HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Electrical and Communications Engineering
Laboratory of Computational Engineering
S-114.500 Basics for Biosystems of the Cell

DNA, RNA and Protein Structure Prediction

Aino Salminen 63343U
Jenni Hulkkonen 61047A

Contents

Abstract

In this study we explore some DNA, RNA and protein structure prediction software, which is found in the Internet [1].

The single-stranded DNA and RNA fold into specific three-dimensional conformations, which are determined by the sequence of nucleotides. The prediction of RNA folding is important in understanding many biological processes, including translation regulation in messenger RNA, replication of single-stranded RNA viruses, and the function of structural RNAs and RNA/protein complexes. The prediction of DNA folding is important for example in PCR, where DNA is in a single-stranded form.

Proteins are generally self-folding. The three-dimensional structures of proteins are therefore also extremely diverse, ranging from completely fibrous to globular. The prediction of protein folding is important because the structure of a protein is related to its function. Experimental structure determination, or structure prediction, aids the elucidation of protein function; conversely, synthetic protein sequences might be designed so that the protein performs a desired function.

The study of protein structure therefore produces very valuable practical benefits for medicine, agriculture and industry. The understanding of enzyme function allows the design of drugs which inhibit specific enzyme targets for therapeutic purposes.

The field of research dealing with the prediction of structure from sequence is generally known as **bioinformatics** [2].

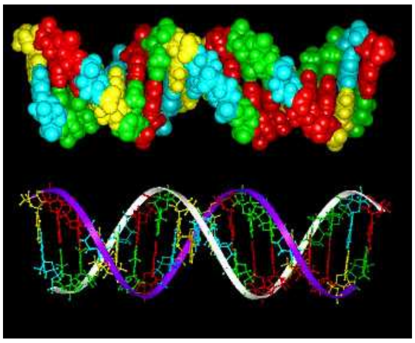
Introduction

DNA

Two-dimensional structure

DNA molecules consist of two complementary chains of nucleotides. Nucleotides are composed of a sugar, phosphate groups and a base. The base may be adenine, cytosine, guanine or thymine. Hydrogen bonds between the base pairs hold the chains together.

Single-stranded DNA can be found for example in PCR or in telomeres, where the 3’ DNA is always longer than the 5’ end. Single-stranded DNA may pair with itself. Programs that predict the folding of single-stranded DNA are introduced later in this presentation.



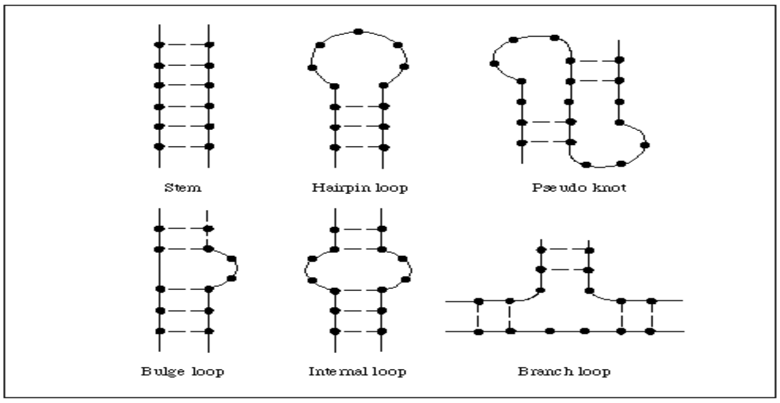
Three-dimensional structure

The two chains of DNA form a double helix. The base pairs are inside the helix and the sugar-phosphate backbones are outside it.

In eycaryotes, the DNA is packaged into chromosomes. DNA is wrapped around histones. DNA and core histones form nucleosomes, which are packed into a compact chromatin fiber.

RNA

RNA is very similar in structure to DNA. It contains the base urasil instead of thymine. RNA is single-stranded. However, it often contains short stretches of nucleotides that can form base-pairs with the same molecule. Programs that predict this folding are introduced later.



There are several types of RNA. MRNA is copied from genes. Then intron sequences are removed by RNA splicing. The 5’ end of mRNA is capped and a poly-A tail is added to the 3’ tail.

TRNA recognizes and binds the codon and the amino acid in protein translation. Four segments of tRNA are double-helical. One of these regions forms the anticodon that pairs with the complementary codon in an mRNA molecule. The other is the site where the amino acid is attached to the tRNA.

Proteins

Proteins are huge polypeptides which consist of amino acid residues. The shape of a protein is specified by its amino acid sequence.

Primary structure

The primary structure of a protein consists of amino acids linked by peptide bonds to form polypeptide chains. The code for the primary structure is in DNA.
There are 20 different amino acids.

Secondary structure

The amino acids in a polypeptide chain form hydrogen bonds between the N-H and C=O groups. The chain twists around on itself and forms a three-dimensional structure. Most common folding patterns are the α-helix and the β-sheet.

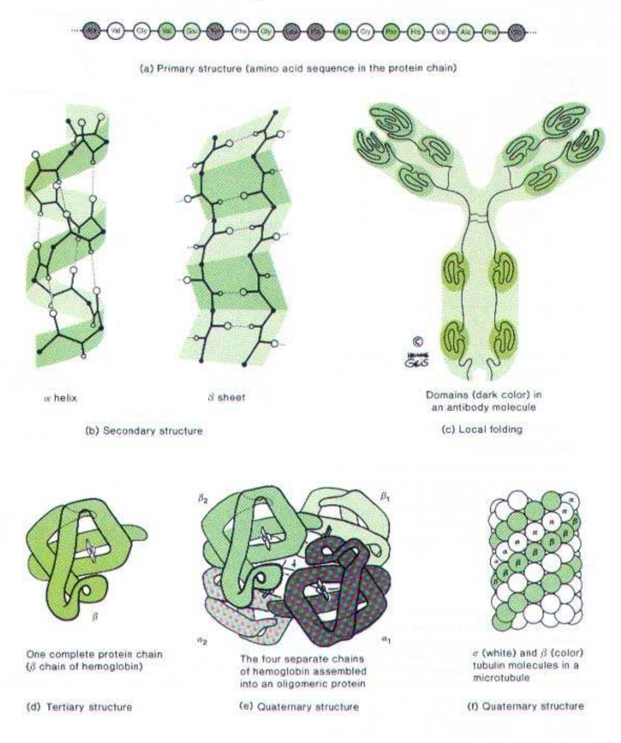
Tertiary structure

The final structure of a protein is the one in which the free energy is minimized. Proteins called chaperons help on protein folding. Hydrophobic amino acid chains are buried on the inside of a protein and hydrophilic amino acid chains gather on the outside. Sulphur bridges stabilize the structure.

Protein functionality is based on its tertiary structure

The prediction of the 3-D structure of proteins is far in the future, because proteins are generally self-folding. One way to try to predict protein tertiary structure is that 3-D structure of some proteins have already been explored and by comparing the sequences we can try to predict the structure.

The four levels of protein structure organization


Val
Gly
Val
Glu
Phe
Gly
Asp
Gly
His
Val
Ala
Phe
(a) Primary structure (amino acid sequence in the protein chain)
α helix
β sheet
(b) Secondary structure
©
Geis
Domains (dark color) in an antibody molecule
(c) Local folding
β
One complete protein chain (β chain of hemoglobin)
(d) Tertiary structure
β₂
β₁
α₂
α₁
The four separate chains of hemoglobin assembled into an oligomeric protein
(e) Quaternary structure
α
β
α (white) and β (color) tubulin molecules in a microtubule
(f) Quaternary structure

Free energy

Some prediction programs use free energy in their algorithms.
The change of free energy is ΔG = ΔG°' + RT*ln*([B]/[A]). (ΔG°' is the standard free energy). A reaction can occur spontaneously only if ΔG is negative.

Methods

DNA and RNA secondary structure prediction

MFOLD [3]

Mfold is a program used for RNA and DNA secondary structure prediction. It uses the nearest neighbor thermodynamic rules [4] in prediction. The program runs in a Unix environment but can be accessed with a web browser.

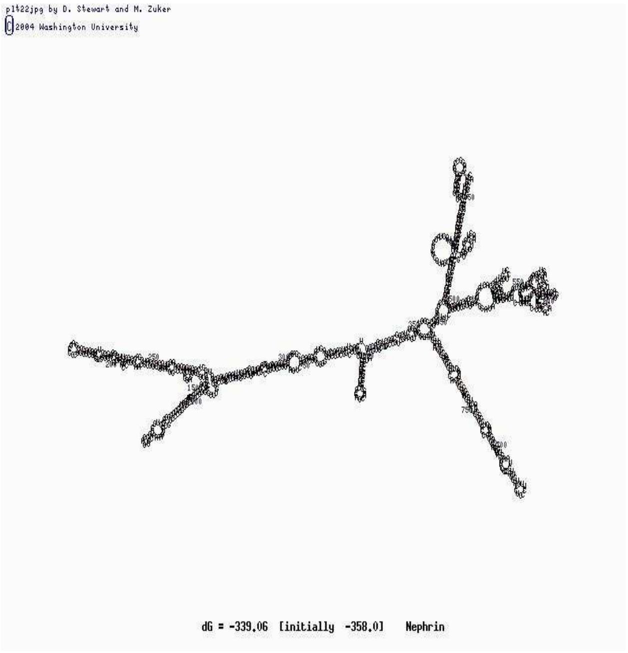
Applications of mfold

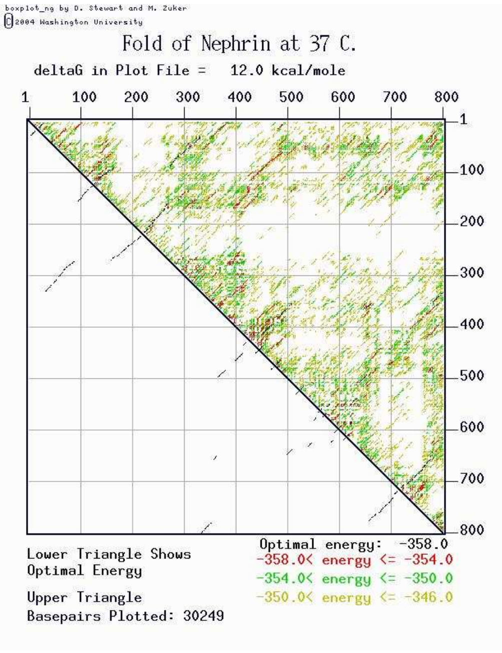
RNA Folding
DNA Folding
Nucleic Acid Quikfold
Zipfold server
Tm server
2-state hybridization server
Free Energy Determination

Using mfold

The folding temperature is fixed at 37° (body temperature). The temperature can be varied in the older version 2.3. You may force certain bases to be double or single stranded, and it is also possible to prohibit certain bases from pairing with others. You may choose whether the RNA sequence is linear or circular. The maximum loop size and the maximum distance between paired bases can be entered. Your job can be prosessed while you wait (up to 800 base pairs) or can de submitted for batch processing (up to 6000 base pairs).

An Example


plt22jpg by D. Stewart and M. Zuker
©2004 Washington University
dG = -339.06 [initially -358.0] Nephrin



VIENNA RNA PACKAGE [5]

The Vienna RNA Package predicts and compares RNA secondary structure. The prediction is based on three algorithms: the minimum free energy algorithm [6], the partition function algorithm [7] and the suboptimal folding algorithm [8]. The comparison is based on measures of distance using string alignment or tree-editing.

The program runs in a Unix or Windows environment.

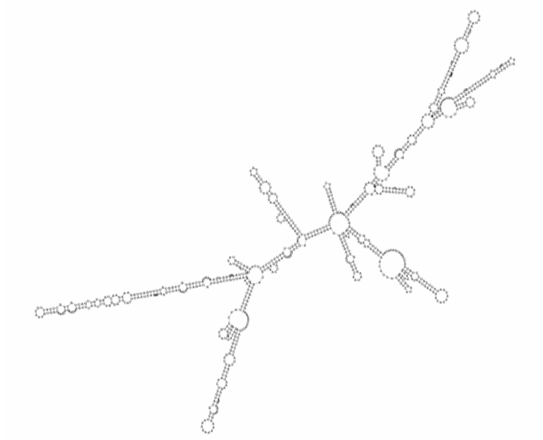
Applications of Vienna RNA Package

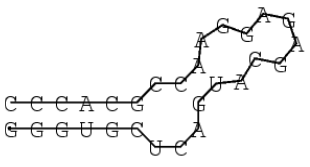
RNAfold
RNAfold predicts minimum energy secondary structures and pair probabilities.
RNAeval
RNAeval evaluates energy of RNA secondary structures.
RNAheat
RNAheat calculates the specific heat (melting curve) of an RNA sequence.
RNAinverse
RNAinverse provides inverse fold (design) sequences with predefined structure.
RNAdistance
RNAdistance compares secondary structures.
RNApdist
RNApdist compares base pair probabilities.
RNAsubopt
RNAsubopt completes suboptimal folding.

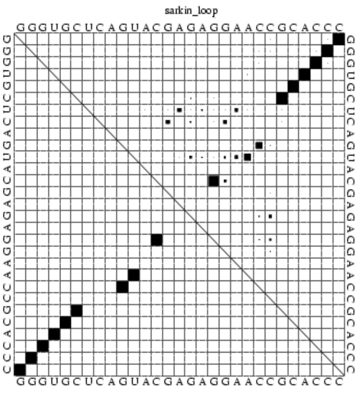
Using Vienna package RNAfold

The program reads RNA sequences from stdin and returns their minimum energy and partition function in stdout. It produces PostScript files with plots of the resulting secondary structure graph and a "dot plot" of the base pairing matrix.

An Example



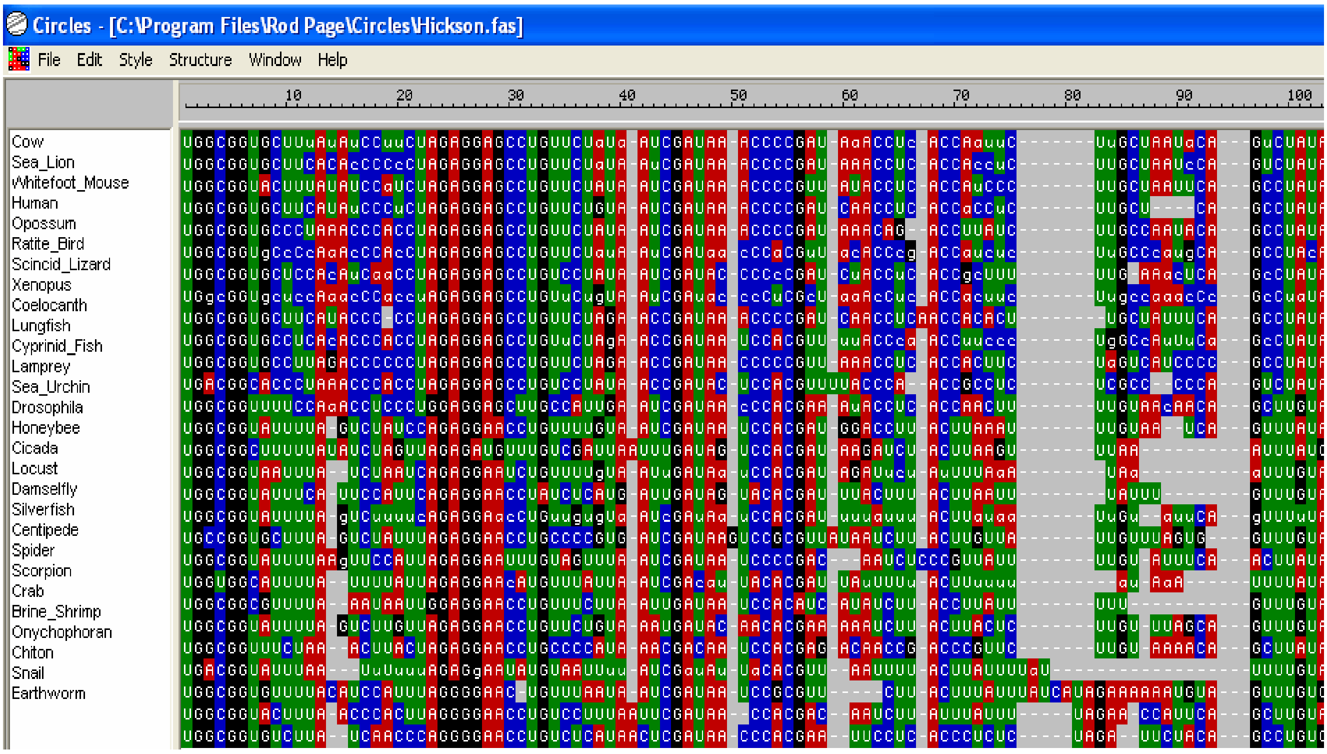




<u>**CIRCLES**</u> **[9]**

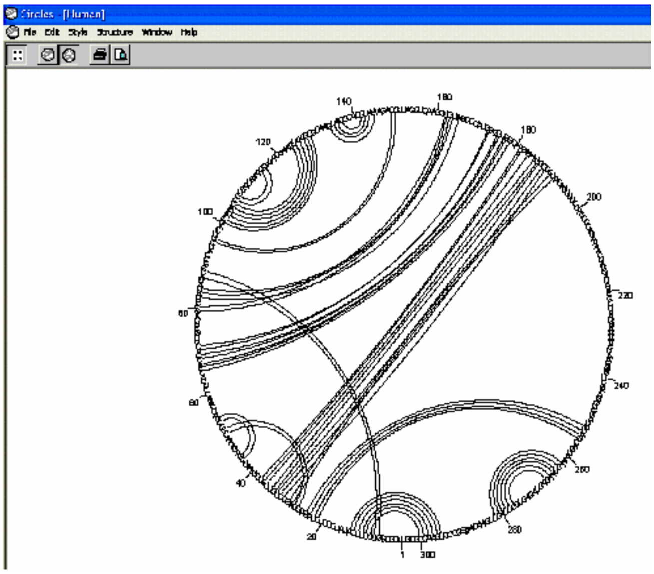
Circles is an experimental Windows 95/98/NT program for inferring RNA secondary structure using the comparative method. It provides a user-friendly interface to maximum weight matching programs. You can give your alignment in FASTA, Clustal, or NEXUS format. Circles computes a maximum weight matching. Maximum Weighted Matching method (MWM) [10] is capable of detecting pseudoknots and other tertiary base-pairing interactions in a computationally efficient manner. Circles exports one or more RNA secondary structures in standard formats.

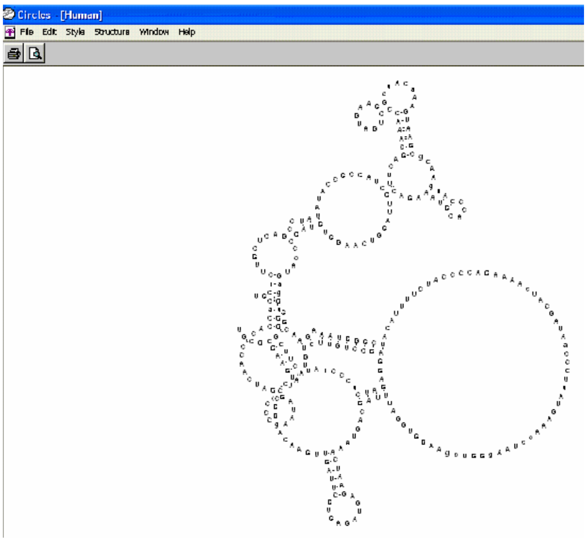
An Example



```
Cow
Sea_Lion
Whitefoot_Mouse
Human
Opossum
Ratite_Bird
Scincid_Lizard
Xenopus
Coelocanth
Lungfish
Cyprinid_Fish
Lamprey
Sea_Urchin
Drosophila
Honeybee
Cicada
Locust
Damselfly
Silverfish
Centipede
Spider
Scorpion
Crab
Brine_Shrimp
Onychophoran
Chiton
Snail
Earthworm
```

```
UGGCGGUGCUUuAuAuCCuuCUAGAGGAGCCUGUUCUaUa-AUCGAUAA-ACCCCGAU-AaACCUc-ACCAauuC-------UuGCUAAUaCA---GuCUAUA
UGGCGGUGCUUCACAcCCCcCUAGAGGAGCCUGUUCUaUA-AuCGAUAA-ACCCCGAU-AAACCUC-ACCAccuC-------UUGCUAAUcCA---GUCUAUA
UGGCGGUACUUUAUAUCCaUCUAGAGGAGCCUGUUCUAUA-AUCGAUAA-ACCCCGUU-AUACCUC-ACCAuCCC-------UUGCUAAUUCA---GCCUAUA
UGGCGGUGCUUCAUAuCCCuCUAGAGGAGCCUGUUCUGUA-AUCGAUAA-ACCCCGAU-CAACCUC-ACCaCCuC-------UUGCU----CA---GCCUAUA
UGGCGGUGCCCUAAACCCACCUAGAGGAGCCUGUUCUAUA-AUCGAUAA-ACCCCGAU-AAACAG--ACCUUAUC-------UUGCCAAUACA---GCCUAUA
UGGCGGUgCcCcAaAcCCAcCUAGAGGAGCCUGUUCUauA-AuCGAUaa-cCCaCGuU-acACCcg-ACCaucuc-------UuGCCcaugCA---GCCUAcA
UGGCGGUGCUCCAcAuCaaCCUAGAGGAGCCUGUCCUAUA-AUCGAUAC-CCCcCGAU-CuACCuC-ACCgcUUU-------UUG-AAacUCA---GcCUAUA
UGgcGGUgcuccAaacCCaccuAGAGGAGCCUGUuCugUA-AuCGAuac-ccCuCGcU-aaAcCuc-ACCacuuc-------UugccaaacCc---GcCuaUA
UGGCGGUGCUUCAUACCC-CCUAGAGGAGCCUGUUCUAGA-ACCGAUAA-ACCCCGAU-CAACCUCAACCACACU--------UGCUAUUUCA---GCCUAUA
UGGCGGUGCCUCAcACCCACCUAGAGGAGCCUGUuCUAgA-ACCGAUAA-UCCACGUU-uuACCca-ACCuuccc-------UgGCcAuUuCa---GcCUAUA
UGGCGGUGCCUUAGACCCCCCUAGAGGAGCCUGUUCUAGA-ACCGAUAA-CCCCCGUU-AAACCUC-ACCACUUC-------UaGUCAUCCCC---GCCUAUA
UGACGGCACCCUAAACCCACCUAGAGGAGCCUGUCCUAUA-ACCGAUAC-UCCACGUUUUACCCA--ACCGCCUC-------UCGCC--CCCA---GUCUAUA
UGGCGGUUUUCCAaACCUCCCUGGAGGAGCUUGCCAUUGA-AUCGAUAA-cCCACGAA-AuACCUC-ACCAACUU-------UUGUAAcAACA---GCUUGUA
UGGCGGUAUUUUA-GUCUAUCCAGAGGAACCUGUUUUGUA-AUCGAUAA-UCCACGAU-GGACCUU-ACUUAAAU-------UUGUAA--UCA---GUUUAUA
UGGCGGCUUUUUAUAUCUAGUUAGAGAUGUUUGUCGAUUAAUUUGAUAG-UCCACGAU-AAGAUCU-ACUUAAGU-------UUAA----------AUUUAUC
UGGCGGUAAUUUA--UCUAAUCAGAGGAAUCUGUUUUgUA-AUuGAUAa-uCCACGAU-AGAUucu-AuUUUAaA--------UAa----------aUUUGUA
UGGCGGUAUUUCA-UUCCAUUCAGAGGAACCUAUCUCAUG-AUUGAUAG-UACACGAU-UUACUUU-ACUUAAUU--------UAUUU--------GUUUGUA
UGGCGGUAUUUUA-gUCuuuucAGAGGAacCUGuugugUa-AUcGAuAa-uCCACGAU-uuuauuu-ACUUauaa-------UuGu--auuCA---gUUUuUA
UGCCGGUGCUUUA-GUCUAUUUAGAGGAACCUGCCCCGUG-AUCGAUAAGUCCGCGUUAUAAUCUU-ACUUGUUA-------UUGUUUAGUG----GUUUGUA
UGGCGGUAUUUUAAgUUCCAUUAGAGGAAUUUGUAGUUUA-AUCGAUAA-UCCCCGAC---AAUCUCCCGUUAUU-------UUGU-AUUUCA---ACUUAUA
UGGUGGCAUUUUA--UUUUAUUAGAGGAAcAUGUUUAUUA-AUCGAcau-UACACGAU-UAuUUUu-ACUUuuuu---------au-AaA------UUUUAUA
UGGCGGCGUUUUA--AAUAAUUGGAGGAACCUGUUUCUUA-AUUGAUAA-UCCACAUC-AUAUCUU-ACCUUAUU-------UUU-----------GUUUGUA
UGGCGGUAUUUUA-GUCUUGUUAGAGGAACCUGUUCUGUA-AAUGAUAC-AACACGAA-AAAUCUU-ACUUACUC-------UUGU-UUAGCA---GUUUGUA
UGGCGGUUUCUAA--ACUUACUAGAGGAACCUGCCCCAUA-AACGACAA-UCCACGAG-ACAACCG-ACCCGUUC-------UUGU-AAAACA---GCUUAUA
UGACGGUAUUUAA---UuUuuuAGAGgAAUAUGUAAUUuu-AUCGauAu-UaCACGUU--AAUUUU-ACUUAUUUUaU------------------UUUUGUA
UGGCGGUGUUUUACAUCCAUUUAGGGGAAC-UGUUUAAUA-AUCGAUAA-UCCGCGUU-----CUU-ACUUUAUUUAUCAUAGAAAAAAUGUA---GUUUGUC
UGGCGGUACUUUA-ACCCACUUAGGGGAACCUGUCCUUUAAUUCGAUAA--CCACGAC--AAUCUU-AUUUAUUU-----UAGAA-CCAUUCA---GCUUGUA
UGGCGGUGUCUUA--UCAACCCAGGGGAACCUGUCUCAUAACUCGAUAA-CCCACGAA--UUCCUC-ACCCUCUC-----UAGA--UUCUACA---GCCUGUC
```





Protein structure prediction

<u>kPROT</u> [11]

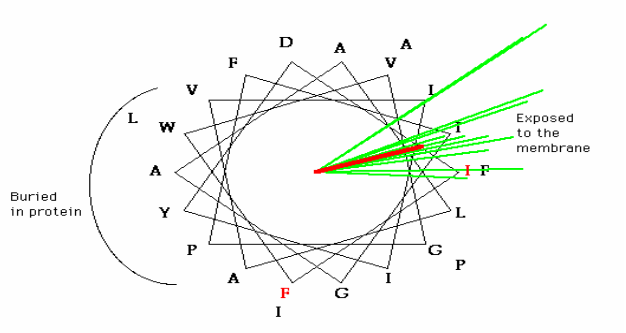
kPROT is a www service for structure prediction of membrane proteins and specifically designed for transmembrane proteins with multiple alpha-helical spans. It provides a prediction of the angular orientation of the segment.

The program submits protein sequence(s) of previously identified transmembrane segment.
It predicts the helical orientation of the segment as expected to be when embedded in a helical bundle within the membrane i.e. lipid-exposed vs. protein-buried faces of alpha-helices.
The Helical orientation predictions are done with the "knowledge-based Propensities for Residue Orientation in Transmembrane segments" (kPROT) energy-like scale.
The kPROT scale was derived from statistics of transmembrane proteins in the SWISS-PROT database.

Predictions and analyses

- Predicts Transmembrane helix orientation
- Plots the kPROT profile of a sequence and it's fourier transform
- Calculates and compares the kPROT profiles of two proteins
- Calculates the kPROT profile of a protein and searches a library for proteins with similar profile

An Example



JNET [12]

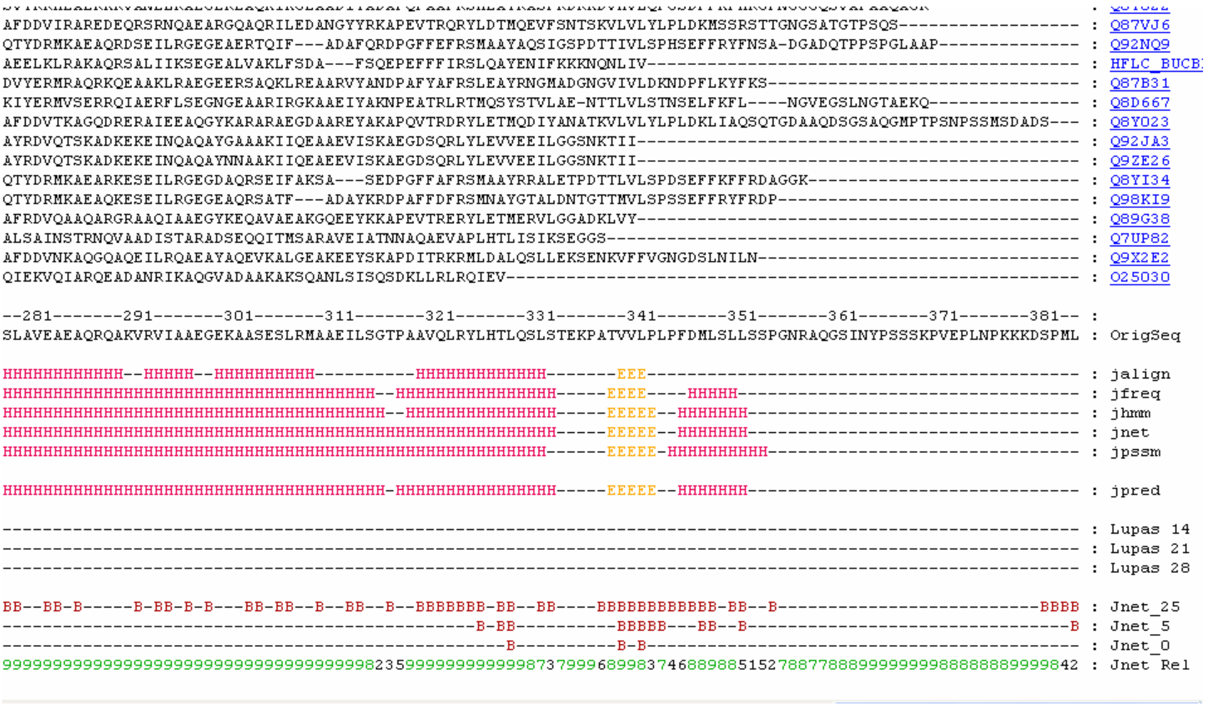
Jnet is a program for UNIX workstations and a www service. It uses Neural Network to do Protein Secondary Structure Prediction Method. The program works by applying multiple sequence alignments, alongside PSIBLAST and HMM profiles. A reliability index indicates which residues are predicted with a high confidence.

Using Jnet

Syntax: jnet -mode <sequence file> [hmm profile] [<psiblast pssm> and <psiblast freq>]

- sequence file (compulsory)
  - Can be generated from a ClustalW MSF file
- hmm profile (optional)
  - can be generated from ClustalW MSF file
- psiblast pssm, psiblast freq (optional)
  - can be generated from a PSIBLAST report
  - If supplied both must be present

An Example



```
AFDDVIRAREDEQRSRNQAEARGQAQRILEDANGYYRKAPEVTRQRYLDTMQEVFSNTSKVLVLYLPLDKMSSRSTTGNGSATGTPSQS------------------ : Q87VJ6
QTYDRMKAEAQRDSEILRGEGEAERTQIF---ADAFQRDPGFFEFRSMAAYAQSIGSPDTTIVLSPHSEFFRYFNSA-DGADQTPPSPGLAAP-------------- : Q92NQ9
AEELKLRAKAQRSALIIKSEGEALVAKLFSDA---FSQEPEFFFIRSLQAYENIFKKKNQNLIV------------------------------------------- : HFLC_BUCB
DVYERMRAQRKQEAAKLRAEGEERSAQKLREAARVYANDPAFYAFRSLEAYRNGMADGNGVIVLDKNDPFLKYFKS------------------------------- : Q87B31
KIYERMVSERRQIAERFLSEGNGEAARIRGKAAEIYAKNPEATRLRTMQSYSTVLAE-NTTLVLSTNSELFKFL----NGVEGSLNGTAEKQ--------------- : Q8D667
AFDDVTKAGQDRERAIEEAQGYKARARAEGDAAREYAKAPQVTRDRYLETMQDIYANATKVLVLYLPLDKLIAQSQTGDAAQDSGSAQGMPTPSNPSSMSDADS--- : Q8Y023
AYRDVQTSKADKEKEINQAQAYGAAAKIIQEAAEVISKAEGDSQRLYLEVVEEILGGSNKTII-------------------------------------------- : Q92JA3
AYRDVQTSKADKEKEINQAQAYNNAAKIIQEAEEVISKAEGDSQRLYLEVVEEILGGSNKTII-------------------------------------------- : Q9ZE26
QTYDRMKAEARKESEILRGEGDAQRSEIFAKSA---SEDPGFFAFRSMAAYRRALETPDTTLVLSPDSEFFKFFRDAGGK--------------------------- : Q8YI34
QTYDRMKAEAQKESEILRGEGEAQRSATF---ADAYKRDPAFFDFRSMNAYGTALDNTGTTMVLSPSSEFFRYFRDP------------------------------ : Q98KI9
AFRDVQAAQARGRAAQIAAEGYKEQAVAEAKGQEEYKKAPEVTRERYLETMERVLGGADKLVY-------------------------------------------- : Q89G38
ALSAINSTRNQVAADISTARADSEQQITMSARAVEIATNNAQAEVAPLHTLISIKSEGGS----------------------------------------------- : Q7UP82
AFDDVNKAQGQAQEILRQAEAYAQEVKALGEAKEEYSKAPDITRKRMLDALQSLLEKSENKVFFVGNGDSLNILN-------------------------------- : Q9X2E2
QIEKVQIARQEADANRIKAQGVADAAKAKSQANLSISQSDKLLRLRQIEV--------------------------------------------------------- : O25030

--281-------291-------301-------311-------321-------331-------341-------351-------361-------371-------381-- :
SLAVEAEAQRQAKVRVIAAEGEKAASESLRMAAEILSGTPAAVQLRYLHTLQSLSTEKPATVVLPLPFDMLSLLSSPGNRAQGSINYPSSSKPVEPLNPKKKDSPML : OrigSeq

HHHHHHHHHHHH--HHHHH--HHHHHHHHHHH---------HHHHHHHHHHHHH-------EEE------------------------------------------- : jalign
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH--HHHHHHHHHHHHHHHH-----EEEE----HHHHH---------------------------------- : jfreq
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH--HHHHHHHHHHHHHHH-----EEEEE--HHHHHHH--------------------------------- : jhmm
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-----EEEEE--HHHHHHH--------------------------------- : jnet
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH------EEEEE-HHHHHHHHHH------------------------------- : jpssm

HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-HHHHHHHHHHHHHHHH-----EEEEE--HHHHHHH--------------------------------- : jpred

----------------------------------------------------------------------------------------------------------- : Lupas 14
----------------------------------------------------------------------------------------------------------- : Lupas 21
----------------------------------------------------------------------------------------------------------- : Lupas 28

BB--BB-B-----B-BB-B-B---BB-BB--B--BB--B--BBBBBBB-BB--BB----BBBBBBBBBBBB-BB--B--------------------------BBBB : Jnet_25
-----------------------------------------------B-BB----------BBBBB---BB--B--------------------------------B : Jnet_5
--------------------------------------------------B----------B-B------------------------------------------- : Jnet_0
99999999999999999999999999999999999823599999999999987379996899837468898851527887788899999999888888889999842 : Jnet Rel
```

SUPERPOSITIONAL STRUCTURE ASSIGNMENT (SSA) [13]

The program SSA automates the assignment of the secondary structure of a peptide from its atomic coordinates. The automation is based on the superposition of the coordinates with sequences of ideal secondary structure. The program SSA is most often used with the program Sequery.

Using SSA

The program takes a file produced by Sequery. It outputs the resulting assignments and summary information.
Syntax: *runssa SequeryOutputFile SSAOutputFile*
SSA will generate an output file containing a line for each *Sequery* match, listing the PDB code, chain identifier, numerical residue range, 4-residue peptide matching sequence (in upper case) along with 4 residues of sequence on either side (in lower case), the sequence pattern matched, and the assigned secondary structure.

An Example

```
# Sequery Filename : Example.sequery

#    C
#    h
#    a
#    i   Residue                              User User
#PDB n      Range      Sequence                  Type  RMSD Match?
#-----------------------------------------------------------------
1pxt B   40 to   43 -> vvivAANRsaig matching A.NR  EXTE
1tiv _   21 to   24 -> qpktACNRchck matching A6NR  EXTE
1pox A  324 to  327 -> iavlADAQktla matching .D5Q  IRRE
2bpa 1  230 to  233 -> tsydADNRpllv matching A.NR  IRRE
1cpc A   12 to   15 -> avaaADSQgrfl matching .D5Q  HELI
1hpm _  131 to  134 -> mkeiAEAYlgkt matching .E.Y  HELI
1aor A  602 to  605 -> elgiAEFY     matching .E.Y  ERROR
1aor A  407 to  410 -> syrlAESYghpe matching .E.Y  HELI
1mxa _  213 to  216 -> pilpAEWLtsat matching .EWL  TUR1
1fnc _  223 to  226 -> mkekAPDNfrld matching AP..  TUR1


####################################STATISTICS#############################
#
# Total number of sequery lines processed is:          10
# Total number of Errors (excluded from statistics):    1  (10.00 %)
# # of Regular Helix Types: 3 (33.33 %)
# # of Irreg Helix Types:   0    (0.00 %)
# Total # of Helix Types:            3        (33.33 %)
#
# # of Extended Types:               2        (22.22 %)
#
# # of Type1 Turns:      2     (22.22 %)
# # of Type2 Turns:      0     (0.00 %)
# # of Type1' Turns:     0     (0.00 %)
# # of Type2' Turns:     0     (0.00 %)
# # of Type8 Turns:      0     (0.00 %)
# # of Irreg Turns:     0      (0.00 %)
# Total # Of Turns:                  2        (22.22 %)
#
# # of Irreg Types *:                2        (22.22 %)
#
#* # of Irreg Types excludes Irreg Helices and Irreg Turns
```

References

[1] http://www.ks.uiuc.edu/Development/biosoftdb/biosoft.cgi?&category=17
[2] http://www.cryst.bbk.ac.uk/PPS2/course/section1/overview.html
[3] http://www.bioinfo.rpi.edu/applications/mfold/
[4] Zuker, M., Mathews, D.H., and Turner, D.H. In *RNA Biochemistry and Biotechnology*. J. Barciszewski and B.F.C. Clark eds. Nato ASI Series. Kluwer Academic Publishing, 1999
[5] http://www.tbi.univie.ac.at/~ivo/RNA/
[6] **M. Zuker** & P. Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**, 133-148 (1981)
[7] J.S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structures, Biopolymers 29: 11051119(1990)
[8] S. Wuchty, W. Fontana, I. L. Hofacker and P. Schuster "Complete Suboptimal Folding of RNA and the Stability of Secondary Structures", Biopolymers, 49, 145-165 (1999)
[9] http://taxonomy.zoology.gla.ac.uk/rod/circles/
[10] http://www.tbi.univie.ac.at/Leere/Slides/814236/ws02/xtina0.pdf
[11] http://bioinformatics.weizmann.ac.il/kPROT/
[12] http://www.compbio.dundee.ac.uk/~www-jpred/jnet/
[13] http://www.bch.msu.edu/labs/kuhn/web/software.html

Berg, J. M., Tymoczko, J.L., Stryer:
Biochemistry.
Fifth Edition. W. H. Freeman and Company. NY. 2002

Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter:
Molecular Biology of the Cell.
Fourth Edition. Garland Science, Taylor & Francis Group. 2002.