18.417 Introduction to Computational Molecular Biology

— Foundations of Structural Bioinformatics —

Sebastian Will

MIT, Math Department

Fall 2011

Credits: Slides borrow from slides of Jérôme Waldispühl
and Dominic Rose/Rolf Backofen

Before we start

Instructor: Sebastian Will

Contact: wills@mit.edu

Office hours: by appointment, Office: 2-155

Lecture: Tuesday, Thursday, 9:30-11:00 am

Room: 8-205

Web: `http://math.mit.edu/classes/18.417/`
(slides, further information)

Credits/Evaluation: *no* assignments, *no* exam, but *Final Project*

Final Project:

- study paper in depth, implement/extend algorithm, **or** theoretical proof
- project report (2-4 pages), talk (20 min)
- find a topic during term

What is Computational Molecular Biology (a.k.a. Bioinformatics)?

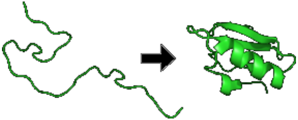
Short answer: study of computational approaches to study of biological systems (at the molecular level)

Today: somewhat longer answer, including

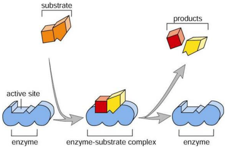
- What are the components of biological systems?
- How do they work together?
- What is their chemistry and structure?
- Which aspects do we want to study in Computational Biology?
- What is *Structural* Bioinformatics?
- What can you learn in this course?

Components of Biological Systems

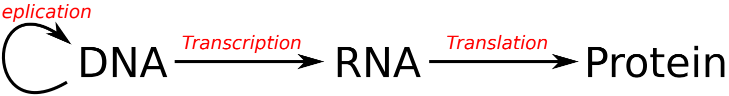
- Three classes of *biological macromolecules*:
    - DNA (= deoxyribonucleic acid)
    - RNA (= ribonucleic acid)
    - Protein
- Single molecules are linear chains of building blocks, specified by *sequence* of their building blocks, e.g. ACTGGAGCGTC.
- Molecules form 3D-*structures*. Folding is a physical process (*minimize energy*)



- “Levinthal Paradox”: fast folding but huge conformation space
- Structure allows macromolecules to interact. *Structure=Function*, e.g. 'lock&key'

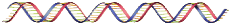


Information Flow — Central Dogma



Replication

DNA *Transcription* RNA *Translation* Protein

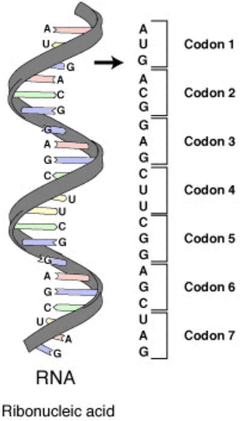
DNA: store genetic information (e.g. in *genome*);
regular double helix structure 
building blocks: 4 nucleotides A,C,G, and T
(Adenine, Cytosine, Guanine, Thymine)

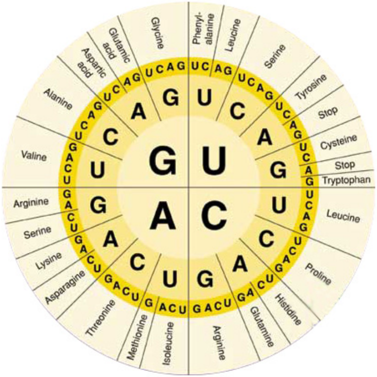
RNA: intermediate for protein synthesis (*messenger RNA*),
catalytic and regulatory function (*non-coding RNA*)
building blocks: 4 nucleotides A,C,G, and U
(U=Uracil) and some rare other nucleotides

Protein: catalytic and regulatory function (*'enzymes'*)
building blocks: 20 amino acids + 1 rare aa

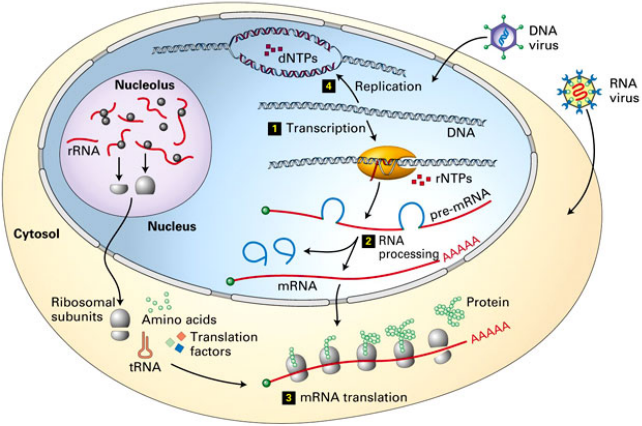
Genetic code

- Transcription: A,C,G,T $\mapsto$ A,C,G,U
- Translation: Tripletts from alphabet $\{$A,C,G,U$\}$ (= *codons*) redundantly code for amino acids

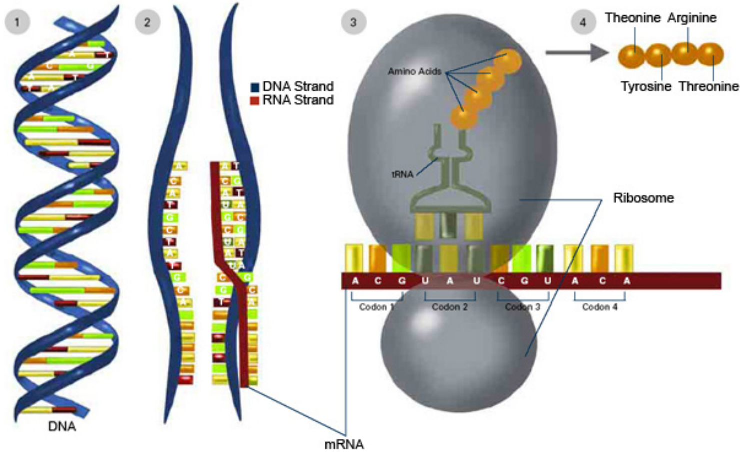




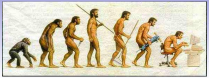
Information Flow (Cell Compartments)

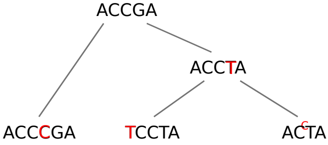


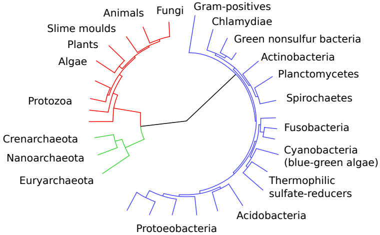
Protein Bio-Synthesis



Important for molecular mechanism: *complementarity* of nucleotides G-C, A-T, A-U

Evolution ()





- variaton (imperfect replication: point mutation, deletion, insertion, ...)
- selection
- homologous sequences

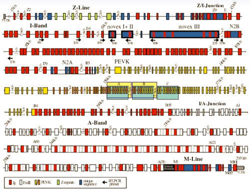
What can we study (computationally)?

What can we study (computationally)?

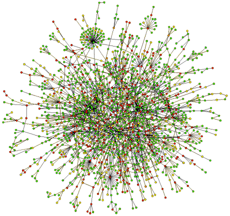
- Evolutionary relation between homologous molecules/fragments of molecules
- Structural relation between molecules
- Relation between sequence and structure
- Interaction between molecules
- Interaction networks, Regulatory networks, Metabolic networks
- Structure of genomes, Relation between genomes
- ...

Areas of Bioinformatics

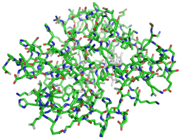
1. *Genomics:* Study of entire genomes. Huge amount of data, fast algorithms, limited to sequence.



2. *Systems Biology:* Study of complex interactions in biological systems. High level of representation.



3. *Structural Bioinformatics:* Study of the folding process of bio-molecules. Less structural data than sequence data available, step toward function, fills gap between genomics and systems biology.



Some Organic Chemistry

Biological macromolecules (and most organic compounds) are built from only few different types of atoms

- C — Carbon
- H — Hydrogen
- O — Oxygen
- N — Nitrogen
- P — Phosphor
- S — Sulfur

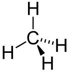
CHNO: 99% of cell mass

Organic Chemistry = Chemistry of Carbon

Special properties of Carbon

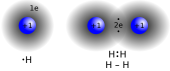
- binds up to 4 other atoms,

  e.g. Methane

  (tetrahedron conformation)
- small size
- strong *covalent* bonds

  covalent bond:

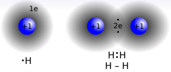
  

- chains and rings

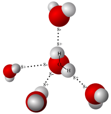
⇒ large, stable, complex molecules

Non-covalent bonds

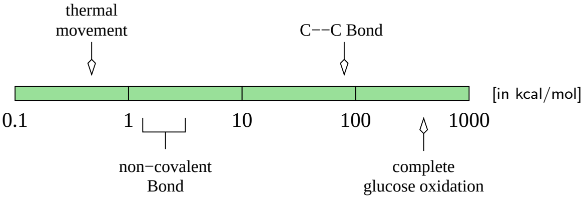
- Covalent



- Non-covalent
  - Van der Waals (sum of the attractive or repulsive forces between molecules, caused by correlations in the fluctuating polarizations of nearby particles)
  - hydrogen bonds (attractive interaction of a hydrogen atom with an electronegative atom)



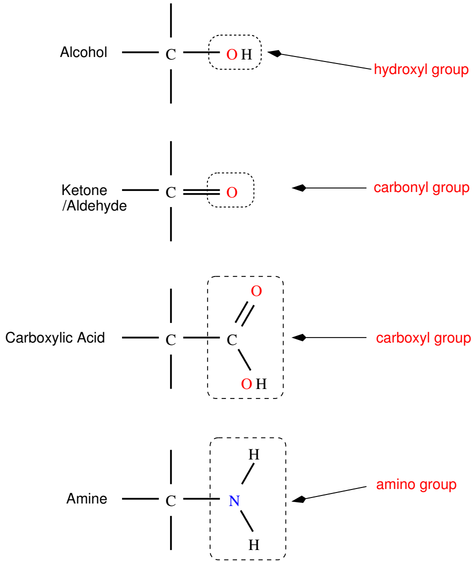
  - ionic bonds (electrostatic attraction between two oppositely charged ions, e.g. Na+ Cl)



Functional groups

organic molecules: carbon skeleton + functional groups

functional groups are involved in specific chemical reactions



Small organic molecules

Small: ≤ 30 atoms

4 families:

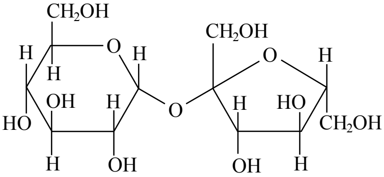
- sugars
  $\Rightarrow$ component of building blocks, main energy source
- fats / fatty acids
  $\Rightarrow$ cell membrane, energy source
- amino acids
  $\Rightarrow$ proteins
- nucleotides
  $\Rightarrow$ DNA + RNA, *energy currency*

Sugars

⇒ component of building blocks, main energy source

- general formula $(CH_2O)_n$,
  different lengths (e.g n=5, n=6)
- linear, cyclic

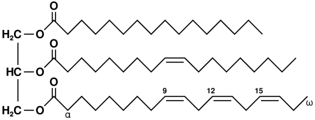
For example, saccharose (glucose+fructose):

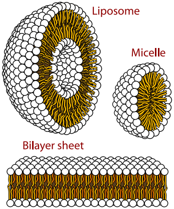


Fats

Fat = Triglyceride of fatty acids

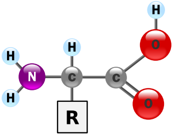
⇒ cell membrane (lipid bilayer), energy source





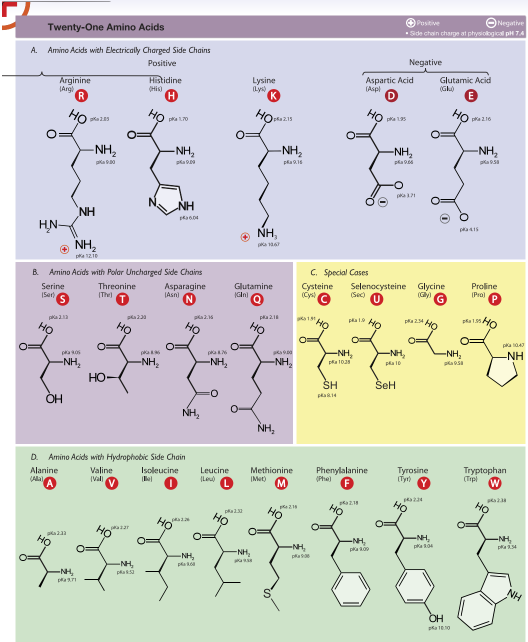
Amino Acids

- all aa same build

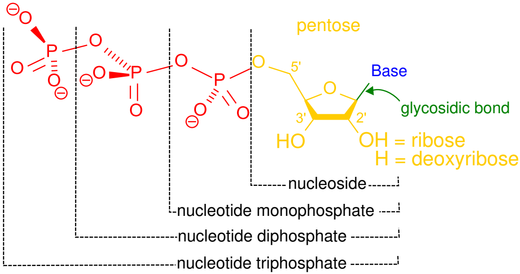


- aa differ in side chains R
  - size
  - charge: positiv/negativ (sauer/basisch)
  - hydrophobicity: hydrophobic/hydrophilic
- in naturally occuring proteins: 21 different amino acids

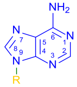
Amino Acids


Twenty-One Amino Acids
Positive
Negative
• Side chain charge at physiological pH 7.4
A. Amino Acids with Electrically Charged Side Chains
Positive
Negative
Arginine (Arg) R
Histidine (His) H
Lysine (Lys) K
Aspartic Acid (Asp) D
Glutamic Acid (Glu) E
pKa 2.03
pKa 9.00
pKa 12.10
pKa 1.70
pKa 9.09
pKa 6.04
pKa 2.15
pKa 9.16
pKa 10.67
pKa 1.95
pKa 9.66
pKa 3.71
pKa 2.16
pKa 9.58
pKa 4.15
B. Amino Acids with Polar Uncharged Side Chains
Serine (Ser) S
Threonine (Thr) T
Asparagine (Asn) N
Glutamine (Gln) Q
pKa 2.13
pKa 9.05
pKa 2.20
pKa 8.96
pKa 2.16
pKa 8.76
pKa 2.18
pKa 9.00
C. Special Cases
Cysteine (Cys) C
Selenocysteine (Sec) U
Glycine (Gly) G
Proline (Pro) P
pKa 1.91
pKa 10.28
pKa 8.14
pKa 1.9
pKa 10
pKa 2.34
pKa 9.58
pKa 1.95
pKa 10.47
D. Amino Acids with Hydrophobic Side Chain
Alanine (Ala) A
Valine (Val) V
Isoleucine (Ile) I
Leucine (Leu) L
Methionine (Met) M
Phenylalanine (Phe) F
Tyrosine (Tyr) Y
Tryptophan (Trp) W
pKa 2.33
pKa 9.71
pKa 2.27
pKa 9.52
pKa 2.26
pKa 9.60
pKa 2.32
pKa 9.58
pKa 2.16
pKa 9.08
pKa 2.18
pKa 9.09
pKa 2.24
pKa 9.04
pKa 10.10
pKa 2.38
pKa 9.34

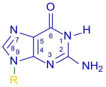
Nucleotides



Purines

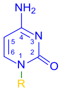


Adenine

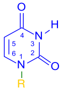


Guanine

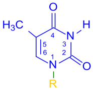
Pyrimidines



Cytosine



Uracil



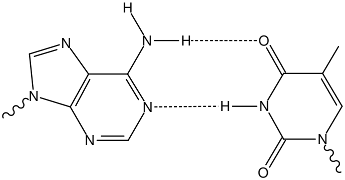
Thymine

Nucleotides work as energy currency of metabolism

NTP ⟶ P + NDP + E

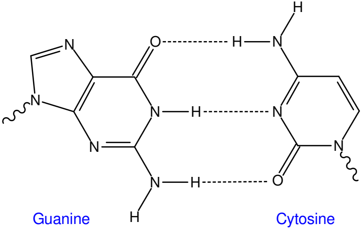
(split of nucleoside triphosphate into phosphate + nucleoside diphosphate releases energy)

Complementarity of Organic Bases



Adenine

Thymine



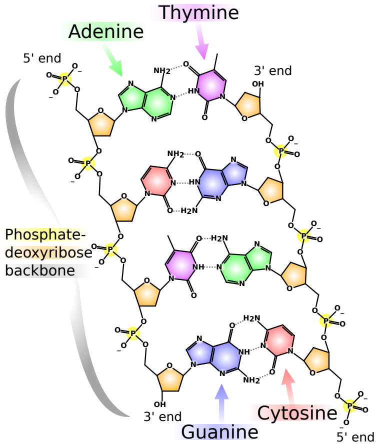
Guanine

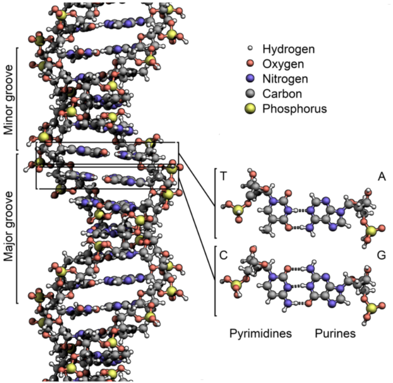
Cytosine

DNA structure

Primary structure: chain of nucleotides

Tertiary Structure: antiparallel double helix



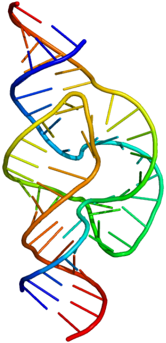


RNA primary structure similar, but

- *ribose not deoxyribose*, - *U not T*, - *single stranded*

RNA structure

tRNA

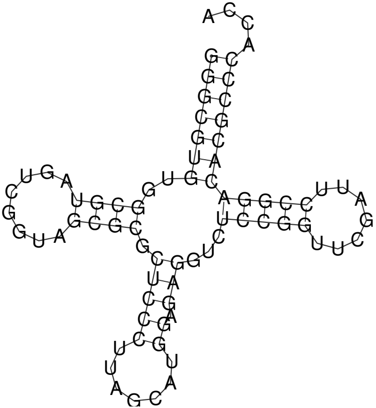


Hammerhead Ribozyme

mainly stabilized by contacts between complementary bases (H-bonds)
⇒ RNA secondary structure = set of base pairs

RNA secondary structure

- set of pairs of (complementary) bases that form H-bonds
- 2D representation (typical tRNA clover-leaf)



- linear representation

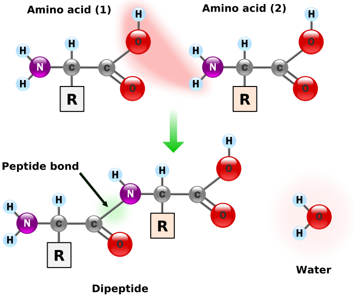
```
GGGCGUGUGGCGUAGUCGGUAGCGCGCUCCCUUAGCAUGGAGAGGUCUCCGGUUCGAUUCCGGACACGCCCACCA

(((((((..((((........)))).(((((.......)).)))...(((((.......))))))))))))....
```

- note: example is pseudoknot-free

Protein Primary Structure

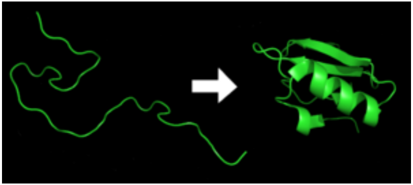
- Protein = chain of amino acids (AA)
- aa connected by peptide bonds



and so on . . .

Protein Structure Formation / Folding

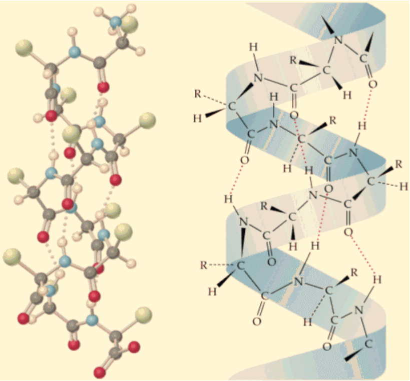
- minimization of free energy
- Forces between amino acid side chains
    - hydrophobic interaction
    - H-bonds
    - electro-static force
    - van-der-Waals force
    - disulfide bonds



Protein secondary structure: α-helix

Features:

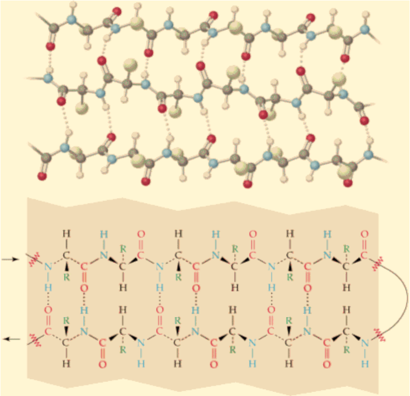
- 3.6 amino acids per turn
- hydrogen bond between residues n and $n+4$
- local motif
- approximately 40% of the structure



Protein secondary structure: β-sheets

Features:

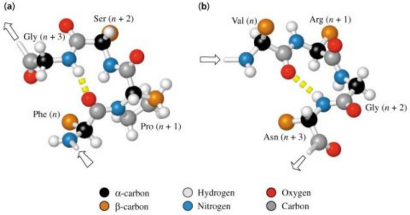
- 2 amino acids per turn
- hydrogen bond between residues of different strands
- involve long-range interactions
- approximately 20% of the structure



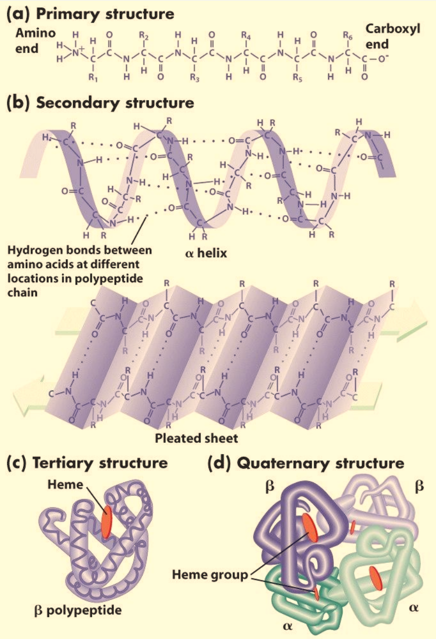
Protein secondary structure: Turns

Features:

- Up to 5 residue length
- hydrogen bonds depend of type
- local interactions
- approximately 5-10% of the structure



Protein structure hierarchy

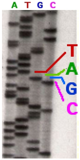


DNA sequencing

A very incomplete overview

= determining the order of nucleotides in DNA

- early 1970s: first DNA sequencing, but 'laborious'
- 1977: Sanger Chain-Termination 'rapid' sequencing



- whole genome sequencing, 2001 draft version of Human genome published
- high throughput sequencing (454, Illumina/Solexa, . . .)

- 2011 sequencing of a human genome costs about USD 10,000
- constant progress in technology (speed & accuracy)

⇒ RNA and protein sequences are usually inferred from DNA

Experimental Structure Determination

- How can we know the 3D structure of a protein/RNA?
    - X-ray cristallography
        - Requires crystalls of macromolecule.
          Often extremely difficult and time-intensive
        - X-rays send through crystall produce specific patterns
        - Angles and intensities allow to construct 3D-electron density
        - From this, one can determine atom positions, bonds, etc.
    - Nuclear magnetic resonance spectroscopy (NMR)
        - uses phenomenon of nuclear magnetic resonance
        - only relatively small molecules
        - does not require crystalls
        - measure distances between pairs of atoms within the molecule
        - structure has to be predicted using these constraints
- Experimentally resolved structures are available in the protein data base (PDB) in a machine-readable format.
- The number of resolved structures grows exponentially, but slower than the one of known sequences.

Topics of the Class

Sequence Alignment

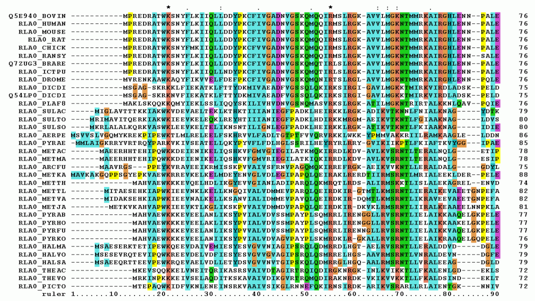
- pairwise alignment

```
Sequence A: ACGTGAACT
Sequence B: AGTGAGT
```

⇓align A and B

```
Sequence A: ACGTGAACT
Sequence B: A-GTGA-GT
```

- global and local alignment
- multiple alignment (NP-complete ⇒ heuristics)

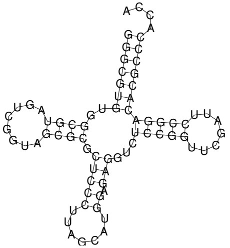


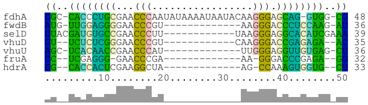
```
                                 *   .      :       .          .   *         : : :      .
Q5E940_BOVIN -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE   76
  RLA0_HUMAN -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE   76
  RLA0_MOUSE -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE   76
    RLA0_RAT -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE   76
  RLA0_CHICK -----------MPREDRATWKSNYFMKIIQLLDDYPKCFVVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE   76
  RLA0_RANSY -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--SALE   76
Q7ZUG3_BRARE -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQTIRLSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE   76
  RLA0_ICTPU -----------MPREDRATWKSNYFLKIIQLLNDYPKCFIVGADNVGSKQMQTIRLSLRGK-AIVLMGKNTMMRKAIRGHLENN--PALE   76
  RLA0_DROME -----------MVRENKAAWKAQYFIKVVELFDEFPKCFIVGADNVGSKQMQNIRTSLRGL-AVVLMGKNTMMRKAIRGHLENN--PQLE   76
  RLA0_DICDI -----------MSGAG-SKRKKLFIEKATKLFTTYDKMIVAEADFVGSSQLQKIRKSIRGI-GAVLMGKKTMIRKVIRDLADSK--PELD   75
Q54LP0_DICDI -----------MSGAG-SKRKNVFIEKATKLFTTYDKMIVAEADFVGSSQLQKIRKSIRGI-GAVLMGKKTMIRKVIRDLADSK--PELD   75
  RLA0_PLAF8 -----------MAKLSKQQKKQMYIEKLSSLIQQYSKILIVHVDNVGSNQMASVRKSLRGK-ATILMGKNTRIRTALKKNLQAV--PQIE   76
  RLA0_SULAC -----MIGLAVTTTKKIAKWKVDEVAELTEKLKTHKTIIIANIEGFPADKLHEIRKKLRGK-ADIKVTKNNLFNIALKNAG-----YDTK   79
  RLA0_SULTO ----MRIMAVITQERKIAKWKIEEVKELEQKLREYHTIIIANIEGFPADKLHDIRKKMRGM-AEIKVTKNTLFGIAAKNAG-----LDVS   80
  RLA0_SULSO ----MKRLALALKQRKVASWKLEEVKELTELIKNSNTILIGNLEGFPADKLHEIRKKLRGK-ATIKVTKNTLFKIAAKNAG-----IDIE   80
  RLA0_AERPE MSVVSLVGQMYKREKPIPEWKTLMLRELEELFSKHRVVLFADLTGTPTFVVQRVRKKLWKK-YPMMVAKKRIILRAMKAAGLE---LDDN   86
  RLA0_PYRAE -MMLAIGKRRYVRTRQYPARKVKIVSEATELLQKYPYVFLFDLHGLSSRILHEYRYRLRRY-GVIKIIKPTLFKIAFTKVYGG---IPAE   85
  RLA0_METAC ------MAEERHHTEHIPQWKKDEIENIKELIQSHKVFGMVGIEGILATKMQKIRRDLKDV-AVLKVSRNTLTERALNQLG-----ETIP   78
  RLA0_METMA ------MAEERHHTEHIPQWKKDEIENIKELIQSHKVFGMVRIEGILATKIQKIRRDLKDV-AVLKVSRNTLTERALNQLG-----ESIP   78
  RLA0_ARCFU ------MAAVRGS---PPEYKVRAVEEIKRMISSKPVVAIVSFRNVPAGQMQKIRREFRGK-AEIKVVKNTLLERALDALG-----GDYL   75
  RLA0_METKA MAVKAKGQPPSGYEPKVAEWKRREVKELKELMDEYENVGLVDLEGIPAPQLQEIRAKLRERDTIIRMSRNTLMRIALEEKLDER--PELE   88
  RLA0_METTH -------------MAHVAEWKKKEVQELHDLIKGYEVVGIANLADIPARQLQKMRQTLRDS-ALIRMSKKTLISLALEKAGREL--ENVD   74
  RLA0_METTL -------MITAESEHKIAPWKIEEVNKLKELLKNGQIVALVDMMEVPARQLQEIRDKIR-GTMTLKMSRNTLIERAIKEVAEETGNPEFA   82
  RLA0_METVA -------MIDAKSEHKIAPWKIEEVNALKELLKSANVIALIDMMEVPAVQLQEIRDKIR-DQMTLKMSRNTLIKRAVEEVAEETGNPEFA   82
  RLA0_METJA --------METKVKAHVAPWKIEEVKTLKGLIKSKPVVAIVDMMDVPAPQLQEIRDKIR-DKVKLRMSRNTLIIRALKEAAEELNNPKLA   81
  RLA0_PYRAB -------------MAHVAEWKKKEVEELANLIKSYPVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSRNTLIELAIKKAAQELGKPELE   77
  RLA0_PYRHO -------------MAHVAEWKKKEVEELAKLIKSYPVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSRNTLIELAIKKAAKELGKPELE   77
  RLA0_PYRFU -------------MAHVAEWKKKEVEELANLIKSYPVVALVDVSSMPAYPLSQMRRLIRENNGLLRVSRNTLIELAIKKVAQELGKPELE   77
  RLA0_PYRKO -------------MAHVAEWKKKEVEELANIIKSYPVIALVDVAGVPAYPLSKMRDKLR-GKALLRVSRNTLIELAIKRAAQELGQPELE   76
  RLA0_HALMA -----MSAESERKTETIPEWKQEEVDAIVEMIESYESVGVVNIAGIPSRQLQDMRRDLHGT-AELRVSRNTLLERALDDVD-----DGLE   79
  RLA0_HALVO -----MSESEVRQTEVIPQWKREEVDELVDFIESYESVGVVGVAGIPSRQLQSMRRELHGS-AAVRMSRNTLVNRALDEVN-----DGFE   79
  RLA0_HALSA -----MSAEEQRTTEEVPEWKRQEVAELVDLLETYDSVGVVNVTGIPSKQLQDMRRGLHGQ-AALRMSRNTLLVRALEEAG-----DGLD   79
  RLA0_THEAC -------------MKEVSQQKKELVNEITQRIKASRSVAIVDTAGIRTRQIQDIRGKNRGK-INLKVIKKTLLFKALENLGD----EKLS   72
  RLA0_THEVO -------------MRKINPKKKEIVSELAQDITKSKAVAIVDIKGVRTRQMQDIRAKNRDK-VKIKVVKKTLLFKALDSIND----EKLT   72
  RLA0_PICTO -------------MTEPAQWKIDFVKNLENEINSRKVAAIVSIKGLRNNEFQKIRNSIRDK-ARIKVSRARLLRLAIENTGK----NNIV   72
       ruler 1.......10........20........30........40........50........60........70........80........90
```

RNA Secondary Structure Prediction

- Predict minimal free energy structure for single sequence
- Predict minimal free energy structure for aligned sequences
- Predict common structure for alignment for **unaligned** sequences:

Simultaneous Alignment and Folding

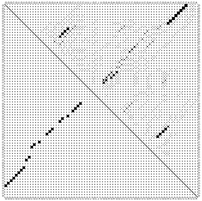


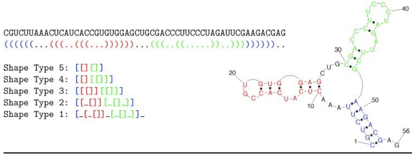


```
      ((..((((((((...(((.................))).))))))))..))
fdhA  CGC-CACCCUGCGAACCCAAUAUAAAAUAAUACAAGGGAGCAG-GUGG-CG 48
fwdB  AUG-UUGGAGGGGAACCCGU-------------AAGGGACCCUCCAAG-AU 36
selD  UUACGAUGUGCCGAACCCUU------------UAAGGGAGGCACAUCGAAA 39
vhuD  GU--UCUCUCGGGAACCCGU------------CAAGGGACCGAGAGA--AC 35
vhuU  AGC-UCACAACCGAACCCAU-------------UUGGGAGGUUGUGAG-CU 36
fruA  CC--UCGAGGG-GAACCCGA-------------AA-GGGACCCGAGA--GG 32
hdrA  GG--CACCACUCGAAGGCUA-------------AG-CCAAAGUGGUG--CU 33
      .........10........20........30........40........50
```

Studying the Structure Ensemble of an RNA

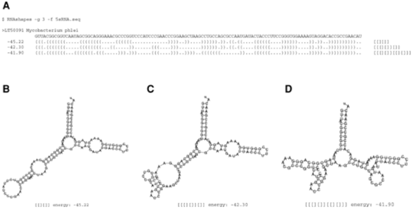
- Prediction of the structure ensemble
  - ⇒ probabilities of structures
  - ⇒ probabilities of structure elements and features
- Suboptimal Structures
- Shape Abstraction of RNA Structure





```
CGUCUUAAACUCAUCACCGUGUGGAGCUGCGACCCUUCCCUAGAUUCGAAGACGAG
(((((...(((..(((...))))))...(((..((.....))..))))))))..

Shape Type 5: [[][]]
Shape Type 4: [[][[]]]
Shape Type 3: [[[]][[]]]
Shape Type 2: [[_[]][_[]_]]
Shape Type 1: [_[_[]]_[_[]_]]_
```



A

```
$ RNAshapes -g 3 -f 5sRNA.seq

>LT50091 Mycobacterium phlei
           GUUACGGCGGUCAAUAGCGGCAGGGAAACGCCCGGUCCCAUCCCGAACCCGGAAGCUAAGCCUGCCAGCGCCAAUGAUACUACCCUUCCGGGUGGAAAAGUAGGACACCGCCGAACAU
  -45.22   (((.((((((((.....((((((((.....((((((..............)))))..))....)))))).)).((.((....((((((....))))))....)).))..)))))))))..    [[][]]
  -42.30   (((.((((((((.....((((((((...((.....))......((....))........)))))).)).((.((....((((((....))))))....)).))..)))))))))..    [[[][]][]]
  -41.90   (((.((((((((...((((((((.(((.....))).))).....((....))..)))))..((((((........))...((((((....))))))....)))))..)))))))))..    [[[][]][[][]]]
```

B

C

D

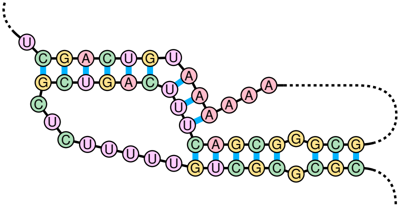
[[][]] energy: -45.22

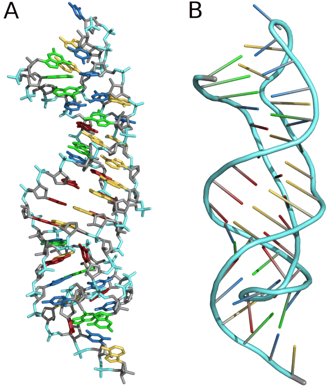
[[[][]][]] energy: -42.30

[[[][]][[][]]] energy: -41.90

RNA Pseudoknot Prediction

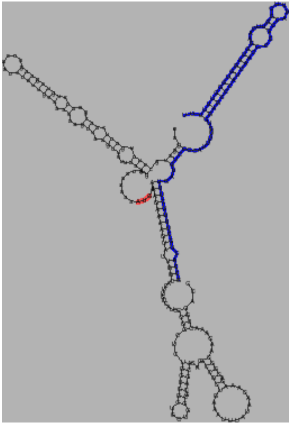
- Usually: for RNA structure analysis, assume no pseudoknots
- Pseudoknot (PK) prediciton is NP-complete
- Efficient PK prediction from restricted classes of PKs

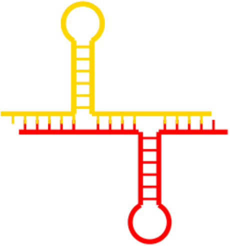


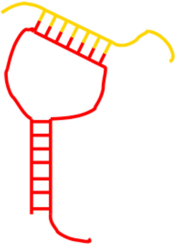


RNA-RNA Interaction

- Prediction of interaction complex of two RNAs
- Similar to Pseudoknot-prediction, the unrestricted problem is NP-complete
- Efficient variants exist for restricted types of interaction



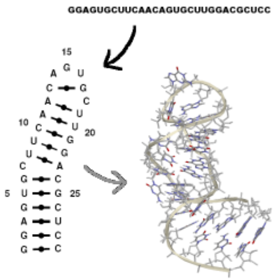




RNA 3D Structure Modeling

- De-novo prediction of 3D structure from sequence

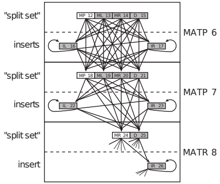
MC-Fold / MC-Sym:

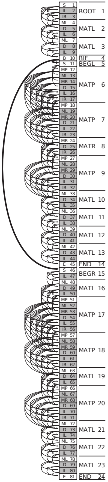


- MC-Fold predicts secondary structure including non-canonical base pairs
- MC-Sym builds tertiary from secondary structure

Stochastic Context-Free Grammars

- SCFGs are a generalization of HMMs, which can model secondary structure
- Consensus Models for describing RNA families.
- Tool Infernal scans database for family members

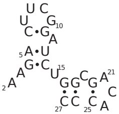




input multiple alignment:

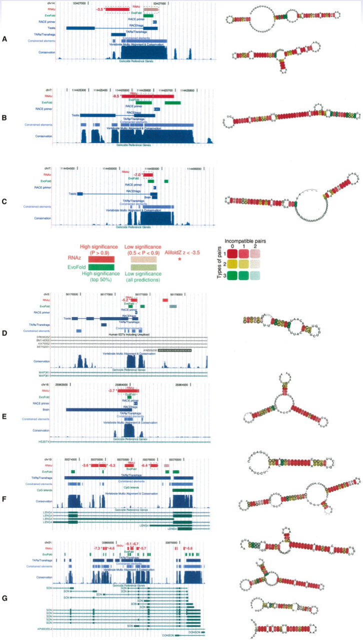
```
[structure]  . : : < < < _ _ _ _ > - > > : < < - < . _ _ _ . > > > .
      human  . A A G A C U U C G G A U C U G G C G . A C A . C C C .
      mouse  a U A C A C U U C G G A U G - C A C C . A A A . G U G a
        orc  . A G G U C U U C - G C A C G G G C A g C C A c U U C .
             1       5         10        15        20        25   28
```

example structure:



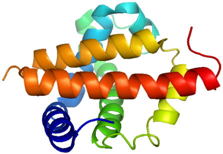
De-novo Prediction of Structural RNA

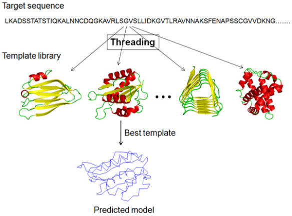
- scan whole genome alignments for potential structural RNA
- structural stability
- conservation of structure
- Fast methods RNAz, EvoFold



Protein Structure Prediction

- De-novo Protein Structure Prediction
- Homology-based prediction: Protein Threading
- Protein-Protein Interaction





Target sequence

LKADSSTATSTIQKALNNCDQGKAVRLSGVSLLIDKGVTLRAVNNAKSFENAPSSCGVVDKNG.......

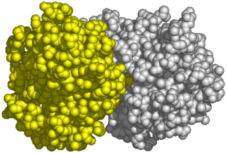
Threading

Template library

...

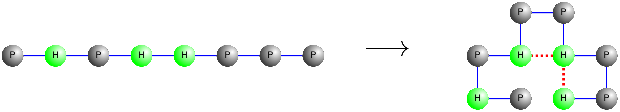
Best template

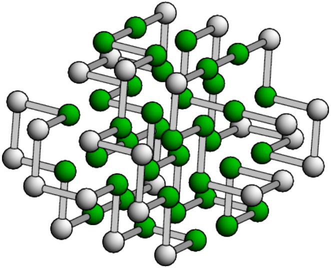
Predicted model



3D Lattice Protein Models

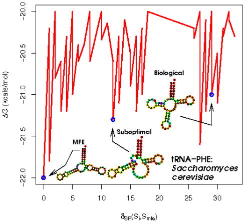
- protein structure prediction is NP-complete even in simple protein models
- optimal ab-initio prediction in HP-lattice protein models (3D cubic and fcc)

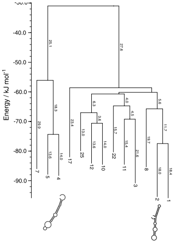


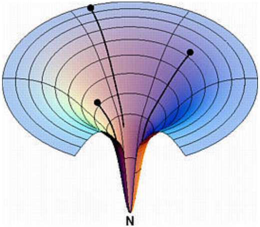


Beyond Energy Minimization:
Kinetiks of Protein and RNA folding

- Predicting Protein Folding-Pathways (Motion Planning)
- Modeling of Folding as Markov Process, Energy Landscapes
- Simulated and Exact Folding Kinetics







vs.