Protein-nucleic acid interactions

General modes of interaction

Non-specific: histones, polymerases, ssb…

Specific: regulatory proteins, transcription factors…

Some factors exhibit both binding modes: non-specific interactions for substrates that lack a specific binding site; specific interactions achieved when sites are located

Protein-nucleic acid interactions

General modes of interaction

- Non-specific: histones, polymerases, ssb…
- Specific: regulatory proteins, transcription factors…

Some factors exhibit both binding modes: non-specific interactions for substrates that lack a specific binding site; specific interactions achieved when sites are located

Non-specific binding can utilize general properties of the nucleic acid, electrostatics and shape recognition

Specific interactions involve direct readout of a specific structure or specific sequences

Some specific interactions:

- Proton donors and acceptors
- van der Waals interactions
- Major versus minor groove recognition
- Distortion of nucleic acid in a sequence/damage dependent manner
- Water/cation mediated interactions

Protein-nucleic acid interactions

Proton donors and acceptors in the bases (hydrogen bonds, dipolar forces)

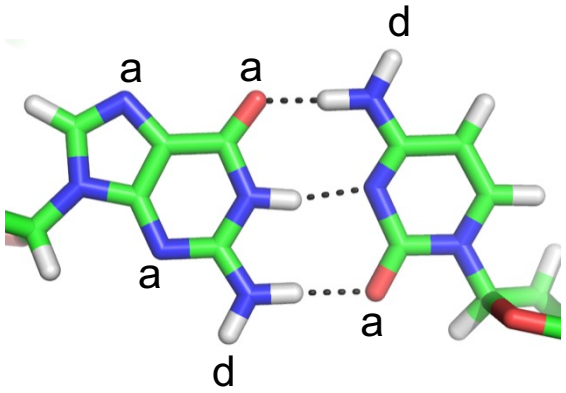
Hydrogen bonds are dipolar interactions that operate at short range

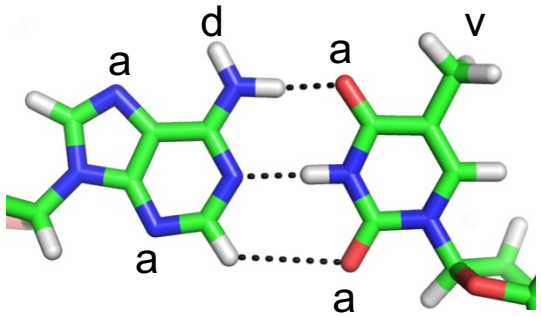
Provide specificity to the interaction

These occur between the protein amino acid side chains, backbone amides and carbonyls and the bases, sugars and backbone oxygen atoms of the nucleic acid

Base stacking interactions utilize dispersion forces and the hydrophobic effect

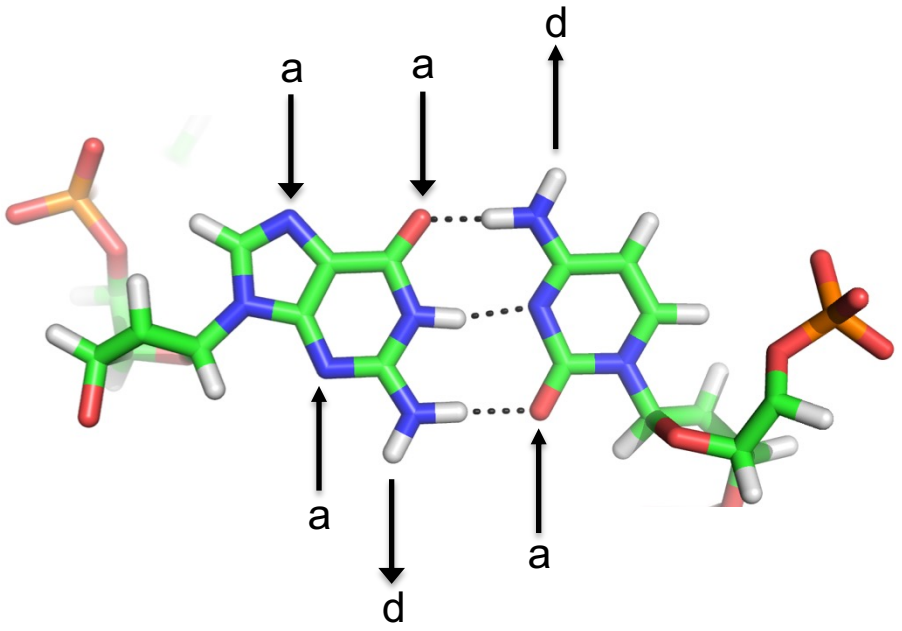
Proteins also interact with bases using amino acid side chains





Protein-nucleic acid interactions

Major groove



G

C

Minor groove

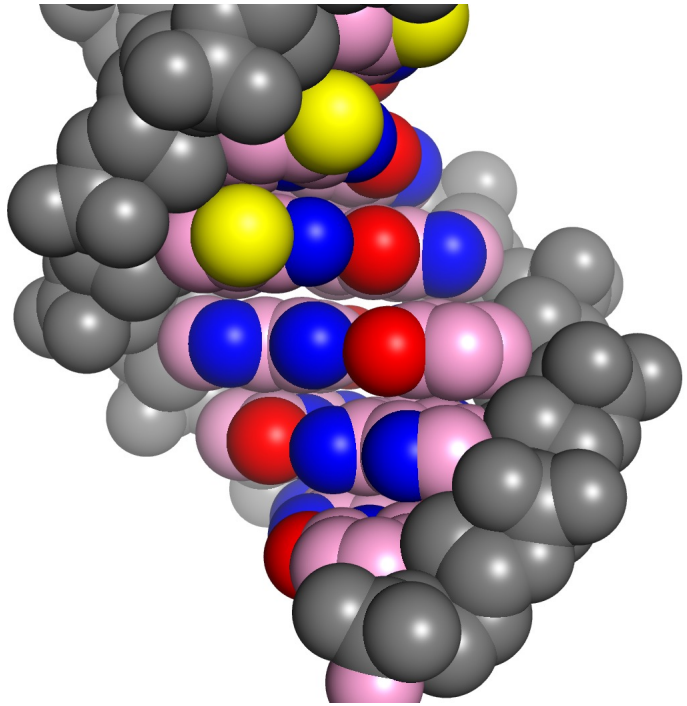
Protein-nucleic acid interactions

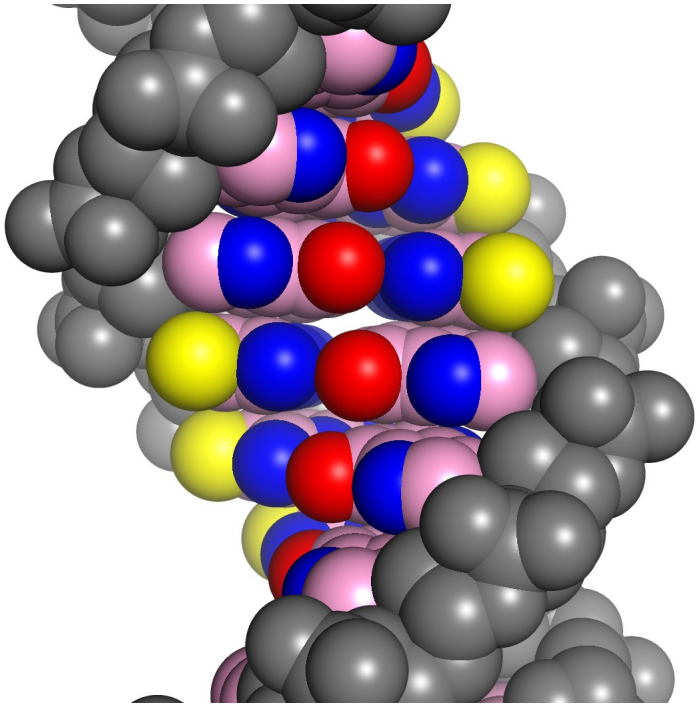
Major groove

T:A

G:C

C:G





A:T

T:A

acceptor donor VDW

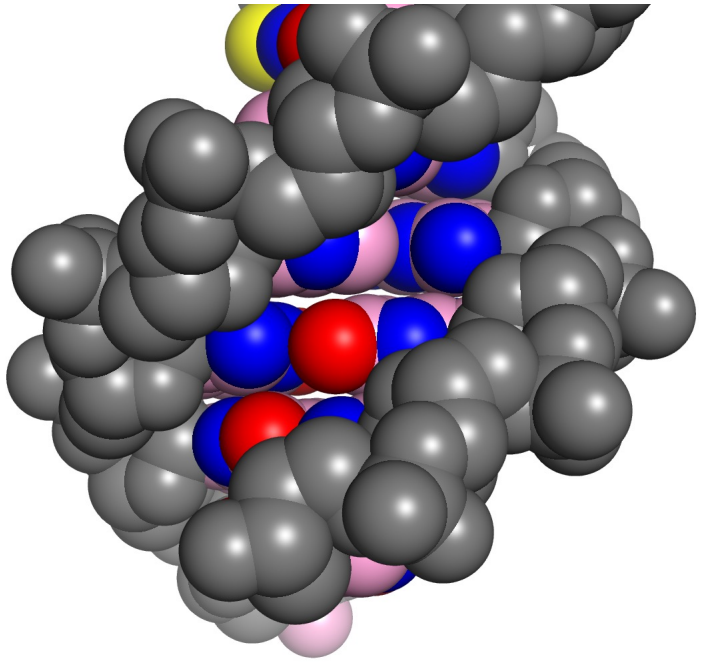
Protein-nucleic acid interactions

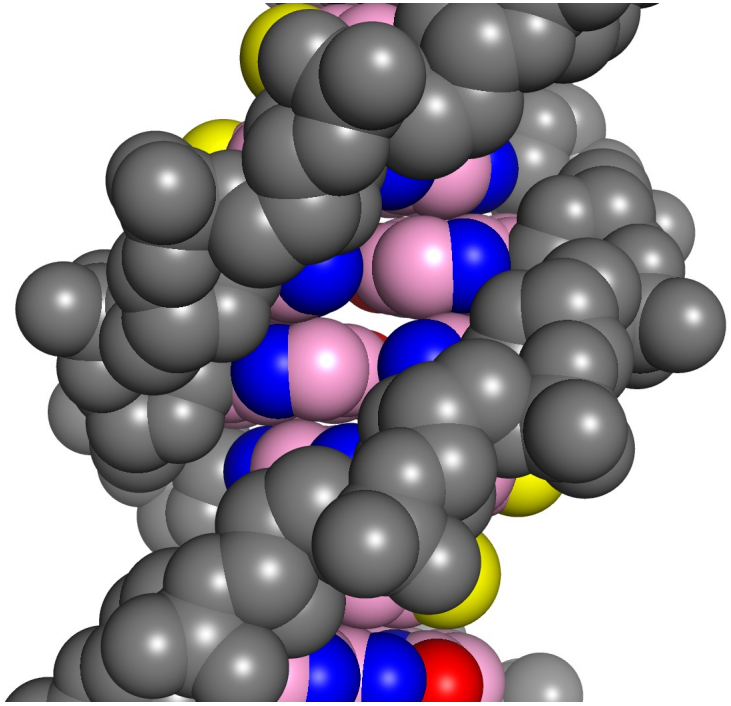
Minor groove

A:T

C:G

G:C



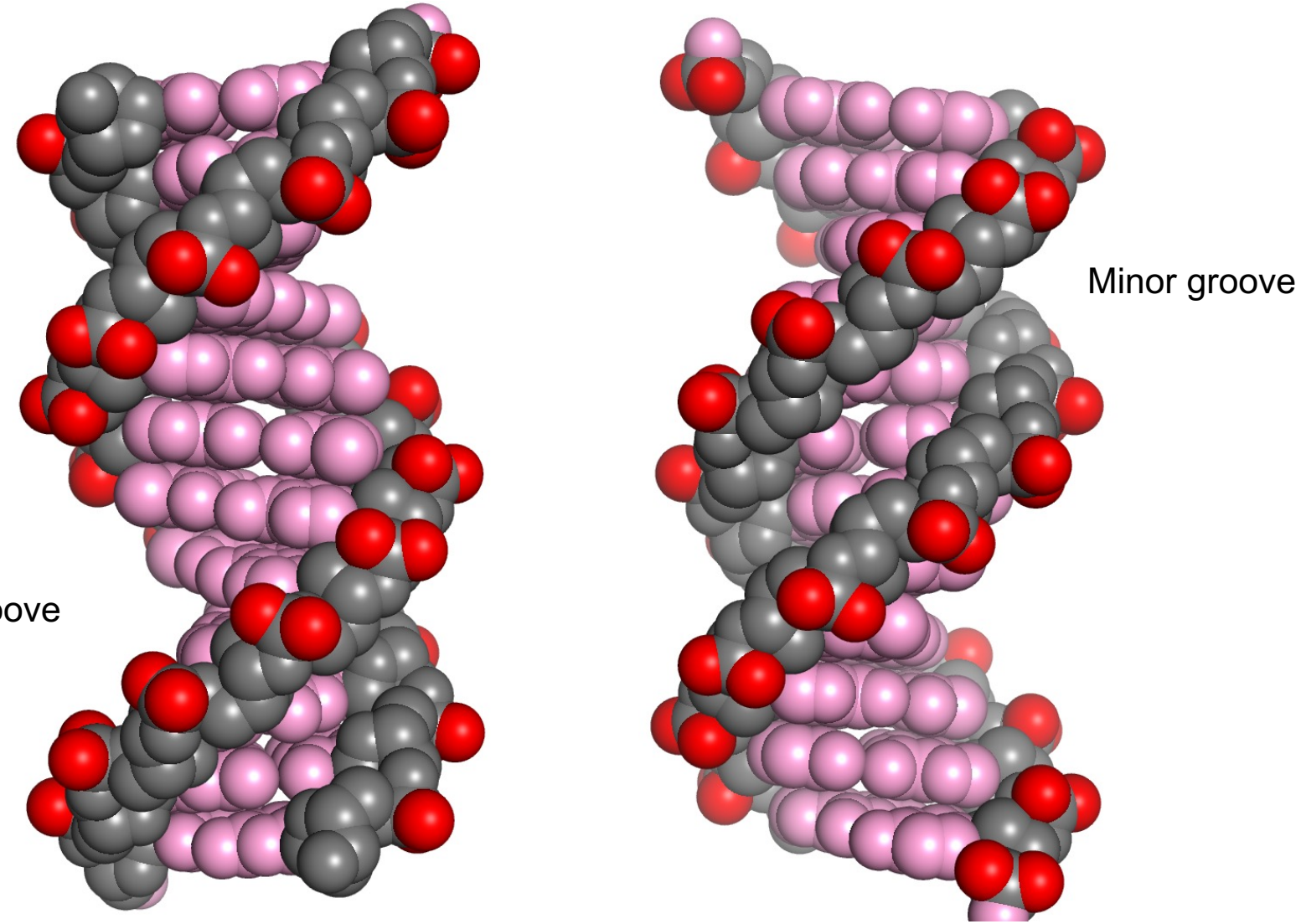


T:A

A:T

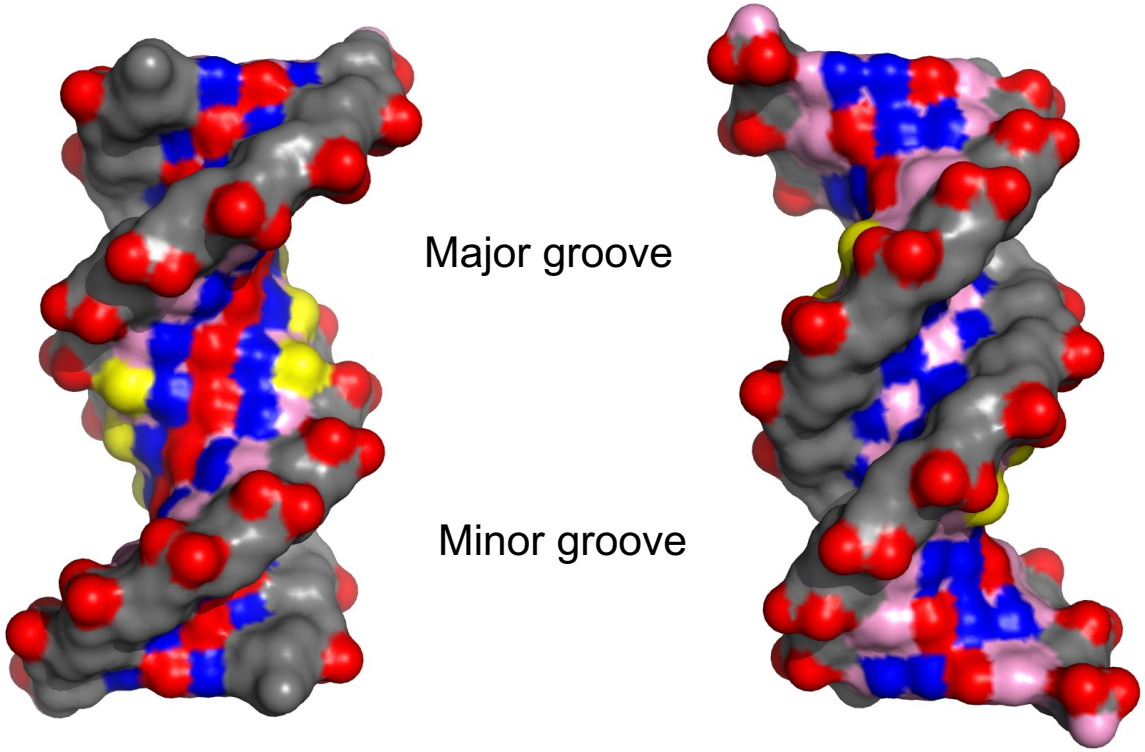
acceptor donor VDW

Protein-nucleic acid interactions



Non-ester oxygen atoms

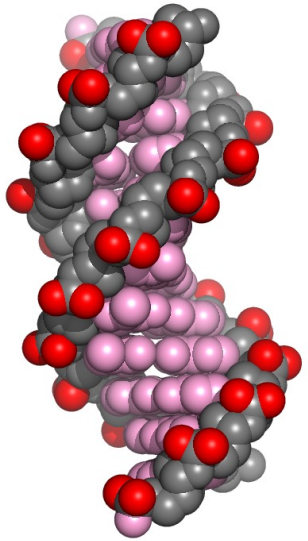
Protein-nucleic acid interactions

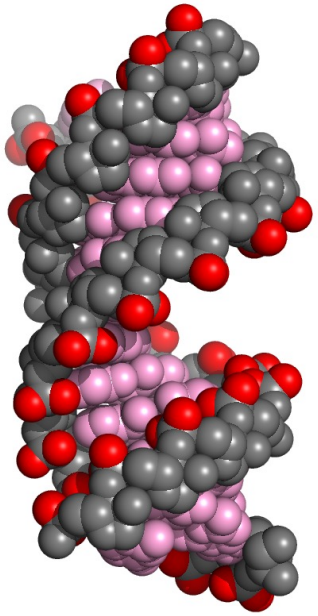


Shape recognition

- Surface dominated by phosphates with bases buried in the interior
- Major groove about 12 Å wide and 8-9 Å deep
- Minor groove about 6 Å wide and 7-8 Å deep
- If a protein is going to bind and specifically recognize the bases it will need structures to reach into the grooves

Protein-nucleic acid interactions





Major and minor grooves

B-DNA	wide major groove
	narrow minor groove
A-DNA/RNA	deep but narrow major groove
	shallow but wide minor groove

Sequence specific DNA binders generally survey the major groove because it can accommodate a protein alpha helix or 2-stranded beta sheet

Some non-sequence specific DNA proteins bind to the minor groove distorting it to accommodate protein interactions (TATA binding protein)

Drugs that intercalate into DNA often bind in the minor groove in A:T rich regions

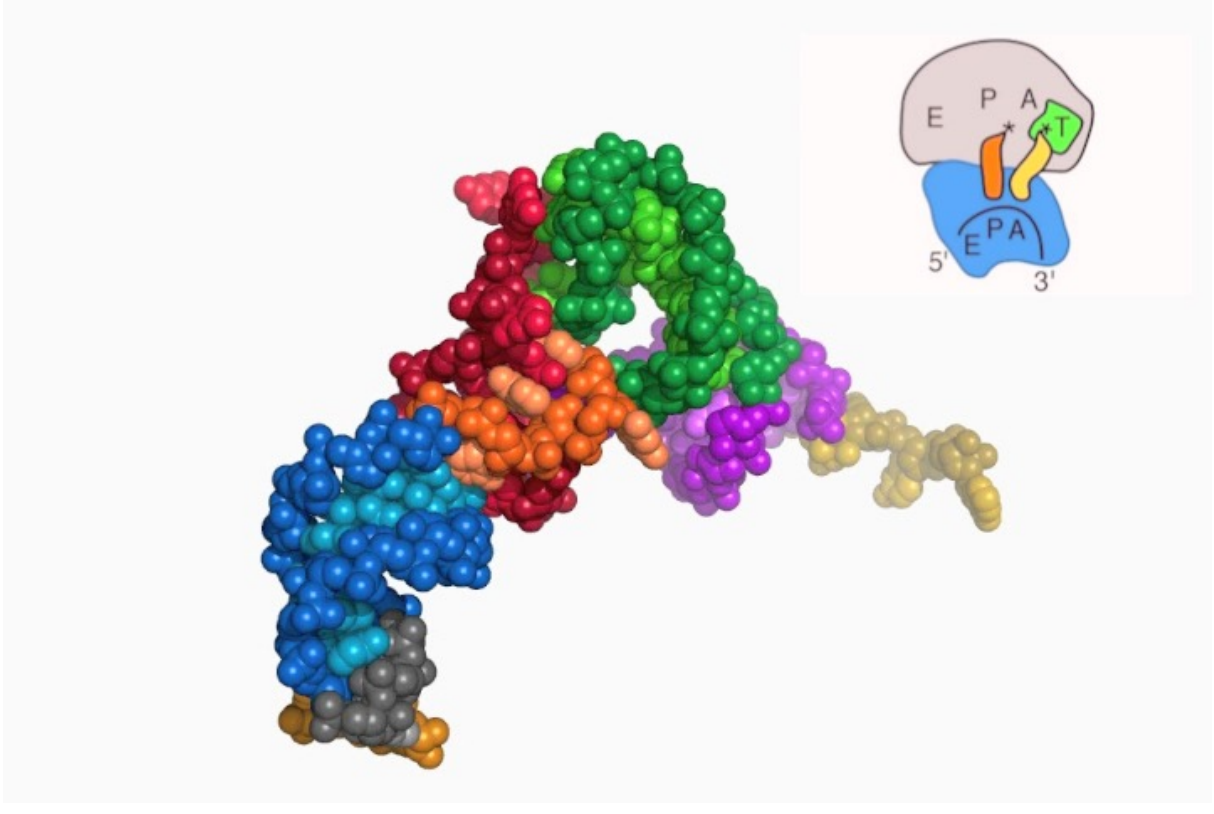
Protein-nucleic acid interactions

Nucleic acids are not static

Structured DNA and RNA are deformable, and this is sequence dependent

Many proteins can sense a sequence indirectly by virtue of distorting the duplex through kinking, bending or unwinding (many examples)

Interactions are also dependent on water and counter ions (most typically cations) that coat the surface of nucleic acids



Dunkle *et al. Science* 2011

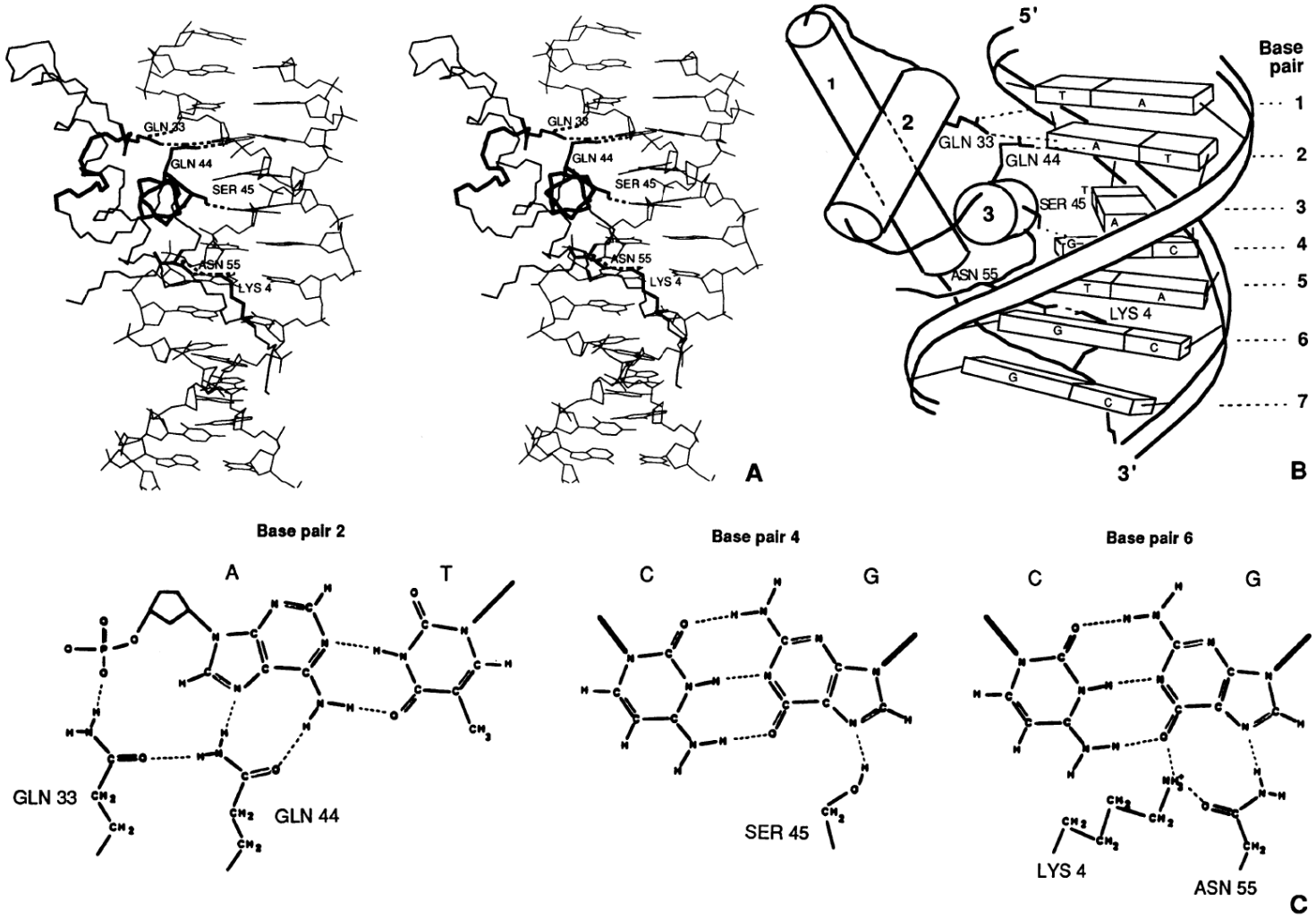
Protein-nucleic acid interactions

- Survey of DNA binding proteins
  - Structural elements that can interact with DNA
    - helix-turn-helix
    - beta-ribbon
    - zinc finger
    - bZip
    - bHLH
  - Some as oligomers or dimers

Protein-nucleic acid interactions

Helix-turn-helix motif

3 helices, one reads, one positions and one stabilizes



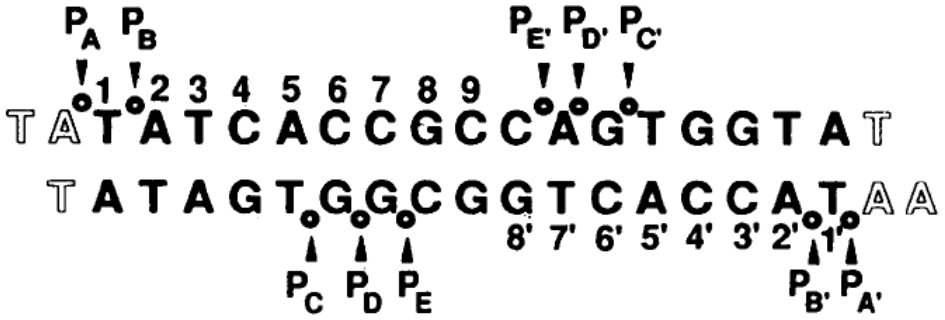
Jordan and Pabo, Science 1988

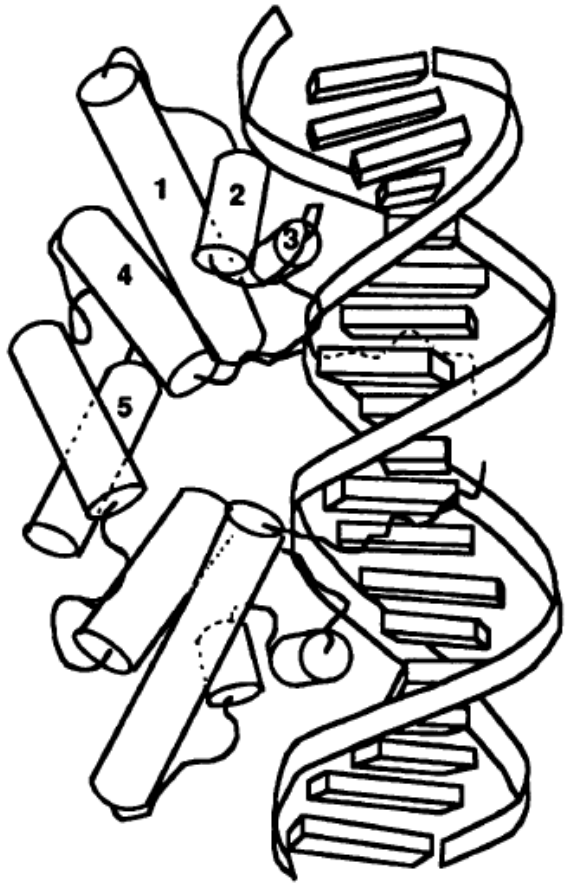
Protein-nucleic acid interactions

Helix-turn-helix motif

3 helices, one reads, one positions and one stabilizes

Dimer reads out palindromic sequence





Jordan and Pabo, Science 1988

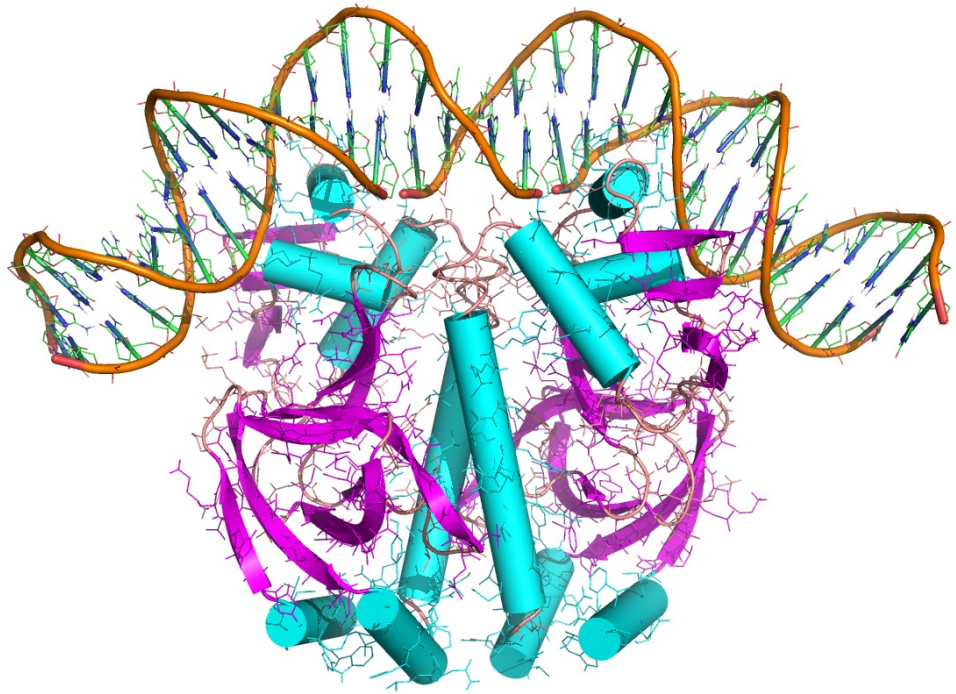
Protein-nucleic acid interactions

Helix-turn-helix motif

3 helices, one reads, one positions and one stabilizes

Dimer reads out palindromic sequence

Can bind in many ways, reading helix is often in different configurations with some HTH proteins capable of bending DNA by as much as 90 degrees (CAP-cAMP)



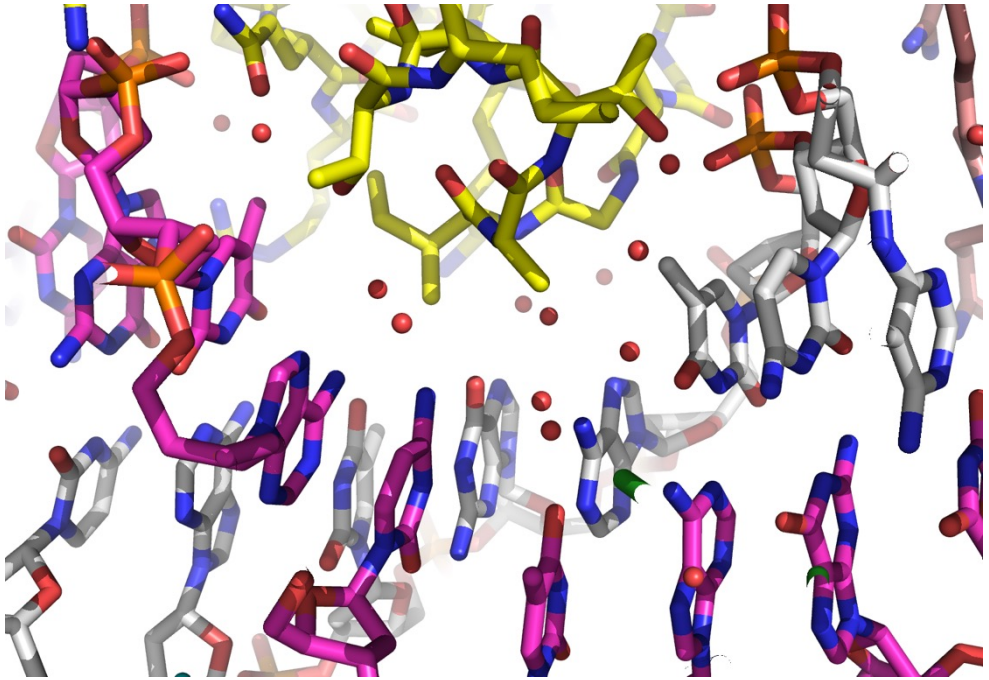
Schultz et al, Science 1991

Protein-nucleic acid interactions

Helix-turn-helix motif

Or contacts can be primarily mediated by water or hydrophobic contacts

Trp repressor shows a bit of both



Lawson and Carey, Nature 1993

Protein-nucleic acid interactions

Direct contacts between protein side chains or backbone to DNA bases

Water mediated or hydrophobic contacts

Minor groove intercalation

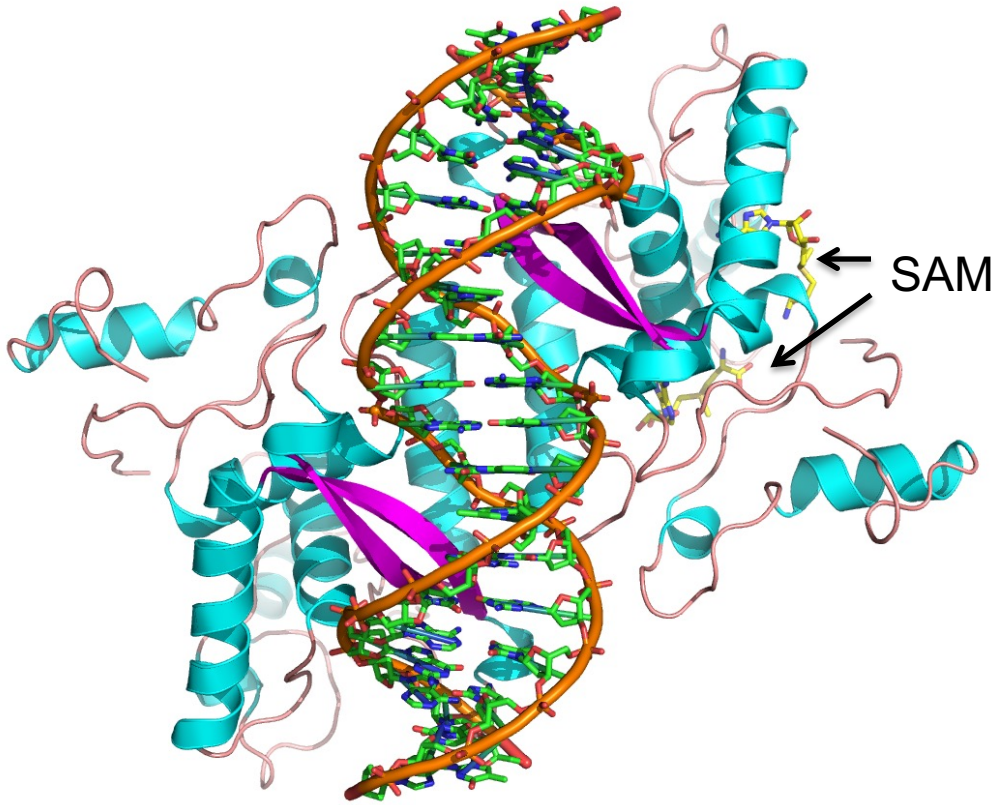
Contacts to phosphate bases

Protein-nucleic acid interactions

Beta sheet motifs

Anti-parallel pair of beta-strands can insert into the major groove

Example: Met Repressor + SAM (co-repressor)



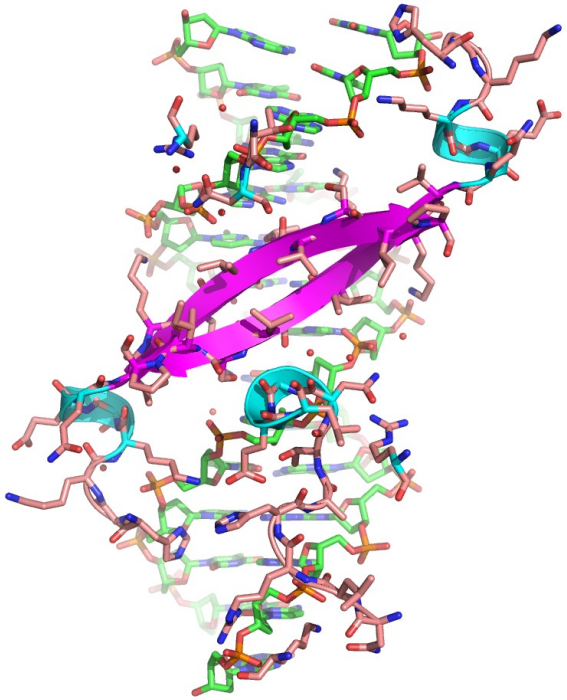
Somers and Phillips, Nature 1992

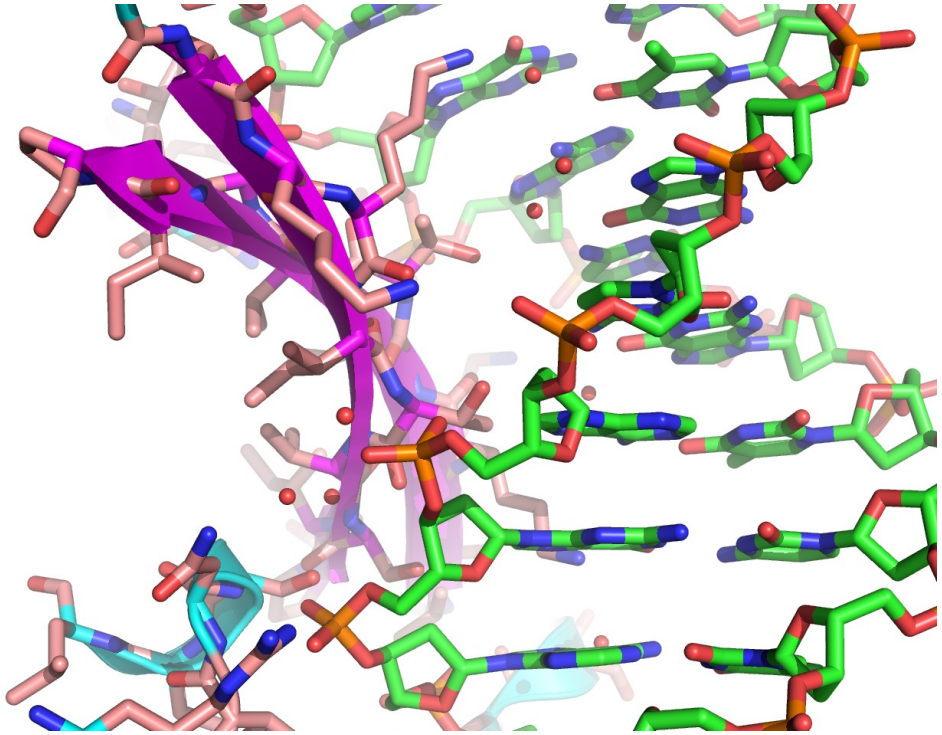
Protein-nucleic acid interactions

Beta sheet motifs

Anti-parallel pair of beta-strands can insert into the major groove

Example: Met Repressor + SAM (co-repressor)

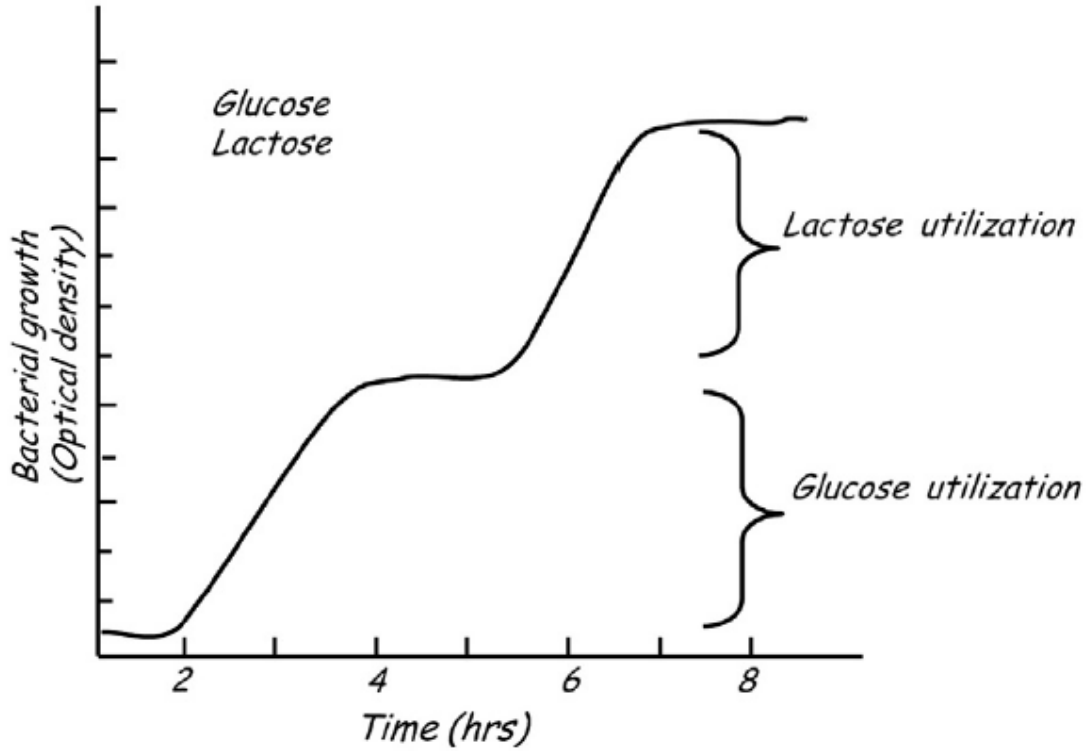




Somers and Phillips, Nature 1992

Protein-nucleic acid interactions

Allosteric control of DNA binding – Lac repressor

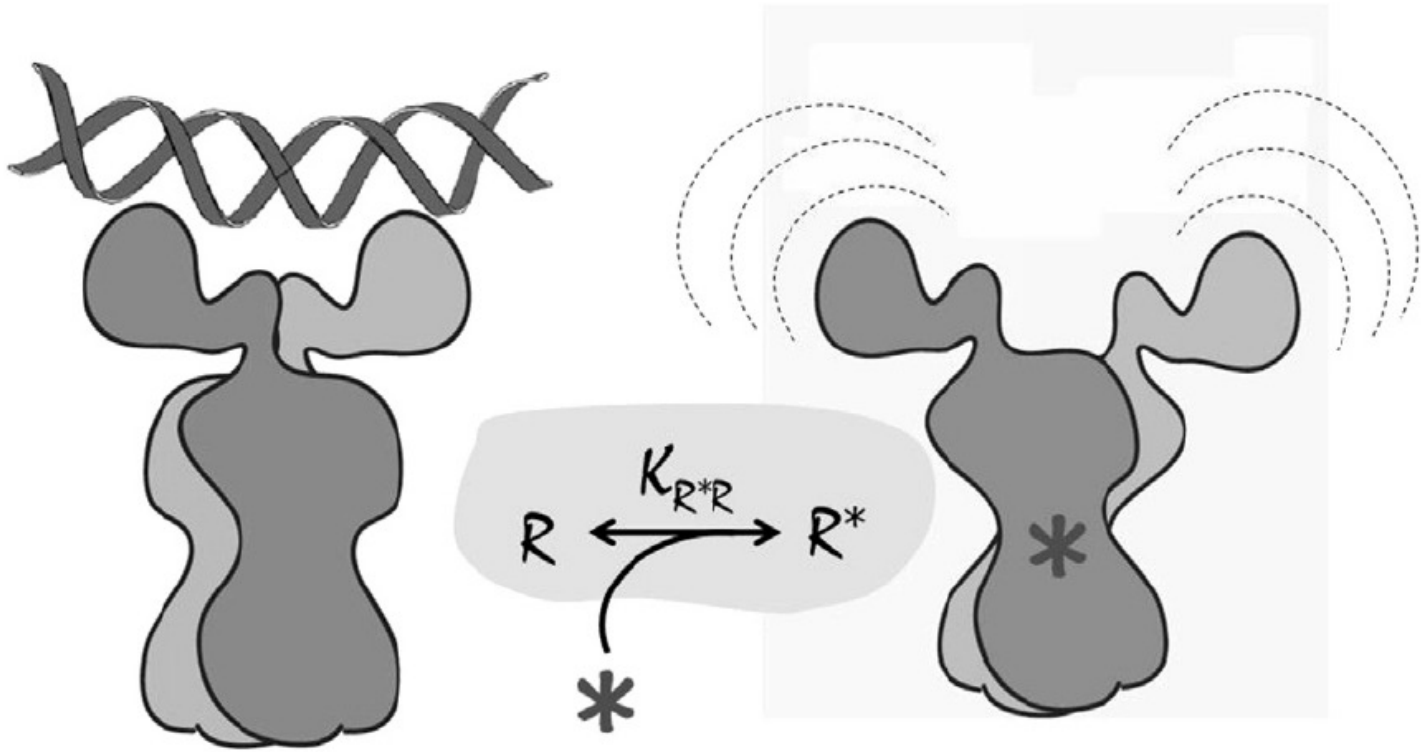


Monod, 1942

Lewis, JMB 2013

Protein-nucleic acid interactions

Allosteric control of DNA binding – Lac repressor

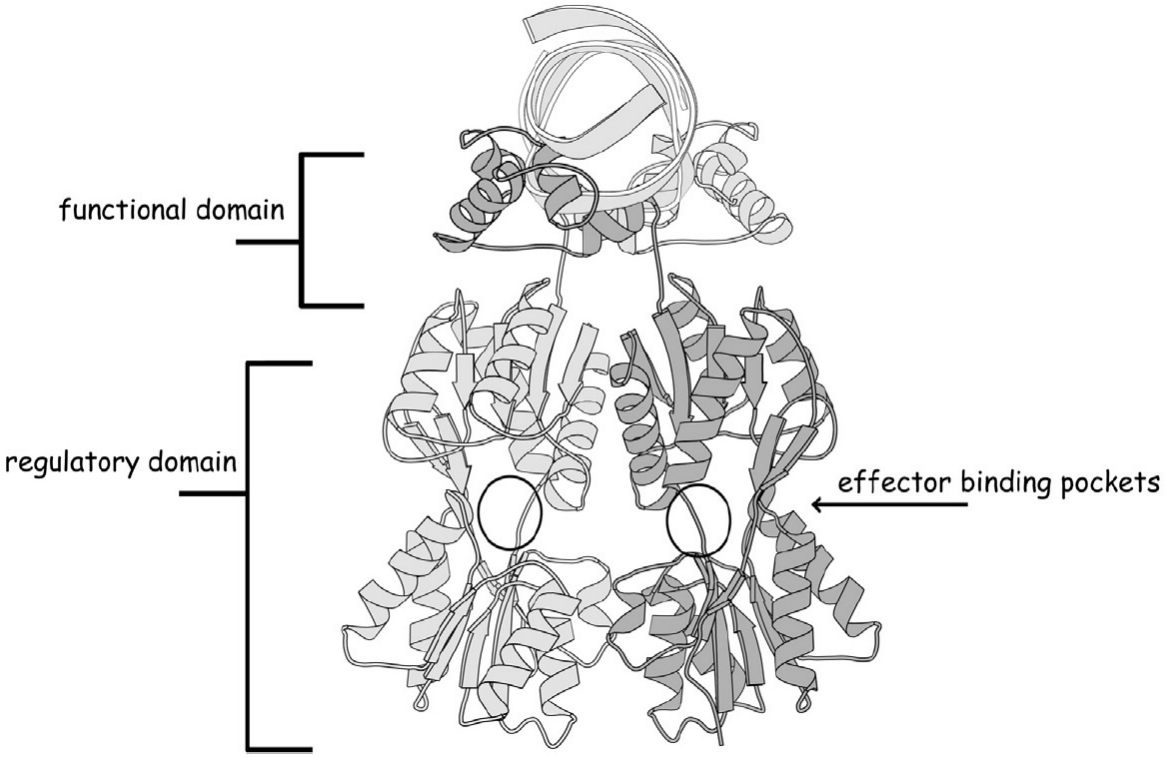


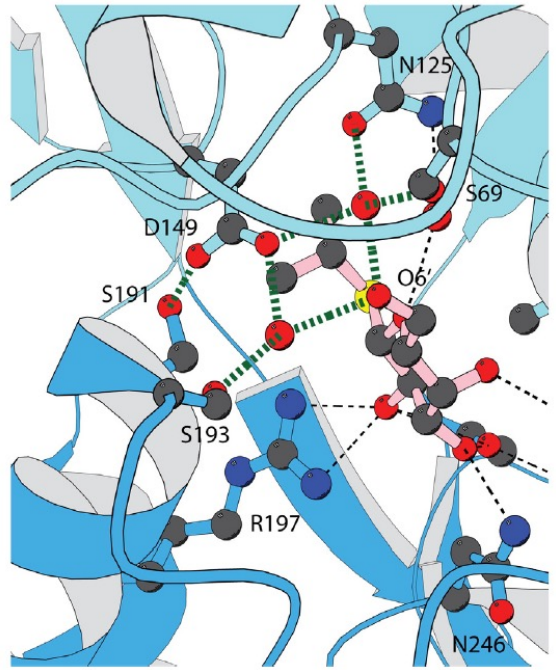
* allolactose, a lactose metabolite that triggers expression of the lac operon

Lewis, JMB 2013

Protein-nucleic acid interactions

Allosteric control of DNA binding – Lac repressor

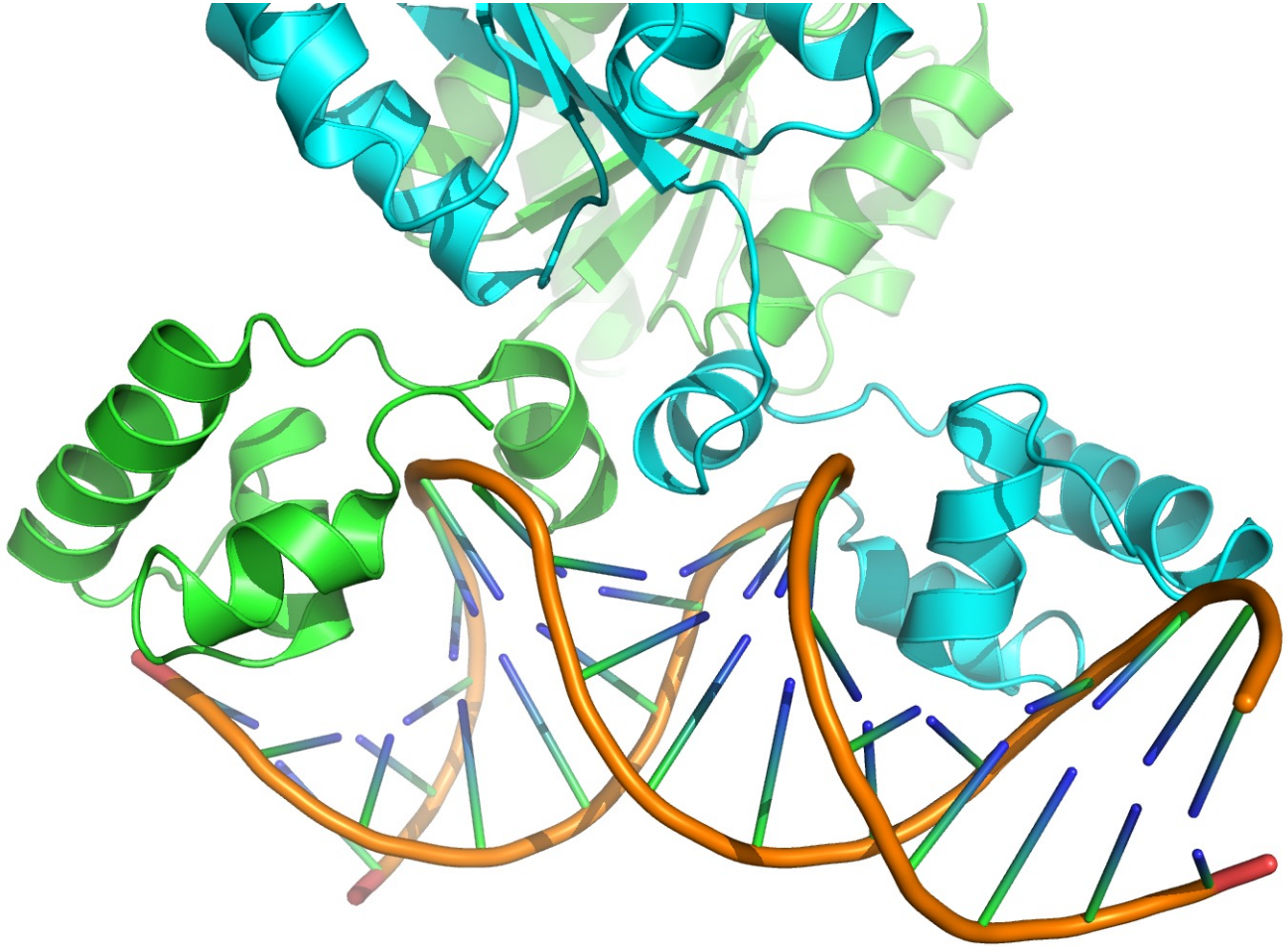




Lewis, JMB 2013

Protein-nucleic acid interactions

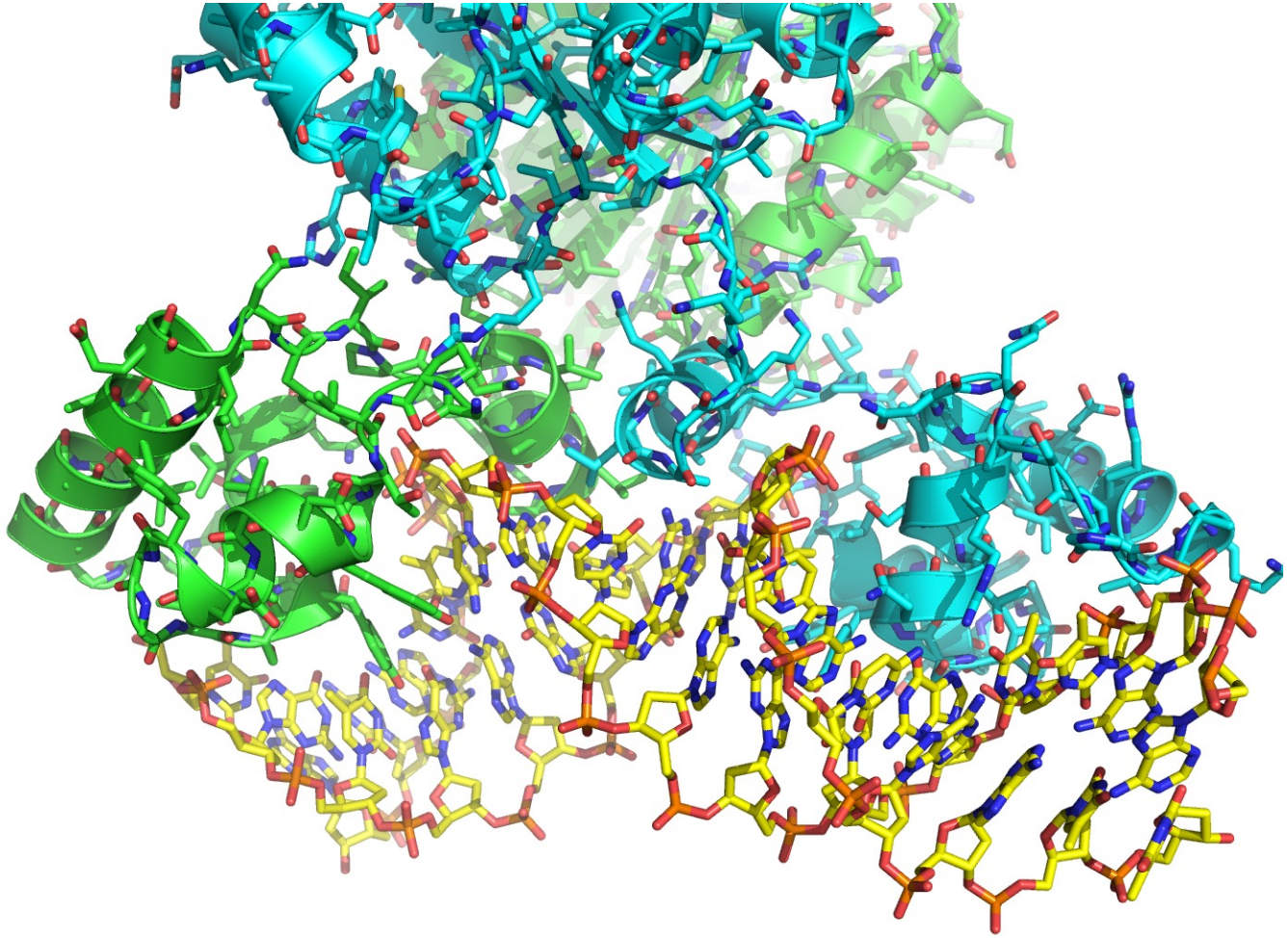
Allosteric control of DNA binding – Lac repressor



Bell and Lewis, NSB 2000

Protein-nucleic acid interactions

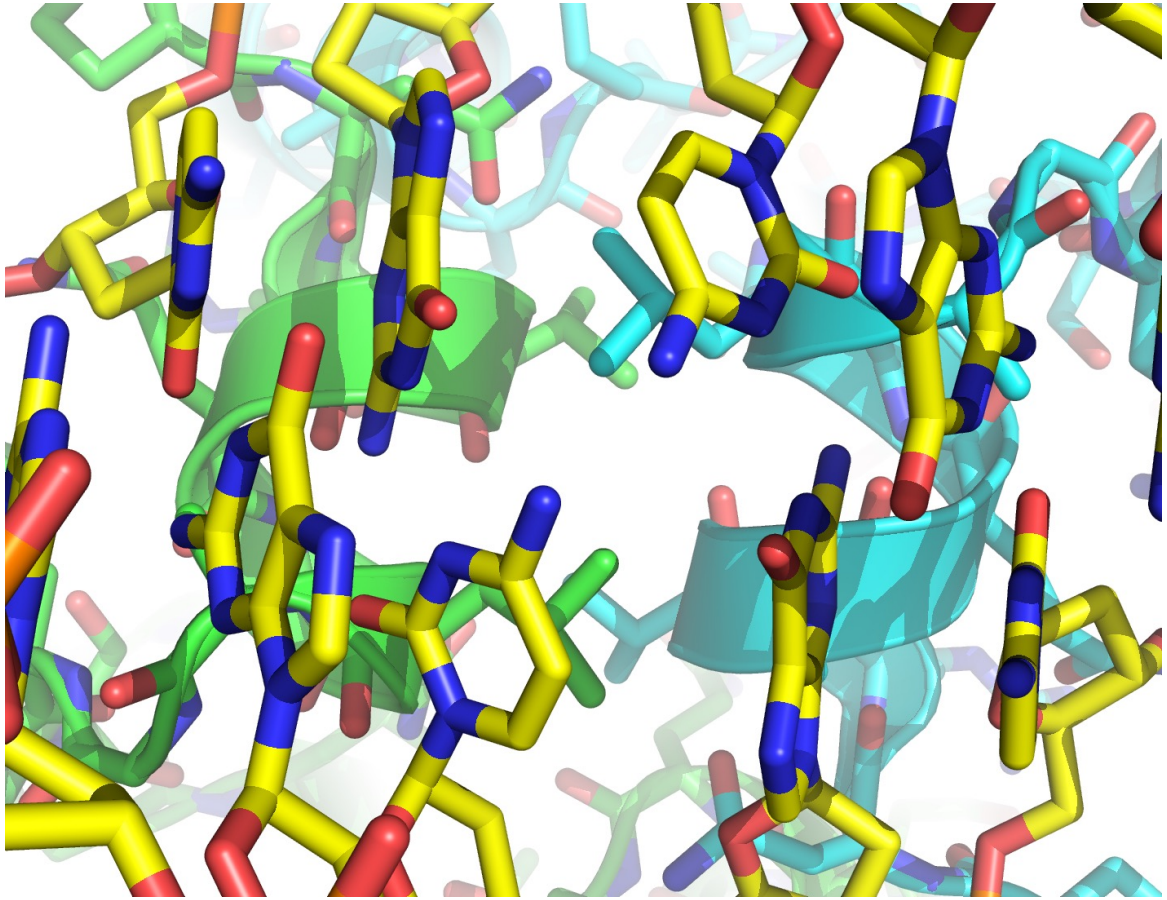
Allosteric control of DNA binding – Lac repressor



Bell and Lewis, NSB 2000

Protein-nucleic acid interactions

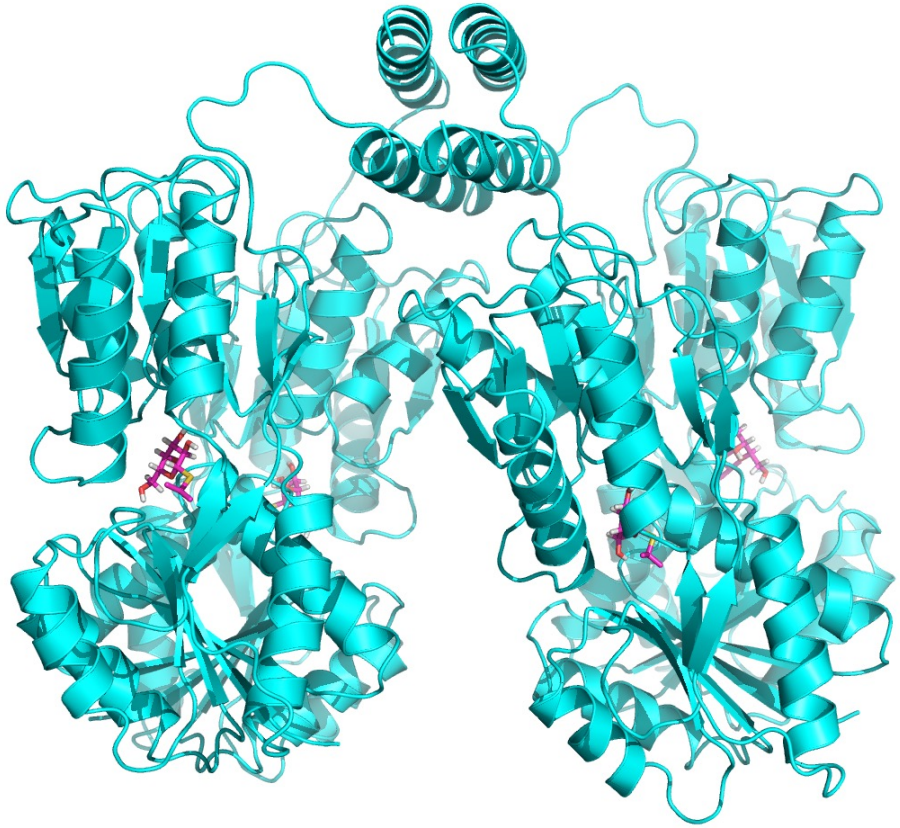
Allosteric control of DNA binding – Lac repressor



Bell and Lewis, NSB 2000

Protein-nucleic acid interactions

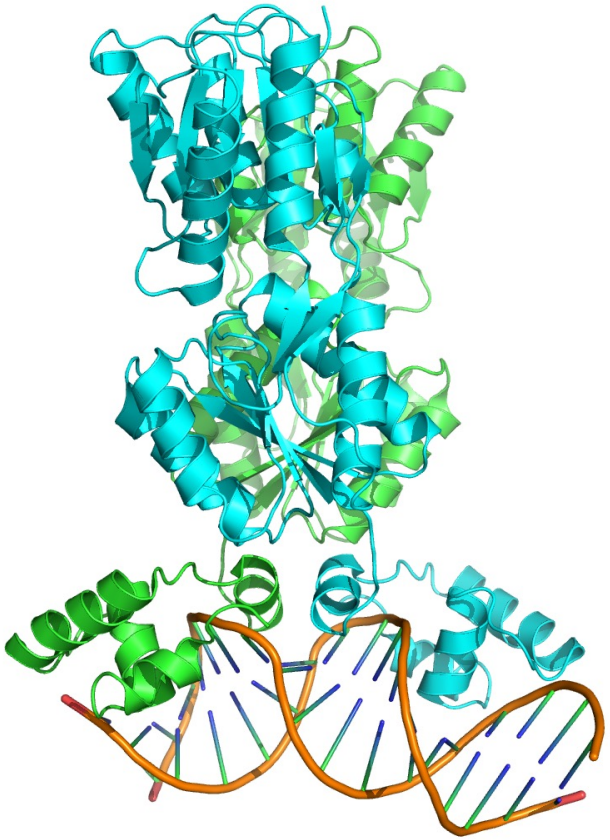
Allosteric control of DNA binding – Lac repressor – tetramerization domain

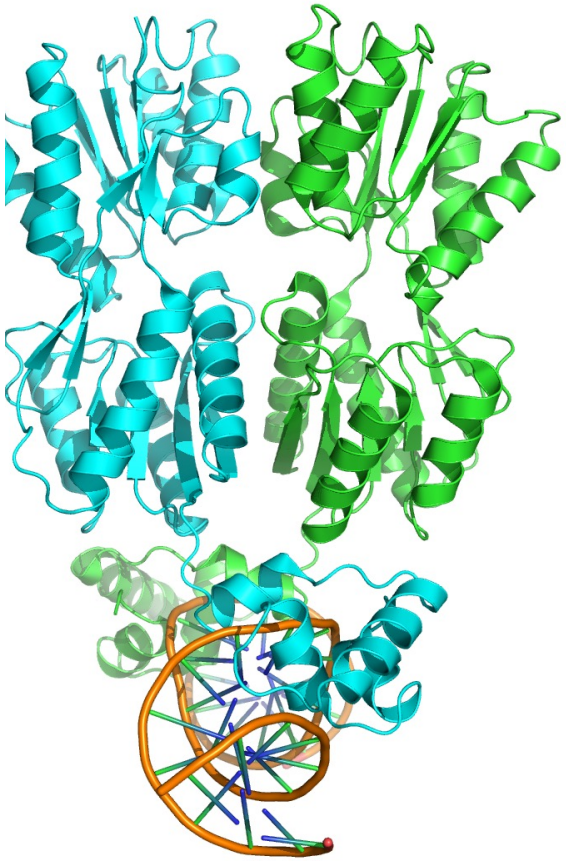


Friedman et al., Science 1995

Protein-nucleic acid interactions

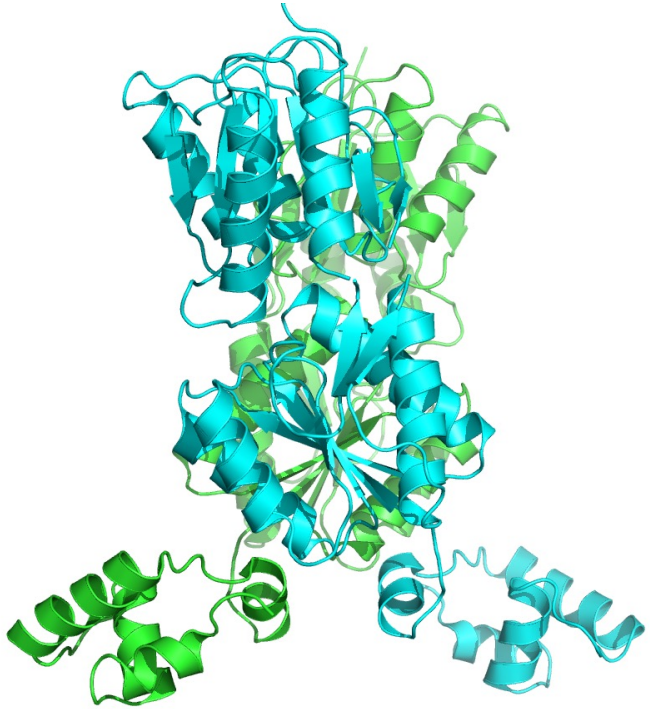
Allosteric control of DNA binding – Lac repressor

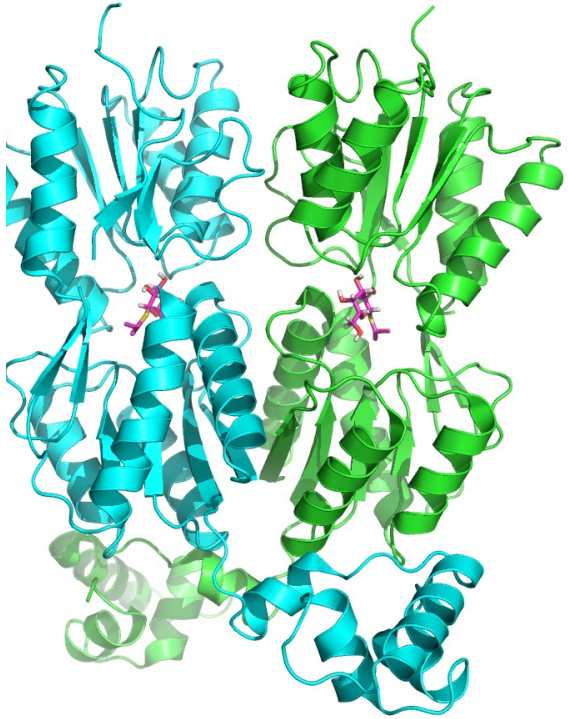




Protein-nucleic acid interactions

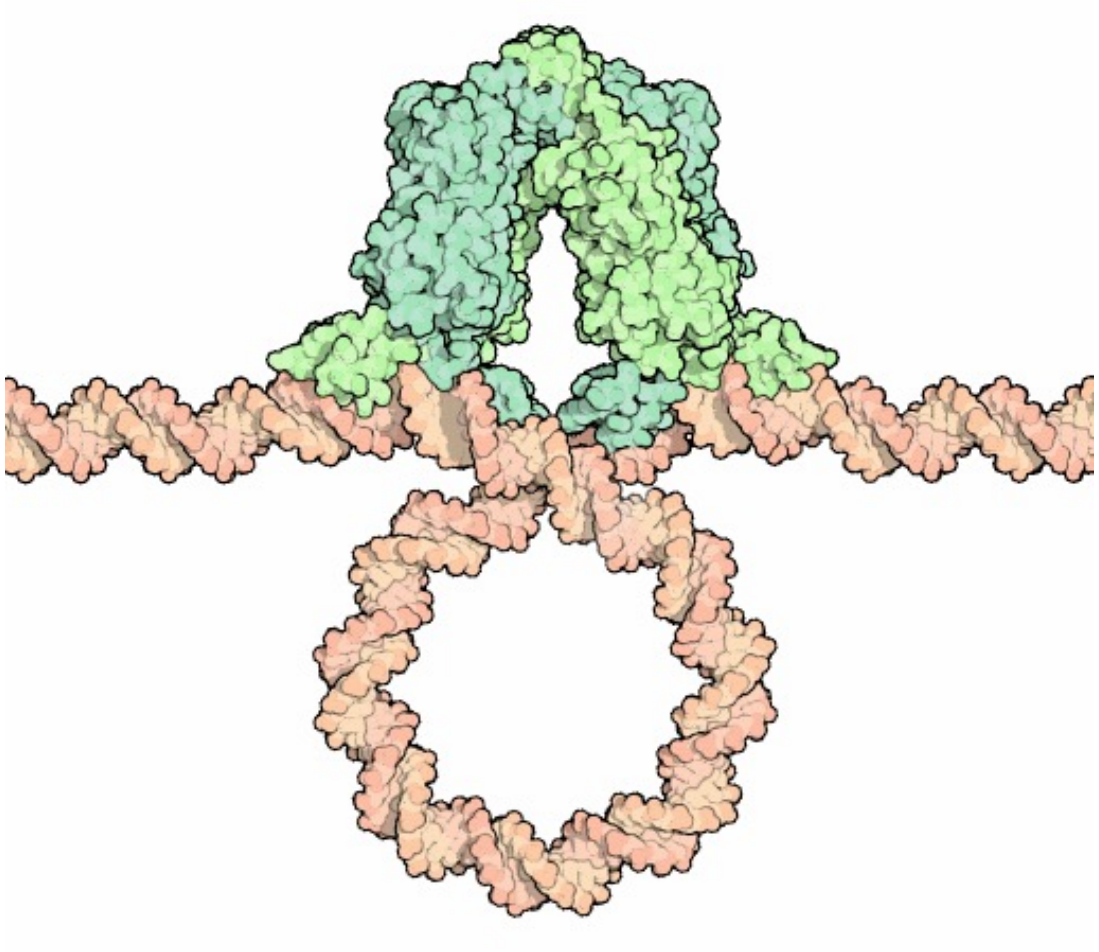
Allosteric control of DNA binding – Lac repressor





Protein-nucleic acid interactions

Allosteric control of DNA binding – Lac repressor



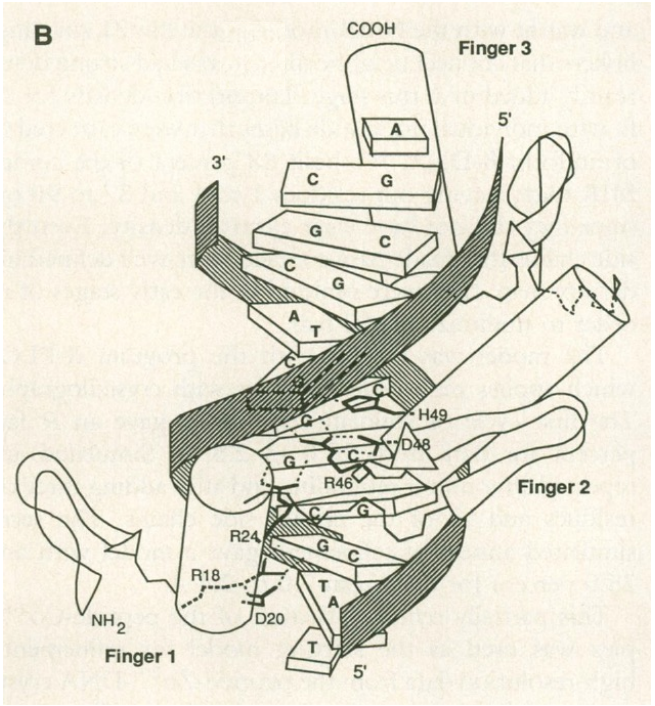
David Goodsell, RCSB

Protein-nucleic acid interactions

Zinc finger motifs

Several types include C2-H2, C2-C2, binuclear C6

Code and readout



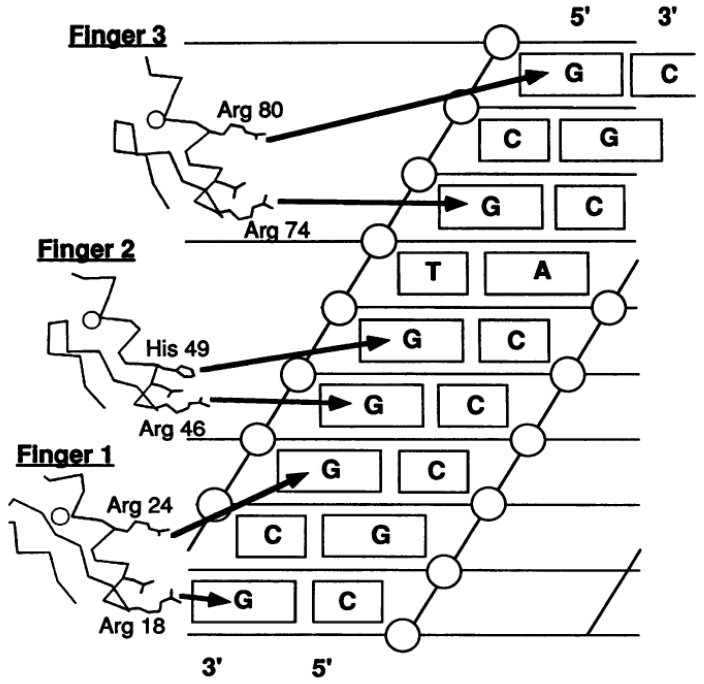


Fig. 5. Sketch summarizing all the base contacts made by the Zif268 peptide. The DNA is represented as a cylindrical projection.

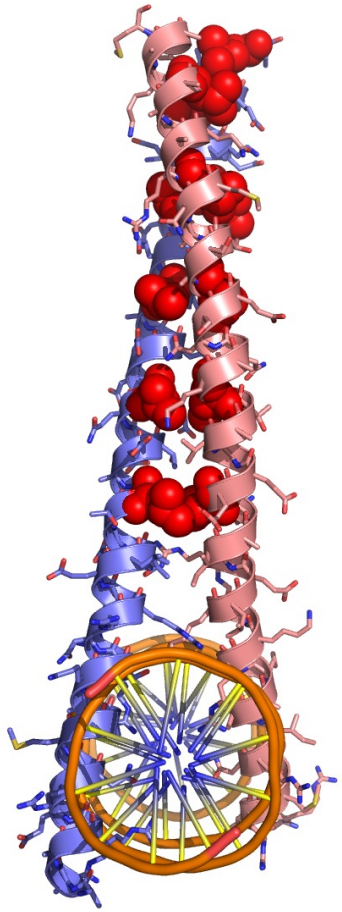
Pavletich and Pabo, Science 1991

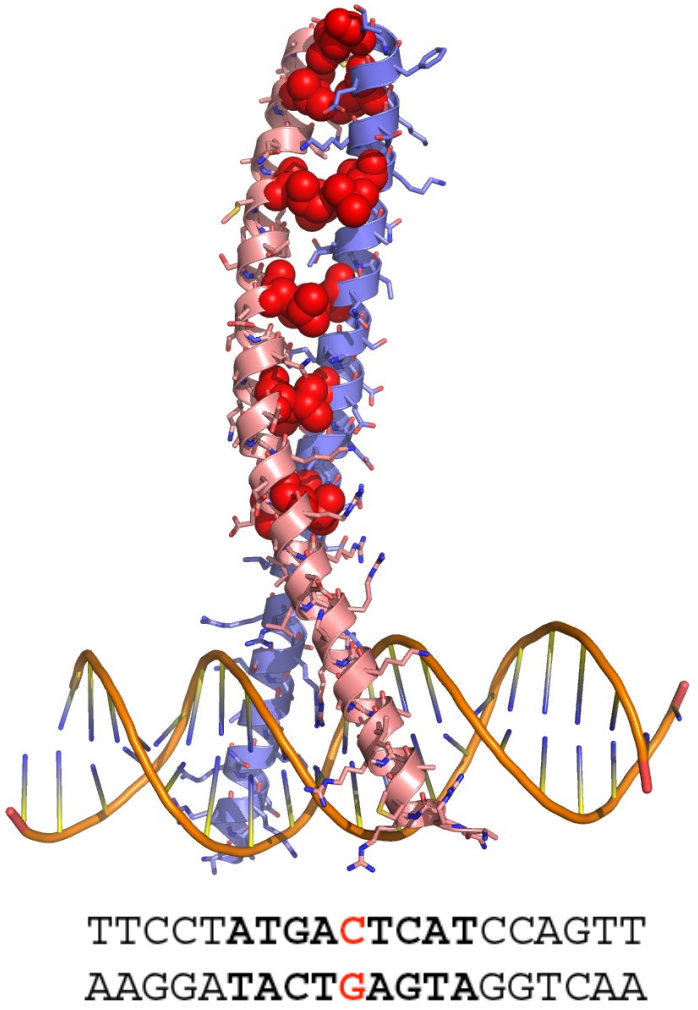
Protein-nucleic acid interactions

Leucine zipper

GCN4 (homodimer)

Jun-Fos (heterodimer)





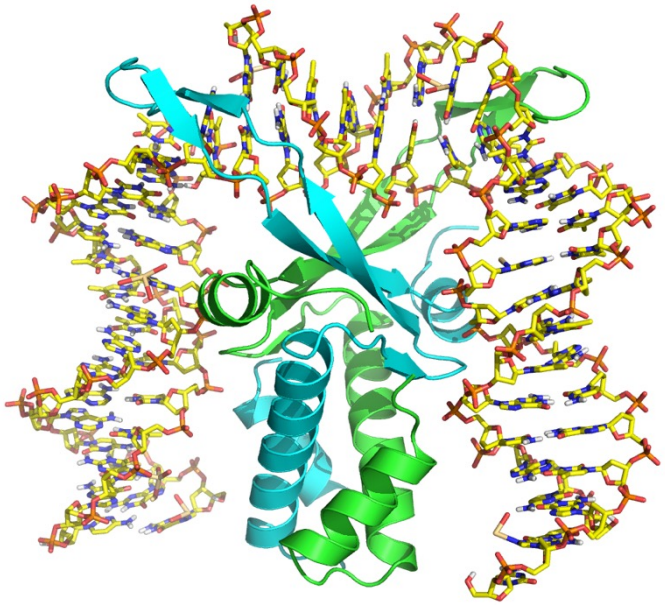
Glover and Harrison, Nature 1995

Protein-nucleic acid interactions

Beta ribbon proteins

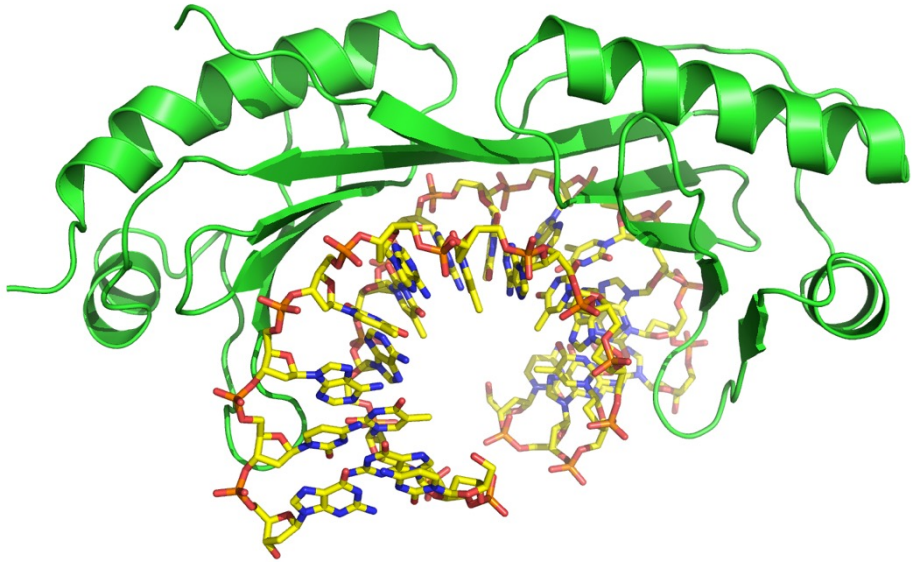
Use in bending DNA

IHF and TATA binding protein



IHF

Rice et al., Cell 1996



Tata-binding protein

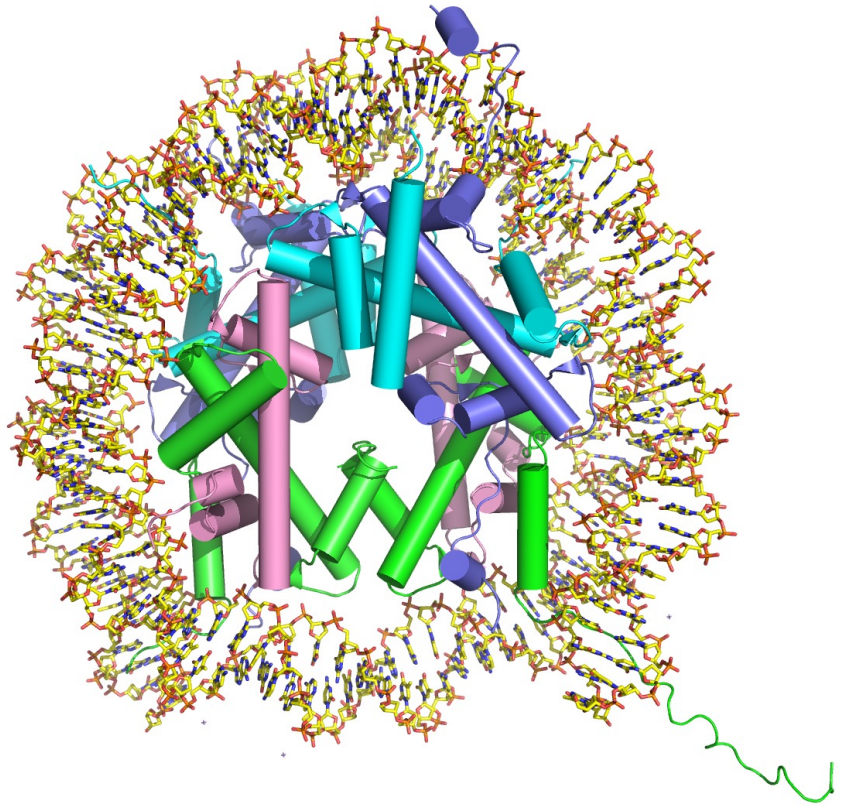
Burley, Richmond, Sigler 1990s

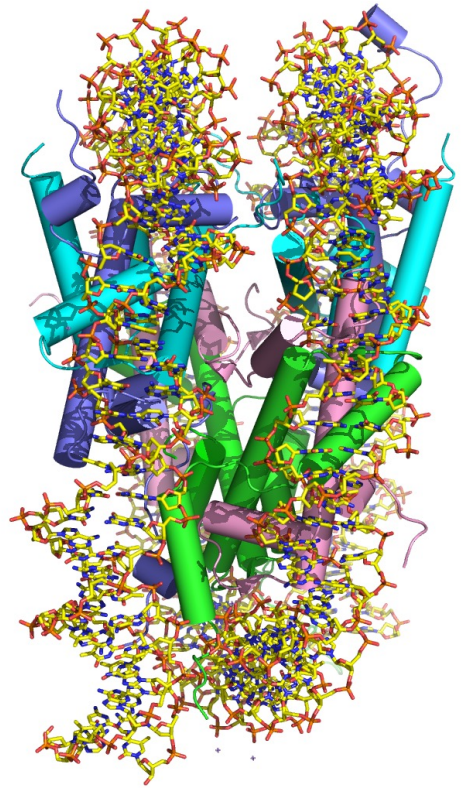
Protein-nucleic acid interactions

Nucleosomes

Bending DNA

Preferred binding sites relies on deformability of the DNA



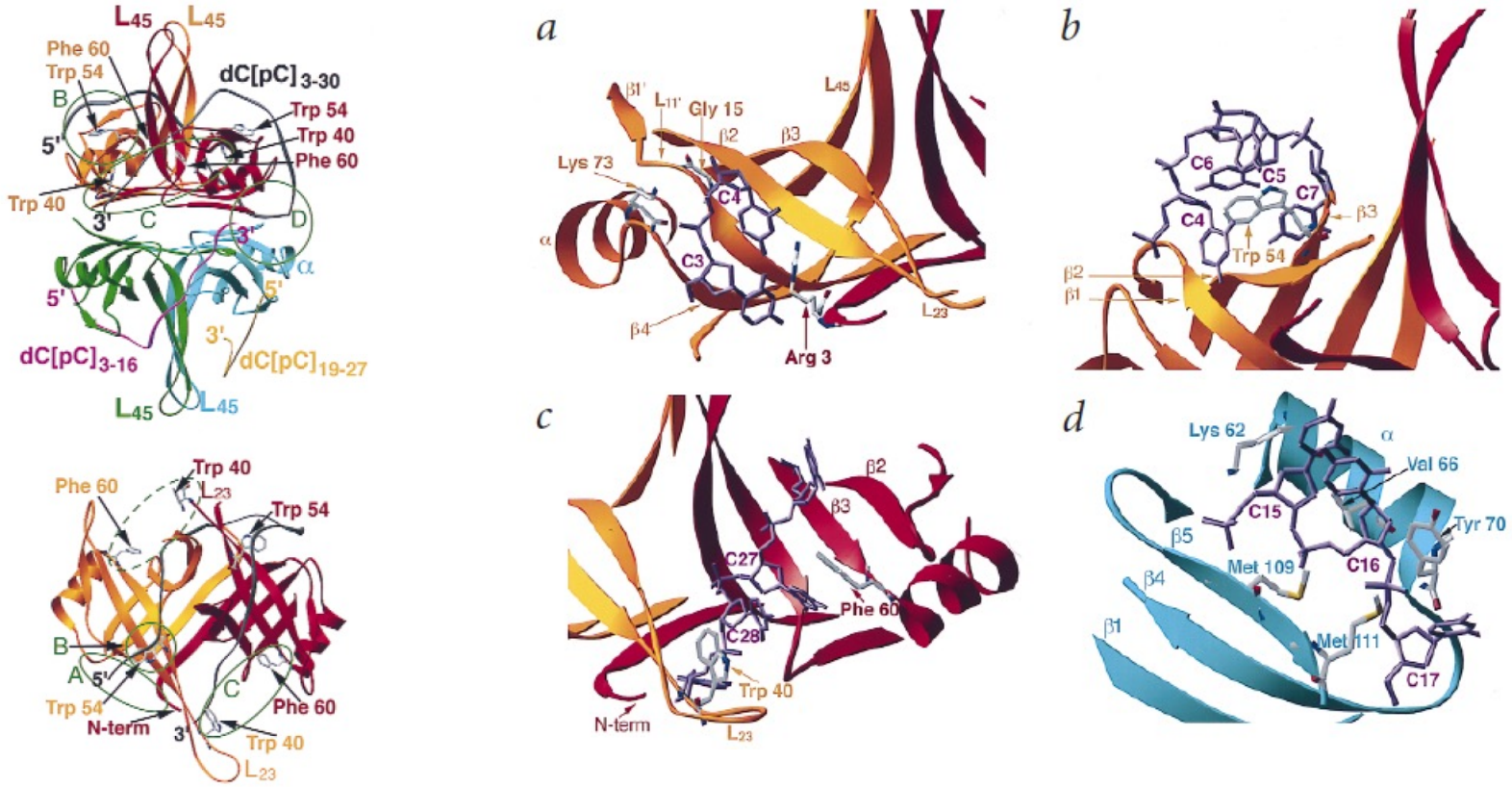


Luger et al., Nature 1998

Protein-nucleic acid interactions

Single stranded DNA binding proteins

SSB interactions with ssDNA engage base edges

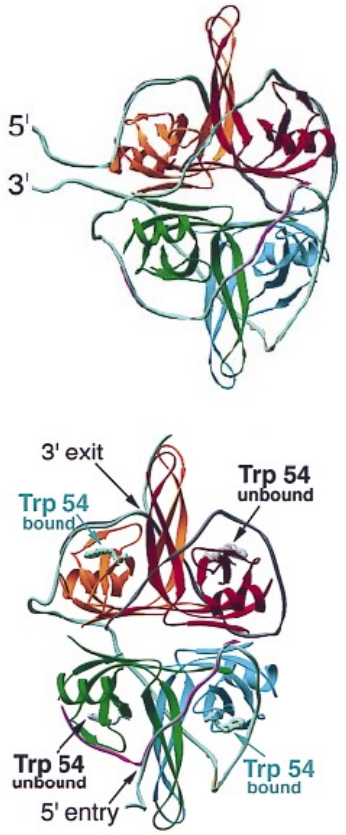


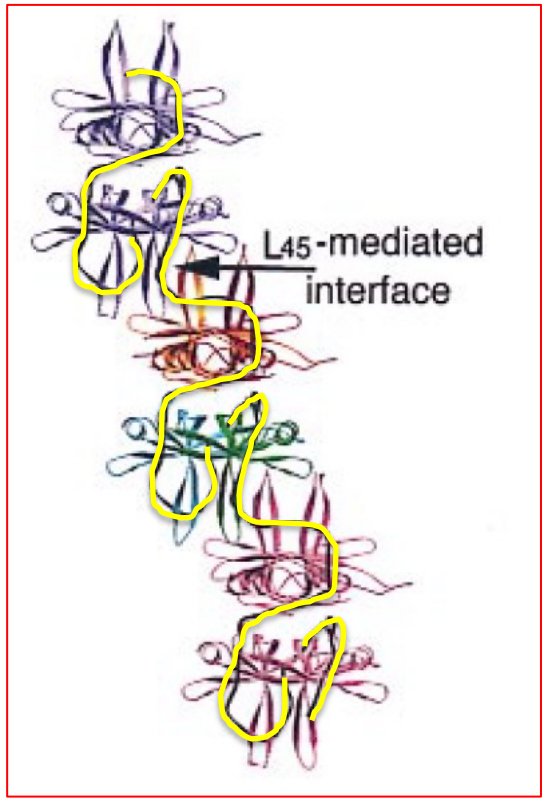
Raghunathan et al., NSMB 2000

Protein-nucleic acid interactions

Single stranded DNA binding proteins

SSB may package ssDNA to prevent interactions



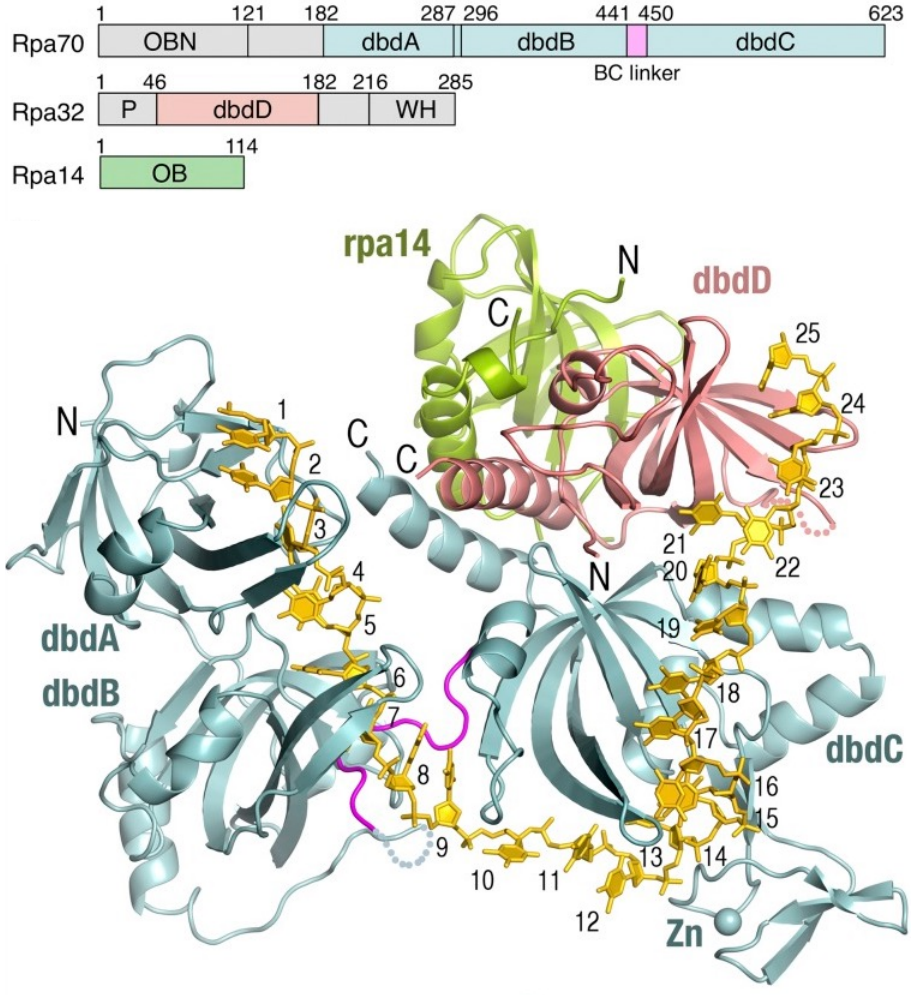


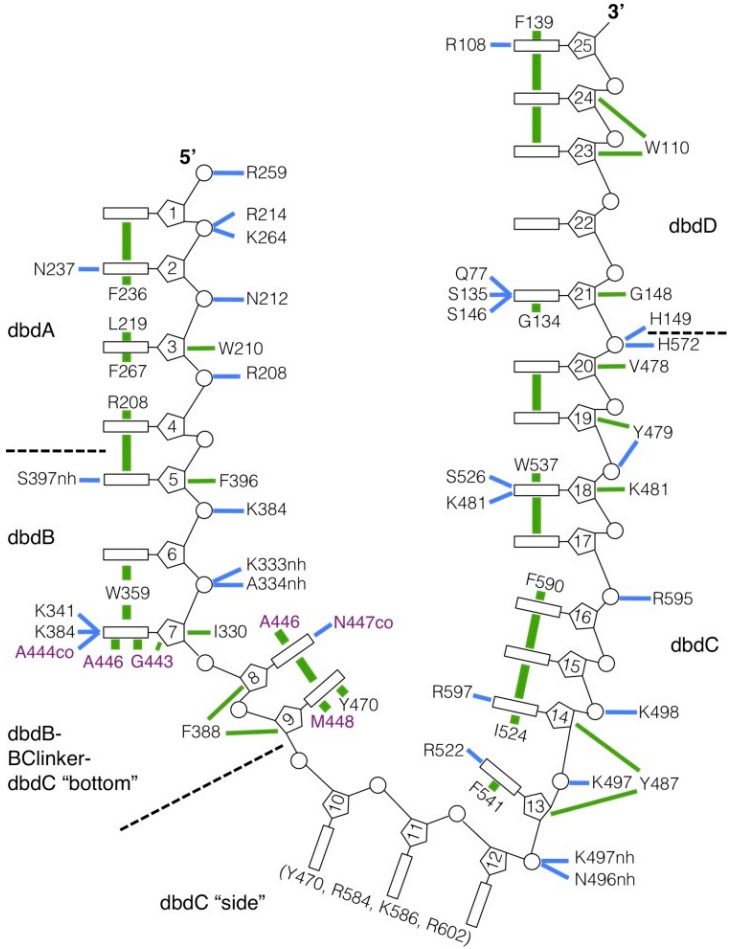
Raghunathan et al., NSMB 2000

Protein-nucleic acid interactions

Single stranded DNA binding proteins

RPA interactions with ssDNA engage base edges



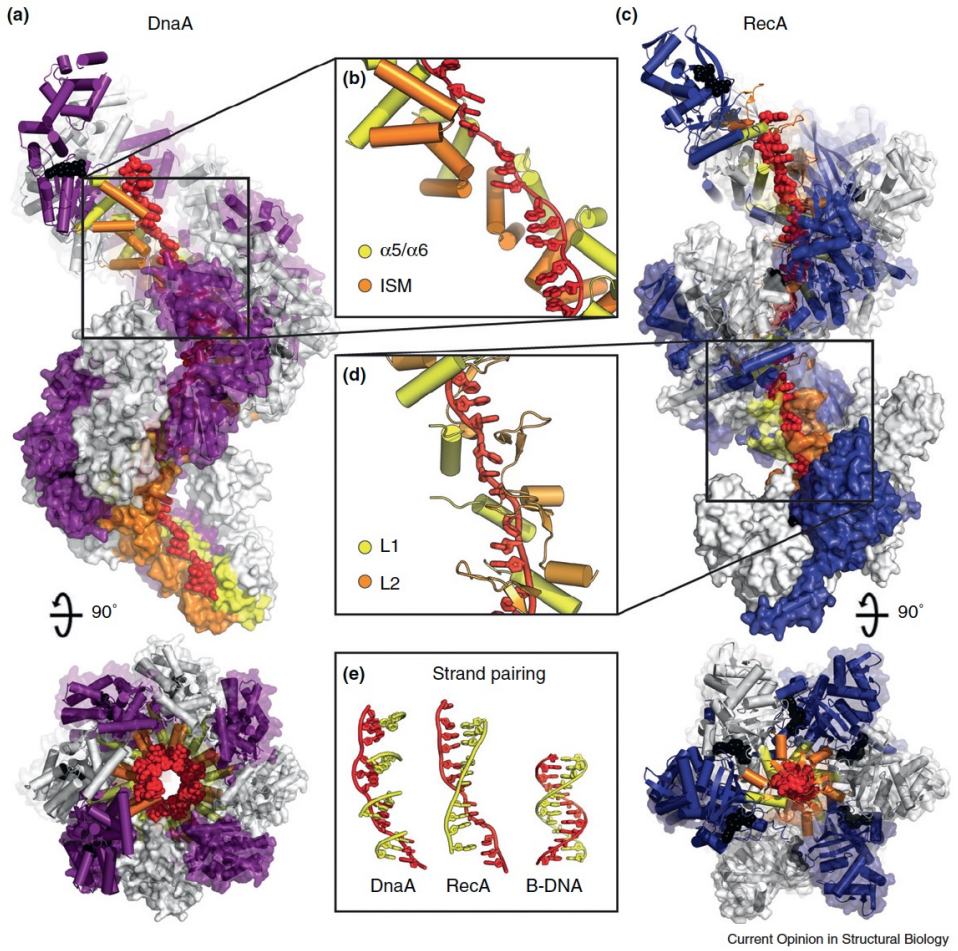


Fan and Pavletich, G&D 2012

Protein-nucleic acid interactions

Single stranded DNA binding proteins

DnaA and RecA – ATPases that bind ssDNA to expose base edges

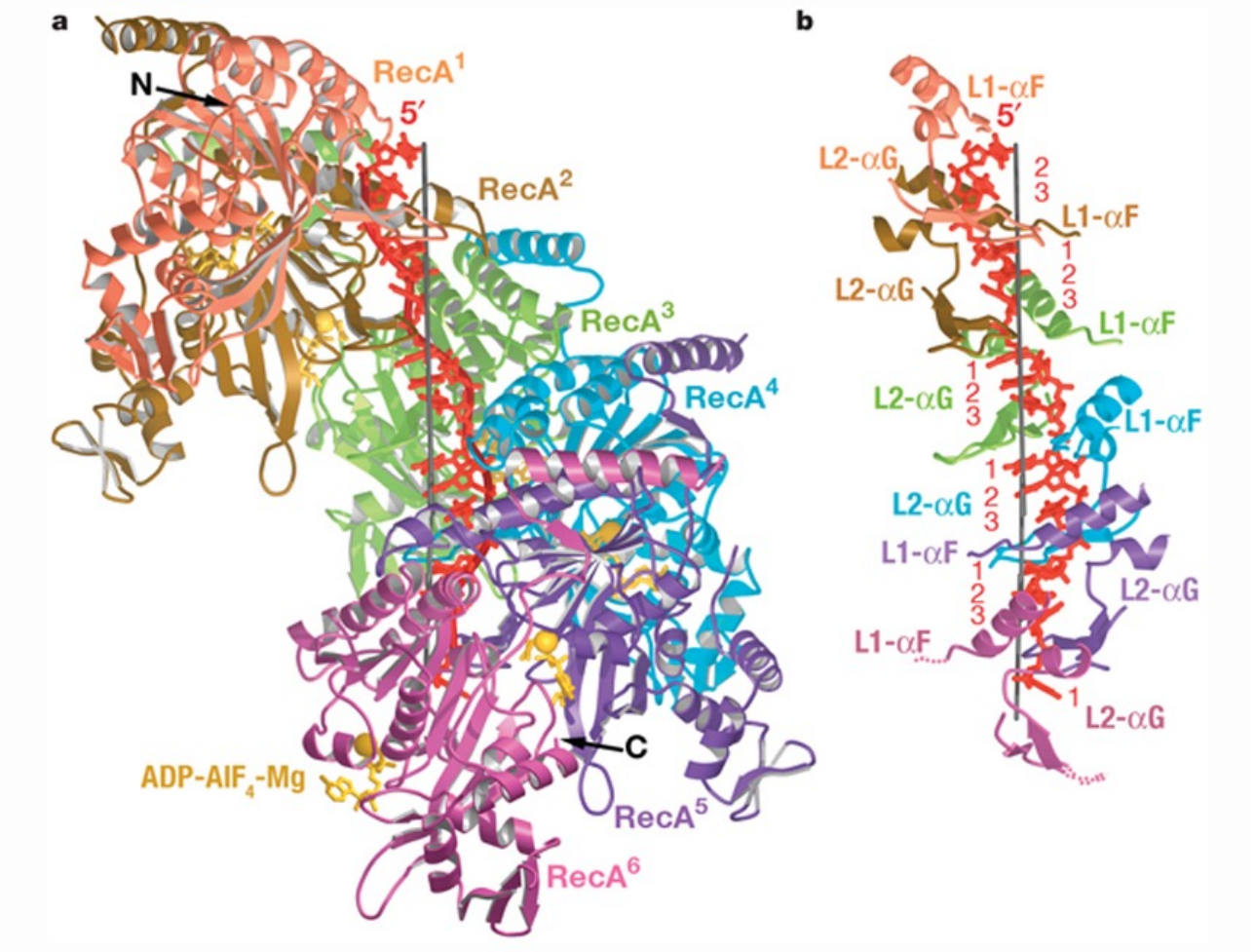


Duderstadt and Berger, 2013

Protein-nucleic acid interactions

Single stranded DNA binding proteins

RecA – Structure of the presynaptic filament

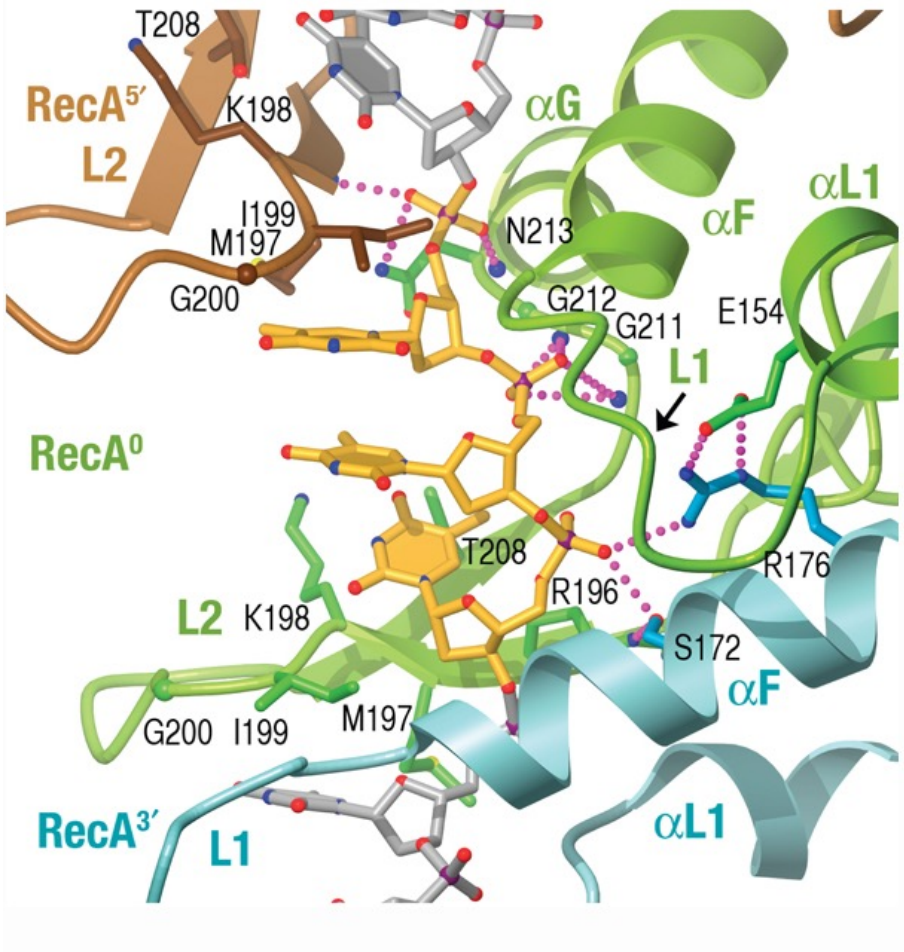


Chen et al., Nature 2008

Protein-nucleic acid interactions

Single stranded DNA binding proteins

RecA – each nucleotide triplet bound by 3 consecutive RecA protomers

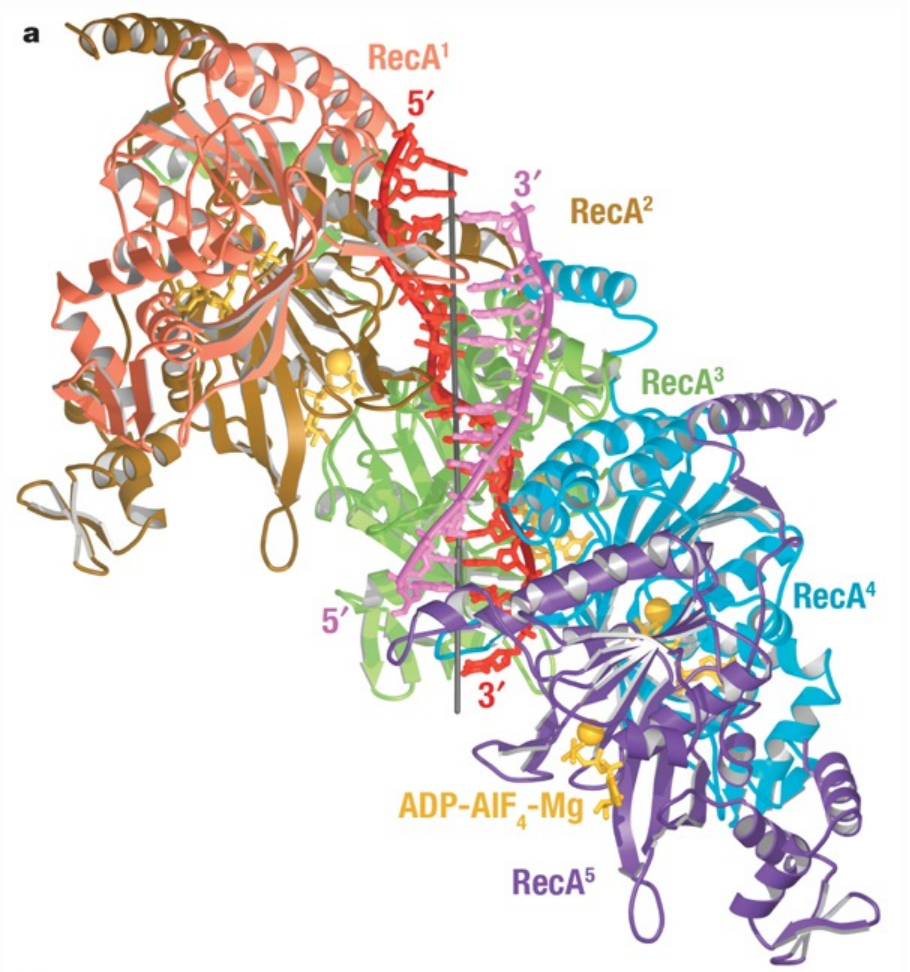


Chen et al., Nature 2008

Protein-nucleic acid interactions

Single stranded DNA binding proteins

RecA – Structure of the postsynaptic nucleoprotein filament

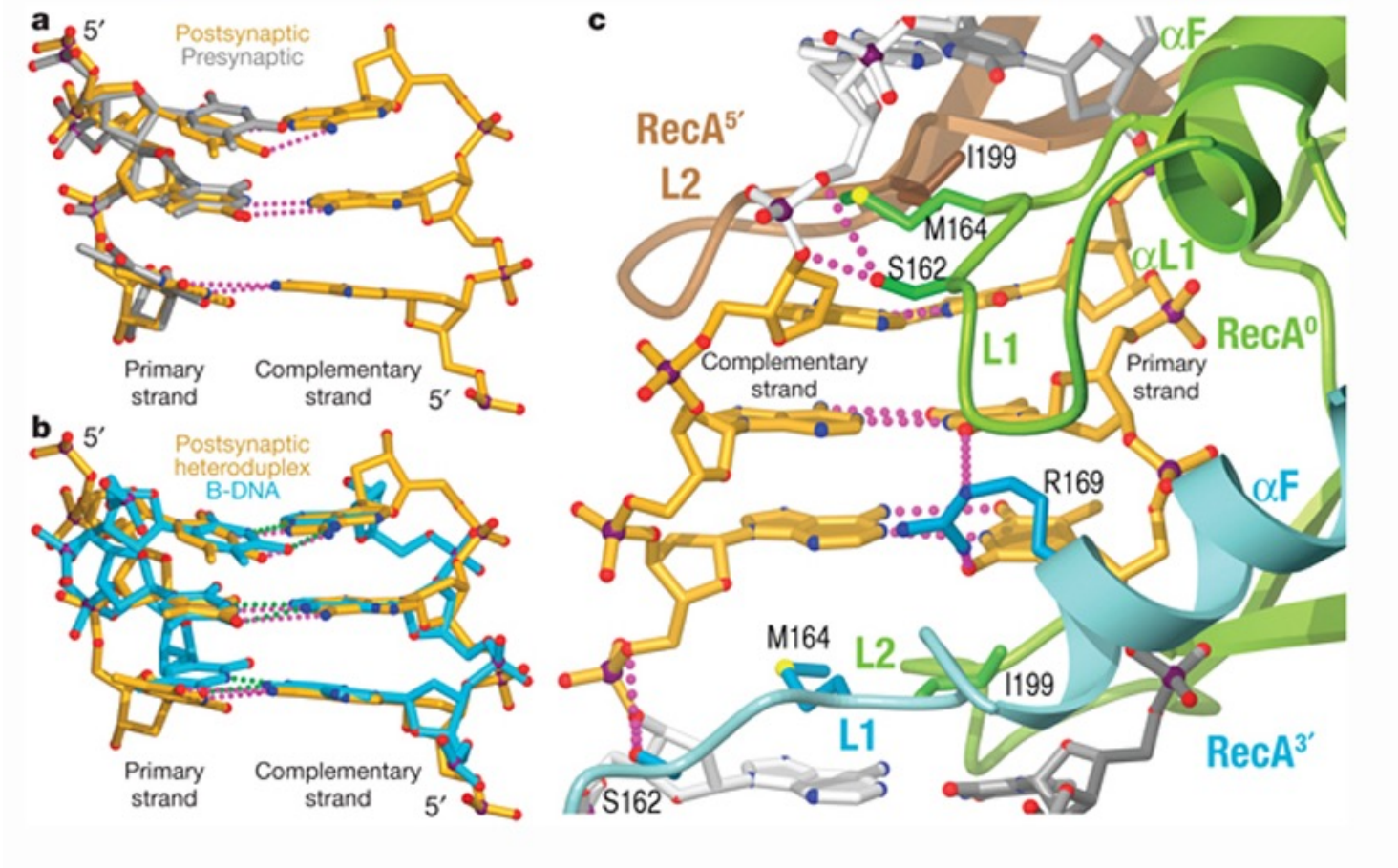


Chen et al., Nature 2008

Protein-nucleic acid interactions

Single stranded DNA binding proteins

RecA – strand pairing

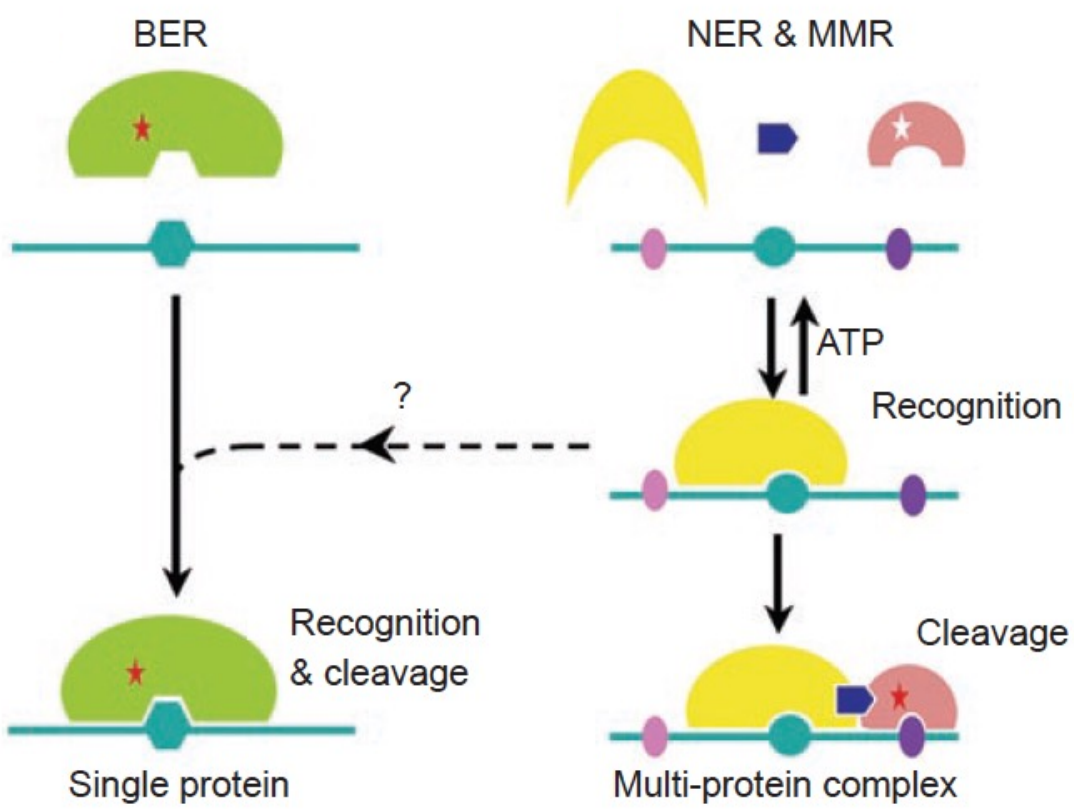


Chen et al., Nature 2008

Protein-nucleic acid interactions

Damage recognition

Base excision repair, nucleotide excision repair, mismatch repair

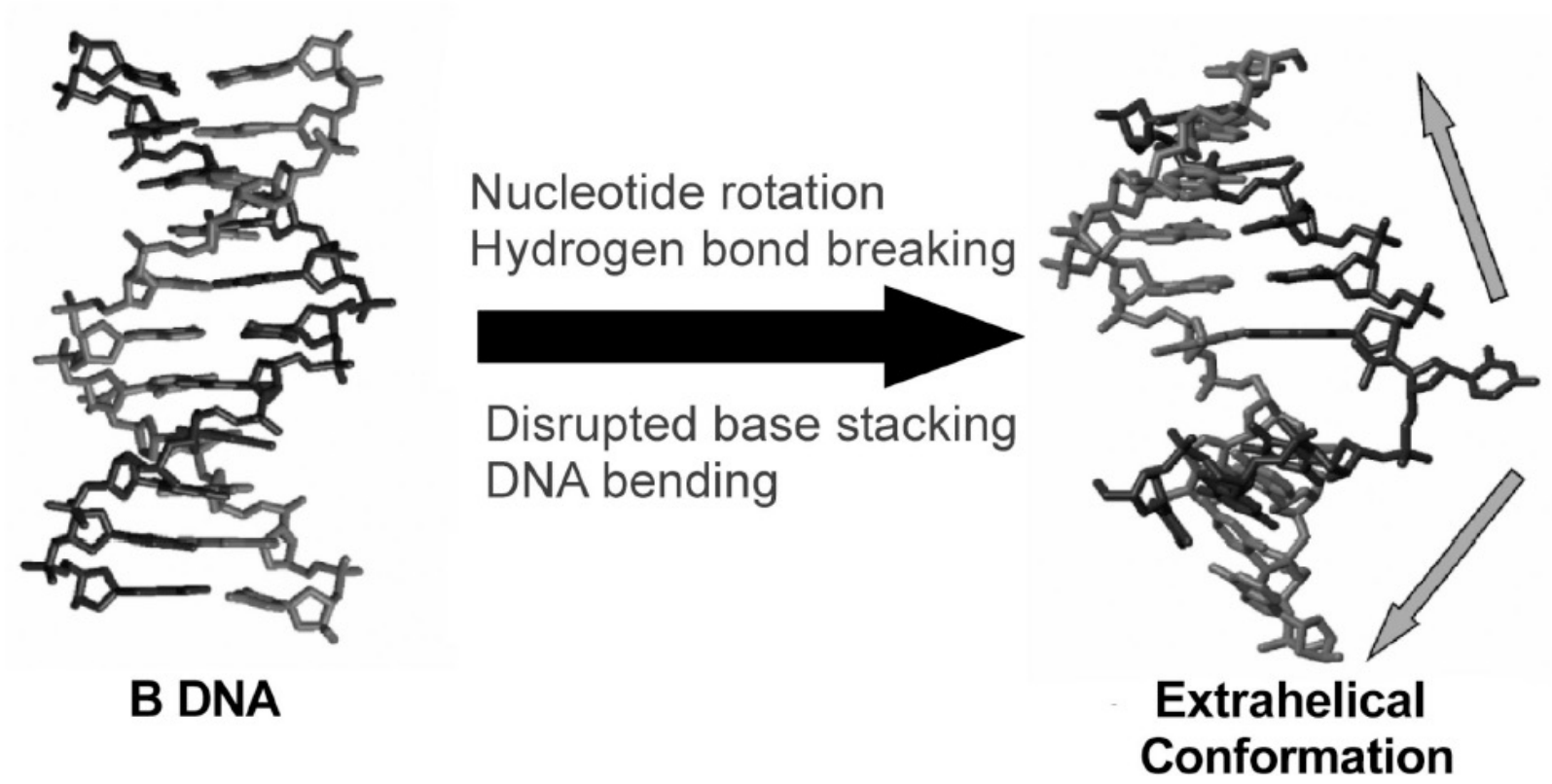


Yang, Cell Research 2008

Protein-nucleic acid interactions

Damage recognition

Conformational states in equilibrium, damage will distort DNA or enhance flipping rates

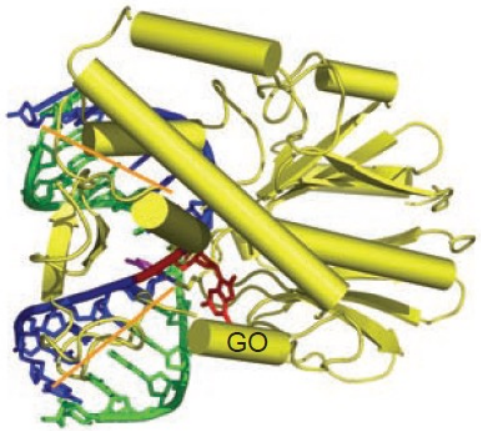


Stivers, Chemistry 2008

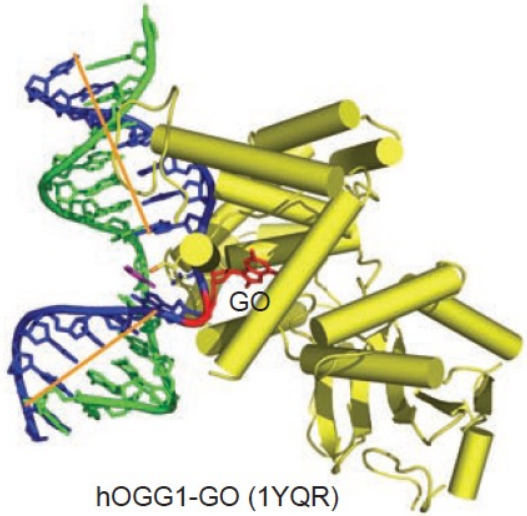
Protein-nucleic acid interactions

Damage recognition

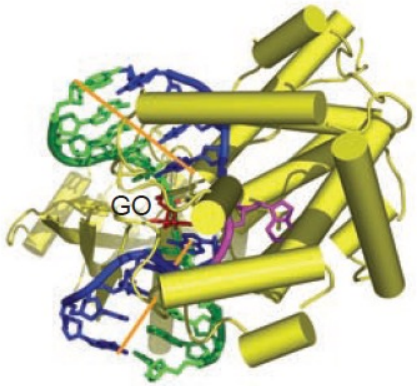
Nucleotide flipping observed for DNA glycosylases



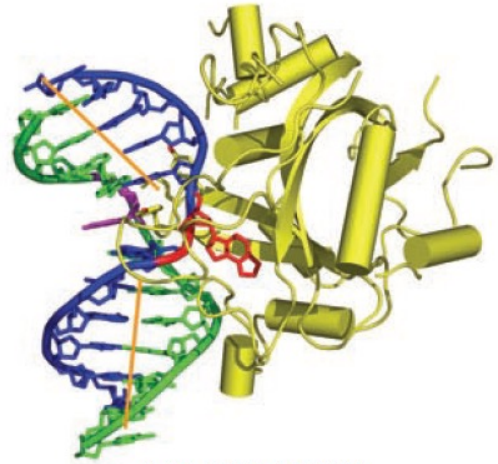
MutM-GO (1R2Y)



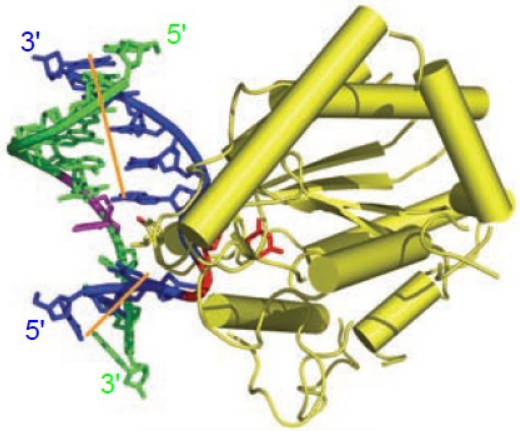
hOGG1-GO (1YQR)



MutY-GO (1RRQ)



AAG-DNA (1EWN)



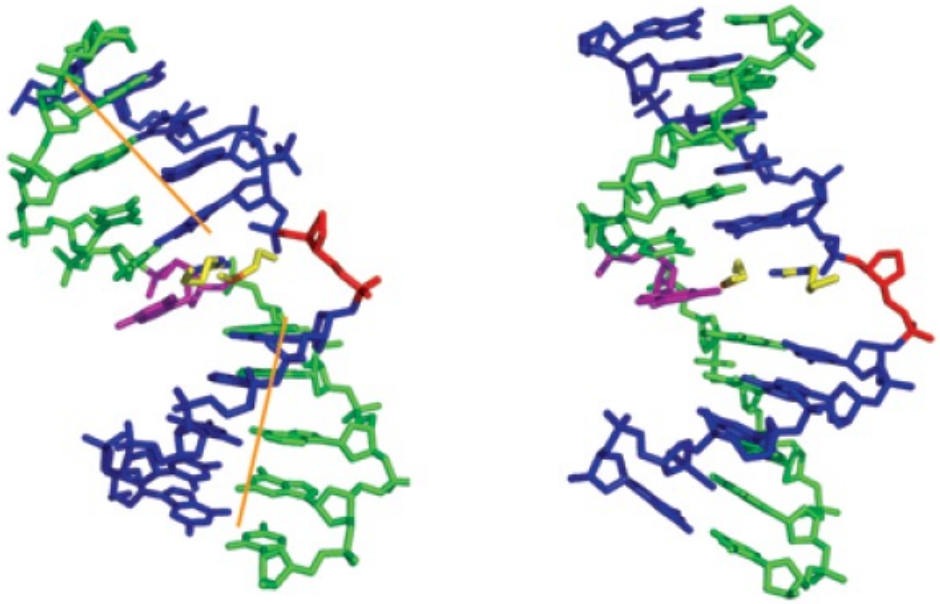
UDG-DNA (1EMJ)

Yang, Cell Research 2008

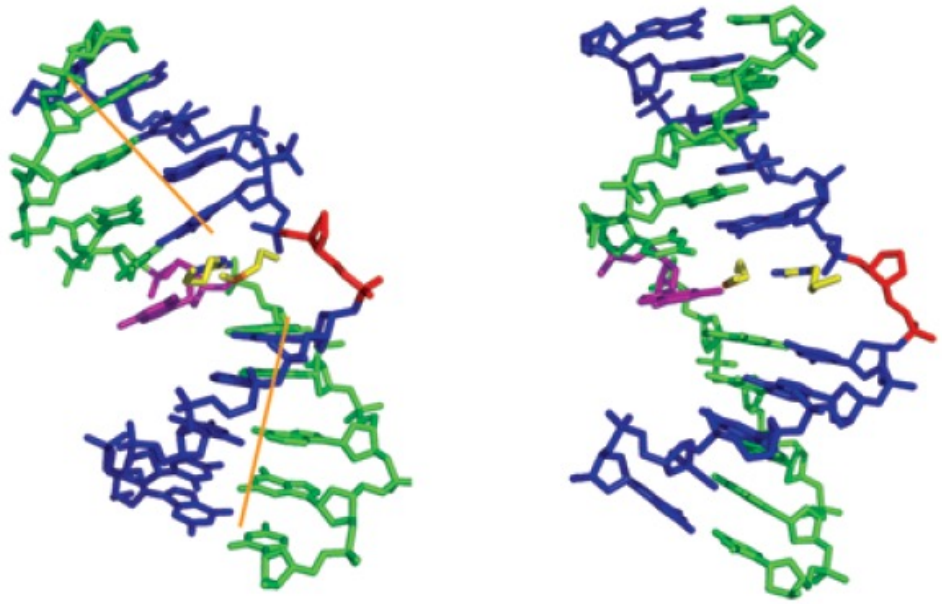
Protein-nucleic acid interactions

Damage recognition

AP endonucleases also utilize nucleotide flipping, side chains insert into DNA



DNA bound to human APE1 (PDB: 1DE8)



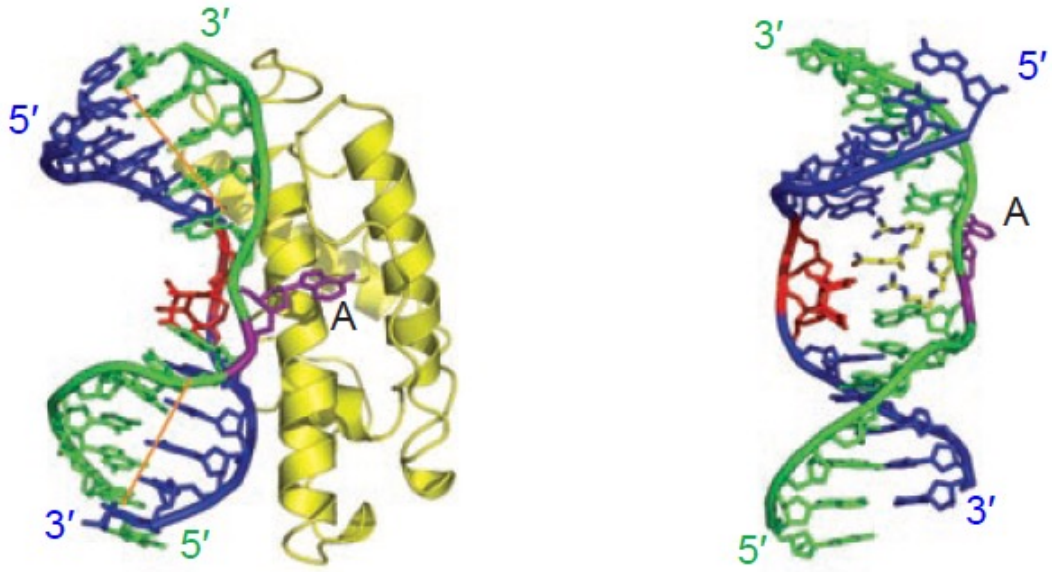
DNA bound to human APE1 (PDB: 1DE8)

Yang, Cell Research 2008

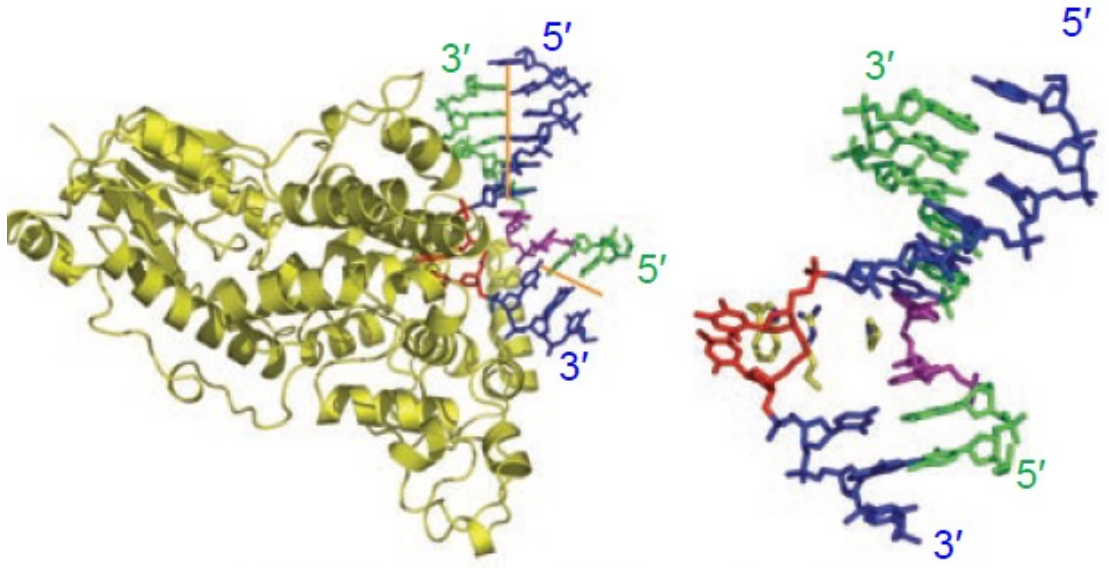
Protein-nucleic acid interactions

Damage recognition

Recognition of thymine dimers (CPDs) by Endo V and DNA photolyase



T4 Endonuclease V - CPD complex (1VAS)



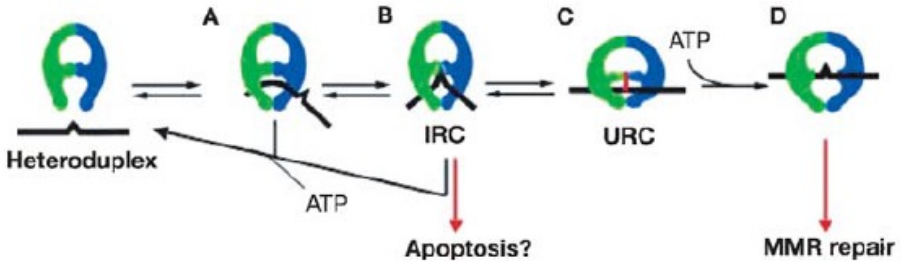
DNA Photolyase - CPD complex (1TEZ)

Yang, Cell Research 2008

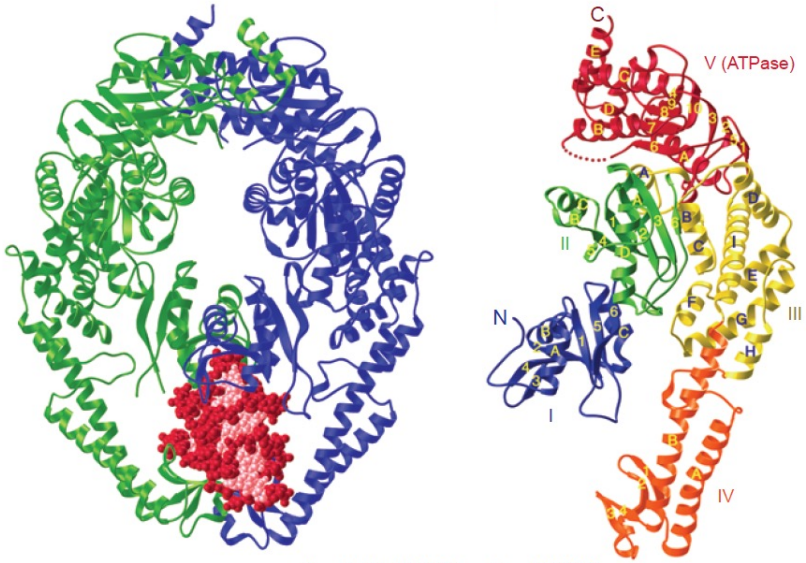
Protein-nucleic acid interactions

Damage recognition

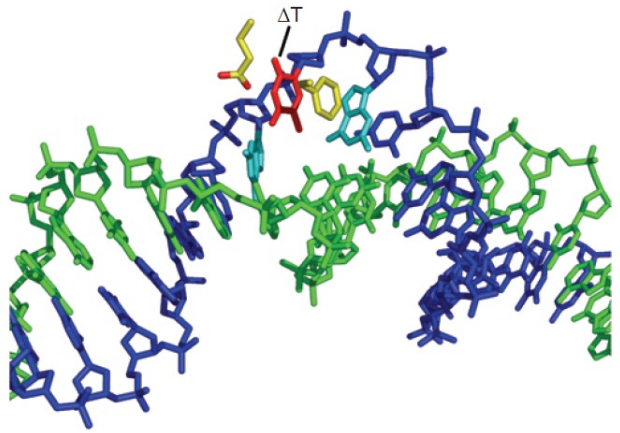
Mismatch repair detected by MutS, sample to disrupt



Kunkel and Erie, Annual Rev 2005



Taq MutS-DNA Complex (1EWQ)

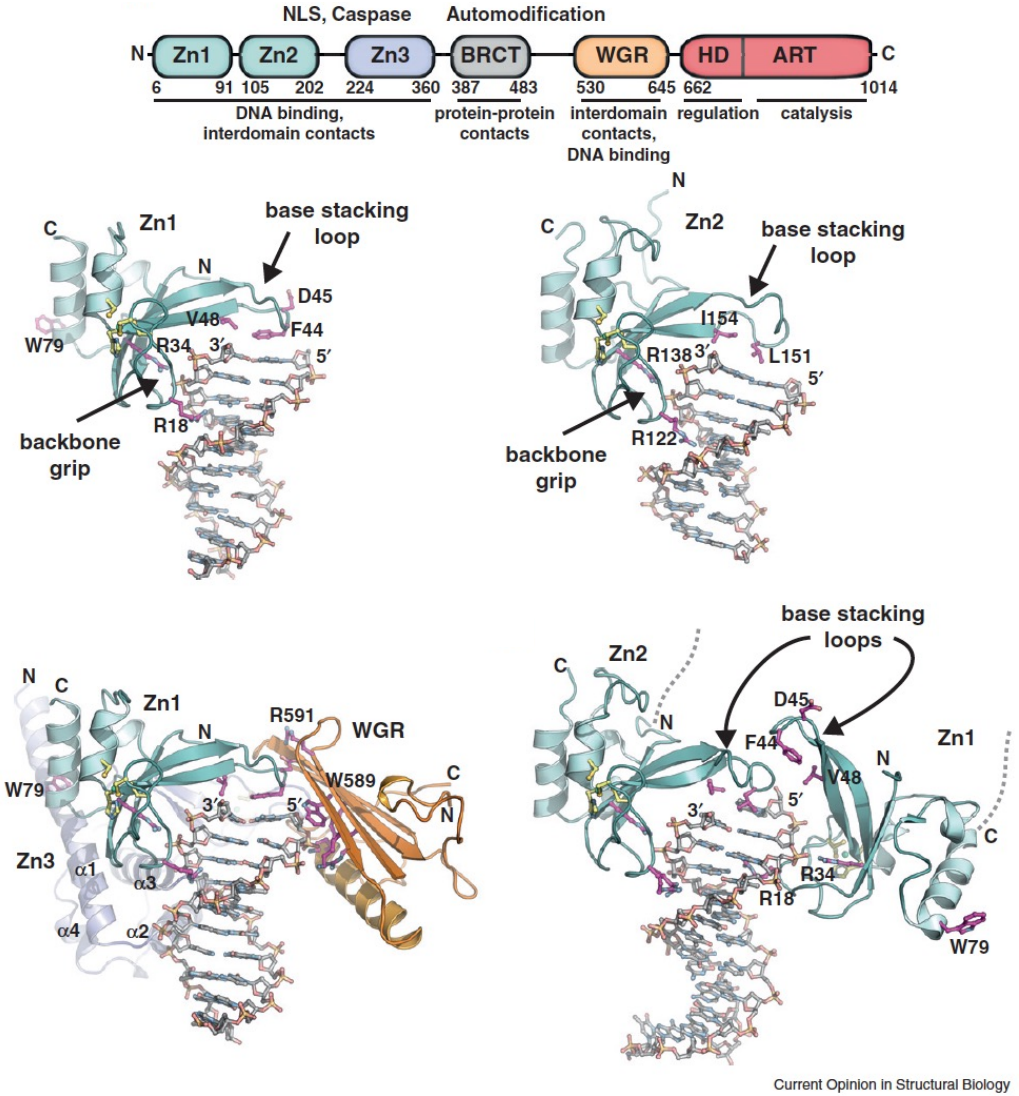


Yang, Cell Research 2008

Protein-nucleic acid interactions

Damage recognition

PARP1 couples DNA damage detection to poly(ADP-ribose) synthesis

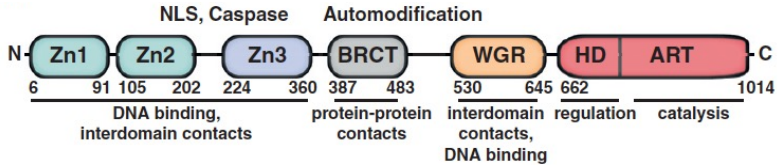


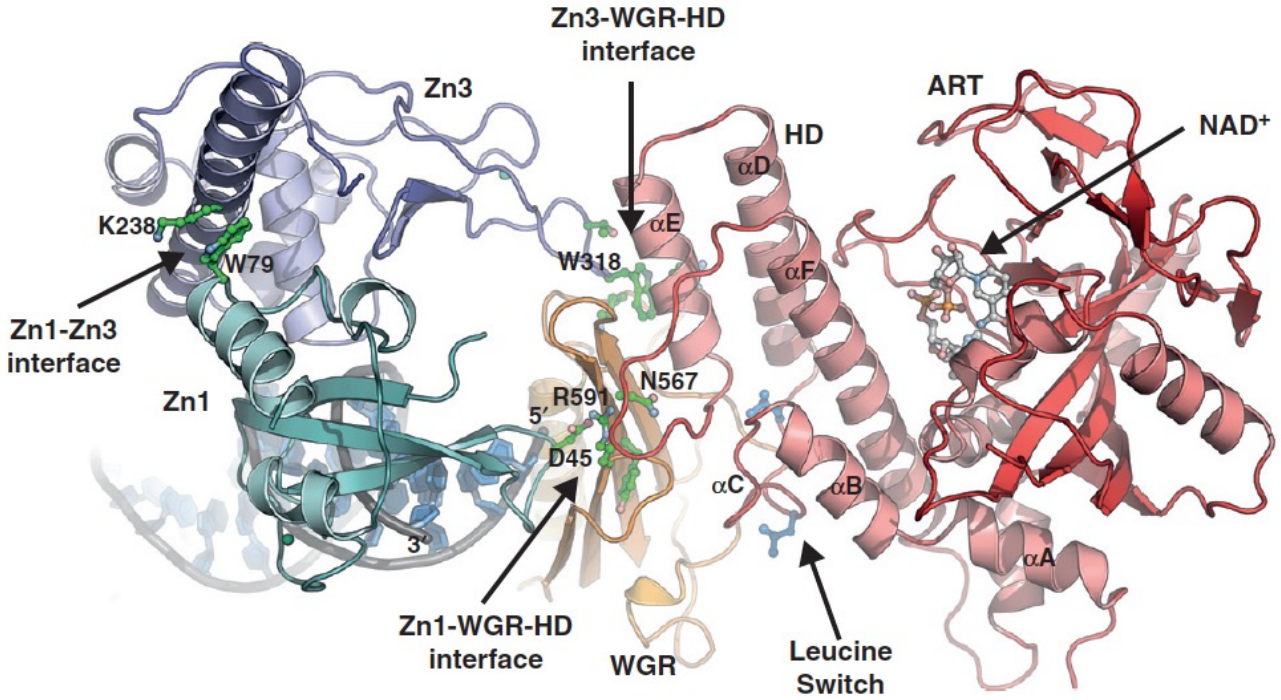
Langelier and Pascal, COSB 2013

Protein-nucleic acid interactions

Damage recognition

PARP1 couples DNA damage detection to poly(ADP-ribose) synthesis



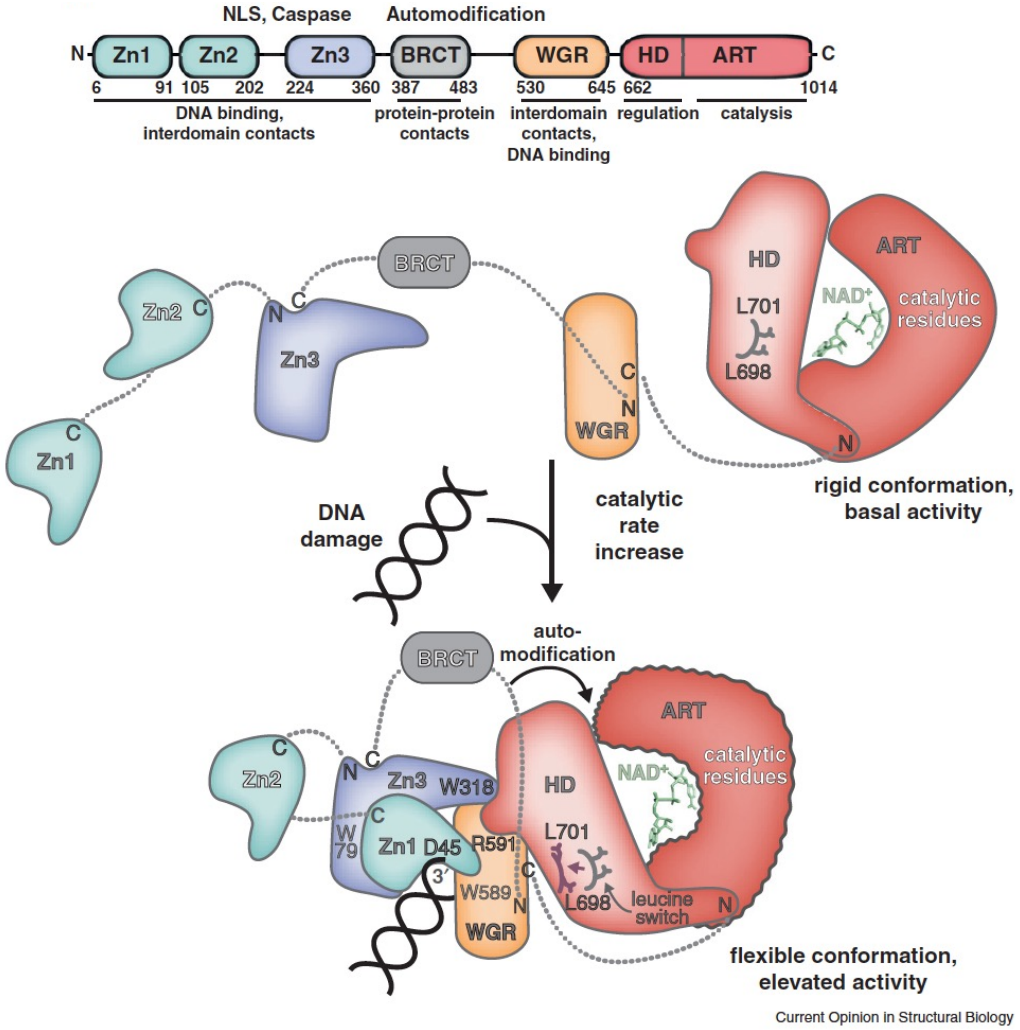


Langelier and Pascal, COSB 2013

Protein-nucleic acid interactions

Damage recognition

PARP1 couples DNA damage detection to poly(ADP-ribose) synthesis



Langelier and Pascal, COSB 2013

Protein-nucleic acid interactions

RNA binding proteins – a multitude of domains

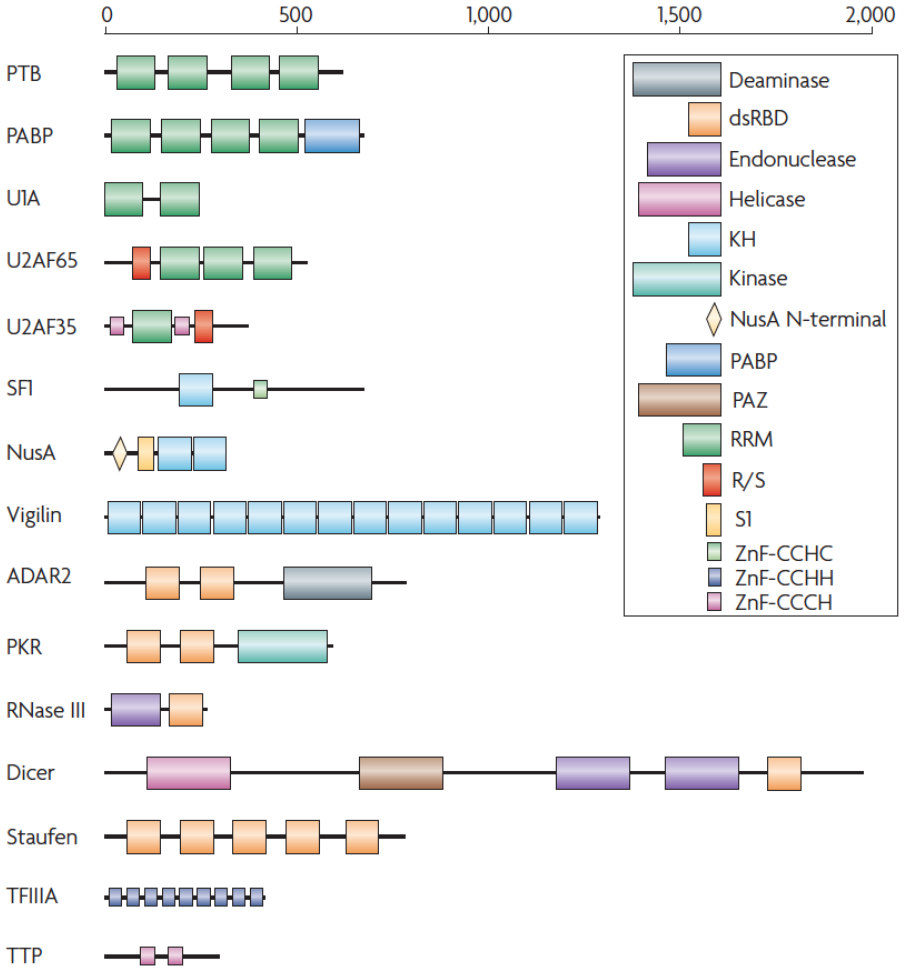
Table 1 | **Common RNA-binding domains and their properties**

Domain	Topology	RNA-recognition surface	Protein–RNA interactions	Representative structures (PDB ID)
RRM	αβ	Surface of β-sheet	Interacts with about four nucleotides of ssRNA through stacking, electrostatics and hydrogen bonding	U1A N-terminal RRM[18] (1URN)
KH (type I and type II)	αβ	Hydrophobic cleft formed by variable loop between β2, β3 and GXXG loop. Type II: same as type I, except variable loop is between α2 and β2	Recognizes about four nucleotides of ssRNA through hydrophobic interactions between non-aromatic residues and the bases; sugar-phosphate backbone contacts from the GXXG loop, and hydrogen bonding to bases	Nova-1 KH3 (type I)[41] (1EC6), NusA (type II)[37] (2ASB)
dsRBD	αβ	Helix α1, N-terminal portion of helix α2, and loop between β1 and β2	Shape-specific recognition of the minor–major–minor groove pattern of dsRNA through contacts to the sugar-phosphate backbone; specific contacts from the N-terminal α-helix to RNA in some proteins	dsRBD3 from Staufen[51] (1EKZ)
ZnF-CCHH	αβ	Primarily residues in α-helices	Protein side chain contacts to bulged bases in loops and through electrostatic interactions between side chains and the RNA backbone	Fingers 4–6 of TFIIIA[56] (1UN6)
ZnF-CCCH	Little regular secondary structure	Aromatic side chains form hydrophobic binding pockets for bases that make direct hydrogen bonds to protein backbone	Stacking interactions between aromatic residues and bases create a kink in RNA that allows for the direct recognition of Watson–Crick edges of the bases by the protein backbone	Fingers 1 and 2 of TIS11d[57] (1RGO)
S1	β	Core formed by two β-strands with contributions from surrounding loops	Stacking interactions between bases and aromatic residues and hydrogen bonding to the bases	Ribonuclease II[121] (2IX1), exosome[99] (2NN6)
PAZ	αβ	Hydrophobic pocket formed by OB-like β-barrel and small αβ motif	Recognizes single-stranded 3′ overhangs of siRNA through stacking interactions and hydrogen bonds	PAZ[73] (1SI3), Argonaute[76] (1U04), Dicer[72] (2FFL)
PIWI	αβ	Highly conserved pocket, including a metal ion that is bound to the exposed C-terminal carboxylate	Recognizes the defining 5′ phosphate group in the siRNA guide strand with a highly conserved binding pocket that includes a metal ion	PIWI[75] (1YTU), Argonaute (1U04)[76]
TRAP	β	Edges of β-sheets between each of the 11 subunits that form the entire protein structure	Recognizes the GAG triplet through stacking interactions and hydrogen bonding to bases; limited contacts to the backbone	TRAP[122] (1C9S)
Pumilio	α	Two repeats combine to form binding pocket for individual bases; helix α2 provides specificity-determining residues	Binding pockets for bases provided by stacking interactions; specificity dictated by hydrogen bonds to the Watson–Crick face of a base by two amino acids in helix α2	Pumilio[84] (1M8Y)
SAM	α	Hydrophobic cavity between three helices surrounded by an electropositive region	Shape-dependent recognition of RNA stem–loop, mainly through interactions with the sugar-phosphate backbone and a single base in the loop	Vts1[123] (2ESE)

Lunde et al., Nat Rev MCB 2007

Protein-nucleic acid interactions

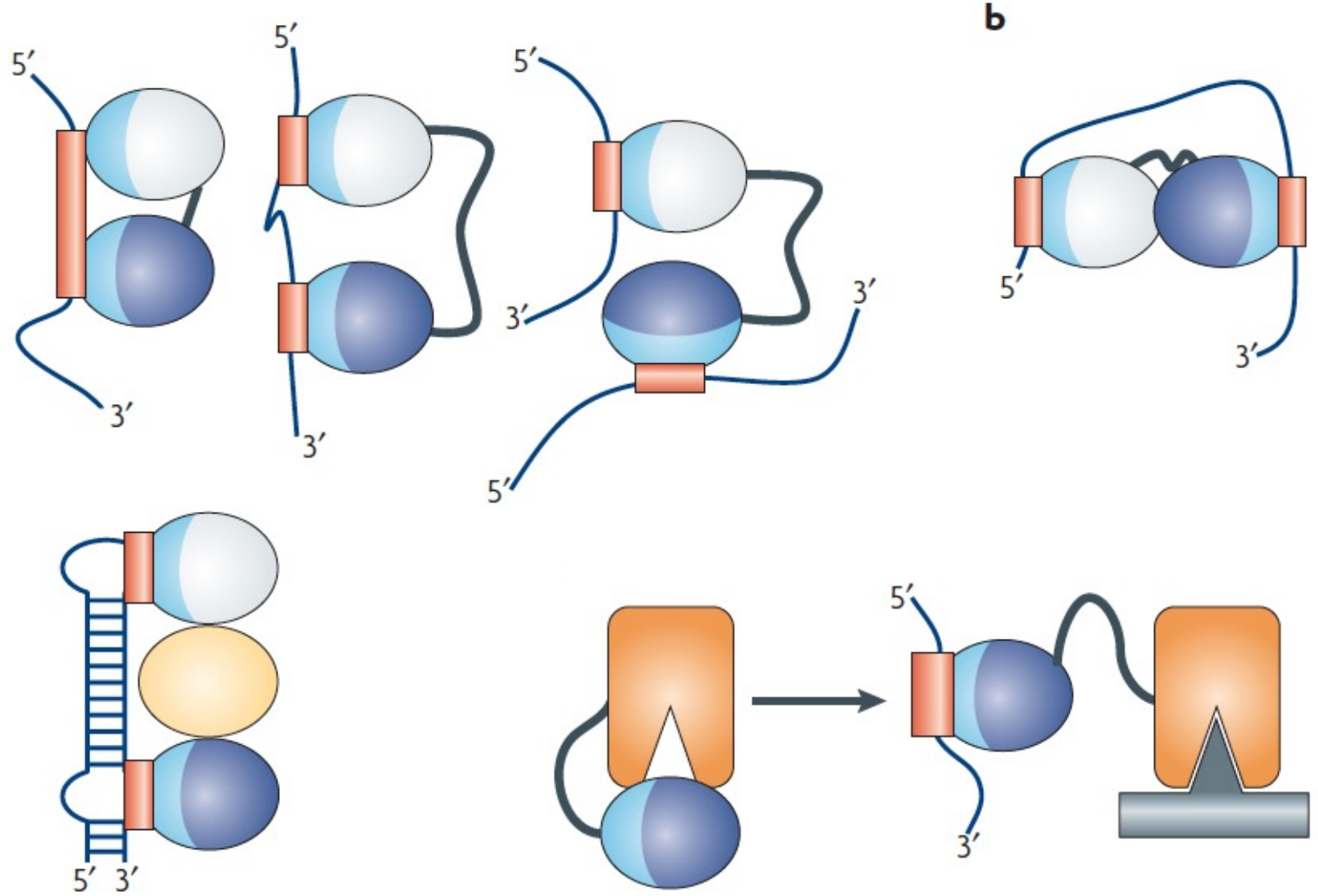
RNA binding proteins – modular structures to recognize unique elements



Lunde et al., Nat Rev MCB 2007

Protein-nucleic acid interactions

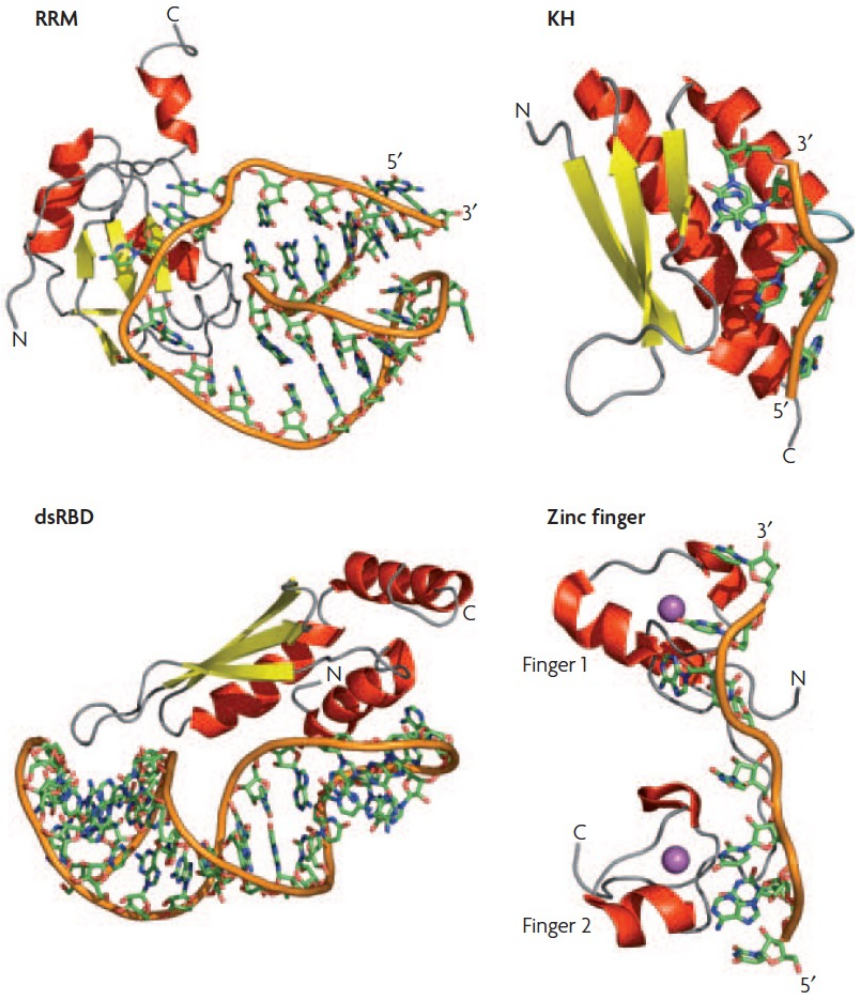
RNA binding proteins – combine modules to recognize specific elements



Lunde et al., Nat Rev MCB 2007

Protein-nucleic acid interactions

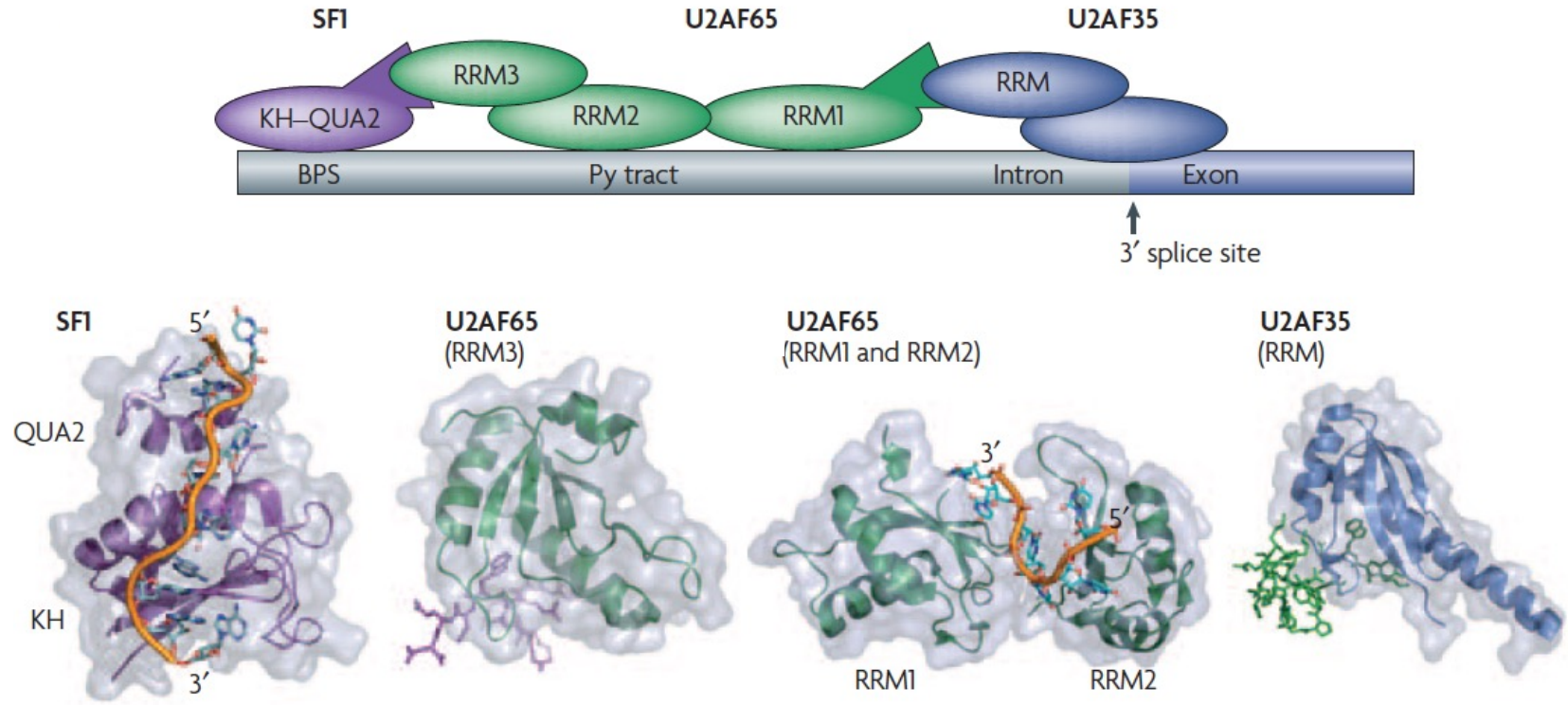
RNA binding proteins – how some binding modules recognize RNA



Lunde et al., Nat Rev MCB 2007

Protein-nucleic acid interactions

RNA binding proteins – protein and RNA-protein interactions define spliceosomal assembly

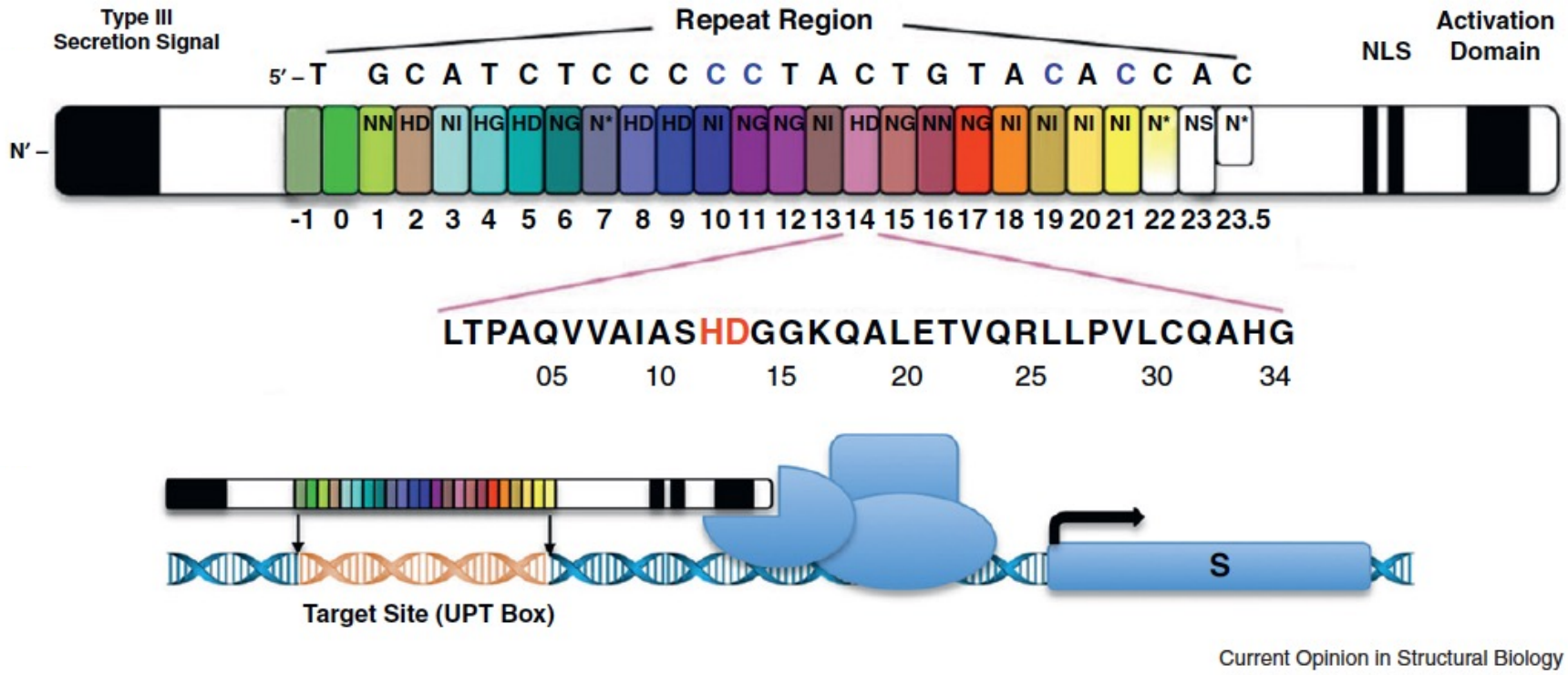


Lunde et al., Nat Rev MCB 2007

Protein-nucleic acid interactions

Nucleases (how do they find their sites)

Talon – TAL effectors are secreted proteins that target host DNA to alter transcription

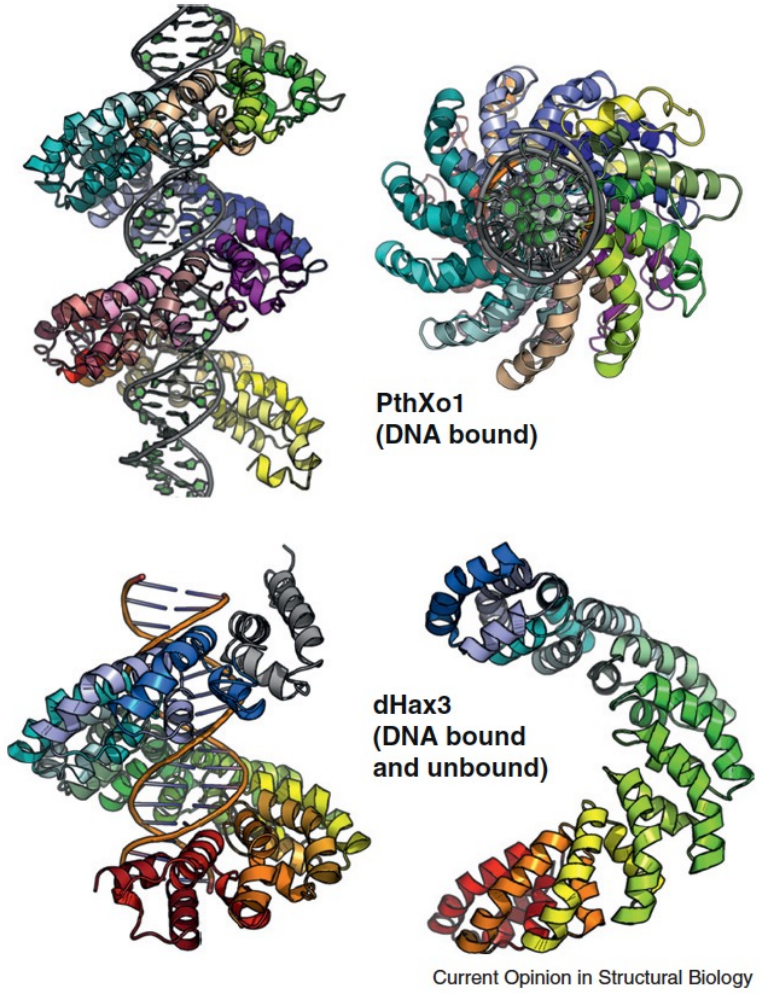


Mak et al., COSB 2013

Protein-nucleic acid interactions

Nucleases (how do they find their sites)

Talon – TAL effectors modules wrap around DNA to specifically recognize sites



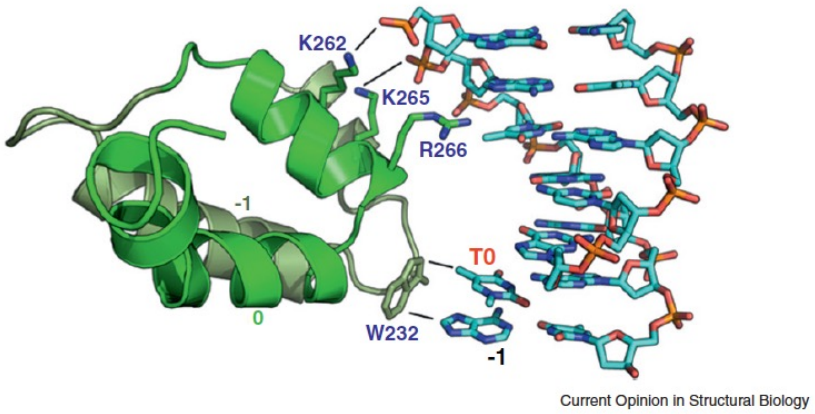
Mak et al., COSB 2013

Protein-nucleic acid interactions

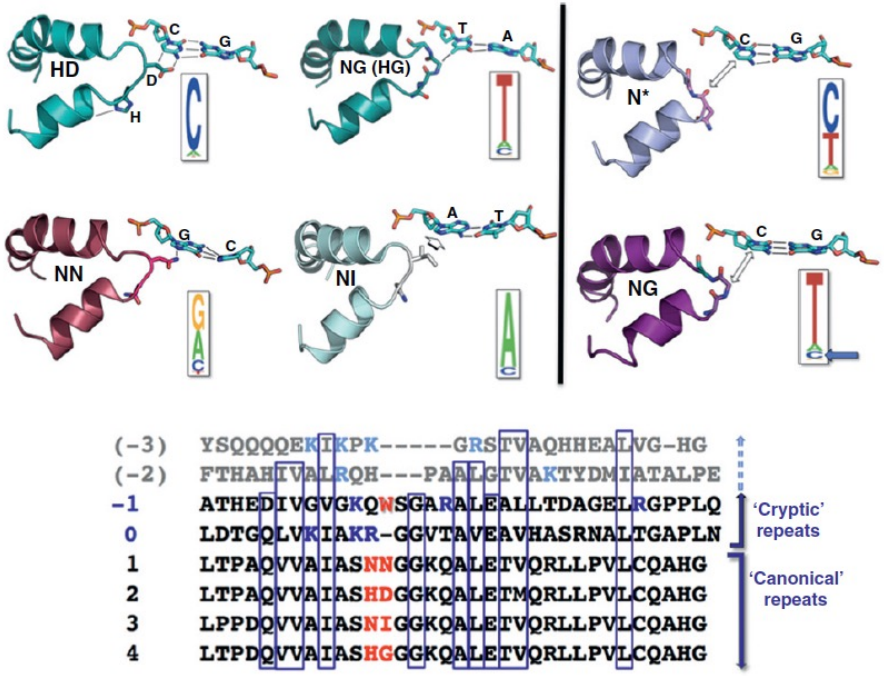
Nucleases (how do they find their sites)

Talon – TAL effector code

Each base can be read by a unique combination of amino and backbone interactions



Current Opinion in Structural Biology

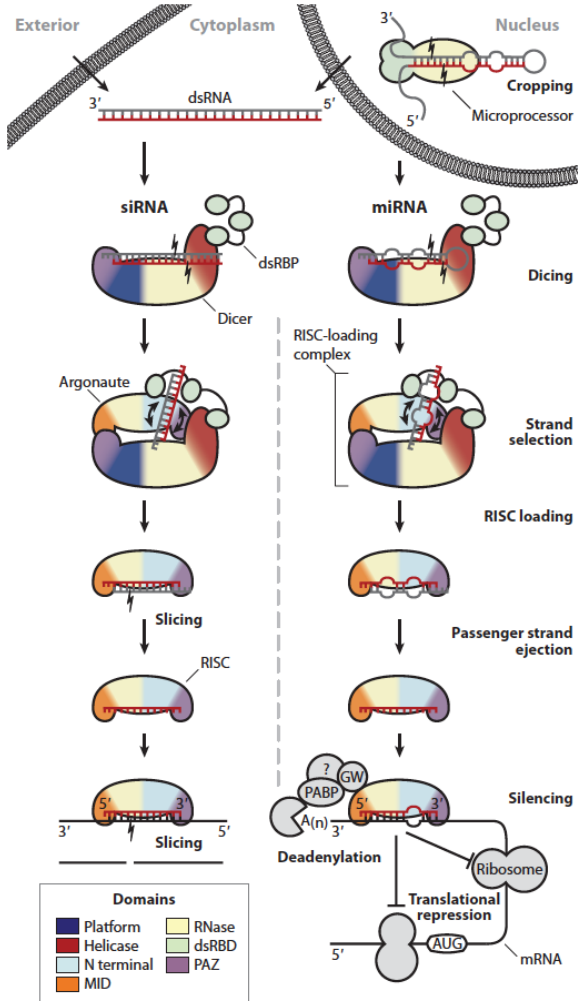


Mak et al., COSB 2013

Protein-nucleic acid interactions

Nucleases (how do they find their sites)

Ago – Ago proteins target RNA through a combination of protein and NA interactions

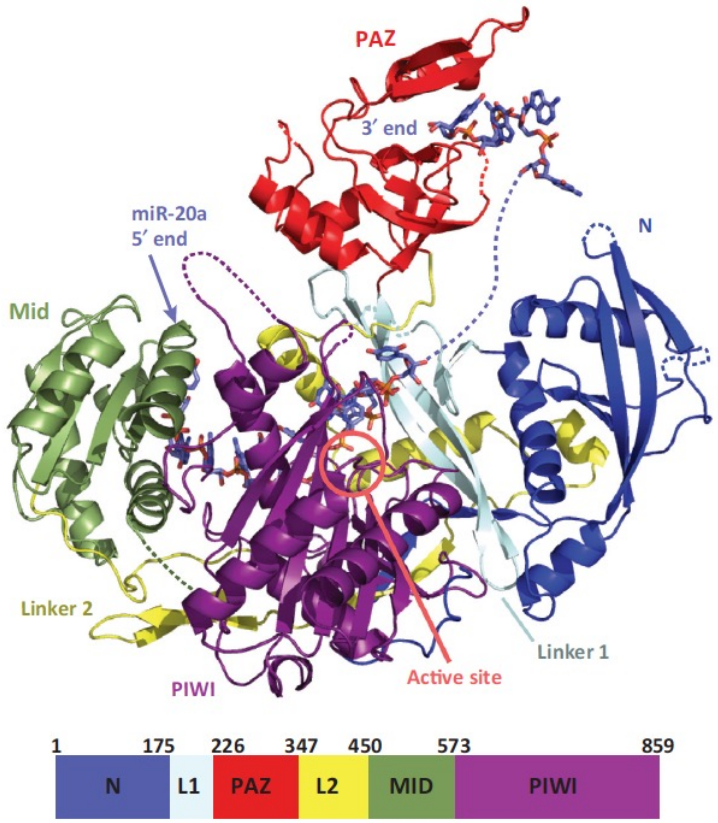


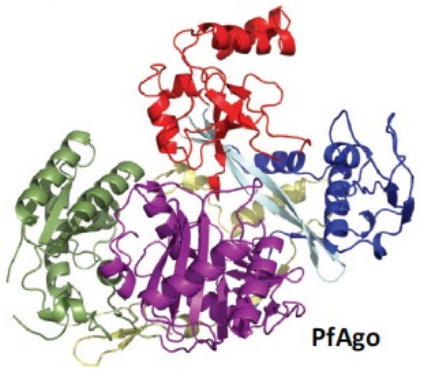
Wilson and Doudna, Annu Rev 2013

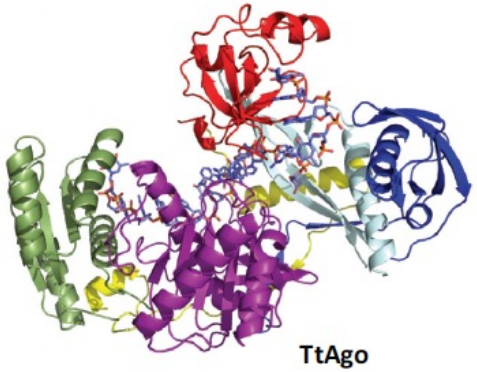
Protein-nucleic acid interactions

Nucleases (how do they find their sites)

Ago – Guide RNA (DNA) interactions with Ago homologs – bases exposed for search





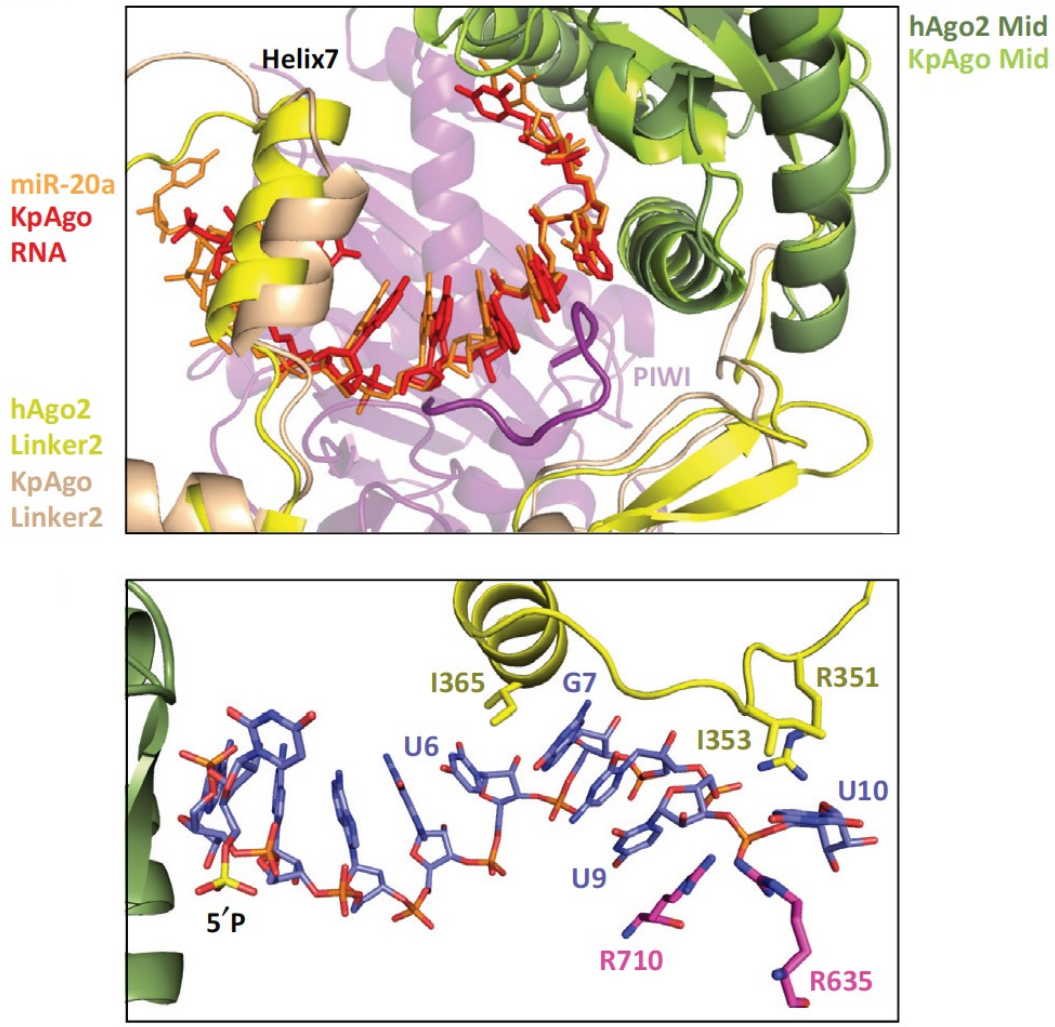


Kuhn and Joshua-Tor, TiBS 2013

Protein-nucleic acid interactions

Nucleases (how do they find their sites)

Ago – Path of the seed sequence in human and yeast Ago

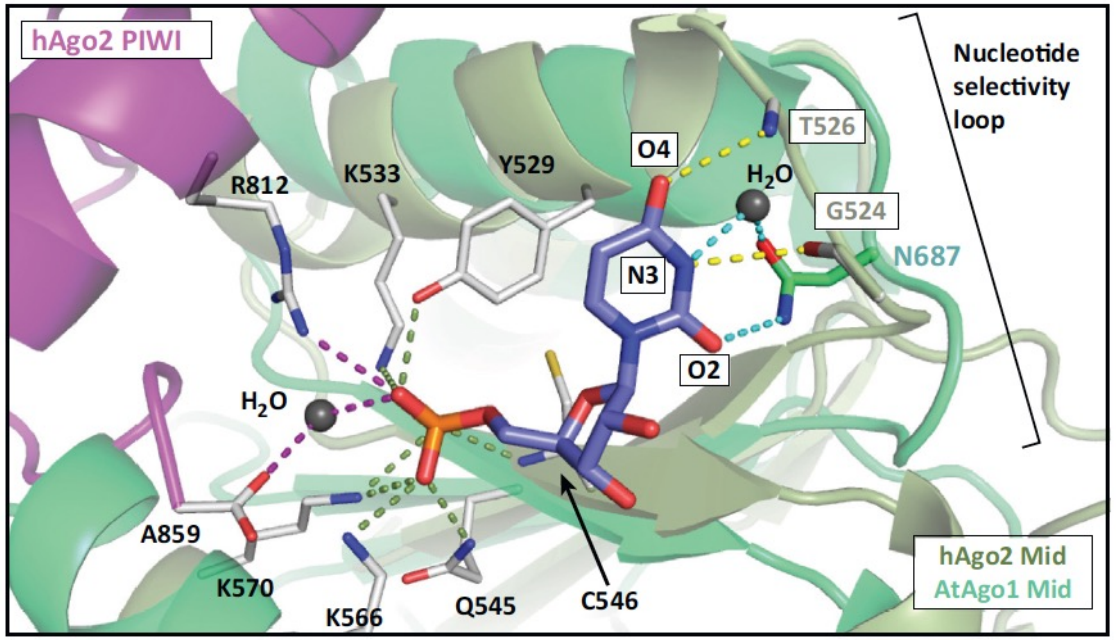


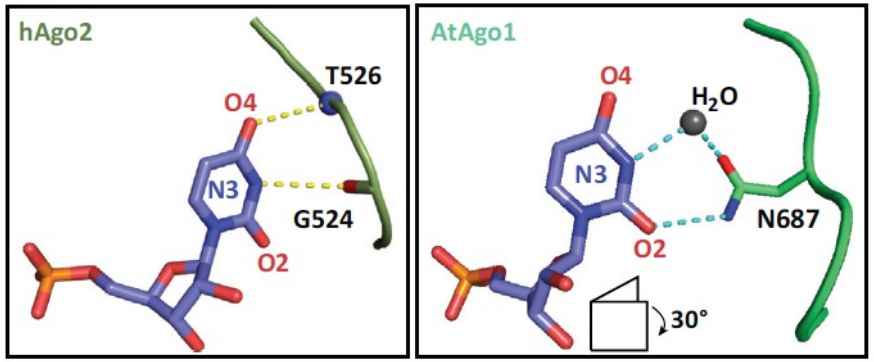
Kuhn and Joshua-Tor, TiBS 2013

Protein-nucleic acid interactions

Nucleases (how do they find their sites)

Ago – Guide RNA 5’ end anchored in the MID domain



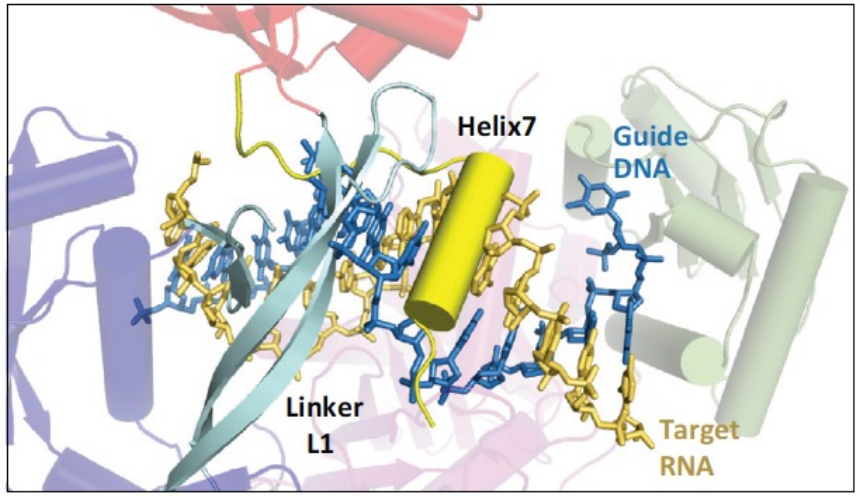


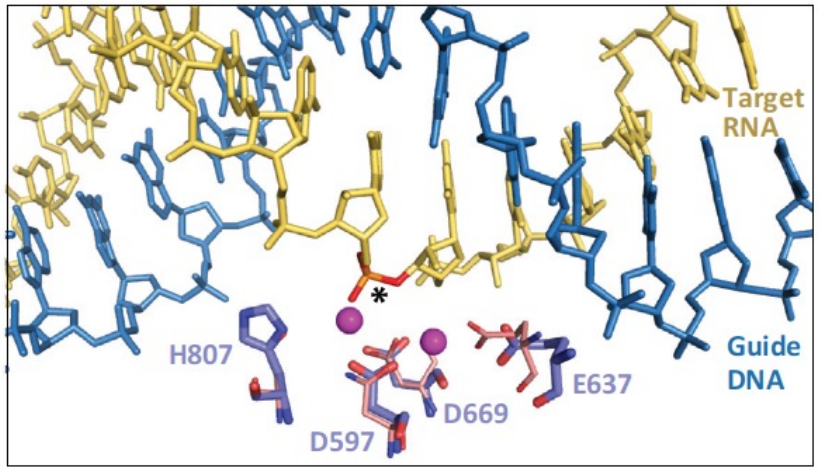
Kuhn and Joshua-Tor, TiBS 2013

Protein-nucleic acid interactions

Nucleases (how do they find their sites)

Ago – Target RNA binding requires conformational changes for cleavage





Kuhn and Joshua-Tor, TiBS 2013

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS

Clustered Regularly Interspaced Short Palindromic Repeats

Associated with cas (CRISPR-associated) genes

System provides bacteria and archaea immunity to phage (viruses) or mobile genetic elements

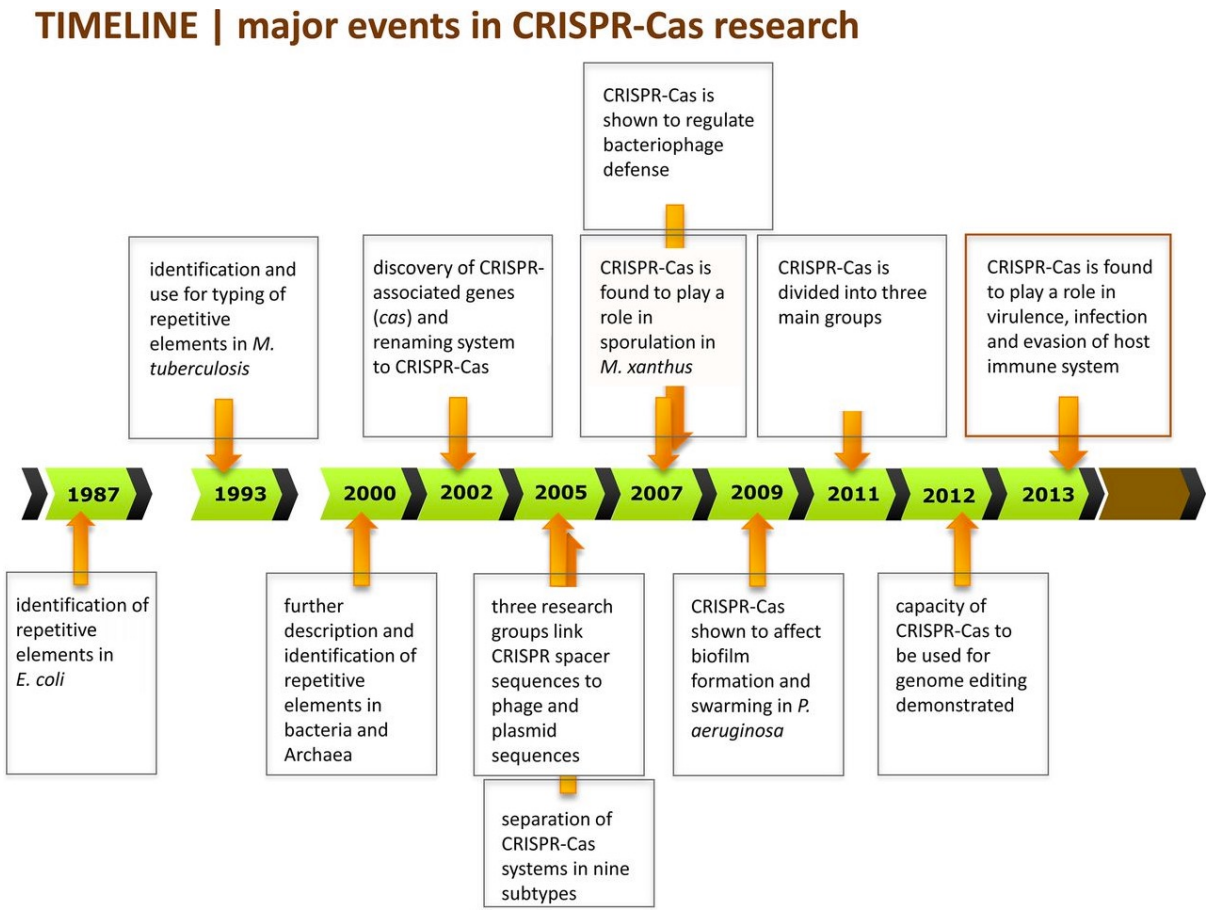
May also be involved in regulated gene expression

Wiedenheft et al., Nature 2012

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS – the timeline

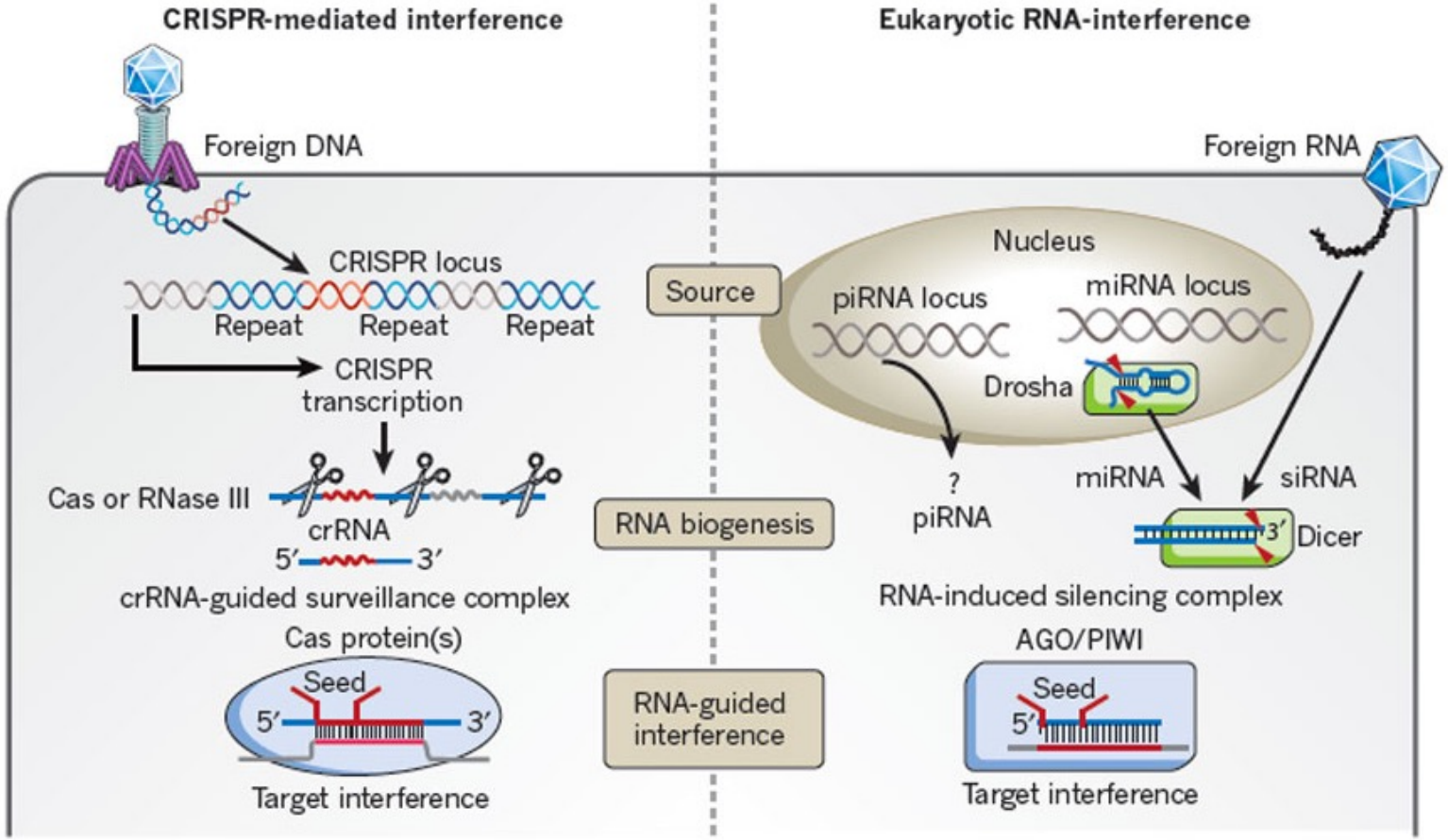


Louwen et al., Micro Mol Biol Rev 2014

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: analogies to eukaryotic interference pathways

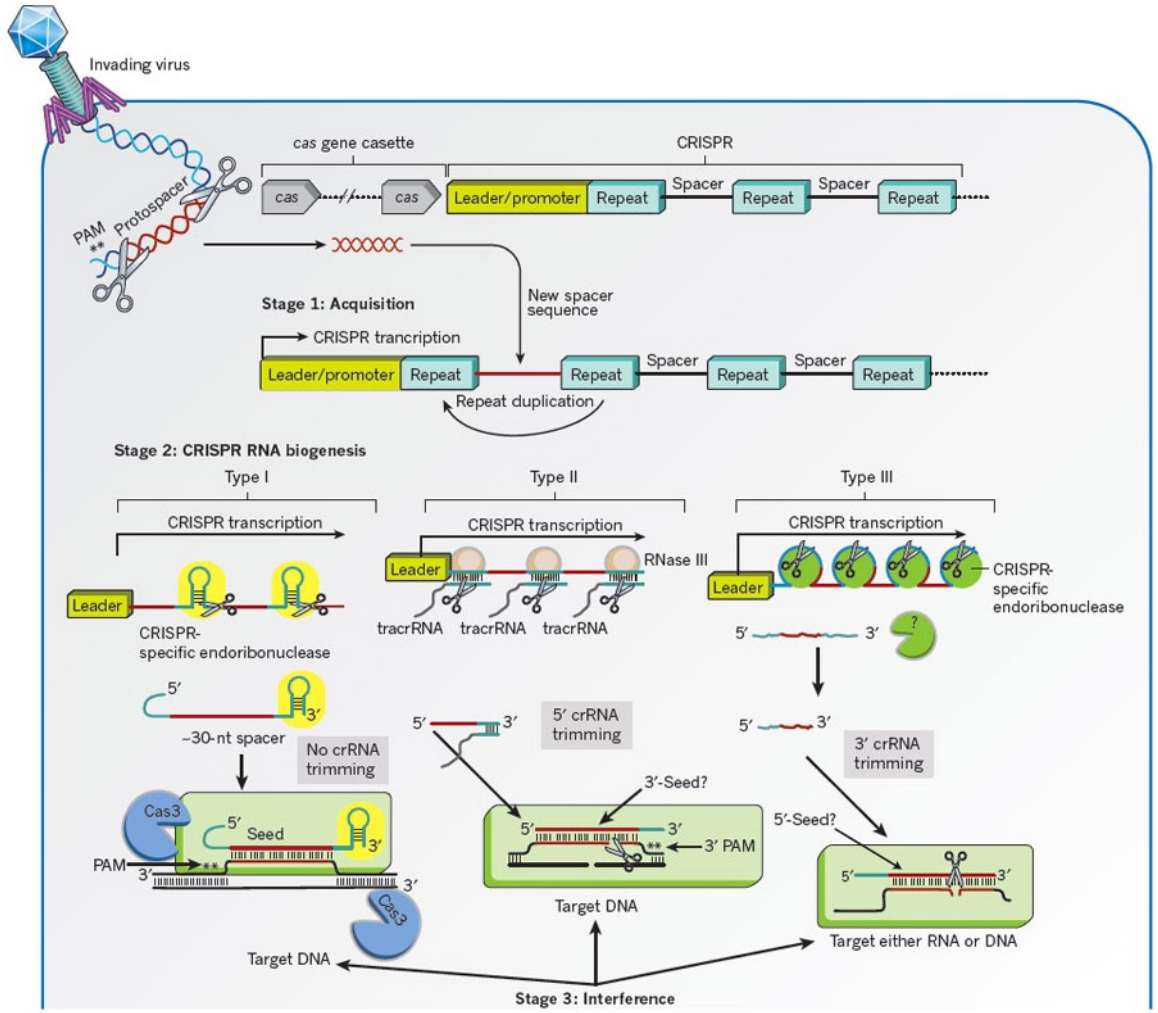


Wiedenheft et al., Nature 2012

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: acquisition of new spacers through integration between repeats

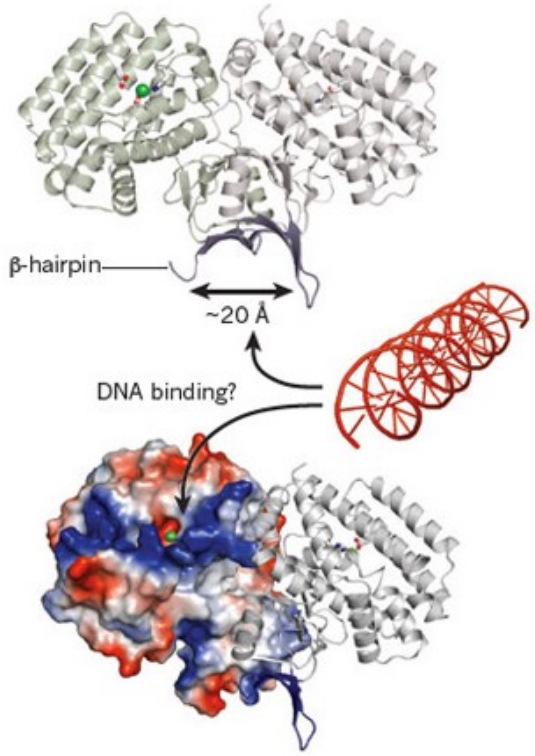


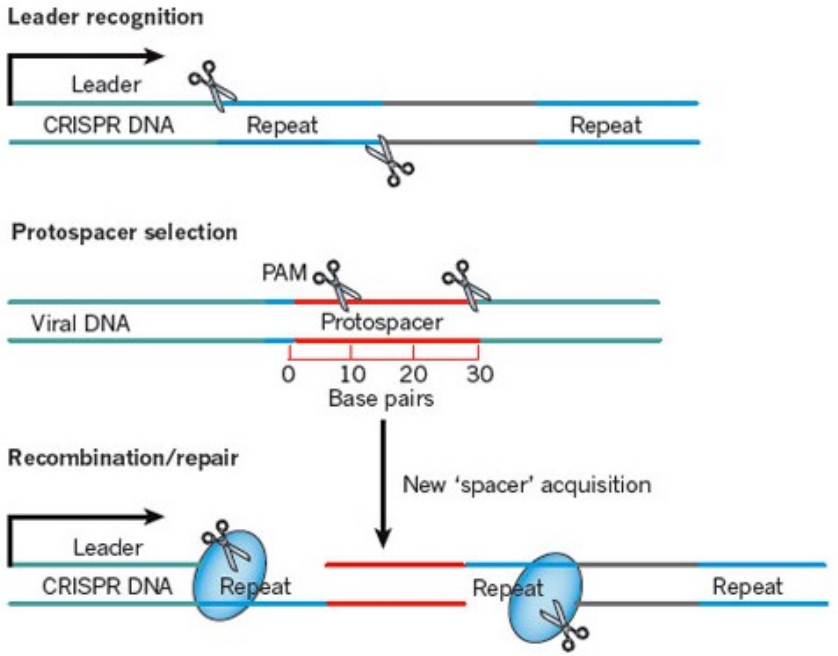
Wiedenheft et al., Nature 2012

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: Cas1 mediates steps prior to integration



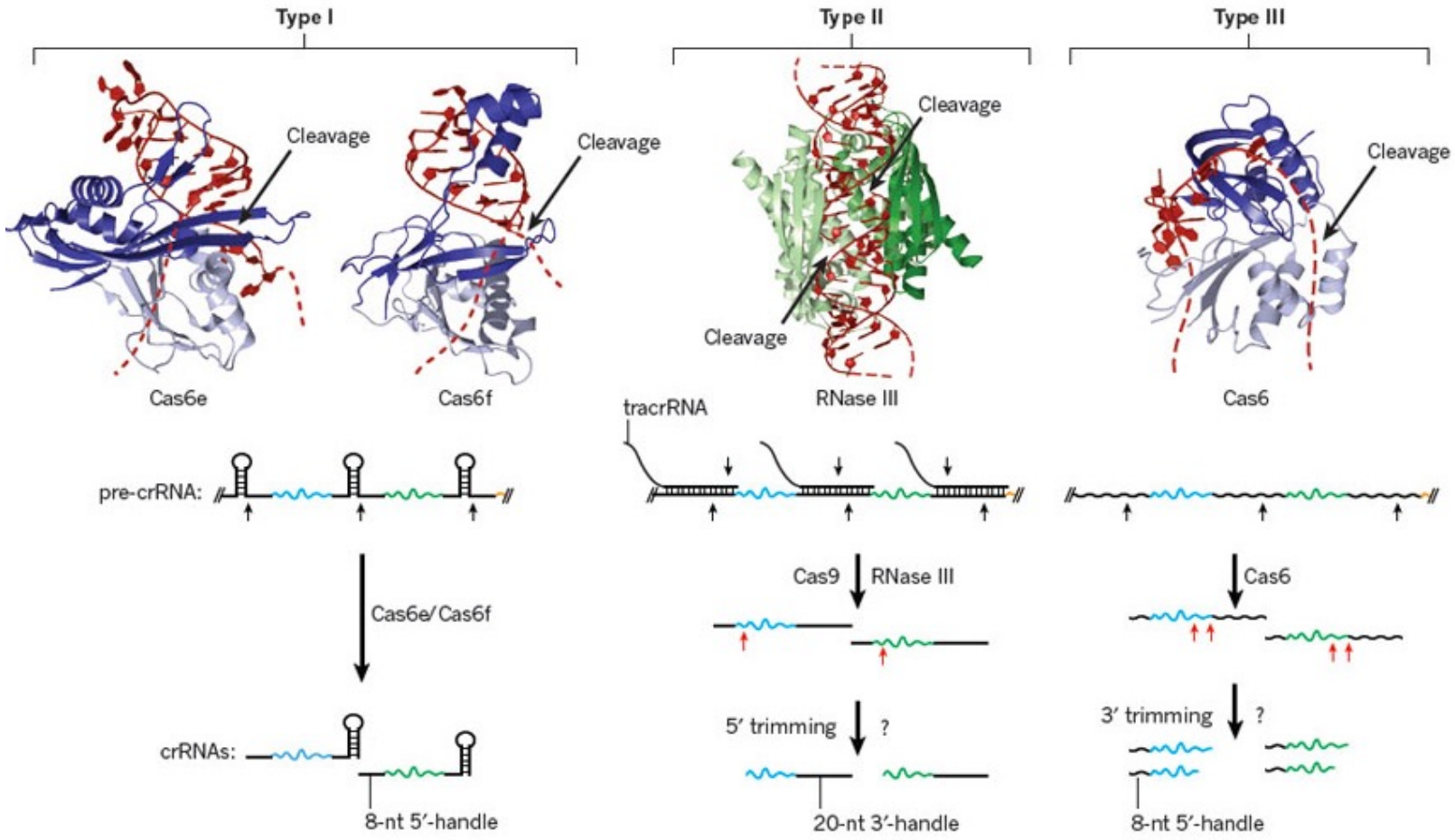


Wiedenheft et al., Nature 2012

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: Diverse mechanisms of CRISPR RNA biogenesis

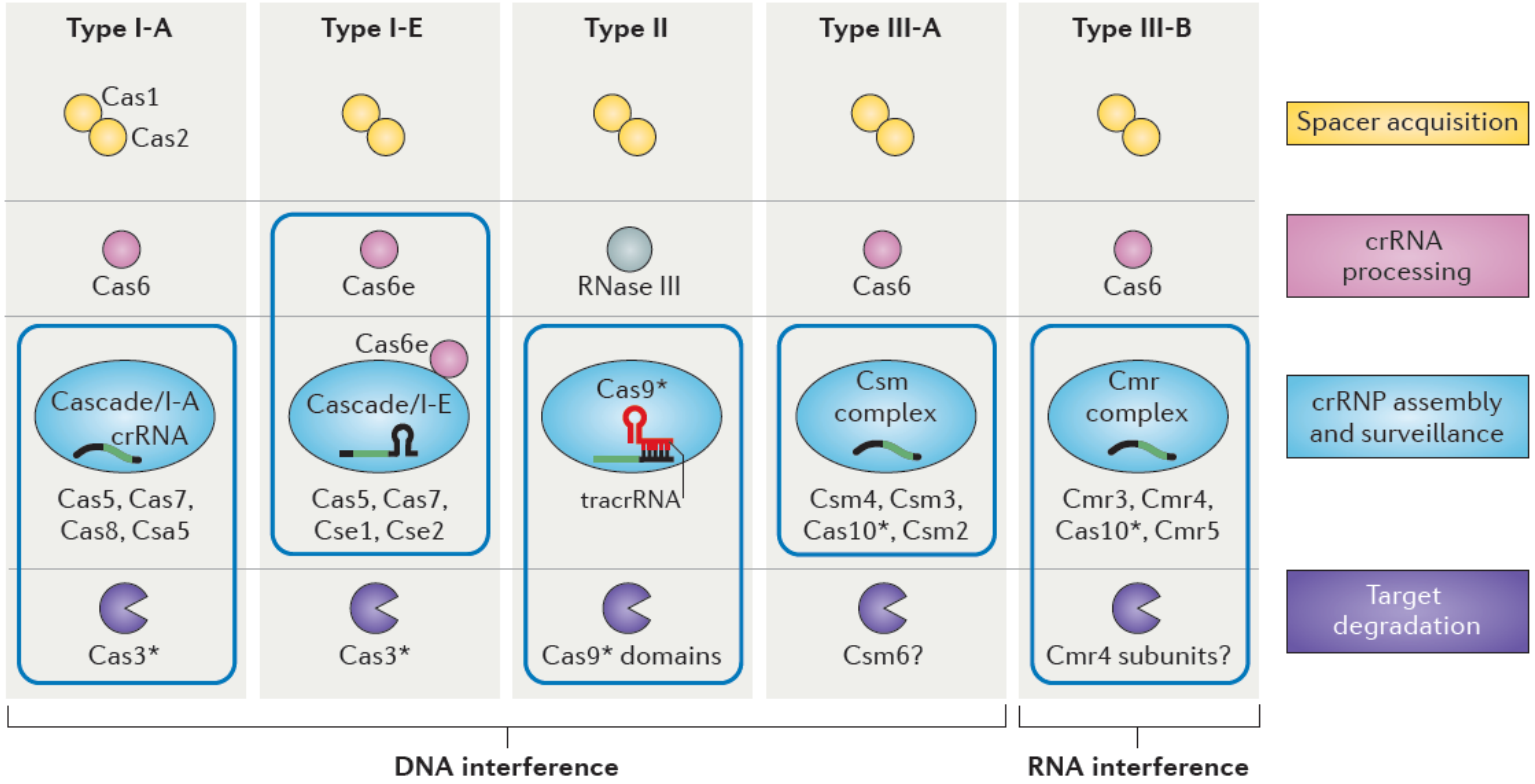


Wiedenheft et al., Nature 2012

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: Diverse mechanisms of CRISPR RNA presentation and cleavage

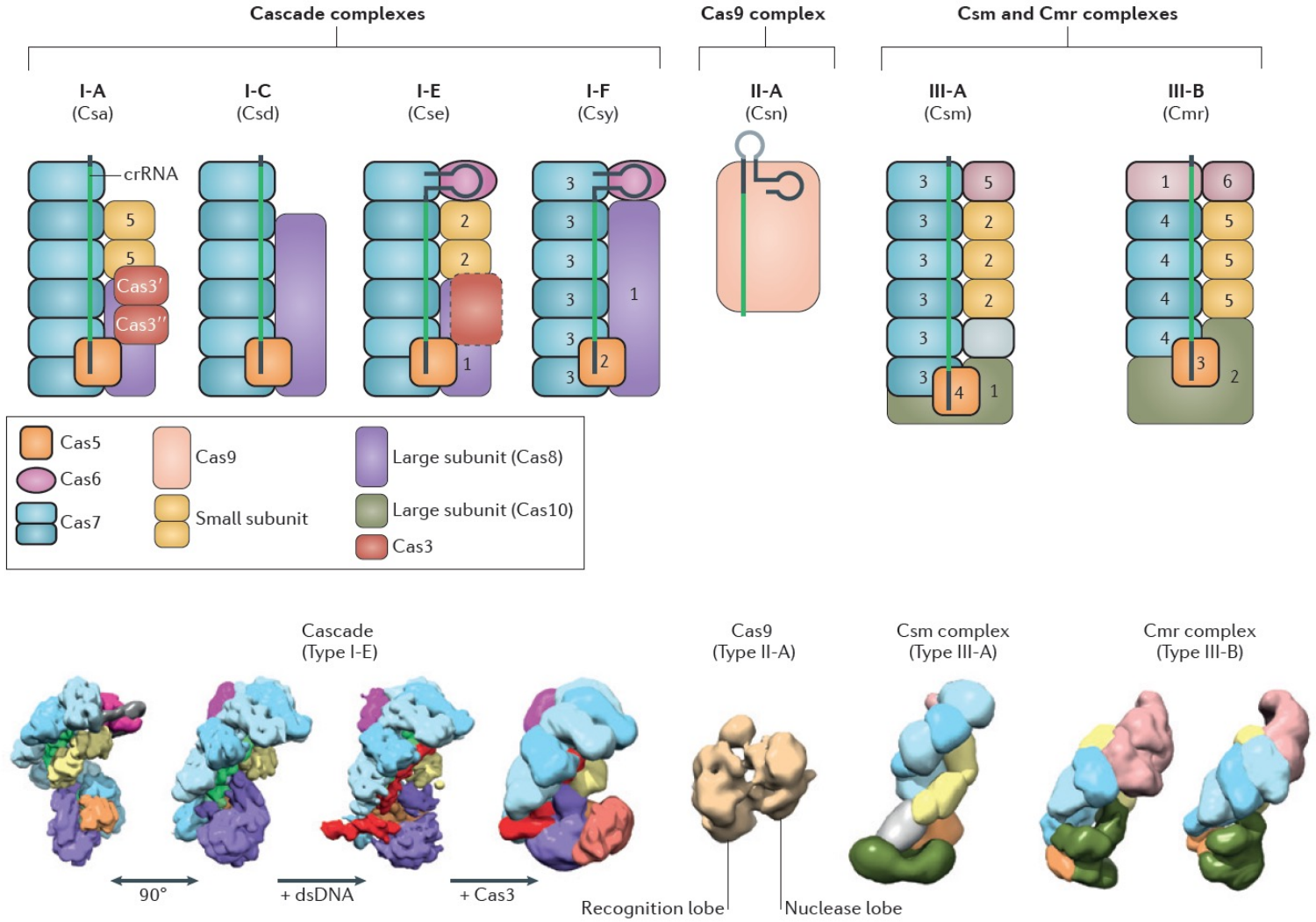


van der Oost et al., Nat Rev Micro 2014

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: Diverse mechanisms of CRISPR RNA presentation and cleavage

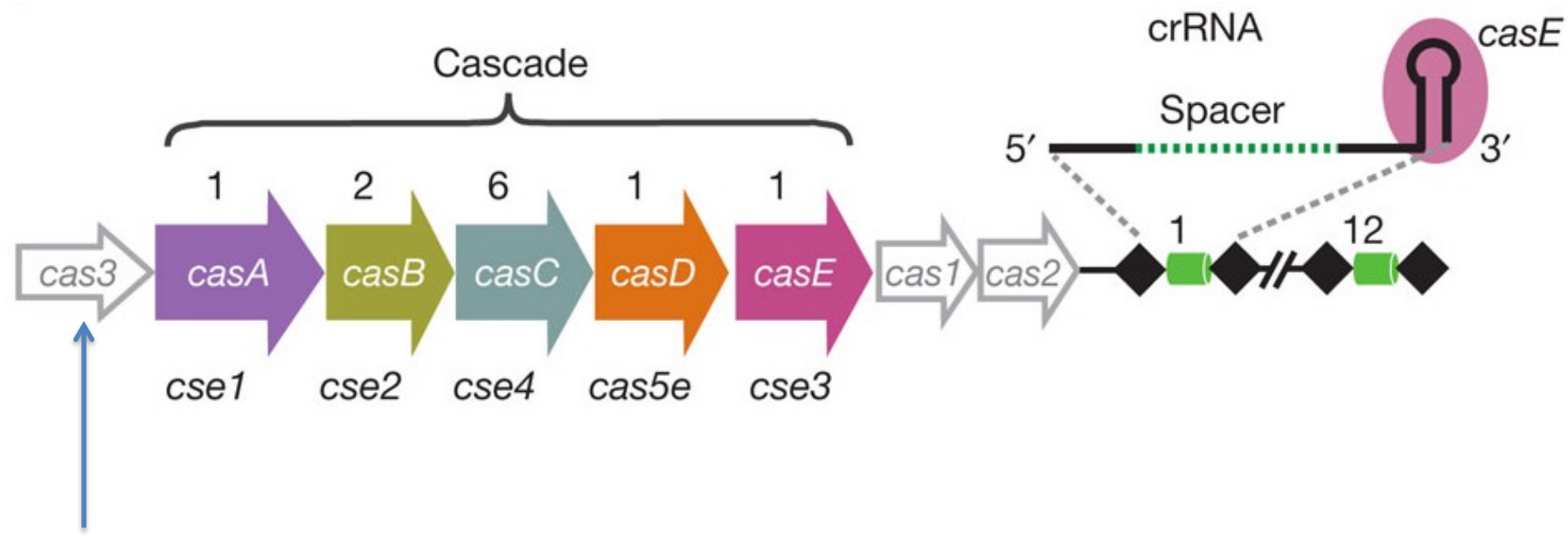


van der Oost et al., Nat Rev Micro 2014

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: The CASCADE system in E. coli

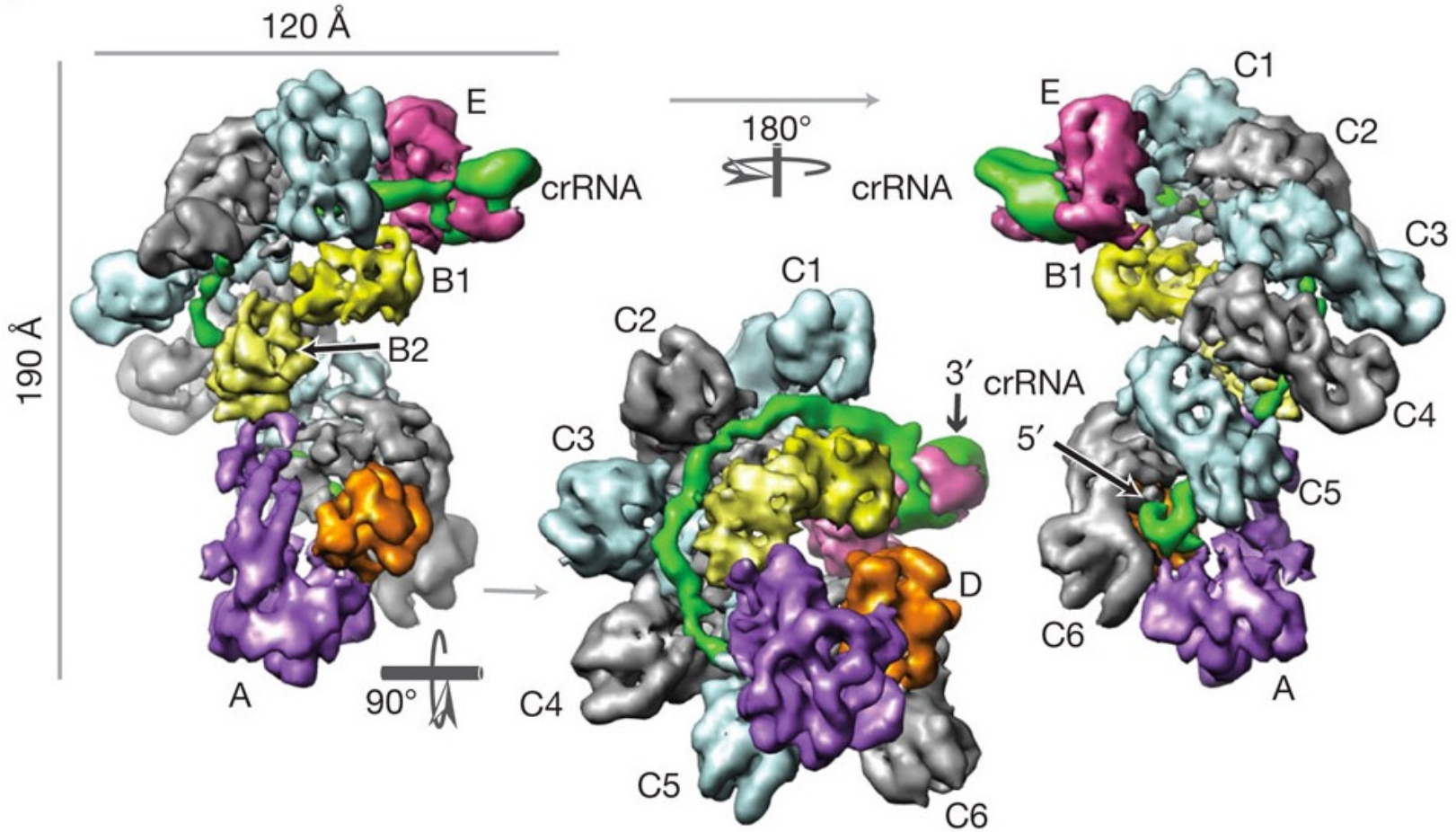


Wiedenheft et al., Nature 2011

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: Organization of the CASCADE complex bound to crRNA



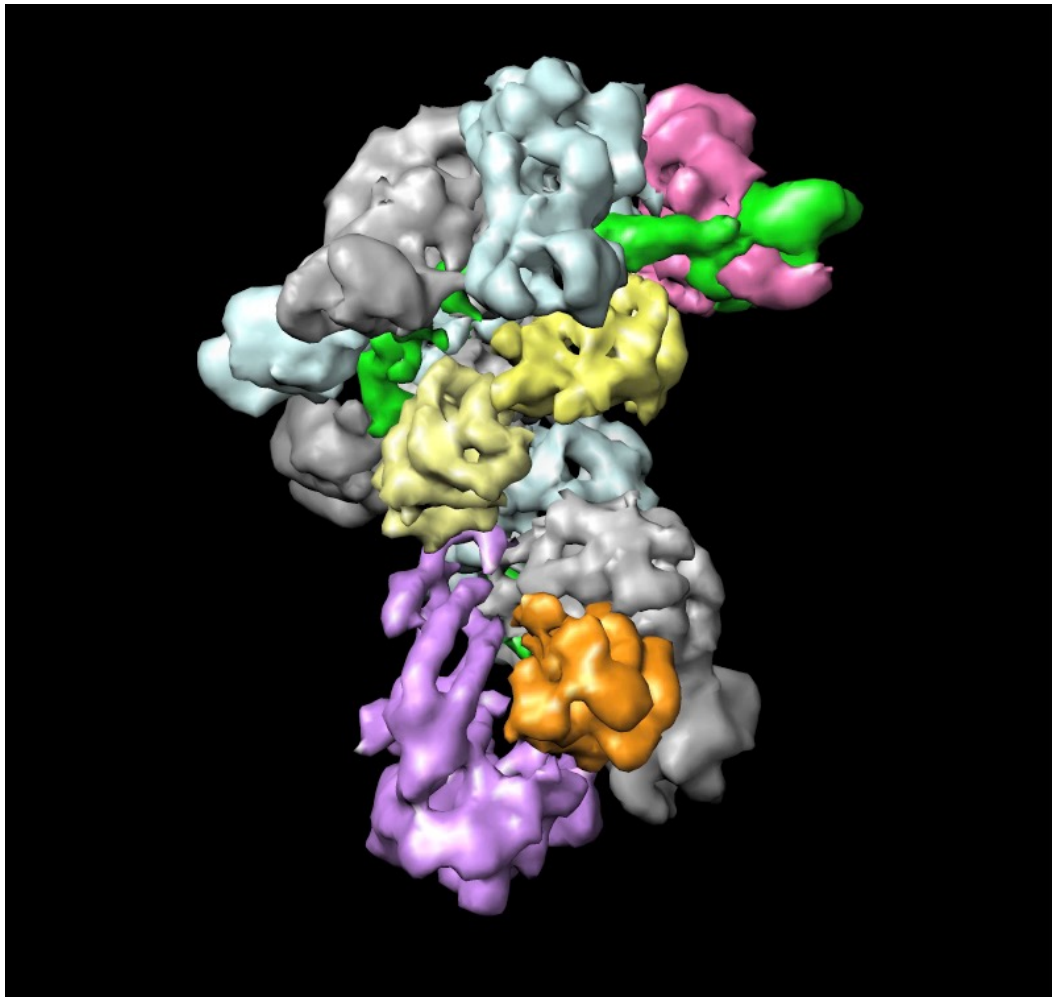
6 CasC, 2 CasB, CasA, CasD and CasE

Wiedenheft et al., Nature 2011

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: Organization of the CASCADE complex bound to crRNA



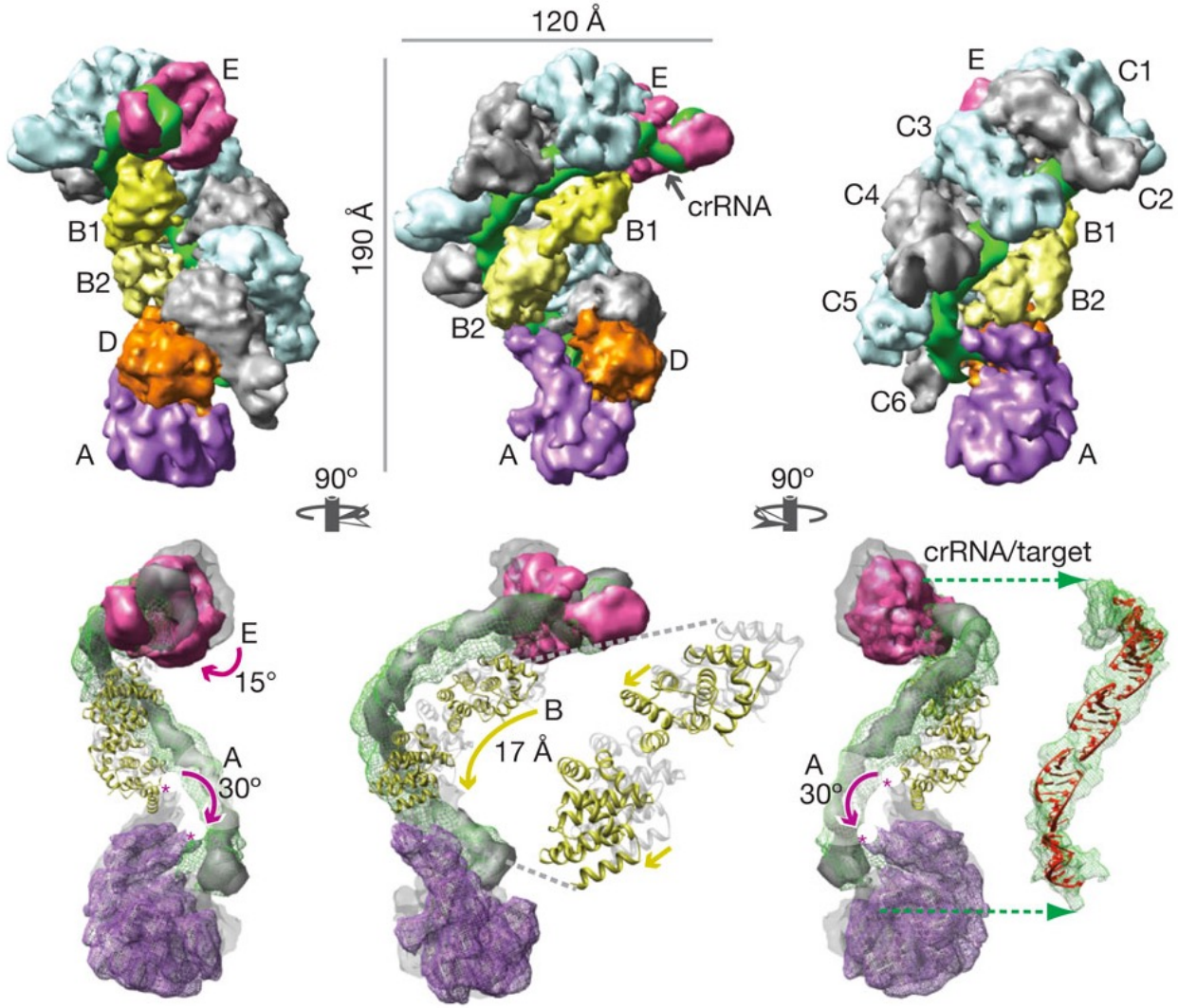
6 CasC, 2 CasB, CasA, CasD and CasE

Wiedenheft et al., Nature 2011

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: CASCADE complex bound to crRNA:ssRNA

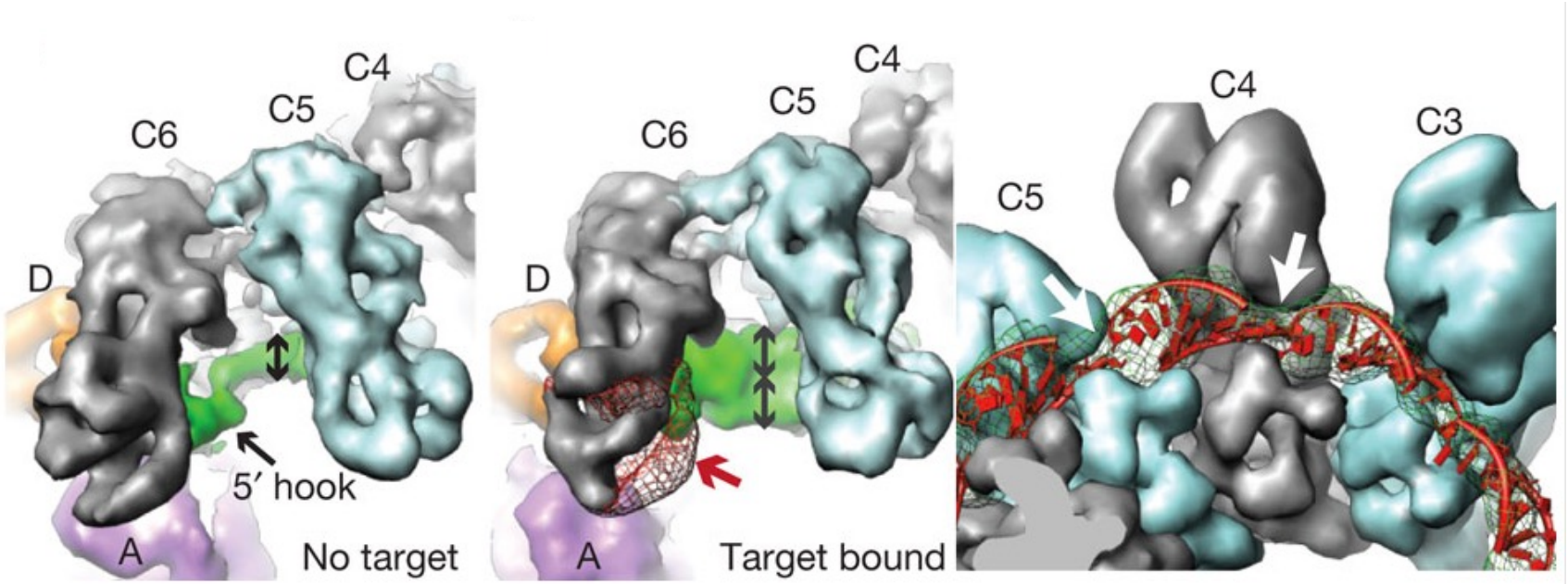


Wiedenheft et al., Nature 2011

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: CASCADE complex bound to crRNA:ssRNA



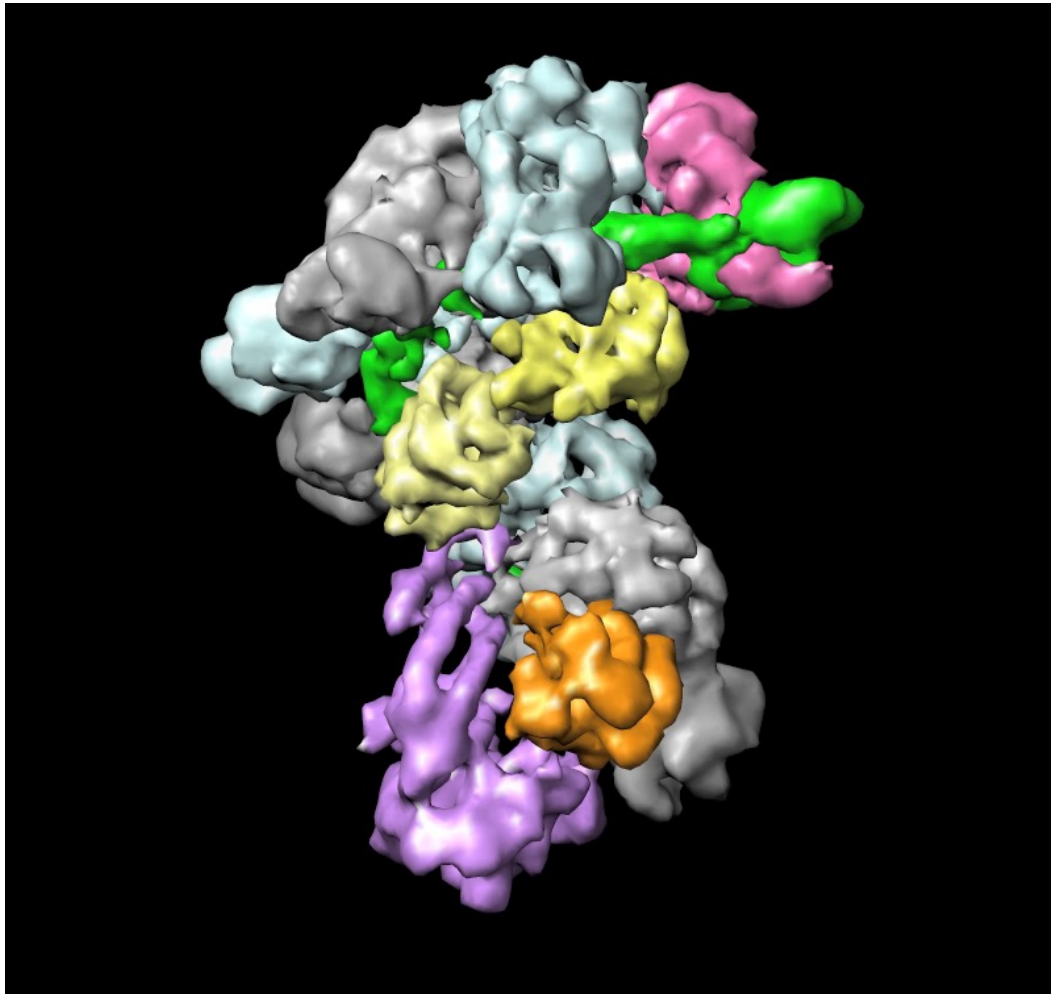
Wider gap at the 5’ end where seed binds

Wiedenheft et al., Nature 2011

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: Conformational changes results from ssRNA binding

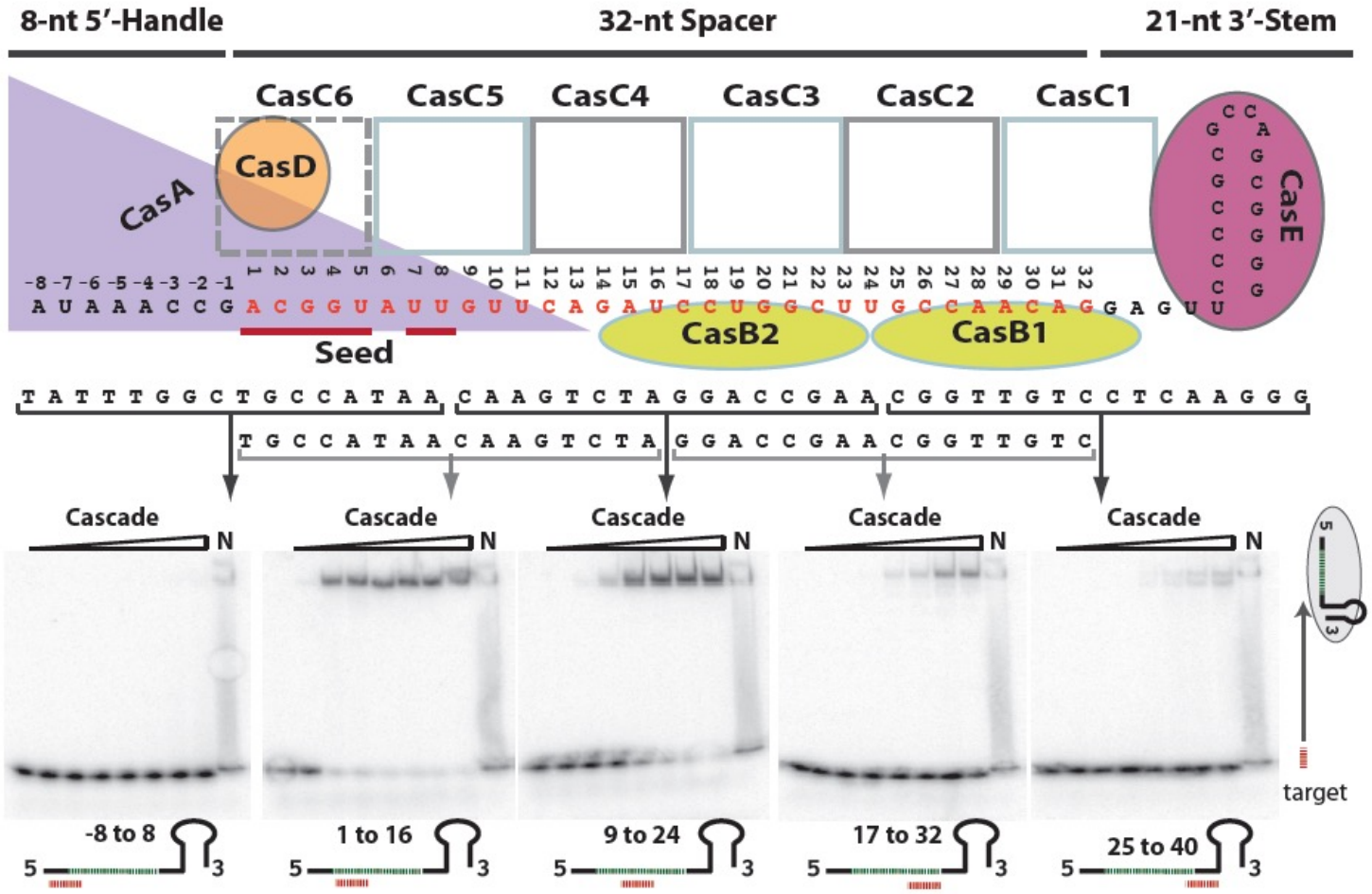


Wiedenheft et al., Nature 2011

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: Affinity for ssRNA stronger near 5’ end where seed binding site



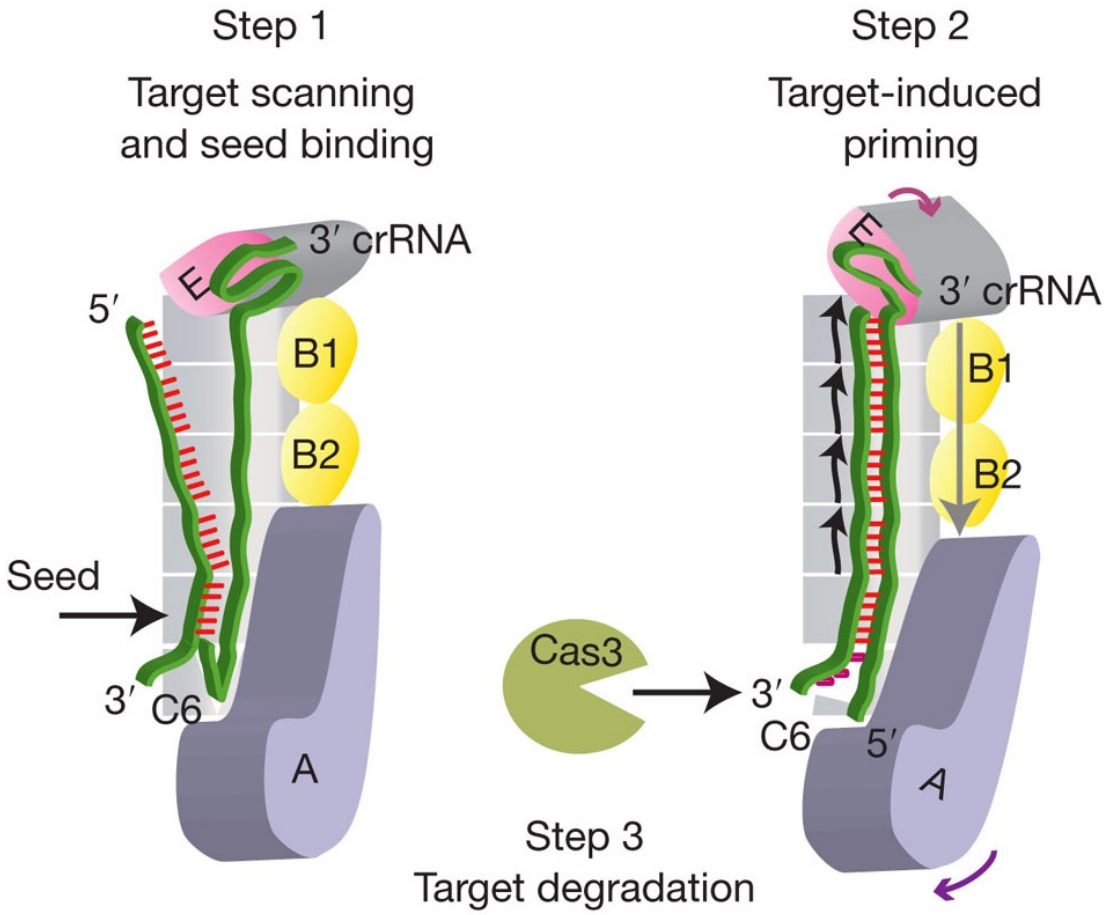
Mutations in seed not tolerated

Wiedenheft et al., Nature 2011

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

CRISPR-CAS: Engagement and scanning followed by conformational changes and Cas3 recruitment



Wiedenheft et al., Nature 2011

Protein-nucleic acid interactions

Nucleases (and how do they find their sites)

Cas9 system (http://www.youtube.com/watch?v=M739wgbcKuA)

Sternberg et al., Nature 2014

Protein-nucleic acid interactions

Discussion Paper:

Jinek et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science. 2014 Mar 14;343(6176):1247997. doi: 10.1126/science.1247997. Epub 2014 Feb 6.

Background Papers:

Wilson and Doudna. Molecular Mechanisms of RNA interference. Annu Rev Biophys. 2013;42:217-39. doi: 10.1146/annurev-biophys-083012-130404.

van der Oost J. et al. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol. 2014 Jul;12(7):479-92. doi: 10.1038/nrmicro3279. Epub 2014 Jun 9.