TERTIARY MOTIF INTERACTIONS ON RNA STRUCTURE

Bioinformatics Senior Project

Wasay Hussain

Spring 2009

Overview of RNA

- The central Dogma of Molecular biology is

$$\text{DNA} \longrightarrow \text{RNA} \longrightarrow \text{Proteins}$$

The RNA (Ribonucleic Acid) is a carrier between the DNA (Deoxyribonucleic Acid) to proteins. It plays a crucial role to control the transfer pathway from DNA genetic code to producing functional proteins.

Overview Cont..

- The three dimensional folds create sites that process as chemical catalyst.

- The backbone helps to stabilize the global structure of RNA. This is critical for guiding the folding process.

Structure

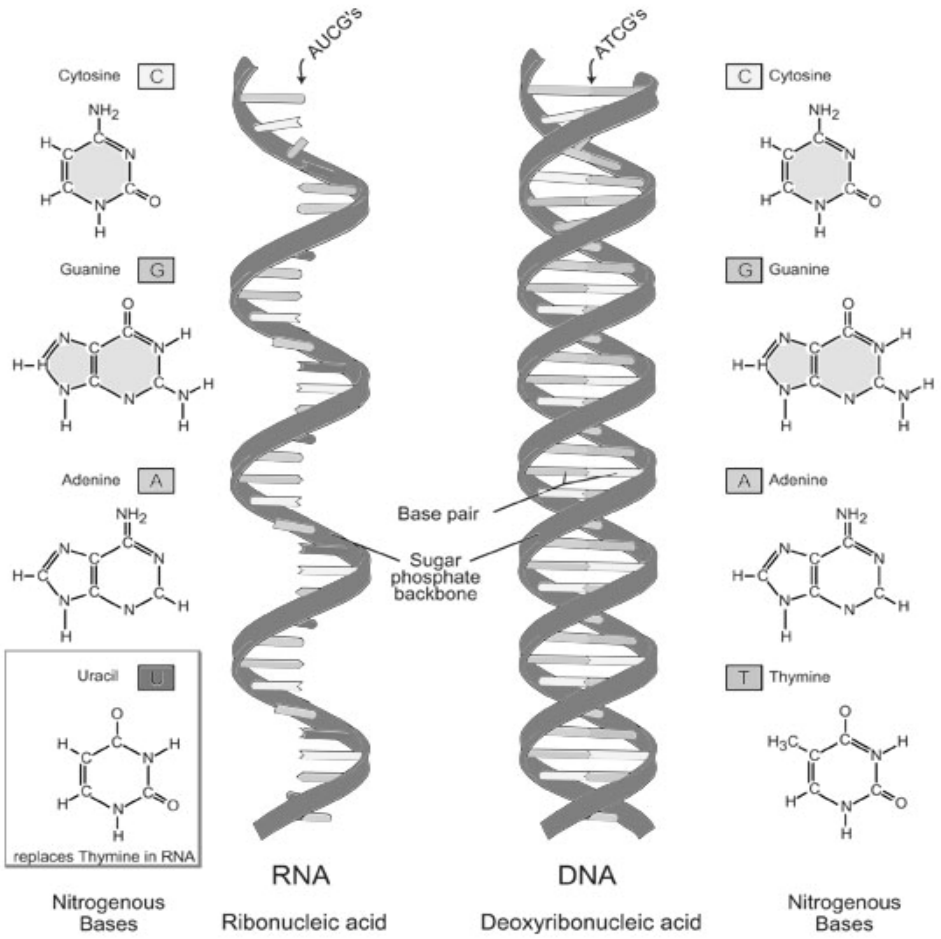


Image adapted from: National Human Genome Research Institute. Talking Glossary of Genetic Terms. Available at: www.genome.gov/Pages/Hyperion//DIR/VIP/Glossary/Illustration/rna.shtml.

Types of RNA Structures

- Primary : Nucleotide sequence of RNA
- Secondary: Watson-Crick base pair 2-dimensional model
- Tertiary: Interactions between distinct secondary structures

Types of RNA's

- tRNA
- mRNA
- rRNA
- snRNA

This is a list of a few different types of RNA that play crucial roles in motif patterns.

What are motifs?

- RNA motifs are short fragments of RNA that have a repeated pattern, which play a crucial role in determining some function.
- It is super imposable with other elements of an RNA structure.
- Sharing of molecular interactions with other elements of an RNA structure

Types of Motifs

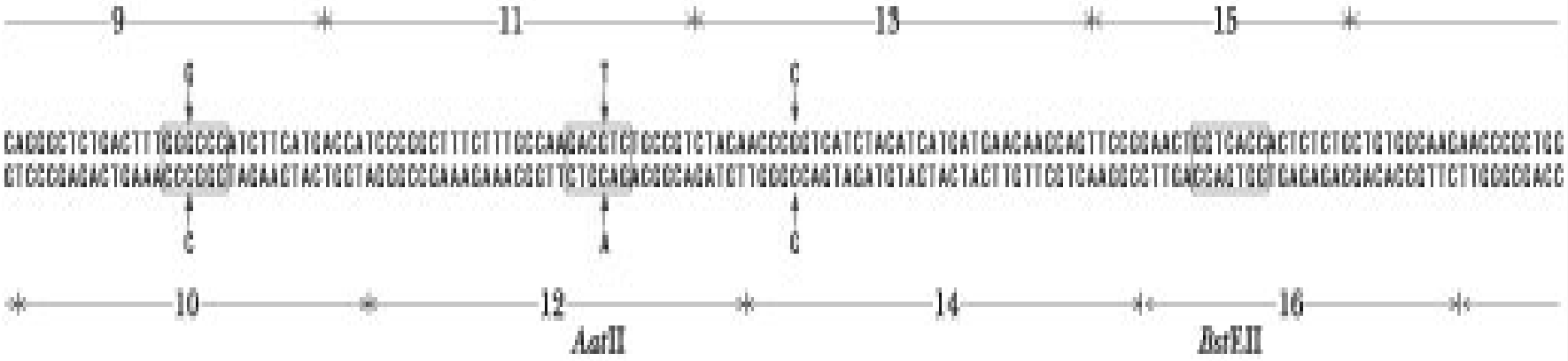
- 3 types of motifs structures:
  - Sequence
  - Secondary
  - Tertiary

Sequence Motif

- An example of a sequence motif is the Shine-Delgarno sequence of bacterial mRNA.

- Sequence motifs always have some functional implications. This motif will have a small arrangement of secondary structure elements to form a stable domain.

Sequence Motif



This is a type of sequence motif where the sequences are a pattern form and become a stable domain with a secondary structure. [13]

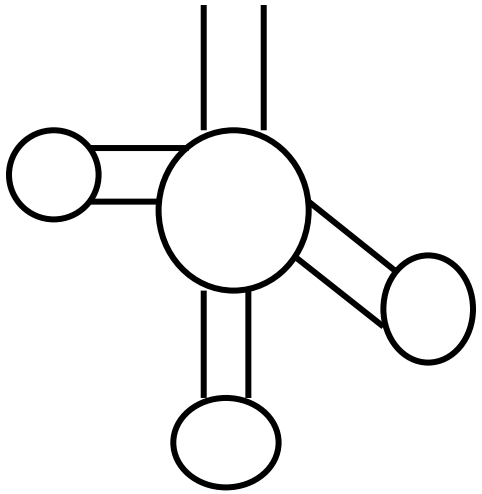
Secondary Motifs

- These motifs can usually be calculated from the sequence, through the aid of comparative sequence analysis. Other motifs at the secondary structures are:

- Hairpin (terminal) loops

- Internal loops

- Junction loops

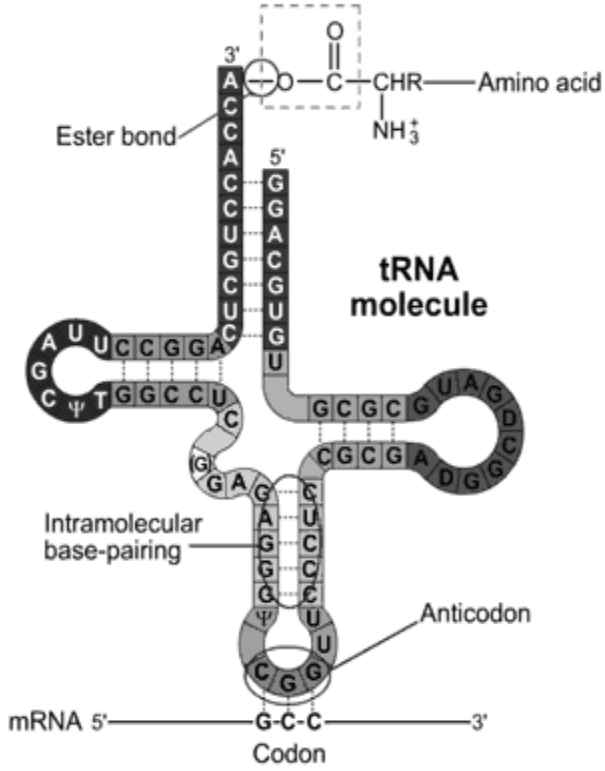


Tertiary Motifs

- These interactions enable the highly anionic double-stranded helices to tightly pack together and create a globular structure.

- An example of tertiary motifs

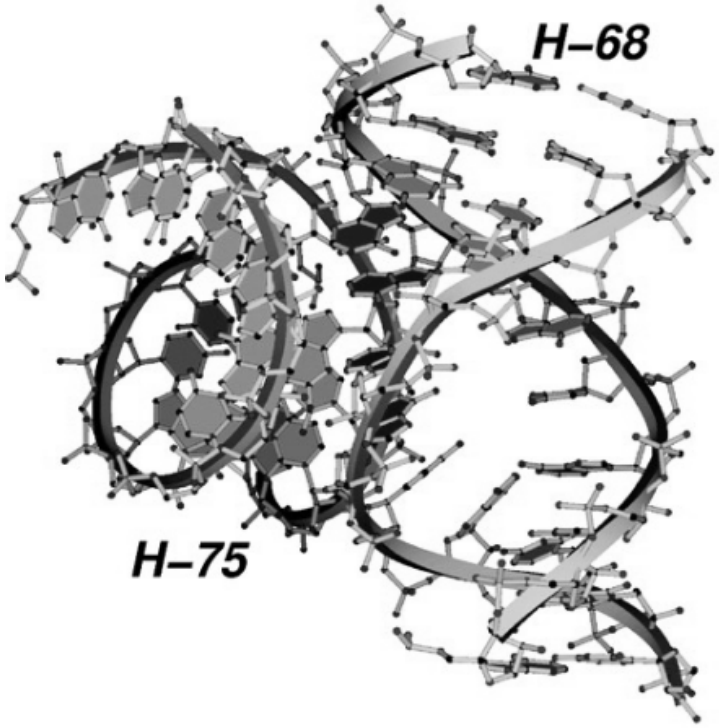
  folding is the t-RNA.



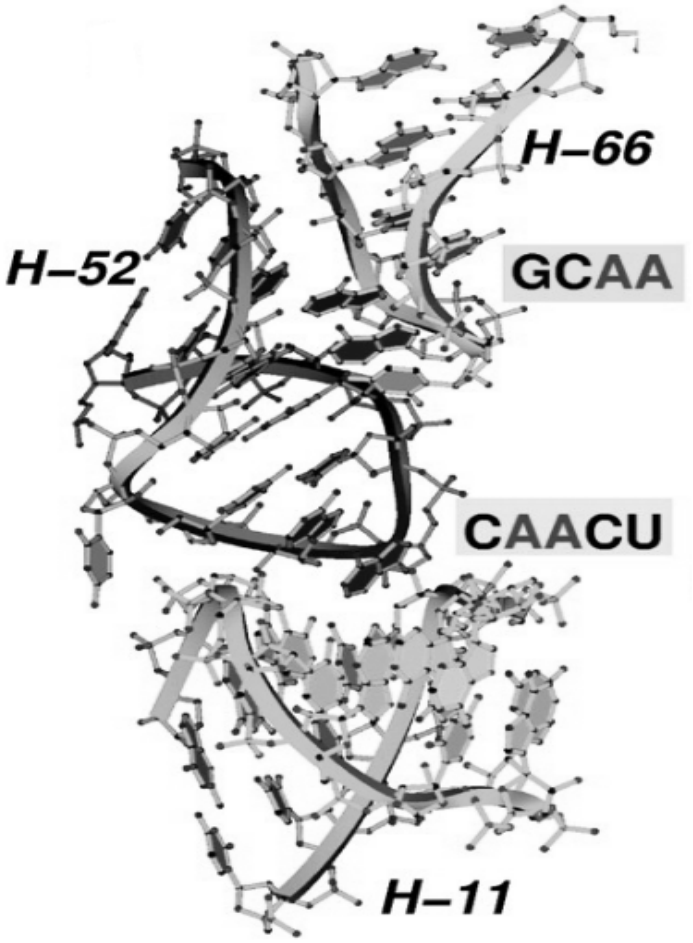
Tertiary Interaction Motifs

- Another example of tertiary interaction motif is the A-minor motifs.
  - This motif interaction involves insertion of minor grooves of Adenosine (A) into the major grooves of Cytosine (C) or Guanine (G) bases.
  - This interaction forms Hydrogen bonds.
  - This interaction is very significant on a large ribosomal subunit due to the 186 Adenines.

A-minor Interactions



- This is a picture of an A-minor interaction that helps in stabilizing the interaction between helix 68 of domain IV and helix 75 of domain V.
- The Adenosine is stacked, which allows it to pack the minor groove on helix 75.



- This A-minor is mediating a loop-loop interaction on the RNA.
- As seen in the figure, the helix 52 and helix 66 have interactions between their stem loops.

Roles of Tertiary Structures

- Tertiary interactions play an important role in establishing a global fold in the molecule.
- It is also composed of conserved building blocks known as “motifs”.
- Formation of motifs are sequence-dependent.

- To identify tertiary structural motifs:

→ Sequence: is needed and necessary in order to find patterns within the structure

→ Structural information: important in identifying recurrent backbone conformations

Sometimes this information is unknown for many RNA structures; we may look at different approaches using various parameters in obtaining motifs.

□ Proposed methods:

Djelloul and Denise use a “graph-based” approach, where bases are represented with it’s vertices and edges are interactions. This approach reduces the field into “isomorphic occurrences of the pattern” known as sub graph isomorphism. This computation is used along with finding “similar” occurrences of the pattern.

- Since RNA motifs are known to be recurrent and ordered stacked arrays of isosteric non- Watson Crick base pairs, that scatter the 2-Dimensional helices and fold into an identical 3-D structure. Two non-Canonical base pairs are isosteric if they belong to the same geometric family and substitute without destroying the 3-D motif.

Motif Identification Algorithms

- COMPADRES: Uses Phosphorus and C4 atoms to represent a nucleotide. Uses fragments to compare RMSD values.

- ARTS (Alignment of RNA tertiary structures):

  Compares and aligns pairs of 3D nucleic acid structures; it identifies common substructures.

COMPADRES

- COMPADRES stands for **Comp**arative **A**lgorithm to **D**iscover **R**ecurring **E**lements of **S**tructure.

- COMPADRES is used to identify recurrent RNA backbone conformations.

- This algorithm compares all short RNA worms in the structure database against each other.

- This algorithm identifies complex motifs from the RNA database of existing structures.

- These motifs are defined mathematically by the RNA worms that characterize their backbones.

COMPADRES Output

- Stage 1: structurally identical stretches of nucleotides are automatically grouped in a pair-wise fashion.
- Stage 2: the worm representation that describes each of the candidate motifs is automatically compared to the worm representations that describe the library of known motifs.

	PDB	Sequence	Chain ID	Residues
π-turns				
Type 1	1JJ2	GUACG	0	1873–1877
	1JJ2	AUAAC	0	408–412
	1JJ2	CGAUA	0	451–455*
	1JJ2	CGCAA	0	1854–1858*
	1DDY	CCUCA	A	21–25*
Type 2	1HR2	GUAUG	A	176–180
	1JJ2	GGUCG	0	2847–2851*
Ω-turns	1JJ2	CGAAG	0	245–249
	1JJ2	GGUUC	0	1416–1420
	1JJ2	GGAAU	0	1744–1748
	1NTB	CUUCU	A	15–19*
	1N32	UGACG	A	751–755*
α-loops	1JJ2	GUCCCCAA	0	1100–1107
	1N32	CCGGCCAA	A	503–510
C2FA	1JJ2	GAGACC	0	446–451
	1JJ2	ACCAGA	0	2019–2024
Hook turns	1JJ2	CAAGU	0	658–662
	1N32	CGAGC	0	545–549

The examples marked with an asterisk (*) were identified with the subsequent PRIMOS search (see Methods). For the π-turn, the two types were distinguished because of differing sugar pucker conformation at nucleotide 2. However, both types shared the conformational characteristics of the canonical π-turn. Here, we also include two previously unreported hook turns that were also discovered with COMPADRES.

COMPADRES Conclusion

- This algorithm can be used to identify new motifs.
- It is essentially proficient in finding new examples of "known motifs".
- Overall this is a powerful tool in motif interaction studies.

ARTS

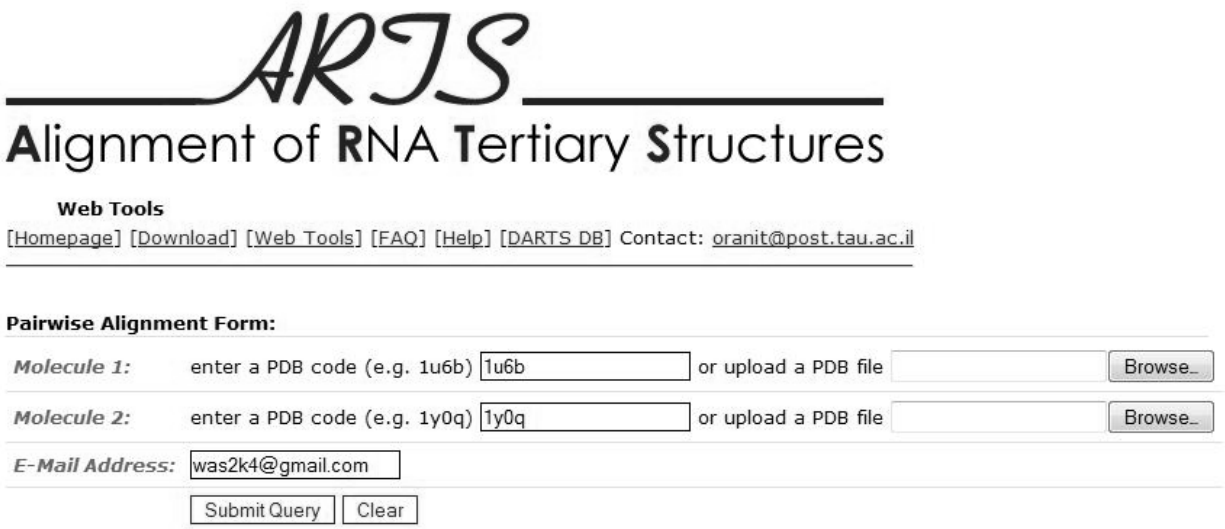
- Uses pair wise alignment of the nucleic acid structures inputted by the user.
- Input: PDB code of nucleic acid structure, or PDB file of the structure.
- Output: A list of top ranking alignments found.

ARTS contd..

- ARTS also allows one to use a DATABASE search which enables one to obtain more advanced alignments and so forth. With this search option one would have to use real parameters.

- ARTS can also be used to discover new motifs.

Alignment tool Input



- The output contains two tables:
  - The upper table shows the compared structures and the number of nucleotides along with base pairs in each structure.
  - The bottom table shows the top 10 ranking alignments sorted in order (descending) by its score.

Output

ARTS

Alignment of RNA Tertiary Structures

Web Tools

[Homepage] [Download] [Web Tools] [FAQ] [Help] [DARTS DB] Contact: oranit@post.tau.ac.il

(This table shows the compared Seq and number of nucleotides)

Structure	Number of Nucleotides	Number of Base Pairs
1u6b.pdb	219	79
1y0q.pdb	230	82

Alignment No.	Score	BP Core Size	Core Size	RMSD	P-Value	PDB Alignment
1	194.00	38	118	1.39	0.0000	alignment1
2	194.00	38	118	1.39	0.0000	alignment2
3	194.00	37	120	1.46	0.0000	alignment3
4	193.00	37	119	1.45	0.0000	alignment4
5	191.00	37	117	1.37	0.0000	alignment5
6	191.00	37	117	1.42	0.0000	alignment6
7	191.00	36	119	1.42	0.0000	alignment7
8	70.00	10	50	1.58	0.1852	alignment8
9	58.00	8	42	1.71	0.3844	alignment9
10	58.00	7	44	1.87	0.3844	alignment10

(Alignment score in decending order)

BP Core Size

- Base pair core size gives a table with matched base pairs of the alignments.
- The table has columns for each structure; each row shows the match between two base-pairs.
- Order : Chain ID-Base-Residue number of two nucleotide on corresponding BP

ARTS

Alignment of **R**NA **T**ertiary **S**tructures

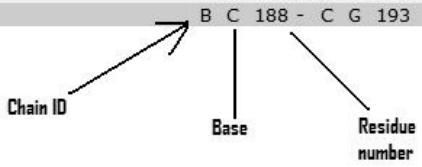
Web Tools

[Homepage] [Download] [Web Tools] [FAQ] [Help] [DARTS DB] Contact: oranit@post.tau.ac.il

Number of Conserved Base Pairs = 7

Alignment syntax: *ChainID Base Residue No. - ChainID Base Residue No.*

1u6b.pdb	1y0q.pdb
B U 176 - B A 131	A G 190 - A C 124
B G 180 - C C 204	A A 195 - A U 251
B G 181 - C C 203	A A 196 - A U 250
B U 185 - C A 196	A A 200 - A U 213
B G 186 - C C 195	A U 201 - A A 212
B G 187 - C C 194	A A 202 - A U 211
B C 188 - C G 193	A U 203 - A A 210



This is one input structure superimposed onto another which can also be viewed for each alignment.

ARTS conclusion

- The output gives the top ranked superposition's between the two input structures.
- A list of matched nucleotides is also generated, with an exact location (i.e. residue number).
- This gives a true comparison of the 3-D structure.

THE GUTELL LAB PROJECT

- Overview:

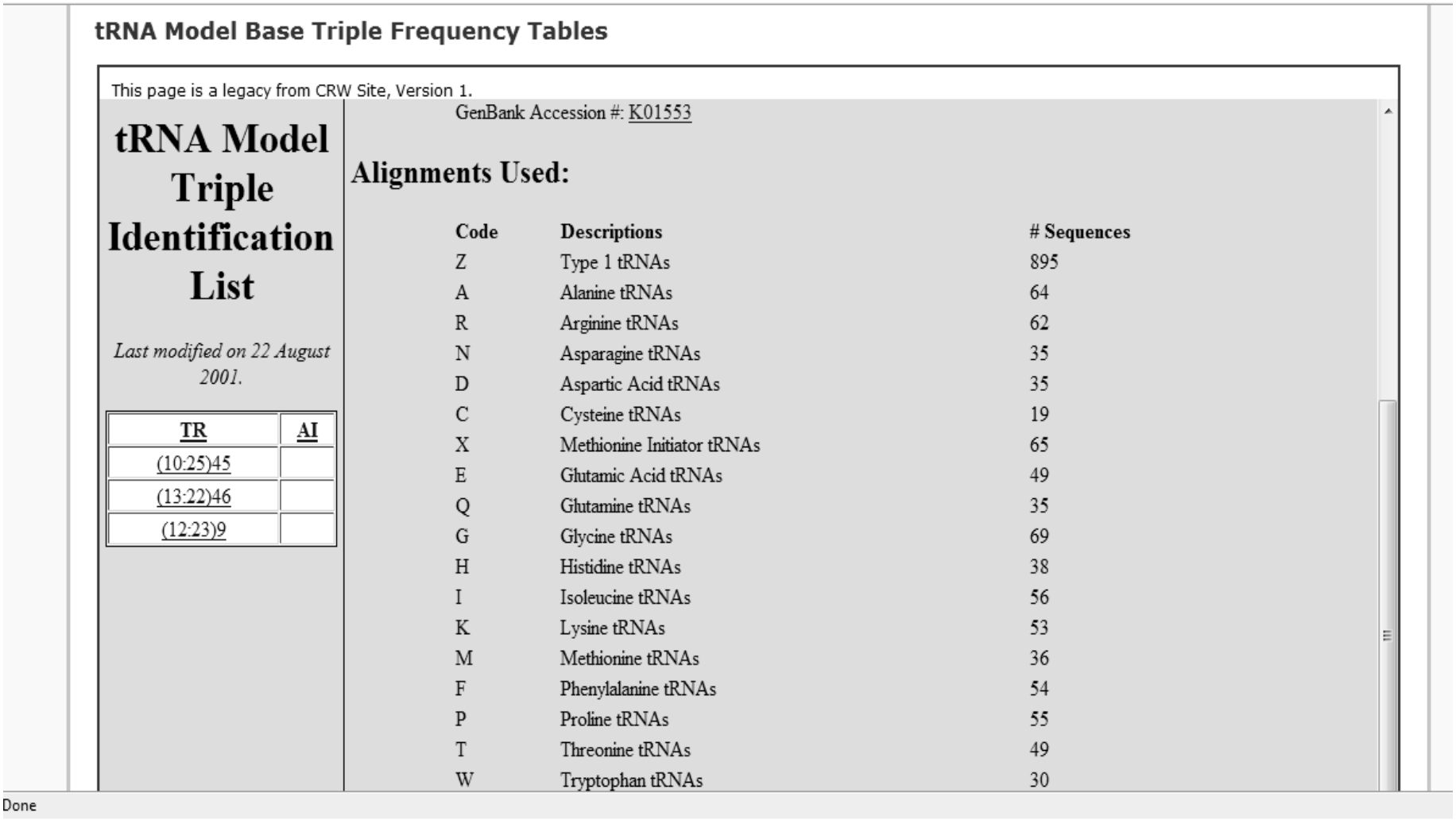
  The Gutell Lab, which is located at University of Texas at Austin, has a vast database of collected and analyzed RNA sequences. The website is a comparative RNA database. This information has been structured into two parts, the raw data (sequence and structure) and the processed data (analysis, accuracy, etc). All this data was determined using the comparative sequence analysis.

Comparative Information Database

Information is available on the following:

- rRNA (ribosomal RNA): 5S, 16S, 23S subunits
- tRNA (transfer RNA)
- Catalytic intron RNA's: Group 1 and Group 2

Sample table (tRNA)



tRNA Model Base Triple Frequency Tables

This page is a legacy from CRW Site, Version 1.

tRNA Model Triple Identification List

Last modified on 22 August 2001.

TR	AI
(10:25)45	
(13:22)46	
(12:23)9	

GenBank Accession #: K01553

Alignments Used:

Code	Descriptions	# Sequences
Z	Type 1 tRNAs	895
A	Alanine tRNAs	64
R	Arginine tRNAs	62
N	Asparagine tRNAs	35
D	Aspartic Acid tRNAs	35
C	Cysteine tRNAs	19
X	Methionine Initiator tRNAs	65
E	Glutamic Acid tRNAs	49
Q	Glutamine tRNAs	35
G	Glycine tRNAs	69
H	Histidine tRNAs	38
I	Isoleucine tRNAs	56
K	Lysine tRNAs	53
M	Methionine tRNAs	36
F	Phenylalanine tRNAs	54
P	Proline tRNAs	55
T	Threonine tRNAs	49
W	Tryptophan tRNAs	30

Done

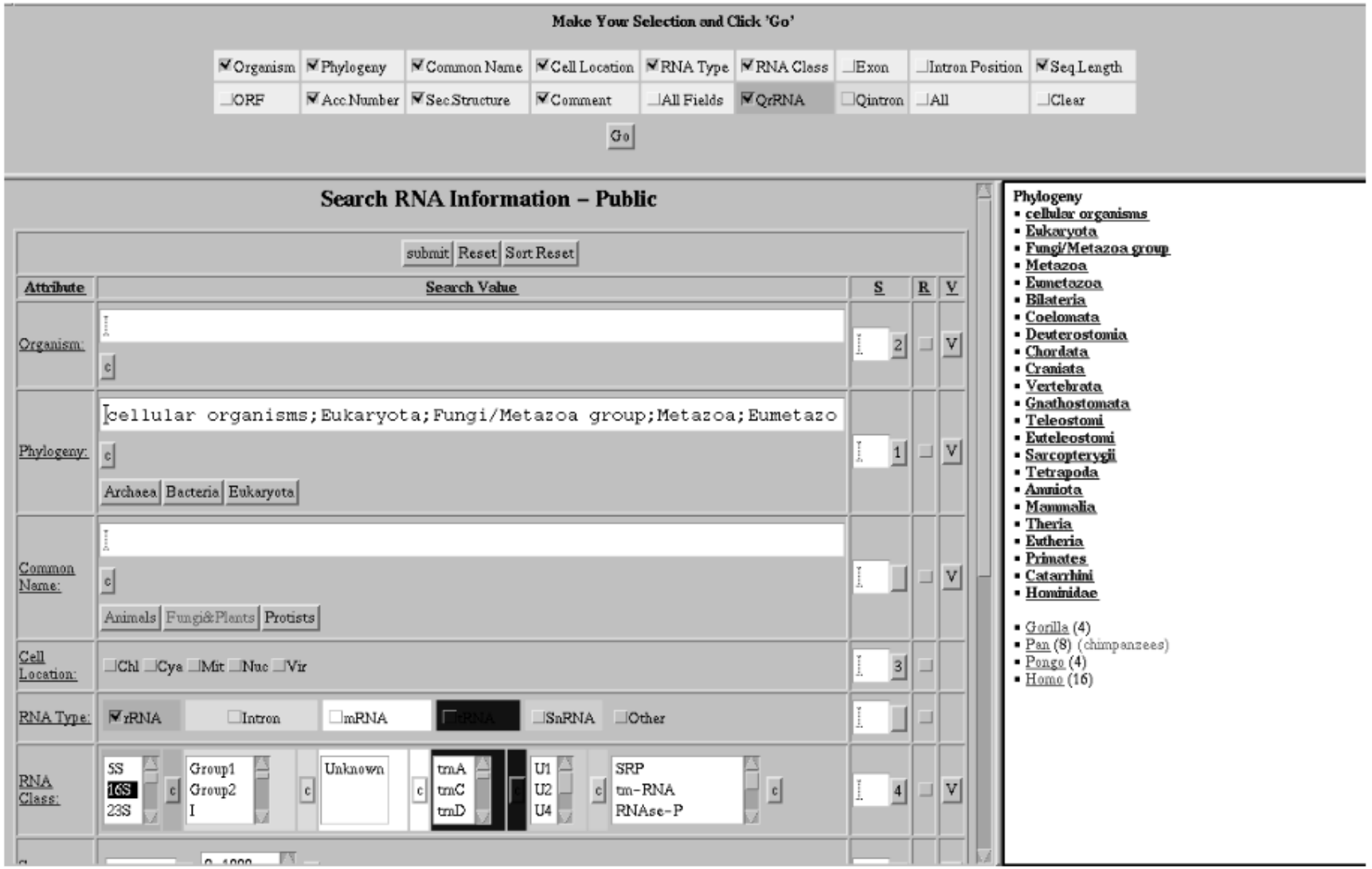
- Each table gives structural information along with the analysis and data collection. As seen in the previous slide, alignments and sequence number at which it took place were also included.

- This type of detailed data and analysis gives a researcher studying RNA structures or phylogenetics an essential tool in comparative studies.

Gutell database

- Gutell's lab also allows one to search RNA information stored in the databases with general and specific information about it.
- Some attributes include:
  - Organism (Order)
  - Phylogeny (Kingdom)
  - Location on the cell
  - Classes

Input


Make Your Selection and Click 'Go'
Organism
Phylogeny
Common Name
Cell Location
RNA Type
RNA Class
Exon
Intron Position
Seq Length
ORF
Acc Number
Sec Structure
Comment
All Fields
QrRNA
Qintron
All
Clear
Go
Search RNA Information – Public
submit
Reset
Sort Reset
Attribute
Search Value
S
R
V
Organism:
Phylogeny:
cellular organisms;Eukaryota;Fungi/Metazoa group;Metazoa;Eumetazo
Archaea
Bacteria
Eukaryota
Common Name:
Animals
Fungi&Plants
Protists
Cell Location:
Chl
Cya
Mit
Nuc
Vir
RNA Type:
rRNA
Intron
mRNA
tRNA
SnRNA
Other
RNA Class:
5S
16S
23S
Group1
Group2
I
Unknown
trnA
trnC
trnD
U1
U2
U4
SRP
tm-RNA
RNAse-P
Phylogeny
cellular organisms
Eukaryota
Fungi/Metazoa group
Metazoa
Eumetazoa
Bilateria
Coelomata
Deuterostomia
Chordata
Craniata
Vertebrata
Gnathostomata
Teleostomi
Euteleostomi
Sarcopterygii
Tetrapoda
Amniota
Mammalia
Theria
Eutheria
Primates
Catarrhini
Hominidae
Gorilla (4)
Pan (8) (chimpanzees)
Pongo (4)
Homo (16)

Results/ Output

close

Search Results – Public

PS | PDF | BPSEQ

Row #	Organism (2)	L (3)	RT (4)	RC	EX	IN	IP	O	Size	Cmp	AccNum	StrDiags	Common Name	Gr.ID	Gr.Class	Comment	Phylogeny M	Row #
1	Gorilla gorilla	M	R	16S		0	0		949	100	D38114		GORILLA				cellular organisms m	1
2	Gorilla gorilla	M	R	16S		0	0		950	100	X93347		GORILLA				cellular organisms m	2
3	Homo sapiens	M	R	16S		0	0		954	100	V00662		HUMAN				cellular organisms m	3
4	Homo sapiens	M	R	16S		0	0		954	100	J01415	d.16	HUMAN				cellular organisms m	4
5	Homo sapiens	M	R	16S		0	0		954	100	X62996		HUMAN				cellular organisms m	5
6	Homo sapiens	M	R	16S		0	0		954	100	V00710		HUMAN				cellular organisms m	6
7	Homo sapiens	M	R	16S		0	0		954	100	D38112		HUMAN				cellular organisms m	7
8	Homo sapiens	M	R	16S		0	0		954	100	X93334		HUMAN				cellular organisms m	8
9	Homo sapiens	N	R	16S		0	0		1868	100	M10098		HUMAN				cellular organisms m	9
10	Homo sapiens	N	R	16S		0	0		1869	100	X03205		HUMAN				cellular organisms m	10
11	Homo sapiens	N	R	16S		0	0		1870	100	K03432	d.16	HUMAN				cellular organisms m	11
12	Pan paniscus	M	R	16S		0	0		950	100	D38116		PYGMY CHIMPANZEE				cellular organisms m	12
13	Pan troglodytes	M	R	16S		0	0		949	100	D38113		CHIMPANZEE				cellular organisms m	13
14	Pan troglodytes	M	R	16S		0	0		956	100	X93335		CHIMPANZEE				cellular organisms m	14
15	Pan troglodytes	M	R	16S		0	0		961	100	X93340		CHIMPANZEE				cellular organisms m	15
16	Pan troglodytes	M	R	16S		0	0		962	100	X93342		CHIMPANZEE				cellular organisms m	16
17	Pongo pygmaeus	M	R	16S		0	0		953	100	D38115						cellular organisms m	17
18	Pongo pygmaeus abelii	M	R	16S		0	0		954	100	X97707						cellular organisms m	18

1 – 18 results displayed.

Query Results : 18
Searched by : Phylogeny=Hominidae, Cell Location=Mitochondrion,Nucleus, RNA Class=16S, Secondary Structure=Entries with Sequences or Structures, RNA Type=R
Sort order = Phylogeny,Organism,Cell Location,RNA Class

Data (cont.)

- There are different databases for various structures of the RNA.
- This differs from the location to the actual conformations.
- For instance there are mass data retrieval interfaces for sequence alignment, secondary structures and base pairs.

Alignments

- These alignments are sorted into tables by their alignment types.
- Depending on the type, the tables can give detailed information such as number of sequences found and sequences on other ribosome's as well.
- For instance the Ribosomal RNA alignment gives the alignments at the 5s, 16s, and 23s rRNA's subunits, with its sequence number for each.

Primary Alignments:

Table 1: Ribosomal RNA Alignments

Alignment (Abbreviation)	5S rRNA					16S rRNA					23S rRNA				
	AE2	GB	alnFA	rawFA	Seqs	AE2	GB	alnFA	rawFA	Seqs	AE2	GB	alnFA	rawFA	Seqs
Archaea (A)	AE2	GB	alnFA	rawFA	147	AE2	GB	alnFA	rawFA	788	AE2	GB	alnFA	rawFA	226
Bacteria (B)	AE2	GB	alnFA	rawFA	2,637	AE2	GB	alnFA	rawFA	35,998	AE2	GB	alnFA	rawFA	5,947
Chloroplast (C)	AE2	GB	alnFA	rawFA	199	AE2	GB	alnFA	rawFA	404	AE2	GB	alnFA	rawFA	442
Eukaryota (E)	AE2	GB	alnFA	rawFA	2,845	AE2	GB	alnFA	rawFA	1,937	AE2	GB	alnFA	rawFA	115
Mitochondria (M)	AE2	GB	alnFA	rawFA	49	AE2	GB	alnFA	rawFA	899	AE2	GB	alnFA	rawFA	285
Three Phylogenetic Domains (3)	AE2	GB	alnFA	rawFA	5,624	AE2	GB	alnFA	rawFA	5,591	AE2	GB	alnFA	rawFA	585
Three Phylogenetic Domains/Two Organelles (T)	AE2	GB	alnFA	rawFA	5,868	AE2	GB	alnFA	rawFA	6,389	AE2	GB	alnFA	rawFA	922

Table 2: Intron RNA Alignments

Alignment (Abbreviation)	Intron RNA				
	AE2	GB	alnFA	rawFA	Seqs
Group IA Introns	AE2	GB	alnFA	rawFA	109
Group IB Introns	AE2	GB	alnFA	rawFA	201
Group IC1 Introns	AE2	GB	alnFA	rawFA	818
Group IC2 Introns	AE2	GB	alnFA	rawFA	31
Group IC3 Introns	AE2	GB	alnFA	rawFA	984
Group ID Introns	AE2	GB	alnFA	rawFA	21
Group IE Introns	AE2	GB	alnFA	rawFA	248
Group IIA Introns	AE2	GB	alnFA	rawFA	174
Group IIB Introns	AE2	GB	alnFA	rawFA	607

Conclusions:

- As more and more motifs are being identified and studied, the easier it will become to label new patterns and fold on the RNA structure.
- The extent of what motifs do are still very much the focus of numerous studies.
- Most work done already is limited to motifs in RNA secondary structures.

Conclusions contd..

- Resources such as The Gutell Lab provides a database of sequenced alignments of the RNA structure, which can be used for comparative analysis.
- With some of the figures and structures in the Gutell database, further studies can be done with the sequences and structural information. These can also give the location of the sequence which is key in determining motifs and patterns.

Algorithms and programs shown in these slides are a few of many programs developed. These algorithms provide the easiest and simplest search on a given sequence, which can ultimately help in discovering a more accurate pattern of motifs. Although there are many other databases or programs developed that may be more in-depth with motifs, those can also complicate searches with extra parameters. However we can always cross validate our data with them as well.

Resources

1) Wadley LM, Pyle AM. The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery. Nucleic Acid Res. 2004;32:6650–6659. [PubMed]

2) Neocles B Leontis, Aurelie Lescoute, Eric Westhof. The building blocks and motifs of RNA architecture. Current Opinion in Structural Biology 2006, 16:279–287

3) Zorn J, Gan HH, Shiffeldrim N, Schlick T: Structural motifs in ribosomal RNAs: implications for RNA design and genomics. Biopolymers 2004, 73:340-347.

4) Olson W.K. (1976) The spatial configuration of ordered polynucleotide chains. I. helix formation and base stacking. Biopolymers, 15:, 859–878.

5) Robert T. Batey, Robert P. Rambo, Jennifer A. Doudna. Tertiary Motifs in RNA Structure and Folding. Angew. Chem. Int. Ed. 1999, 38, 2326 ± 2343

Resources

6) Mira Abraham, Oranit Dror, Ruth Nussinov. Analysis and classification of RNA tertiary structures. RNA (2008), 14:1–16.

7) Mahassine Djelloul, Alain Denise. Automated motif extraction and classification in RNA tertiary structures. RNA (2008), 14:2489–2497

8) Yurong Xin, Christian Laing, Neocles B. Leontis …. Annotation of tertiary interactions in RNA structures reveals …. RNA (2008), 14:2465–2477

9) Cannone J.J., Subramanian S., Schnare M.N., Collett J.R., D'Souza L.M., Du Y., Feng B., Lin N., Madabusi L.V., MÜller K.M., Pande N., Shang Z., Yu N., and Gutell R.R. (2002). The Comparative RNA Web (CRW) Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and Other RNAs. *BioMed Central Bioinformatics*, 3:2. [Correction: *BioMed Central Bioinformatics*. 3:15.]

10) The Gutell Lab: http://www.rna.ccbb.utexas.edu/

Resources

11) Poul Nissen , Joseph A. Ippolito,Nenad Ban,Peter B. Moore, Thomas A. Steitz. **RNA tertiary interactions in the large ribosomal subunit: The A-minor motif .** *PNAS April 24, 2001 vol. 98 no. 9 4899-4903*

12) ***The Pyle Lab: http://www.pylelab.org/***

13) ***Web reference: http://web.virginia.edu/Heidi/chapter12/chp12.htm#f13***