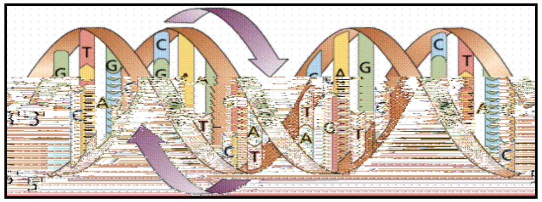
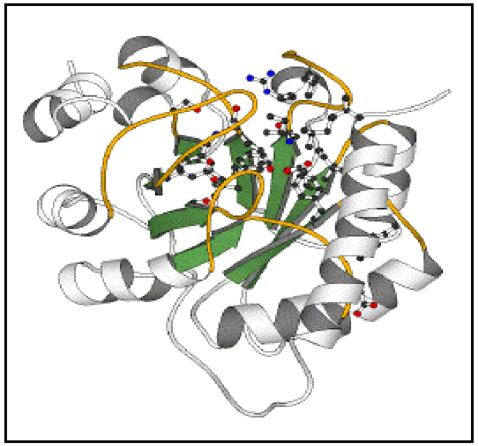
Hydration of protein-RNA recognition sites

Ranjit P. Bahadur
Assistant Professor
Department of Biotechnology
Indian Institute of Technology Kharagpur, India

1st November, 2013

Central Dogma of life





Storage Medium

DNA

CCTGAGCCAACTATTGATGAA

transcription

Transmission Medium

mRNA

CCUGAGCCAACUAUUGAUGAA

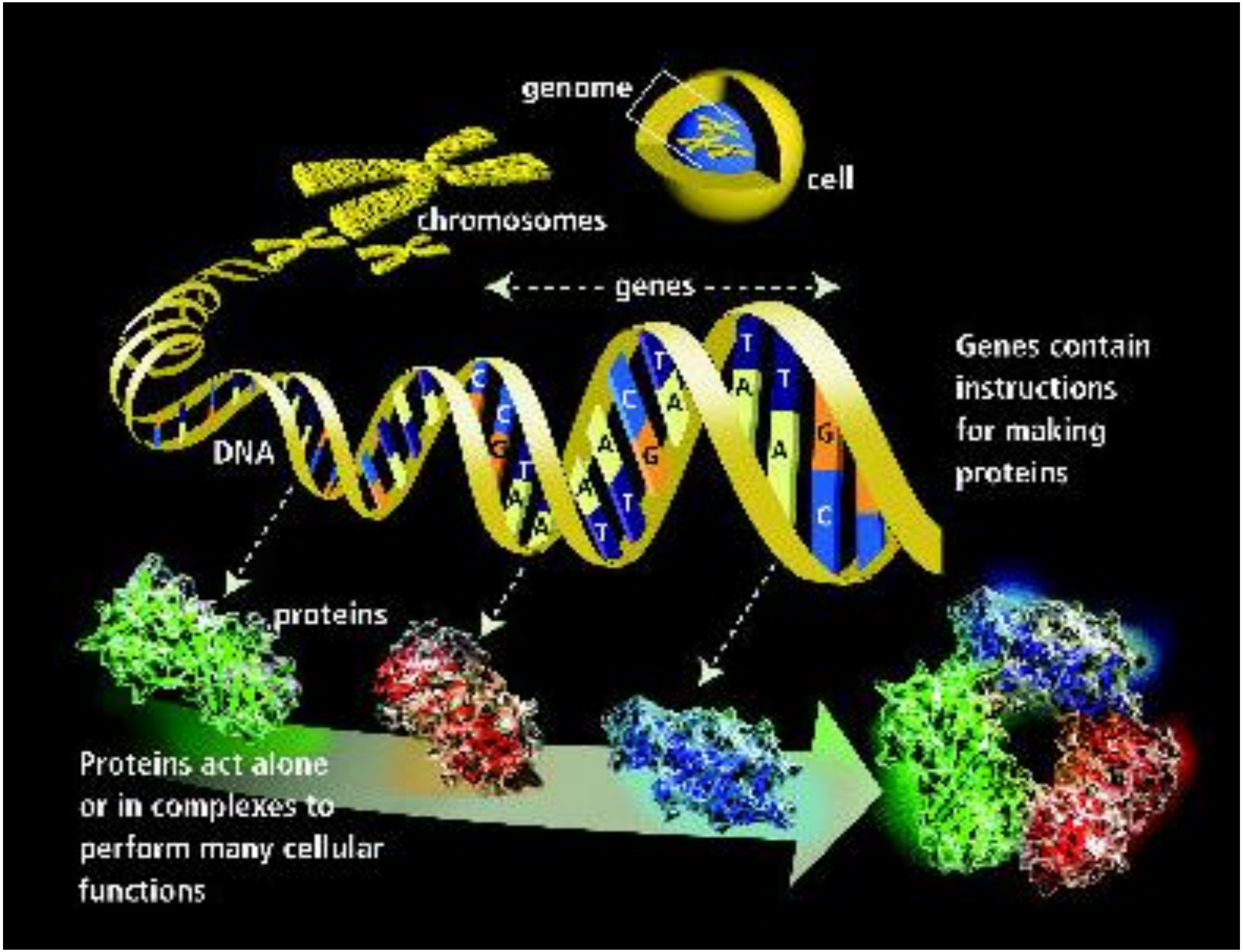
translation

Text composed of 3-letter words (with 4 unique letters) gets mapped into a seq of 20 unique letters

Protein

PEPTIDE

Flow of genetic information



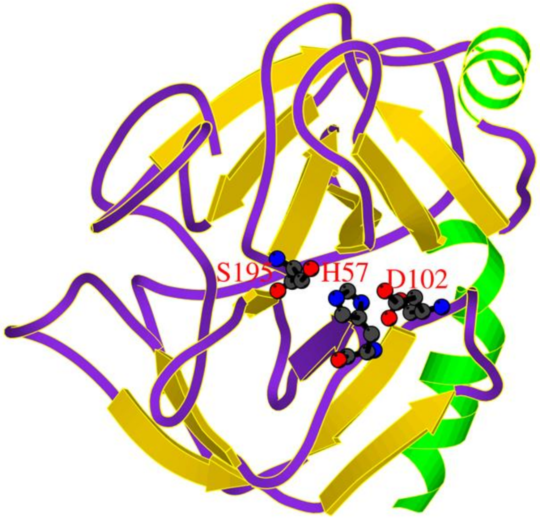
What is a protein

Proteins are polymers built up from 20 different amino-acids linked end to end by peptide bonds, which folded into a 3D structure with cellular function

```
CGVPAIQPVLSGLXXIVNGEEAVPGSWPWQVSLQDKTGFHFCGGSLINENWVVTAAHCGV
TTSDVVVAGEFDQGSSSEKIQKLKIAKVFKNSKYNSLTINNDITLLKLSTAASFSQTVSA
VCLPSASDDFAAGTTCVTTGWGLTRYXXANTPDRLQQASLPLLSNTNCKKYWGTKIKDAM
ICAGASGVSSCMGDSGGPLVCKKNGAWTLVGIVSWGSSTCSTSTPGVYARVTALVNWVQQ
TLAAN*
```

More directly linked to function

Better conserved than sequence



Amino acids: building blocks

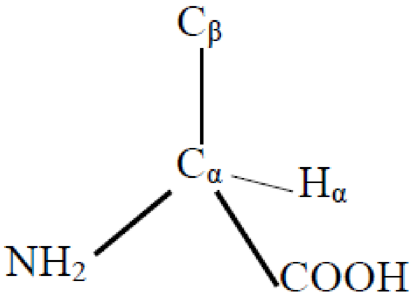
Primary structure of proteins

- Amino acid sequence of a protein: primary protein structure.
- 20 naturally amino acids occur in proteins.
- Amino acids have a common chemical structure:

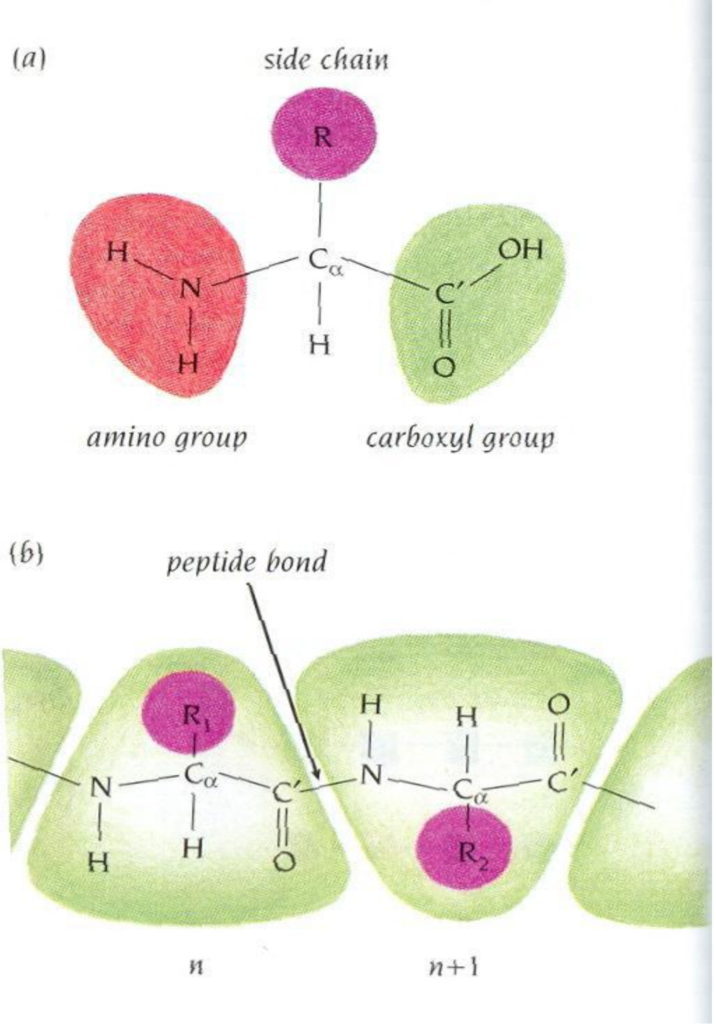
  Tetrahedral (sp3) carbon atom ($C\alpha$) bound to four asymmetric groups:

1. Amino group (NH_2)
2. Carboxy group (COOH)
3. H atom
4. Functional chemical group: characteristic for the amino acid

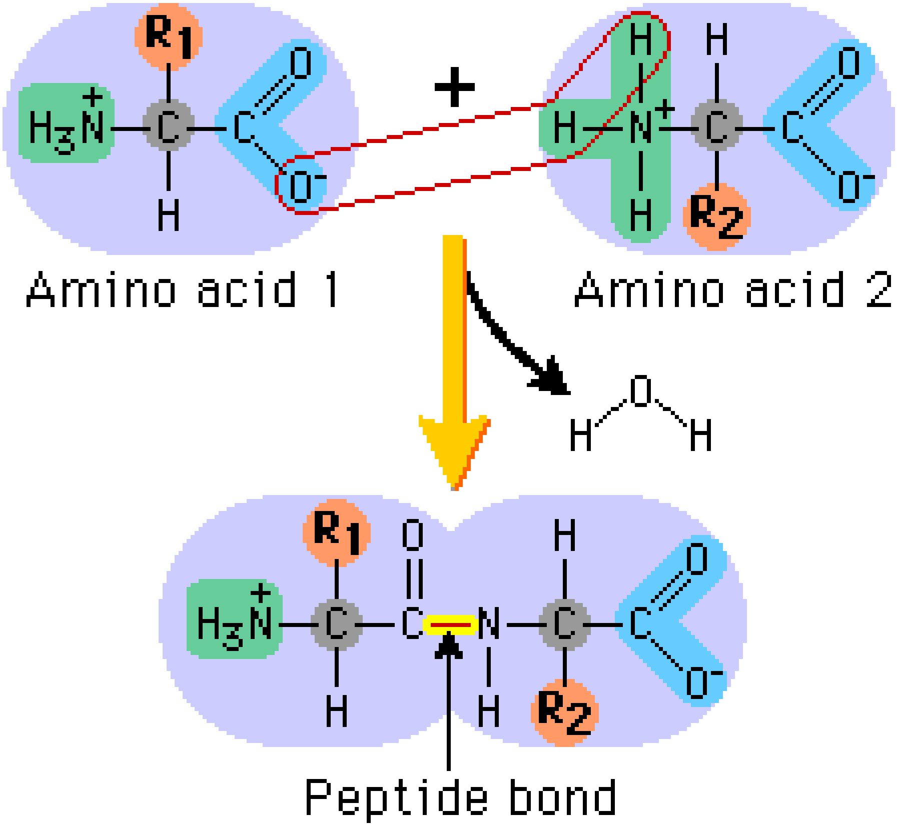
- Naturally occuring amino acids are typically L amino acids.



Polypeptide bonds: main chain or backbone



The peptide bond



Dehydration synthesis or condensation reaction

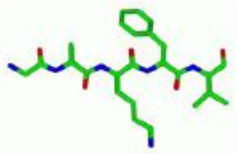
(----ALA-HIS-GLY-ILE-LEU-PHE-TYR-LYSGLY---)$_n$

Protein structure hierarchy

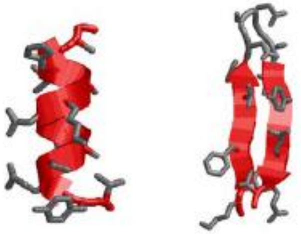
Overview on protein structure (folded state)

- **Primary structure**
  - Sequence of the protein

-Gly-Ala-Arg-Phe-Val-
G A R F V

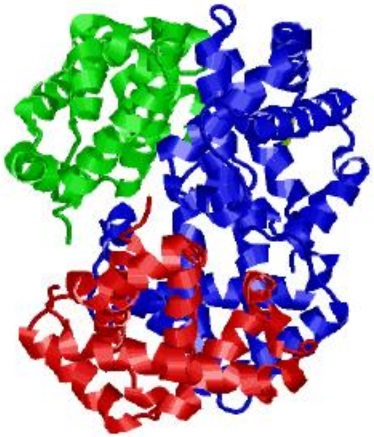


- **Secondary structure**
  - Regular local structural elements stabilized by hydrogen bonds



- **Tertiary structure**
  - Three-dimensional fold of the protein

- **Quaternary structure:**
  - Non-covalent assembly of several proteins to form a functional complex



Amino acid residues:: chemical properties

Amino acids can be grouped based on side chain character

Non-polar side-chains (hydrophobic):

Alanine (Ala), Valine (Val), Isoleucine (Ile), Leucin (Leu), Methionine (Met), Phenylalanine (Phe), Proline (Pro) and Tryptophan (Trp), also Glycine (Gly)

Polar side-chains (hydrophilic):

Serine (Ser), Threonine (Thr), Cysteine (Cys), Asparagine (Asn), Glutamine (Gln), Tyrosine (Tyr) and Histidine (His)

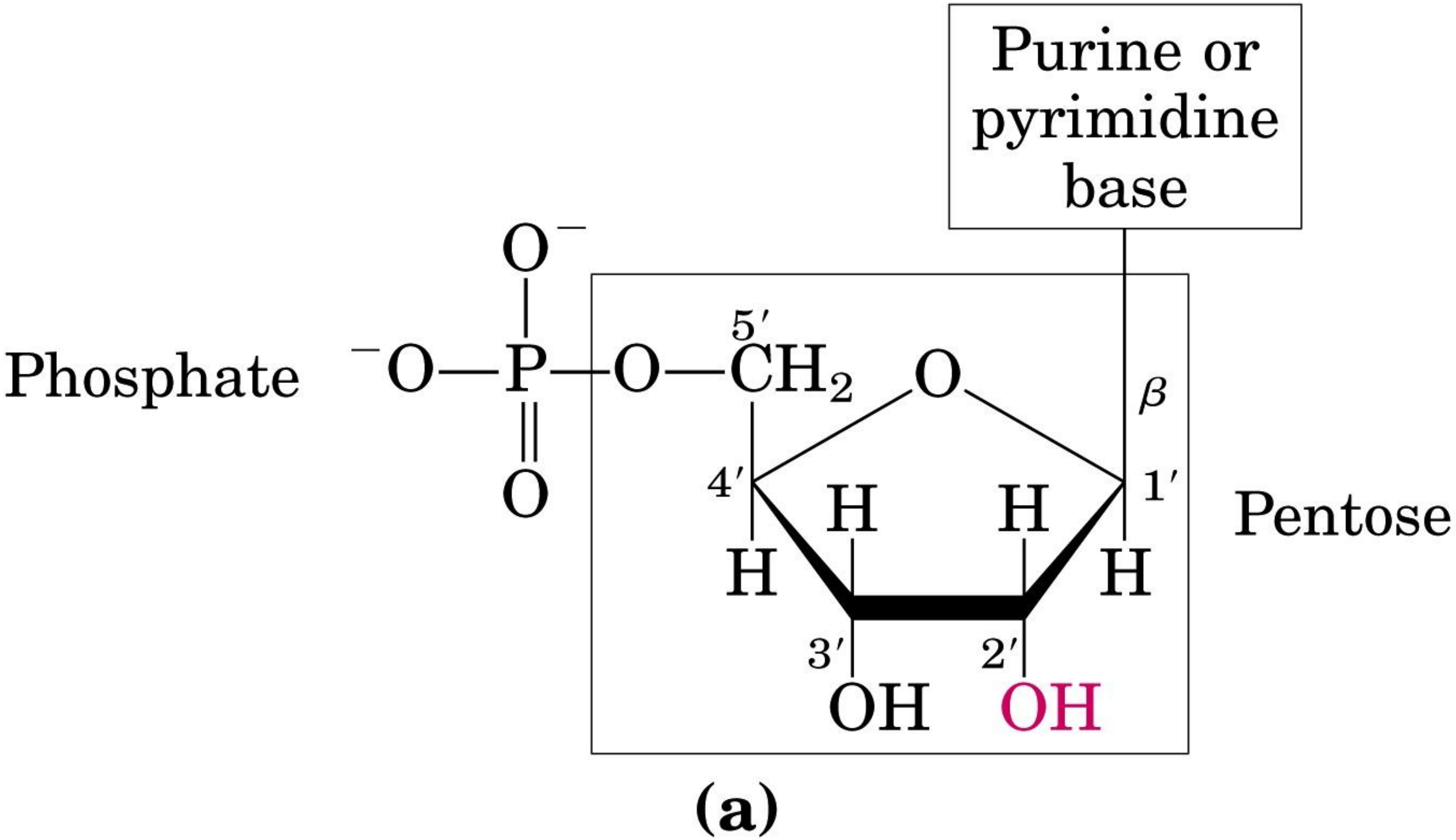
Charged side-chains (hydrophilic):

Positive charge: Arginine (Arg) and Lysine (Lys)

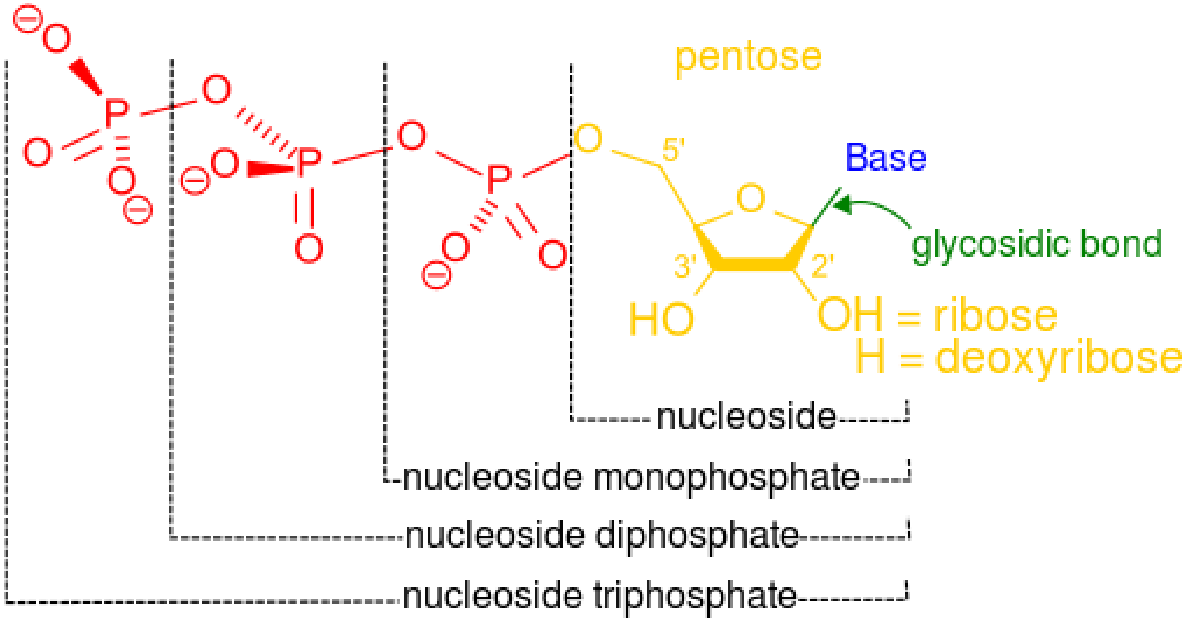
Depending on the pH, Histidine (His) can also be positively charged.

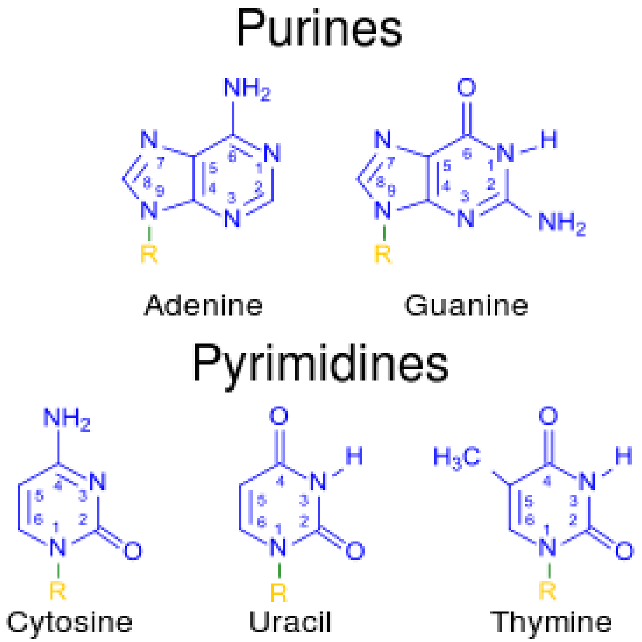
Negative charge: Glutamic acid (Glu) and Aspartic acid (Asp)

Nucleic acids structure:: Sugar-Phosphate backbone

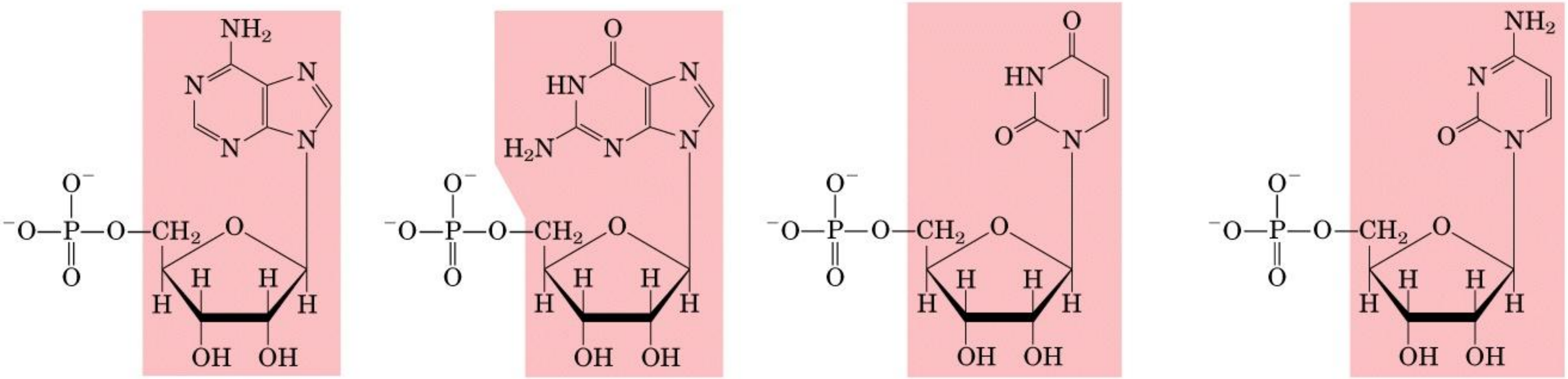


Nucleic acids structure:: Bases





Nucleic acids structure:: Nucleotide and Nucleoside



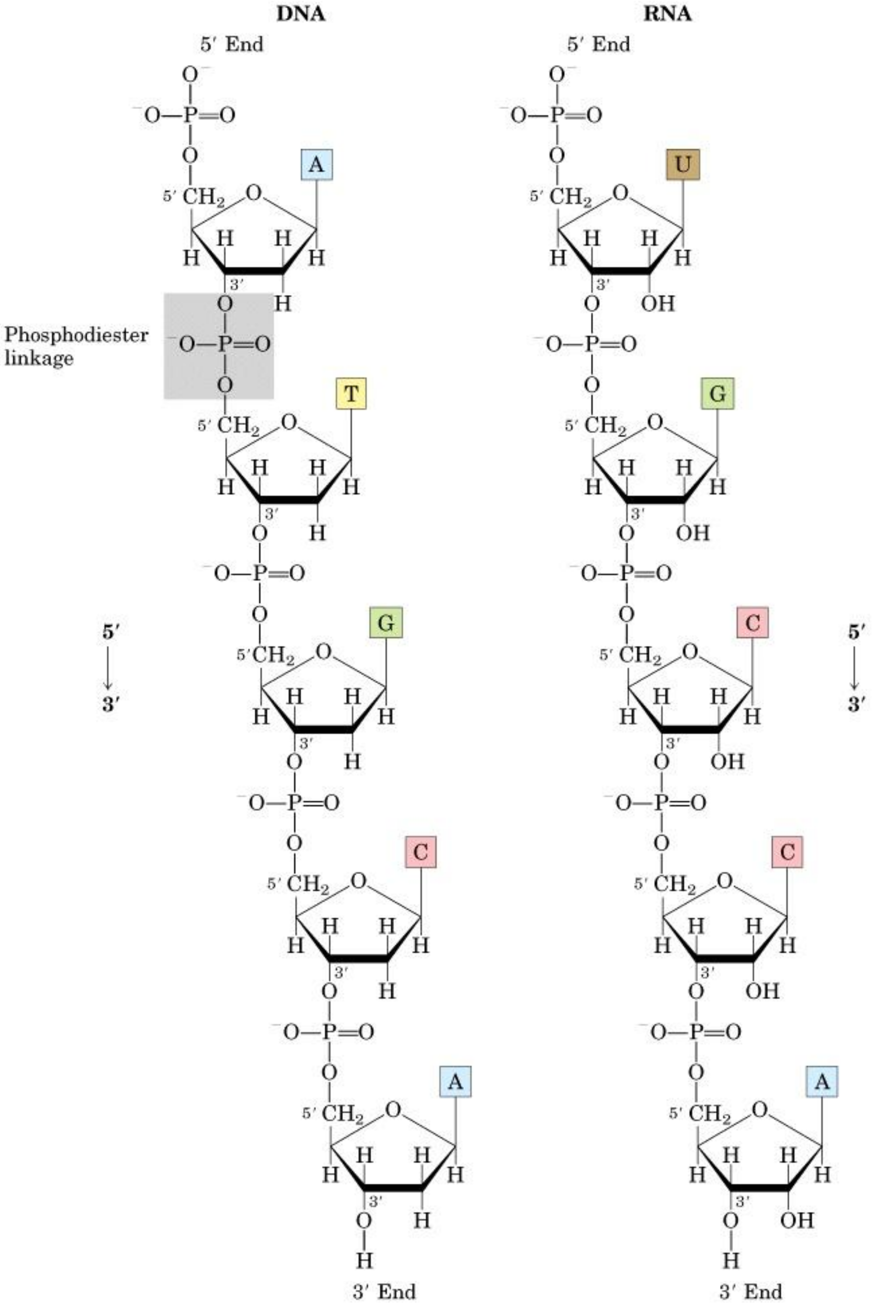
Nucleotide:	Adenylate (adenosine 5′-monophosphate)	Guanylate (guanosine 5′-monophosphate)	Uridylate (uridine 5′-monophosphate)	Cytidylate (cytidine 5′-monophosphate)
Symbols:	A, AMP	G, GMP	U, UMP	C, CMP
Nucleoside:	Adenosine	Guanosine	Uridine	Cytidine

(b) Ribonucleotides

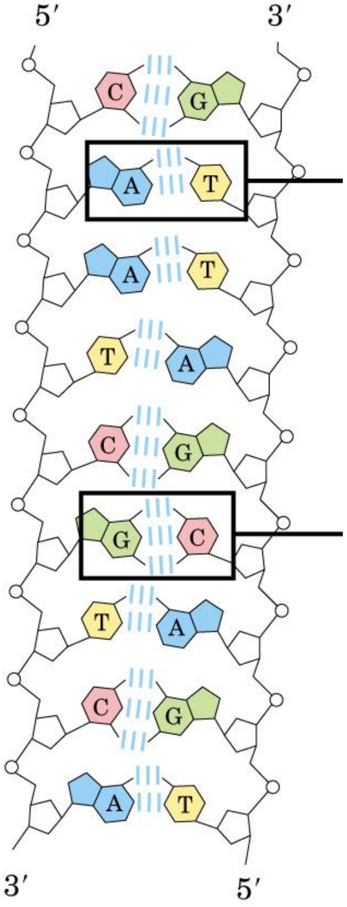
A **nucleotide** is composed of a nucleobase (also termed a nitrogenous base), a five-carbon sugar (either ribose or 2-deoxyribose), and one or more phosphate groups.

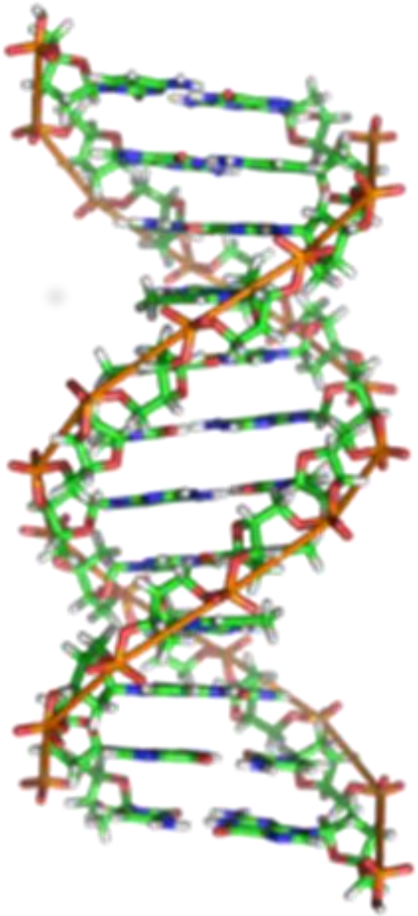
Without the phosphate group, the nucleobase and sugar compose a **nucleoside**.

Nucleic acids structure:: Polyribonucleotide (DNA and RNA)

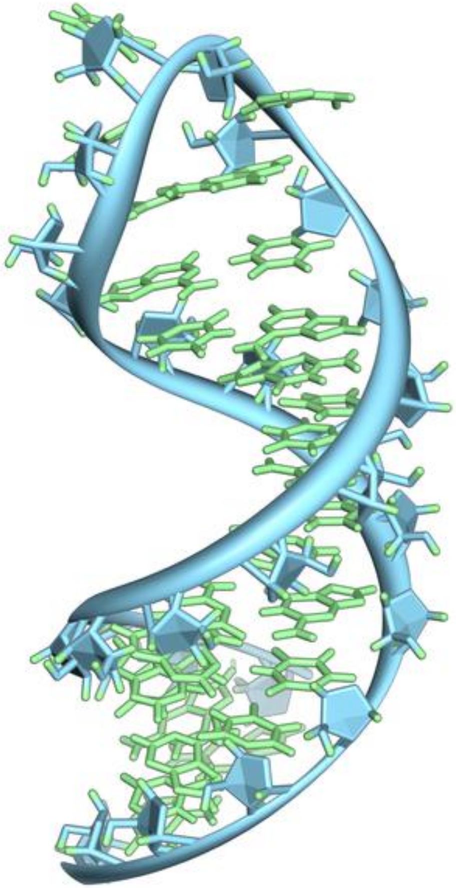


Nucleic acids structure:: base pairing in DNA and in RNA



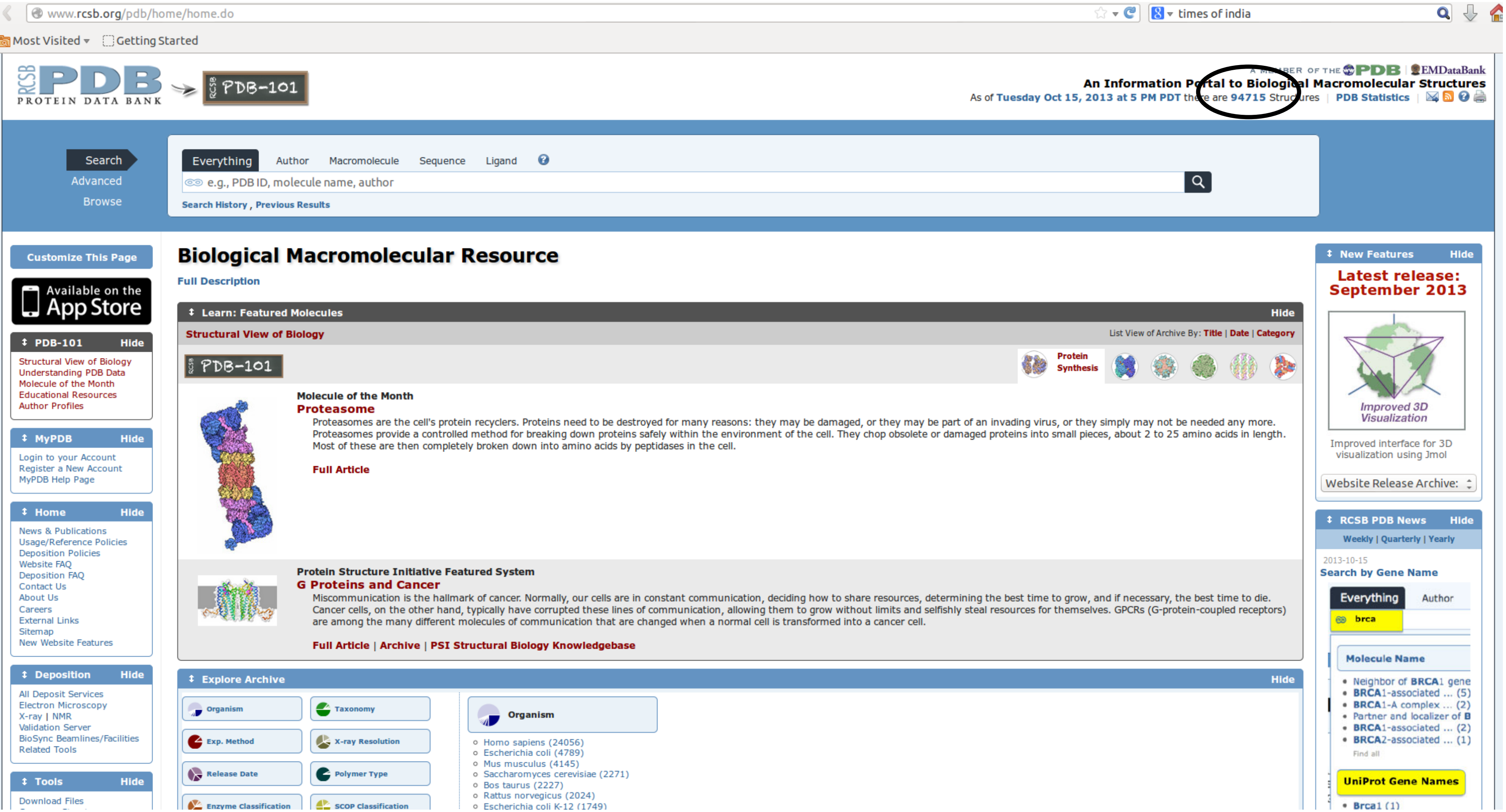


DNA- B form



Duplex RNA

The Protein Data Bank (PDB)



An Information Portal to Biological Macromolecular Structures

As of **Tuesday Oct 15, 2013 at 5 PM PDT** there are **94715** Structures

Biological Macromolecular Resource

Full Description

Molecule of the Month

Proteasome

Proteasomes are the cell's protein recyclers. Proteins need to be destroyed for many reasons: they may be damaged, or they may be part of an invading virus, or they simply may not be needed any more. Proteasomes provide a controlled method for breaking down proteins safely within the environment of the cell. They chop obsolete or damaged proteins into small pieces, about 2 to 25 amino acids in length. Most of these are then completely broken down into amino acids by peptidases in the cell.

Protein Structure Initiative Featured System

G Proteins and Cancer

Miscommunication is the hallmark of cancer. Normally, our cells are in constant communication, deciding how to share resources, determining the best time to grow, and if necessary, the best time to die. Cancer cells, on the other hand, typically have corrupted these lines of communication, allowing them to grow without limits and selfishly steal resources for themselves. GPCRs (G-protein-coupled receptors) are among the many different molecules of communication that are changed when a normal cell is transformed into a cancer cell.

Organism

- Homo sapiens (24056)
- Escherichia coli (4789)
- Mus musculus (4145)
- Saccharomyces cerevisiae (2271)
- Bos taurus (2227)
- Rattus norvegicus (2024)
- Escherichia coli K-12 (1749)

Latest release: September 2013

Improved interface for 3D visualization using Jmol

2013-10-15

Search by Gene Name

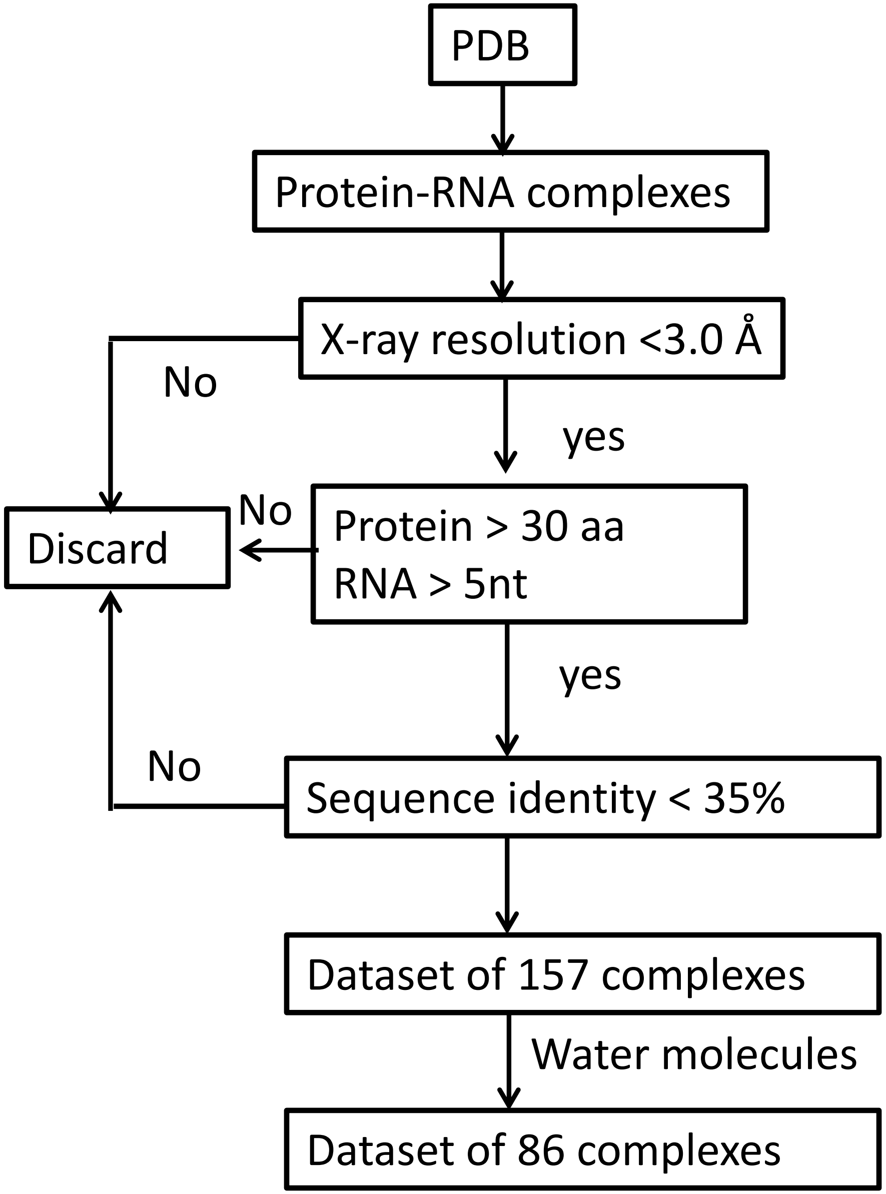
- Neighbor of **BRCA**1 gene
- **BRCA**1-associated ... (5)
- **BRCA**1-A complex ... (2)
- Partner and localizer of B
- **BRCA**1-associated ... (2)
- **BRCA**2-associated ... (1)

- **Brca**1 (1)

Objective

Understand the role of interface water molecules in protein-RNA recognition

Non redundant dataset of protein-RNA complexes



PDB

Protein-RNA complexes

X-ray resolution <3.0 Å

No

yes

Discard

No

Protein > 30 aa
RNA > 5nt

yes

No

Sequence identity < 35%

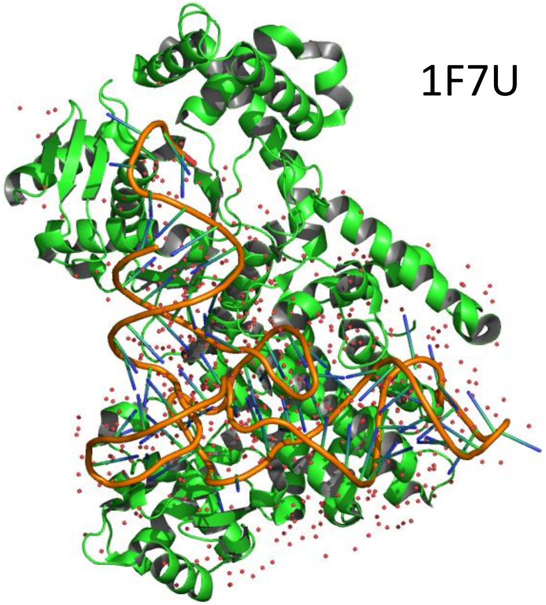
Dataset of 157 complexes

Water molecules

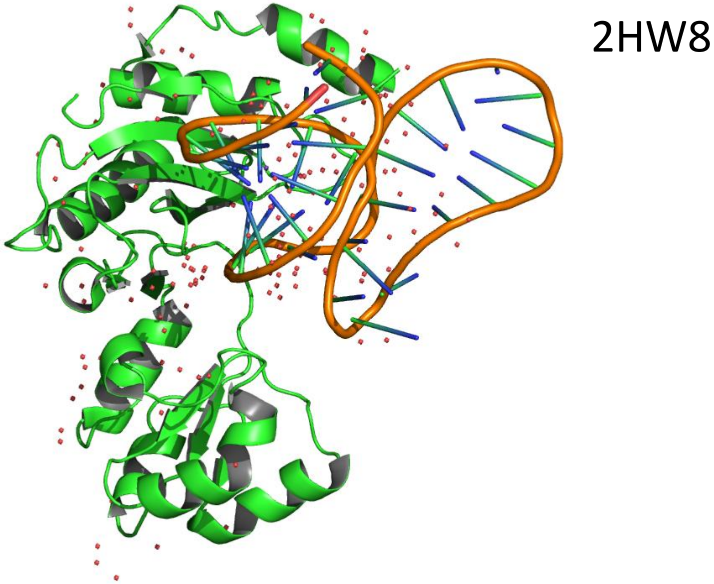
Dataset of 86 complexes

Protein-RNA complexes

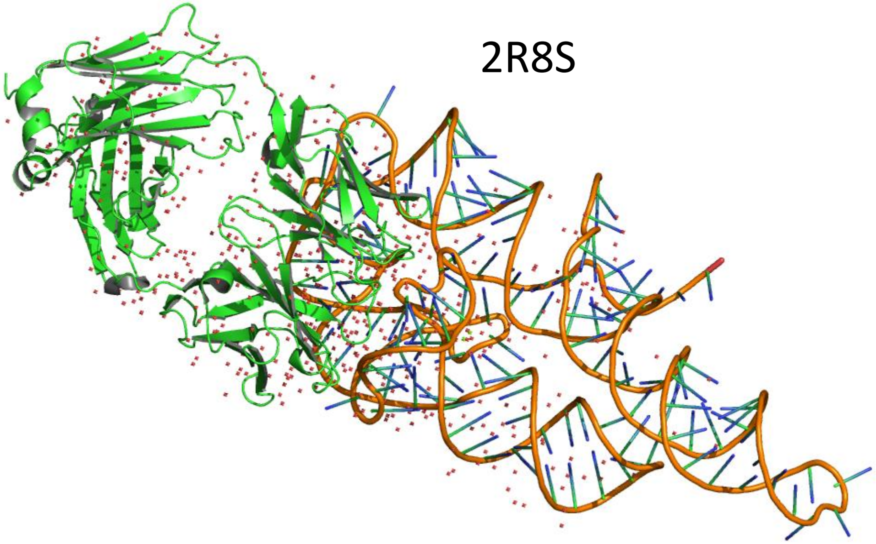
(A) Complexes with tRNA



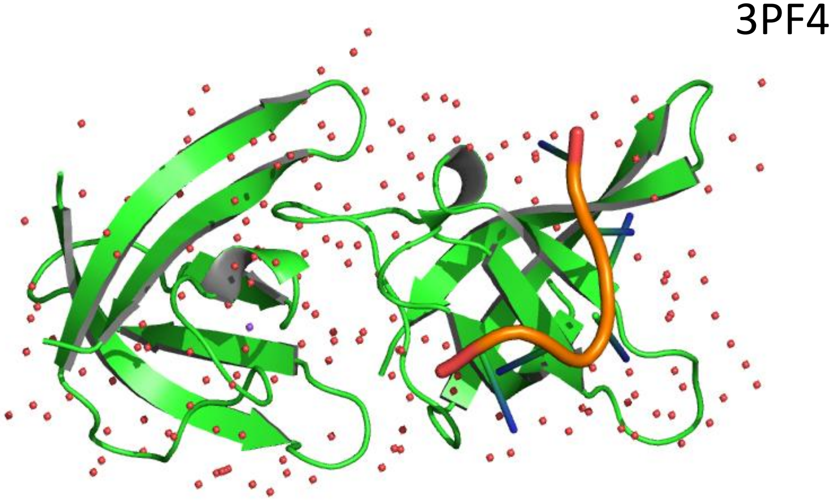
(B) Complexes with ribosomal proteins



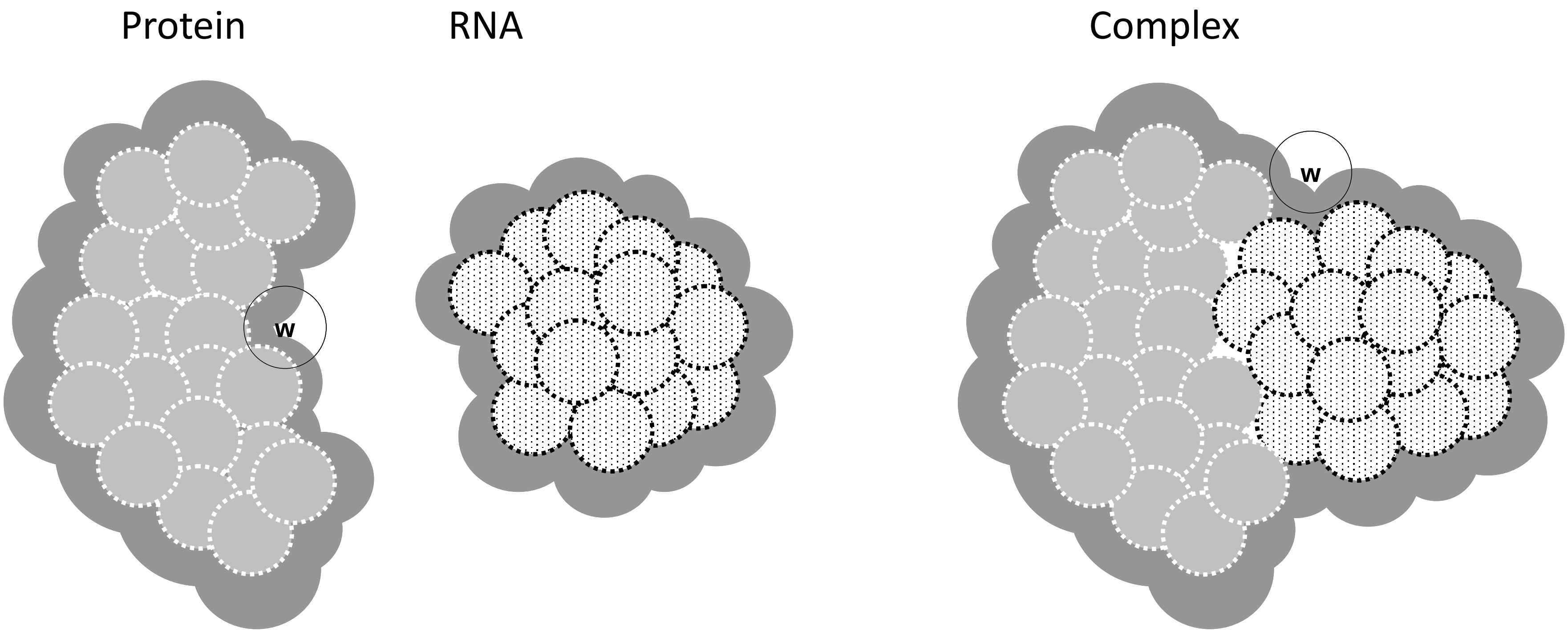
(C) Complexes with duplex RNA



(D) Complexes with single-stranded RNA



Protein-RNA interface definition



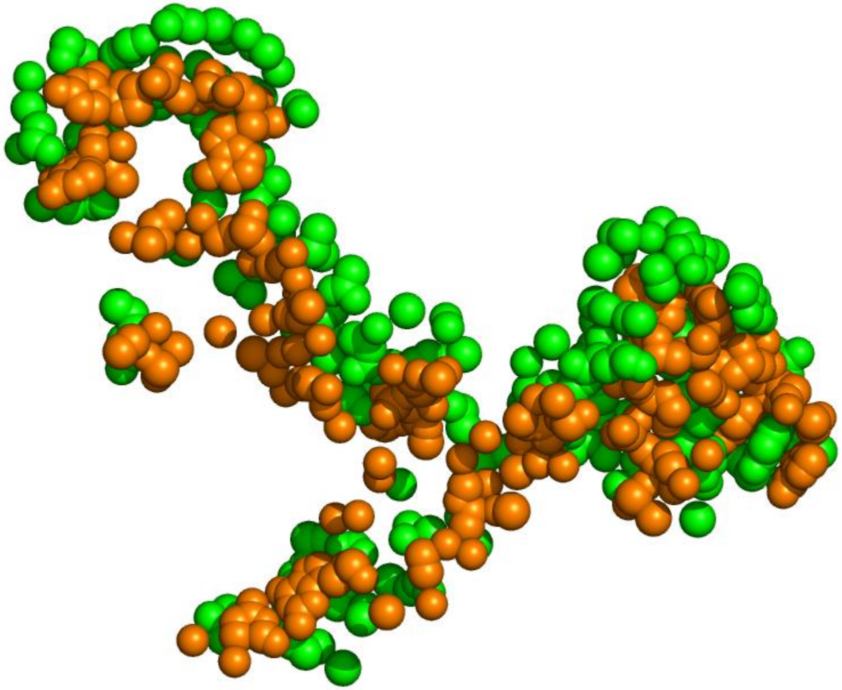
Interface area (*B*) = ASA(protein) + ASA(RNA) – ASA(complex)

‘NACCESS’ by Hubbard SJ, 1992

Interface atoms and residues/nt are all atoms and residues/nt that lose ASA in the complex and contribute to *B*

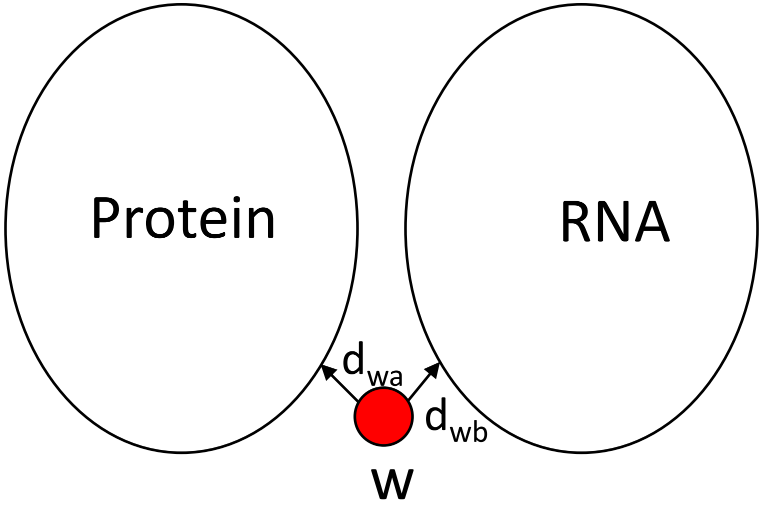
Protein-RNA interface: Arginyl-tRNA synthetase (1F7U)

Complex



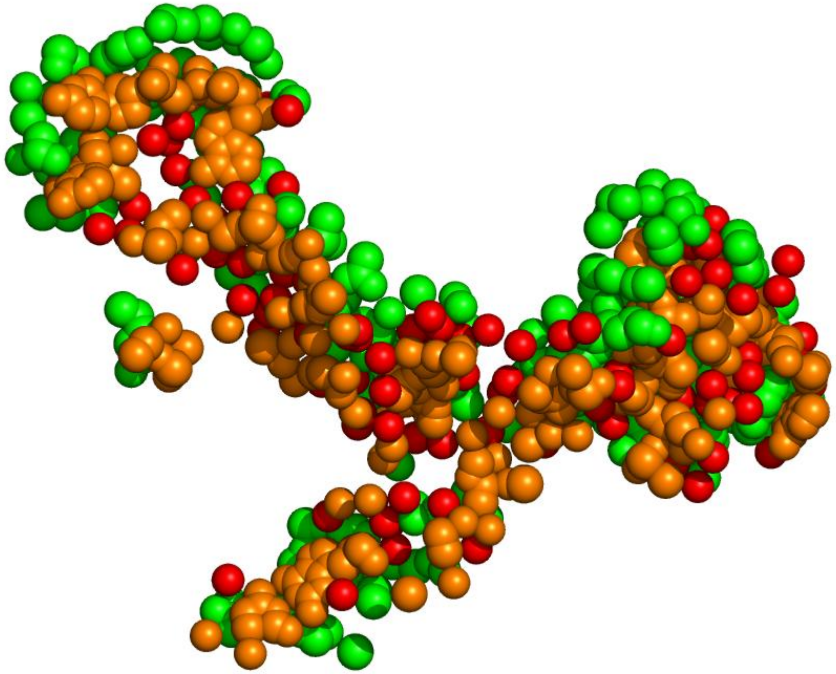
Interface atoms

Definition of interface water

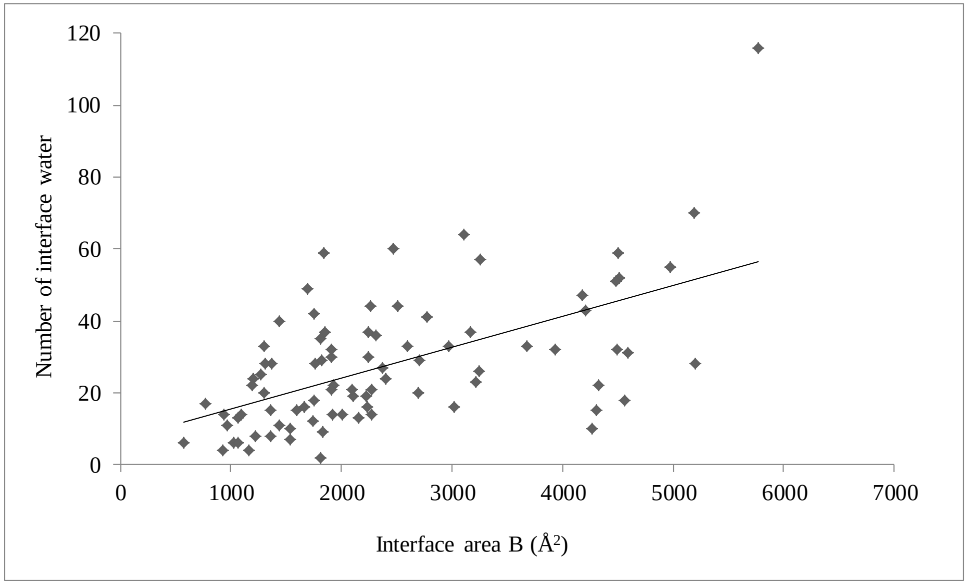


d_{wp} and d_{wr} < 4.5Å

1F7U



Interface water increases with the size of the interface

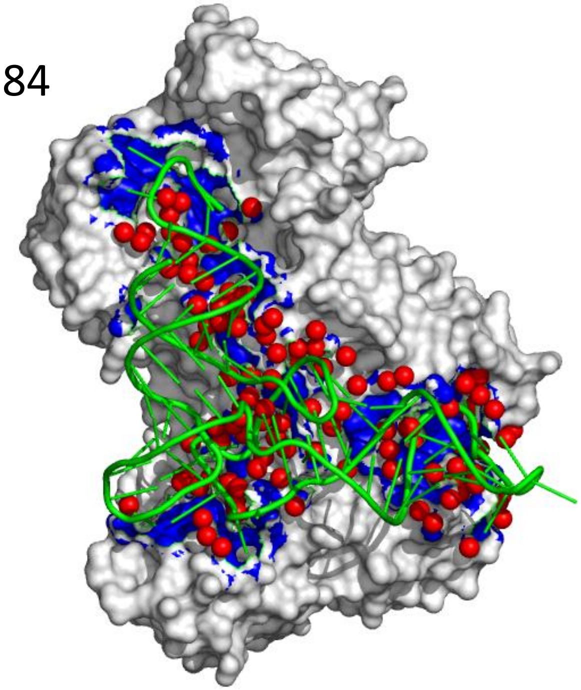


- 86 interfaces
- 2376 interface water molecules

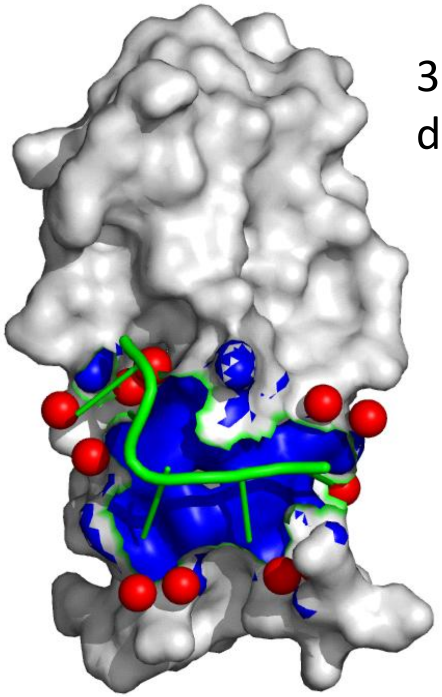
Interface statistics	protein-RNA				
	tRNA	ribosomal	duplex	single stranded	All
Number of complexes	12	5	26	43	86
Interface area B (Å²)	4185±1226	1725±434	2612±1066	1866±851	2407±1237
Interface water					
range	8-116	12-49	9-70	2-51	2-116
mean number per interface	39±29	25 ±15	33±18	21±12	28±18
per 1000 Å² of B	9.3	14.5	12.6	11.3	11.6

- Interfaces with tRNA are more dehydrated compare to single stranded RNA or duplex RNA.
- Interfaces with ribosomal proteins are least dehydrated.

Quantify the 'dry' and 'wet' interfaces



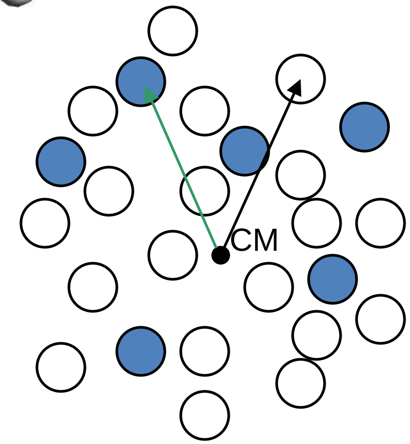
1F7U
$d_r = 0.84$



3PF4
$d_r = 1.46$

$\langle r_{wat} \rangle$ = average distances of interface water molecules from the center of mass of the interface

$\langle r_{atom} \rangle$ = average distances of interface atoms from the center of mass of the interface



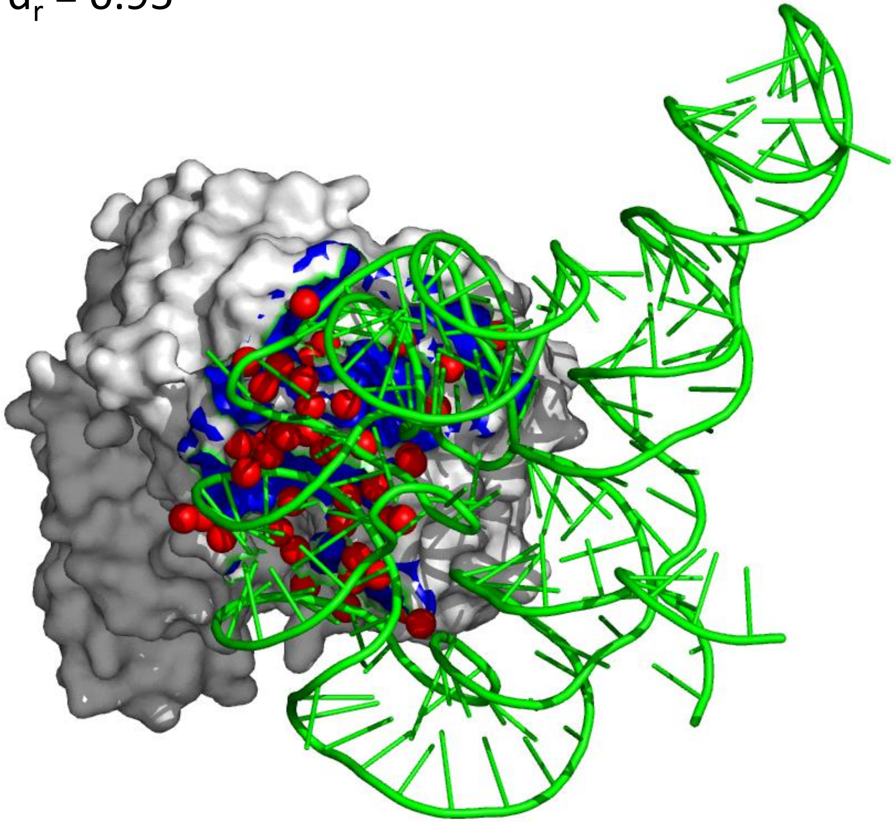
$$d_r = \langle r_{wat} \rangle / \langle r_{atom} \rangle$$

$d_r > 1.0$; "dry" interface

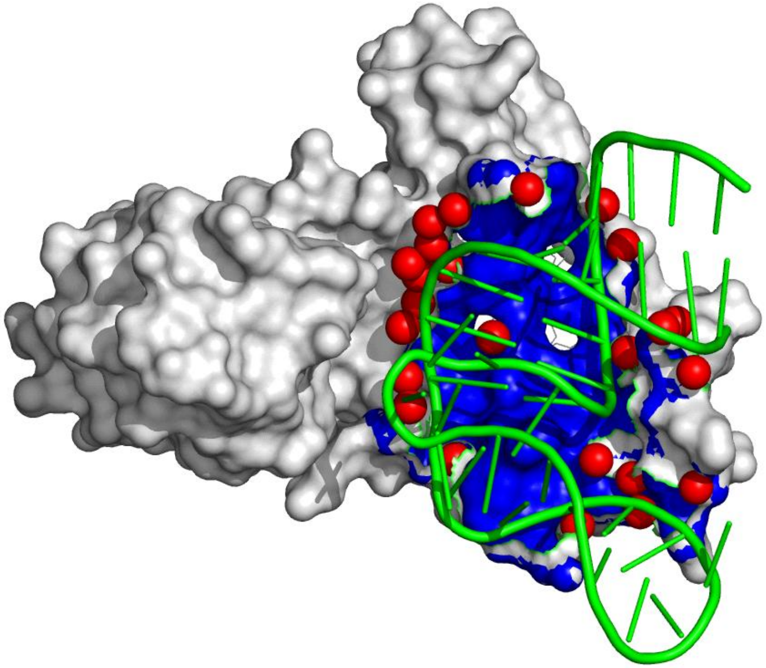
$d_r \leq 1.0$; "wet" interface

The 'dry' and 'wet' protein-RNA interfaces

2R8S
$d_r = 0.95$



2HW8
$d_r = 1.18$



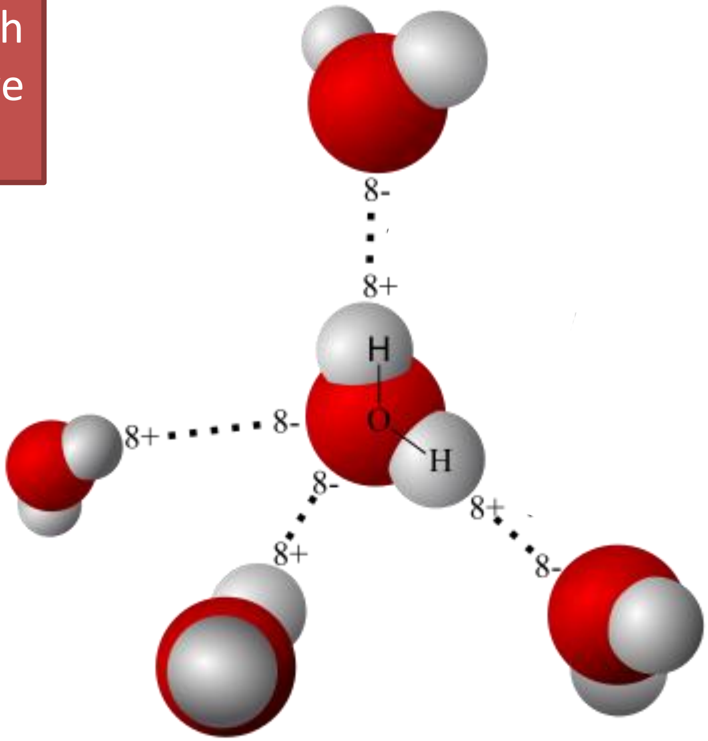
Interfaces with tRNA and duplex RNA relatively more 'wet' than interfaces with ribosomal proteins and single stranded RNA.

Hydrogen bond

A hydrogen bond is the electrostatic attractive interaction between polar molecules in which hydrogen (H) is bound to a highly electronegative atom, such as nitrogen (N), oxygen (O) or fluorine (F).

Bond dissociation energy 20KJ/mol (water)

Weaker than covalent bond but stronger



Hydrogen bonds between molecules of water

Hydrogen bonds involving interface water molecules

Interface statistics	protein-RNA				
	tRNA	ribosomal	duplex	single stranded	All
Water mediated H-bonds					
range	16-198	23-78	17-134	5-130	5-198
Mean number per interface	76±50	45±22	62±34	40±26	52±35
Per 1000 Å^2 of B	18.2	26.1	23.7	21.4	21.6

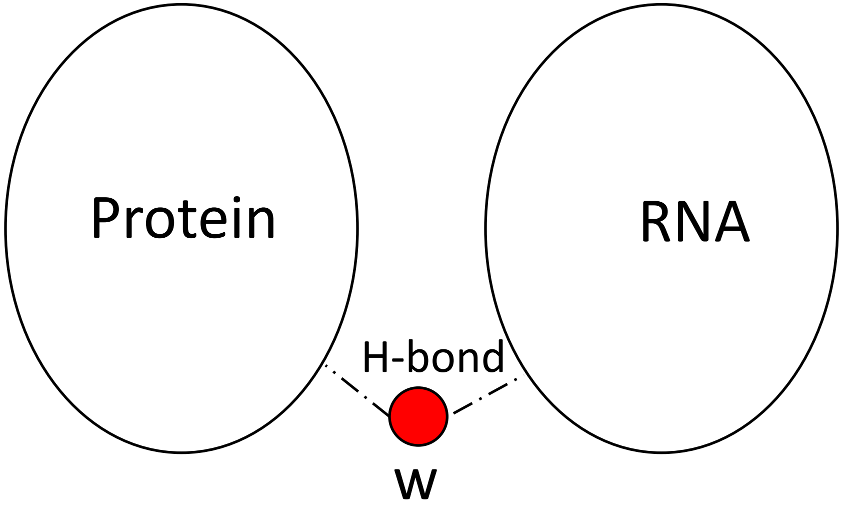
- 86 interfaces
- 2376 interface water molecules
- 4492 H-bonds
- One H-bond per 46 Å^2 of B

H-bonds distribution is biased: 42% with protein vs. 58% with RNA

Each protein-RNA interface : 52 water mediated H-bonds
20 direct H-bonds (Bahadur et al., 2008)

- Interfaces with ribosomal proteins have the highest H-bond density.
- Interfaces with tRNA have the lowest density.
- Interfaces with duplex RNA and single stranded RNA have intermediate density.

Bridging interface water molecules



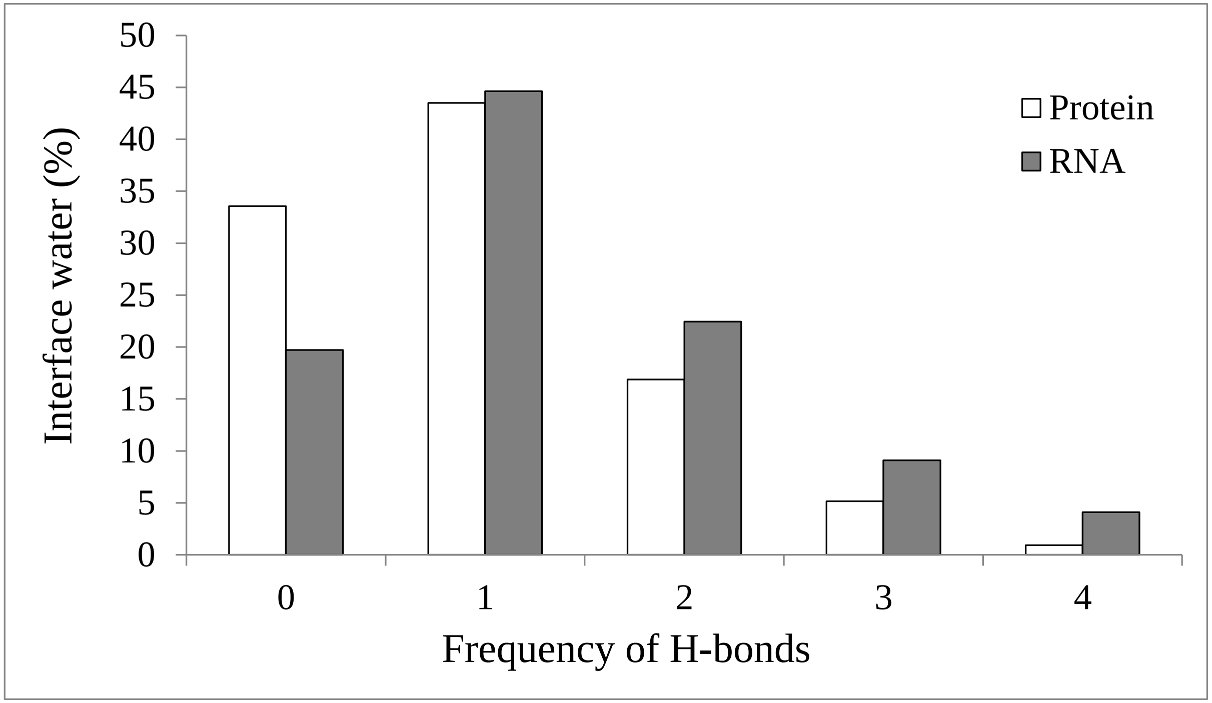
Water molecules making H-bonds with both sides of the interface

Protein-RNA interfaces contain 10 bridging water molecules

One third of the interface water molecules are bridging

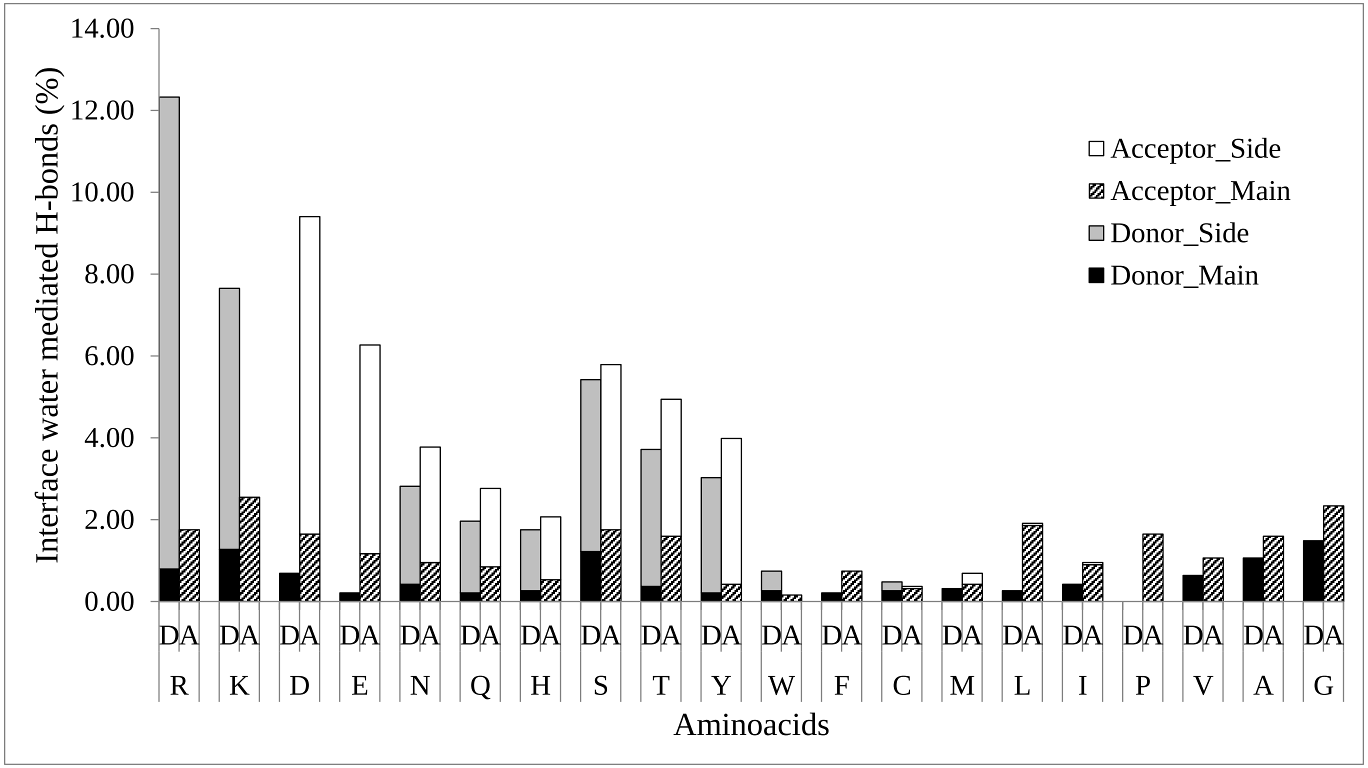
Interface statistics	protein-RNA				
	tRNA	ribosomal	duplex	single stranded	All
Bridging water molecules					
range	3-34	4-12	2-23	0-26	0-34
mean number per interface	15±9	8±3	12±6	8±6	10±7
per 1000 Å^2 of B	3.6	4.6	4.6	4.3	4.2

Water involving multiple H-bonds



Water molecules not involved in polar interaction mainly contribute to the van der Waals interaction.

Protein polar groups involving water mediated polar interactions



H-bond contribution

Neutral polar side chain (34%)

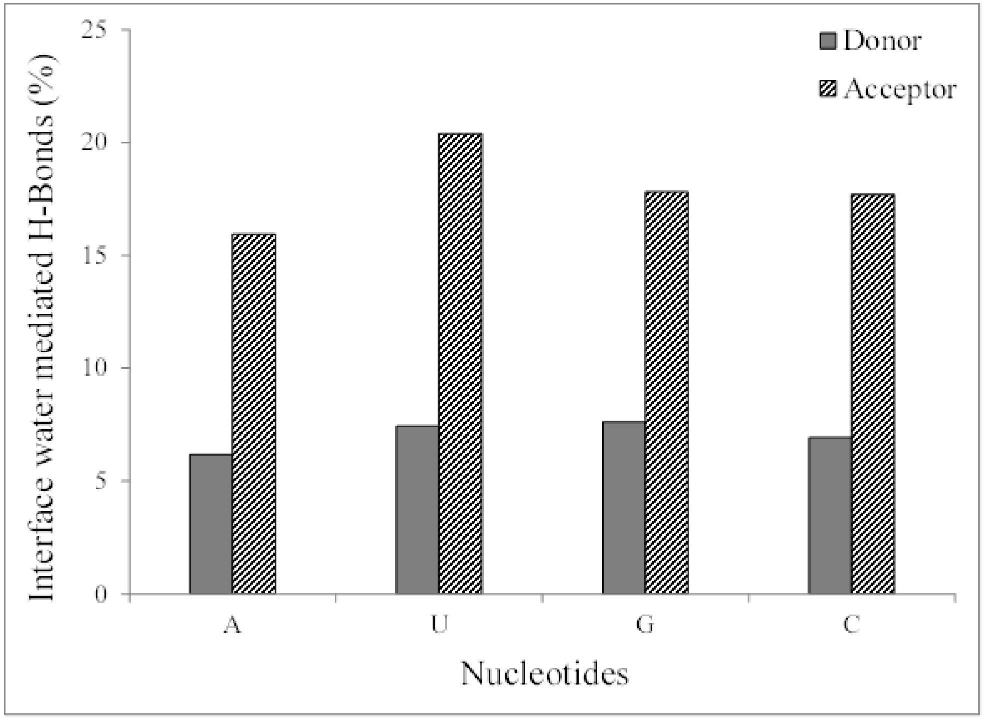
Charged side chains (31%)

Main chain O (24%)

Main chain N (11%)

	Donor (major)	Acceptor (major)
Charged side chain::	Arg, Lys	Asp, Glu
Neutral polar side chain::	Asn, Gln, Thr, Tyr	

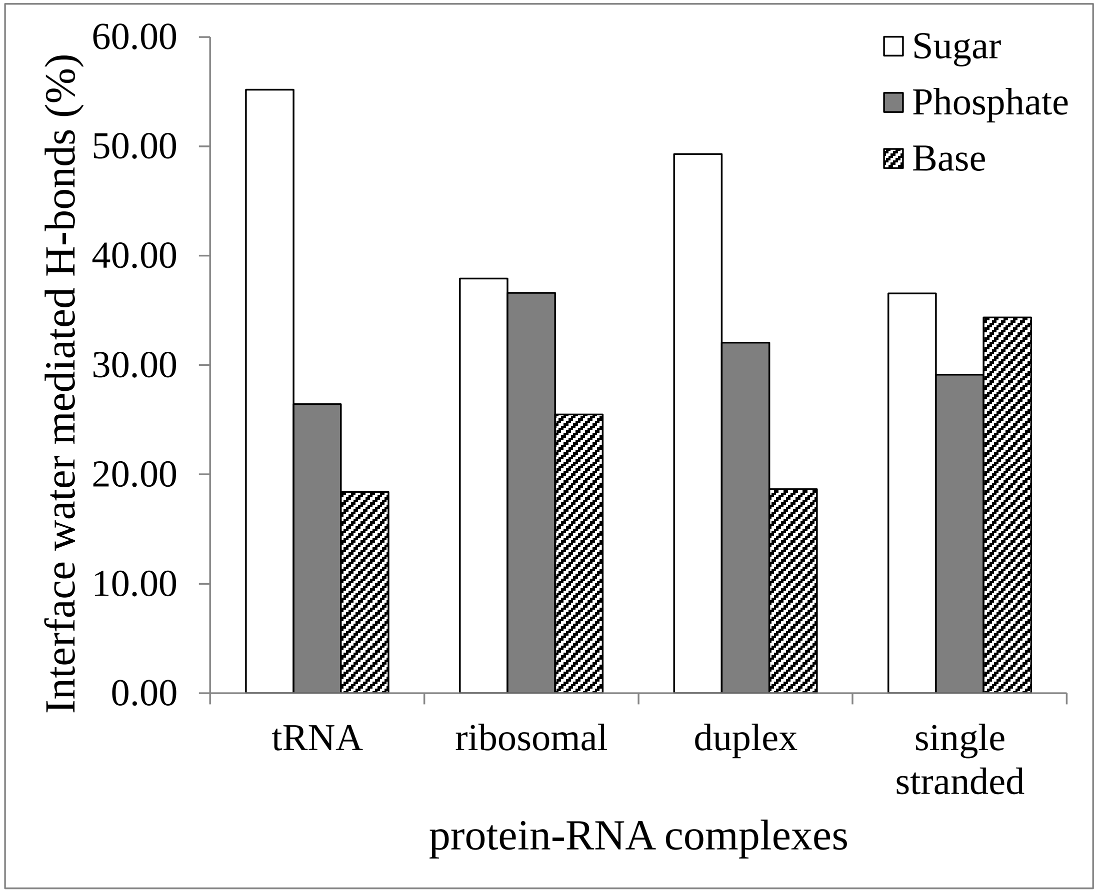
RNA groups involving water mediated polar interactions



H-bond contribution	
Ribose	(45%)
Phosphate	(30%)
Base	(25%)

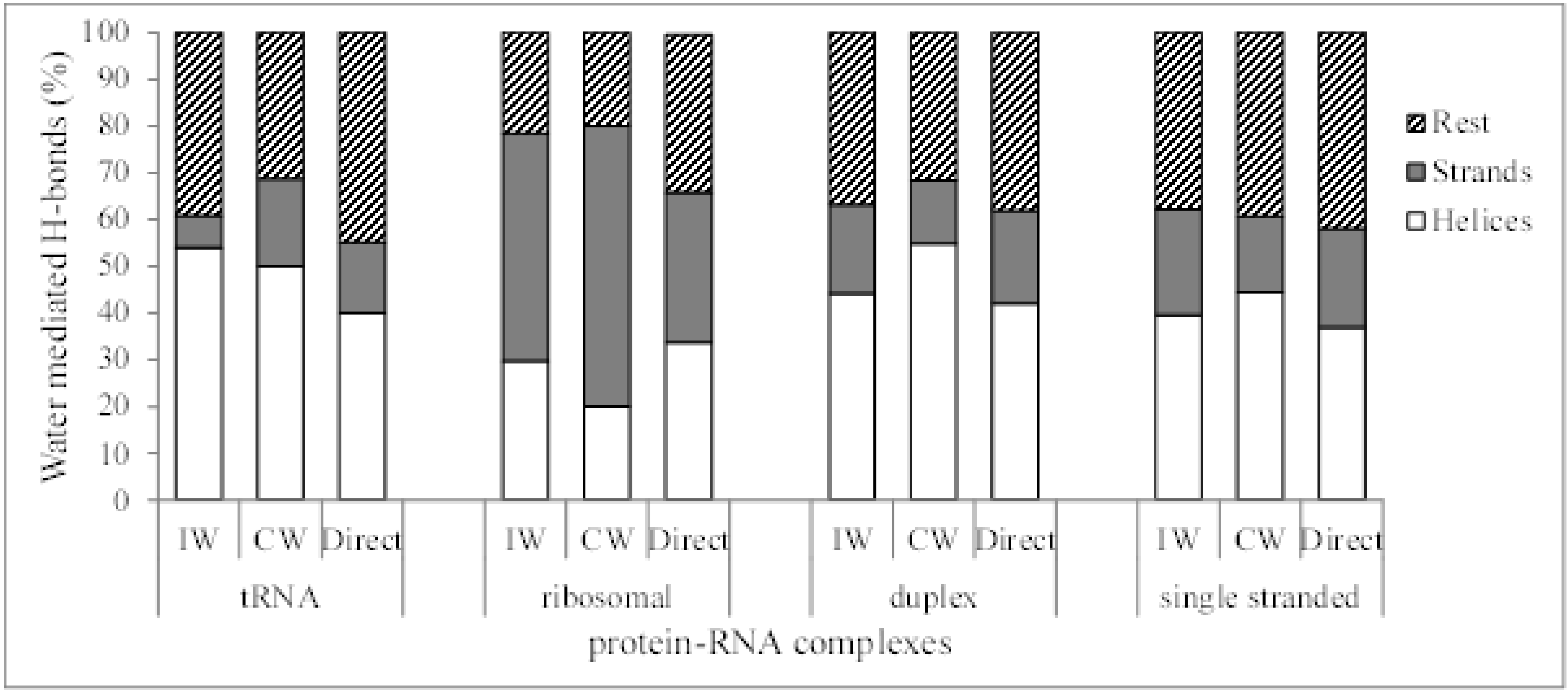
- Between different bases, guanine and uracil are frequently found in making H-bonds with interface water molecules than adenine and cytosine.
- 2’ OH involved in 31% of all interface water mediated H-bonds.

Water-RNA H-bonds in different structural classes

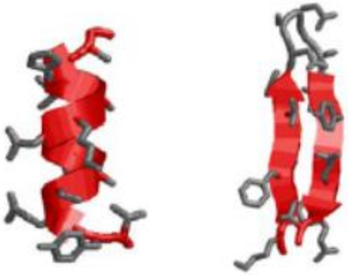


tRNA, duplex:: sugar-water > phosphate-water > base-water
ribosomal:: (sugar-water = phosphate-water) > base-water
single stranded:: (sugar-water = base-water) > phosphate-water

Hydration of the protein secondary structural elements



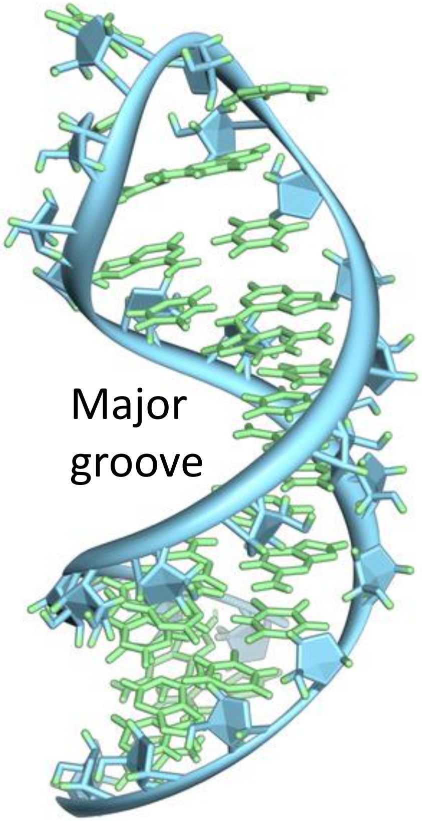
Helices---water:: 44%
Strands---water:: 19%
Nonregular---water:: 37%



While in complexes with tRNA, duplex RNA and single stranded RNA, interface water molecules prefer to form H-bonds with helices over strands, in complexes with ribosomal proteins, it is just opposite.

Interaction with major and minor groove

Complexes with duplex RNA (26)

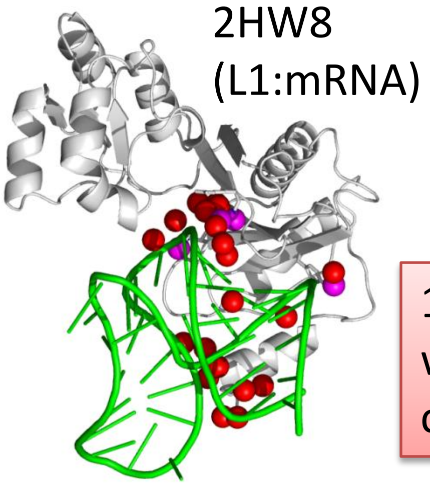


- Double helix of RNA resemble A form of DNA
- Minor groove in RNA is wide and shallow, whereas the major groove is narrow and deep.

Minor groove is more hydrated (49%) than major groove (37%) in protein-RNA complexes.

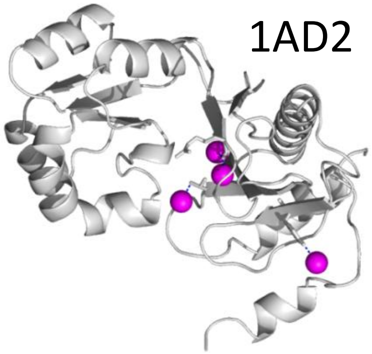
Conserved interface water molecules

An interface water molecule is defined as conserved water if it makes at least one same H-bond with the donor or acceptor atoms of protein or RNA chains in the bound and unbound structures.



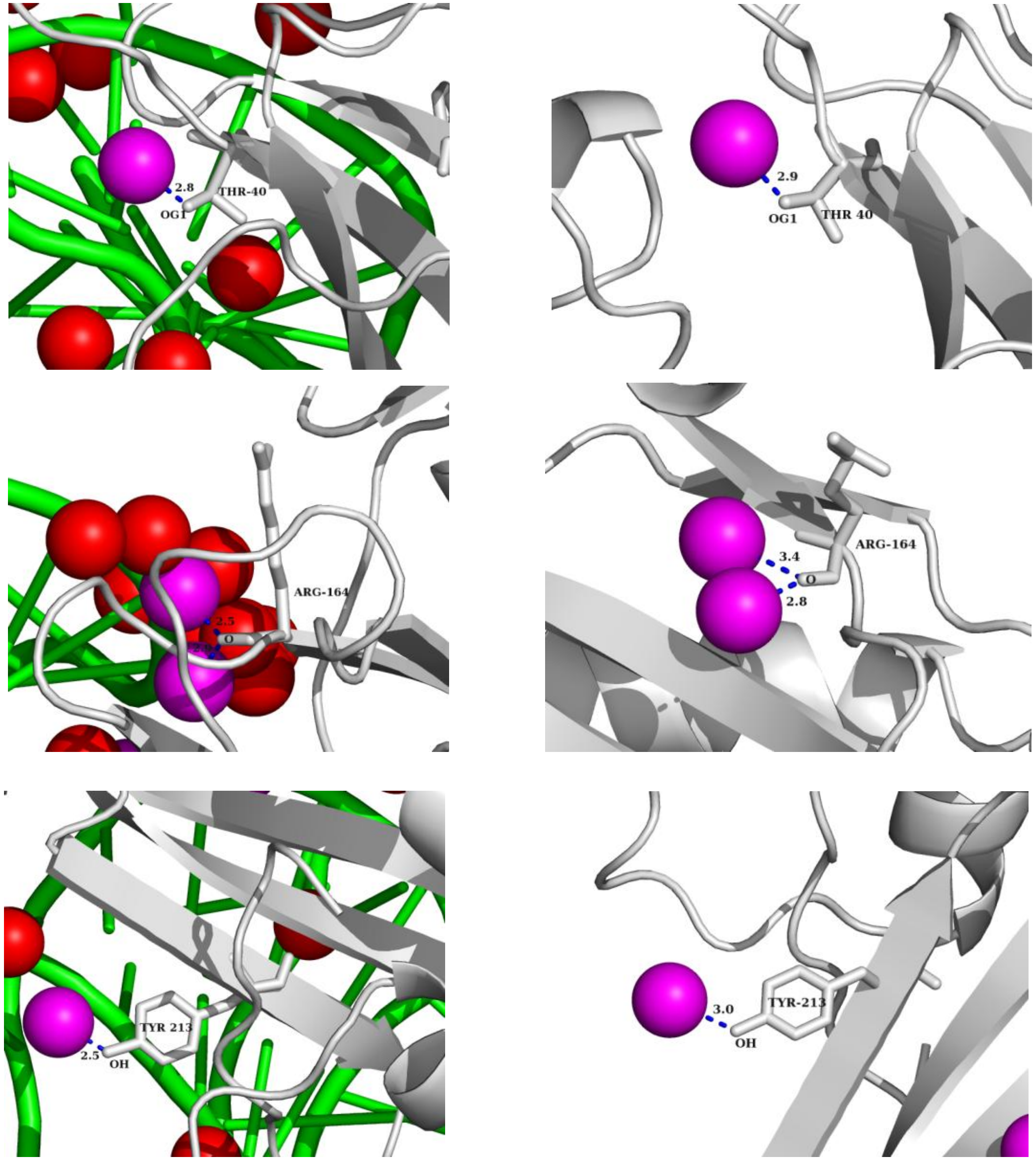
2HW8
(L1:mRNA)

11% interface water are conserved



1AD2

Conserved interface water makes more H-bonds than other interface water



Summery:: role of water in protein-RNA recognition

- Interface water prefers to make polar interaction with RNA than its partner protein.
- Interfaces with tRNA are more dehydrated compare to single stranded RNA or duplex RNA, whereas, interfaces with ribosomal proteins are least dehydrated.
- Interface water mediated H-bonds are numerous than direct protein-RNA H-bonds.
- One third of the interface water are involve to bridge the interface.
- Protein side, neutral polar side chains are most preferable to interact with water. RNA side, ribose is most preferred due to the presence of O2’.
- Helices are more frequent to interact with water than other sec. str. elements.
- In complexes with duplex RNA, minor groove is more hydrated than major groove.
- Conserved interface water molecules play an important role in the recognition. They make more H-bonds than other interface water molecules, and one should be careful about them while engineering protein-RNA interfaces.

Conclusion and future direction

- Protein RNA docking algorithm will improve incorporating interface water in their methodology.
- Will improve the prediction of binding affinity of protein-RNA complexes.
- Will help in drug design by knowing which interface water not to be displaced.
- Engineering protein-RNA interfaces.