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?**

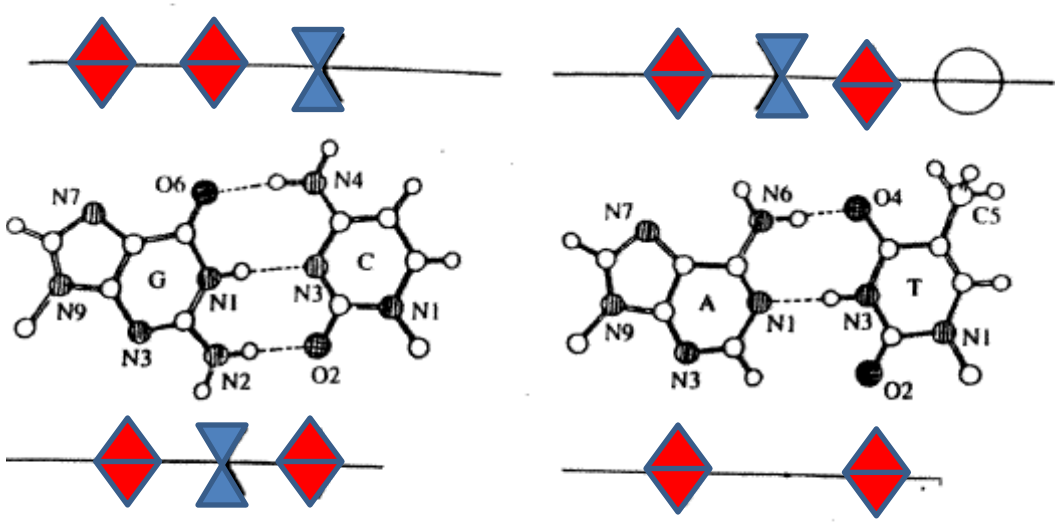
3. **What are some common protein motifs used to bind DNA duplexes?**

4. **What are common motifs in RNA binding proteins?**

1. What features of DNA duplexes are used by proteins to make specific contacts?

Major Groove

Duplex DNA



Minor Groove

The arrangement of hydrogen bond donors  and acceptors 

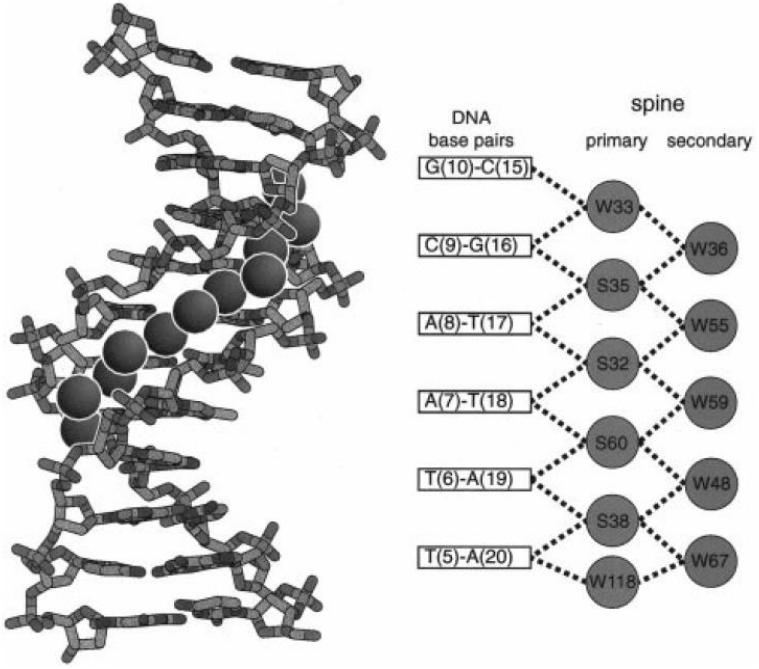
allows a protein to distinguish among AT, TA, CG, and GC in the major groove.

In the minor groove, only AT and GC pairs can be discriminated.

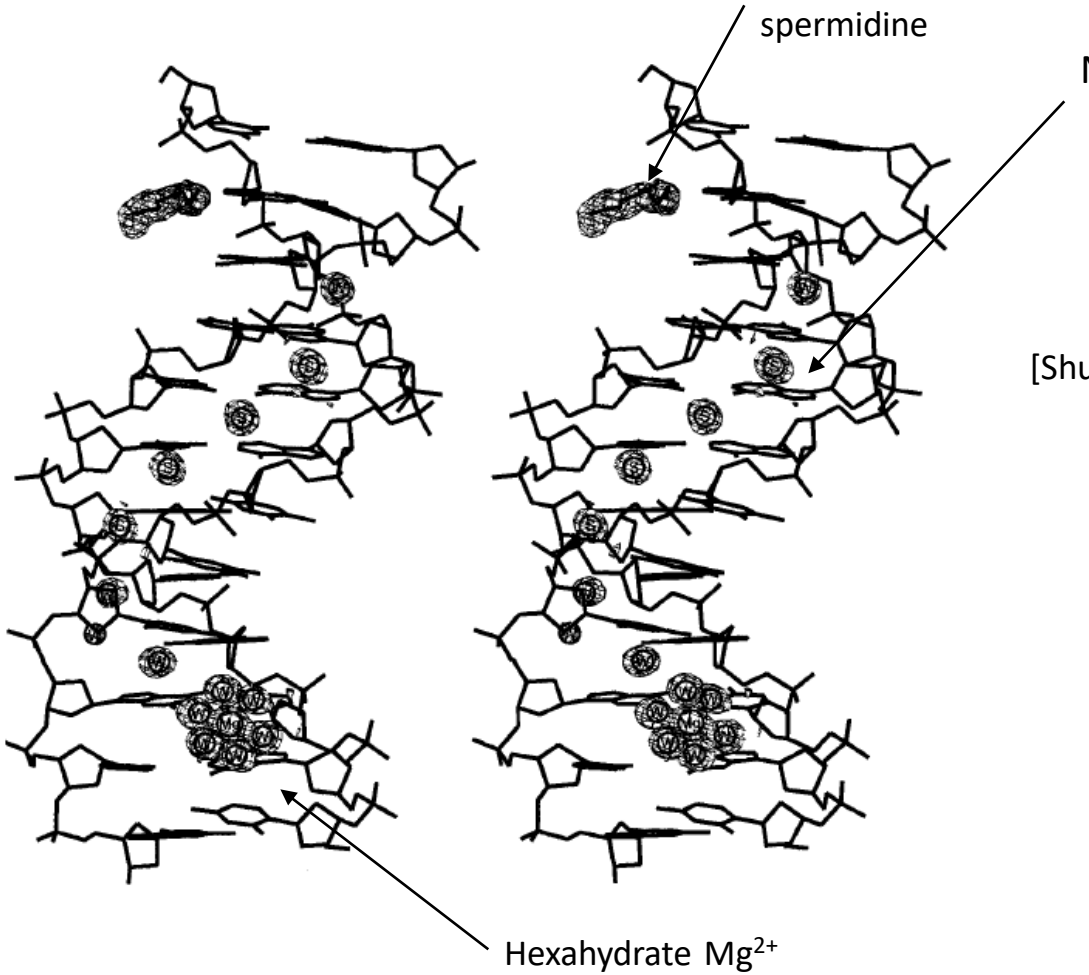
The “Spine of Hydration”

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

[Shui et al (Williams lab) 1998. Biochemistry 37:8341]

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

- Coulombic interactions (with consequential ion release)
- van der Waals (dipole-dipole and induced dipole)
- Solvent driven (hydrophobic effect)
- 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?

a) Coulombic interactions (with consequential ion release)

Protein + NA <=> Protein:NA

K_{obs} = [PNA]/[P][NA]

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).

$P(aM^+, bX^-, cH^+, dH_2O) + NA(eM^+, fH_2O) <=> PNA(gM^+, hX^-, jH_2O)$

So a more accurate equilibrium reaction is

$P + NA <=> PNA + xM^+$ (x = g-(e+a))

And therefore increasing the concentration of M^+ will shift the equilibrium to the left (free P and NA)

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).

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

Δc = differential cation binding
Δa = differential anion binding

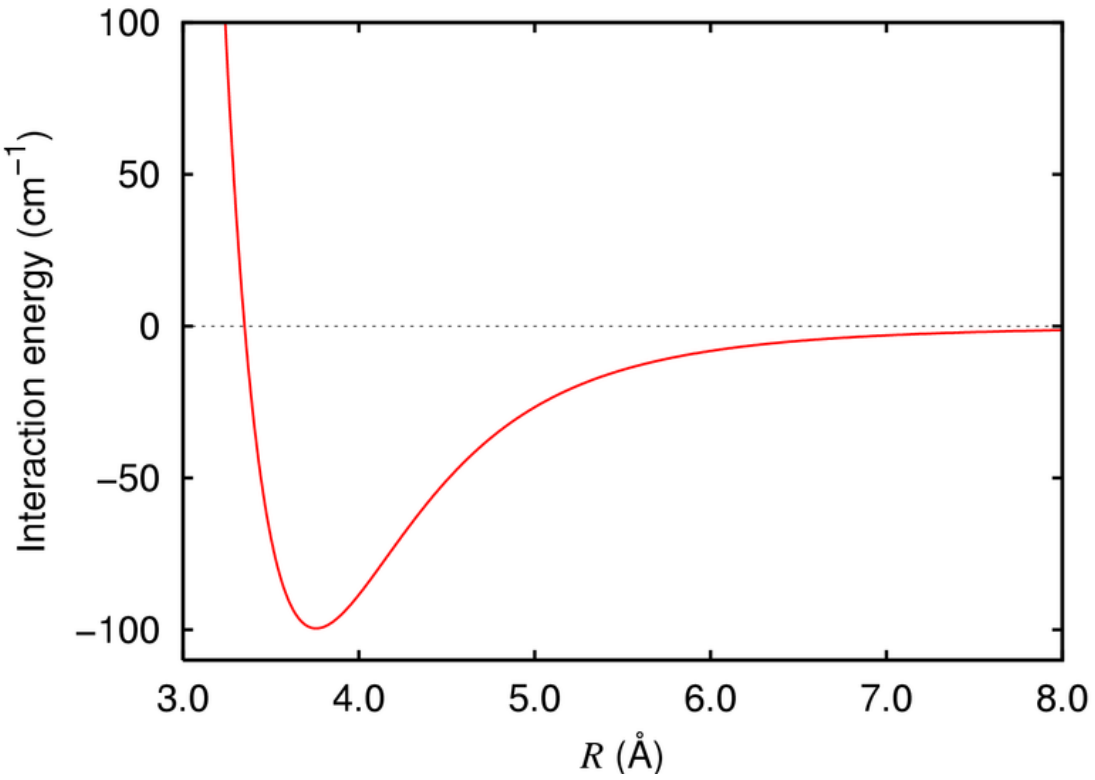
so, the slope of a plot of $\log K_{obs}$ vs log[MX] provides a measure of the NET number of ions (cations + anions) released or taken up upon formation of PD.

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

if $\frac{\partial \log K_{obs}}{\partial \log [MX]} > 0$, then there is a net uptake of ions

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).

d) Hydrogen bonding

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.

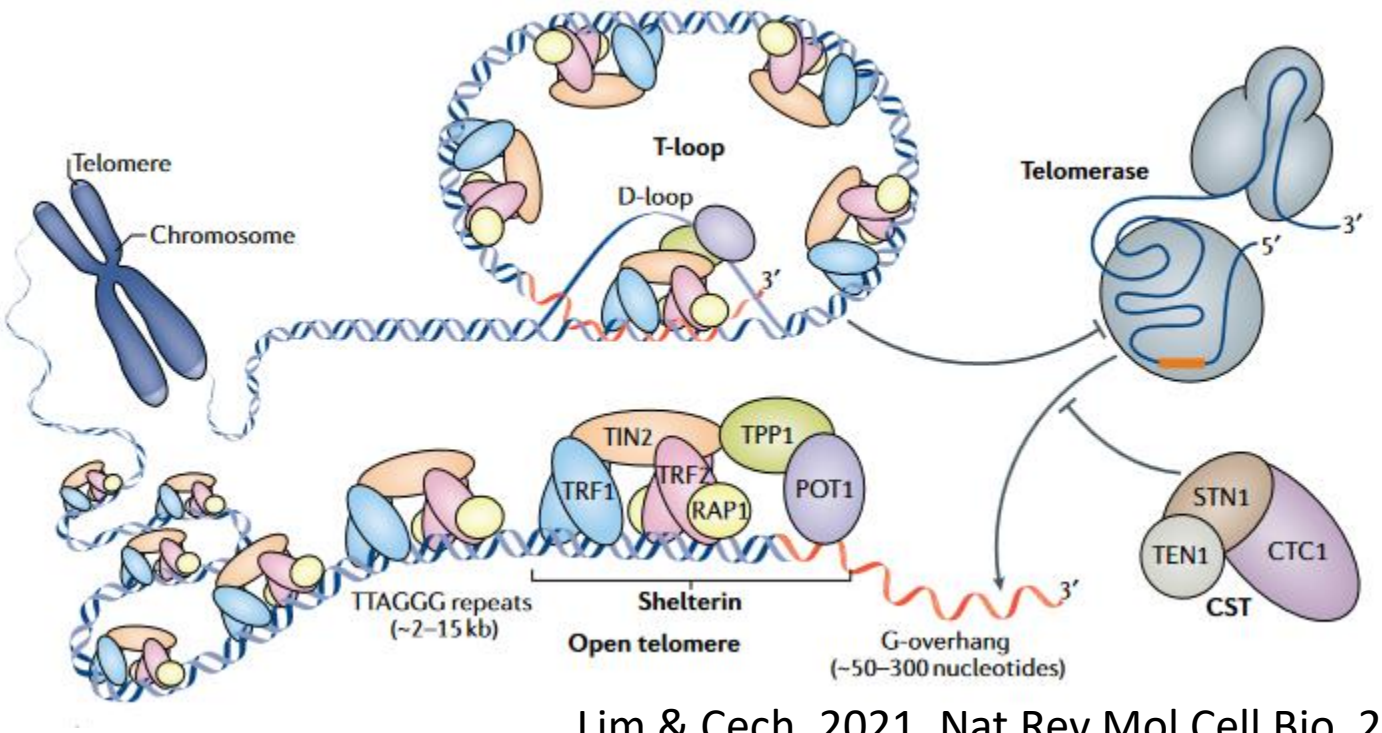
To describe complex formation, it is necessary to know the binding affinity and rates of binding and dissociation.

In practical terms, in order to understand regulation by a protein:nucleic acid interaction, it is necessary to know the binding mechanism.

An arbitrary example, but think of general principles.

TELOMERES

Telomeres are at the ends of linear chromosomes.



Lim & Cech, 2021. Nat Rev Mol Cell Bio. 22:283-94

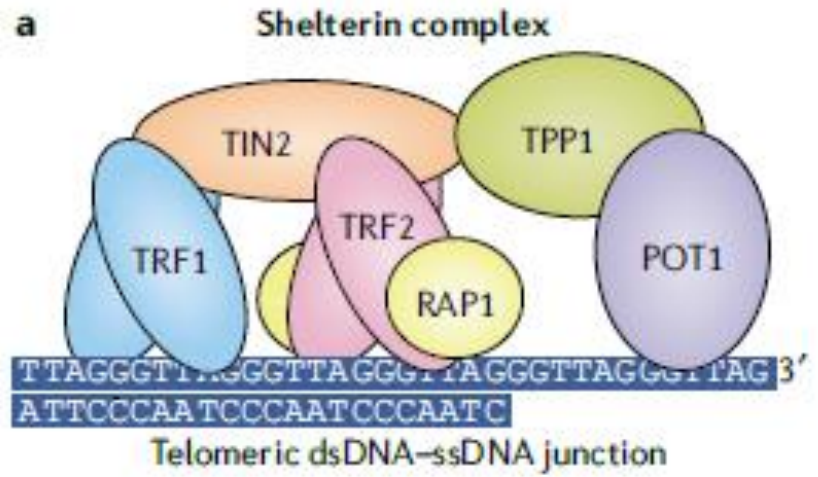
Telomeric DNA in humans is dsDNA $(TTAGGG)_n$ extending from 7-15kb [5'TTAGGG/3'AATCCC]

Shelterin complex contains up to six proteins – it's not a static entity

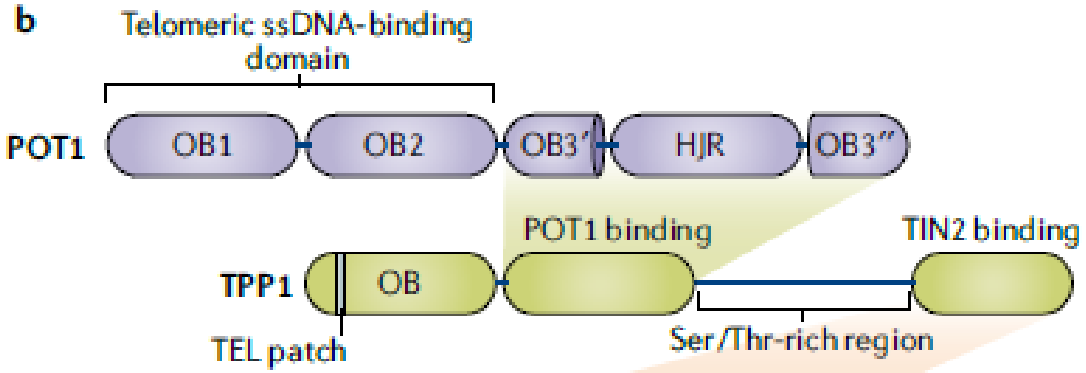
CST: terminates the repeat and recruits DNA polymerase α

Telomerase: RNA + Proteins templates the telomeric repeat and adds the dNTPs

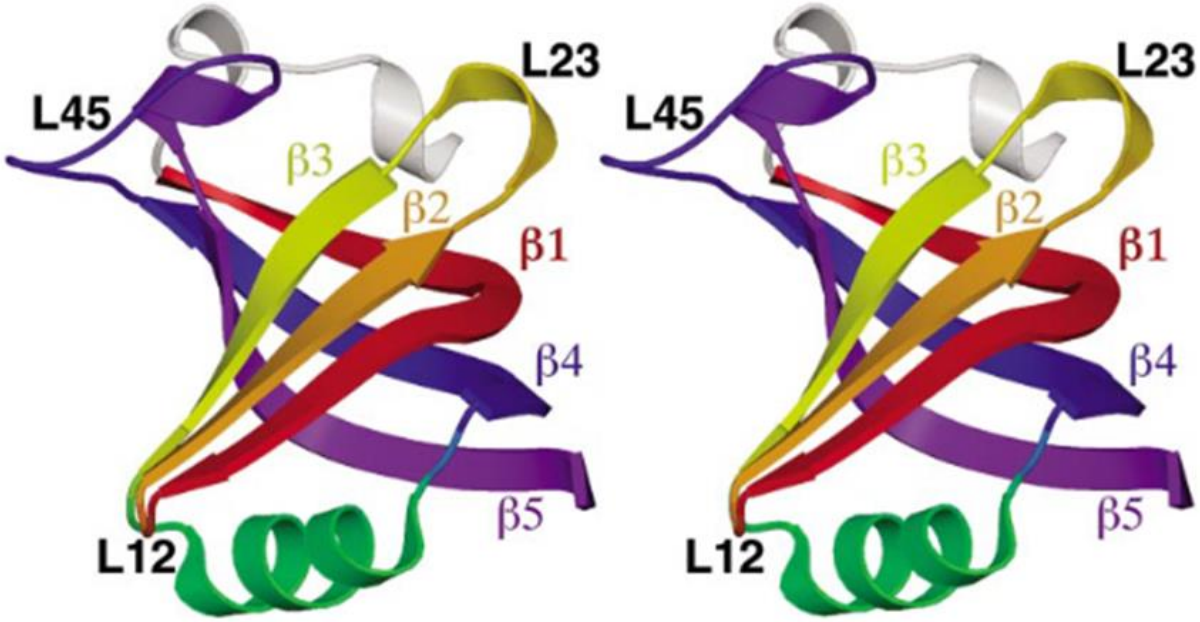
Several proteins within Shelterin make contact with DNA.



POT1 uses an OB fold to read the ssDNA sequence

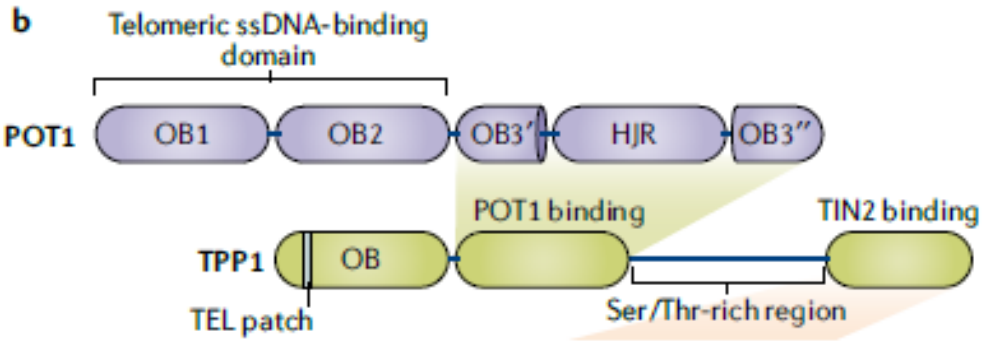


The canonical OB-fold.

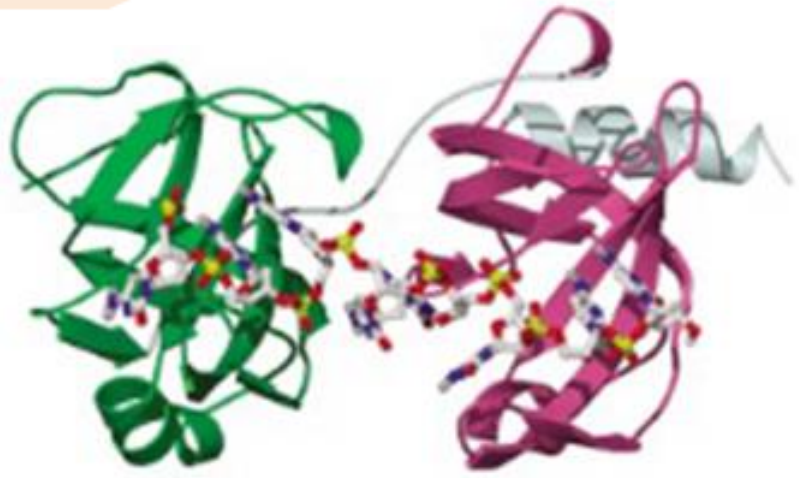


Theobald et al., 2003. NUCLEIC ACID RECOGNITION BY OB-FOLD PROTEINS. Annu. Rev. Biophys. Biomol. Struct. 32:115–33

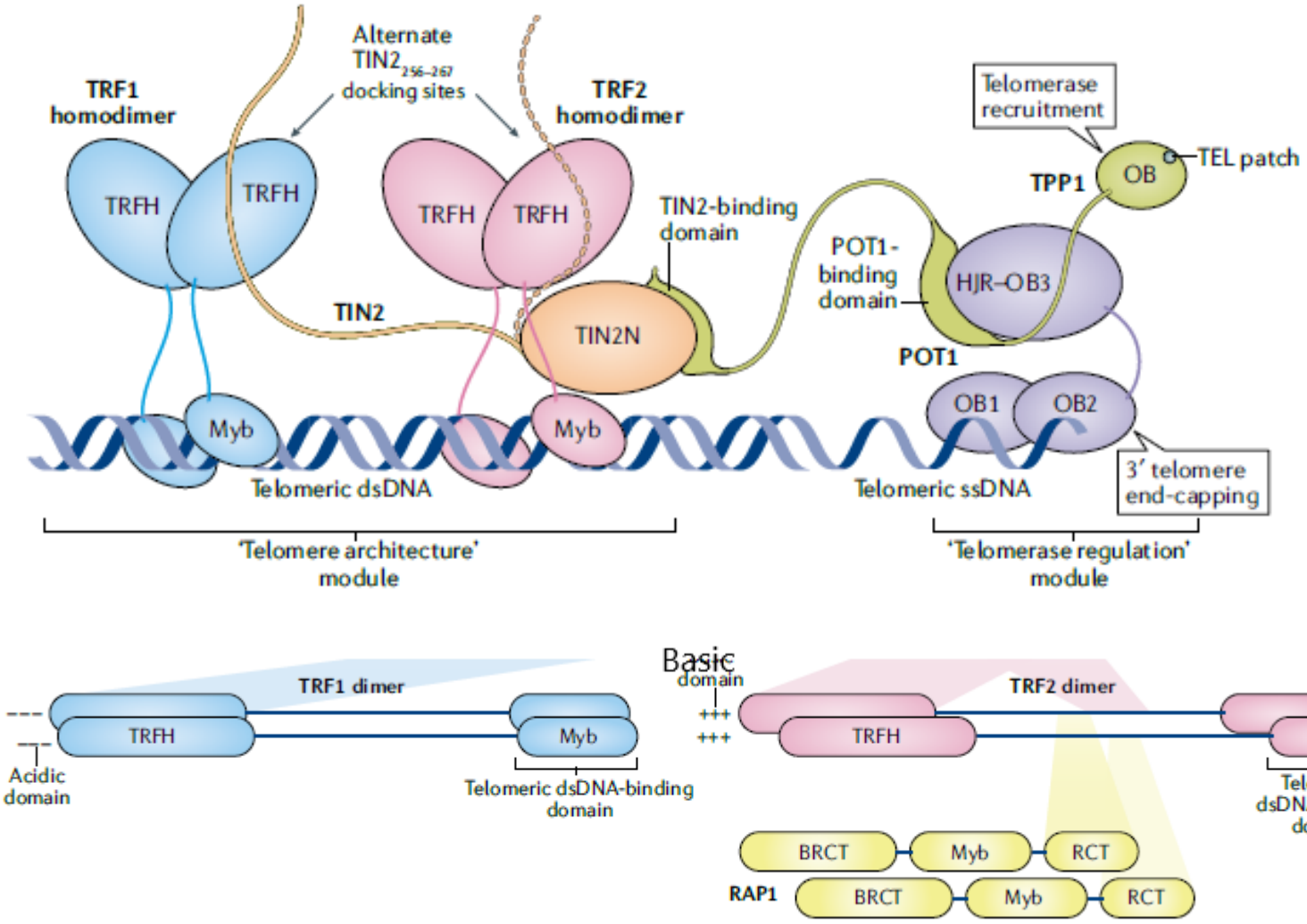
OB folds have floppy structures. The DNA sits on the surface of the β-sheet. If this is sequence-specific binding, where are those intermolecular interactions?

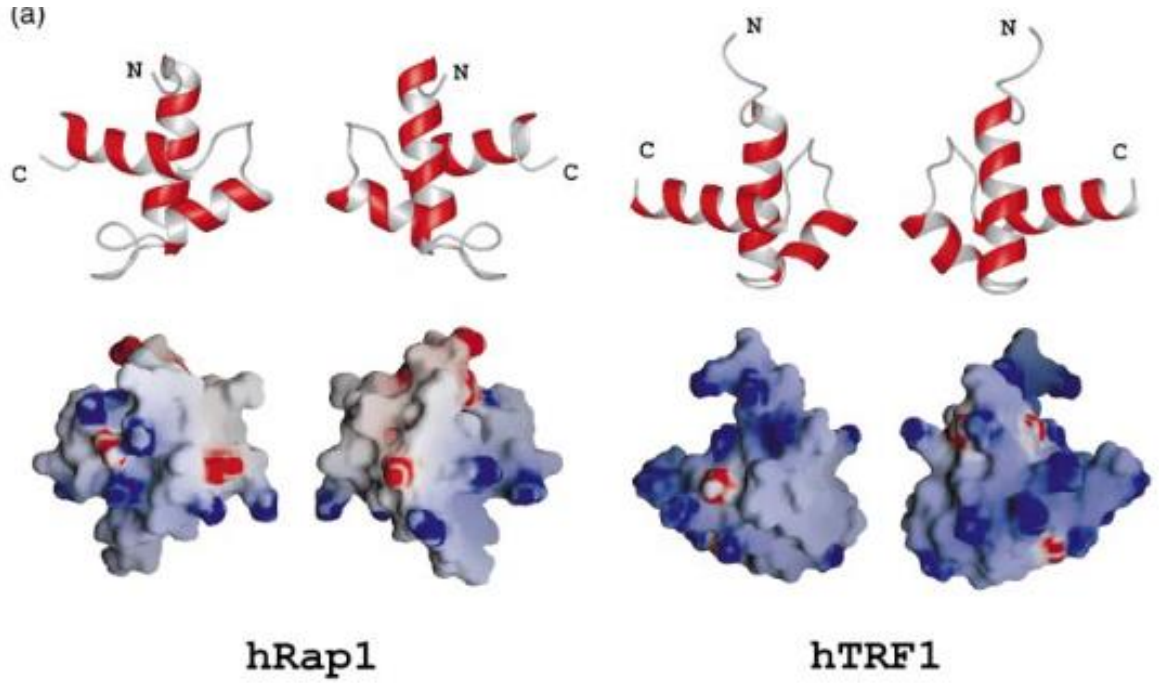


Two known examples

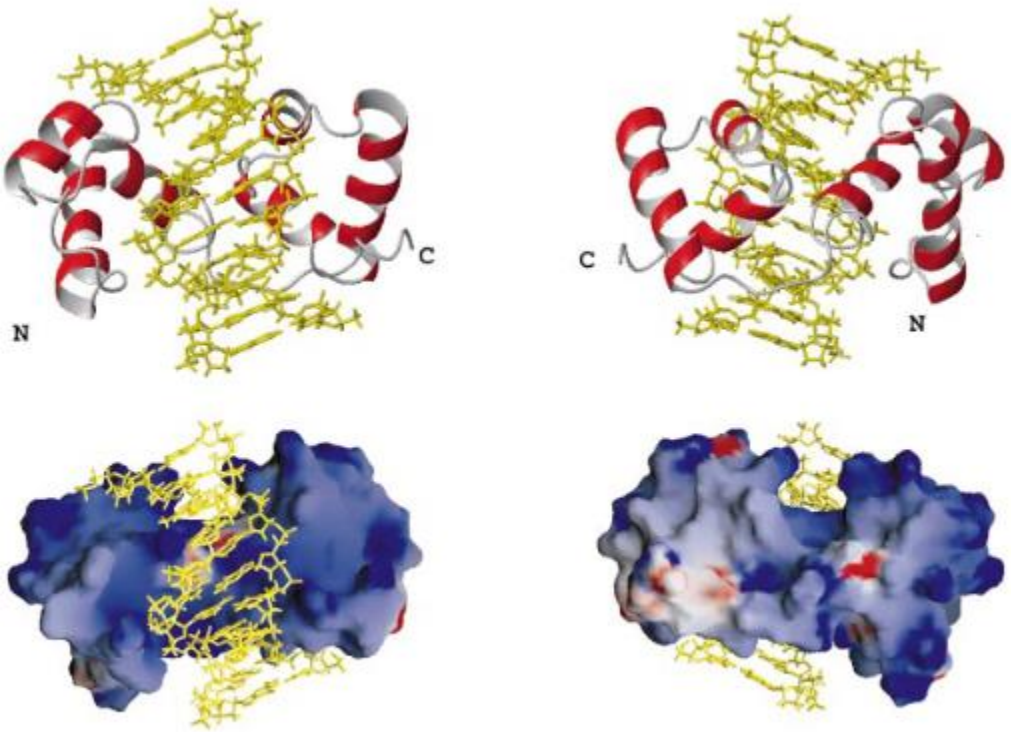


Other proteins in Shelterin use an Myb domain to contact dsDNA





Human hTRF1 and hTRF2 contain a similar C-terminal Myb domain that mediates sequence-specific binding to telomeric DNA.
Human RAP1 (hRAP1) also contains a Myb domain.
blue: positive electrostatic potential surface; red: negative surface



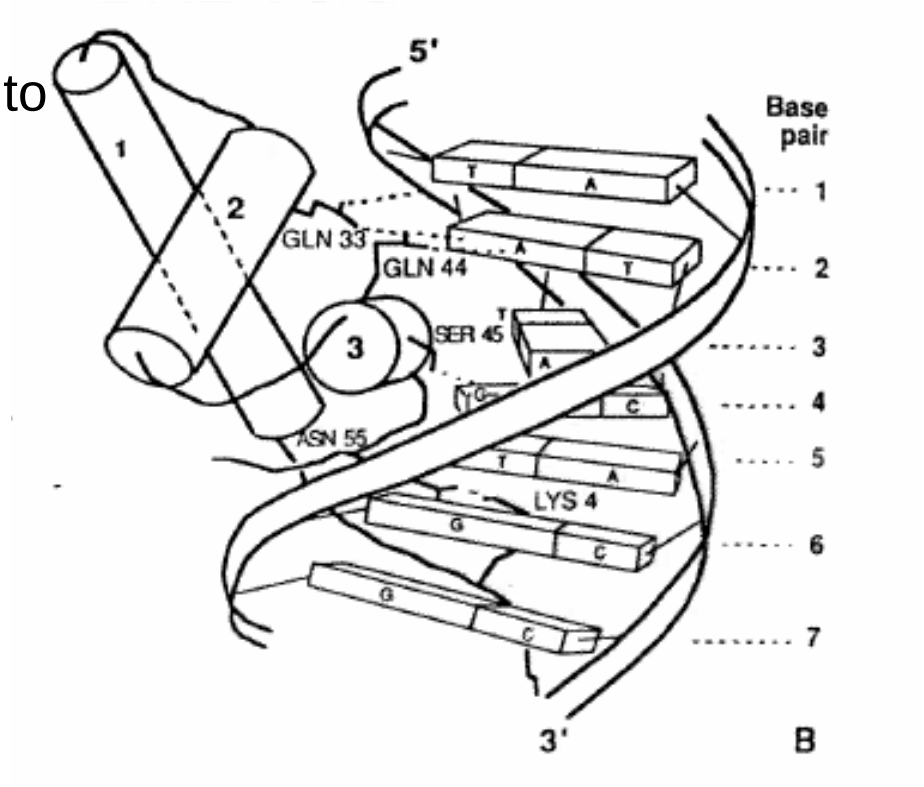
c-Myb R2R3

One possible mode of interaction of two tandem Myb domains with dsDNA

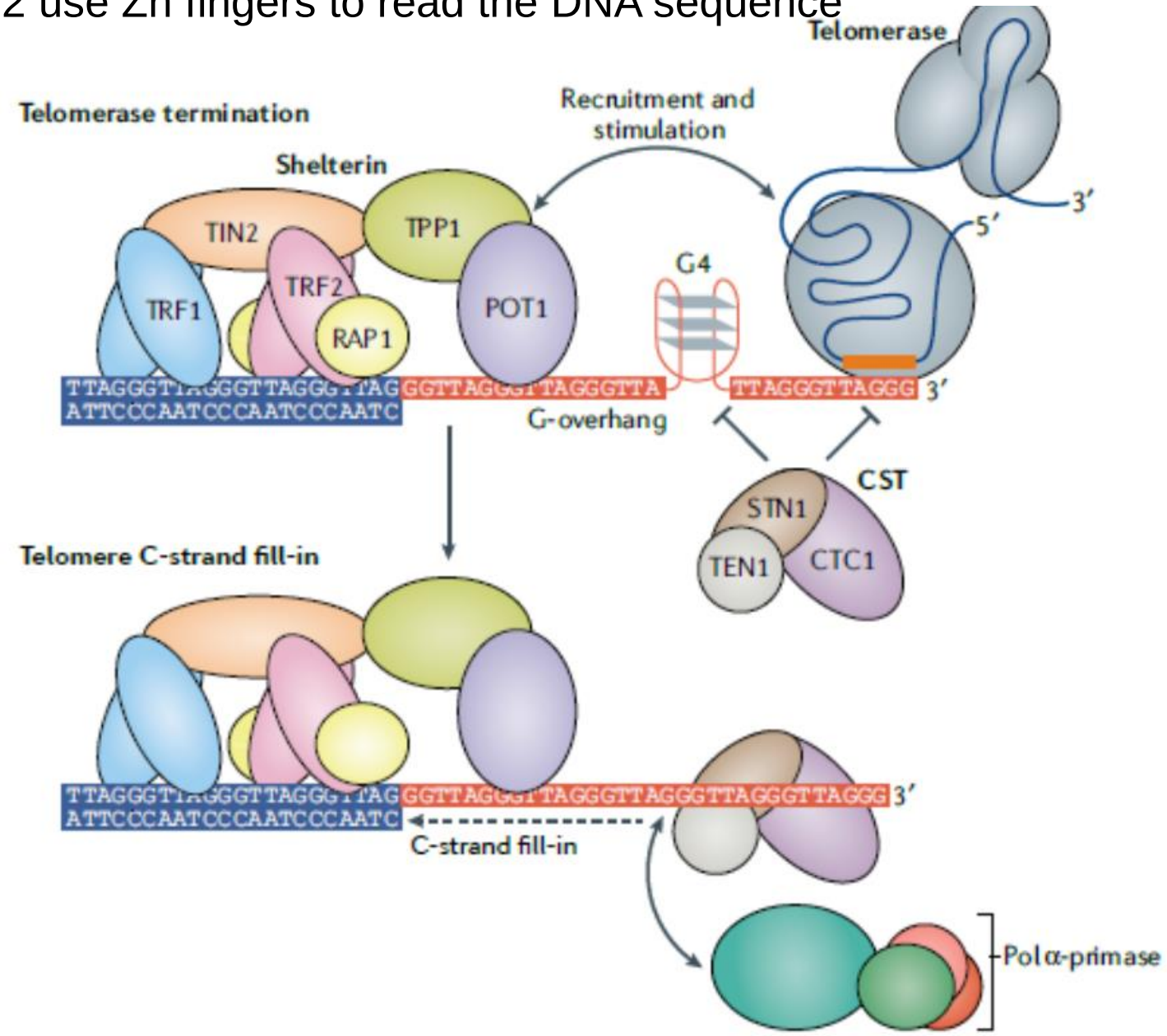
Mybs are examples of Helix-Turn-Helix (HTH) motifs, first identified in bacteria

The helix-turn-helix motif binds to DNA.

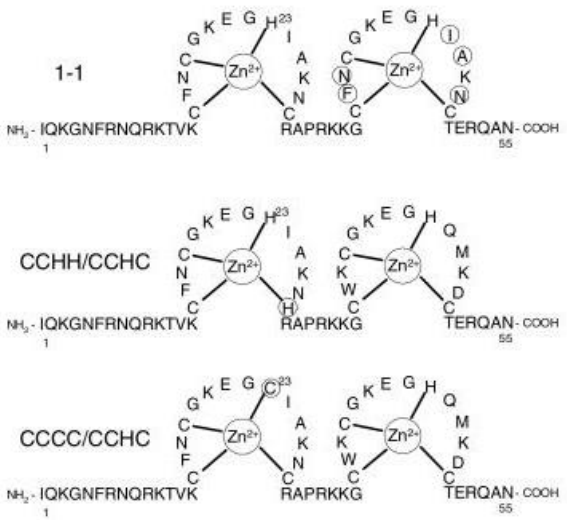
These aren't stable out of the context of the whole protein.

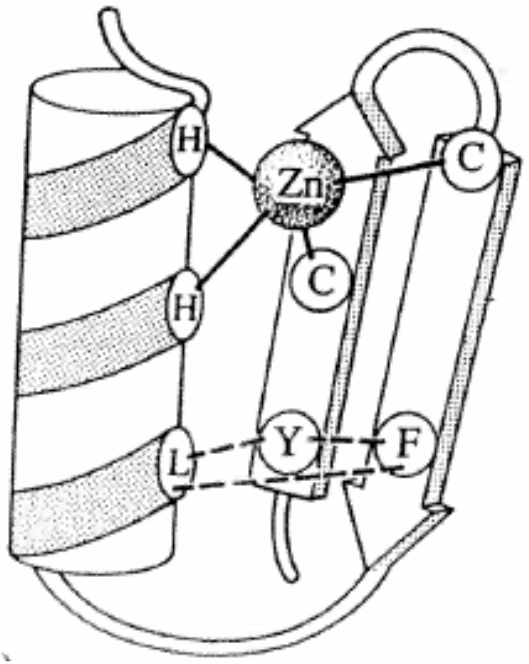


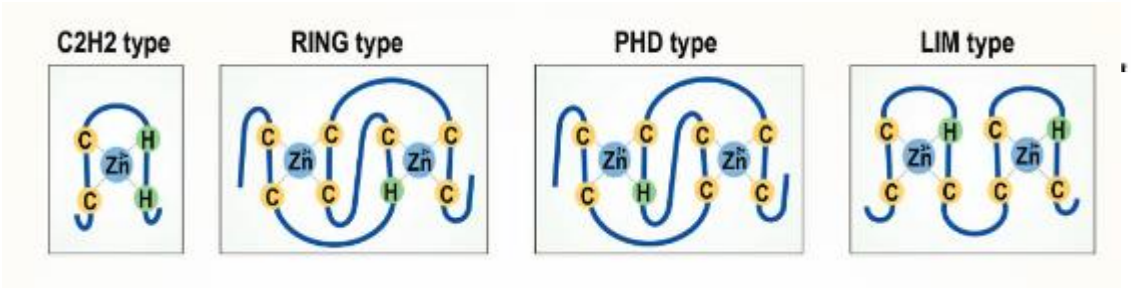
TRF1 & 2 use Zn fingers to read the DNA sequence

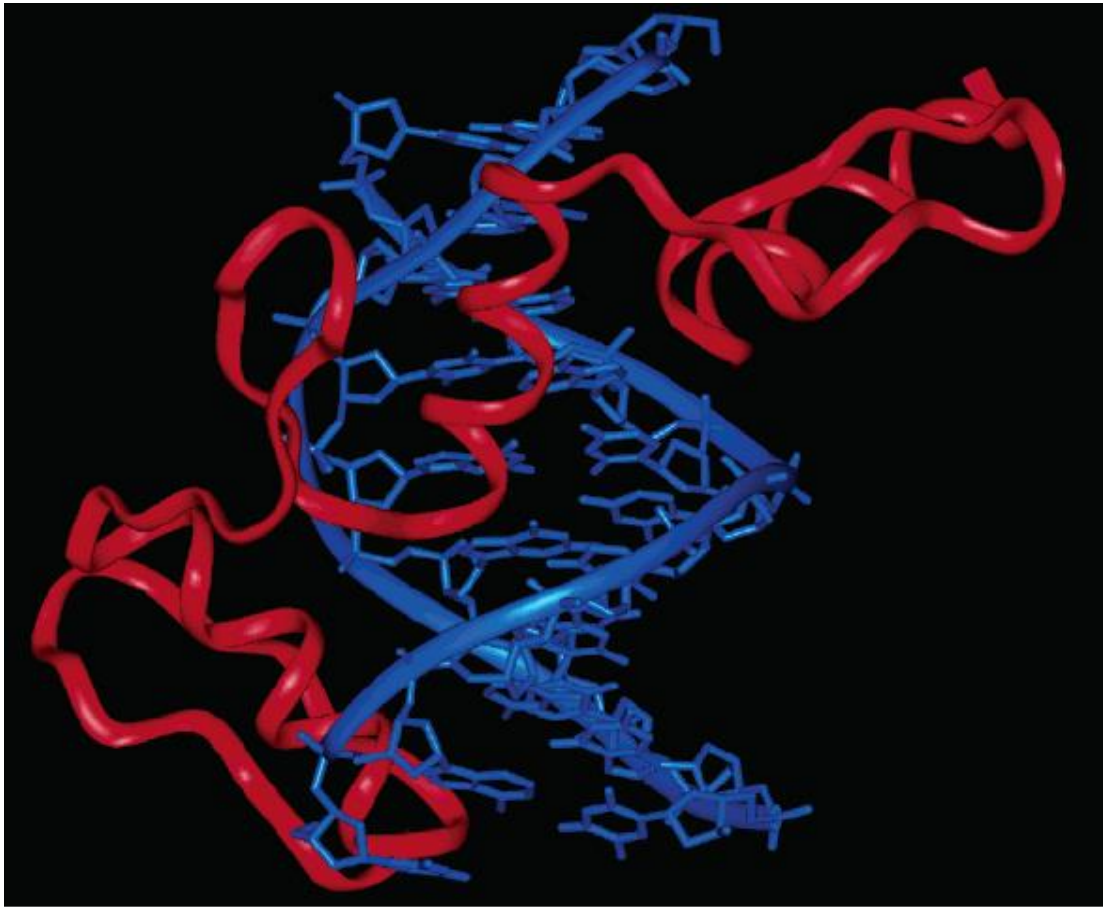


Zinc finger specificity can be modulated. (Talens)





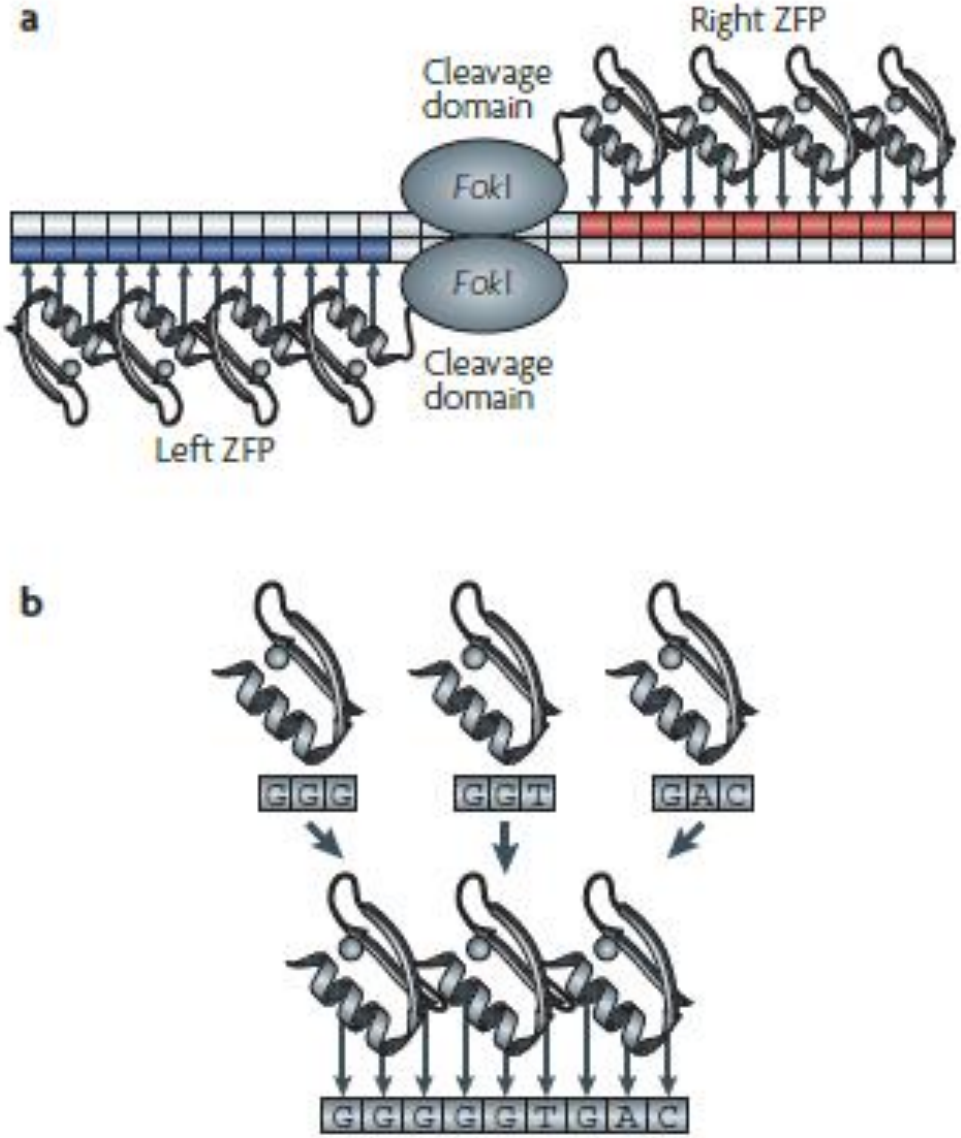


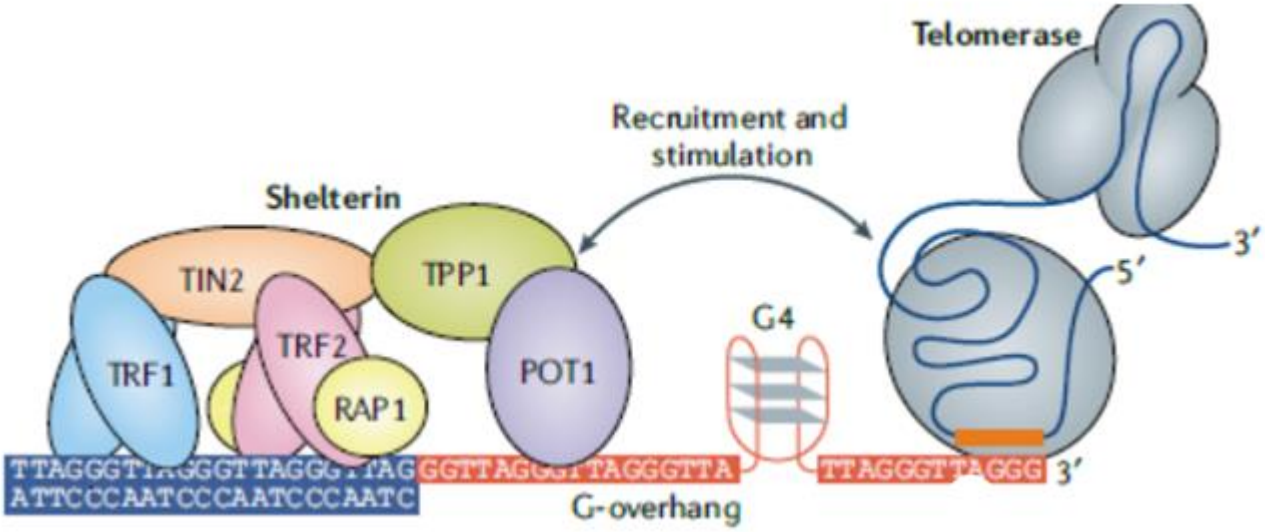


Three tandem fingers bind to DNA. What's the advantage of having more than one finger?

Genome editing with engineered zinc finger nucleases

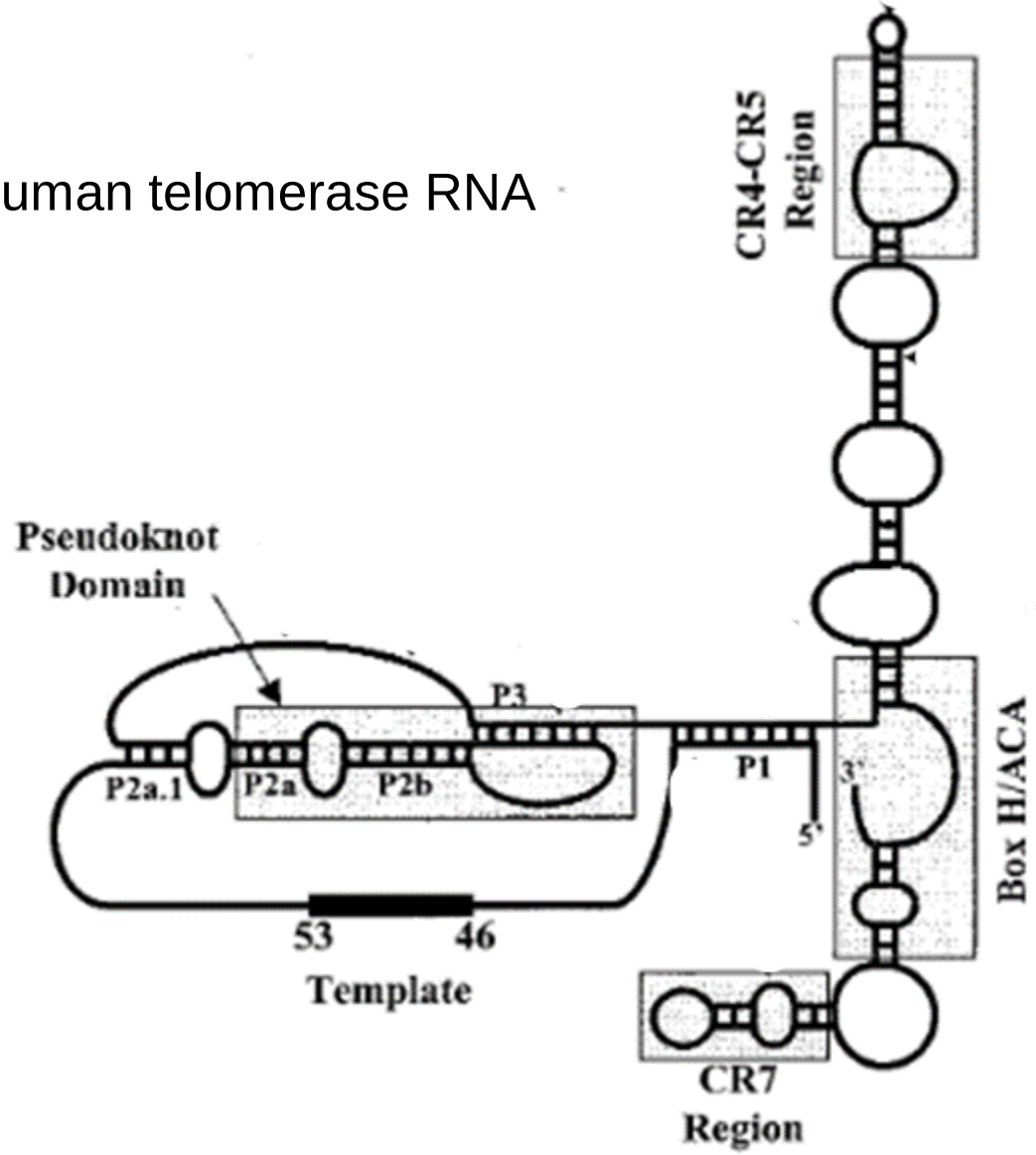
Urnov et al., 2010, Nature reviews/genetics 11:636-646.





Telomerase has two critical components:
1. The RNA is the template for addition of the DNA repeats.
2. The TERT enzyme (Telomerase Reverse Transcriptase) synthesizes the telomeric DNA repeats from reading the Telomerase RNA. TERT can read through G4. It is a classical RT.

Human telomerase RNA



The Stanley J. Korsmeyer Memorial Lectureship

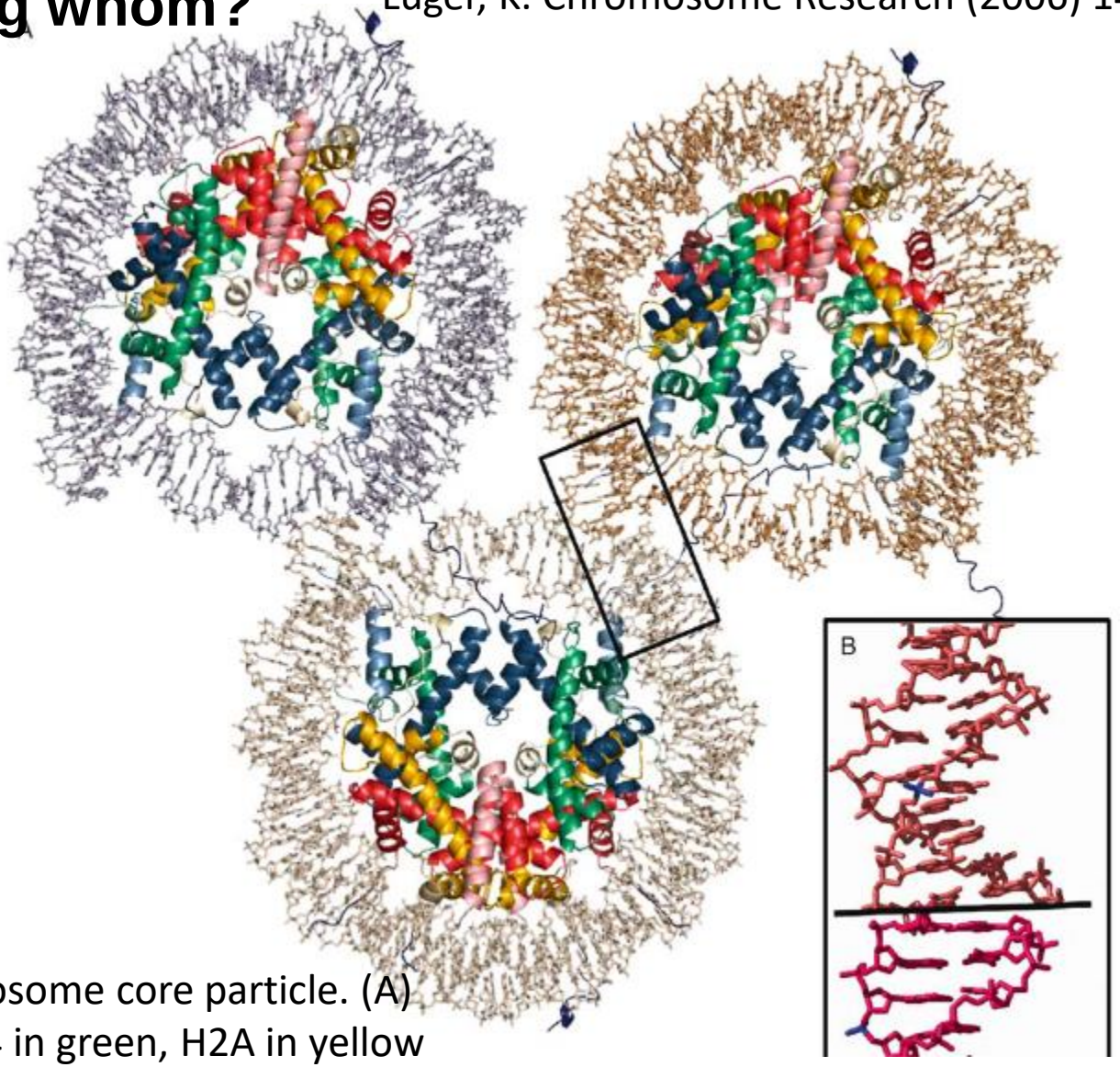
Wednesday, October 26, 2023
3:00pm - 4:00pm CST
Eric P. Newman Education Center, Floor 1
EPNEC Auditorium

"Telomerase: Is the Fountain of Youth a Death Trap?"

Thomas R. Cech, Ph.D.
University of Colorado Boulder
Distinguished Professor of Biochemistry
Investigator, Howard Hughes Medical Institute
Nobel Prize in Chemistry, 1989

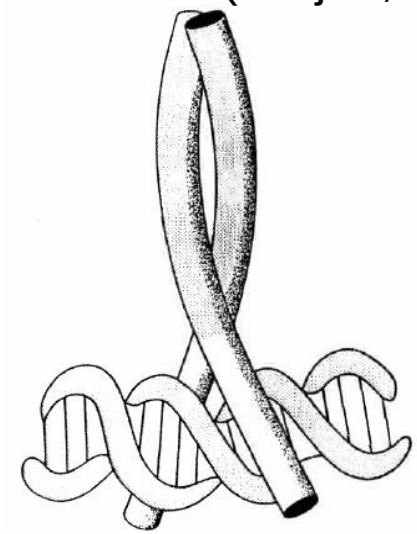
Who's binding whom?

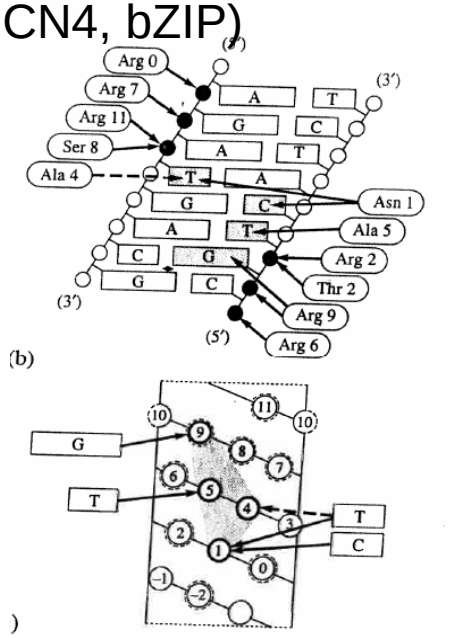
Luger, K. Chromosome Research (2006) 14:5Y16



Structure of the nucleosome core particle. (A) H3 is shown in blue, H4 in green, H2A in yellow and H2B in red. The particles are viewed down the superhelical axis of the DNA (grey).

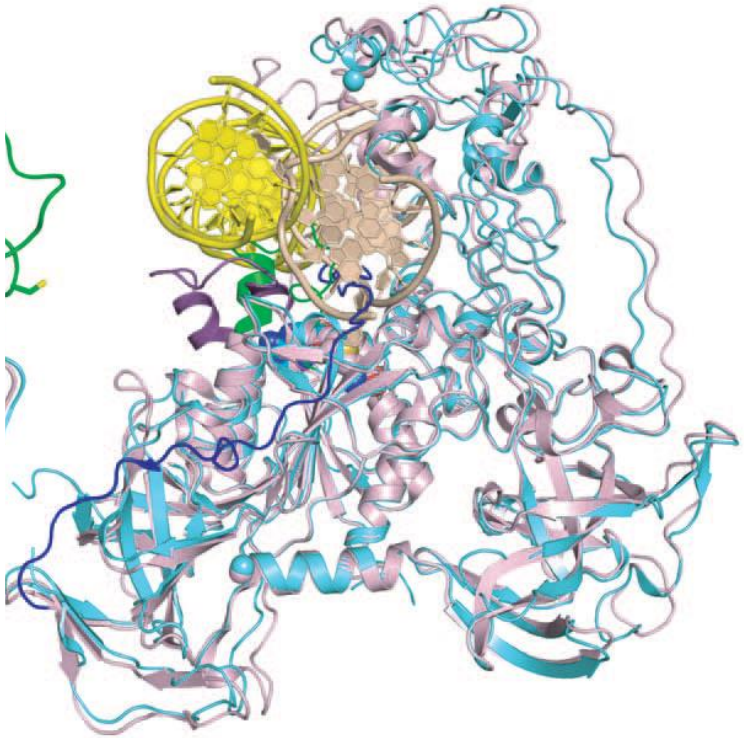
Leucine zippers themselves can bind DNA (fos/jun, GCN4, bZIP)

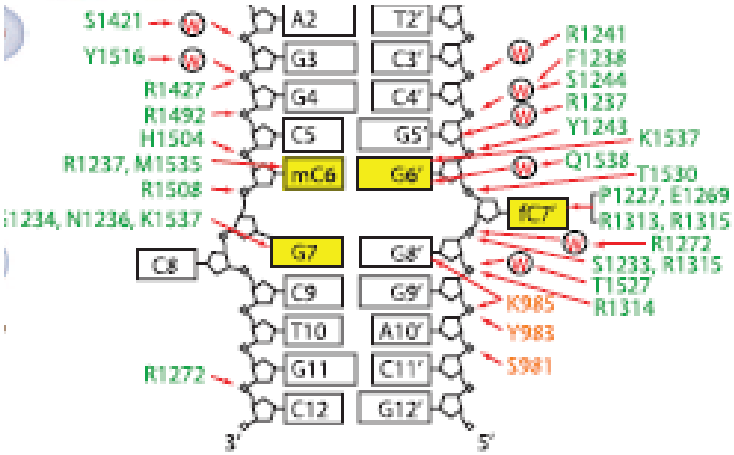




^{5m}C DNA methyltransferase

70-80% of human CpG dinucleotides





Song et al., 2012. Science 335:709-712

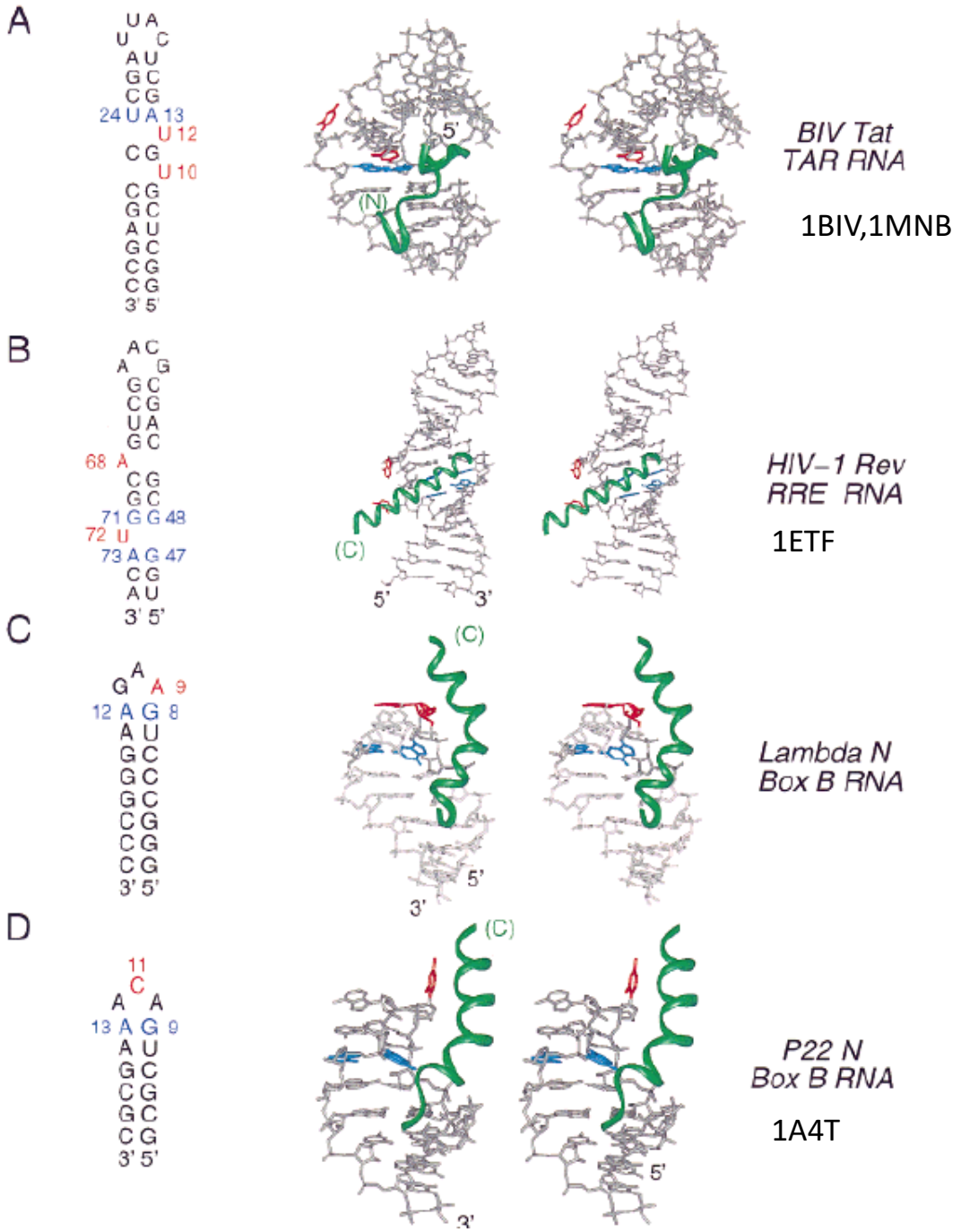
Proteins that bind RNA.

First, The Arginine-rich motif (ARM)
These ‘motifs’ are intrinsically disordered regions!

Peptide–RNA complex name		Peptide/RNA Sequence
BIV Tat–TAR	Tat peptide:	SGPRP RGTRG KGRRI RR 65 81
	TAR RNA:	4 U U 5' GGCUCG G AGCU C A 3' CCGAGC C UCGA U U 31
HIV–1 Rev peptide–RRE RNA	Rev peptide:	DTRQA RRNRR RRWRE RQRAA AAR 33 55
	RRE RNA:	41 5' GGUCUGG GCG CAGC G C 3' CC AGACA GGC GUCG A A U A 79
HIV–1 Rev peptide–RRE aptamer	Rev peptide:	TRQAR RNRRR RWRER QR 34 50
	RRE aptamer:	1 G A 5' GGCUG CUC GUAC U U 3' CCGA C GAG CAUG G C A A A 35 G U
λ N peptide–boxB RNA	λ N Peptide:	MDAQT RRRER RAEKQ AQWKA AN 1 22
	λ boxB RNA:	1 5' GGGCCCUG A 3' CCCGGGAA G A 19
P22 N peptide–boxB RNA	P22 N Peptide:	NAKTR RHERR RKLAI ERDTI 1 20
	P22 boxB RNA:	4 5' GCGCU G A C 3' CGCGA A A 18

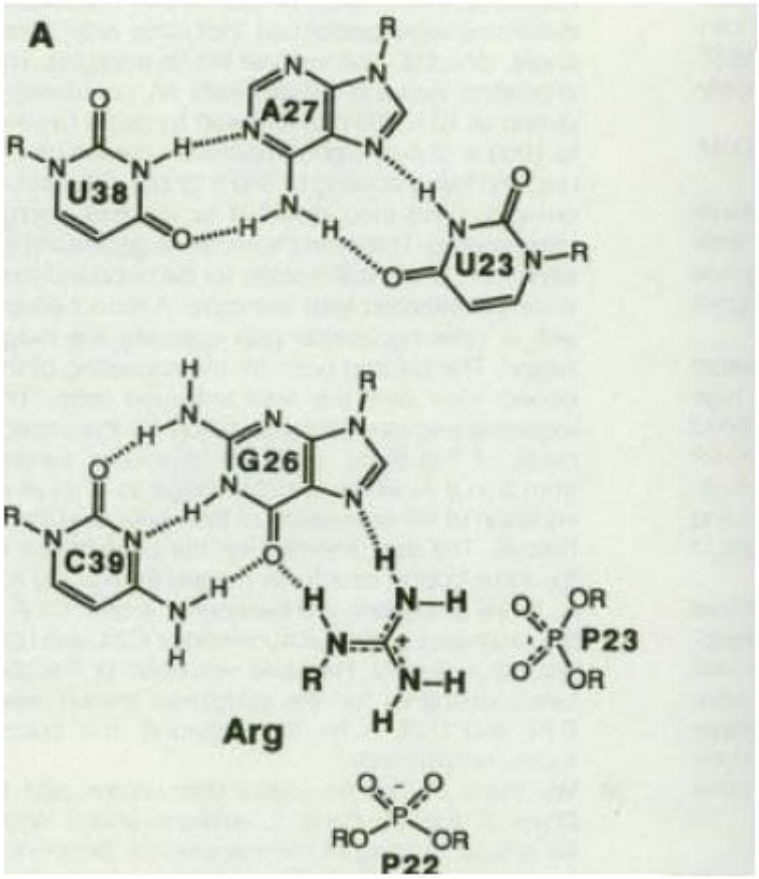
Most peptides can fit into these major grooves where there is a dramatic deformation of the A-form duplex by the bulged nucleotides.

Due to the number of interactions (hydrogen bonds, electrostatics, and stacking hydrophobic amino acids), the dissociation constants of these small complexes can be nanomolar.



Weiss & Narayana (1998) Biopolymers 48:167

Binding of the Tat protein ARM: ITKALQISYRKKRRGRRAPQGSQTHQVSL can be reduced to binding of a single R.



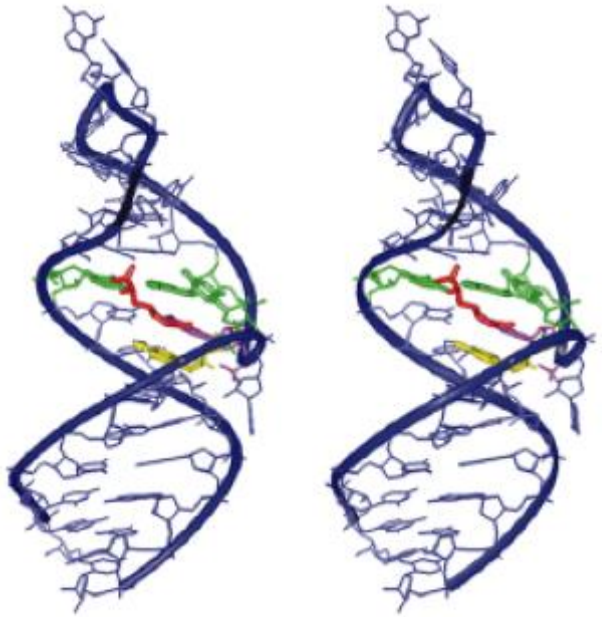
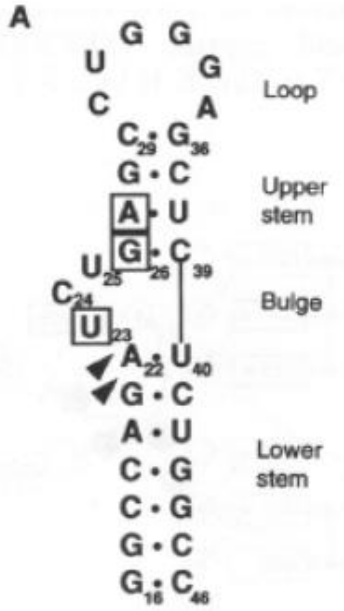
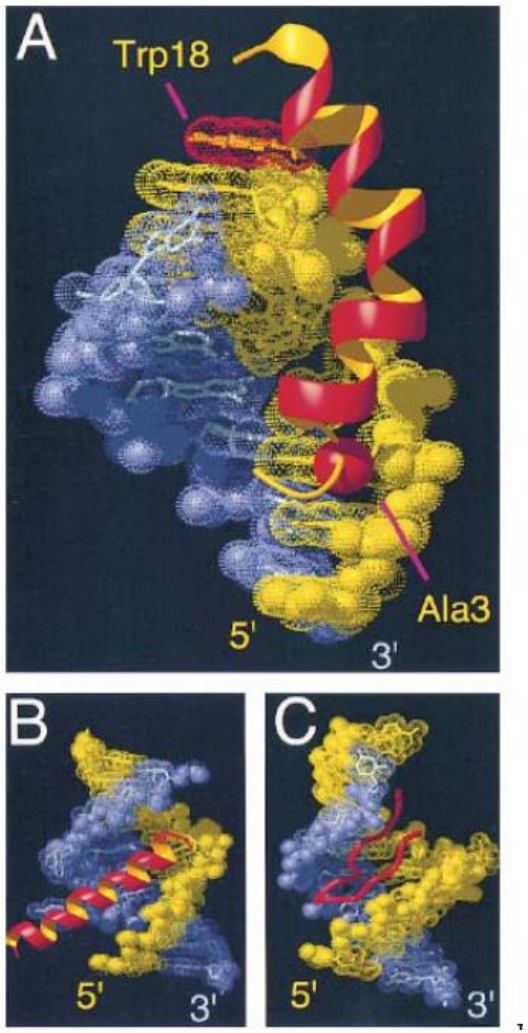


Figure 6. The average structure of the TAR-argininamide complex is shown with U23 highlighted in green, G26 in purple, A22 in yellow and the argininamide in red. The ribbon traces the backbone showing the distortion in the bulge region.



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. And remember: in the context of the proteins, the peptides are intrinsically disordered, so binding to RNA is accompanied by a conformational change.

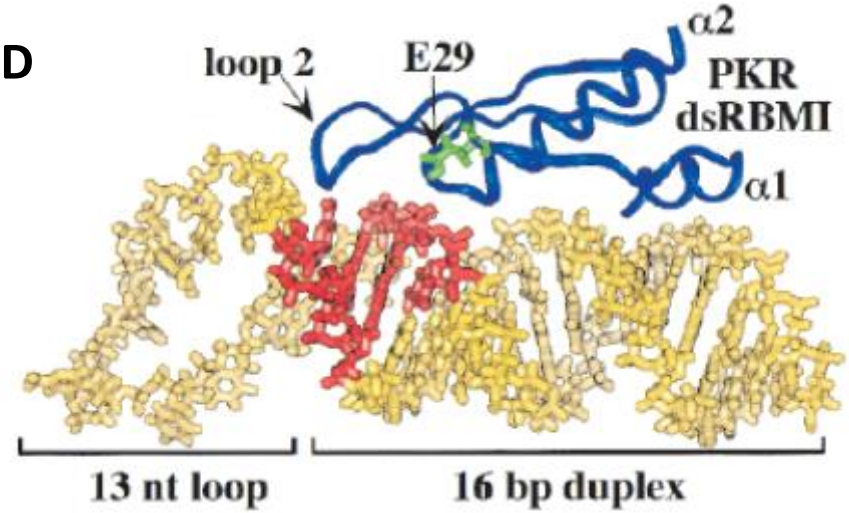


A is P22/BoxB. The α-helix fits into the groove of the RNA and bends over the GAAAA loop where its Trp stacks with an adenosine. The peptide bends at R11 to allow the helical sidechains to stack with the nucleobases.

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

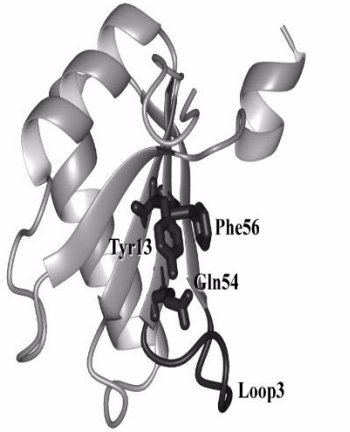


Double-stranded RNA binding domains (dsRBM) are nonspecific, but sensitive to A-form structure.
This is the domain from PKR (Protein Kinase R).

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).

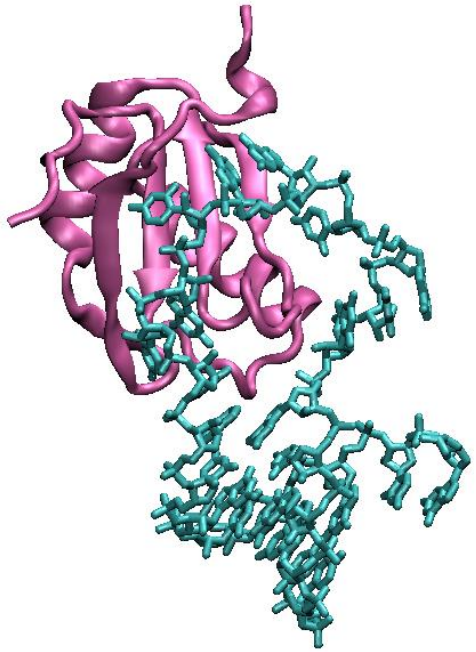
3. **RNA Recognition Motif (RRM).** It is the most common eukaryotic RNA binding domain. RRMs are identified by their conserved sequences



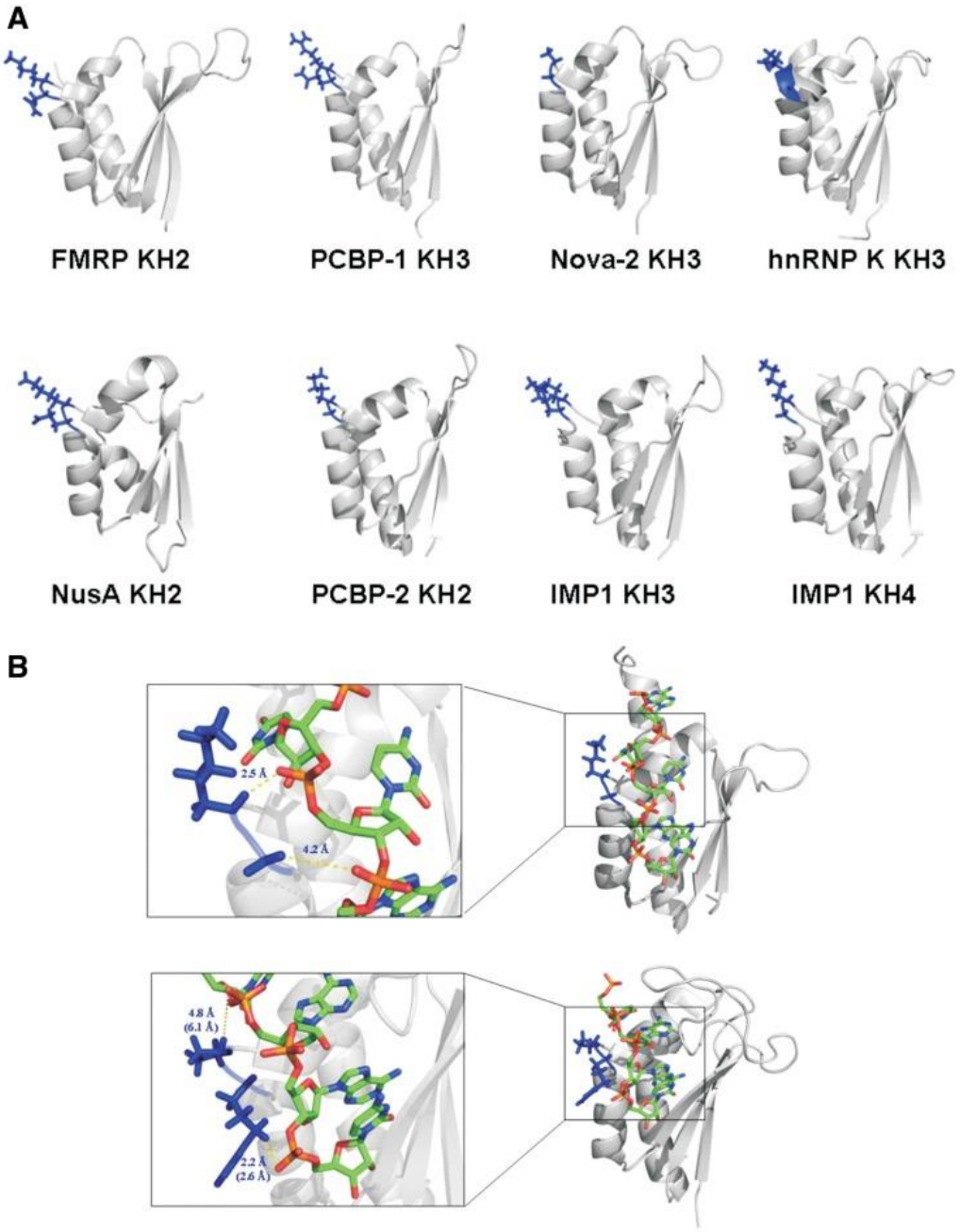
	RNP-2	RNP-1
Consensus	LFVGNL	KGFGFVXF
	IY I KL	R YA Y

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.

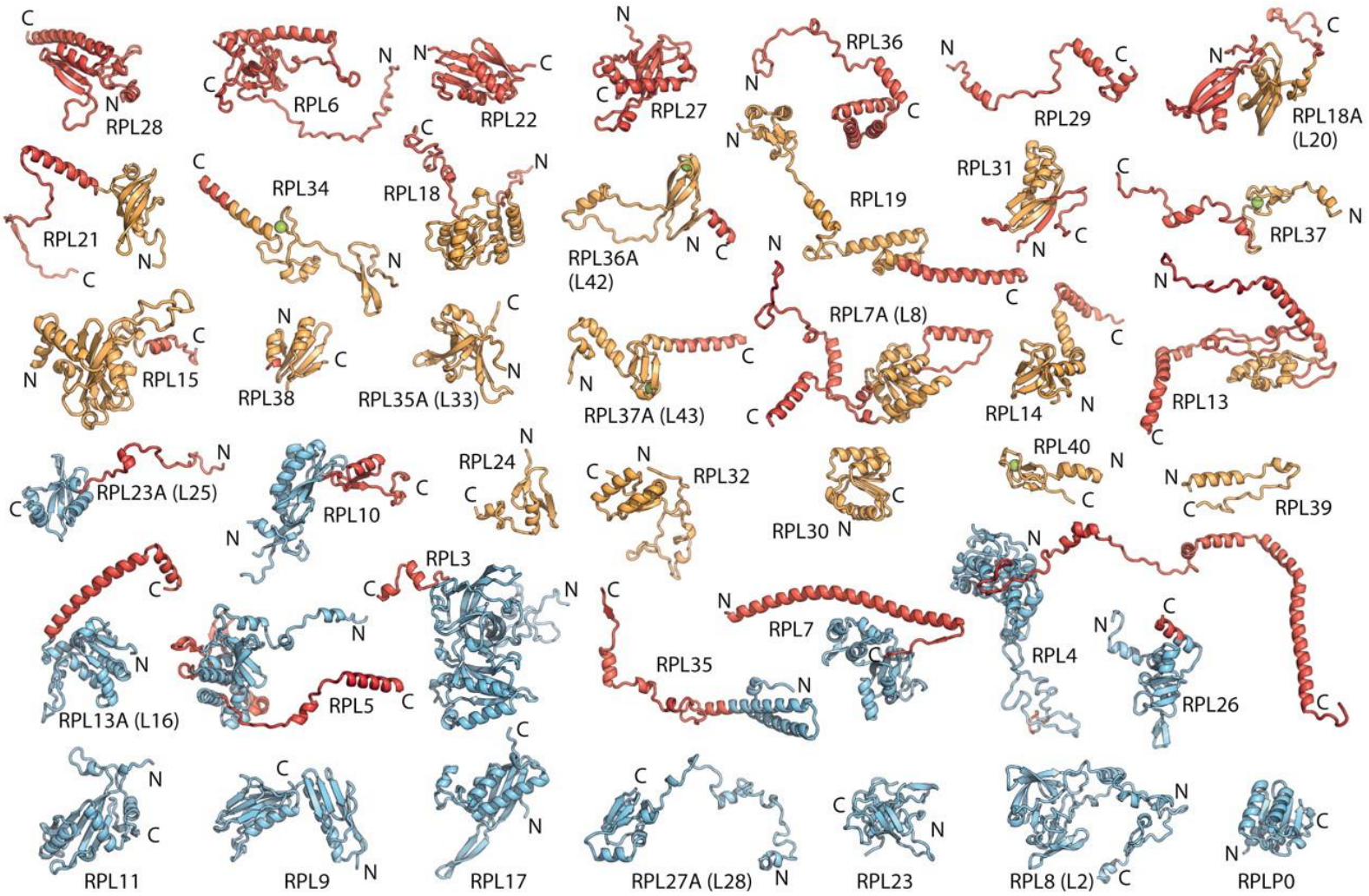


4. KH domains
use amino acids
on their α-helical
side to bind RNA



Hollingworth et al., 2012. Nucleic Acids Res 40:6873-6886.

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