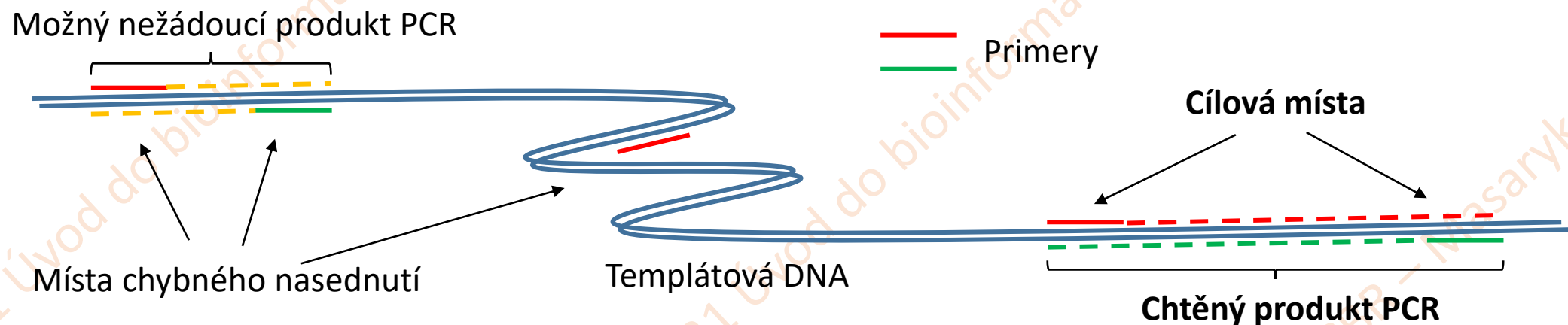


Specifita primeru

- Primer většinou cílíme na konkrétní místo na DNA
- Templátová DNA je většinou výrazně delší než cílová sekvence
- Sekvence je volena tak, aby možnost nasednutí na jiné než požadované místo byla co nejmenší
 - Prodloužení primeru zvyšuje specifitu
 - Posun cílové sekvence primeru vpřed/vzad, je-li to možné
 - Snaha nezařazovat na 3 konec G+C bohatou sekvenci
- Pokud je známa sekvence vstupní DNA (plazmid, genom), je vhodné otestovat možná místa „falešného nasedání primeru“ (***false priming sites***)

Specifita primeru – false priming sites

- Nasednutí na nežádoucí místa snižuje množství primerů pro požadovanou reakci
- V případě nežádoucího nasednutí dvou primerů v blízké oblasti může dojít k tvorbě **nežádoucích produktů**



Design primeru – obecné parametry

Optimální parametry pro namnožení kopií požadovaného úseku DNA

- Délka primeru: 17 – 28 bp
- Obsah G+C: 40 – 60 %
- Teplota tání T_m jednotlivých primerů: 50 – 65°C
- Rozdíl T_m levého a pravého primeru: $< 2^\circ\text{C}$
- Sekundární struktury na 3' konci: žádné
- Sekundární struktury: žádné nebo nestabilní
- Možná místa chybného nasedání: žádná

Design primeru – namnožení úseku

Potřebujeme-li **namnožit** konkrétní úsek DNA (např. gen):

- Začátek a konec sekvence je striktně daný – pozici primerů lze měnit jen omezeně

Možný rozsah pozic levého primeru



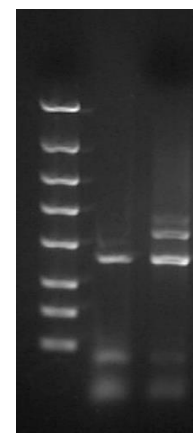
Cílový úsek

```
...TTCCCCGCGGATGCCTACCGCTCGGGGAACATCGACATCTCTGTGTTCTTCCAAGCTAGCGGCGTCTCCTTGCAGCAGTTGGTCCATCACGGTTGGATGAATGGATCTCCGGCAC...  
...AAGGGGGCGCCTACGGATGGCGAGCCCCTTGTAGCTGTAGAGACACAAGAAGGTTTCGATCGCCGCAGAGGAACGTCGTCAACCAGGTAGTGCCAACCTACTTACCTAGAGGCCGTG...
```



Možný rozsah pozic pravého primeru

- Špatné nasednutí primerů není žádoucí ale nemusí být kritické – případné nežádoucí produkty lze často odseparovat



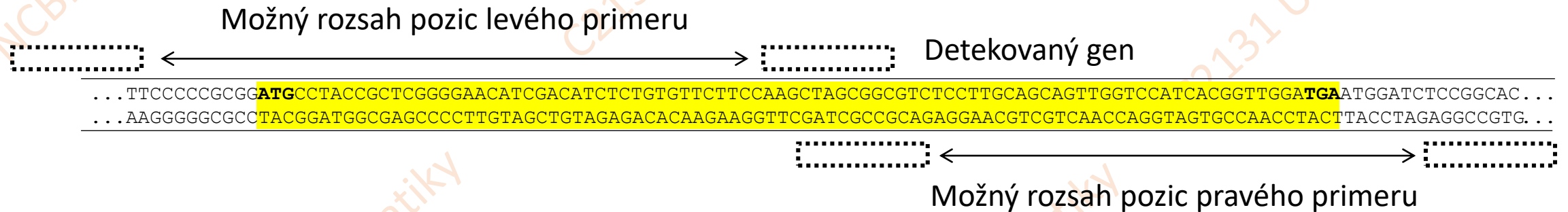
Nežádoucí produkty

Žádaný produkt

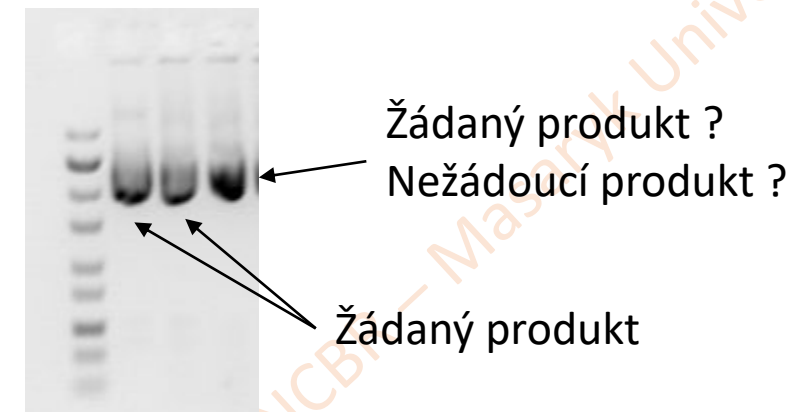
Design primeru – detekce genu

Je-li cílem **určit**, zda je či není **přítomen** gen (varianta genu):

- Primery mohou nasedat jak uvnitř genu tak vně – velká variabilita



- Možné nežádoucí produkty jsou rizikové – možnost falešně pozitivních výsledků

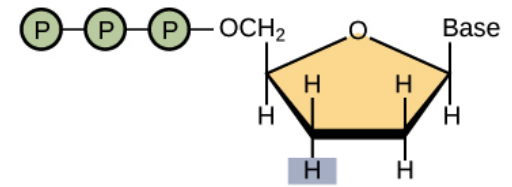


Možnosti modifikace primerů – aplikace

Velké množství různých aplikací. Např.:

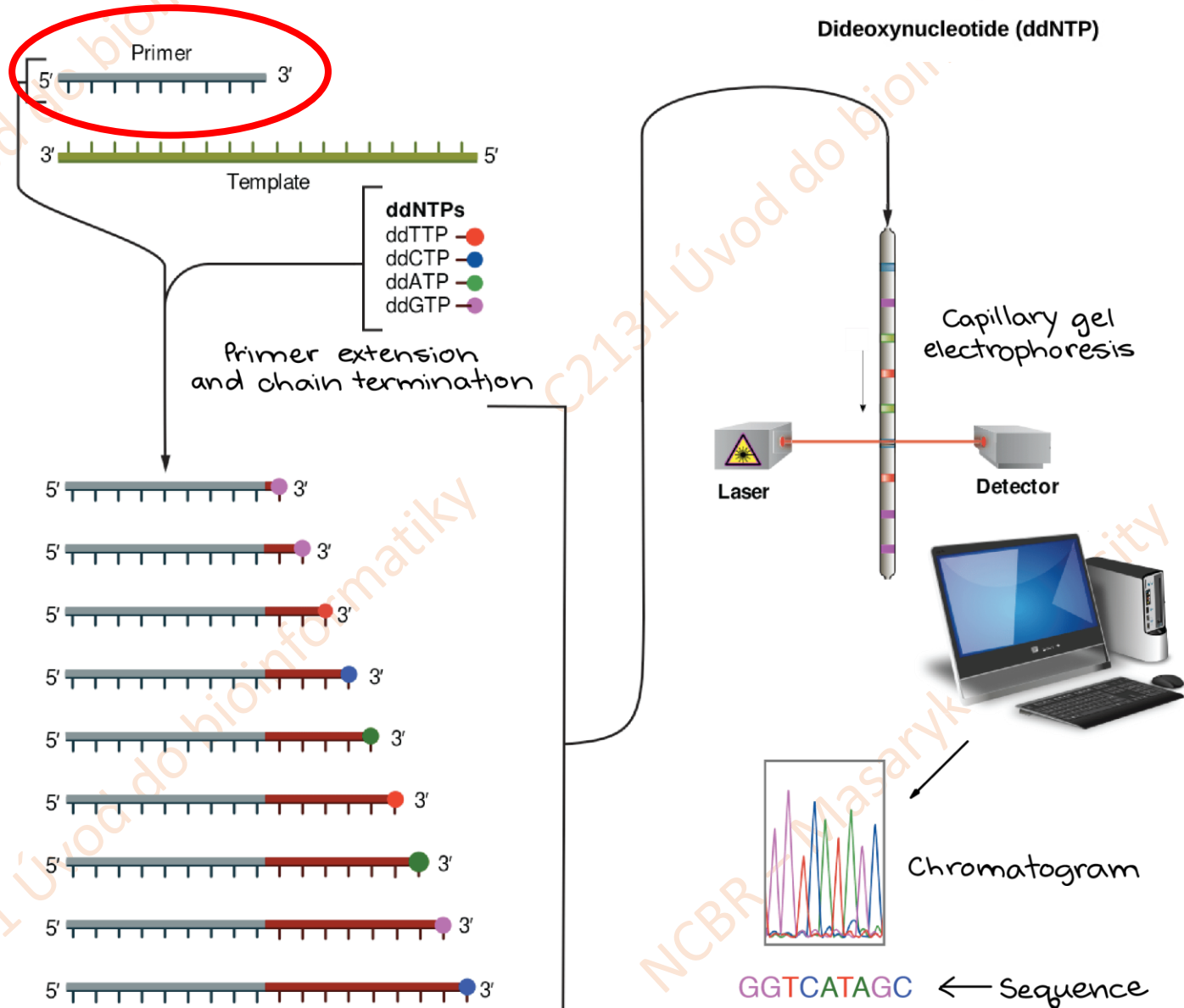
- **Sekvence DNA** – použití jediného primeru
- **Mutageneze** – použití primeru se změněnou sekvencí
- **Modifikace množené DNA** – primery z několika částí z nichž jedna je komplementární k templátové DNA, zatímco zbytek má jinou funkci (např. usnadňuje vložení namnoženého genu do cílové molekuly DNA nebo ukotvení na povrch)

Sekvence nukleových kyselin



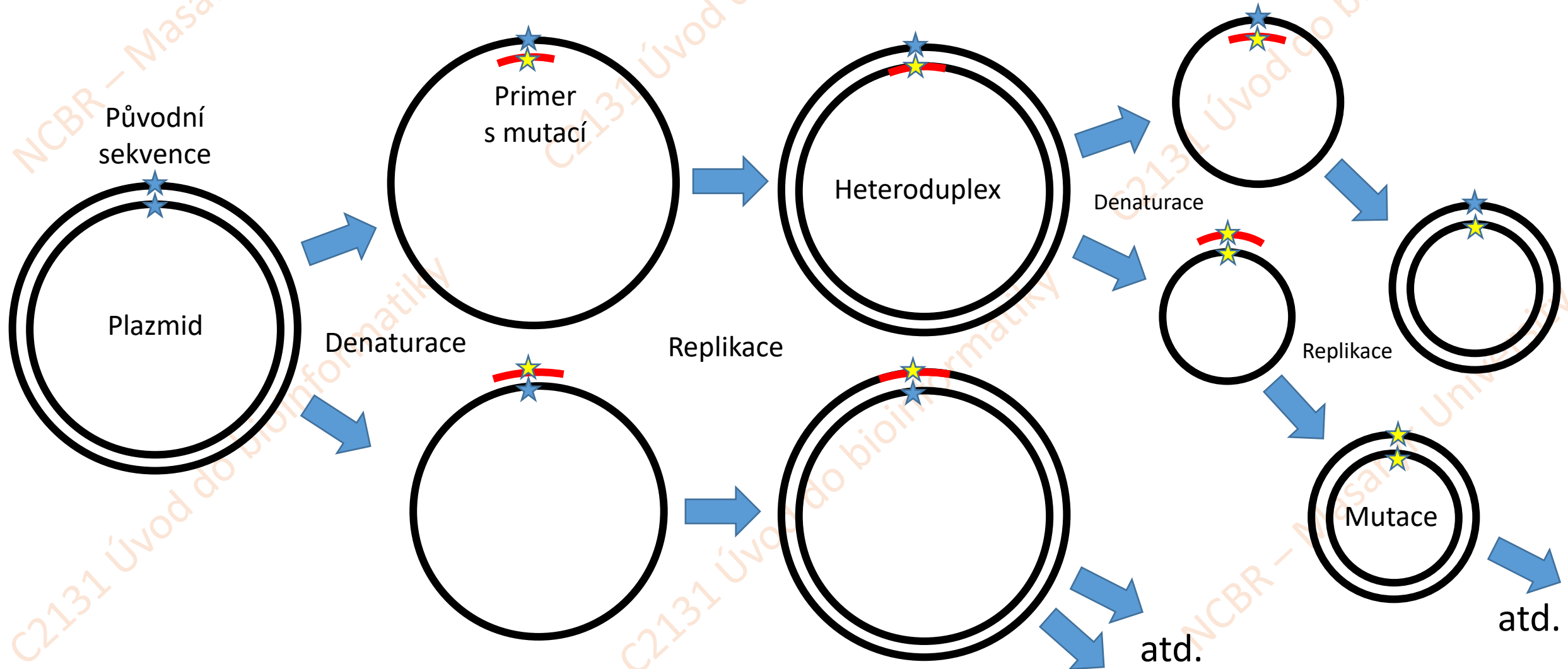
Sangerova sekvenace

- Zavedena 1977
- PCR syntéza řetězce DNA s náhodným vložením značených dideoxynukleotidů
- Separace produktů pomocí gelové elektroforézy



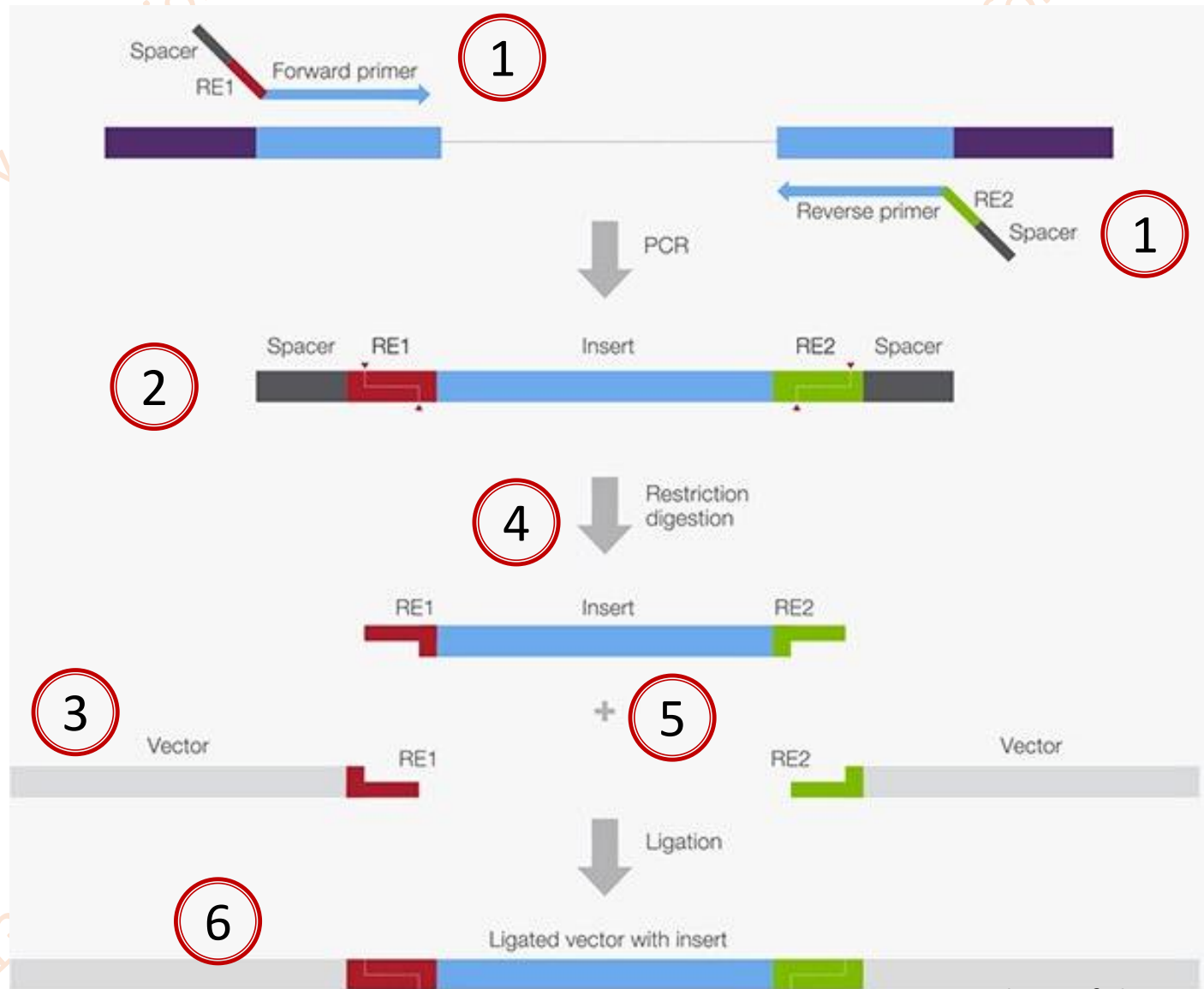
Mutageneze pomocí PCR

Dvojice primerů s vloženou mutací + plazmid nesoucí cílovou sekvencí



Modifikace DNA – schéma klonování

1. Primery obsahují kromě komplementární sekvence ještě sekvenci pro restriční enzym
2. Vložení restričních míst na okraj genu pomocí PCR
3. Upravený plazmid (vektor) se stejnými restričními místy
4. Štěpení produktu PCR a cílového plazmidu pomocí restričních endonukleas
5. Smíchání naštěpených molekul DNA
6. Spojení (ligace) plazmidu a genu



SW nástroje pro design primeru

- Navržení dobrých primerů je náročné – často se používají **bioinformatické nástroje**
- Placené – většinou součástí komplexních programů
- Omezeně dostupné (registrace)
- Volně dostupné – poměrně velké množství, různé aplikace

Oligo Analysis

<https://www.euofinsgenomics.eu/en/ecom/tools/oligo-analysis/>

- Jednoduchá analýza základních parametrů oligonukleotidů (primerů)

OLIGO ANALYSIS TOOL

Specify name and sequence of the oligo you would like to analyse.

Oligo Information

Type: DNA

Name: primer_left

Sequence: CCGATCAAATCACAGGTCA

Length: 19 mer MW: 5765.79 [g/mol]

Properties: OD Calc | Dilution Calc | Self-Dimer Check | PCR Check

Calculate the physical properties like GC content, T_m and extinction coefficient of your oligo sequence as well as reverse and complement sequences.

Analyse

Calculate

Result

Sequence (5' -> 3'):	CCG ATC AAA TCA CAG GTC A
Reverse complement (5' -> 3'):	TGA CCT GTG ATT TGA TCG C
Reverse sequence (5' -> 3'):	ACT GGA CAC TAA ACT AGC C
Complement sequence (5' -> 3'):	GGC TAG TTT AGT GTC CAG T
Base composition:	A x 7, C x 6, G x 3, T x 3, U x 0, Wobbles x 0
GC content [%]:	47.4 %
Melting temperature [°C]:	54.5 °C
Extinction coefficient:	213,500.00 L x mol ⁻¹ x cm ⁻¹

Order as

Velikost

G+C obsah

T_m

Homodimer

OLIGO ANALYSIS TOOL

Specify name and sequence of the oligo you would like to analyse.

Oligo Information

Type: DNA

Name: primer_left

Sequence: CCGATCAAATCACAGGTCA

Length: 19 mer MW: 5765.79 [g/mol]

Properties: OD Calc | Dilution Calc | Self-Dimer Check | PCR Check

Check if a self-dimer can be formed using this primer sequence in your PCR. Self-dimers are formed by intermolecular interactions between same sense primers, where the primer is homologous to itself.

Analyse

Check Self-Dimer

Result

Max annealing score (opt. <= 14): 10

Maximum stack size: 4

```

5'      CCGATCAAATCACAGGTCA
      ||||
3'  ACTGGACACTAAACTAGCC
    
```

Order as

Primer 3

<http://bioinfo.ut.ee/primer3/>

- Velké množství nastavitelných parametrů pro návrh primerů

Primer3web version 4.1.0 - Pick primers from a DNA sequence.

Select the **Task** for primer selection

[Template masking](#) before primer design ([available species](#))

Select species Nucleotides to mask in 5' direction
Primer failure rate cutoff Nucleotides to mask in 3' direction

Paste source sequence below (5'->3', string of ACGTnacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (very important for [Mispriming Library \(repeat library\)](#))

Pick left primer, or use left primer below Pick hybridization probe (internal oligo), or use oligo below Pick right primer, or use right primer below (5' to 3' on opposite strand)

[Sequence Id](#) A string to identify your output.

[Targets](#) E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or mark the [source sequence](#) with [and]: e.g. ...ATCT[CCCC]TCAT...

[Overlap Junction List](#) E.g. 27 requires one primer to overlap the junction between positions 27 and 28. Or mark the [source sequence](#) with -: e.g. ...ATCTAC-TGTCAT..

[Excluded Regions](#) E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the [source sequence](#) with < and >: e.g. ...A<7TTC...TCT>...

[Pair OK Region List](#) See manual for help.

[Included Region](#) E.g. 20,400: only pick primers in the 400 base region starting at position 20. Or use { and } in the [source sequence](#) to mark the beginning and end of the included region: e.g. in ATC{TTC...TCT}AT the included region is TTC...TCT.

disclaimer **code**

General Primer Picking Conditions

Upload the settings from a file Structure reversion

Primer Size	Min	18	Opt	20	Max	22
Primer Gap	Min	5	Opt	5	Max	5
Primer Gap	Min	100000	Opt	50	Max	1000000
Primer GC	Min	50.0	Opt	50.0	Max	50.0

Primer Tm Difference: Side of strand/oligo/primer:

Product Size Range: Number to Return: Max % Similarity: Max Library Mispriming: Pair Max Library Mispriming:

Thermodynamic Secondary Structure Alignments

HH Self Self Complementarity	0.0	HH Self Complementarity	0.00
HH Max 3 Self Complementarity	0.0	HH Max 3 Self Complementarity	0.00
HH Max Hairpin Complementarity	0.0	HH Max Hairpin Complementarity	0.00
HH Max 3 Pair Self Complementarity	0.0	HH Max 3 Pair Self Complementarity	0.00
HH Max Primer Hairpin	0.0	HH Max Primer Hairpin	0.00

Thermodynamic Template Alignments

HH Max Library Mispriming	12.00	HH Max Template Mispriming	12.00
HH Pair Max Template Mispriming	50.00	HH Pair Max Template Mispriming	50.00

Objective Function Penalty Weights for Primer Pairs

HH Self Complementarity	0.0	HH Self Complementarity	0.0
HH Max 3 Self Complementarity	0.0	HH Max 3 Self Complementarity	0.0
HH Max Hairpin Complementarity	0.0	HH Max Hairpin Complementarity	0.0
HH Max 3 Pair Self Complementarity	0.0	HH Max 3 Pair Self Complementarity	0.0
HH Max Primer Hairpin	0.0	HH Max Primer Hairpin	0.0

Objective Function Penalty Weights for Internal Oligo (Hyb Oligo)

Internal Oligo Size	Min	18	Opt	20	Max	22
Internal Oligo Gap	Min	5	Opt	5	Max	5
Internal Oligo GC	Min	50.0	Opt	50.0	Max	50.0

Primer-BLAST

- Jednoduché rozhraní

<https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>

- Lze využít pro **návrh primerů** (využívá postupy Primeru 3), ale též pro **kontrolu špatného nasedání** vlastních primerů

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

Primer-BLAST

A tool for finding specific primers

Finding primers specific to your PCR template (using Primer3 and BLAST).

PCR Template [Reset page](#) [Save search parameters](#) [Retrieve recent results](#) [Publication](#) [Tips for finding specific primers](#)

Enter accession, gi, or FASTA sequence (A refseq record is preferred) [Clear](#)

Vložení sekvence

Or, upload FASTA file Soubor nevybrán

Range

Forward primer [Clear](#)

Reverse primer

Primer Parameters

Use my own forward primer (5'->3' on plus strand) [Clear](#)

Use my own reverse primer (5'->3' on minus strand) [Clear](#)

PCR product size Min 70 Max 1000

of primers to return

Primer melting temperatures (T_m) Min 57.0 Opt 60.0 Max 63.0 Max T_m difference [Clear](#)

Primer-BLAST: Specifita primeru

Výběr databáze
pro kontrolu
falešného
nasedání

Specifikace
organismu

Exon/intron selection

A refseq mRNA sequence as PCR template input is required for options in the section

Exon junction span: No preference

Exon junction match: Exon at 5' side: 7, Exon at 3' side: 4

Minimal number of bases that must anneal to exons at the 5' or 3' side of the junction

Intron inclusion: Primer pair must be separated by at least one intron on the corresponding genomic DNA

Intron length range: Min: 1000, Max: 1000000

Primer Pair Specificity Checking Parameters

Specificity check: Enable search for primer pairs specific to the intended PCR template

Search mode: Automatic

Database: Refseq mRNA

Exclusion: Exclude predicted Refseq transcripts (accession with XM, XR prefix) Exclude uncultured/environmental sample sequences

Organism: Homo sapiens

Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggestion list as you type.

Add more organisms

Entrez query (optional):

Primer specificity stringency: Primer must have at least 2 total mismatches to unintended targets, including at least 2 mismatches within the last 5 bps at the 3' end. Ignore targets that have 6 or more mismatches to the primer.

Max target size: 4000

Allow splice variants: Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input)

Get Primers Show results in a new window Use new graphic view

Advanced parameters

Primer-BLAST: Specifita primeru

Primer-BLAST >> JOB ID:WICFVGzxYVIGZ2RiaQJAUBMZUWI-Ckp_Pw

Primer-BLAST Results

Input PCR template none
Specificity of primers Target templates were found in selected database: Refseq mRNA (Organism limited to Homo sapiens)
Other reports [Search Summary](#)

Detailed primer reports

Primer pair 1 **Základní charakterizace navržených (nebo vložených) primerů**

	Sequence (5'->3')	Length	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CCGATCAAATCACAGGT	17	51.05	47.06	4.00	2.00
Reverse primer	CCGATTAATCTCGGAT	17	47.72	41.18	7.00	4.00

Products on target templates

>[XM_024446112.1](#) PREDICTED: Homo sapiens phosphodiesterase 4D (PDE4D), transcript variant X7, mRNA

```
product length = 2106
Forward primer 1   CCGATCAAATCACAGGT 17
Template       292 G.A.GT..... 308

Forward primer 1   CCGATCAAATCACAGGT 17
Template       2397 GAA.G..... 2381
```

Program detekuje i nedokonalé nasednutí

>[XM_024446110.1](#) PREDICTED: Homo sapiens phosphodiesterase 4D (PDE4D), transcript variant X4, mRNA

```
product length = 2106
Forward primer 1   CCGATCAAATCACAGGT 17
Template       292 G.A.GT..... 308

Forward primer 1   CCGATCAAATCACAGGT 17
Template       2397 GAA.G..... 2381
```

SW nástroje – plusy a mínusy

- **Výhody:**

- Usnadnění návrhu primerů
- Často minimální nároky na znalost uživatele
- U některých propojení s komerční firmou – možnost objednat syntézu

- **Nevýhody:**

- Pro některé cíle automatický design selhává
- Použití bez znalostí parametrů může vést k chybám

Závěrem

- Hledání restrikčních míst i design primerů je možný **manuálně**, ale **SW nástroje** práci významně usnadňují
- **Využití** těchto postupů je velmi široké a nové aplikace se objevují každý rok
- Pochopení **principů** těchto technik je vysoce žádoucí pro každého pracovníka v oblasti biochemie a molekulární biologie

