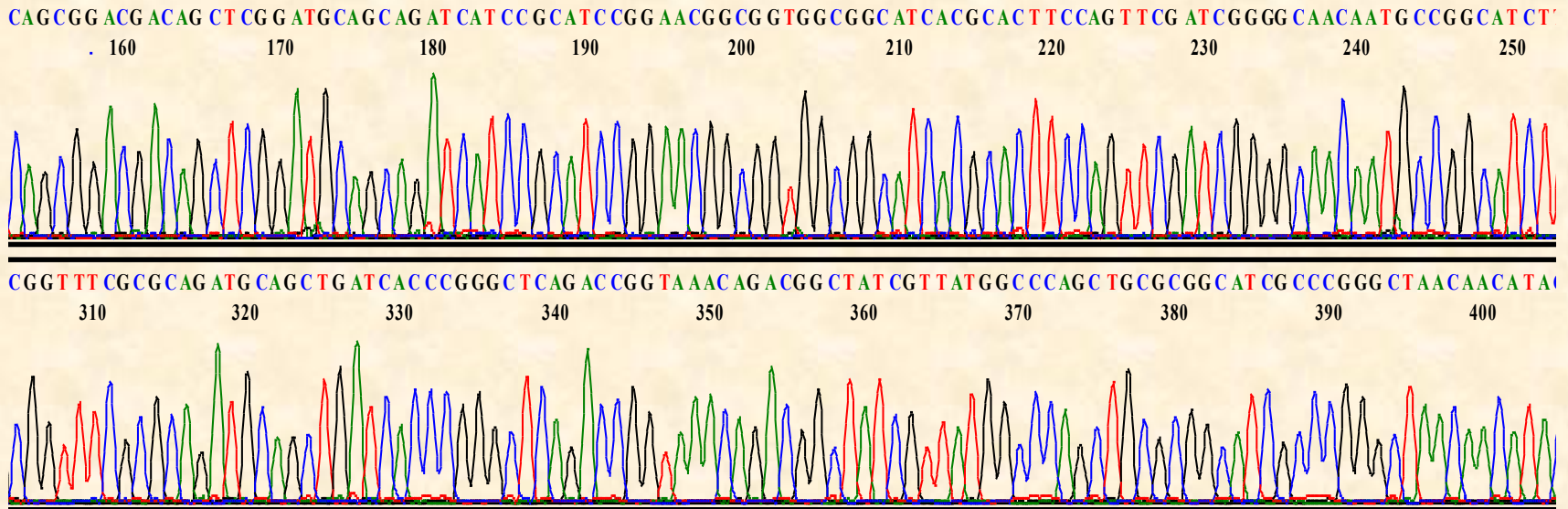


# Predikce genů



GATAGCGTAATGATCGGCTGGCTGCCGATTTTCATGCTGGTTTCCCAACGAAAATAACCGCTCACGGTGCCATCACGATCGCACACCGCAAATCGGCGG  
TACAGGTGGTCGCGCCCGCCGACACATCGCTGCGCCAATAATGATCTTTCAGCGGACGACAGCTCGGATGCAGCAGATCATCCGCATCCGGAACGGC  
GGTGGCGGCATCACGCACCTCCAGTTCGATCGGGGCAACAATGCCGGCATCTTTCAGGGCAAAGCGAATAAACAGCACGCTCACCTTCCGCGGCAGCGCC  
AGCGCGGTTTCGCGCAGATGCAGCTGATCACCCGGGCTCAGACCGGTAAACAGACGGCTATCGTTATGGCCAGCTGCGCGGCATCGCCCGGGCTAACAA  
CATACAGGTGGCGACCATCAATCACGGTCGGGGCGCCGGATCACGGCTGGCTTCCGGATAGGCGCTCAGCAGGGTAACGGCATCCACAATCACCAGCAT

GATAGCGTAATGATCGGCTGGCTGCCGATTTTCATGCTGGTTTCCCAACGAAAAAACCCTCACGGTGCCATCACGATCGCACACCCGAAAAATCGGCGG  
TACAGGTGGTCGCGCCCGCCGCCAGCACATCGCTGCGCCAATAATGATCTTTTCAGCGGACGACAGCTCGGATGCAGCAGATCATCCGCATCCGGAACGGC  
GGTGGCGGCATCACGCACCTCCAGTTCGATCGGGGCAACAATGCCGGCATTTTCAGGGCAAAGCGAATAAACAGCACGCTCACTTCGCGCGCAGCGCC  
AGCGCGGTTTCGCGCAGATGCAGCTGATCACCCGGGCTCAGACC GGTAACAGACGGCTATCGTTATGGCCCAGCTGCGCGGCATCGCCCGGGCTAACAA  
CATACAGGTGGCGACCATCAATCACGGTCGGGGCGGCCGGATCACGGCTGGCTTCCGGATAGGCGCTCAGCAGGGTAACGGCATCCACAAATCACCAGCAT

**„Syrové“ sekvence DNA**

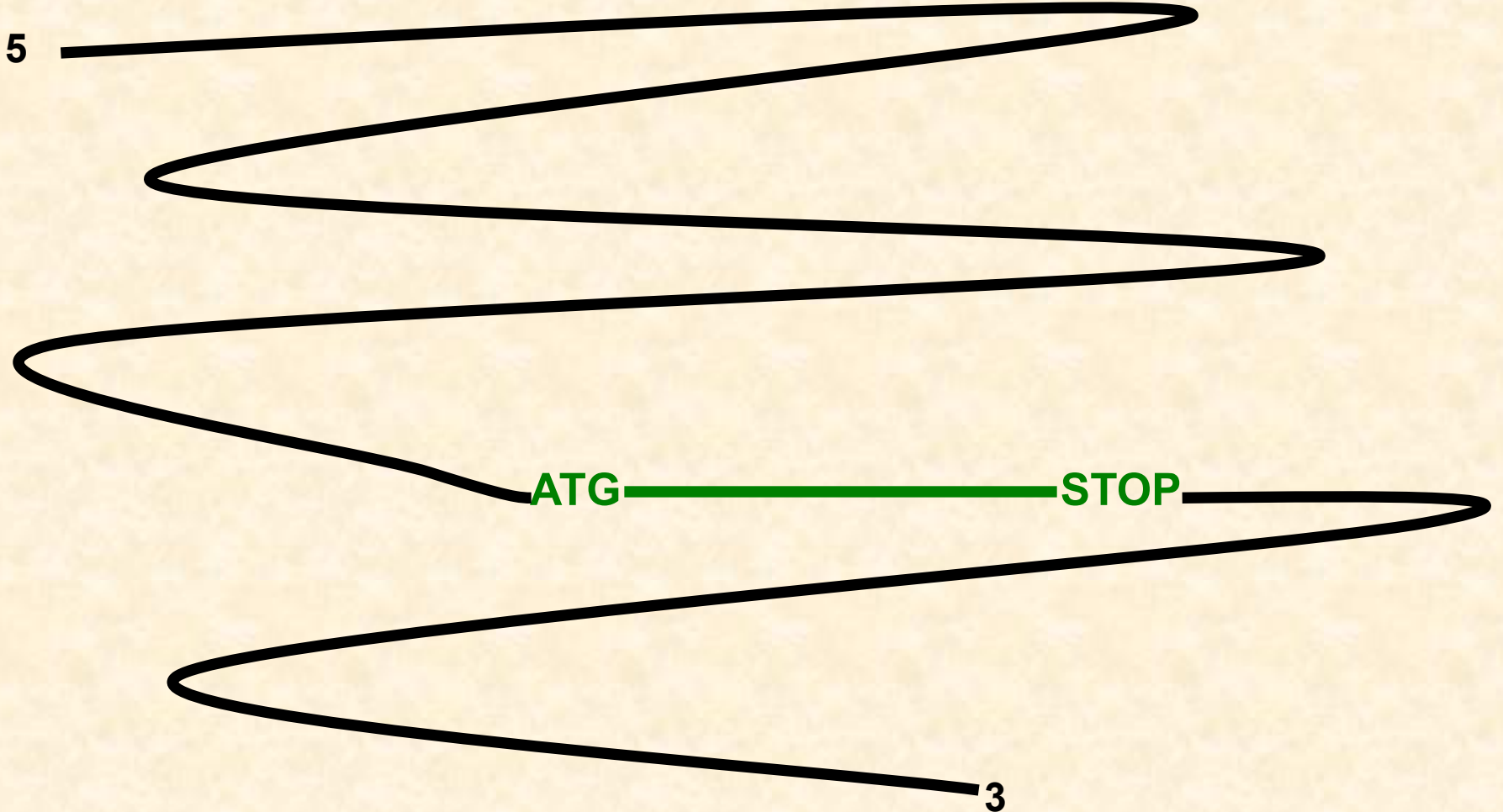


**Identifikace a anotace genů a proteinů**

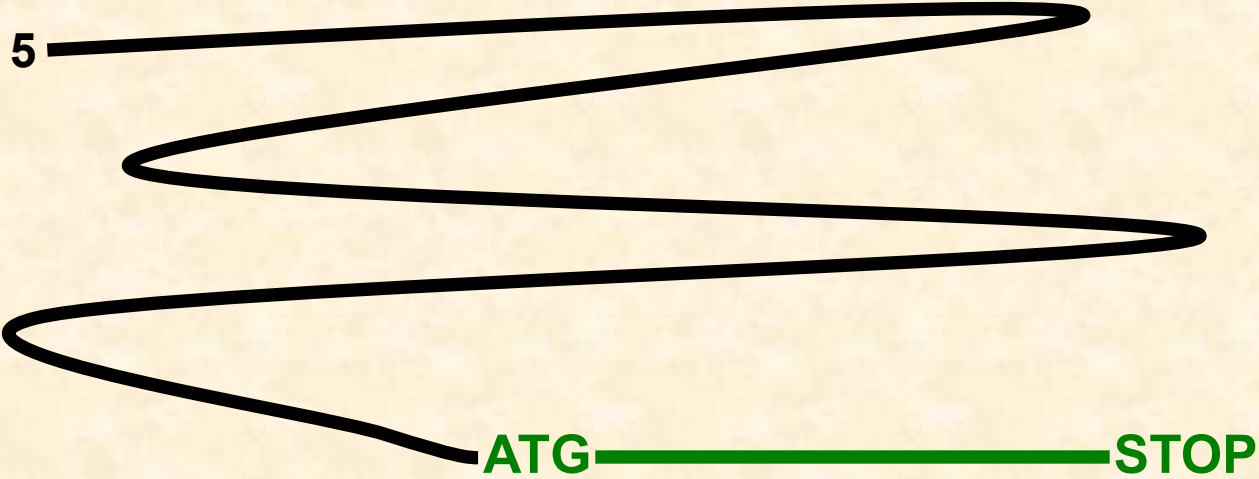
Table 1  
Software commonly used for bacterial genome annotation and comparison

<i>DNA level annotation</i>		
GeneMark	<a href="http://exon.gatech.edu/genemark/">http://exon.gatech.edu/genemark/</a>	Protein gene prediction
Glimmer	<a href="http://www.genomics.jhu.edu/Glimmer/">http://www.genomics.jhu.edu/Glimmer/</a>	Protein gene prediction
SHOW	<a href="http://genome.jouy.inra.fr/ssb/SHOW/">http://genome.jouy.inra.fr/ssb/SHOW/</a>	Protein gene prediction
tRNAscan-SE	<a href="http://lowelab.ucsc.edu/tRNAscan-SE/">http://lowelab.ucsc.edu/tRNAscan-SE/</a>	tRNA gene prediction
RNAmmer	<a href="http://www.cbs.dtu.dk/services/RNAmmer/">http://www.cbs.dtu.dk/services/RNAmmer/</a>	rRNA gene prediction
RepSeek	<a href="http://www.abi.snv.jussieu.fr/%98public/RepSeek/">http://www.abi.snv.jussieu.fr/%98public/RepSeek/</a>	Search for approximate repeats in complete DNA sequences
IslandPath	<a href="http://www.pathogenomics.sfu.ca/islandpath/">http://www.pathogenomics.sfu.ca/islandpath/</a>	Identification of genomic islands
<i>Protein level annotation</i>		
BLAST	<a href="http://www.ncbi.nlm.nih.gov/blast/">http://www.ncbi.nlm.nih.gov/blast/</a>	Compare a novel sequence with those contained in nucleotide and protein databases
InterProScan	<a href="http://www.ebi.ac.uk/InterProScan/">http://www.ebi.ac.uk/InterProScan/</a>	Search for domains/motifs in the InterPro database
COGNITOR	<a href="http://www.ncbi.nlm.nih.gov/COG/old/xognitor.html">http://www.ncbi.nlm.nih.gov/COG/old/xognitor.html</a>	Compare a query sequence to the COG (Cluster of Orthologous Groups of proteins) database
PRIAM	<a href="http://bioinfo.genopole-toulouse.prd.fr/priam/">http://bioinfo.genopole-toulouse.prd.fr/priam/</a>	Detection of enzymatic function in a fully sequenced genome, based on all sequences available in the ENZYME database
GOAnno	<a href="http://bips.u-strasbg.fr/GOAnno/">http://bips.u-strasbg.fr/GOAnno/</a>	BLAST search on the Gene Ontology database
PSORTb	<a href="http://www.psort.org/psortb/">http://www.psort.org/psortb/</a>	Prediction of bacterial protein subcellular localization
TMHMM	<a href="http://www.cbs.dtu.dk/services/TMHMM/">http://www.cbs.dtu.dk/services/TMHMM/</a>	Prediction of transmembrane helices in protein sequences
SignalP	<a href="http://www.cbs.dtu.dk/services/SignalP/">http://www.cbs.dtu.dk/services/SignalP/</a>	Prediction of signal peptide cleavage sites in protein sequences
<i>Comparative genomic tools</i>		
Mauve	<a href="http://gel.ahabs.wisc.edu/mauve/">http://gel.ahabs.wisc.edu/mauve/</a>	Multiple genome alignments in the presence of large-scale evolutionary events
MOSAIC	<a href="http://mig.jouy.inra.fr/mig/mig_eng/presentation/project/mosaic">http://mig.jouy.inra.fr/mig/mig_eng/presentation/project/mosaic</a>	Define the set of backbones and loops in closely related bacterial genomes
ACT	<a href="http://www.sanger.ac.uk/Software/ACT/">http://www.sanger.ac.uk/Software/ACT/</a>	Comparative genome analysis and visualization tools for multiple genome alignments
CGAT	<a href="http://mbgd.genome.ad.jp/CGAT/">http://mbgd.genome.ad.jp/CGAT/</a>	
MaGe	<a href="http://www.genoscope.cns.fr/agc/mage/">http://www.genoscope.cns.fr/agc/mage/</a>	Computation of gene order conservation (syntenies) between available bacterial genomes
Pathologic	<a href="http://biocyc.org/">http://biocyc.org/</a>	Metabolic network reconstruction and comparative pathway analysis
PUMA2	<a href="http://compbio.mcs.anl.gov/puma2/">http://compbio.mcs.anl.gov/puma2/</a>	Metabolic pathway reconstruction
The SEED	<a href="http://theseed.ucchicago.edu/FIG/">http://theseed.ucchicago.edu/FIG/</a>	Comparative analysis and annotation tools using the subsystem approach
STRING	<a href="http://string.embl.de/">http://string.embl.de/</a>	Search Tool for the Retrieval of Interacting Proteins
PyPhy	<a href="http://www.cbs.dtu.dk/staff/thomas/pyphy/">http://www.cbs.dtu.dk/staff/thomas/pyphy/</a>	Reconstruction of phylogenetic relationships of complete microbial genomes
HoSeqI	<a href="http://pbil.univ-lyon1.fr/software/HoSeqI/">http://pbil.univ-lyon1.fr/software/HoSeqI/</a>	Automatically assign sequences to homologous gene families from the HOGENOM database

# Gen



# Gen



Met

DNA

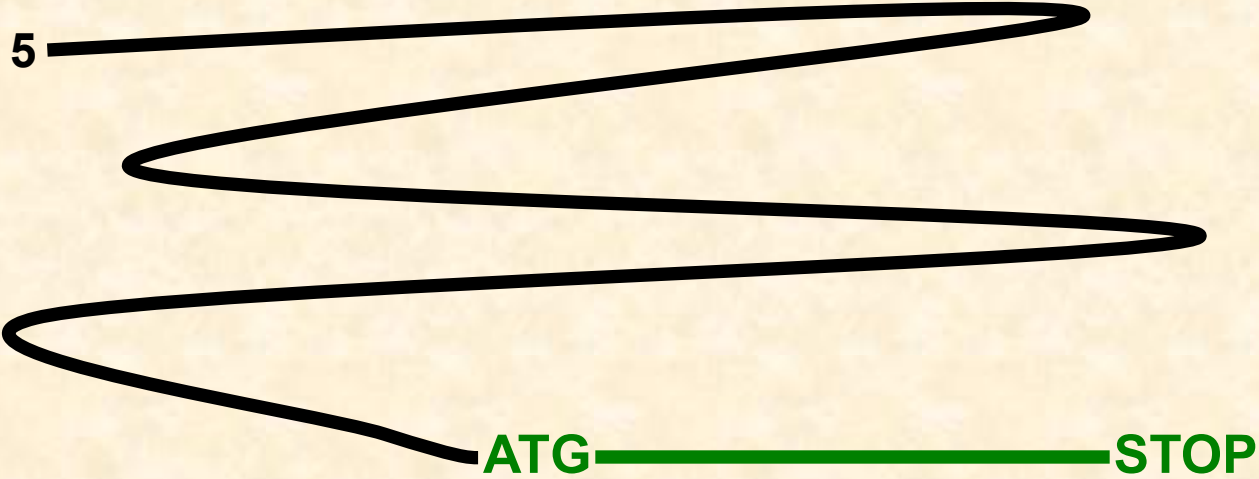


RNA



Protein

# Gen



DNA



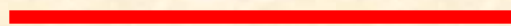
RNA



Protein



Met



# Predikce genů kódujících proteiny

- **Prokaryotické geny**

Nepřerušované úseky DNA mezi **startovním kodonem** (ATG, GTG, TTG, CTG) a **stop kodonem** (TAA, TGA, TAG).

- **Úsek kóduje protein, který je podobný již dříve popsanému proteinu (prohledávání DATABÁZÍ pomocí ALIGNMENTU).**

# Překlad DNA sekvence

- **ExPASy**

<http://www.expasy.org/tools/dna.html>

- **ORF Finder (NCBI)**

<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>



nonpolar polar basic acidic (stop codon)

# Překlad DNA sekvence

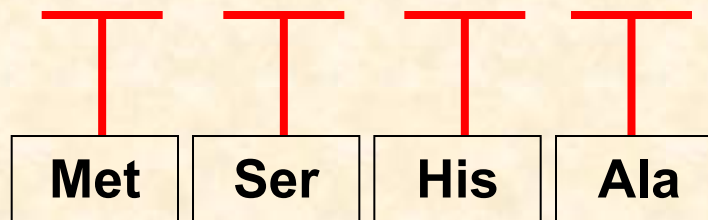
The table shows the 64 codons and the amino acid for each. The **direction** of the mRNA is 5' to 3'.

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

# Překlad DNA sekvence

ATG ————— STOP

ATG    TCG    CAT    GCC    STOP



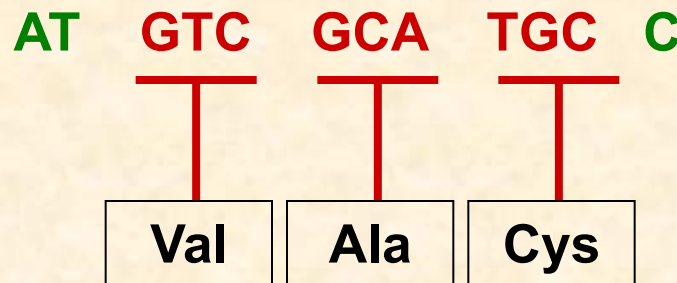
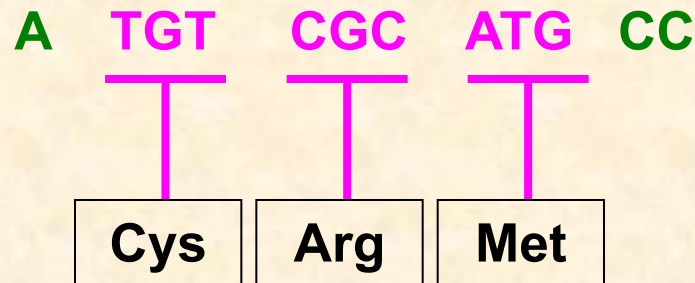
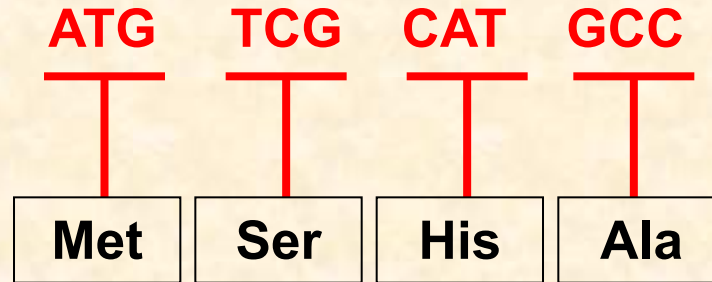
N-konec

Met —————

C-konec

NH<sub>2</sub> ————— COOH

↓ ↓ ↓  
**ATGTCGCATGCC**



*Čtení tripletů závisí na tom, u kterého nukleotidu stanovíme počátek čtení.*

**ATGCGCAGGAATGCATAG** sekvence DNA

<b>Met</b>	<b>His</b>	<b>?</b>
------------	------------	----------

**Protein 2**

**ATGCGCAGGAATGCATAG**

<b>Met</b>	<b>Arg</b>	<b>Arg</b>	<b>Asn</b>	<b>Ala</b>	<b>Stop</b>
------------	------------	------------	------------	------------	-------------

**Protein 1**



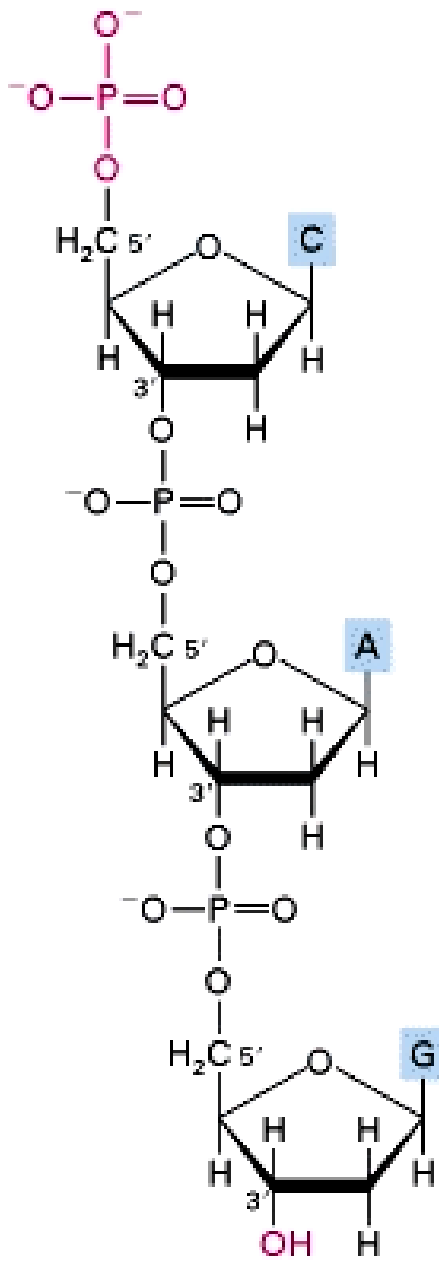


# Opravdu ORF kóduje protein?

- ORF kóduje protein, který je podobný již dříve popsanému proteinu (prohledávání DATABÁZÍ pomocí ALIGNMENTU) = **nejspolehlivější ověření.**
- **Nástroje pro překlad DNA jsou propojeny s prohledáváním databází.**

# Polynukleotidový řetězec

5' end



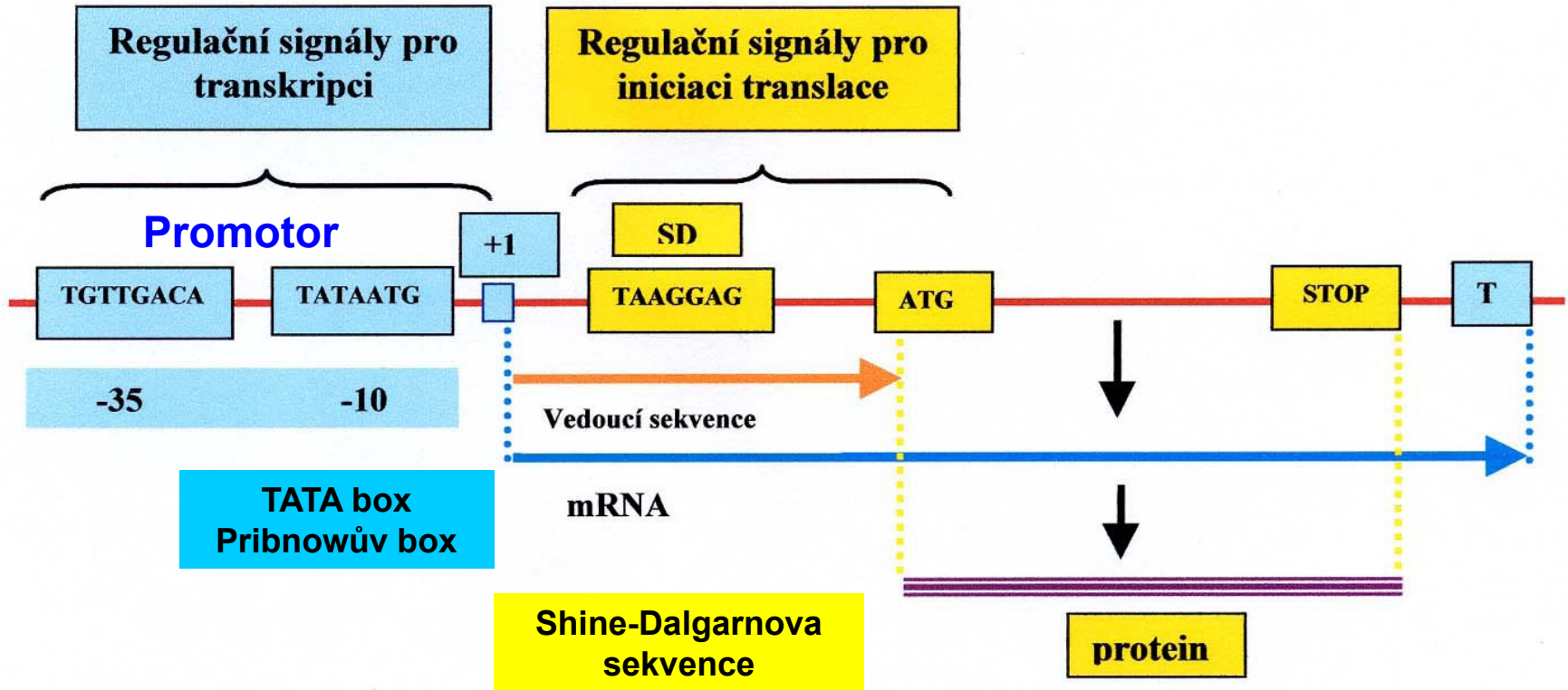
3' end

**5 -P-C-A-G-3 -OH**

**5** ATGCGTTGACGTTGCACGTGCAA  
GTCGCAGTCGATGCGATGCTGAC  
GTACGTGCGTACGATGCGTCGTA  
CGTGCTGACGTCGTACGT **3**

**5** → **3**

# Translační a transkripční signální sekvence



## Prokaryota