Nucleic acids are bound by proteins. What are some common protein motifs used to bind DNA duplexes and RNA?
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.
The “Spine of Hydration”

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$^+$

Reducing the dimensionality of the binding site challenge
Protein:NA recognition mechanisms.
1) Coulombic interactions (with consequential ion release)
2) van der Waals (dipole-dipole and induced dipole)
3) Solvent driven (hydrophobic effect)
4) 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.
1) Coulombic interactions (with consequential ion release)

Protein + DNA <=> Protein:DNA

\[ K_{obs} = \frac{[PD]}{[P][D]} \]

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).

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

So a more accurate equilibrium reaction is

\[
P + D \rightleftharpoons PD + xM^+ \quad (x = g-(e+a))\]

so increasing the concentration of M\(^+\) will shift the equilibrium to the left (free P and D).
Most protein:DNA interactions also show anion effects (which almost always means that there is an interaction of the protein with the anion, in addition to cation binding to the DNA).

\[
\frac{\partial \log K_{\text{obs}}}{\partial \log [\text{MX}]} = \Delta c + \Delta a
\]

\(\Delta c = \) differential cation binding  
\(\Delta a = \) differential anion binding

so, the slope of a plot of \(\log K_{\text{obs}}\) vs \(\log [\text{MX}]\) provides a measure of the NET number of ions (cations + anions) released or taken up upon formation of PD.

if \(\frac{\partial \log K_{\text{obs}}}{\partial \log [\text{MX}]} < 0\), then there is a net release of ions

if \(\frac{\partial \log K_{\text{obs}}}{\partial \log [\text{MX}]} > 0\), then there is a net uptake of ions
2) van der Waals (dipole-dipole and induced dipole)

London dispersion forces are weak interactions that are typically induced-dipole.
4. 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.

However, it is not sufficient to consider direct interactions between protein and nucleic acid since 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.
What proteins bind to DNA and RNA?

The helix-turn-helix motif binds to DNA. These aren’t stable out of the context of the whole protein.

The sequence-specific contacts between the helix and the bases can be manipulated, based on the known orientation of helix3 with respect to the DNA. [Wharton & Ptashne (1985) Changing the binding specificity of a repressor by redesigning an α-helix. Nature 316:601.]
Zinc finger specificity can be modulated. (Talons)

Three tandem fingers bind to DNA. What’s the advantage of having more than one finger?
Leucine zippers themselves can bind DNA (fos/jun, GCN4, bZIP)

EcoR1 restriction enzyme + DNA (Rosenberg lab, U Pitt)

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

Complex DNA binding proteins
Hexameric helicases function in DNA recombination and replication. They unzip duplex DNA to repair a lesion or to begin DNA replication. Their binding sites can be single-stranded DNA tails or internal regions, single strand breaks, ... They have DNA-dependent ATP hydrolysis activity that is necessary for them to unzip the double strand.
Proteins that bind RNA.

1. The Arginine-rich motif (ARM)

<table>
<thead>
<tr>
<th>Peptide–RNA complex name</th>
<th>Peptide/RNA sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIV Tat–TAR</td>
<td>Tat peptide: SGPRP RGTRG KGRRI RRR</td>
</tr>
<tr>
<td></td>
<td>TAR RNA: 5' GCCUCG G AGCU C A 3'</td>
</tr>
<tr>
<td></td>
<td>5' CCGAGC C UCGA U U 3'</td>
</tr>
<tr>
<td>HIV–1 Rev peptide–RRE RNA</td>
<td>Rev peptide: DTRQA RRNRR RRWRE RQRAA AAR</td>
</tr>
<tr>
<td></td>
<td>RRE RNA: 5' GGUCUGG GCG CAGC G C 3'</td>
</tr>
<tr>
<td></td>
<td>5' CC AGACA GGC ,GUCG A A 79</td>
</tr>
<tr>
<td>HIV–1 Rev peptide–RRE aptamer</td>
<td>Rev peptide: TRQAR RNRRR RWRE RQRAA AAR</td>
</tr>
<tr>
<td></td>
<td>RRE aptamer: 5' GGUCUG GCG CAGC G C 3'</td>
</tr>
<tr>
<td></td>
<td>5' CCGA C GAG CAUG G C</td>
</tr>
<tr>
<td></td>
<td>35 A A G U</td>
</tr>
<tr>
<td>λ N peptide–boxB RNA</td>
<td>λ N Peptide: MDAQT RRRER RAEKQ AQWKA AN</td>
</tr>
<tr>
<td></td>
<td>λ boxB RNA: 5' GGGCCUG A 3'</td>
</tr>
<tr>
<td></td>
<td>3' CCCGGGA A G</td>
</tr>
<tr>
<td>P22 N peptide–boxB RNA</td>
<td>P22 N Peptide: NAKTR RHERR RKLAI ERDTI</td>
</tr>
<tr>
<td></td>
<td>P22 boxB RNA: 5' GCGCU G A 3'</td>
</tr>
<tr>
<td></td>
<td>3' CCGCA A A 18</td>
</tr>
</tbody>
</table>
Binding of the Tat protein ARM: ITKALQISYRKRRGRRAPQGSQTHQVSL can be reduced to binding of a single R.

Figure 6. The average structure of the TAR-argininamide complex is shown with U23 highlighted in green, G26 in purple, A22 in yellow and the argininamide in red. The ribbon traces the backbone showing the distortion in the bulge region.
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.

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.
Double-stranded RNA binding domains (dsRBM) 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>LFVGLNL</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 β 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 $\alpha$-helical side to bind RNA.

Many RNA binding proteins use multiple domains to interact with different RNA structures.
Fig. 2 Evolutionary representation of ribosomal proteins of the 60S subunit. 60S large ribosomal subunit proteins are colored according to conservation.

Sebastian Klinge et al. Science 2011;334:941-948

Published by AAAS
The **RecA motif** with its ATP binding site is used in RNA helicases.

RNA helicase proteins as chaperones and remodelers
Helicases function as RNA chaperones for folding and remodelers for misfolding.

DEAH box family unwind RNA helices via translocation, but few have been carefully studied to understand their processivity.

DEAD box family do not translocate. They remodel internal duplexes.

In prokaryotes, these proteins are dedicated to rRNA folding and ribosome assembly.

In eukaryotes, they function in the spliceosome and the ribosome, and to fold rRNA and snRNAs.

Can the DEXH proteins displace bound proteins? Do they promote RNA annealing? Can they disrupt noncanonical RNA structures?