Nucleic acids are bound by proteins.

How and why?

1. What features of RNA and DNA are recognized by proteins?

2. What are the principles of binding?

3. What are some common protein motifs used to bind DNA duplexes?

4. What are common motifs in RNA binding proteins?
1. What features of DNA duplexes are used by proteins to make specific contacts?

The arrangement of hydrogen bond donors and acceptors allows a protein to distinguish among AT, TA, CG, and GC in the major groove.

In the minor groove, only AT and GC pairs can be discriminated.
Nucleic acids are extensively hydrated, and the concept that there is a network of ordered water molecules held in place through hydrogen bonding to bases and phosphates is generally accepted. The “spine of hydration” was thought to occur in the minor groove of B-form DNA based on early crystal structures of the Dickerson dodecamer.

Shui et al. revisited that study, and concluded that many of those waters were in fact ions (Na+) that constituted the first layer of the spine. The waters were in a second ordered layer.
Proteins must compete with ions to bind nucleic acids.

The B-DNA dodecamer at high resolution reveals a spine of water on sodium and spermidine.

2. What are the principles of Protein:nucleic acid binding?

a) Coulombic interactions (with consequential ion release)
b) van der Waals (dipole-dipole and induced dipole)
c) Solvent driven (hydrophobic effect)
d) Hydrogen bonding

These interactions will be highly dependent on solution conditions of temperature, salt concentrations, and pH.

These conditions must be explicitly stated in any description of protein binding to RNA or DNA!!!!!

What are the principles of protein:protein binding?
Protein + NA \rightleftharpoons \text{Protein:NA}

K_{obs} = \frac{[PN]}{[P][N]}

This is too simple, since DNA and RNA are polyanions and bind counterions. Logically, since Protein binds a nucleic acid, it must also ‘bind’ anions. When the nucleic acid binds protein, it must release its counterions and waters from sites that will interact with protein (vice versa for the protein).

\[
P(aM^+, bX^-, cH^+, dH_2O) + N(eM^+, fH_2O) \rightleftharpoons PD(gM^+, hX^-, jH_2O)
\]

The more accurate equilibrium reaction is

\[
P + N \rightleftharpoons PN + xM^+ (x = g-(e+a))
\]

so increasing the concentration of M\(^+\) will shift the equilibrium to the left (free P and N).
b) van der Waals (dipole-dipole and induced dipole)

London dispersion forces are weak interactions that are typically induced-dipole.

$$E_{vdw} \approx \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}$$
d) Hydrogen bonding

**Recognition of a specific site** is often described in terms of ‘direct readout’ – amino acids of the protein ‘recognize’ the 3D arrangement of hydrogen bond donors and acceptors on the nucleic acid. ‘Indirect readout’ – the protein recognizes conformational features of the nucleic acid.

Hydrogen bonding is the most common devise to obtain specificity of interactions, since hydrogen bonding has preferences for length and bond angle.

But note: hydrogens are not visible in crystal structures. They are added to structures and the presence of hydrogen bonding is modeled by geometry and proximity.

And: many specific interactions are mediated by water molecules (not necessarily visible in crystal structures).
**Energetically**, residues that are involved in intermolecular hydrogen bonding are often hydrogen bonded to water in the free state, so there is not a large energy gain in formation of the protein:nucleic acid hydrogen bond (about -1.1 to -1.7 kcal/mol H-bond).

But, if a hydrogen bond to water is not replaced by an equivalent hydrogen bond, then there is an energy loss associated with complex formation.

**Specificity** due to hydrogen bonding is more related to losing a hydrogen bond than forming one, although the opposing effects are often impossible to separate.
Essential features for modulating the binding of a protein to a nucleic acid are:
1> Reversible binding.
2> Competitive binding.
   The same protein for different sites or many proteins for the same site.
3> Modulation of binding affinity and specificity by small effector ligands
4> Competition between different protein subunits.

Binding can be modulated in two ways:
1> Thermodynamic, or equilibrium control.
   In this case, regulation is achieved by equilibrium binding affinities of various proteins for their DNA/RNA sites, and so the percent site occupancy by a given protein is the key.

2> Kinetic control.
   The rates of complex formation or dissociation are most important.

To describe complex formation, it is necessary to know the binding affinity and rates of binding and dissociation.

In practical terms, in order to understand regulation by a protein:nucleic acid interaction, it is necessary to know the binding mechanism.
3. What are some common protein motifs used to bind DNA duplexes?

The helix-turn-helix motif binds to DNA.

These aren’t stable out of the context of the whole protein.
Zinc finger specificity can be modulated. (Talens)

Three tandem fingers bind to DNA. What’s the advantage of having more than one finger?
Genome editing with engineered zinc finger nucleases

Urnov et al., 2010, Nature reviews/genetics 11:636-646.
Leucine zippers themselves can bind DNA (fos/jun, GCN4, bZIP)

E. coli Single Strand Binding protein (Lohman lab, WUMS)

EcoR1 restriction enzyme + DNA (Rosenberg lab, U Pitt)
First, The Arginine-rich motif (ARM)
These ‘motifs’ are intrinsically disordered regions!
Most peptides can fit into these major grooves where there is a dramatic deformation of the A-form duplex by the bulged nucleotides.

Due to the number of interactions (hydrogen bonds, electrostatics, and stacking hydrophobic amino acids), the dissociation constants of these small complexes can be nanomolar.
The point is: there is not one unique way for an ARM peptide to make specific contact with an RNA. It is almost impossible to model. And remember: in the context of the proteins, the peptides are intrinsically disordered, so binding to RNA is accompanied by a conformational change.

A is P22/BoxB. The α-helix fits into the groove of the RNA and bends over the GAAAA loop where its Trp stacks with an adenosine. The peptide bends at R11 to allow the helical sidechains to stack with the nucleobases.

B is HIV Rev/RRE. The REV peptide forms an α-helix, but positions itself in the RNA bulge. The peptide contacts both RNA strands.

C is BIV Tat/TAR. The Tat peptide forms a β hairpin as it positions itself in the RNA bulge. The peptide contacts both RNA strands.
2. dsRBD

Double-stranded RNA binding domains (dsRBMs) are nonspecific, but sensitive to A-form structure. This is the domain from PKR (Protein Kinase R).

The affinity comes from many contacts between the protein and 2’ OH groups in the minor groove. If a DNA were A-form, these contacts would be missing, but the protein could still make electrostatic contacts with the phosphates. dsRBD binding is structure-selective, but not sequence-specific.

Other very important proteins have this motif:
ADAR1, the RNA-specific adenosine deaminase that converts adenosine to inosine in duplexes contains three dsRBDS.
DCR (Dicer) is the enzyme that cleaves double-stranded RNAs into 21 base-pair pieces. These small duplex RNAs go on to become incorporated into the RISC, where they are bound by Ago and become the templates for RNAi cleavage of mRNAs. DCR has one dsRBD.
Argonaute (Ago) proteins have two dsRBDS. They bind to miRNA and siRNA as part of the process of gene regulation by translation repression (the current model for miRNA activity) or mRNA degradation (RNAi).
3. RNA Recognition Motif (RRM). It is the most common eukaryotic RNA binding domain. RRM's are identified by their conserved sequences

<table>
<thead>
<tr>
<th></th>
<th>RNP-2</th>
<th>RNP-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus</td>
<td>LFVGNL</td>
<td>KGFGFVXF</td>
</tr>
<tr>
<td></td>
<td>IY I KL</td>
<td>R YA Y</td>
</tr>
</tbody>
</table>

Two or three aromatic residues are solvent-exposed on the surface of the \(\beta\) sheet.

While nucleobases are not hydrophobic, the aromatic amino acids in RNP1 and RNP2 are hydrophobic, and do not ‘like’ to be exposed to solvent. They will stack with nucleobases to remove themselves from water.
4. KH domains use amino acids on their α-helical side to bind RNA.
Many RNA binding proteins use multiple domains to interact with different RNA structures.
The **RecA motif** with its ATP binding site is used in RNA helicases.
NOW: RNA is an ENZYME!

Group I and Group II introns, the ribosome, the spliceosome, and RNase P are all ribozymes; and then there are small RNAs that cleave themselves that are also ribozymes.

Five nucleolytic ribozymes

**FIGURE 1.** Secondary structures of the five nucleolytic ribozymes: (A) hammerhead; (B) hairpin; (C) glmS; (D) HDV; (E) VS. Key nucleotides are
Hairpin Ribozyme

Hammerhead Ribozyme

glmS Ribozyme-riboswitch

Hepatitis Delta Virus (HDV) Ribozyme

Residues thought to be part of general acid/base catalysis are green/red. Scissile phosphate and 2’OH are magenta. Substrates in light blue.
In **RNase A**, the unprotonated Histidine 12 acts as a general base catalyst, extracting the H from the 2’ OH, creating a reactive Oxygen that attacks the phosphate PO$_4^-$ and produces a 2’,3’ cyclic phosphate.

The protonated Histidine 119 acts as a general acid and donates its H to the 5’ OCH$_2$ leaving group.

The ammonium group of Lys41 stabilizes the excess negative charge of the trigonal bipyramidal oxyphosphorane transition-state.

Hairpin, hammerhead, and *glmS* ribozymes use the same acid/base chemistry.

We will ask: Where are the acid and the base?
The Hammerhead can be removed from the viral genome and studied on its own. Uhlenbeck [Nature 1987 328:596] showed that it could be engineered to become a true enzyme.
The hammerhead RNA can be synthesized in two pieces.

The top strand O2 is the “substrate”

The bottom strand O1 is the “enzyme”

Add equimolar amounts of O1 and O2.
Incubate in 50 mM Tris, pH 7.5, 10 mM MgCl₂, 37°C, 1 hr
85% of O2 “substrate” is cleaved to give two products.

The O1 “enzyme” is unchanged.
The products were characterized and found to have a 5′ OH on the short product $\text{OH}_{\text{short}}\text{UCGAGC}_{\text{OH}}$ and a 2′,3′ cyclic phosphate on the long product.
What features of the ribozyme are critical for its activity?

Look for mutants. First, identify conservation and variation, using phylogenetic comparison.

The Hammerhead consensus sequence. There is a conserved core with variable arms. N17 can be C or A in this context.
Each nucleotide in the central core was replaced. The mutant enzyme or substrate was combined with their respective wild type substrate/enzyme, and assayed for cleavage. The numbers are the relative cleavage rates with respect to wild-type enzyme/substrate. NC = no cleavage after 2 hr

Conclusion: No mutations in the core are tolerated (Ruffner et al., (1990) Biochem 29:10695-10702). (except here)
There are two G:A interactions and one C:C interaction, in addition to an abnormal A:U pair.
The original hammerhead enzyme had very short stems. It was later understood that it lacked stabilizing tertiary interactions.

![Diagram of the hammerhead enzyme](image)

**Fig. 2.** The structure of the hammerhead ribozyme. (a) shows schematic diagrams of the secondary structures of the minimal and full-length hammerheads, respectively, and (b) shows an all-atom representation of the tertiary structure of the full-length hammerhead. A distal contact between Stems I and II in the full-length hammerhead, shown schematically in (a), stabilizes the active site structure, the details of which are shown in (c). (d) shows a proposed transition state structure extrapolated from (c).
A schematic pathway of the hammerhead reaction illustrates both how a conformational transition positions the active site and how it might allow for the reverse ligation reaction. In the inactive, uncleaved state, a mixture of structures could be present; an even more heterogeneous mixture could be present in the inactive cleaved state.

If the RNA persists for some time (how long?) in its active cleaved state, it could re-ligate the product since this is a reversible transesterification reaction. If the tertiary interaction is too strong, then ligation would also be favored. If the product:enzyme helix is too long, product release would be dis-favored and so lead to ligation. This is a dynamic process!
G12 aromatic N hydrogen bonds to the nucleophilic 2′-OH. G12 is invariant in HH ribozymes, where it serves as the general base in the cleavage reaction. The ribose 2′ OH of G8 makes a hydrogen bond to the 5′-oxo leaving group (it acts as the general acid).
The **Hairpin Ribozyme** was discovered in a viroid RNA where it undergoes self-cleavage to produce a single copy of the RNA strand.

**Figure 7:** Minimum reaction sequence for hairpin ligation, consisting of substrate binding, the formation of a docked state requiring specific interactions between residues in loops A and B (dashed lines), and ligation between G(+1) and N(-1). All steps are reversible, with the equilibrium dependent on the energetics of substrate binding, the stability of the docked state, and the ribozyme-catalyzed equilibrium between ligation and cleavage.
Engineer a three-piece RNA. Tether one strand to glass via biotin/streptavidin.

On one end of the enzyme strand, add a donor and on the other end an acceptor FRET pair.
Single Molecule study of the Hairpin Ribozyme.

dA38 accelerates undocking, since there is a loss of a hydrogen bond. But the rate and extent of cleavage was preserved.
A38 cannot serve as the general acid when it is replaced by 2F-A or n⁸A, providing chemical evidence (in contrast to crystal evidence) that it functions like a histidine in the cleavage/ligation reaction.
The Hepatitis Delta Virus ribozyme requires Mg$^{2+}$ for cleavage.
C75 is particularly interesting. It serves as the general acid catalyst in the cleavage reaction.

In the crystal structure of the HDV ribozyme in its postcleavage state, C75 is proposed to make a hydrogen bond between its N3 atom and the protonated leaving group, the 5′-hydroxyl of G1. This conformation suggested that C75 might be protonated in the reactant state, donate its proton to the 5′-O leaving group, and thereby serve as a general acid in the cleavage reaction.

Removing the C75 base (creating an abasic site) inactivates HDV, but it can be rescued by addition of exogenous imidazole!
In the hairpin (A), hammerhead (B), and glmS (C) ribozymes, the substrate is bound as part of a deformed helix that docks with other structures in the enzyme. The backbone is stretched to align it with residues that cleave it.

In the hairpin (A), G8 and A38 play the part of the histidines in RNase A: G8 is the base (pKa = 9.0), and A38 the acid (pKa=5.0). A9 is in the position of Lys41 where the exocyclic amines of A9 and A38 provide electrostatic stabilization of the transition state.

Figure 4. Active sites of the (A) hairpin, (B) hammerhead, (C) glmS, and (D) HDV ribozymes color-coding as in Figure 3. The moieties thought to contribute to catalysis are labeled.
Conclusion:

These are small ribozymes, but they have evolved different global structures and sequences. In particular, three of the four use the same catalytic mechanism in different active sites.

The chemical contexts of the active sites allows bases and riboses to act as general acids/bases, which involves local changes of the pKa.

The structures flex: the active sites are in junction regions that are formed by docking of duplex regions through tertiary interactions. Docking is not stable, so the conformations fluctuate between an open inactive form and a closed active form.

In two of the ribozymes, the reactions can go forward and backward (cleavage and ligation), since product release can be delayed by salt and temperature.

Hammerhead ribozymes appear in all kingdoms in unexpected places!
This x-ray crystal structure of a bacterial large ribosomal subunit shows the side that sits on the small subunit. The RNA is white, proteins are gold. Green is a puromycin bound at the active site where catalysis occurs. The nearest protein is 18 Å away from the catalytic center.

Ribosomes are RNA enzymes – the RNA does the chemistry.