

C2131 Úvod do bioinformatiky
Masarykova univerzita, PřF

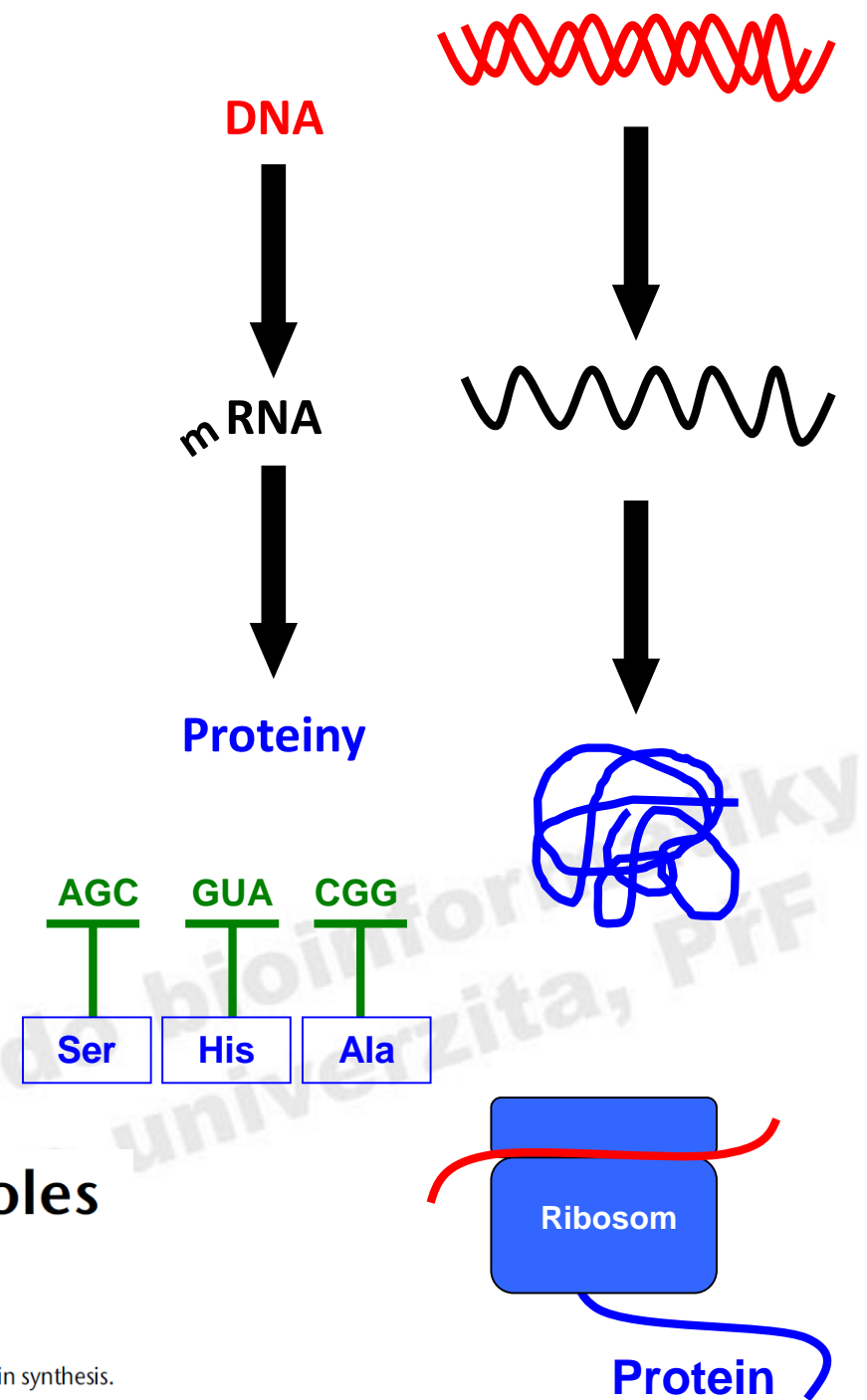
RNA bioinformatika

C2131 Úvod do bioinformatiky, jaro 2023

C2131 Úvod do bioinformatiky
Masarykova univerzita, PřF

RNA

- **RNA:** tři hlavní formy zapojené do proteosyntézy, mediátorová RNA (**mRNA**), ribozomální RNA (**rRNA**), transferová RNA (**tRNA**).
- **mRNA:** přenos informace pro syntézu proteinů.
- **rRNA:** nekódující RNA, složka ribozomů. Využívána pro studium evolučních vztahů. Nezbytná pro správnou funkci ribozomů, tj. evolučně konzervovaná.
- **tRNA:** transport aminokyselin pro proteosyntézu. Obsahují místo pro připojení aminokyseliny a antikodon pro interakci s mRNA.



Cellular RNAs: Varied Roles

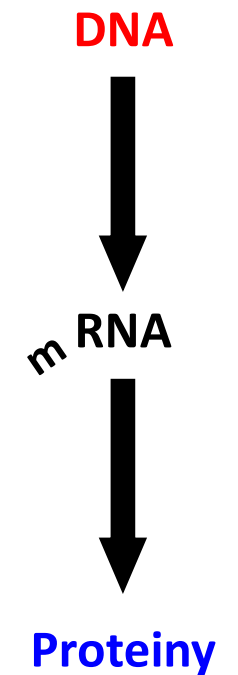
Jian Gu, Baylor College of Medicine, Houston, Texas, USA

Ram Reddy, Baylor College of Medicine, Houston, Texas, USA

In the cell, there are three major types of RNA directly involved in protein synthesis. In addition, many other cellular RNAs also play important functions.

mRNA – nástroje a databáze

- **Prokaryotická** mRNA je často **polycistronická** (multigenní, polygenní) – obsahuje přepisy několika genů. Transkripce je u prokaryot úzce spřažena s translací. Životnost bakteriální mRNA je krátká.
- **Eukaryotická** mRNA je **monocistronická**. Vzniká jako pre-mRNA (prekurzorová mRNA), podléhá **modifikacím** (tvorba komplexů s proteiny, úprava 5' konce čepičkou, polyadenylace 3' konce) a **sestřihu**.
- **Predikční nástroje**: predikce míst sestřihu (souvislost s predikcí genů) a dalších specifických míst a oblastí, predikce interakcí (s proteiny a další RNA), predikce modifikací, predikce lokalizace.



Messenger RNA in Eukaryotes

Göran Akusjärvi, *University of Uppsala, Uppsala, Sweden*

Jan-Peter Kreivi, *University of Uppsala, Uppsala, Sweden*

Svend Petersen-Mahrt, *London Research Institute, London, UK*

Posttranscriptional regulation of gene expression represents an important level at which eukaryotes can expand the coding capacity of their genomes. The concept that one gene makes one protein does not apply to higher eukaryotes. Thus, a eukaryotic cell can use alternative ribonucleic acid (RNA) splicing, alternative polyadenylation and RNA editing to produce hundreds or even several thousands of protein isoforms from a single gene.

Messenger RNA in Prokaryotes

Sidney R Kushner, *University of Georgia, Athens, Georgia, USA*

Messenger RNAs (mRNAs) are molecules that represent the intermediate step in converting the genetic information carried in a cell's DNA to functional proteins. Structural features of mRNAs control both how effectively their contained data are translated into functional proteins and how rapidly they are destroyed.

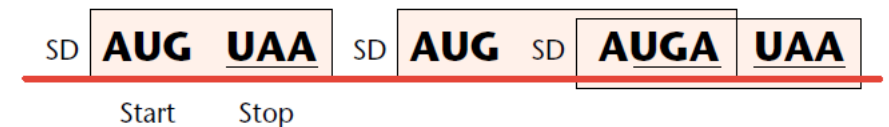
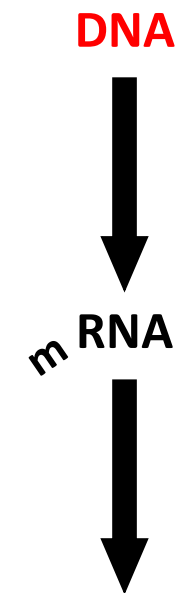


Figure 3 Prokaryotic mRNAs are polycistronic, containing multiple open reading frames (ORFs, boxes) where each initiation codon is downstream of the Shine–Dalgarno sequence (SD) and the region between the ORFs can vary; it may even overlap with the termination codon of the prior ORF.

mRNA – predikce míst sestřihu

SpliceFinder: ab initio prediction of splice sites using convolutional neural network

Ruohan Wang, Zishuai Wang, Jianping Wang* and Shuaicheng Li*



Abstract

Background: Identifying splice sites is a necessary step to analyze the location and structure of genes. Two dinucleotides, GT and AG, are highly frequent on splice sites, and many other patterns are also on splice sites with important biological functions. Meanwhile, the dinucleotides occur frequently at the sequences without splice sites, which makes the prediction prone to generate false positives. Most existing tools select all the sequences with the two dimers and then focus on distinguishing the true splice sites from those pseudo ones. Such an approach will lead to a decrease in false positives; however, it will result in non-canonical splice sites missing.

Related work

The existing splice site prediction tools work on data from either RNA sequences or DNA sequences. For RNA-seq based tools, TopHat [12], SpliceMap [13], and MapSplice [14] apply the alignment-based approach by mapping the reads from RNA-seq experiments to the reference genome and discovering the exon-exon junctions. The alignment-based approach makes it easier to avoid false positives since it relies on the junction signals, instead of patterns. However, the need for a reference genome limits its application. Recently, deep neural networks have been employed to predict splice sites from arbitrary pre-mRNA transcript sequences [15].

SpliceAI

„Compared with other state-of-the-art splice site prediction tools, *SpliceFinder* generates results in about half lower false positive while keeping recall higher than 0.8. Also, *SpliceFinder* captures the non-canonical splice sites.“

„We demonstrate that *MapSplice* achieves higher sensitivity and specificity than *TopHat* and *SpliceMap* on a set of simulated RNA-seq data. Experimental studies also support the accuracy of the algorithm.“

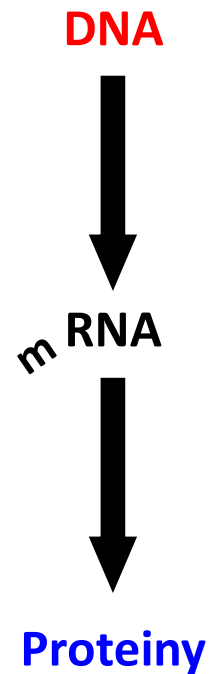
„Compared to current methods, *SpliceMap* can achieve 12% higher sensitivity without sacrificing specificity.“

„The *TopHat* pipeline is much faster than previous systems, mapping nearly 2.2 million reads per CPU hour, which is sufficient to process an entire RNA-Seq experiment in less than a day on a standard desktop computer.“



mRNA – predikce polyA (pA) míst

- Polyadenylace 3' konce – připojení poly(A) sekvence na 3' konec pre-mRNA. Štěpení poblíž polyadenylačního signálu, poté připojení 50-250 A.
- Význam pro stabilitu, export a translaci mRNA.
- Geny mohou mít více polyadenylačních míst (alternativní polyadenylace). 1 gen = více transkriptů (mRNA isoformy) s různou stabilitou a efektivitou translace; popřípadě mohou vznikat různé formy proteinů.



Genome analysis

DeepPASTA: deep neural network based polyadenylation site analysis

Deep neural network-based PolyA SiTe Analysis

Ashrafal Arefeen¹, Xinshu Xiao^{2,*} and Tao Jiang^{1,3,4,*}

¹Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA, ²Department of Integrative Biology and Physiology, University of California, Los Angeles, CA 90095, USA, ³Institute of Integrative Genome Biology, University of California, Riverside, CA 92521, USA and ⁴Bioinformatics Division, BNRIST, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

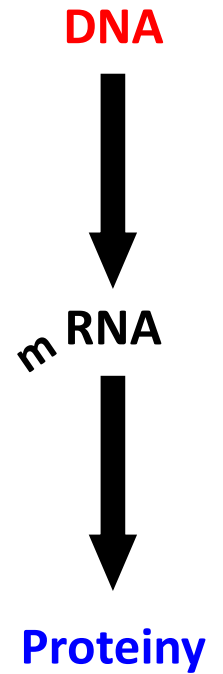


<https://acronymify.com/>

| Acronym | Expanded | Score |
|---------|--|-------|
| PROVE | Predikce úRovně Obědů V mEnze | 5.40 |
| PROVEN | Predikce úRovně Obědů V mENze | 5.40 |
| PROMO | Predikce úRovně Obědů eMpirickém pOzorování | 7.20 |
| MEANER | MEnze zAložeNá Na Empirickém pozoRování | 10.80 |
| MANNER | Menze zAložeNá Na Empirickém pozoRování | 10.80 |
| PROBER | Predikce úRovně OBědů Empirickém pozoRování | 10.80 |
| DROVE | preDikce úRovně Obědů V mEnze | 10.80 |
| DRONER | preDikce úRovně Obědů Na Empirickém pozoRování | 13.50 |
| RODEO | pRedikce úRovně oběDů Empirickém pOzorování | 14.40 |
| IRONER | predlkce úRovně Obědů Na Empirickém pozoRování | 15.00 |

mRNA – predikce mRNA:miRNA interakcí

- **miRNA** – microRNA. Krátké nekódující jednovláknové řetězce RNA. Regulace genové exprese – vazba na komplementární úseky mRNA vede ke snížení produkce proteinu (inhibice translace, destabilizace mRNA).
- Chybná (snížená x zvýšená) produkce microRNA pozorována u různých nemocí.
- Léčiva budoucnosti?



How close are miRNAs from clinical practice? A perspective on the diagnostic and therapeutic market

2019

Bonneau E.¹, Neveu B.², Kostantin E.³, Tsongalis G.J.⁴, De Guire V.²

¹ Department of Laboratory Medicine, Biochemistry Division, Saint-Eustache Hospital, Québec, Canada

² Biochemistry Division, Maisonneuve-Rosemont Hospital, Montreal, Québec, Canada

³ Department of Biochemistry, McGill University, Montreal, Québec, Canada

⁴ Department of Pathology and Laboratory Medicine, Geisel School of Medicine at Dartmouth and Dartmouth Hitchcock Medical Center, Lebanon, NH, United States of America

Fáze 1 – testování na malé skupině osob, nejčastěji zdravých dobrovolníků. Stanovení nejvyšší tolerované dávky, sledování vedlejších účinků.

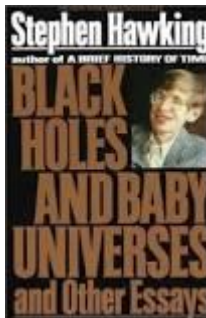
Fáze 2 – testování na malém počtu vybraných pacientů. Teprve tady se sleduje, jestli léčivo skutečně funguje – prokázání léčebných účinků.

Fáze 3 – testování na velkém množství pacientů, ověření účinnosti. V případě úspěchu končí schválením a registrací léčiva, lze použít pro zdravotní péči.

Fáze 4 – sledování nežádoucích účinků při dlouhodobém užívání, interakcí s jinými léčivy, vzácných vedlejších účinků.

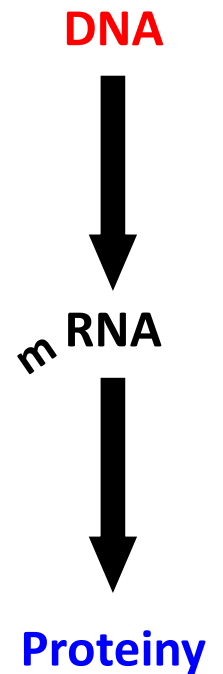
Table 2 Currently active companies working on miRNA-based therapeutics

| Companies | Product | Targeted miR | Disease type | Development phase | Reference |
|----------------------|------------|--------------|-----------------------|---------------------|-------------------------|
| Roche/Santaris | Miravirsen | miR-122 | HCV | Phase 2 | roche.com |
| Regulus Therapeutics | RG-101 | miR-122 | HCV | Phase 2 (hold) | regulusrx.com |
| | RG-012 | miR-21 | Alport syndrome | Phase 1 | |
| | RG-125 | miR-103/107 | NASH | Phase 1 | |
| MiRagen Therapeutics | MRG-201 | miR-29b | Fibrosis | Phase 2 | miragentherapeutics.com |
| | MRG-106 | miR-155 | Lymphoma and leukemia | Phase 1 and Phase 2 | |
| | MRG-107 | miR-155 | ALS | Pre-Clinical | |
| | MRG-110 | miR-92 | Ischemia | Phase 1 | |
| ENGeneIC | Mesomir | miR-16 | Mesothelioma | Phase 2 | engeneic.com |
| Abivax | ABX464 | miR-124 | IBD | Phase 2 | abivax.com |



mRNA – predikce mRNA:miRNA interakcí

- **miRNA** – microRNA. Krátké nekódující jednovláknové řetězce RNA. Regulace genové exprese – vazba na komplementární úseky mRNA vede ke snížení produkce proteinu (inhibice translace, destabilizace mRNA).
- Chybná (snížená x zvýšená) produkce microRNA pozorována u různých nemocí.
- Léčiva budoucnosti?



| Therapeutic molecule | Disease | Target miRNA | Biotechnology or Biopharmaceutical Company | Stage of Development (Clinical trial/preclinical trial) |
|----------------------|--|----------------------|---|---|
| Miravirsen(SPC3649) | For the treatment of hepatitis C virus (HCV) infection | miR-122 | SantarisPharma | Phase II clinical trials |
| MRX34 | For the treatment of different types of cancers | miR-34a | miRNATherapeutics | Phase I clinical trial |
| RG-101 | For the treatment viral effect | miR-122 | Regulus Therapeutics | Phase 1B clinical trial |
| RG-012 | To prevent alport nephropathy | miRNA-21 | Regulus therapeutics (with the strategic alliance with Genzyme) | Preclinical stage |
| RGLS4326 | For the treatment of Polycystic kidney disease (PKD) | miR-17 | Regulus Therapeutics | Phase I clinical study |
| MGN-1374 | For the treatment of post-myocardial infarction | miRNA-15 and miR-195 | miRagen therapeutics | Preclinical stage |
| MGN-2677 | For the treatment of vascular disease | miR-143/145 | miRagen therapeutics | Preclinical stage |
| MGN-4220 | For the treatment of cardiac fibrosis | miR-29 | miRagen therapeutics | Preclinical stage |
| MGN-4893 | For the treatment of disorders like abnormal red blood cell production | miR-451 | miRagen therapeutics. | Preclinical stage |
| MGN-5804 | For the treatment of cardiometabolic disease | miR-378 | miRagen therapeutics | Preclinical stage |
| MGN-6114 | For the treatment of peripheral arterial disease | miR-92 | miRagen therapeutics | Preclinical stage |
| MGN-9103 | For the treatment of chronic heart failure | miR-208 | miRagen therapeutics | Preclinical stage |
| Cobomarsen (MRG-106) | For the treatment of cutaneous T-cell lymphoma (CTCL) | miR-155 | miRagen therapeutics | Phase-I clinical trial |
| MRG-107 | For the treatment of amyotrophic lateral sclerosis (ALS) | miR-155 | miRagen therapeutics | Completed preclinical trial and entering in clinical trial |
| MRG-110 | Target blood vessel growth and to control ischemia | miR-92a | miRagen therapeutics | Phase-I clinical trial |
| Remlarsen (MRG-201) | For the treatment of different type of fibrosis such as cutaneous fibrosis, idiopathic pulmonary fibrosis etc. | miR-29 | miRagen therapeutics | Phase-I clinical trial (for idiopathic pulmonary fibrosis and other fibrosis) |



Contents lists available at [ScienceDirect](https://www.sciencedirect.com/journal/advanced-research)

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare



Therapeutic advances of miRNAs: A preclinical and clinical update

Chiranjib Chakraborty^{a,b,*}, Ashish Ranjan Sharma^{b,1}, Garima Sharma^c, Sang-Soo Lee^{b,*}

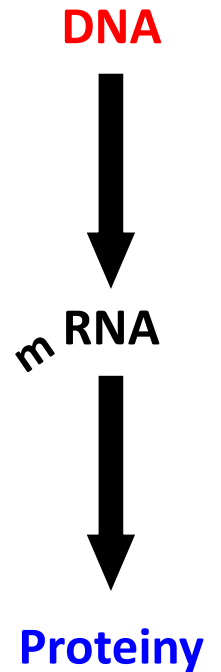
^a Department of Biotechnology, School of Life Science and Biotechnology, Adamas University, Barasat-Barrackpore Rd, Kolkata, West Bengal 700126, India
^b Institute for Skeletal Aging & Orthopedic Surgery, Hallym University-Chuncheon Sacred Heart Hospital, Chuncheon, Gangwon-Do 24252, Republic of Korea
^c Neuropsychopharmacology and Toxicology Program, College of Pharmacy, Kangwon National University, Chuncheon 24341, Republic of Korea



2021

mRNA – predikce mRNA:miRNA interakcí

- **miRNA** – microRNA. Krátké nekódující jednovláknové řetězce RNA. Regulace genové exprese – vazba na komplementární úseky mRNA vede ke snížení produkce proteinu (inhibice translace, destabilizace mRNA).
- Chybná (snížená x zvýšená) produkce microRNA pozorována u různých nemocí.
- Léčiva budoucnosti?



<http://mirdb.org/>

miRDB is an online database for miRNA target prediction and functional annotations. All the targets in miRDB were predicted by a bioinformatics tool, MirTarget, which was developed by analyzing thousands of miRNA-target interactions from high-throughput sequencing experiments. Common features associated with miRNA binding and target downregulation have been identified and used to predict miRNA targets with machine learning methods. miRDB hosts predicted miRNA targets in five species: human, mouse, rat, dog and chicken. Users may also provide their own sequences for custom target prediction using the updated prediction algorithm. In addition, through combined computational analyses and literature mining, functionally active miRNAs in humans and mice were identified. These miRNAs, as well as associated functional annotations, are presented in the FuncMir Collection in miRDB. As a recent update, miRDB presents the expression profiles of hundreds of cell lines and the user may limit their search for miRNA targets that are expressed in a cell line of interest. To facilitate the prediction of miRNA functions, miRDB presents a new web interface for integrative analysis of target prediction and Gene Ontology data.

Why miRDB?

miRDB is an online database for predicted microRNA targets in animals. MicroRNAs are involved in many diverse biological processes and they may potentially regulate the functions of thousands of genes. One major issue in miRNA studies is the lack of bioinformatics programs to accurately predict miRNA targets. Animal miRNAs have limited sequence complementarity to their gene targets, which makes it challenging to select relevant biological features to build target prediction models with high specificity. We have developed a new miRNA target prediction program based on support vector machines (SVMs) and high-throughput training datasets. By systematically analyzing high-throughput experimental data, we have identified novel features that are important to miRNA target binding and expression downregulation. These new features as well as other known features have been integrated in an SVM machine learning framework for the training of our target prediction model. Our prediction algorithm has been validated by independent experimental data for its improved selectivity on predicting a large number of miRNA downregulated gene targets.

1) Target prediction with user-provided miRNA or gene target sequence

Species:

Submission type:

Example miRNA sub mRNA submission

```
GCCCTCCCTCGGCGAAGCCCTCCAGCAGCCACACAGCTGTTCCACCTGCTGTGCAGTCACCTTAGATG
TTCAGAAGCCGCTTCTCTCCACTGAAGAGCTATGAGATACGCACAATACTGTCTTACGTTATGACTCTC
ATGCAAGCAAATACACAGTTTCATTGTTCTGAATCCTGTGGTTTCTTCAGCCCACTTTTATCGCCTTAACCT
AGTTAATGTATATTTGAATTGTGTATGACCTCAGAAGTGAATGATAATGAAGTTGCAAGTTTGTATAGC
CCGTGAAGTGCATAAGTATCTAATTTTACCTGAATTGATTGGGGGAAATTACCAGTAGAATGCCTTGGTCT
GAATATTTGATAGAACCAATTGTTGTACATAAAACAGATTGCCATATATATATGTATAAAAAATAAAAAATA
ATGGAAGATGATGGTGTCTCTAGTAAAAA
```


rRNA – nástroje a databáze

- **rRNA** tvoří 2/3 ribozomu. Ribozomy jsou komplexní makromolekulární struktury, skládají se z malé a velké podjednotky, obsahují RNA a proteiny.
- rRNA se vyskytuje v buňce ve velkém množství (80 % RNA).
- Jednotky **S**: prvotní analýzy ribozomů byly dělány pomocí ultracentrifugace. Svedberg (S) – jednotka **sedimentačního koeficientu** pojmenovaná podle Theodora Svedberga. **Svedberg** byl plodný vědec – měl Nobelovu cenu, čtyři manželky (postupně) a dvanáct dětí. Jeho dalším koníčkem byla botanika a malování.
- **rRNA** je vysoce konzervovaná (především 2D, 3D struktura). Využívána pro **fylogenetické analýzy a studium biodiverzity** (konzervované úseky, variabilní úseky genů).

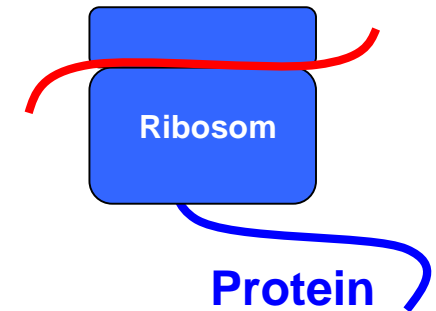


Table 1 RNAs in prokaryotic and eukaryotic ribosomes

| Ribosomes | Ribosomal RNAs (rRNA) |
|--|---|
| Prokaryotic 70S (average mass: 2.5 million daltons, of which about 66% is RNA) | Small subunit: 16S rRNA (1542 nucleotides) ^a Large subunit: 5S rRNA (120 nucleotides) ^a 23S rRNA (2904 nucleotides) ^a |
| Eukaryotic 80S (average mass: 4.2 million daltons, of which about 60% is RNA) | Small subunit: 18S rRNA (1874 nucleotides) ^b Large subunit: 5S rRNA (120 nucleotides) ^b 5.8S rRNA (160 nucleotides) ^b 28S rRNA (4718 nucleotides) ^b |

^a*Escherichia coli*.

^bTypical values for mammalian cytoplasmic ribosomes.

rRNA Structure

Arun Malhotra, University of Miami School of Medicine, Miami, Florida, USA

Ribosomal RNAs (rRNA) are essential components of ribosomes which are responsible for protein synthesis in all living cells. rRNA structure can be described at several levels – at the level of their primary sequences and secondary structures, as well as in terms of their tertiary folds of macromolecules within the ribosomal subunits.

Molekulární fylogeneze

- Fylogenetická data jsou získávána zkoumáním charakteristických znaků studovaných organismů.
- Prvotně používány pouze **MORFOLOGICKÉ** znaky. Problém – fosilní pozůstatky většinou **NEKVALITNÍ**, neposkytují žádané informace nebo se **VŮBEC** nedochovají.
- Kde se dá inteligentně umřít:

Histology of a Woolly Mammoth (*Mammuthus primigenius*) Preserved in Permafrost, Yamal Peninsula, Northwest Siberia

CHRISTINA PAPAGEORGOPOULOU,¹ KARL LINK,² AND FRANK J. RÜHLI^{2*}

¹Laboratory of Anthropology, Department of History and Ethnology, Demokritos University of Thrace, Komotini, GR-69100, Greece

²Centre for Evolutionary Medicine, Institute of Anatomy, University of Zurich, 8057 Zurich, Switzerland



Mamušátko „Lyuba“ (protože když to pojmenujete, asi se to lépe pitvá?).

Did saber-tooth kittens grow up musclebound? A study of postnatal limb bone allometry in felids from the Pleistocene of Rancho La Brea

Katherine Long¹, Donald Prothero^{1,2*}, Meena Madan³, Valerie J. P. Syverson⁴

Rancho La Brea tar pits –
asfaltová jezírka



Fig 2. Growth series of juvenile limb bones of *Smilodon fatalis*, showing the dramatic transformation in size and shape during ontogeny. A. Tibia. B. Femora. Photos by DRP.

Kosti šavlozubých koťátek.

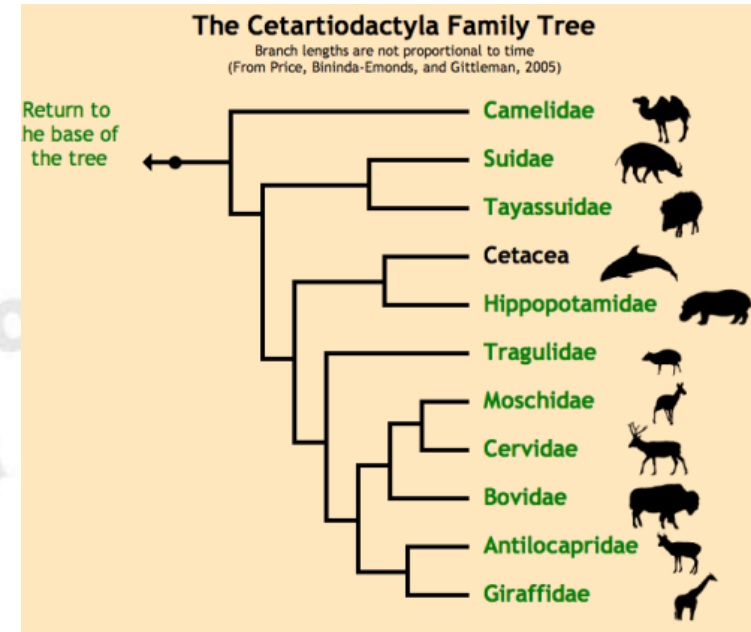
Molekulární fylogeneze

- Fylogenetická data jsou získávána zkoumáním charakteristických znaků studovaných organismů.
- Prvotně používány pouze **MORFOLOGICKÉ** znaky. Problém – fosilní pozůstatky většinou **NEKVALITNÍ**, neposkytují žádané informace nebo se **VŮBEC** nedochovají.
- Molekulární fylogenetická data = **sekvence** nukleových kyselin, proteinů ze současných taxonů.
- Jednotlivé stavy (nukleotidy, aminokyseliny) jsou jednoznačné a nezaměnitelné a vhodné pro matematické a statistické analýzy.
- Příklad: kdo jsou nejbližší příbuzní hrochů?



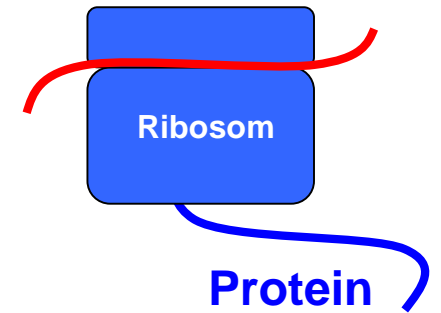
Molekulární fylogeneze

- Fylogenetická data jsou získávána zkoumáním charakteristických znaků studovaných organismů.
- Prvotně používány pouze **MORFOLOGICKÉ** znaky. Problém – fosilní pozůstatky většinou **NEKVALITNÍ**, neposkytují žádané informace nebo se **VŮBEC** nedochovávají.
- Molekulární fylogenetická data = **sekvence** nukleových kyselin, proteinů ze současných taxonů.
- Jednotlivé stavy (nukleotidy, aminokyseliny) jsou jednoznačné a nezaměnitelné a vhodné pro matematické a statistické analýzy.
- Příklad: kdo jsou nejbližší příbuzní hrochů?



rRNA – nástroje a databáze

- Využití pro fylogenetické analýzy: small subunit ribosomal RNA (SSU rRNA) – rRNA z malé ribozomální podjednotky
- 16S rRNA – prokaryota
- 18S rRNA – eukaryota



Silva = latinsky les

<https://www.arb-silva.de/>

SILVA > Bacteria > Proteobacteria > Gammaproteobacteria > Enterobacterales > Morganellaceae > Photorhabdus > Photorhabdus sp. CbKj163

| | | | |
|---|--|---|---|
| Enterobacterales (10) (3) Budviciaceae Enterobacteriaceae Erwiniaceae Gibbsiella Hafniaceae Morganellaceae Pectobacteriaceae Thorselliaceae uncultured Yersiniaceae | Morganellaceae (31) (323) Candidatus Ecksteina Candidatus Hamiltonella Candidatus Hartigia Candidatus Ishikawaella Candidatus Kleidoceria Candidatus Moranella Candidatus Puchtella Candidatus Purcellia Candidatus Regiella Candidatus Riesia Candidatus Rosenkranzia Candidatus Schneideria Candidatus Westeberhardia Cosenzaea endosymbionts Incertae Sedis Moellerella Morganella Photorhabdus Proteus Providencia uncultured Wigglesworthia Xenorhabdus | Photorhabdus (1/46) next ● Photorhabdus sp. CbKj163 ● Photorhabdus sp. OnIr40 ● Photorhabdus sp. Fslw96 ● Photorhabdus sp. WkSs43 ● Photorhabdus sp. KcTs129 ● Photorhabdus sp. OnKn2 ● Photorhabdus akhurstii ● Photorhabdus sp. TyKb140 ● Photorhabdus sp. NnMt2h ● Photorhabdus sp. IT | Photorhabdus sp. CbKj163 Accession Nr AB222085 Description Photorhabdus sp. CbKj163 gene for 16S ribosomal RNA. Regions 1 Length 1499 Quality Sequence — Alignment — Pintail — Links Details ▶ Link to EBI/ENA ▶ this page (permalink) |
|---|--|---|---|

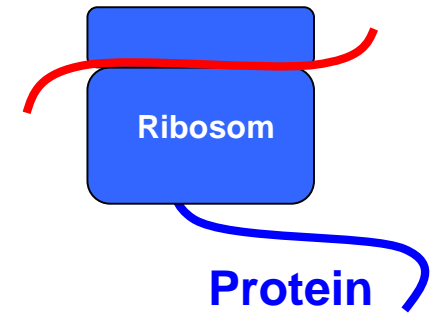
Vyhledávání podle názvu

Motivation

Sequencing the ribosomal RNA (rRNA) genes is currently the method of choice for phylogenetic reconstruction and nucleic acid based detection and quantification of microbial diversity. The ARB software suite with its corresponding rRNA databases has been accepted by researchers worldwide as their standard tool for large scale ribosomal RNA analysis. More than 20 years of development have already been invested to extend and maintain the system. To provide high quality and comprehensive rRNA databases comprising *Bacteria*, *Archaea* and *Eukarya* the SILVA (from Latin *silva*, forest) system has been implemented in 2007. It is designed as an automatic software pipeline for sequence retrieval, quality assignment and the alignment of nucleic acid sequences based on the latest comprehensive ARB alignments.

rRNA – nástroje a databáze

- Využití pro fylogenetické analýzy: small subunit ribosomal RNA (SSU rRNA) – rRNA z malé ribozomální podjednotky
- 16S rRNA – prokaryota
- 18S rRNA – eukaryota



Silva = latinsky les

<https://www.arb-silva.de/>

The screenshot shows the SILVA database search interface. The left sidebar contains a tree view of taxonomic levels: SILVA > Bacteria > Proteobacteria > Gammaproteobacteria > Enterobacteriales > Morganellaceae > Photorhabdus > Photorhabdus sp. CbKj163. The main content area shows details for *Photorhabdus sp. CbKj163*, including accession number AB222085, description "Photorhabdus sp. CbKj163 gene for 16S ribosomal RNA.", length 1499, and quality scores for Sequence, Alignment, and Pintail. A green circle highlights the "Details" link.

Vyhledávání
podle názvu

The screenshot shows the ACT: Alignment, Classification and Tree Service interface. The title "ACT: Alignment, Classification and Tree Service" is circled in green. The "SINA 1.2.11" section shows the "Input data" field with a sequence of nucleotides. Below the sequence is a "Select file" button. At the bottom, a phylogenetic tree is shown with four branches labeled 1, 4, 3, and 2.

ACT: Alignment, Classification and Tree Service

SINA 1.2.11

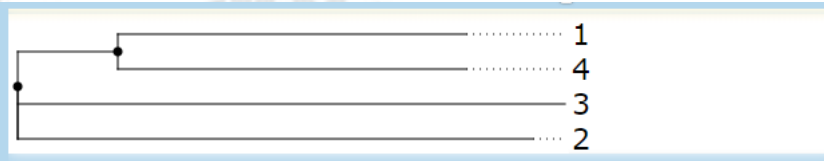
Input data

```
>1
GCGTGCTTAACACATGCAAGTCGAACGAGAACGGTTATAGCTTGCCTATAACTGTCAGCT
AAGTGGCGCACGGGTGAGTAATGTATAGTTAACCTGCCCTCGAGAAAGGAATAACAGTTG
GAAACGACTGTAATGCCCTATATGCCCAATAACAAAAGTATGCTGGGAAACGCTTTA
GTGCTCGAGGATGGGACTGTATGGTATCAGCTAGTTGGTGAGGTAATGGCTCACCAAGGC
AATGACGCCTAACTGGTTTGAGAGGATGATCAGTCACACTGGAAGTGAAGACACGGTCCAG
ACTCCTACGGGAGGCAGCAGTGGGGAATTTGCACAATGGAGGAACTCTGATCGAGCAA
CGCCGCTGGAGGATGACACTTTTCGGAGCGTAAACTCCTTTATATAGGAAGATAATGA
CGGTACTATATGAATAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGTAATACGGAGGG
TGCAAGCGTTACTCGGAATCACTGGGCGTAAAGAGCGTGTAGGCGG
```

or

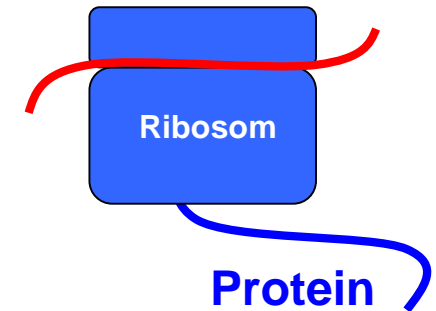
upload an FASTA file

Select file



rRNA – nástroje a databáze

- Small subunit ribosomal RNA (SSU rRNA) – rRNA z malé ribozomální podjednotky
- 16S rRNA – prokaryota
- 18S rRNA – eukaryota
- 16s rRNA – studium **biodiverzity** (metagenomika)



A Comparative Metagenome Survey of the Fecal Microbiota of a Breast- and a Plant-Fed Asian Elephant Reveals an Unexpectedly High Diversity of Glycoside Hydrolase Family Enzymes

Nele Ilmberger¹, Simon Güllert¹, Joana Dannenberg¹, Ulrich Rabausch¹, Jeremy Torres¹, Bernd Wemheuer², Malik Alawi³, Anja Poehein², Jennifer Chow¹, Dimitrij Turaev⁴, Thomas Rattei⁴, Christel Schmeisser¹, Jesper Salomon⁵, Peter B. Olsen⁵, Rolf Daniel², Adam Grundhoff⁶, Martin S. Borchert⁵, Wolfgang R. Streit^{1*}

Materials and Methods

DNA isolation

No specific permissions were required for these activities. The study did not involve endangered or protected species. Coordinates are: Longitude/Latitude 9.941572/9.941572. Fresh feces samples derived from a six-years-old female Asian elephant ("Kandy", *2003 in Hamburg) and from a three-weeks-old male elephant ("Assam", *2012 in Hamburg), both living in the zoo 'Hagenbecks Tierpark' in Hamburg (Germany), were collected from the zoo staff directly after defecation. The samples were transported, directly and on ice, to the laboratory in Hamburg Klein Flottbek for further analysis. The elephants were not treated with antibiotics. The older Asian elephant was mainly fed with grass, hay, leaves and twigs, with additional fruits and vegetables. In contrast, the three-weeks-old male elephant was breast-fed. DNA isolation was performed with the QIAamp DNA Stool kit from Qiagen (Hilden, Germany) as described previously [27].



Kandy už má potomky...



Assam bude moci za chvílku poskytovat vzorek...

<https://ngs.arb-silva.de/silvangs/>

Processing and analysis of 16S rRNA gene data sets

Via pyrosequencing generated raw sequences were processed according to [33], with the following modifications: After raw data extraction, reads shorter than 300 bp and those possessing long homopolymer stretches (>8 bp) or primer mismatches (>5 bp) were removed. The sequences were denoised employing Acacia [34]. Chimeric sequences were removed using UCHIME in reference mode with the most recent SILVA SSU database as reference dataset (SSURef 115 NR) [35]. The processed 16S rRNA gene sequences were uploaded to the SILVA NGS (SILVA next-generation sequencing) server for taxonomic classification [36]. Microbial taxonomy was determined using default settings with two adjustments: The cluster sequence identity threshold was decreased to 0.97 and the maximal taxonomic depth was increased to 30. Rarefaction curves were calculated employing the QIIME 1.8 software package [37].

silvangs

QIAamp® DNA Stool Handbook

For DNA purification from stool samples

Introduction

The QIAamp DNA Stool Mini Kit provides fast and easy purification of total DNA from fresh or frozen stool samples. QIAamp purified DNA is of high quality and well suited for use in PCR and other downstream enzymatic reactions.

The simple QIAamp spin procedure yields pure DNA ready for direct use in less than 1 hour. The procedure can be automated on the QIAcube® for increased standardization and ease of use (see page 9). Purification requires no phenol–chloroform extraction or alcohol precipitation, and involves minimal handling. DNA is eluted in low-salt buffer and is free of protein, nucleases, and other impurities or inhibitors. The purified DNA is ready for use in PCR and other enzymatic reactions, or can be stored at –20°C for later use.

Stool samples typically contain many compounds that can degrade DNA and inhibit downstream enzymatic reactions. To ensure removal of these substances, the QIAamp DNA Stool Mini Kit contains InhibitEX Tablets, a unique reagent provided in a convenient tablet form. InhibitEX Tablets efficiently adsorb these substances early in the purification process so that they can easily be removed by a quick centrifugation step. In addition, the kit contains Buffer ASL, which is specially developed to remove inhibitory substances from stool samples.

DNA of up to 20 kb is purified by the QIAamp DNA Stool Mini Kit. DNA of this length denatures completely during thermal cycling and can be amplified with the highest efficiency.

Sample size

The QIAamp DNA Stool Mini Kit is optimized for use with 180–220 mg fresh or frozen stool, but can also be used with larger amounts of stool. Starting with larger amounts of stool is recommended when the target DNA is not distributed homogeneously throughout the stool and/or is at a low concentration; a larger amount of starting material will increase the likelihood of purifying DNA from low-titer sources in stool samples. See “Protocol: Isolation of DNA from Larger Volumes of Stool”, page 30.

The QIAamp DNA Stool Mini protocols can also be used for samples of less than 180 mg (e.g., forensic samples). In such cases, follow one of the standard protocols, without reducing the amounts of buffers or InhibitEX matrix used.

For maximum flexibility, protocols are provided for use with samples collected in both 2 ml microcentrifuge tubes and 15 ml stool tubes.

Metagenomika

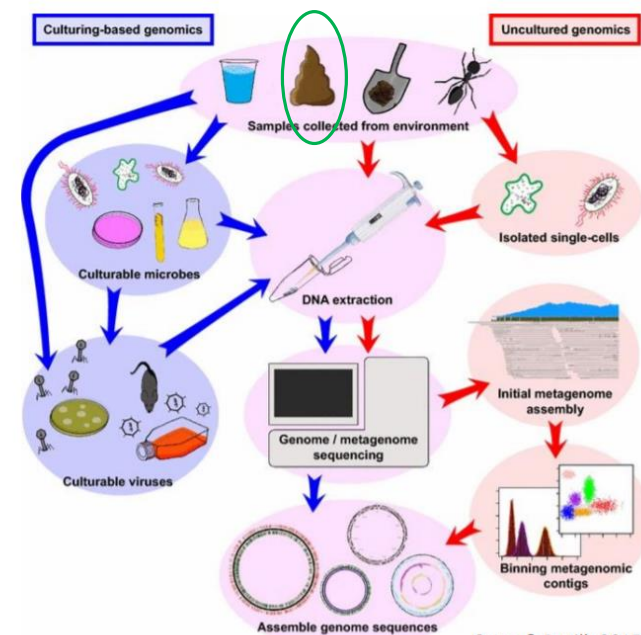
- Sekvence jednotlivého genomu je „out“

- Mnoho organismů nelze získat v izolovaném stavu

- Půdní společenstva
- Mořská společenstva
- Střevní mikrobiom

- Metagenomika se zabývá sekvenacemi

komplexních vzorků



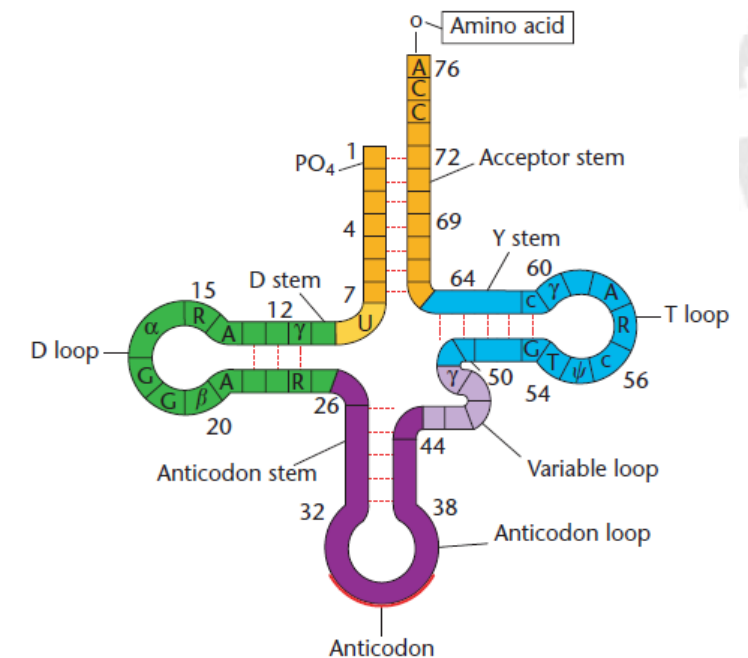
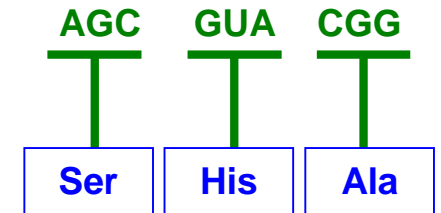
About the Human Microbiome

The Human Microbiome is the collection of all the microorganisms living in association with the human body. These communities include eukaryotes, archaea, bacteria and viruses. Bacteria in an average human body number ten times more than human cells, for a total of about 1000 more genes than are present in the human genome. Because of their small size, however, microorganisms make up only about 1 to 3 percent of our body mass (that's 2 to 6 pounds of bacteria in a 200-pound adult). These microbes are generally not harmful to us, in fact they are essential for maintaining health. For example, they produce certain vitamins that we do not have the genes to make, break down our food to extract nutrients we need to survive, teach our immune systems how to recognize dangerous invaders and even produce helpful anti-inflammatory compounds that fight off other disease-causing microbes. An ever-growing number of studies have demonstrated that changes in the composition of our microbiomes correlate with numerous disease states, raising the possibility that manipulation of these communities could be used to treat disease.

<https://hmpdacc.org/hmp/overview/>

tRNA – nástroje a databáze

- **tRNA** - transport aminokyselin pro proteosyntézu. Obsahují místo pro připojení aminokyseliny a antikodon pro interakci s mRNA. Existují ve dvou stavech, s/bez aminokyseliny.
- Na 3' konci je vždycky sekvence **CCA** (kódovaná nebo připojená enzymaticky) – sloužící k připojení aminokyseliny.
- tRNA podléhá mnoha posttranskripčním **modifikacím**, obsahuje neobvyklé **nukleotidy** (důležité pro strukturu a funkci).
- Typická **sekundární** struktura: jetelový list („cloverleaf“).



Transfer RNA

Emanuel Goldman, *New Jersey Medical School, Newark, New Jersey, USA*

Transfer ribonucleic acid (tRNA) is the class of molecules that decode the genetic code and link the coded information to their attached amino acids. These become constituents of the protein specified by the gene.

tRNA Structure

Eric Westhof, *Institute of Molecular and Cellular Biology, Strasbourg, France*

Pascal Auffinger, *Institute of Molecular and Cellular Biology, Strasbourg, France*

Transfer ribonucleic acid (tRNA) molecules that participate in the elongation step of protein synthesis on the ribosome have a conserved secondary structure, known as the cloverleaf, and fold into a common three-dimensional architecture.

tRNA – databáze

T-psi-C: user friendly database of tRNA sequences and structures

Marcin Piotr Sajek^{1,*}, Tomasz Woźniak^{1,*}, Mathias Sprinzl^{2,†}, Jadwiga Jaruzelska¹ and Jan Barciszewski^{3,4,*}

<http://tpsic.igcz.poznan.pl/>

-Psi-C Database

Discover the best tRNA database available!

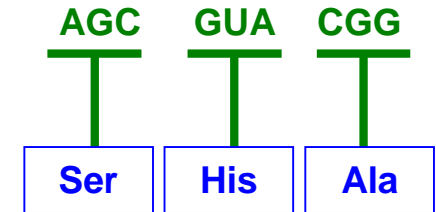
There is simplified list of what you can do:

- Browse sequences
- Search by sequence, structure and other properties
- Just BLAST it
- See available and generated models
- Analyze visualized secondary structure
- Upload your own tRNA
- Download data
- Gain API access to do the above

Statistics

Data version 1 (06.09.2019):

- total number of sequences: 1030
- mitochondrial tRNAs: 122
- plastid tRNAs: 38
- viral tRNAs: 17
- tRNAs obtained from hydro-tRNAseq: 251
- tRNAs with structural data from crystal structure: 54
- tRNAs obtained only from PDB: 46



tdbR00000492

| NAME | VALUE |
|---|---|
| Coding AA: | Trp |
| Anticodon sequence: | NCA |
| Organism: | Rattus norvegicus NCBI:txid10116 |
| RNA type: | mitochondrial |
| Sequence: | AAGAAGUUU"GGAUUACAGUCCAAGAGCCUNCA*AGCCUUAGAAAACAAACAAGUUUAACUUCGCCA |
| Unmodified sequence: | AAGAAGUUUAGGAUUAUACAGUCCAAGAGCCUUCAAAGCCUUAGAAAACAAACAAGUUUAACUUCGCCA |
| Secondary structure: | .(((((((.....))))))(((.....))).....((((.....))))))..... |
| Homology model or structure extracted from PDB: | Model Download (last update Sept. 6, 2019) |
| Prepared by: | Database creators |
| Version: | 1 (Sept. 6, 2019), created: Feb. 25, 2019, modified: Sept. 23, 2019 |
| Folding free energy [kcal/mol]: | -10.90 |
| Pre-aligned sequence: | AAGAAGUUU"GGAUUUA---AC---AGUCCAAGAGCCUNCA*AGCCUUUA-----G-AAAAC-AAACAA--GUUUUAACUUCGCCA |
| Pre-aligned secondary structure: | .(((((((.....))))))(((.....))).....((((.....))))))..... |
| Publications: | <ul style="list-style-type: none">• tRNA alterations in cancer. K Randerath, H P Agrawal, E Randerath <i>Recent results in cancer research. Fortschritte der Krebsforschung. Progres dans les recherches sur le cancer</i> volume: 84 PUBMED ID: 6844691 |

tRNA – databáze

T-psi-C: user friendly database of tRNA sequences and structures

Marcin Piotr Sajek^{1,*}, Tomasz Woźniak^{1,*}, Mathias Sprinzl^{2,†}, Jadwiga Jaruzelska¹ and Jan Barciszewski^{3,4,*}

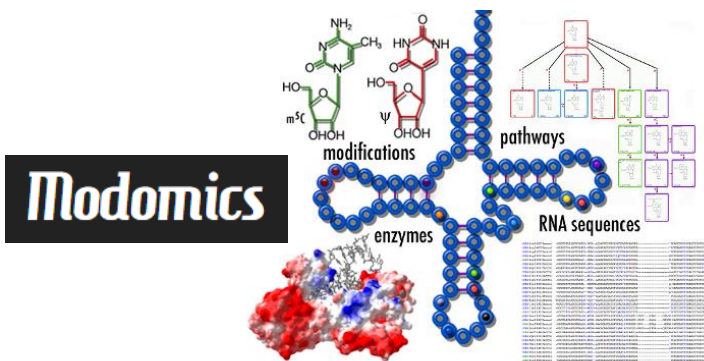
<http://tpsic.igcz.poznan.pl/>

-Psi-C Database

Discover the best tRNA database available!

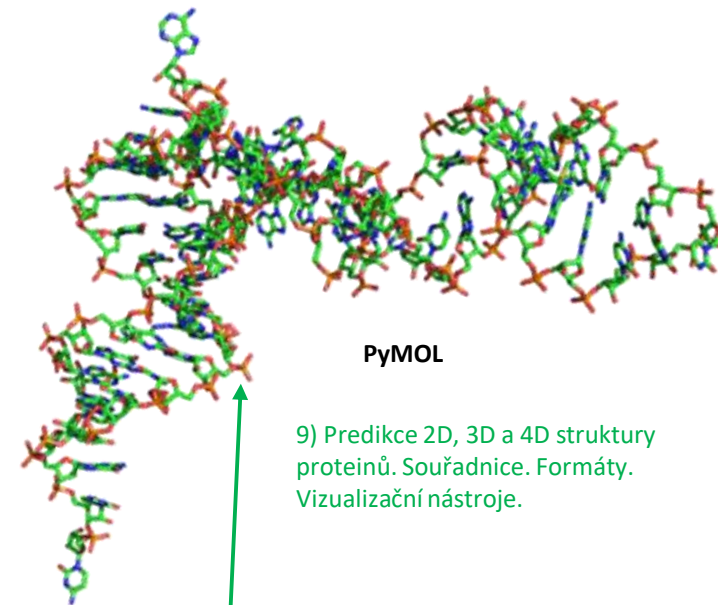
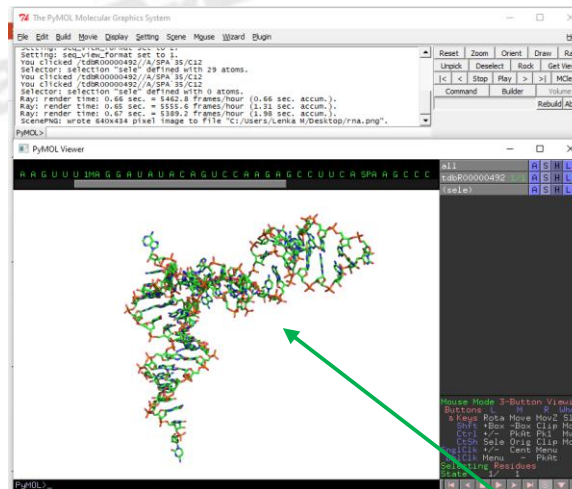
There is simplified list of what you can do:

- Browse sequences
- Search by sequence, structure and other properties
- Just BLAST it
- See available and generated models
- Analyze visualized secondary structure
- Upload your own tRNA
- Download data
- Gain API access to do the above



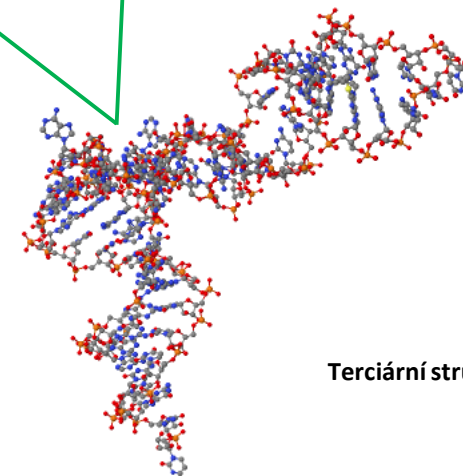
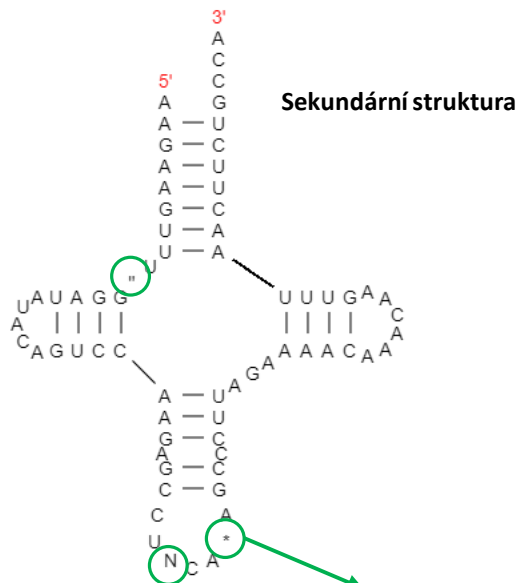
Modomics

modifications, pathways, enzymes, RNA sequences



9) Predikce 2D, 3D a 4D struktury proteinů. Souřadnice. Formáty. Vizualizační nástroje.

tdbR00000492



JSmol

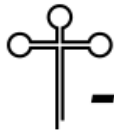
A -> *
Modomics: ms2i6A

tRNA – databáze

T-psi-C: user friendly database of tRNA sequences and structures

Marcin Piotr Sajek^{1,†}, Tomasz Woźniak^{1,†}, Mathias Sprinzl^{2,†}, Jadwiga Jaruzelska¹ and Jan Barciszewski^{3,4,*}

<http://tpsic.igcz.poznan.pl/>



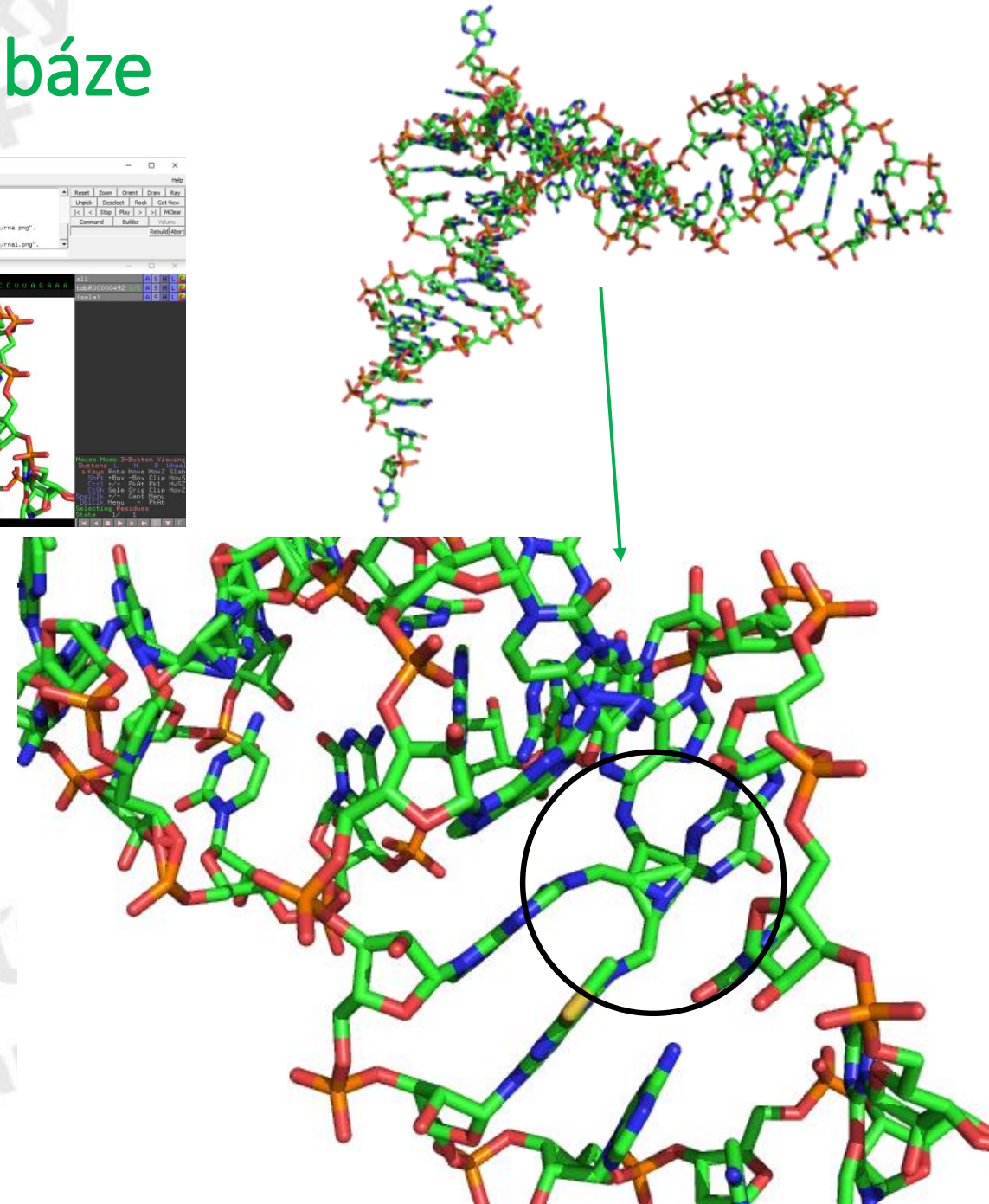
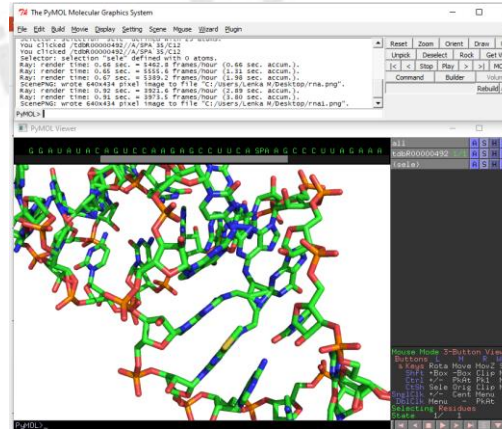
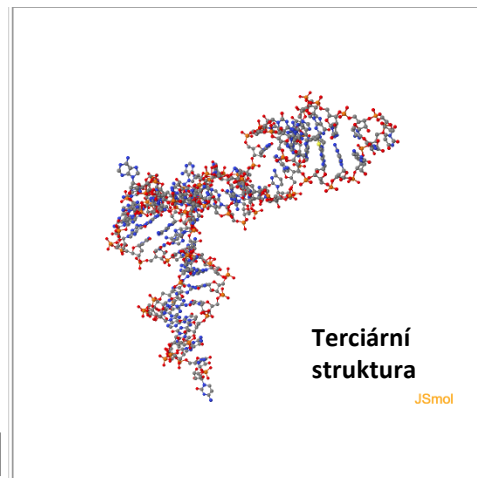
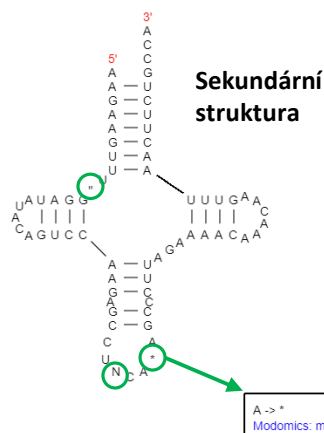
-Psi-C Database

Discover the best tRNA database available!

There is simplified list of what you can do:

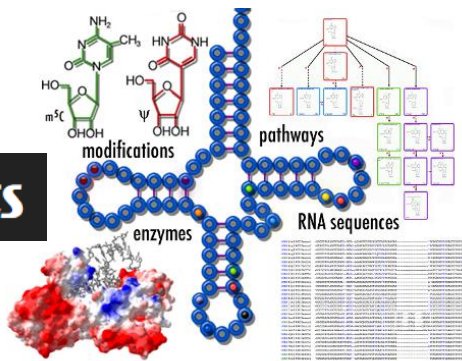
- Browse sequences
- Search by sequence, structure and other properties
- Just BLAST it
- See available and generated models
- Analyze visualized secondary structure
- Upload your own tRNA
- Download data
- Gain API access to do the above

tdbR00000492



tRNA – modifikace

Modomics



RNA sequences with modifications

RNA type: tRNA Subtype: Gly Organism: Bombyx mori

Visualise as: Modomics symbols Display Draw Modification Profile

PDF CSV Print

| Type | AA type | Anticodon | Organism | Organelle | accept | D-domain | anticodon domain | variable region | T-domain | accept | | |
|------|---------|-----------|-------------|-----------|--------------------------|---------------------------|------------------------|-----------------------|----------|---------|--------------|---------------|
| | | | | | 10 | 20 | 30 40 | | 50 60 | 70 | | |
| tRNA | Gly | GCC | Bombyx mori | cytosolic | 012345678901234567-890-- | 1234567890123456789012345 | eeeeeeeeeeeeeeeeeeee | 678901234567890123456 | | | | |
| tRNA | Gly | ,CC | Bombyx mori | cytosolic | -GCAJCGGUUGUUCAGU--GGD-- | AAGUUCUGCCUGCCA?GC | GGCG | -----G-??? | GGGTPCG | UUC | CGCCGAUCACCA | |
| | | | | | -GC | JUGGUGUGPAAD--GGDC- | AGCAUGPUGCCU,CCAAGCAGU | G----- | A-U?? | GGGTPCG | UUC | CGCCGAACGACCA |

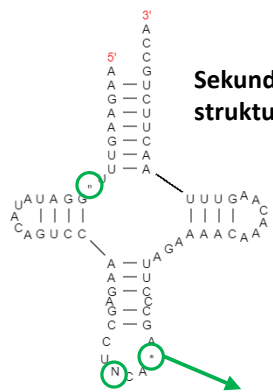
Showing 1 to 2 of 2 entries

<https://iimcb.genesilico.pl/modomics/>

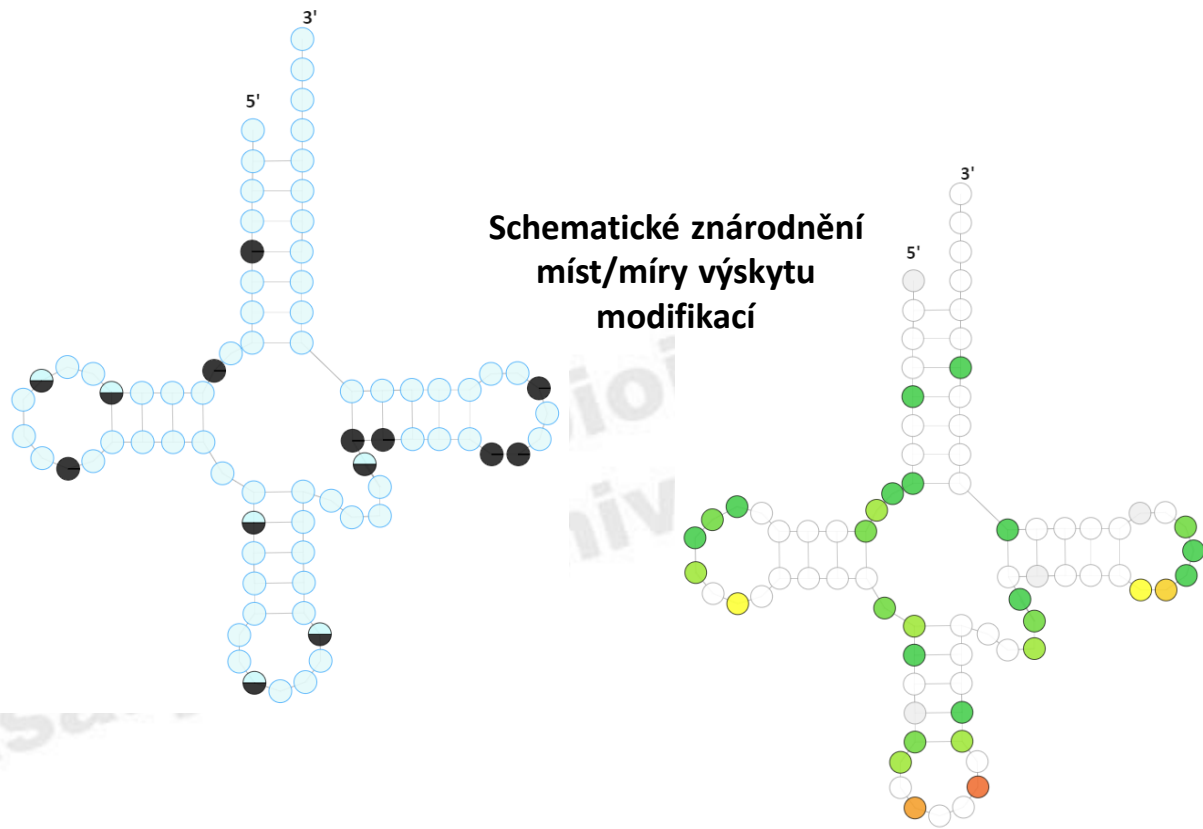
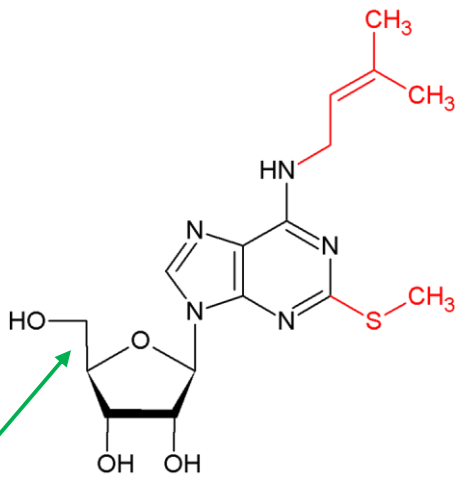
Modomics

MODOMICS is a database of RNA modifications that provides comprehensive information concerning the chemical structures of modified ribonucleosides, their biosynthetic pathways, the location of modified residues in RNA sequences, and RNA modifying enzymes.

tdbR0000492



A -> *
Modomics: ms2i6A



RNA

- **RNA:** tři hlavní formy zapojené do proteosyntézy, mediátorová RNA (**mRNA**), ribozomální RNA (**rRNA**), transferová RNA (**tRNA**).

Mnoho dalších typů a funkcí, **např.:**

- **miRNA** (micro RNA) – mikroRNA, krátké nekódující jednovláknové řetězce RNA. Regulace genové exprese.
- **snRNA** (small nuclear RNA) – malá jaderná RNA, asociovány s proteiny, zapojení do sestřihu.
- **snoRNA** (small nucleolar RNA) – malá jadérková RNA, úprava pre-rRNA. Většina savčích snoRNA pochází z intronových sekvencí pre-mRNA. Geny obsahující tyto introny většinou kódují proteiny zapojené do biosyntézy ribozomů.
- **scRNA** (small cytoplasmic RNA) – malá cytoplazmatická RNA, obecně jakákoliv malá RNA v cytoplasmě, která není přímo zapojená do proteosyntézy.
- **lncRNA** (long non-coding RNA) – dlouhé nekódující RNA, RNA > 200 nukleotidů, regulační funkce. Mohou vznikat přepisem intergenových nebo intronových sekvencí.

On the classification of long non-coding RNAs

Lina Ma,¹ Vladimír B. Bajic² and Zhang Zhang^{1*}

Cellular RNAs: Varied Roles

Jian Gu, Baylor College of Medicine, Houston, Texas, USA

Ram Reddy, Baylor College of Medicine, Houston, Texas, USA

In the cell, there are three major types of RNA directly involved in protein synthesis. In addition, many other cellular RNAs also play important functions.

Dlouhé nekódující RNA a jejich význam u nádorových onemocnění

Long Non-Coding RNAs And Their Relevance in Cancer

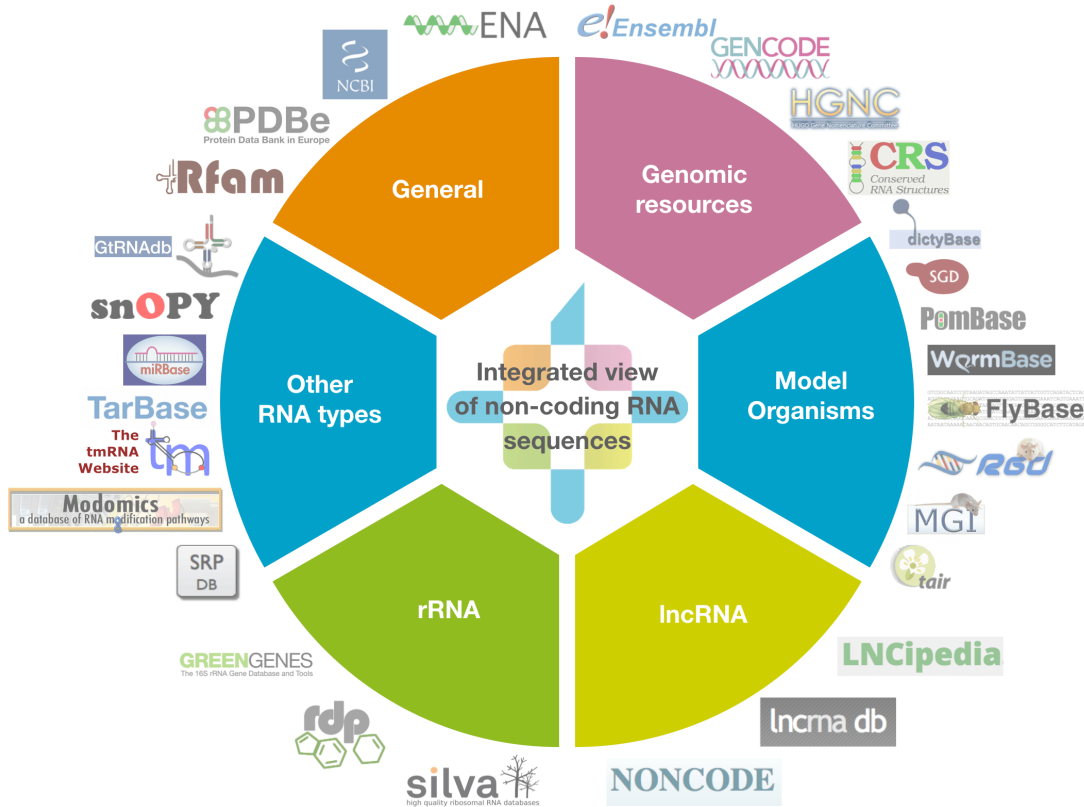
Šána J.^{1,2}, Faltejsová P.^{1,2}, Svoboda M.^{1,2,3}, Slabý O.^{1,2}

Tab. 1. Hlavní třídy dlouhých nekódujících RNA. Upraveno podle [17].

| Třída | Charakteristika | Asociace s onemocněním / biologická funkce |
|---|--|---|
| dlouhé mezigenové nekódující RNA | délka od několika stovek až po desítky tisíc nt; v genomu umístěny mezi dvěma geny | zapojení v tumorigenezi a nádorovém metastazování / regulace genové exprese; imprinting |
| dlouhé intronové nekódující RNA | v genomu umístěny uvnitř intronů; evolučně konzervované; tkáňová a subbuněčná specifita | aberrantní exprese u nádorových onemocnění / možné zapojení do posttranskripčního umlčování genové exprese |
| nekódující RNA asociované s telomerami | 100 bp až 9 kb; konzervovány napříč eukaryoty; syntetizovány z oblastí bohatých na konstitutivní heterochromatin; polyadenylovány | možný význam u onemocnění asociovaných s délkou telomer, včetně nádorových onemocnění / negativní regulátor kompenzační kapacity telomer |
| dlouhé nekódující RNA s dvojitou funkcí pseudogeny | regulační funkce nekódujících RNA; za určitých okolností schopnost kódovat proteiny; kopie genů, které ztratily schopnost kódovat proteiny; vzniklé retrotranspozicí; tkáňově specifické | deregulovány u nádorů prsu a vaječníků / regulace genové exprese deregulovány v průběhu nádorové iniciace a následné progresu / regulace genové exprese skrze funkční umlčení mikroRNA |
| přepisované vysoce konzervované oblasti | delší než 200 bp; absolutně konzervované mezi ortologními oblastmi genomů člověka, krysy a myši; umístěny v inter- i intragenových oblastech | aberrantní exprese u mnoha nádorových onemocnění / inhibice protein-kódujících genů a některých nekódujících RNA |

RNAcentral

<https://rnacentral.org/>



RNAcentral is a free, public resource that offers **integrated** access to a **comprehensive** and **up-to-date** set of non-coding RNA sequences provided by a collaborating group of **Expert Databases** representing a **broad range of organisms and RNA types**.

The development of RNAcentral is coordinated by [European Bioinformatics Institute](#) and is supported by [Wellcome](#). Initial funding was provided by [BBSRC](#).



Text search

Search by *gene*, *species*, *ncRNA type*, *accession*, or any other keyword

[Browse sequences](#)

Sequence search

Search for similar sequences - *now with facets and 2D diagrams*

[Search by sequence](#)

Genome browser

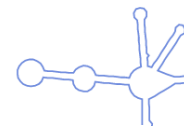
Explore RNAcentral sequences in your favorite genome locations

[Explore genomes](#)

Secondary structures

RNAcentral contains **over 22 million secondary structures** displayed using standard layouts, including SSU and LSU rRNA, and over 2,600 Rfam families. [Learn more](#)

Lysine riboswitch RNA from *Thermotoga maritima*



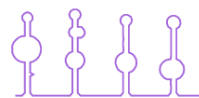
Human 18S ribosomal N2 (RNA18SN2-4)



Parcubacteria LOOT RNA



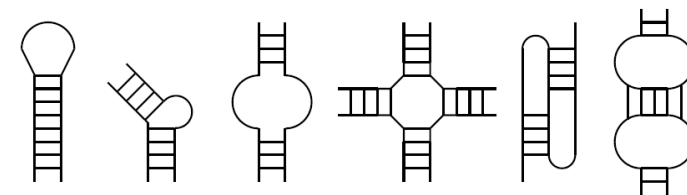
Human small Cajal body-specific RNA 13 (SCARNA13)



RNA – struktura

Lukáš Židek
Skripta předmětu C9530 Strukturní biochemie

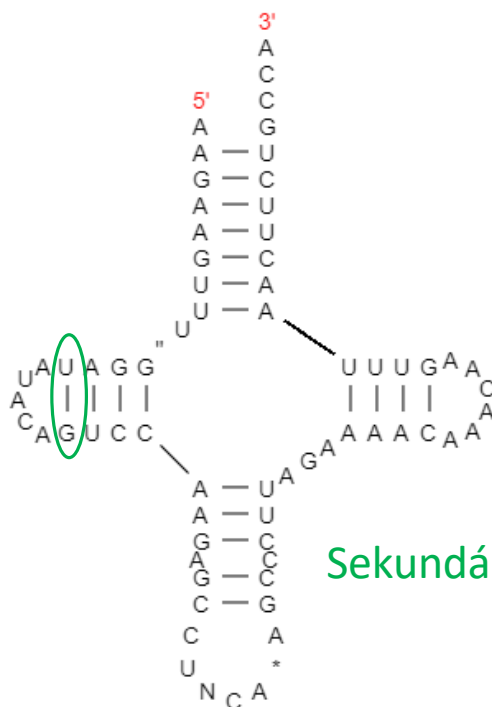
- **RNA:** primární, sekundární a terciární struktura.
- **Primární:** lineární sekvence bází.
- **Sekundární:** základním prvkem jsou dvouvláknové úseky vzniklé párováním bází (A-U, G-C, G-U).
- **Terciární:** 3D (prostorové uspořádání) molekuly RNA.



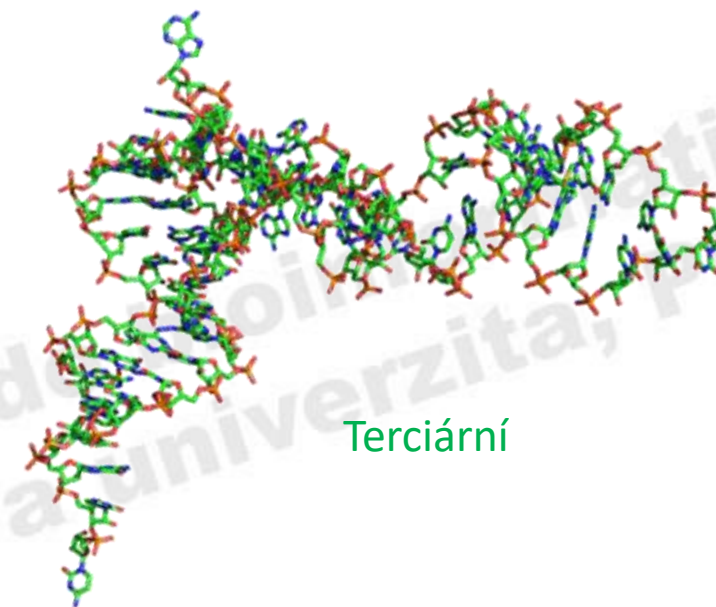
Prvky sekundární struktury

AAGAAGUUUAGGAUUAUACAGUCCAA
GAGCCUUCAAAGCCCUUAGAAAACA
AACAAGUUUAACUUCUGCCA

Primární



Sekundární



Terciární

RNA – predikce 2D struktury

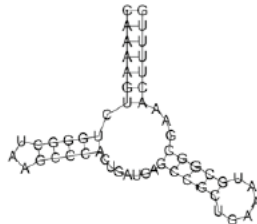
- Predikce sekundární struktury RNA je významný bioinformatický problém.
- Metody *ab initio*/ metody komparativní (srovnávací).
- *Ab initio* – predikce je založena na jediné (zkoumané) RNA sekvenci. Vyhledávání **nejstabilnější** RNA struktury = struktury s **nejnižší** volnou energií. Tvorba párů bází je energeticky výhodná, zjednodušeně řečeno hledáme tedy strukturu s maximálním možným počtem párů bází.
- **Komparativní** metody – využívají evolučně **příbuzné** RNA. Předpoklad: homologní sekvence RNA mají stejnou 2D strukturu. Multiple alignment sekvencí RNA, identifikace evolučně konzervované sekundární struktury. Teoreticky lepší, problémem je nutnost mít sadu homologních sekvencí.

a Secondary structure prediction

CAAAAGUCUGGGCUAAGCCCACUGAUGAGCCGCUGAAAUGCGGGCGAAACUUUUG



CAAAAGUCUGGGCUAAGCCCACUGAUGAGCCGCUGAAAUGCGGGCGAAACUUUUG
(((((((.((((((...)))))) (((((.....))))))))))))

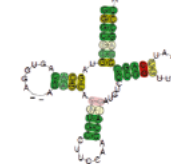


b Common secondary structure prediction

GGGCCUGUAGCUCAGAGGAUAGAGCACGUGGCUACGAACCACGGUGUCGGGGUUCGAAUCCUCCUGGCCA
GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCUUAAGGGUGAGGUCGUGAUUUCGAAUCCUCCUGGCCA
GGCGCCGUGGCAGUGGA--AGCGCGCAGGGCUCAUAACCCUGAUGUCCUCGGAUCGAAACCGAGCGGCUA
GCGUUGGUGUAUAGUGGUG-AGCAUAGCUGCCUCCAAGCA-GUUGACCOSGGUUCGAUCCCGGCCAAGCA
ACUCCUUAGUAUAUU---AAUAUAACUGACUCCAUAUA-GUAGAUUCGAAU-AAACCCAGAAGAGAGUA



GGGCCUGUAGCUCAGUGGA__AGAGCACAUCCUCCAACCA_GAUGUCCOSGGUUCGAAUCCAGCGGCCA
(((((((.((((((.....)))))) (((((.....))))))))))))



21
M.

Ab initio – výlet do historie

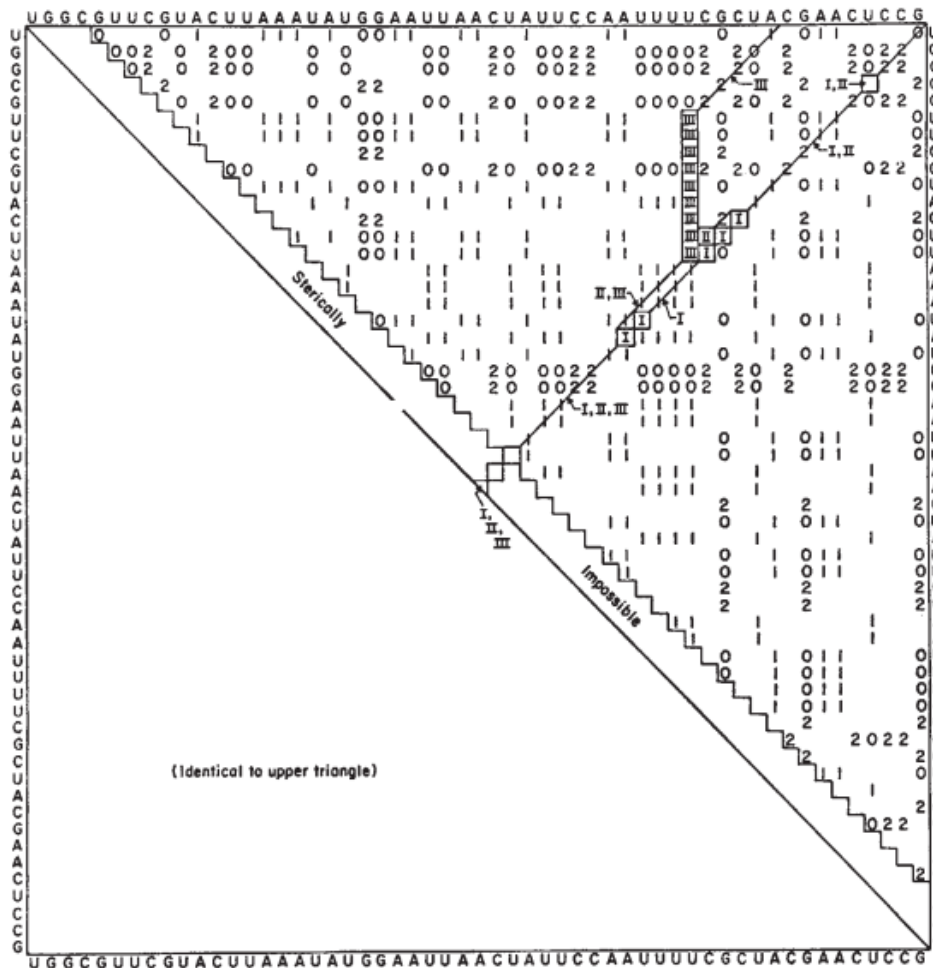


Fig. 2 The base pairing matrix showing all possible base pairs in a sequence of fifty-five bases from R17 viral RNA³. An upper right to lower left diagonal line joining base pairs represents an anti-parallel helical region. Squares represent loops between helical regions. A bulge is seen as a vertical or horizontal line. The three most stable structures (I, II, III) are indicated here and in Fig. 3.

Estimation of Secondary Structure in Ribonucleic Acids

IGNACIO TINOCO, jun., OLKE C. UHLENBECK & MARK D. LEVINE

Department of Chemistry, University of California, Berkeley, California 94720

1971

A simple method for estimating the secondary structure of an RNA molecule has been proposed on the basis of the knowledge of its sequence.

Alberts and Doty¹². By allowing looped out bases and putting the sequence in register in order to maximize base pairing, they were able to suggest optimal secondary structures from a given sequence. When this approach was applied to transfer RNA sequences, a wide variety of sequences could all be put into a similar cloverleaf pattern with the anticodon in the same position, thus demonstrating the interchangeability of the tRNAs in protein biosynthesis on the ribosome²¹. Know-

- Identifikace všem možných párů bází, které může zkoumaná sekvence sama se sebou tvořit.
- A-U, G-C, G-U
- Páry jsou různě „obodované“: A-U (1), G-C (2), G-U (0).
- **Cíl: Složit sekvenci do struktury, kterou získáte nejvíc bodů.**

Ab initio – výlet do historie

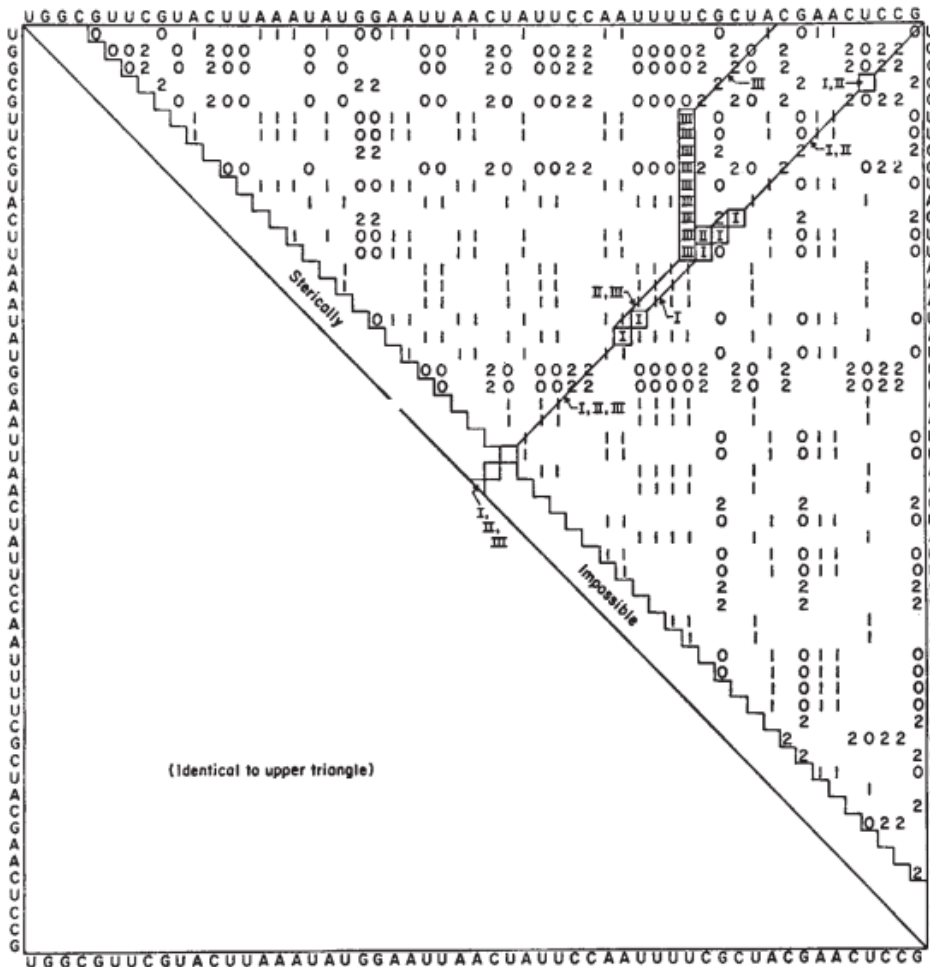
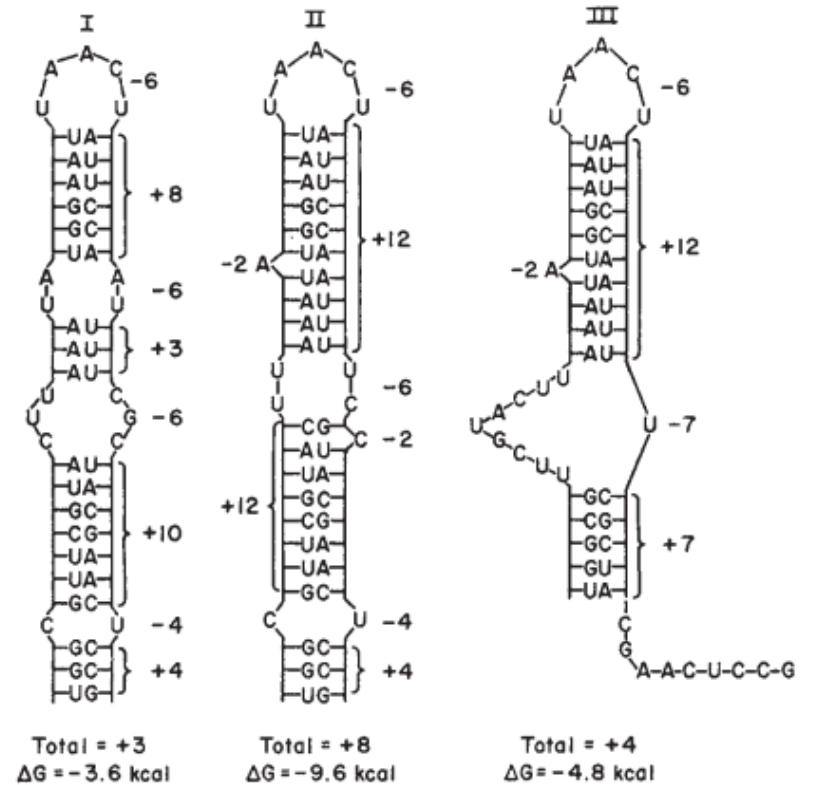


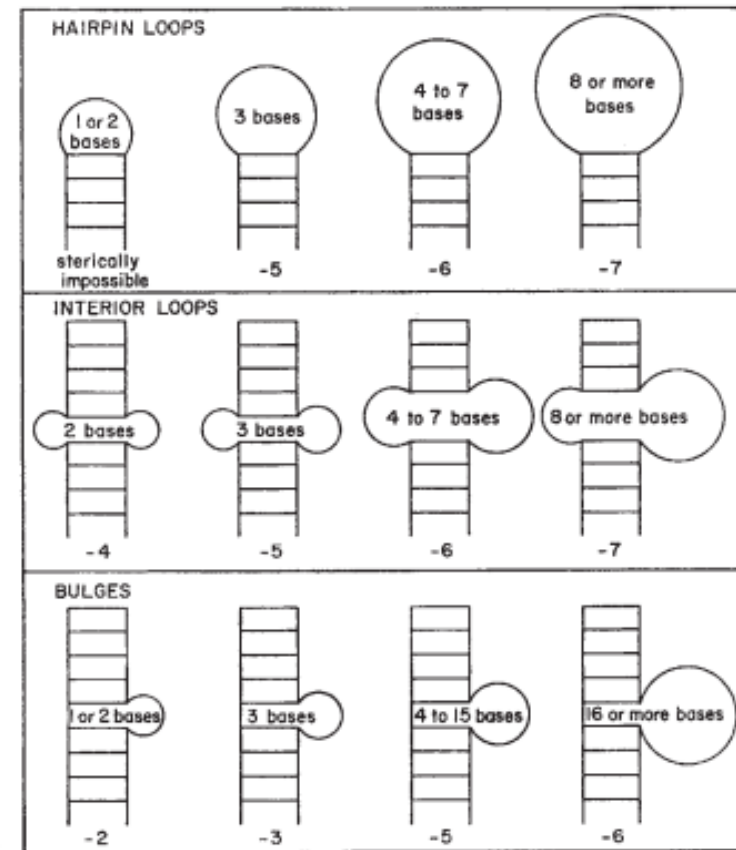
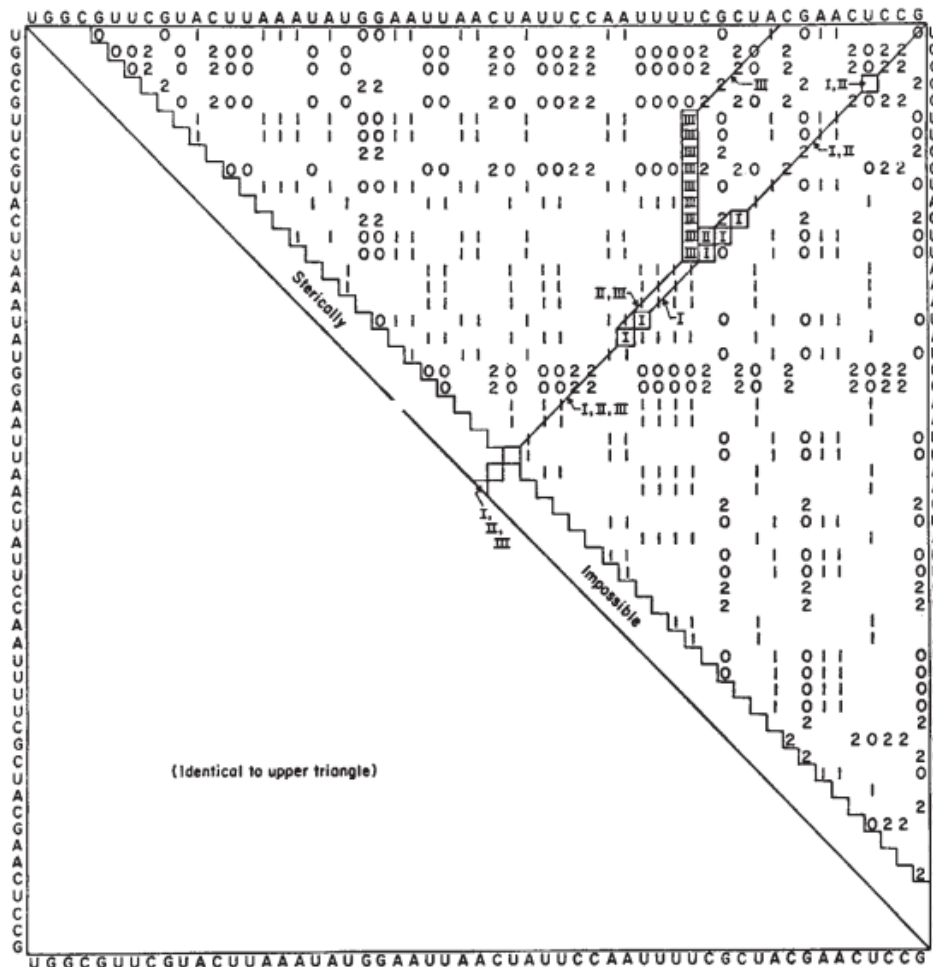
Fig. 2 The base pairing matrix showing all possible base pairs in a sequence of fifty-five bases from R17 viral RNA³. An upper right to lower left diagonal line joining base pairs represents an anti-parallel helical region. Squares represent loops between helical regions. A bulge is seen as a vertical or horizontal line. The three most stable structures (I, II, III) are indicated here and in Fig. 3.



Tři nejlepší řešení podle autorů. Všimněte si kladných bodů za páry bází a záporných bodů za smyčky a výdutě.

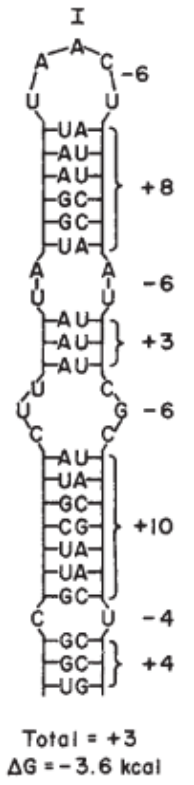
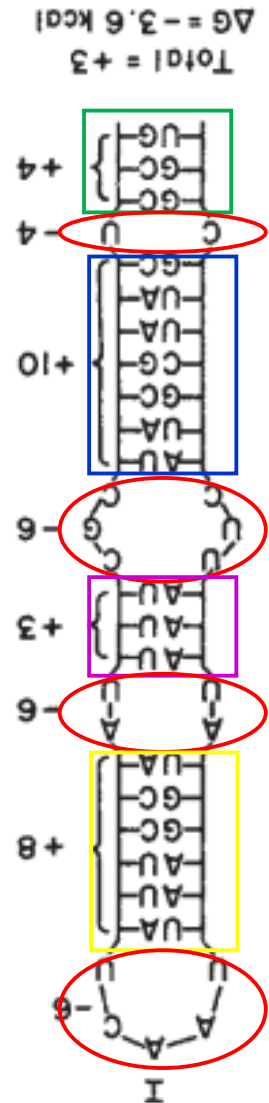
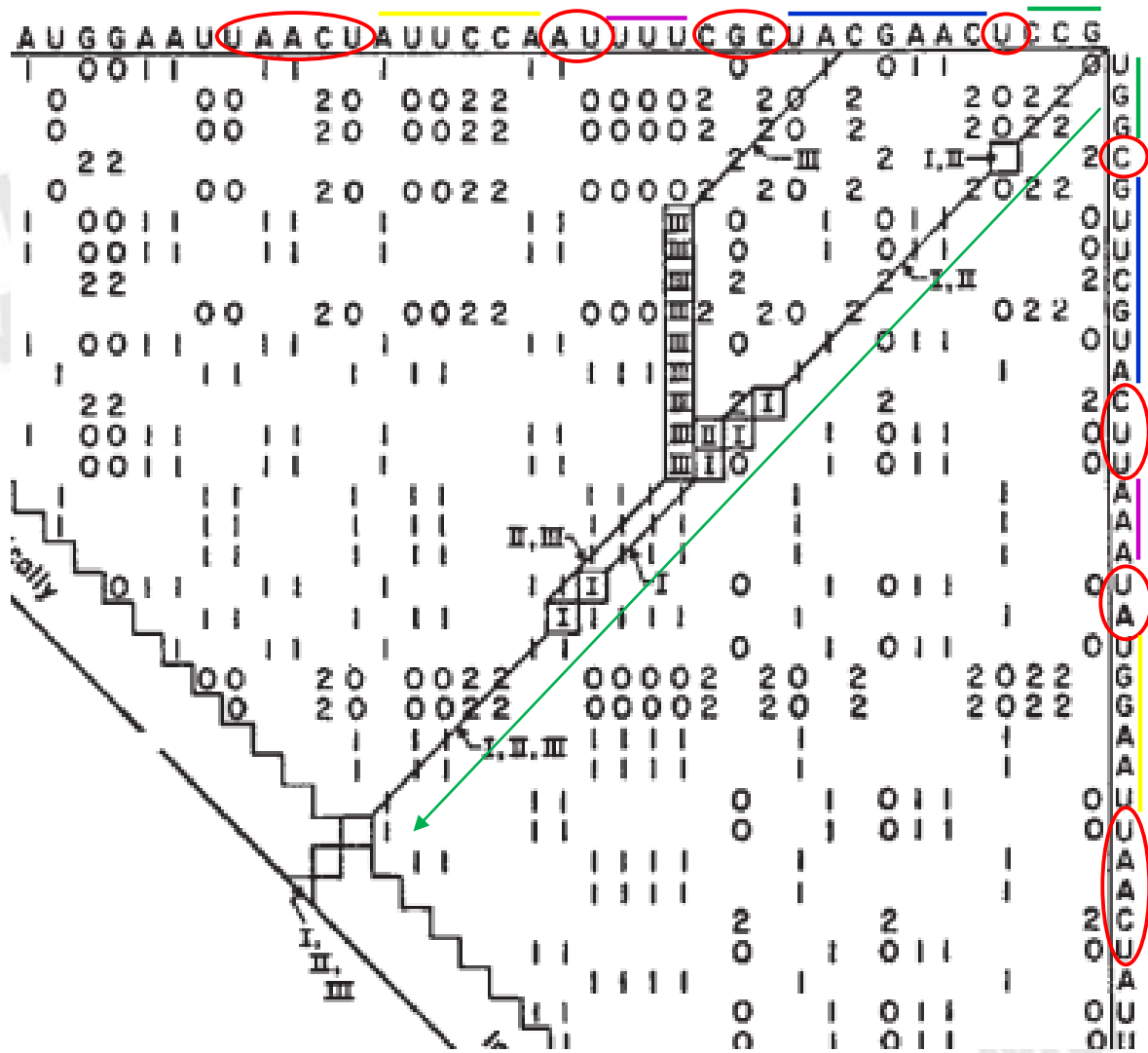
Ab initio – výlet do historie

Fig. 2 The base pairing matrix showing all possible base pairs in a sequence of fifty-five bases from R17 viral RNA³. An upper right to lower left diagonal line joining base pairs represents an anti-parallel helical region. Squares represent loops between helical regions. A bulge is seen as a vertical or horizontal line. The three most stable structures (I, II, III) are indicated here and in Fig. 3.



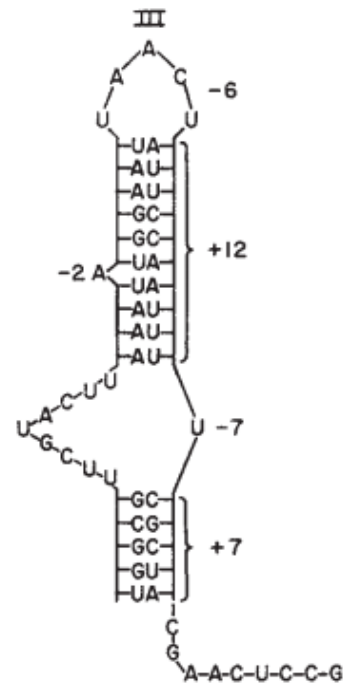
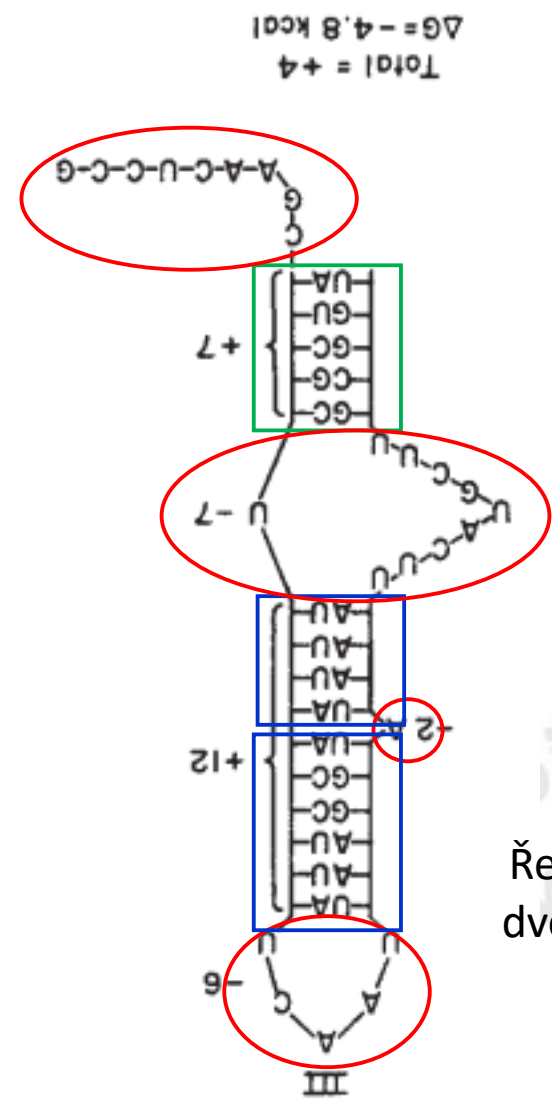
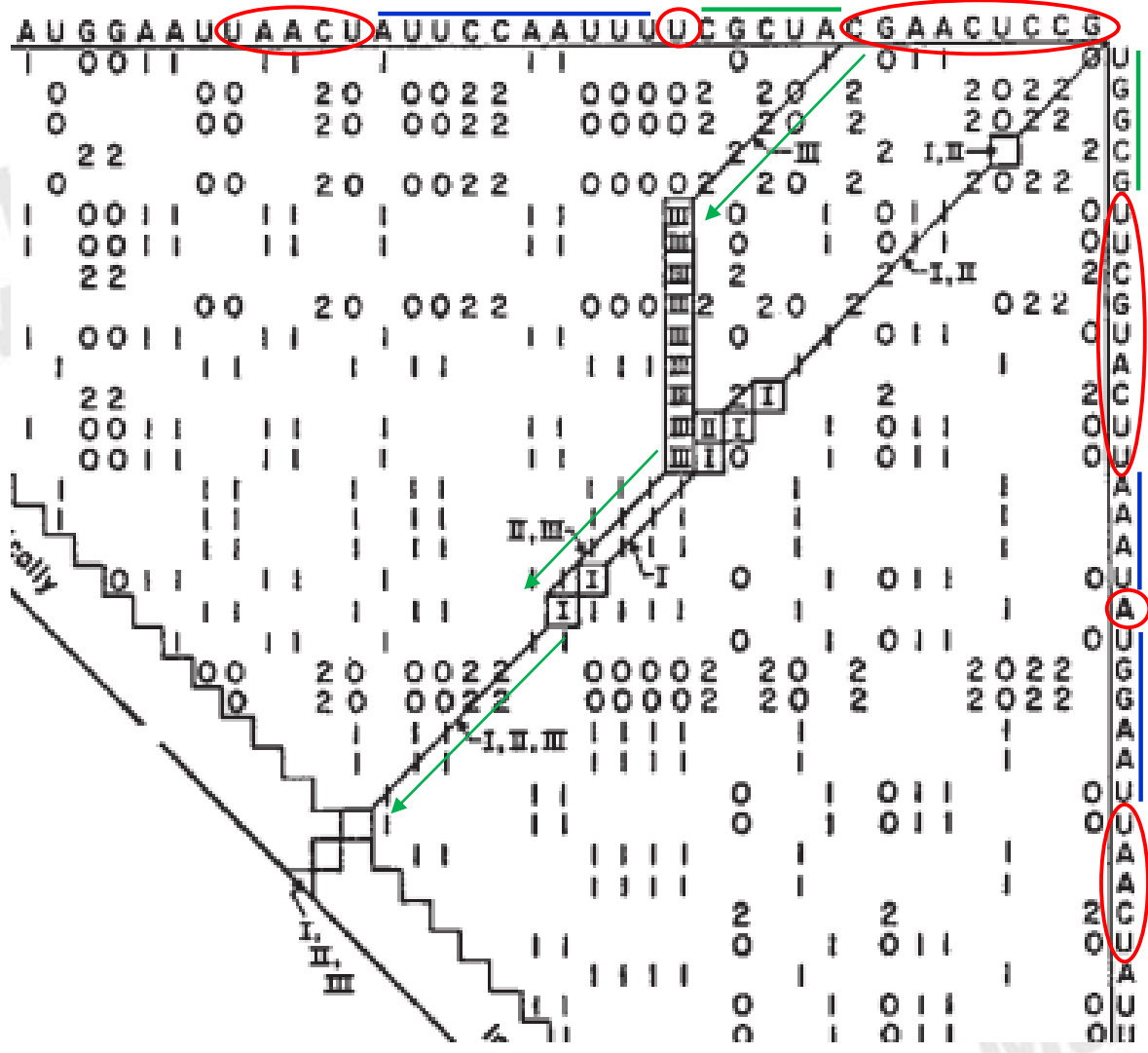
Tři nejlepší řešení podle autorů. Všimněte si kladných bodů za páry bází a záporných bodů za smyčky a výdutě.

Ab initio – výlet do historie



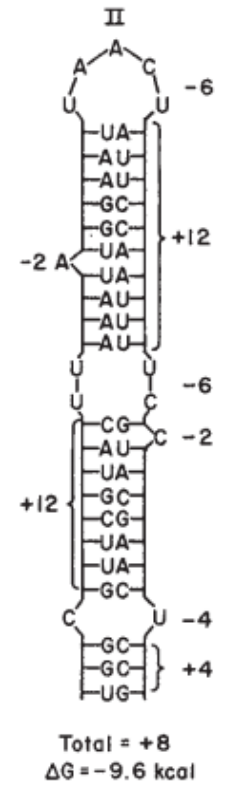
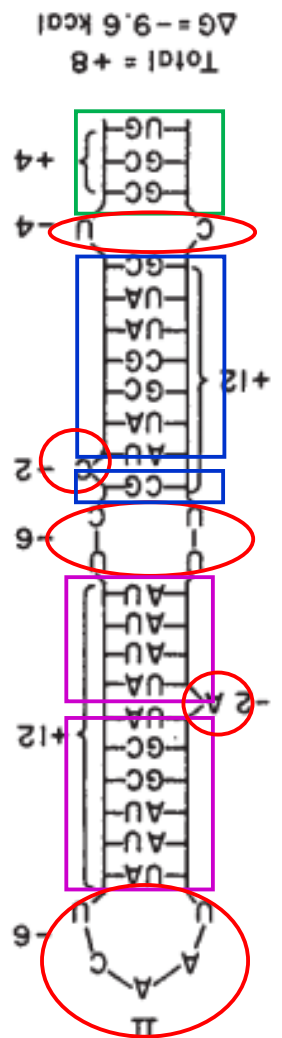
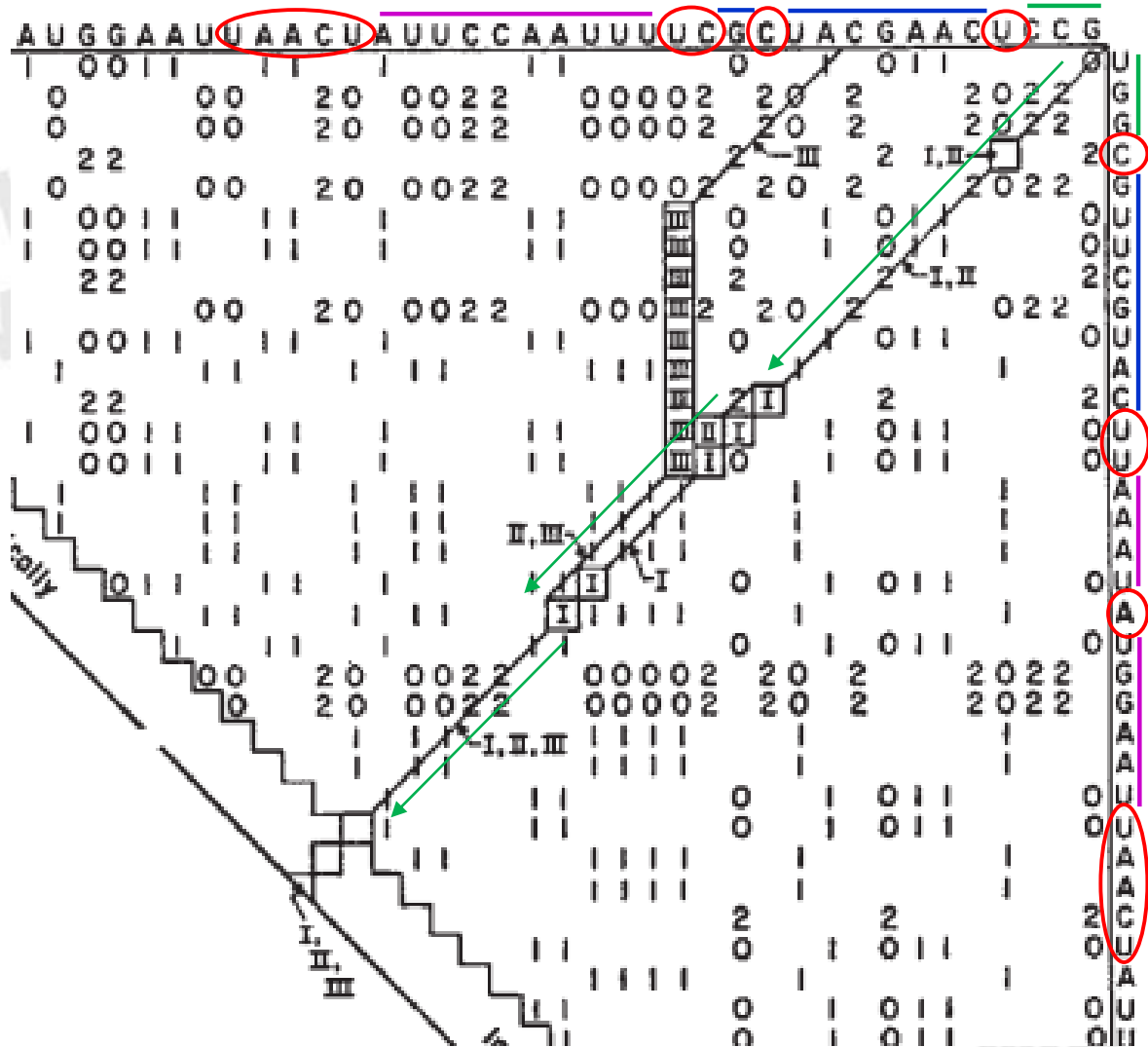
Řešení I se zbaběle drží nejdelší, byť přerušované, diagonály.

Ab initio – výlet do historie



Řešení III zkouší zapojit další dvě diagonály...vyplatí se to?

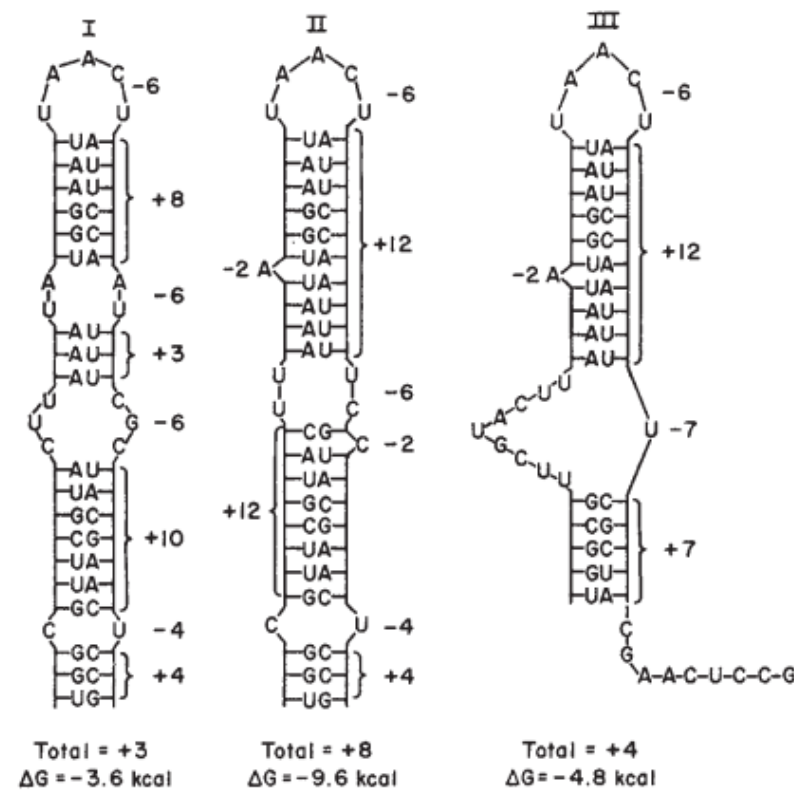
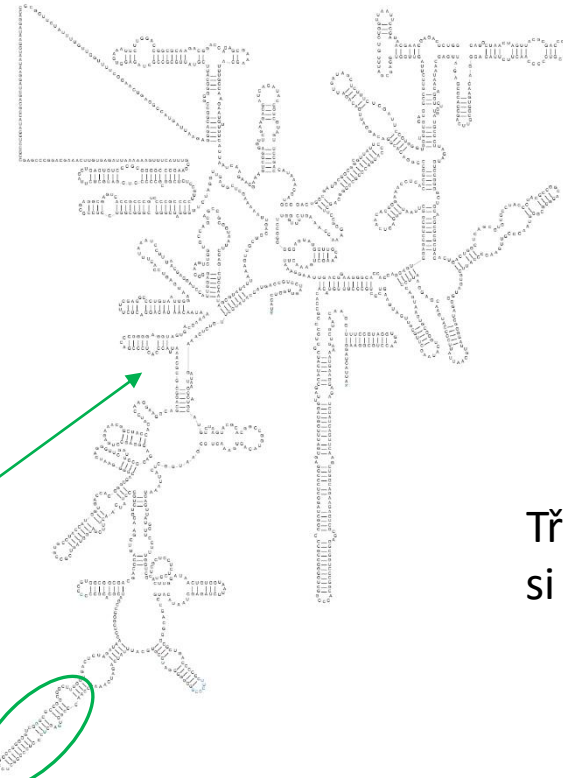
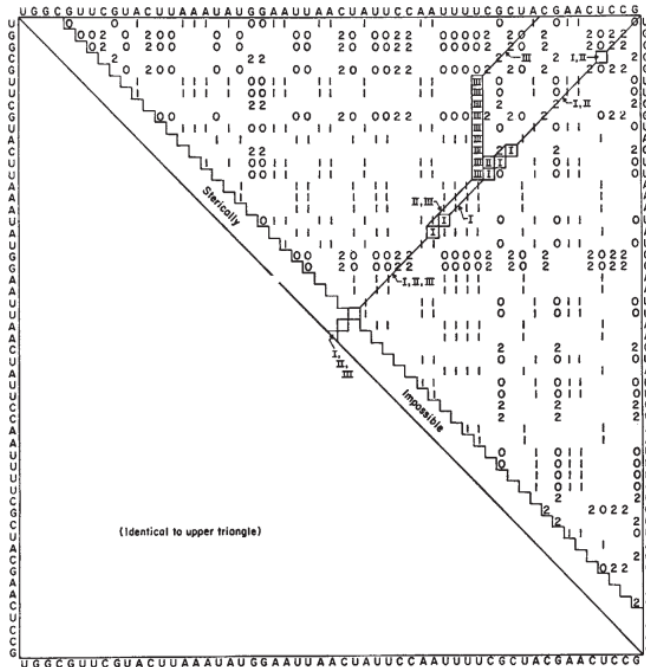
Ab initio – výlet do historie



Řešení II...kombinace obou?

Ab initio – výlet do historie

Fig. 2 The base pairing matrix showing all possible base pairs in a sequence of fifty-five bases from R17 viral RNA³. An upper right to lower left diagonal line joining base pairs represents an anti-parallel helical region. Squares represent loops between helical regions. A bulge is seen as a vertical or horizontal line. The three most stable structures (I, II, III) are indicated here and in Fig. 3.



- Přímý výběr může rozumně fungovat pro krátké sekvence, ideálně s jedním typem sekundární struktury (např. identifikace vlásenek).
- Výběr energeticky nejlepší kombinace možností pro větší molekuly by byl extrémně obtížný.
- Jednoduchá na provedení, náročná na vyhodnocení.

Tři nejlepší řešení podle autorů. Všimněte si kladných bodů za páry bází a záporných bodů za smyčky a výdutě.

RNA – predikce 2D struktury

- Predikce sekundární struktury RNA je významný bioinformatický problém.
- Metody *ab initio*/ metody komparativní (srovnávací).
- *Ab initio* – predikce je založena na jediné (zkoumané) RNA sekvenci. Vyhledávání **nejstabilnější** RNA struktury = struktury s **nejnižší** volnou energií. Tvorba párů bází je energeticky výhodná, zjednodušeně řečeno hledáme tedy strukturu s maximálním možným počtem párů bází.

RNAfold WebServer

Paste or type your **sequence** here:

```
uggcgguucguacuuaaaauuggaauuaacuauuccaauuuucgcuacgaacuccg
```

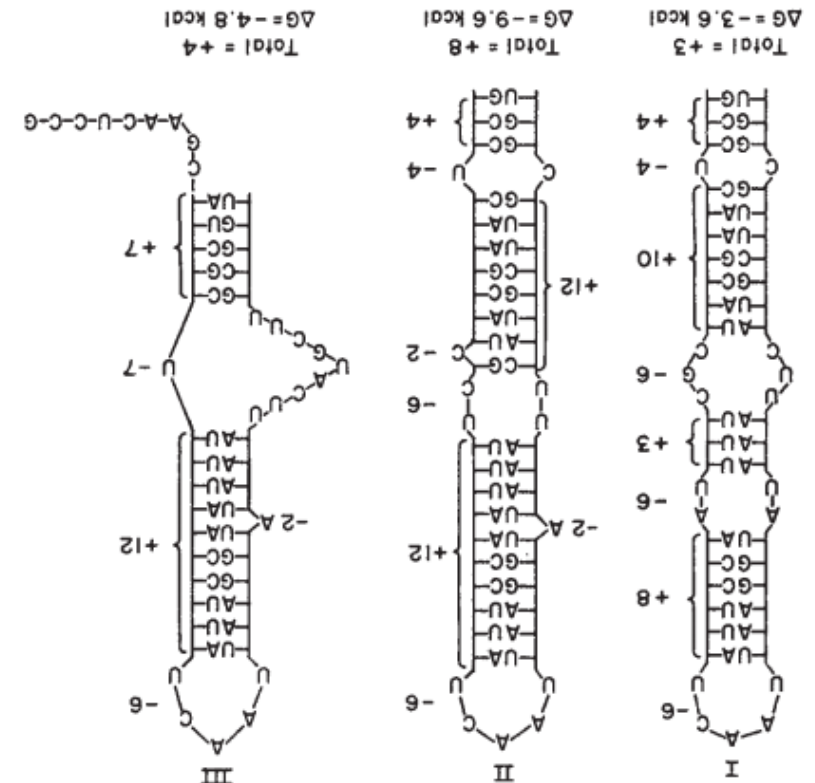
2003

Vienna RNA WebServers

Minimum free energy (MFE) structure

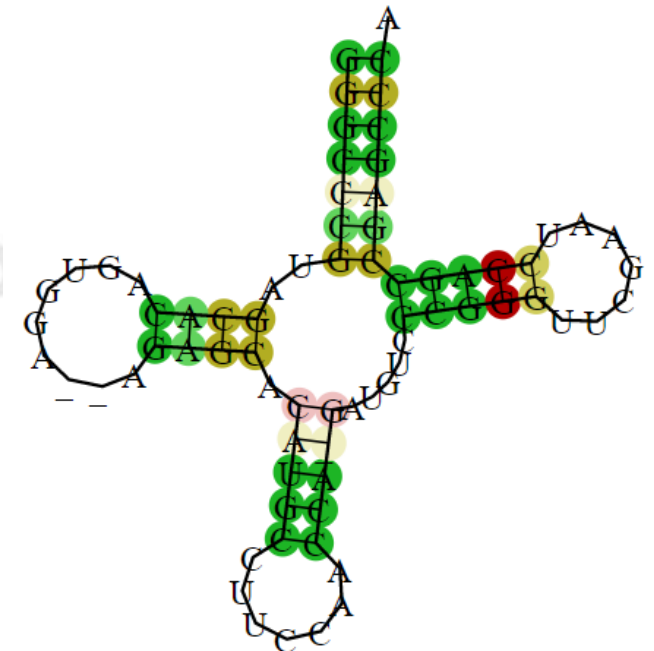
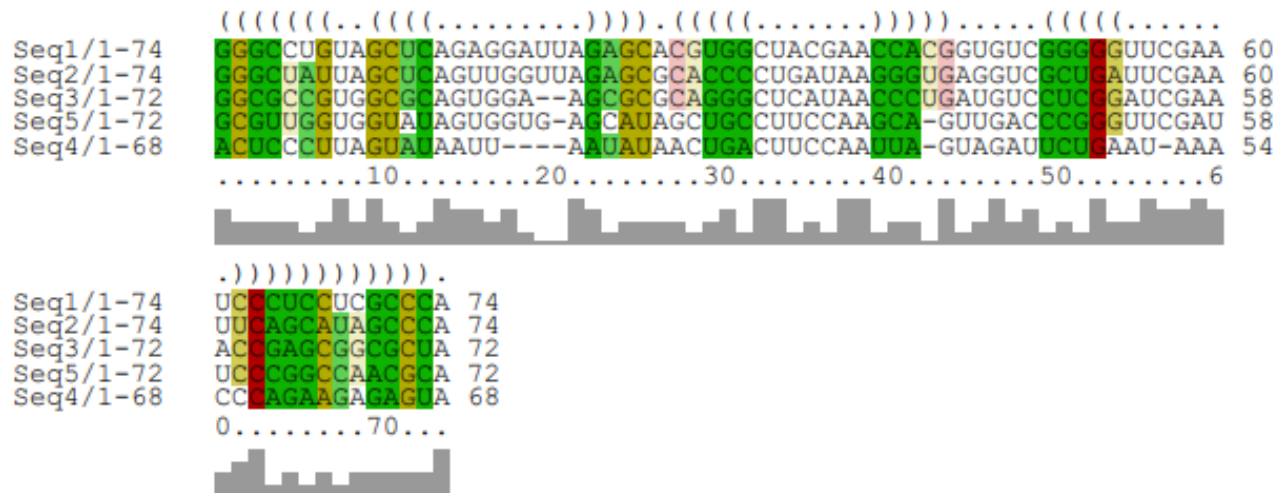
The MFE structure of an RNA sequence is the secondary structure that contributes a minimum of free energy. This structure is predicted using a loop-based energy model and the dynamic programming algorithm introduced by Zuker et al. [1]. As an RNA secondary structure can be uniquely decomposed into loops and external bases the loop-based energy model treats the free energy $F(s)$ of an RNA secondary structure s as the sum of the contributing free energies F_L of the loops L contained in s . According to the chosen energy parameter set and a given temperature (defaults to 37 °C) the secondary structure s that minimizes $F(s)$ is computed.

<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>



RNA – predikce 2D struktury

- Predikce sekundární struktury RNA je významný bioinformatický problém.
- Metody *ab initio*/ metody komparativní (srovnávací).
- Komparativní metody – využívají evolučně příbuzné RNA. Předpoklad: homologní sekvence RNA mají stejnou 2D strukturu. Multiple alignment sekvencí RNA, identifikace evolučně konzervované sekundární struktury. Teoreticky lepší, problémem je nutnost mít sadu homologních sekvencí.
- Jak identifikovat konzervované sekundární struktury u skupiny příbuzných RNA?
- Např. využití kovariance. Princip: funkční motivy RNA jsou strukturně konzervované. Mutace v pozici, která tvoří pár bází, musí být kompenzována mutací v odpovídající komplementární pozici.



RNA – predikce 2D struktury

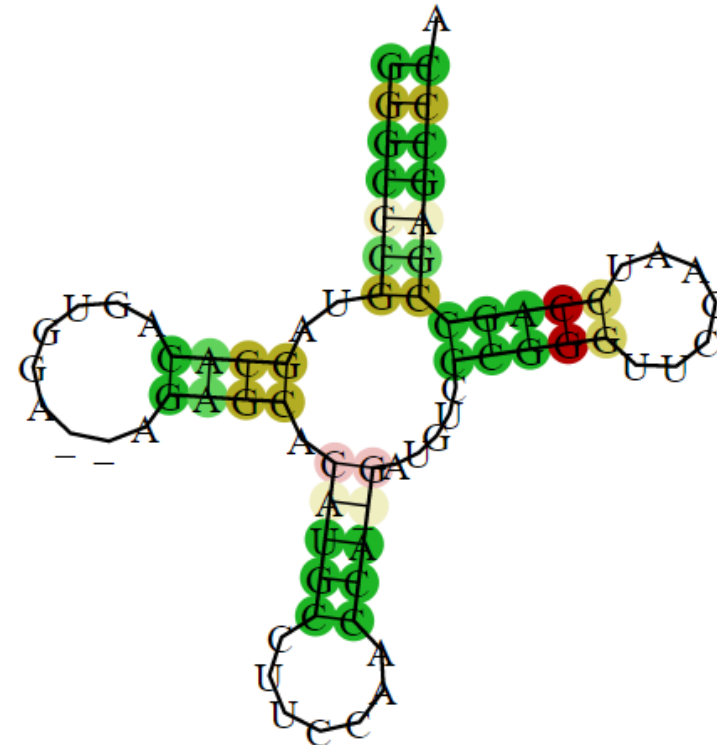
- Predikce sekundární struktury RNA je významný bioinformatický problém.
- Metody *ab initio*/ metody komparativní (srovnávací).
- Komparativní metody – využívají evolučně příbuzné RNA. Předpoklad: homologní sekvence RNA mají stejnou 2D strukturu. Multiple alignment sekvencí RNA, identifikace evolučně konzervované sekundární struktury. Teoreticky lepší, problémem je nutnost mít sadu homologních sekvencí.

RNAalifold WebServer

Paste your alignment(s) here:

CLUSTAL W (1.83) multiple sequence alignment

```
Seq1      GGGCCUGUAGCUCAGAGGAUUAGAGCACGUGGCUACGAACACGGUGUCGGGGUUCGAA
Seq2      GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCUGAUAAAGGGUGAGGUCGCUGAUUCGAA
Seq3      GGC GCCUGGGCGCAGUGGA--AGCGCGCAGGGCUCAUAAACCCUGAUGUCCUGGAUCGAA
Seq5      GCGUUGGUGGUUAGUGGUG-AGCAUAGCUGCCUCCAAGCA-GUUGACCCGGGUUCGAU
Seq4      ACUCCCUUAGUAUAUU---AAUAUAACUGACUCCAUAUA-GUAGAUUCUGAAU-AAA
          * * * * * ** ** * * * * *
```



Vienna RNA WebServers

Minimum free energy (MFE) structure

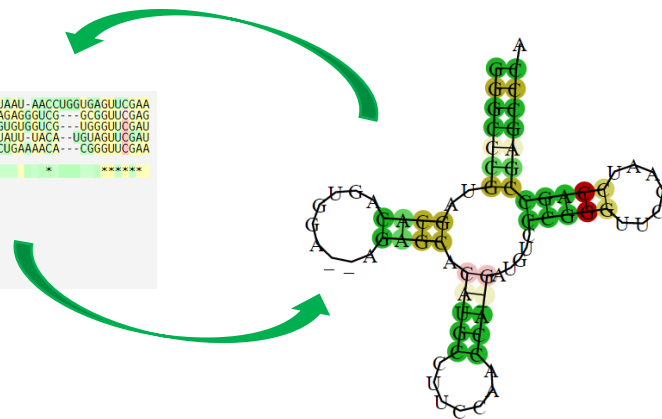
The MFE structure of an RNA sequence is the secondary structure that contributes a minimum of free energy. This structure is predicted using a loop-based energy model and the dynamic programming algorithm introduced by Zuker et al. [1]. As an RNA secondary structure can be uniquely decomposed into loops and external bases the loop-based energy model treats the free energy $F(s)$ of an RNA secondary structure s as the sum of the contributing free energies F_L of the loops L contained in s . According to the chosen energy parameter set and a given temperature (defaults to 37 °C) the secondary structure s that minimizes $F(s)$ is computed.

RNA – alignment

- RNA – nízká chemická komplexita, sekvence obsahuje pouze čtyři znaky.
- Nízká citlivost RNA alignmentu – zhoršená možnost použití **sekvenční** podobnosti pro odhad **biologického** významu alignmentu.
- RNA – evolučně konzervovaná **struktura** (kompenzační mutace).
- Sekvenční identita/podobnost může být u RNA špatným ukazatelem biologické podobnosti a evolučně konzervované struktury.
- **Řešení:** zohlednění sekundární struktury RNA při alignmentu.

| | |
|----------|--|
| D10744 | GGAAAAUUGAU - CAUCGGCAAGAU - AAGUUUUUACUAAAUAUAGGAAUUUAAU - AACCUUGGUGAUUCGAA |
| AF008220 | GGAGGAUAGCUCA - - GCUGGGA - GAGCAUCGCCU - - - - UA - CAAGCAGAGGGUUCG - - - GCGGUUCGAG |
| Z11880 | GCCUUCUAGCUCA - - GU - GGUA - GAGCGCACGGCU - - - - UU - UAACCGUUGUGUCG - - - UGGUUUCGAU |
| X02172 | GCCUUUUAJAGC - - - UUAG - UGUUA - AAGCGAUAAACU - - - - G - AAGAUUUUUU - UACA - - - UGUAGUUCGAU |
| M68929 | GCGGAUUAACUUA - - - GG - GGUAAAAGUUGCAGAUU - - - - G - UGGCUCUGAAAACA - - - CGGGUUCGAA |
| cons | * * * * * |

| | |
|----------|----------------|
| D10744 | UCUCACAUUUUCCG |
| AF008220 | CCCGUCAUCUCCA |
| Z11880 | CCCCACGGAAGGCG |
| X02172 | UCUCAUUAAGGCG |
| M68929 | UCCCGUUAUUCGCC |
| cons | * * * |



Alignment a predikce sekundární struktury RNA jsou velmi úzce související problémy.

Paste your alignment(s) here:

```
CLUSTAL W (1.03) multiple sequence alignment

Seq1      GGGCCUGUAGCUCAGAGGAUUAGAGCACGUGGCCUACGAACCACGGUGCGGGGUUCGAA
Seq2      GGGCUAUUAGCUCAGUUGGUUAGAGCGACCCUGAAUAGGGUGAGGUCGUGAUUCGAA
Seq3      GGC GCCUGGGCGCAGUGGA - - AGCGCGCAGGGCUCUAACCCUGAUUCCUGGUAUCGAA
Seq5      GCGUUGGUGUUAUAGUGUG - AGCAUAGCUGCCUCCAAGCA - GUUGACCCGGGUUCGAU
Seq4      ACUCCCUUAGUUAUUU - - - AAUUAACUGACUCCAUAUA - GUAGAUUCUGAAU - AAA
                * * * * * * * * * * * * * * * * * *
```

R-Coffee

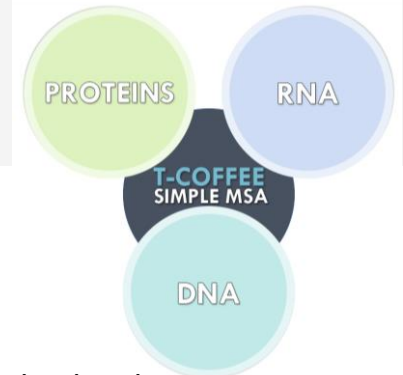
Aligns RNA sequences using predicted secondary structures

T-COFFEE, Version_11.00.d625267 (2016-01-11 15:25:41 - Revision d625267 - Build 507)
 Cedric Notredame -
 SCORE=343

| | BAD | AVG | GOOD |
|----------|-----|-----|------|
| D10744 | 24 | | |
| AF008220 | 40 | | |
| Z11880 | 41 | | |
| X02172 | 33 | | |
| M68929 | 32 | | |
| cons | 34 | | |

| | |
|----------|--|
| D10744 | GGAAAAUUGAU - CAUCGGCAAGAU - AAGUUUUUACUAAAUAUAGGAAUUUAAU - AACCUUGGUGAUUCGAA |
| AF008220 | GGAGGAUAGCUCA - - GCUGGGA - GAGCAUCGCCU - - - - UA - CAAGCAGAGGGUUCG - - - GCGGUUCGAG |
| Z11880 | GCCUUCUAGCUCA - - GU - GGUA - GAGCGCACGGCU - - - - UU - UAACCGUUGUGUCG - - - UGGUUUCGAU |
| X02172 | GCCUUUUAJAGC - - - UUAG - UGUUA - AAGCGAUAAACU - - - - G - AAGAUUUUUU - UACA - - - UGUAGUUCGAU |
| M68929 | GCGGAUUAACUUA - - - GG - GGUAAAAGUUGCAGAUU - - - - G - UGGCUCUGAAAACA - - - CGGGUUCGAA |
| cons | * * * * * |

| | |
|----------|----------------|
| D10744 | UCUCACAUUUUCCG |
| AF008220 | CCCGUCAUCUCCA |
| Z11880 | CCCCACGGAAGGCG |
| X02172 | UCUCAUUAAGGCG |
| M68929 | UCCCGUUAUUCGCC |
| cons | * * * |

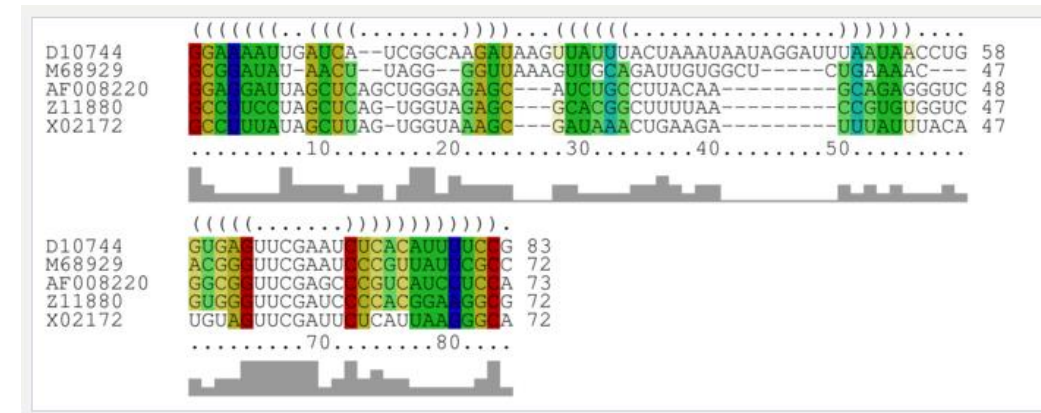
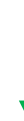


RNA – alignment

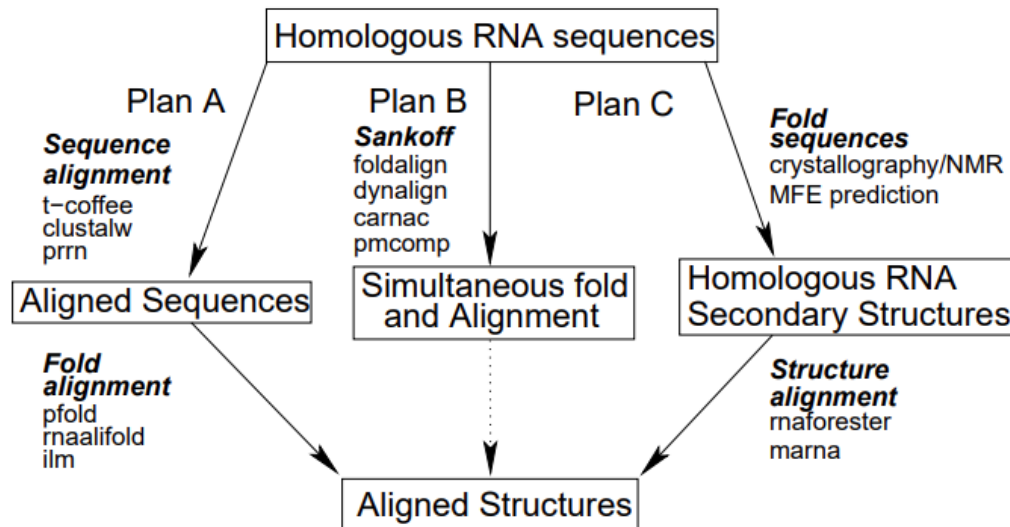
- RNA – nízká chemická komplexita, sekvence obsahuje pouze čtyři znaky.
- Nízká citlivost RNA alignmentu – zhoršená možnost použití **sekvenční** podobnosti pro odhad **biologického** významu alignmentu.
- RNA – evolučně konzervovaná **struktura** (kompenzační mutace).
- Alignment a predikce sekundární struktury RNA jsou velmi úzce související problémy.
- **Fold and align: souběžné řešení obou problémů.**



LocARNA is a tool for multiple alignment of RNA molecules. LocARNA requires **only RNA sequences** as input and will **simultaneously fold** and align the input sequences. Specifications of additional constraints or fixed input structures are possible. LocARNA outputs a multiple alignment together with a consensus structure. For the folding it makes use of a very realistic energy model for RNAs as it is by RNAfold of the Vienna RNA package (or Zuker's mfold).

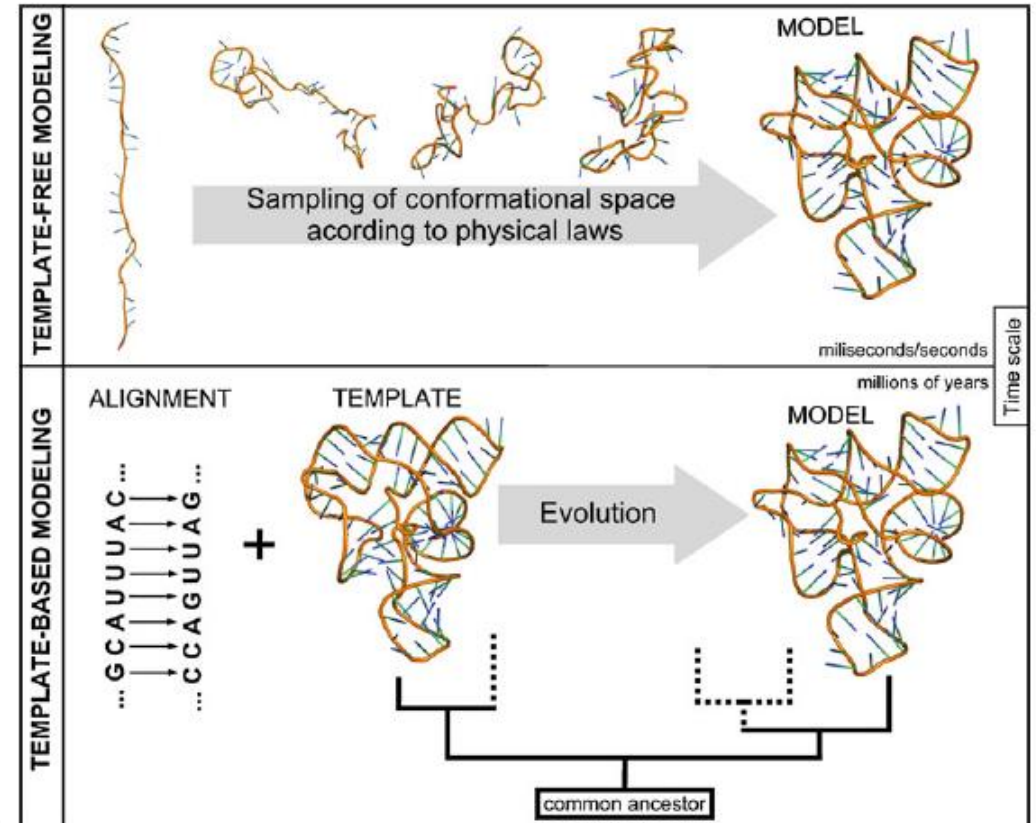


Alignment + **konsenzuální** sekundární struktura



RNA – predikce 3D struktury

- Vznik 3D struktury RNA *in vivo* (folding) mnohdy zahrnuje interakce s proteiny, ionty kovů, jinou RNA a dalšími typy molekul.
- RNA může tvořit mnoho dalších interakcí kromě kanonického párování bází.
- Komparativní metody vycházející z experimentálně získaných struktur (**template-based**).
- Simulace procesu foldingu (**template-free**):
Sekundární struktura \longrightarrow terciární struktura (model).
- **Hybridní metody** – kombinace obou přístupů (*de novo* modeling), významné zapojení experimentálně získaných poznatků (konformace ze známých struktur, fragment assembly). Nejúspěšnější.



RNA and protein 3D structure modeling: similarities and differences

Kristian Rother • Magdalena Rother •
Michał Boniecki • Tomasz Puton • Janusz M. Bujnicki

RNA – predikce 3D struktury

- Vznik 3D struktury RNA *in vivo* (folding) mnohdy zahrnuje interakce s proteiny, ionty kovů, jinou RNA a dalšími typy molekul.
- RNA může tvořit mnoho dalších interakcí kromě kanonického párování bází.
- Komparativní metody vycházející z experimentálně získaných struktur (**template-based**).
- Simulace procesu foldingu (**template-free**):

Sekundární struktura → terciární struktura (model).

- **Hybridní metody** – kombinace obou přístupů (*de novo* modeling), významné zapojení experimentálně získaných poznatků (konformace ze známých struktur, fragment assembly). Nejúspěšnější.

What is ModeRNA



ModeRNA is a program for comparative modeling of RNA 3D structures. It requires a pairwise sequence alignment and a structural template to generate a 3D structural model of the target RNA sequence via either a fully automated or script-based approaches. ModeRNA is capable of handling 115 different nucleotide modifications and bridging gaps using fragments derived from an extensive fragment library.

Apart from that, ModeRNA offers a multitude of functions to examine and modify RNA structure files. ModeRNA can be integrated into other programs via Python scripting.

Assemble2

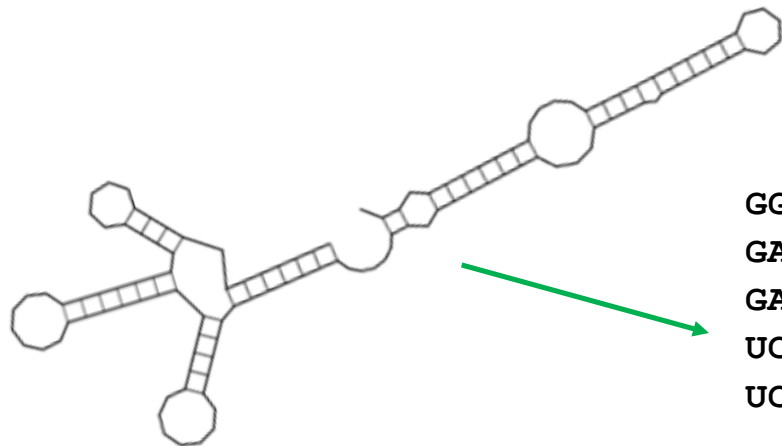
Study and construct RNA 3D architectures interactively...

Assemble2 allows you to design your RNA 2D structure interactively and to create and assemble the corresponding RNA 3D modules directly in UCSF Chimera:

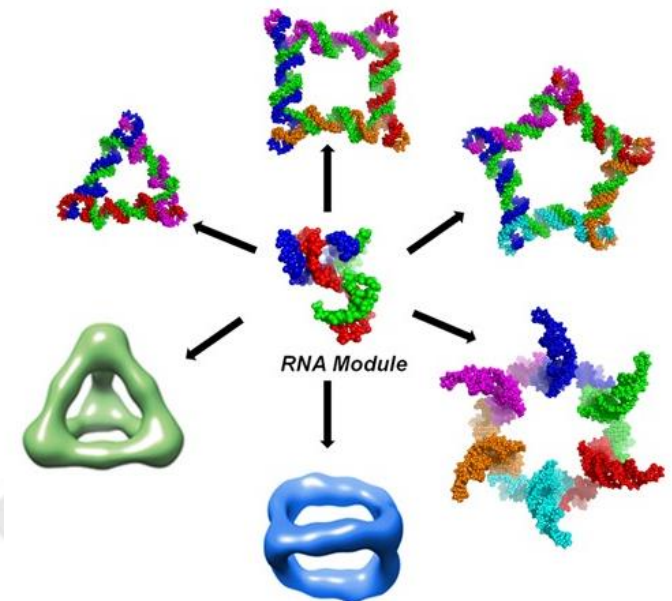
The approach that allows the expert user to rearrange and recombine multiple template structures has been particularly widely used in the RNA modeling field, with methods such as S2S/Assemble [68, 69], ERNA-3D [70], or RNA2D3D [71]. However, similar methods including the 'Frankenstein's Monster approach' [72] and the "protein lego" approach [73] have also been applied to model protein structures (review: [74]).

RNA – „inverse folding“

- **RNA:** může se vyskytovat v mnoha různých strukturních uspořádáních, což odráží množství nejrůznějších **funkcí**, které plní.
- **RNA desing:** návrh a syntéza sekvence RNA, která se složí do požadovaného tvaru a může vykonávat požadovanou funkci.
- Cílené doručování léčiv – RNA aptamery (specifická vazba na receptory buněk).
- Ribozymy (katalyticky aktivní molekuly RNA) – štěpení cílové RNA.
- Regulace exprese genů.
- Imunoterapie s využitím RNA nanočástic.



```
GGGAACCUGAUGCAUGUAGAUCGAAUG
GACUCUAAAUCCGUUCAGCCGGGUUA
GAUUCCTCGGGUUCGCAUGGCGU
UCGUACUAAAUAUGGAAUUAACUAU
UCCAAUUUUCGCUACGAACUCCG
```

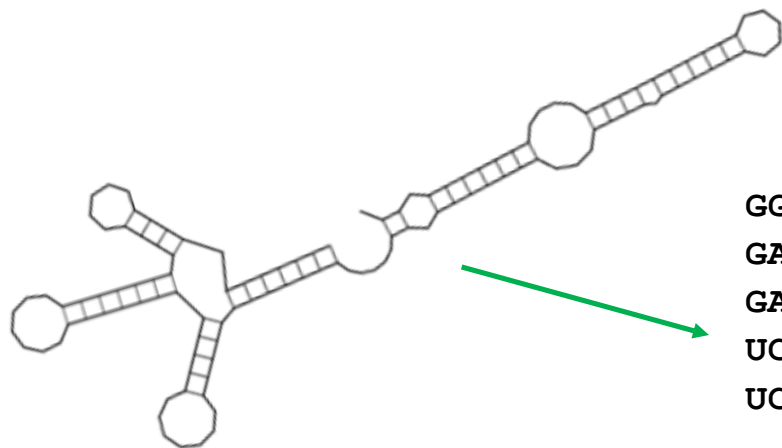


Advancement of the Emerging Field of RNA Nanotechnology

Daniel Jasinski,[†] Farzin Haque,[†] Daniel W Binzel, and Peixuan Guo^{*✉}

RNA – „inverse folding“

- **RNA:** může se vyskytovat v mnoha různých strukturních uspořádáních, což odráží množství nejrůznějších **funkcí**, které plní.
- **RNA desing:** návrh a syntéza sekvence RNA, která se složí do požadovaného tvaru a může vykonávat požadovanou funkci.
- Cílené doručování léčiv – RNA aptamery (specifická vazba na receptory buněk).
- Ribozymy (katalyticky aktivní molekuly RNA) – štěpení cílové RNA.
- Regulace exprese genů.
- Imunoterapie s využitím RNA nanočástic.



```
GGGAACCUGAUGCAUGUAGAUCGAAUG  
GACUCUAAAUCCGUUCAGCCGGGUUA  
GAUUCCTGGGGUUUCCGCCAUGGCGU  
UCGUACUAAAUAUGGAAUUAACUAU  
UCCAAUUUUCGCUACGAACUCCG
```



<https://eternagame.org/home/>

Eterna empowers citizen scientists to invent medicine through **molecular design**.

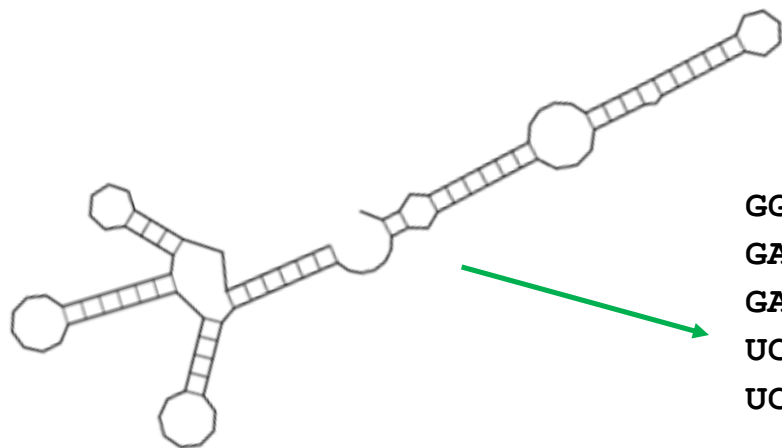
You play by solving puzzles using RNAs, tiny molecules at the heart of every cell.

Play Now

About Eterna

RNA – „inverse folding“

- **RNA:** může se vyskytovat v mnoha různých strukturních uspořádáních, což odráží množství nejrůznějších **funkcí**, které plní.
- **RNA desing:** návrh a syntéza sekvence RNA, která se složí do požadovaného tvaru a může vykonávat požadovanou funkci.
- Cílené doručování léčiv – RNA aptamery (specifická vazba na receptory buněk).
- Ribozymy (katalyticky aktivní molekuly RNA) – štěpení cílové RNA.
- Regulace exprese genů.
- Imunoterapie s využitím RNA nanočástic.



```
GGAACCUGAUGAUGAUGAUGAUGAUG  
GACUCUAAAUCCGUUCAGCCGGGUUA  
GAUUCCTGGGGUUUCGCAUGGCGU  
UCGUACUAAAUAUGGAAUUAACUAU  
UCCAAUUUUCGCUACGAACUCCG
```



10 Years of Discovery

This year we celebrate 10 years of citizen science through Eterna. Together we've synthesized thousands of player-designed molecules, published over 20 academic papers tackling fundamental RNA design questions, and pioneered the open development of RNA medicines. Over the course of the year we'll be rolling out new features and game mechanics as we continue to build the future of Eterna.

OpenVaccine: Eterna takes on COVID-19

With the onset of the COVID-19 pandemic, the Eterna project is looking to harness online gamers toward a solution. Their mission? To develop mRNA vaccines stable enough to be deployed to everyone in the world and not just a privileged few. For more info, see the [announcement](#) and [crowdfunding campaign](#).



RNA – „inverse folding“

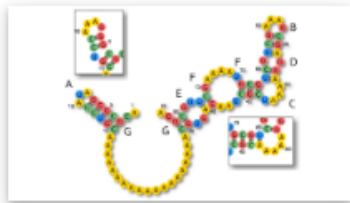


NEMO

Solving the RNA inverse folding problem, also known as the RNA design problem, is critical to advance several scientific fields like bioengineering, yet existing approaches have had limited success. NEMO combines a different technique, Nested Monte Carlo Search (NMCS), with domain-specific knowledge to create an algorithm that outperforms all prior published methods by wide margins and solves 95 of the 100 puzzles listed in a recently proposed RNA solving difficulty benchmark.

↓ DOWNLOAD

📖 TUTORIAL



READ THE PAPER

An unexpectedly effective Monte Carlo technique for the RNA inverse folding problem

bioRxiv - June 2018

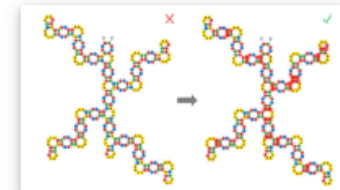


SentRNA

SentRNA is a fully-connected neural network trained using the eternasolves dataset. The agent first predicts an initial sequence for a target using the trained network, and then refines that solution if necessary using a short adaptive walk utilizing a canon of standard design moves. Through this approach, SentRNA can learn and apply human-like design strategies to solve several complex targets previously unsolvable by any computational approach.

↓ DOWNLOAD

📖 TUTORIAL



READ THE PAPER

SentRNA: Improving computational RNA design by incorporating a prior of human design strategies

arXiv - Mar 2018



Solve puzzles to design molecular medicines.



Get feedback from real experiments at Stanford's School of Medicine.



Work together to **write papers** for scientific peer review.



Propose your own puzzles to advance research and **invent medicine**.

RNA – „inverse folding“

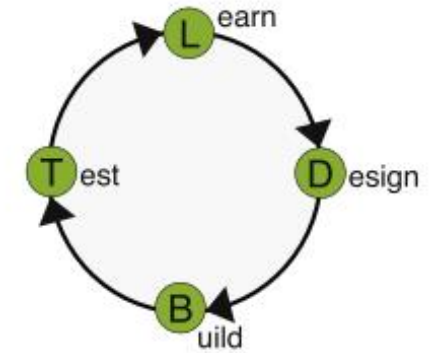
Article

<https://doi.org/10.1038/s41467-023-35827-3>

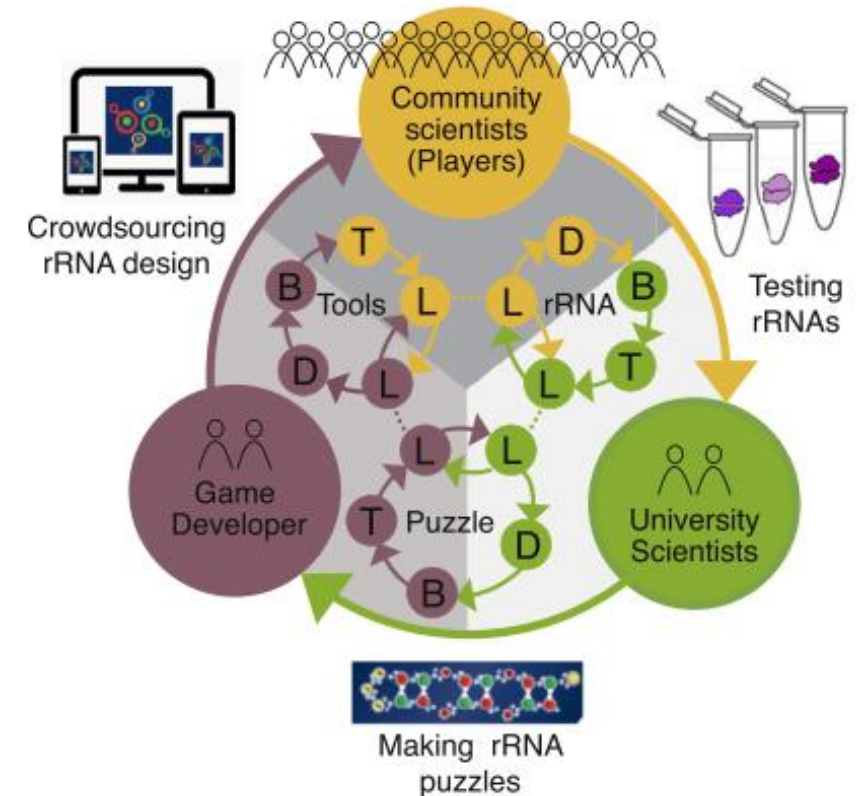
Community science designed ribosomes with beneficial phenotypes

Functional design of ribosomes with mutant ribosomal RNA (rRNA) can expand opportunities for understanding molecular translation, building cells from the bottom-up, and engineering ribosomes with altered capabilities. However, such efforts are hampered by cell viability constraints, an enormous combinatorial sequence space, and limitations on large-scale, 3D design of RNA structures and functions. To address these challenges, we develop an integrated community science and experimental screening approach for rational design of ribosomes. This approach couples Eterna, an online video game that crowdsources RNA sequence design to community scientists in the form of puzzles, with in vitro ribosome synthesis, assembly, and translation in multiple design-build-test-learn cycles. We apply our framework to discover mutant rRNA sequences that improve protein synthesis in vitro and cell growth in vivo, relative to wild type ribosomes, under diverse environmental conditions. This work provides insights into rRNA sequence-function relationships and has implications for synthetic biology.

TRADITIONAL APPROACH



ETERNA APPROACH

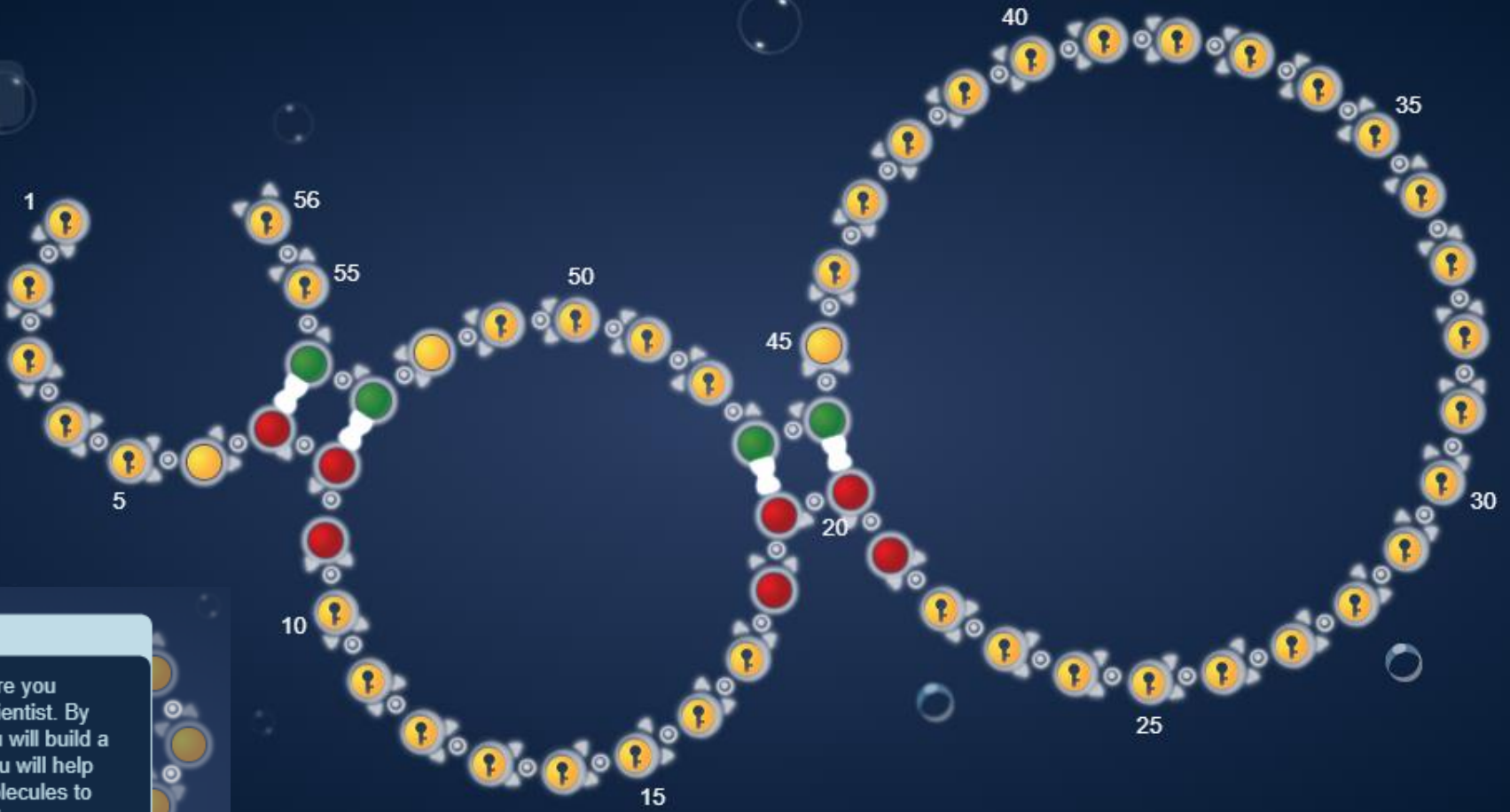


C2131 Úvod
Masaryko

[RNA Polymerase] Level 3 of 9



Total
-0.17 kcal



Welcome to Eterna

This is a game where you become an RNA scientist. By solving puzzles, you will build a virtual lab. Then you will help invent new RNA molecules to combat infectious diseases, including viral pandemics. Ready?

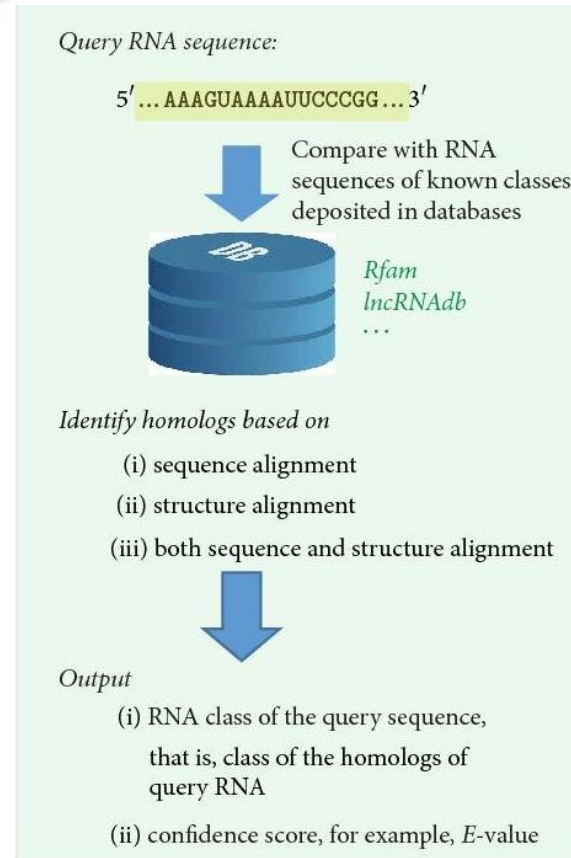
NEXT

Navigation and interaction controls including a leaf icon, a target icon, a sequence editor with nucleotide counts (A: 0, U: 0, G: 4, C: 0), a refresh icon, zoom in (+) and zoom out (-) icons, left and right arrow icons, and a chat icon.

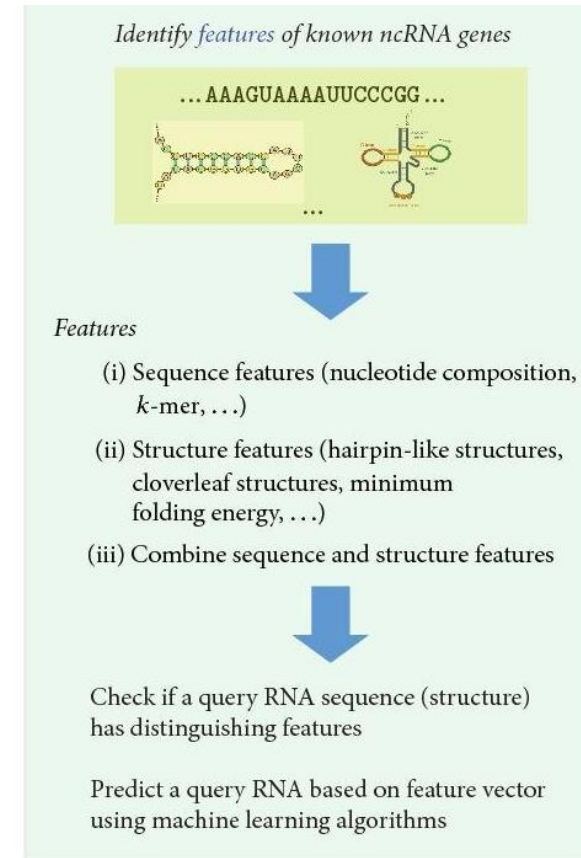


RNA – predikce ncRNA

- Metody založené na **homologii**. Musíme brát v úvahu sekvenční i strukturní homologii. Rychlé, ale málo citlivé.
- Metody využívající typické **vlastnosti a parametry** RNA. Vysoká falešná pozitivita.
- Metody založené na **sekvenování**. Izolace a sekvenace všech RNA, odstranění transkriptů kódujících proteiny a známých ncRNA.
- **Specifické** metody používané pro konkrétní rodiny ncRNA.
- miRNA – velmi krátké (22 nt), málo sekvenčně konzervované. Relativně konzervovaná sekundární struktura.
- lncRNA – nízká sekvenční i strukturní konzervovanost, mohou podléhat sestřihu.



(a)



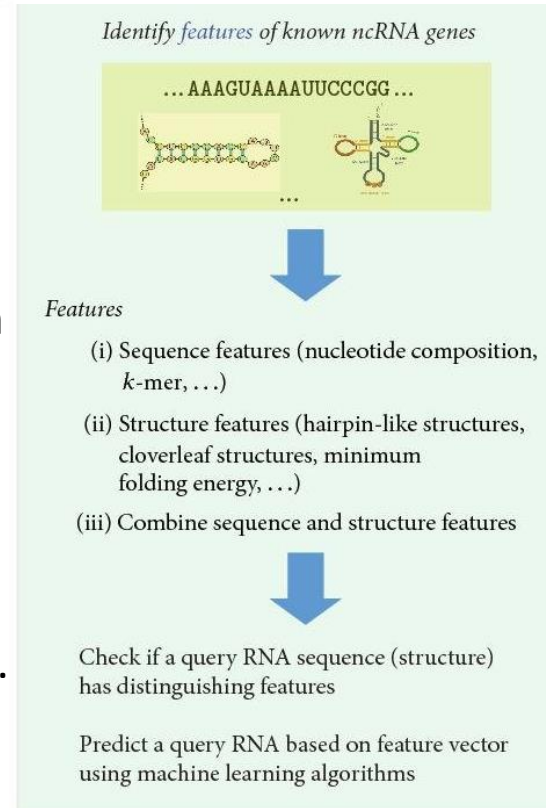
(b)

Review Article

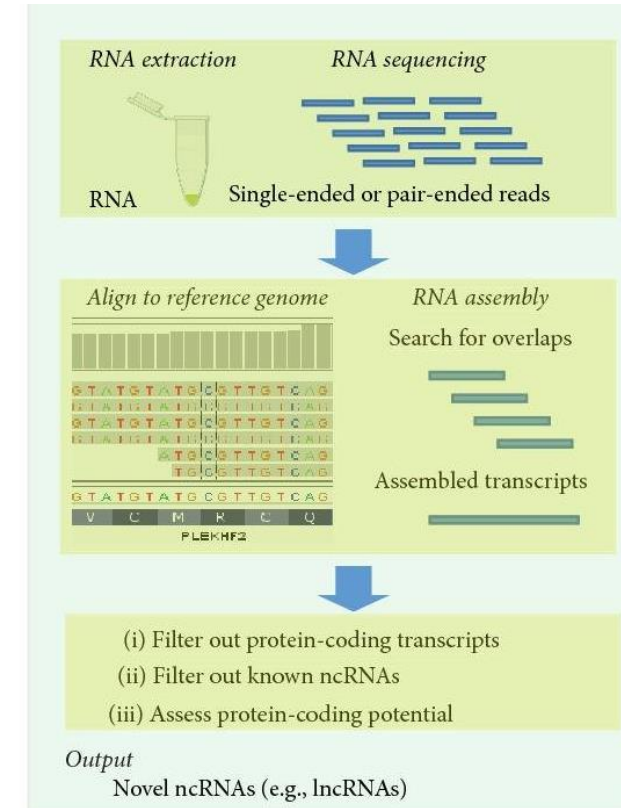
A Review on Recent Computational Methods for Predicting Noncoding RNAs

RNA – predikce ncRNA

- Metody založené na **homologii**. Musíme brát v úvahu sekvenční i strukturní homologii. Rychlé, ale málo citlivé.
- Metody využívající typické **vlastnosti a parametry** RNA. Vysoká falešná pozitivita.
- Metody založené na **sekvenování**. Izolace a sekvenace všech RNA, odstranění transkriptů kódujících proteiny a známých ncRNA.
- **Specifické** metody používané pro konkrétní rodiny ncRNA.
- miRNA – velmi krátké (22 nt), málo sekvenčně konzervované. Relativně konzervovaná sekundární struktura.
- lncRNA – nízká sekvenční i strukturní konzervovanost, mohou podléhat sestřihu.



(b)



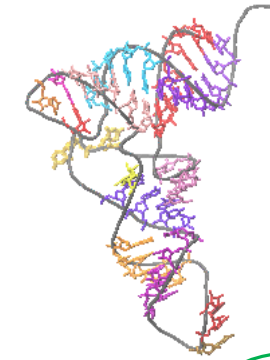
(c)

Review Article

A Review on Recent Computational Methods for Predicting Noncoding RNAs

RNA – predikce ncRNA

- Metody založené na **homologii**. Musíme brát v úvahu sekvenční i strukturní homologii. Rychlé, ale málo citlivé.
- Metody využívající typické **vlastnosti a parametry** RNA. Vysoká falešná pozitivita.
- Metody založené na **sekvenování**. Izolace a sekvenace všech RNA, odstranění transkriptů kódujících proteiny a známých ncRNA.
- Predikce tRNA – největší ncRNA rodina, výskyt u všech živých organismů. Charakteristická sekundární struktura.
- tRNAscan-SE – využití kovariančních modelů, predikce sekundární struktury identifikovaných tRNA a funkce (antikodon).



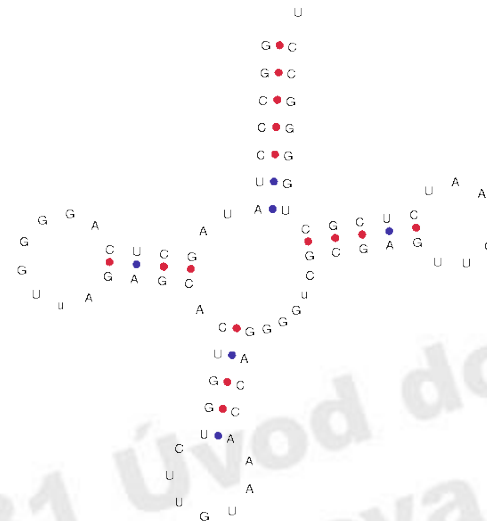
(Total length of sequence(s) = 36534 bp)

Please Wait...

Search Completed

| Predicted tRNA genes | |
|--|----------|
| tRNAs decoding Standard 20 AA | 4 |
| Selenocysteine tRNAs (TCA) | 1 |
| Possible suppressor tRNAs (CTA,TTA,TCA) | 0 |
| tRNAs with undetermined/unknown isotypes | 0 |
| tRNAs with mismatch isotypes | 1 |
| Predicted pseudogenes | 0 |
| Total tRNAs | 6 |

| |
|------------------------------|
| tRNAs with introns: 1 |
| Ser-CGA: 1 |



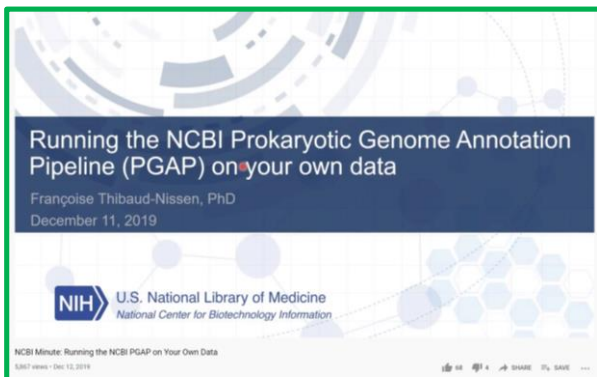
GtRNAdb

tRNAscan-SE analysis of complete genomes

The genomic tRNA database contains tRNA gene predictions made by tRNAscan-SE on complete or nearly complete genomes. Unless otherwise noted, all annotation is automated, and has not been inspected for agreement with published literature.

Vyhledání homologních sekvencí v databázi

Anotace genomů



PGAP is now available as a [stand-alone software package](#). You can annotate your genomes on your own machine, local cluster or the Cloud! Get started by watching a [short video](#)!

NCBI Prokaryotic Genome Annotation Pipeline

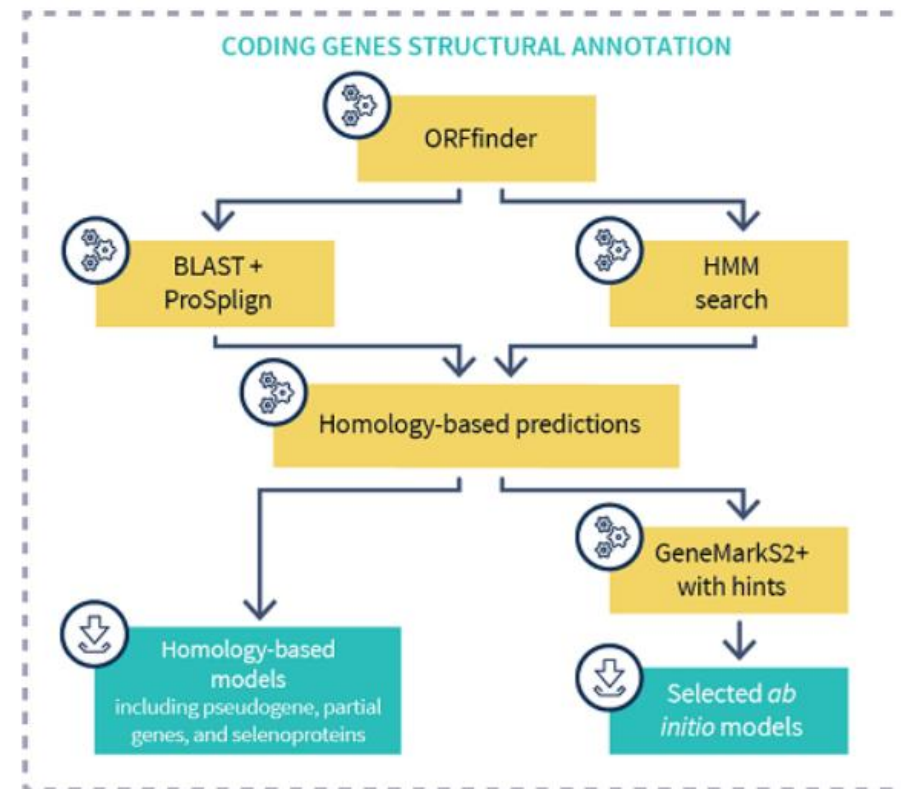
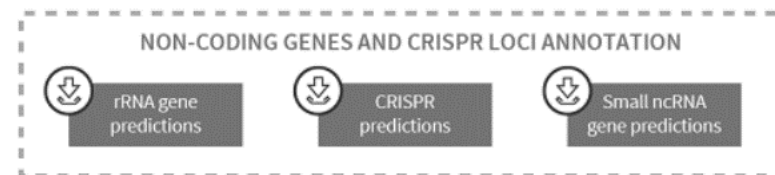
The **NCBI Prokaryotic Genome Annotation Pipeline** (PGAP) is designed to annotate bacterial and archaeal genomes (chromosomes and plasmids).

Genome annotation is a multi-level process that includes prediction of protein-coding genes, as well as other functional genome units such as structural RNAs, tRNAs, small RNAs, pseudogenes, control regions, direct and inverted repeats, insertion sequences, transposons and other mobile elements.

tRNAs

To identify tRNA genes, the input genome sequence is split into ~200nt windows with overlap of ~100nt and passed through tRNAscan-SE. [tRNAscan-SE](#) identifies 99–100% of transfer RNA genes in DNA sequence while giving less than one false positive per 15 gigabases. It is currently one of the most powerful tRNA identification tools, and uses different, targeted parameter sets for Archaea and Bacteria. tRNA predictions below a tRNAscan-SE score of 20 are discarded.

https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

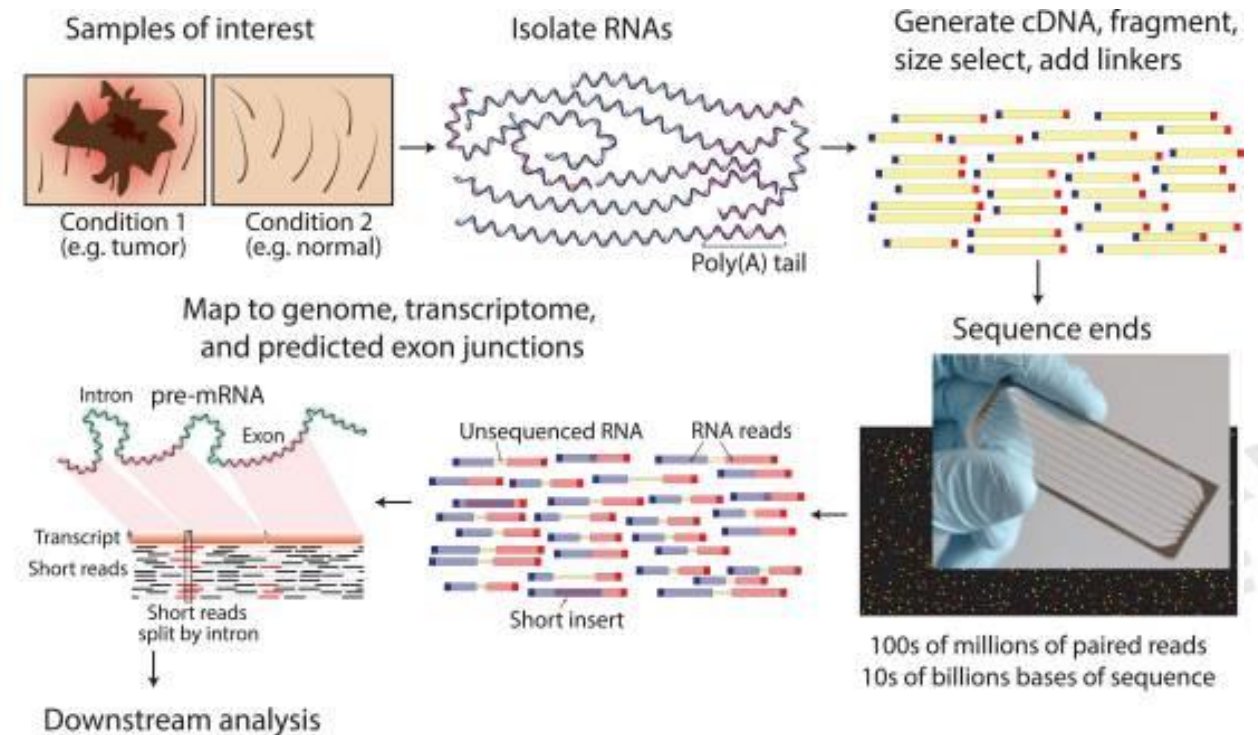


Structural annotation of coding genes feeds into functional annotation



RNA-Seq („RNA-sequencing“)

- Sekvenování **transkriptomu** s využitím moderních „high-throughput“ technologií.
- Expresní profilování, studium alternativního sestřihu, identifikace nových typů transkriptů (RNA).
- Náročný vícekrokový proces, nutná důsledná kontrola dat.
- **Bioinformatika**: kontrola kvality dat, rekonstrukce transkriptomu (např. mapování na referenční genom/transkriptom), kvantifikace exprese, identifikace posttranskripčních modifikací.
- **Metatranskriptomika** – studium celkové RNA skupiny organismů získané přímo z jejich přirozeného životního prostředí. Identifikace druhů, chování celé komunity za konkrétních životních podmínek.

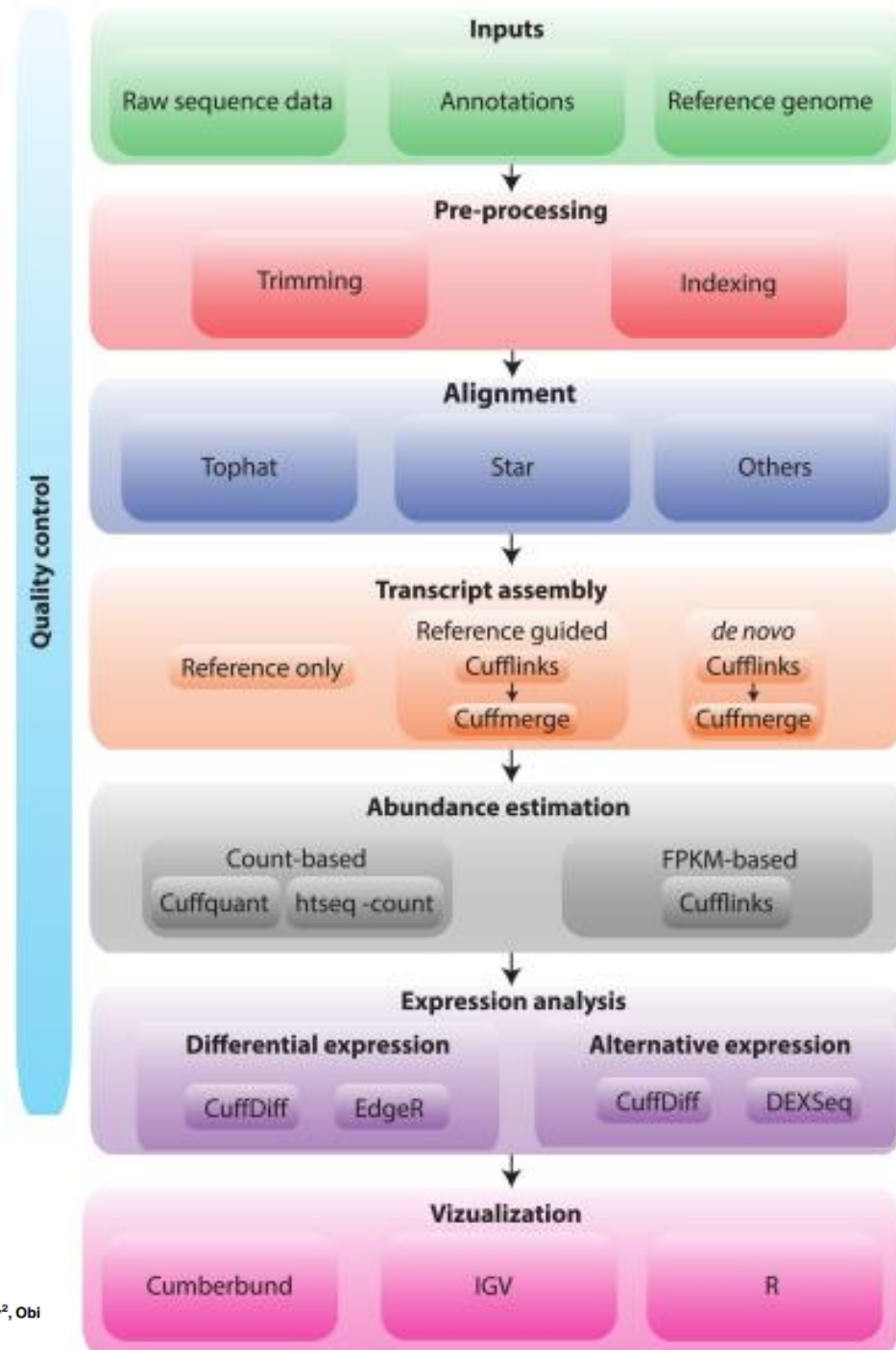


Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud

Malachi Griffith^{1,2,3*}, Jason R. Walker¹, Nicholas C. Spies¹, Benjamin J. Ainscough^{1,2}, Obi L. Griffith^{1,2,3,4*}

RNA-Seq („RNA-sequencing“)

- Sekvenování **transkriptomu** s využitím moderních „high-throughput“ technologií.
- Expresní profilování, studium alternativního sestřihu, identifikace nových typů transkriptů (RNA).
- Náročný vícekrokový proces, nutná důsledná kontrola dat.
- **Bioinformatika**: kontrola kvality dat, rekonstrukce transkriptomu (např. mapování na referenční genom/transkriptom), kvantifikace exprese, identifikace posttranskripčních modifikací.
- **Metatranskriptomika** – studium celkové RNA skupiny organismů získané přímo z jejich přirozeného životního prostředí. Identifikace druhů, chování celé komunity za konkrétních životních podmínek.

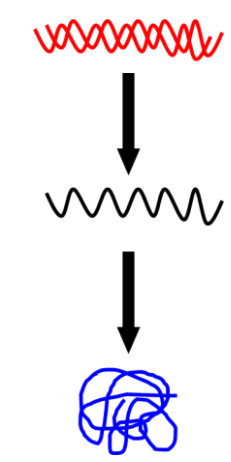
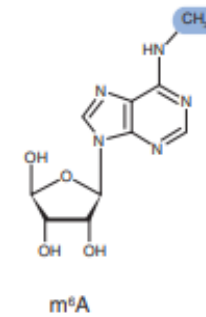
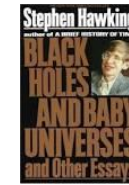


Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud

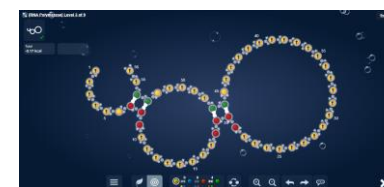
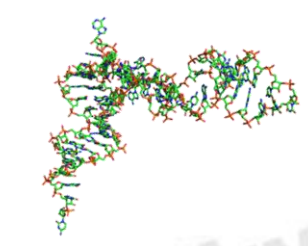
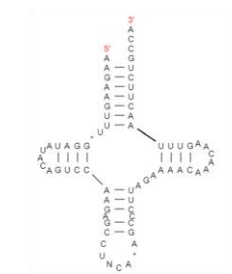
Malachi Griffith^{1,2,3*}, Jason R. Walker¹, Nicholas C. Spies¹, Benjamin J. Ainscough^{1,2}, Obi L. Griffith^{1,2,3,4*}

Shrnutí

- Tři hlavní formy RNA (mRNA, rRNA, tRNA) jsou zapojené do proteosyntézy.
- + mnoho dalších typů RNA, plnicích rozličné funkce (často regulační).
- Geny pro rRNA – fylogenetické analýzy.
- Predikční nástroje často využívané pro predikce modifikací a interakcí RNA.
- Primární/sekundární/terciární struktura RNA.
- Predikce 2D struktury: *ab initio*, komparativní metody.
- Predikce 3D struktury: template-based, template-free, hybridní metody.
- „Inverse folding“ – design RNA pro specifické účely.
- Alignment: nutné/vhodné zahrnout strukturní informace.
- Existuje transkriptomika a metatranskriptomika 😊



```
AAGAAGUUAGGAUUAACAGUCCA
GAGCCUCAAAGCCCUUAGAAAACA
AACAAAGUUUAACUUCUGCCA
```



...a zahrajte si Eternu!

Použitá a doporučená literatura

ADVANCED REVIEW

Cellular RNAs: Varied Roles

Jian Gu, *Baylor College of Medicine, Houston, Texas, USA*

Ram Reddy, *Baylor College of Medicine, Houston, Texas, USA*

RNA methylation in nuclear pre-mRNA processing

Helena Covelo-Molares | Marek Bartosovic[†] | Stepanka Vanacova

Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud

Malachi Griffith^{1,2,3,*}, Jason R. Walker¹, Nicholas C. Spies¹, Benjamin J. Ainscough^{1,2}, Obi L. Griffith^{1,2,3,4,*}

Messenger RNA in Eukaryotes

Göran Akusjärvi, *University of Uppsala, Uppsala, Sweden*

Jan-Peter Kreivi, *University of Uppsala, Uppsala, Sweden*

Svend Petersen-Mahrt, *London Research Institute, London, UK*

rRNA Structure

Arun Malhotra, *University of Miami School of Medicine, Miami, Florida, USA*

Estimation of Secondary Structure in Ribonucleic Acids

IGNACIO TINOCO, jun., OLKE C. UHLENBECK & MARK D. LEVINE

Department of Chemistry, University of California, Berkeley, California 94720

tRNA Structure

Eric Westhof, *Institute of Molecular and Cellular Biology, Strasbourg, France*

Pascal Auffinger, *Institute of Molecular and Cellular Biology, Strasbourg, France*

A simple method for estimating the secondary structure of an RNA molecule has been proposed on the basis of the knowledge of its sequence.

Alberts and Doty^{1,2}. By allowing looped out bases and putting the sequence in register in order to maximize base pairing, they were able to suggest optimal secondary structures from a given sequence. When this approach was applied to transfer RNA sequences, a wide variety of sequences could all be put into a similar cloverleaf pattern with the anticodon in the same position, thus demonstrating the interchangeability of the tRNAs in protein biosynthesis on the ribosome^{2,1}. Know-

Messenger RNA in Prokaryotes

How close are miRNAs from clinical practice? A perspective on the diagnostic and therapeutic market

Bonneau E.¹, Neveu B.², Kostantin E.³, Tsongalis G.J.⁴, De Guire V.²

¹ Department of Laboratory Medicine, Biochemistry Division, Saint-Eustache Hospital, Québec, Canada

² Biochemistry Division, Maisonneuve-Rosemont Hospital, Montreal, Québec, Canada

³ Department of Biochemistry, McGill University, Montreal, Québec, Canada

⁴ Department of Pathology and Laboratory Medicine, Geisel School of Medicine at Dartmouth and Dartmouth Hitchcock Medical Center, Lebanon, NH, United States of America

Transfer RNA

Emanuel Goldman, *New Jersey Medical School, Newark, New Jersey, USA*

Advancement of the Emerging Field of RNA Nanotechnology

Daniel Jasinski,[†] Farzin Haque,[†] Daniel W Binzel, and Peixuan Guo^{*@}

RNA and protein 3D structure modeling: similarities and differences

Kristian Rother · Magdalena Rother · Michal Boniecki · Tomasz Puton · Janusz M. Bujnicki

On the classification of long non-coding RNAs

Lina Ma,[†] Vladimir B. Bajic² and Zhang Zhang^{1,*}

Review Article

A Review on Recent Computational Methods for Predicting Noncoding RNAs