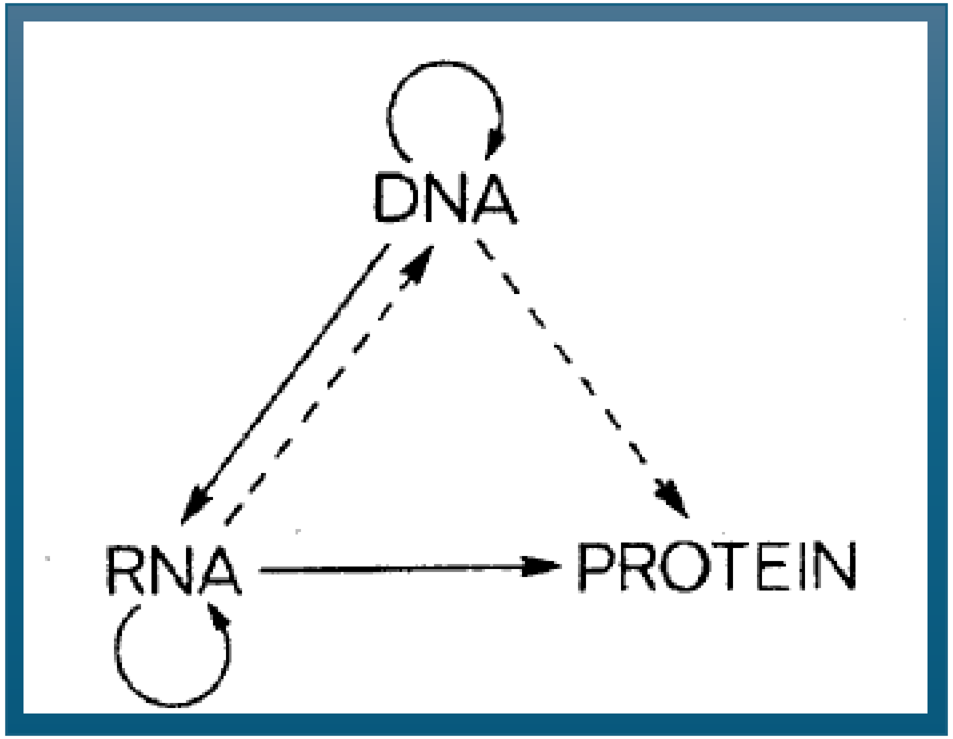
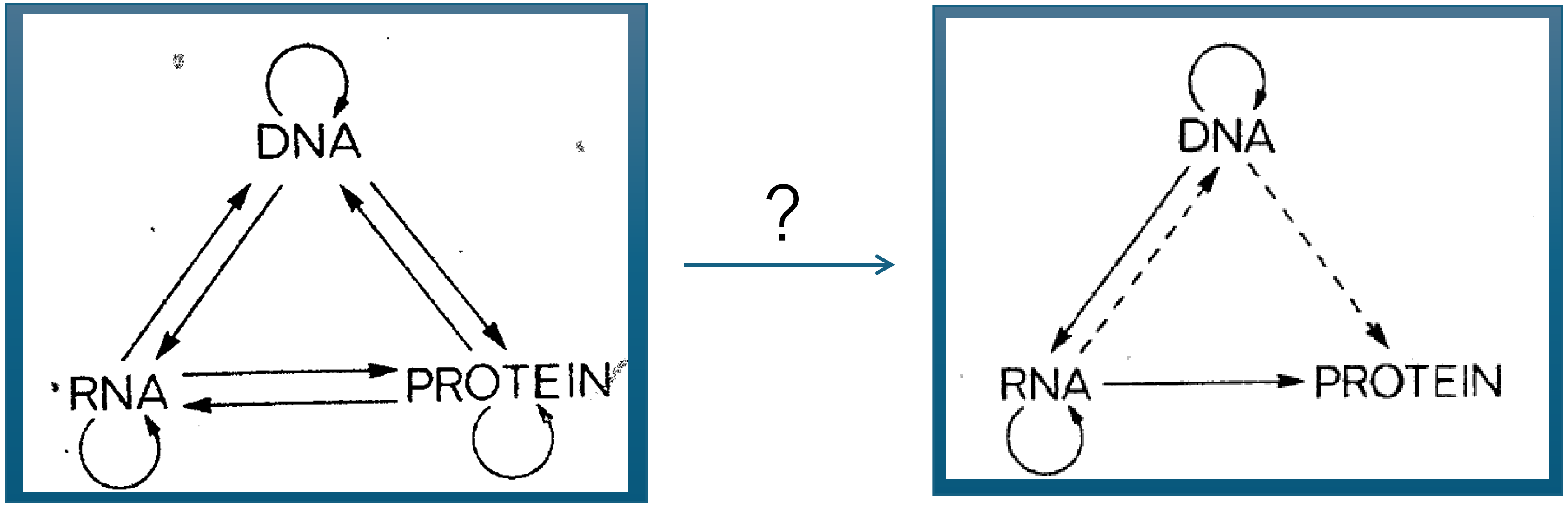
Transcription

Francis Crick's Central Dogma of Molecular Biology circa 1958



Crick Nature 1970

Francis Crick’s Central Dogma of Molecular Biology circa 1958



Crick Nature 1970

DNA encodes the information and RNA is the carrier

1949: Seymour Cohen, 1953: Al Hershey

T2 phage:

When T2 infects E. coli:

bacterial DNA is degraded,
this blocks DNA and protein synthesis in host,
phage DNA is synthesized
phage RNA is rapidly synthesized and degraded
phage proteins are made

1956: Volkin and Astrachan

Infects E. coli with T2 phage in presence of ^{32}P
Phage RNA species has **SAME BASE RATIO** as the phage DNA

Cohen, S. S. (1949) Bacteriol. Rev. 13, 1–24
Hershey, A. D. (1953) J. Gen. Physiol. 36, 777–789
Volkin, E., and Astrachan, L. (1956) Virology 3, 149–161

The hunt for a DNA-dependent RNA polymerase

How would you find an RNA polymerase enzyme?

Grind up living cells to make ‘extracts’

What activity would you look for:

- Elongation of RNA chain
- Use DNA template
- Sequence of product depend on the template

Elongation of RNA chain
Grunberg-Manago and Ochoa 1955

Use DNA template

$$(NMP)_n + NDP \rightleftharpoons (NMP)_{n+1} + P_i$$

Sequence of product depend on the template

Polynucleotide Phosphorylase!

Next Candidates:

1953-1958 various authors

Enzyme uses rNTPs instead of rNDP
Adds only –CCA to the ends of RNA molecules

Terminal tRNA –CCA adding enzyme!

1960 Edmonds and Abrams

dATPs can form a poly(A) chain

Poly(A) polymerase!

First isolation of RNA polymerase:

1959 Weiss and Gladstone
1960 Audrey Stevens
1960 Jerry Hurwitz

1960 Jerry Hurwitz

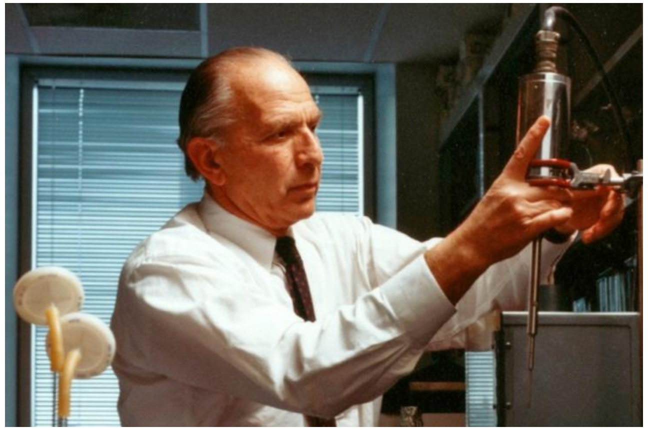


Table I

REQUIREMENTS FOR UMP INCORPORATION

	mμmoles incorporated
Complete System	0.81
-ATP	0.05
-GTP	0.07
-CTP	0.05
-ATP, GTP or CTP	<0.02
- Nucleic acid	0.17
Complete + DNase (1 μg)	<0.02
Complete + RNase (1 μg)	<0.02

The complete system (0.5 ml) contained uridine P^*-P-P (30 mμmoles, 1 x 10^6 cpm/μmole), GTP, ATP and CTP (50 mμmoles) E. coli nucleic acid (2 optical density units at 260 mμ), $MgCl_2$ (2 μmoles), $MnCl_2$ (2 μmoles), mercaptoethanol (1 μmole), acetyl phosphate (4 μmoles), 0.04 units of acetokinase (Rose et al.), and 40 μg of an ammonium sulfate fraction obtained from E. coli W. After 20 minutes at 38°, the reaction was stopped with 0.2 ml of 7 per cent $HClO_4$ and 0.05 ml of 10 per cent albumin was added as carrier. The acid-insoluble material was washed 3 times with 1 per cent $HClO_4$ and dissolved with NH_4OH, plated and the radioactivity measured.

16

Different sources of DNA

Table II

NUCLEIC ACID REQUIREMENT

Exp.	Nucleic Acid Fraction	mμmoles incorporated
1	*E. coli* nucleic acid	0.32
2	T2-DNA	0.52
3	Thymus DNA	0.40
4	*E. coli* RNA	0.03
5	No addition	0.04
6	1 or 3 + DNase	<0.02
7	1 or 3 + RNase	<0.02

The additions were as in Table I with the exception that T2 DNA (1 optical density unit at 260 mμ) and thymus DNA (2 optical density units at 260 mμ) replaced the *E. coli* nucleic acid fraction where indicated and Mg^{++} (8 μmoles) was added instead of the Mg^{++} - Mn^{++} mixture. The enzyme preparation was pretreated with DNase prior to the assay (0.005 μg of DNase per ml of enzyme solution); this treatment reduces the amount of incorporation in the absence of added DNA.

Different sources of DNA

Table II

NUCLEIC ACID REQUIREMENT

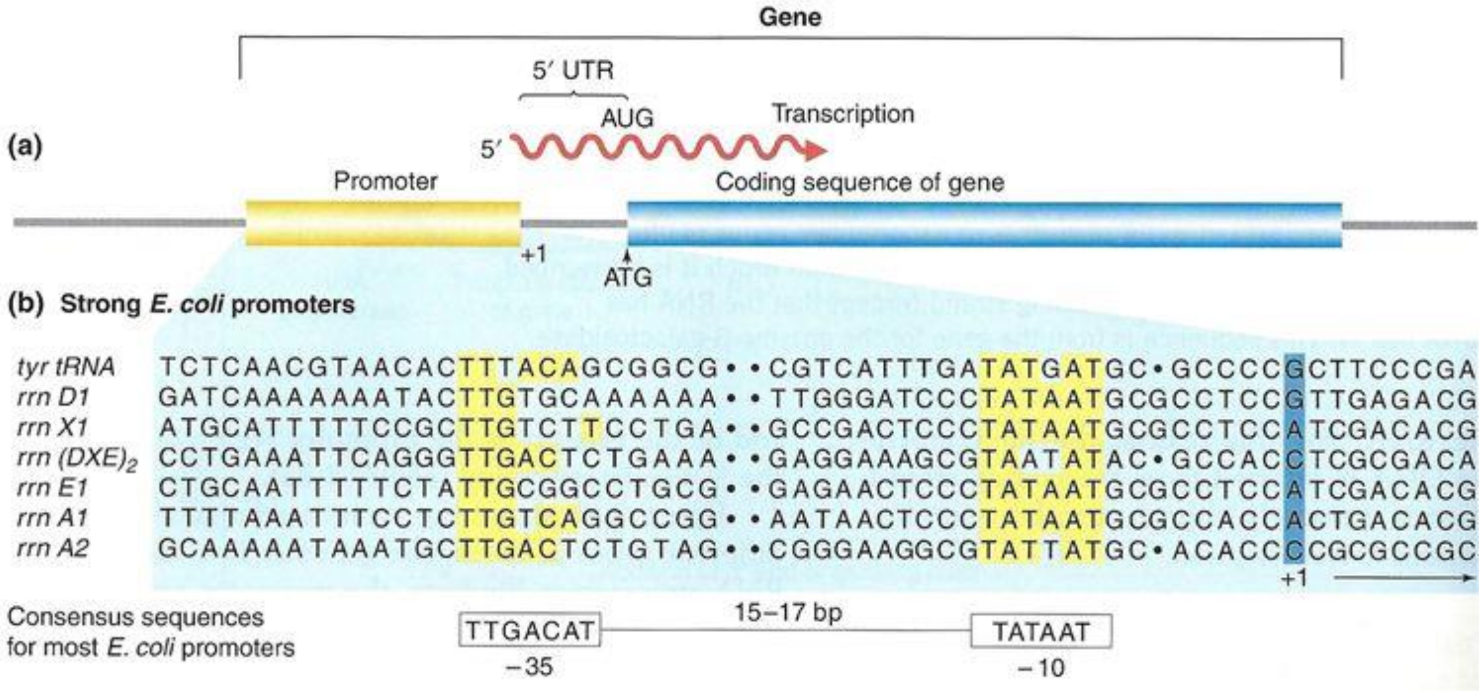
Exp.	Nucleic Acid Fraction	mμmoles incorporated
1	*E. coli* nucleic acid	0.32
2	T2-DNA	0.52
3	Thymus DNA	0.40
4	*E. coli* RNA	0.03
5	No addition	0.04
6	1 or 3 + DNase	<0.02
7	1 or 3 + RNase	<0.02

The additions were as in Table I with the exception that T2 DNA (1 optical density unit at 260 mμ) and thymus DNA (2 optical density units at 260 mμ) replaced the *E. coli* nucleic acid fraction where indicated and Mg^{++} (8 μmoles) was added instead of the Mg^{++} - Mn^{++} mixture. The enzyme preparation was pretreated with DNase prior to the assay (0.005 μg of DNase per ml of enzyme solution); this treatment reduces the amount of incorporation in the absence of added DNA.

How do you impart specificity on an enzyme that can transcribe any DNA?

Promoter = DNA sequence coding for transcription initiation

E. Coli consensus promoter

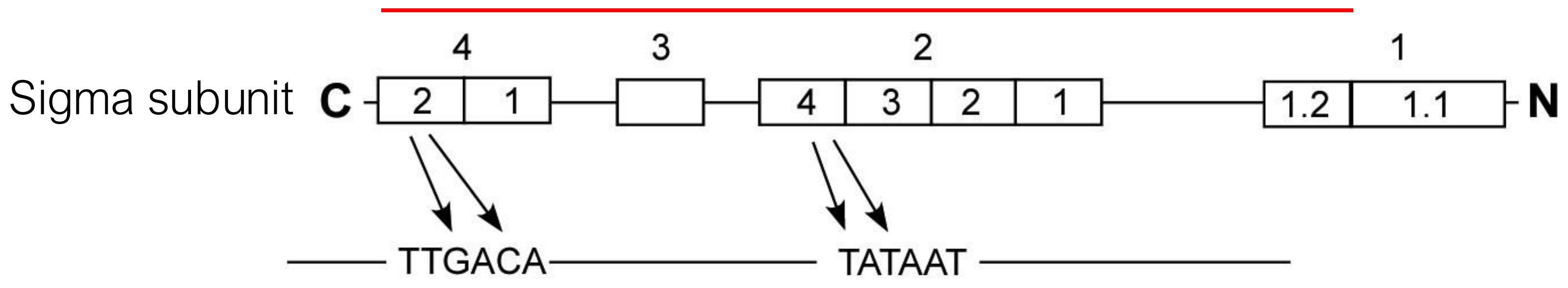


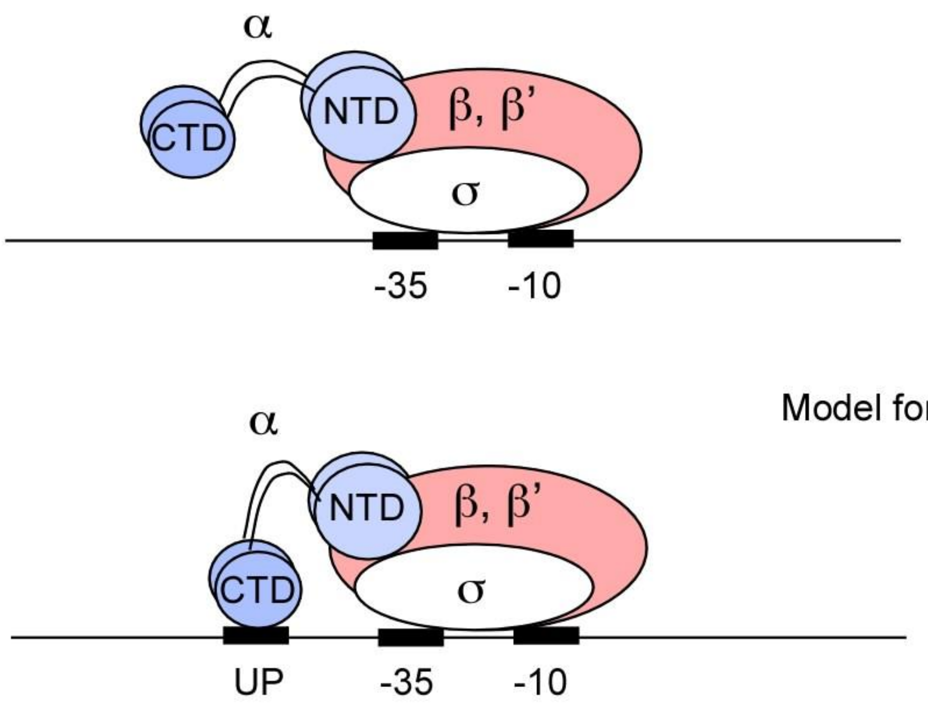
Sigma factors impart specificity on bacterial polymerase

E. coli RNA Polymerase:

β, β', $\alpha 2$, ω	Core	
β, β', $\alpha 2$, ω, σ		Holoenzyme

			Promoter Consensus
	σ^{70}	Housekeeping genes	-35, -10
Some alternative σ factors	σ^{54}	Genes controlled by nitrogen levels	-24, -12
	σ^{32}	Heat shock genes	-35, -10



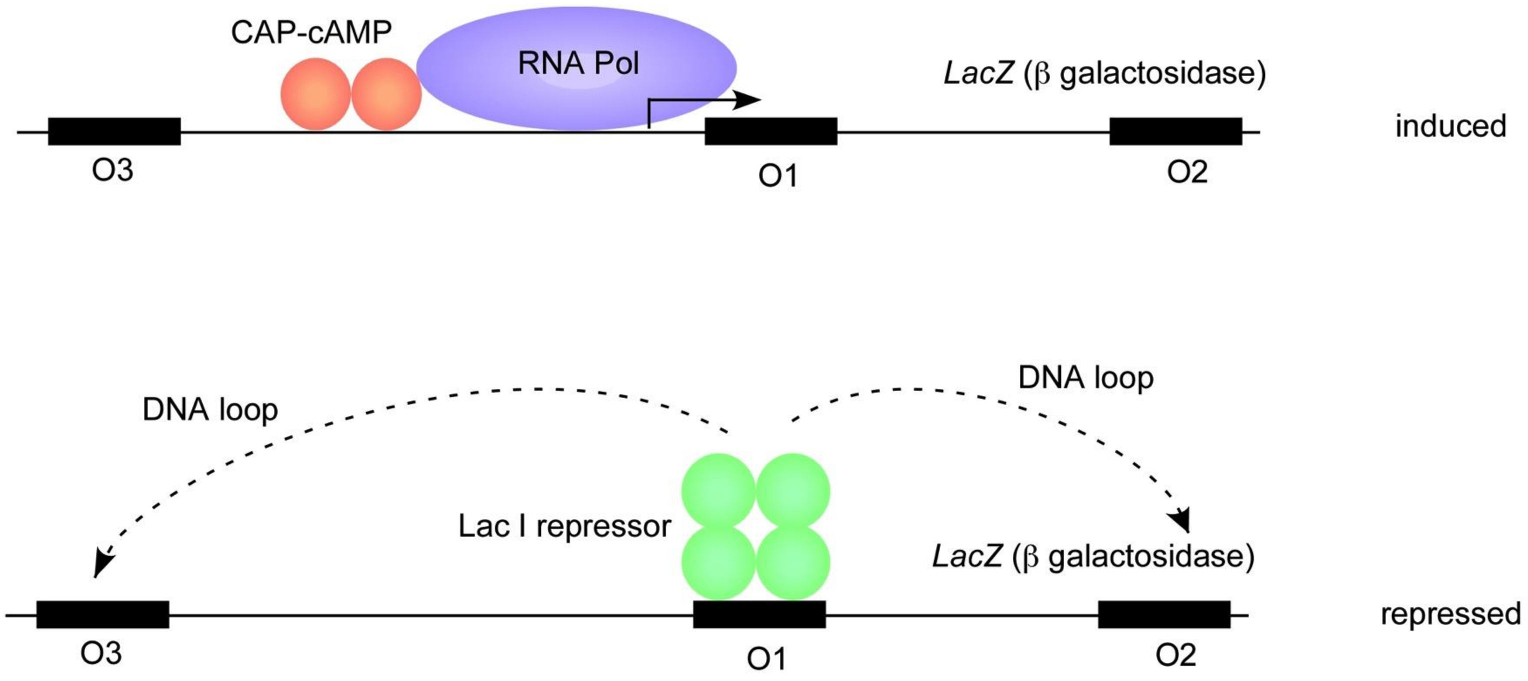


Model for Up element function

Abortive initiation

$$\text{RNA Pol + Promoter} \rightleftharpoons \text{RPc} \longrightarrow \text{RPo} \xrightarrow[\text{NTPs}]{} \text{RNA}$$

Lac Promoter

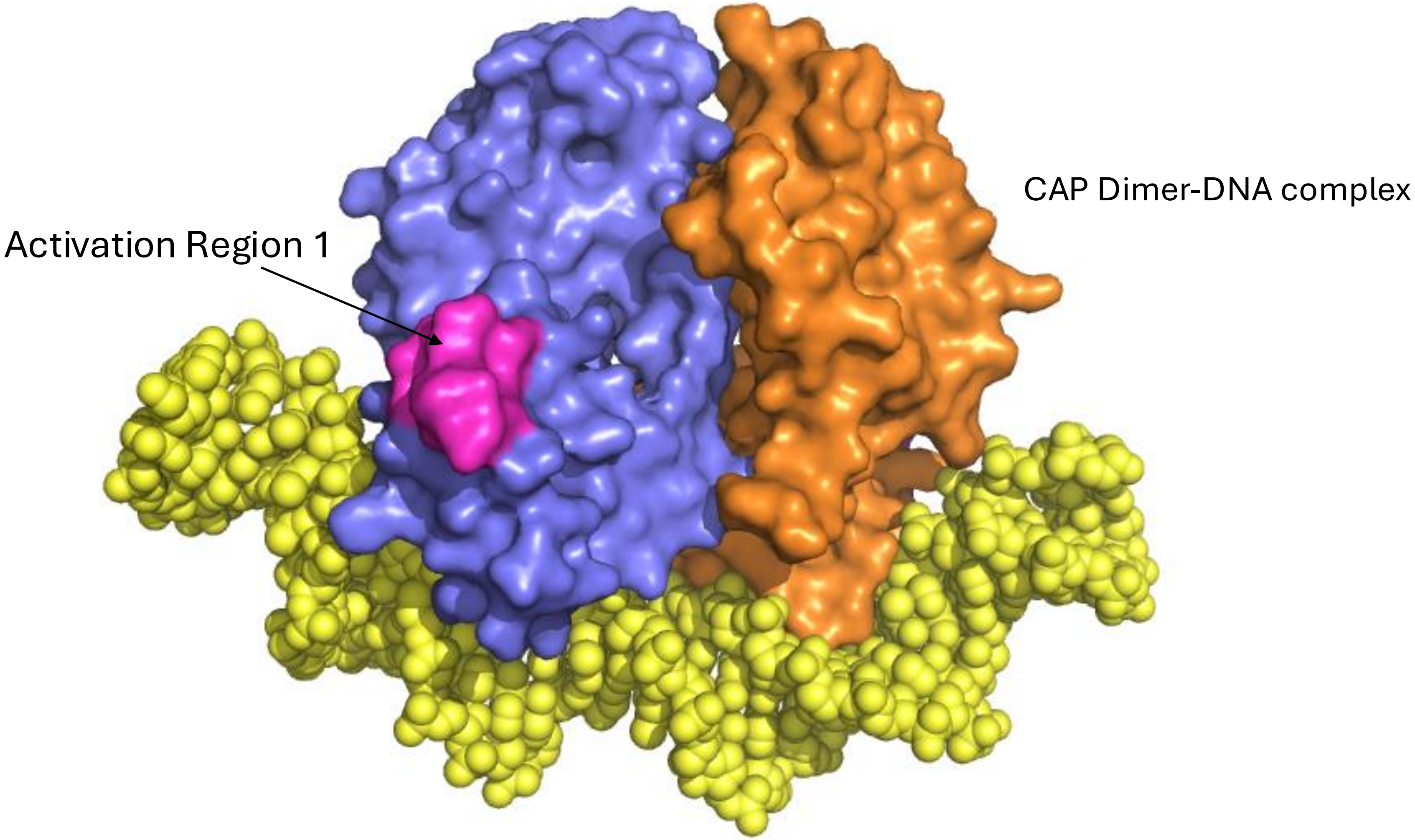


Lac Repressor binding inhibits binding of RNA Pol. Loop between two operators creates stable operator-DNA complex.

DNA binding of repressor regulated by inducer (allolactose; IPTG)

cAMP regulates CAP binding to DNA

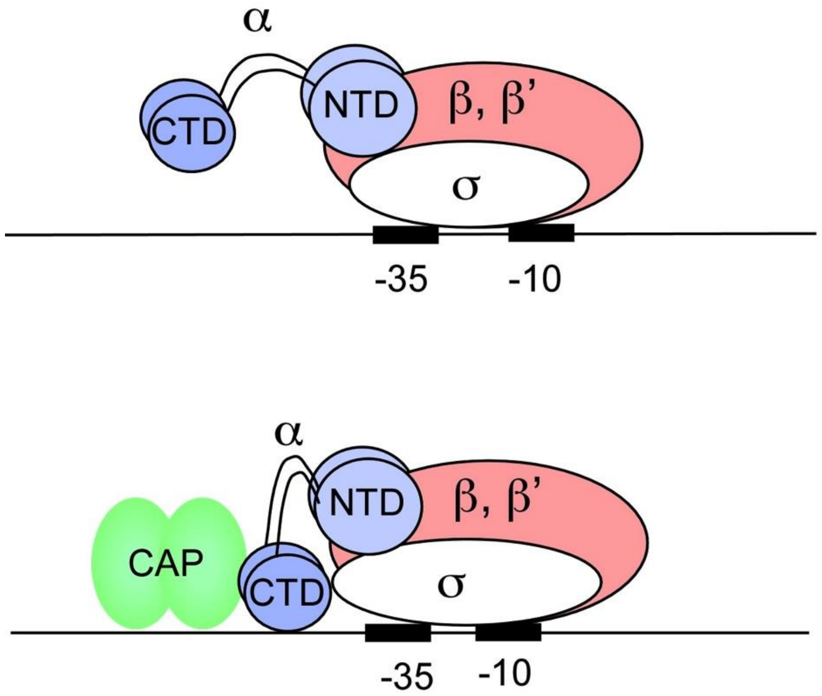
Activating region required for CAP activity at class I promoters



Ebright et al

Genetic selection for Pol alpha subunit mutations that are defective for activation by CAP

α mutations found with the predicted phenotype. Mutations at positions 258, 261, 264, 265 (all in the C-terminal domain)

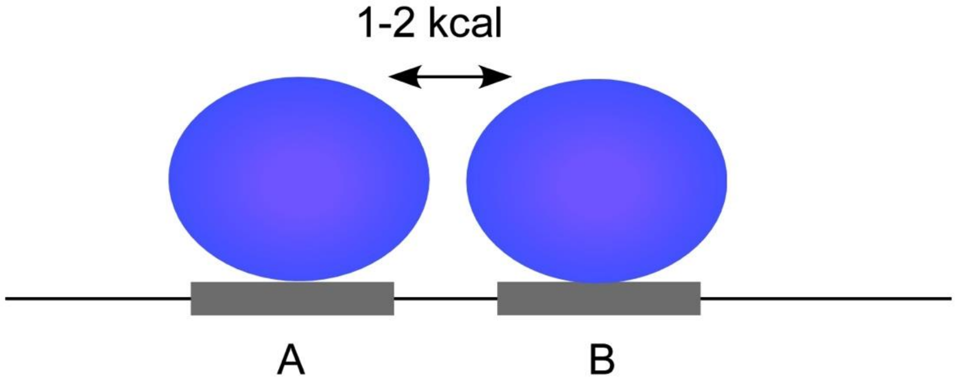


Model for CAP activation at Class 1 promoters

Tang and Ebright, 1994

Cooperative DNA Binding

Energy of protein-protein interactions is used to increase
increase the affinity of a protein for its binding site on DNA



$$\text{P-DNA} \rightleftharpoons \text{P + D} \qquad Kd = \frac{[P]\,[D]}{[P\text{-}D]} \qquad Kd = e^{\Delta G/RT}$$

Example: Kd = 6 nM at room temperature; ΔG = -11 kCal/mole

1 kCal/mole from protein-protein interaction would increase the affinity ~5X

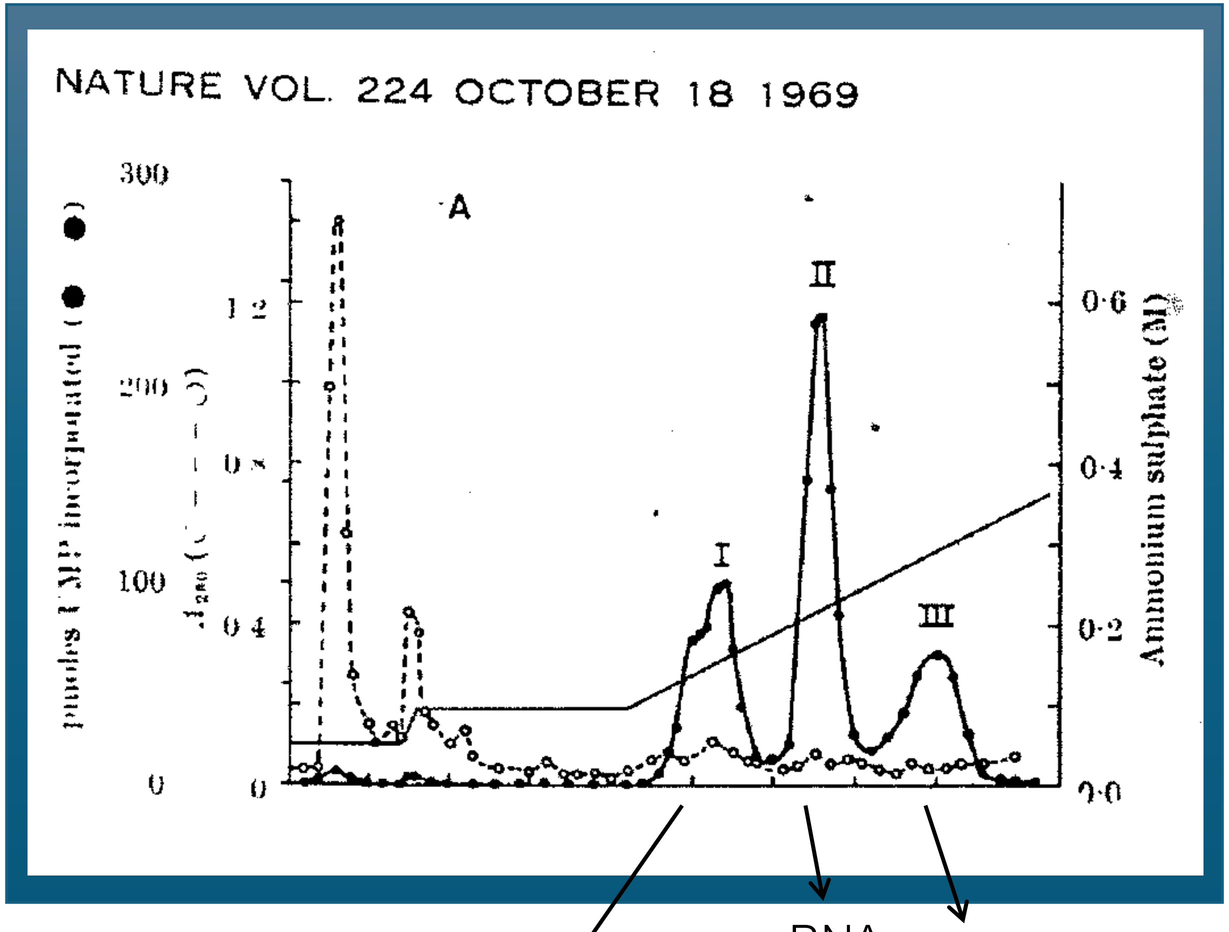
2 kCal/mole from protein protein interaction would increase the affinity ~30x

How do you find RNA polymerases in eukaryotes?

Make an extract that has RNA polymerase activity

Separate the proteins of the extract using columns with different resins
- proteins stick to resin with different affinities
- wash off – those that don't stick as well will wash out earlier

First isolation of multiple eukaryotic RNA polymerases:



rRNA

mRNA
snRNA
miRNA
PiRNA

tRNA
U6 snRNA
5s rRNA

Roeder and Rutter Nature 1969

Two types of RNA polymerases, single subunit and multi subunit

Single subunit Pols:
(mitochondrial and viral RNA polymerases)

generally resemble DNA polymerases

probably been “invented” several times during evolution
based on evolutionarily unrelated folds of active sites

Multi subunit Pols:
(all cellular organisms)

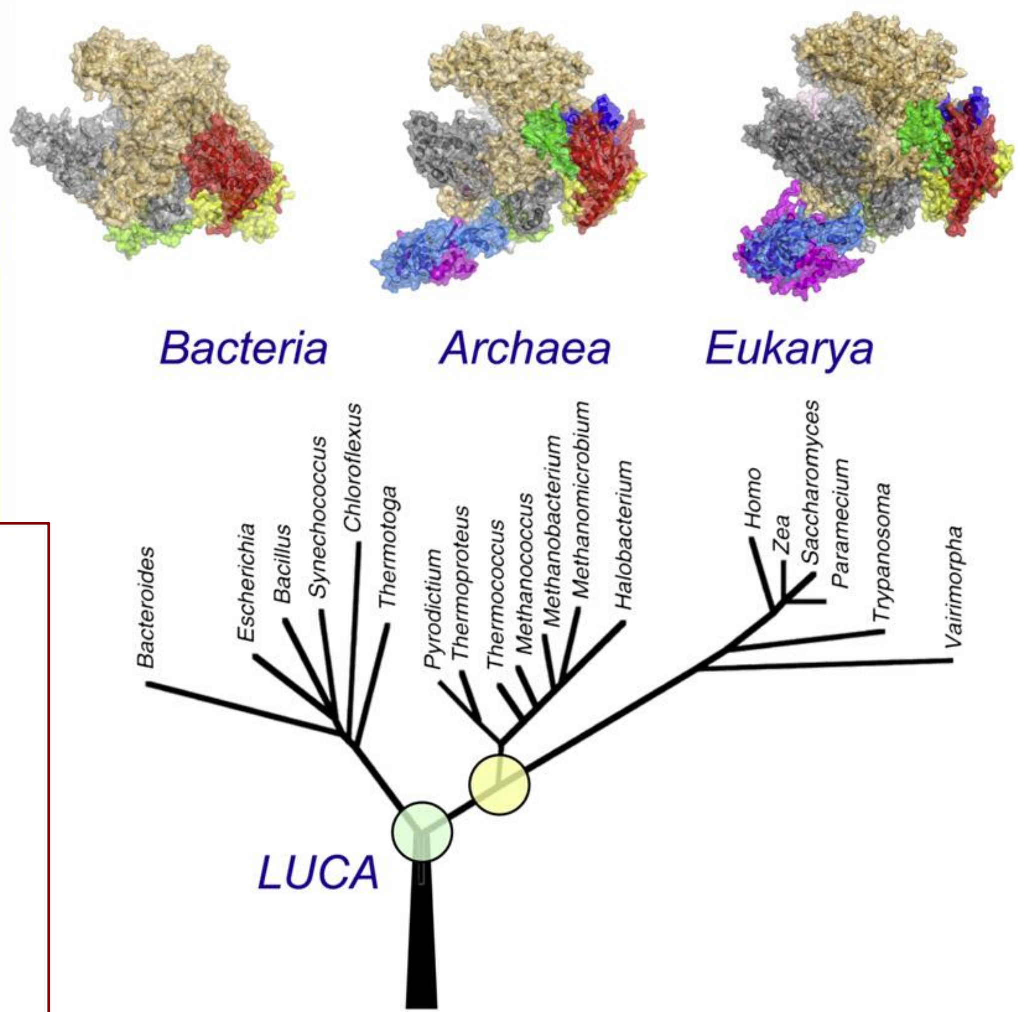
have a common structural framework and act by near
identical mechanisms.

Conservation of multi subunit RNA polymerases

A

	Bacteria	Archaea	Eukarya
RNAP subunits	β'	A'/A"	RPB1
	β	B"/B'	RPB2
	α	D	RPB3
	α	L	RPB11
	ω	K	RPB6
		F	RPB4
		H	RPB5
		E	RPB7
		G	RPB8
		*TFS**	RPB9
		N	RPB10
		P	RPB12
basal transcription factors		*TBP*	*TBP*
		TFB	*TFIIB*
		*TFS**	*TFIIS*
		Spt4	*Spt4*
	NusG	*Spt5*	*Spt5*
	NusA	*NusA*	
		TFEα	*TFIIE*α
	$\sigma_{2.3\text{-}3.2}$	*TFE*β*/C34*	*TFIIE*β
			TFIIFα
	$\sigma_{1.2\text{-}2.3}$		*TFIIF*β

B



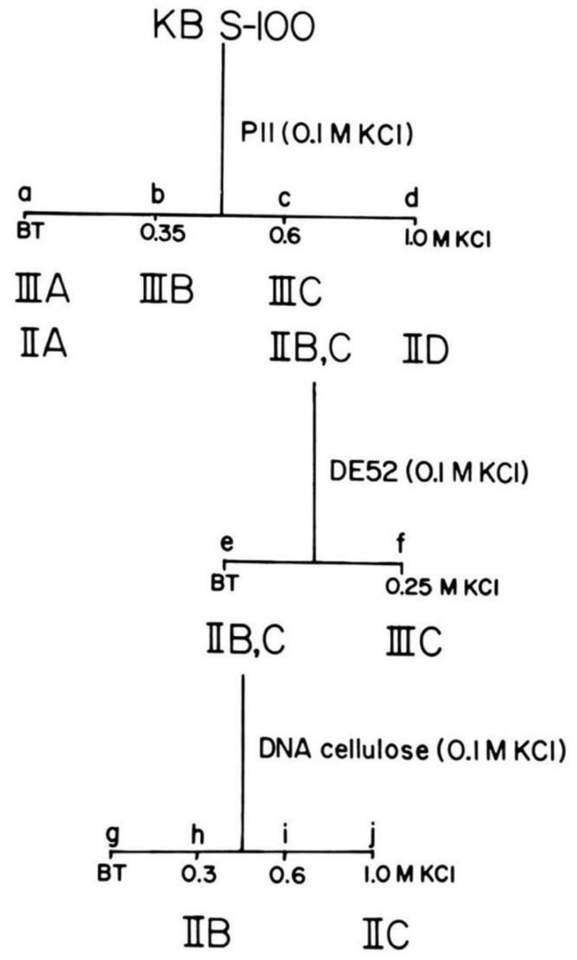
Werner, Grohmann 2011

How to find General Transcription Factors?

Purified polymerase initiates transcription
from a promoter in a random fashion

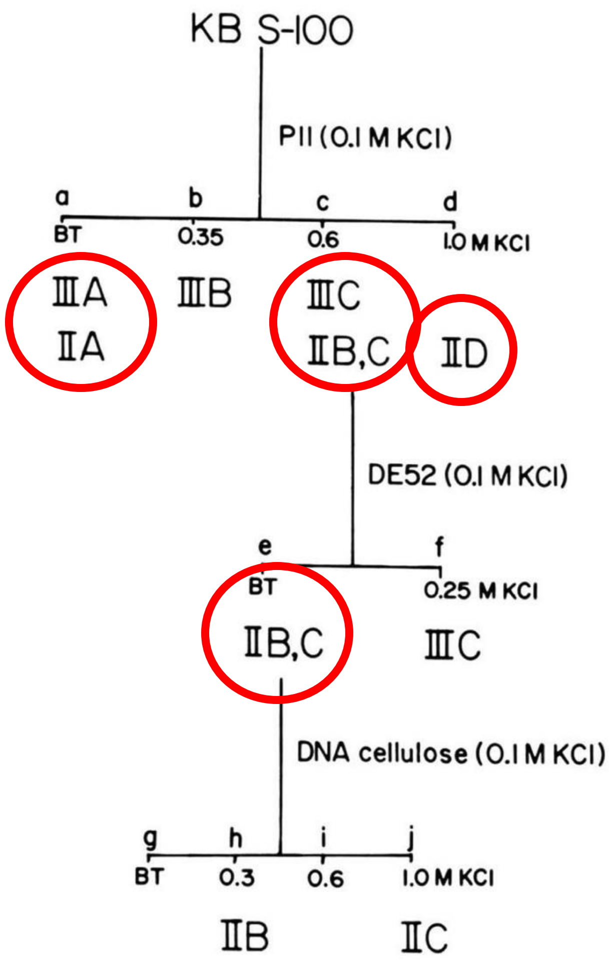
Look for proteins that give accurate
transcription from a promoter

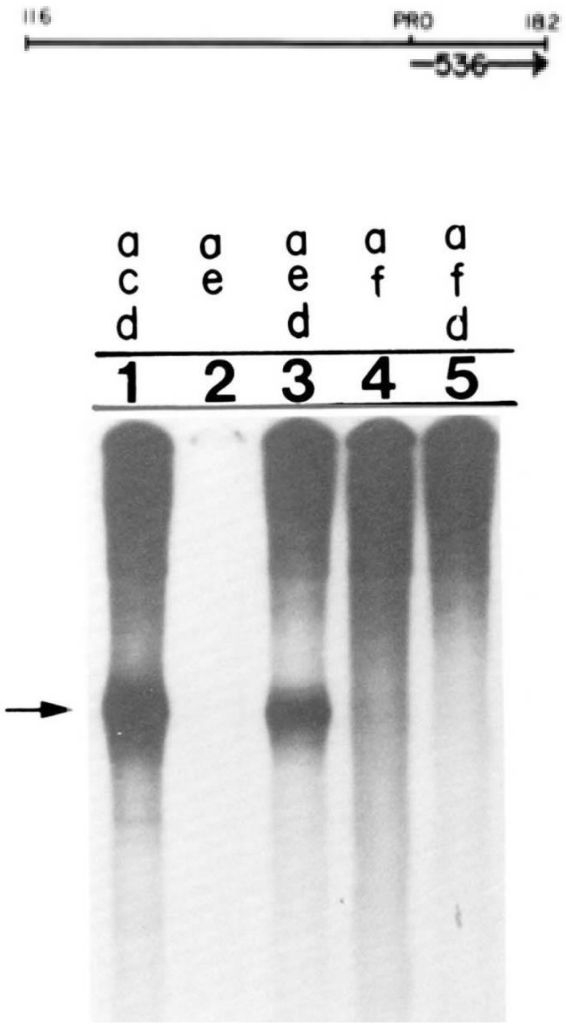
Make cell-free extracts
Run them over charged columns
 some proteins stick more than others
 wash them off with increasing salt
 collect these 'fractions' and test them in assay



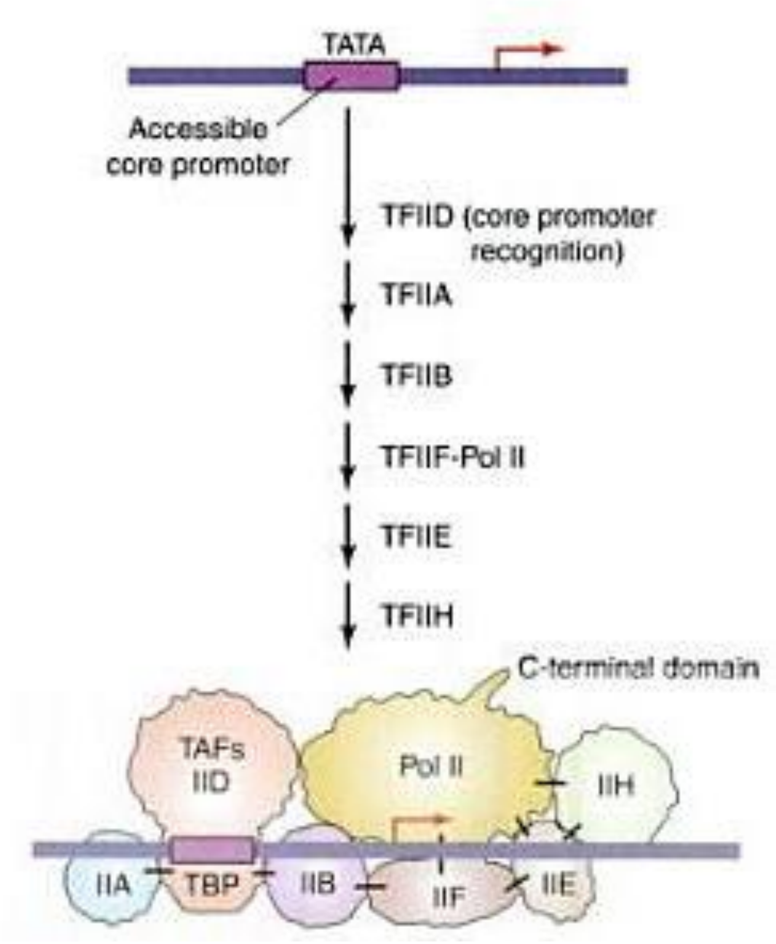
Matsui, Segall, Weil, Roeder (1980) JBC 255:11992

Assay for General Transcription Factors





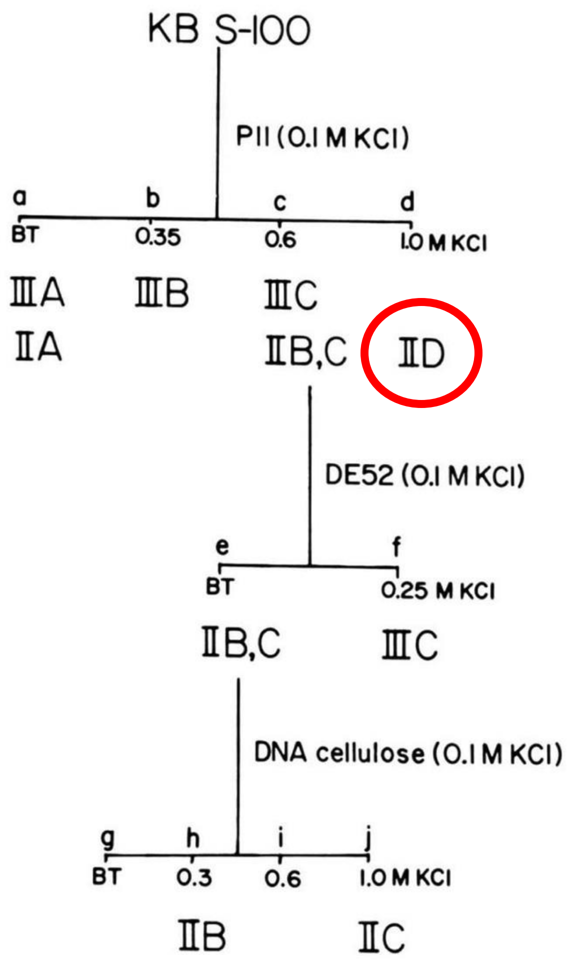
Order of recruitment of the general transcription factors

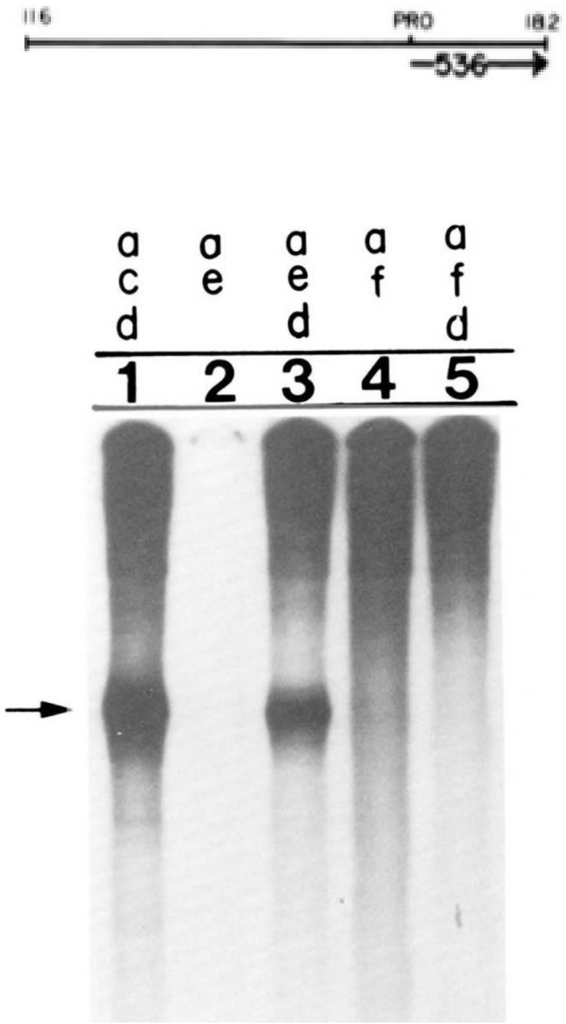


Sawadogo and Roeder 1985

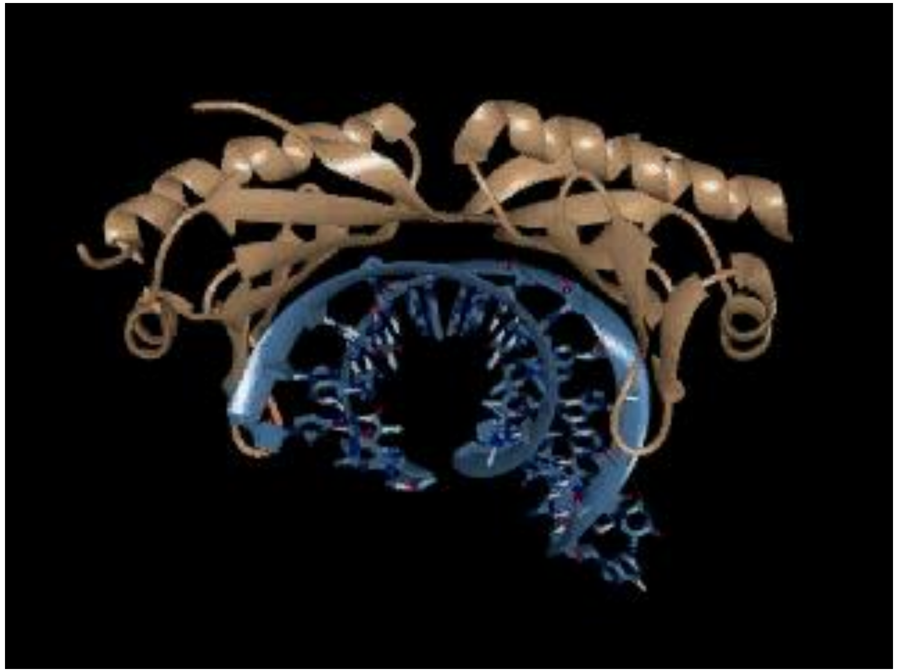
Figure from Tropp and Freifelder

Assay for General Transcription Factors





IID = TBP plus coactivator TAFs (TBP-Associated factors)



Tan lab 2005

Kim et al (Sigler); Kim et al (Burley);(1993)

The TBP structure appears symmetric.

Why could this be a problem?

TATA binding Protein (TBP)

180 amino acid core that binds the minor groove of the TATA box recognizing 8 bp sequence.

When binding TATA, TBP unwinds the TATA, kinking the DNA at two positions and bending the DNA path about 90 degrees.

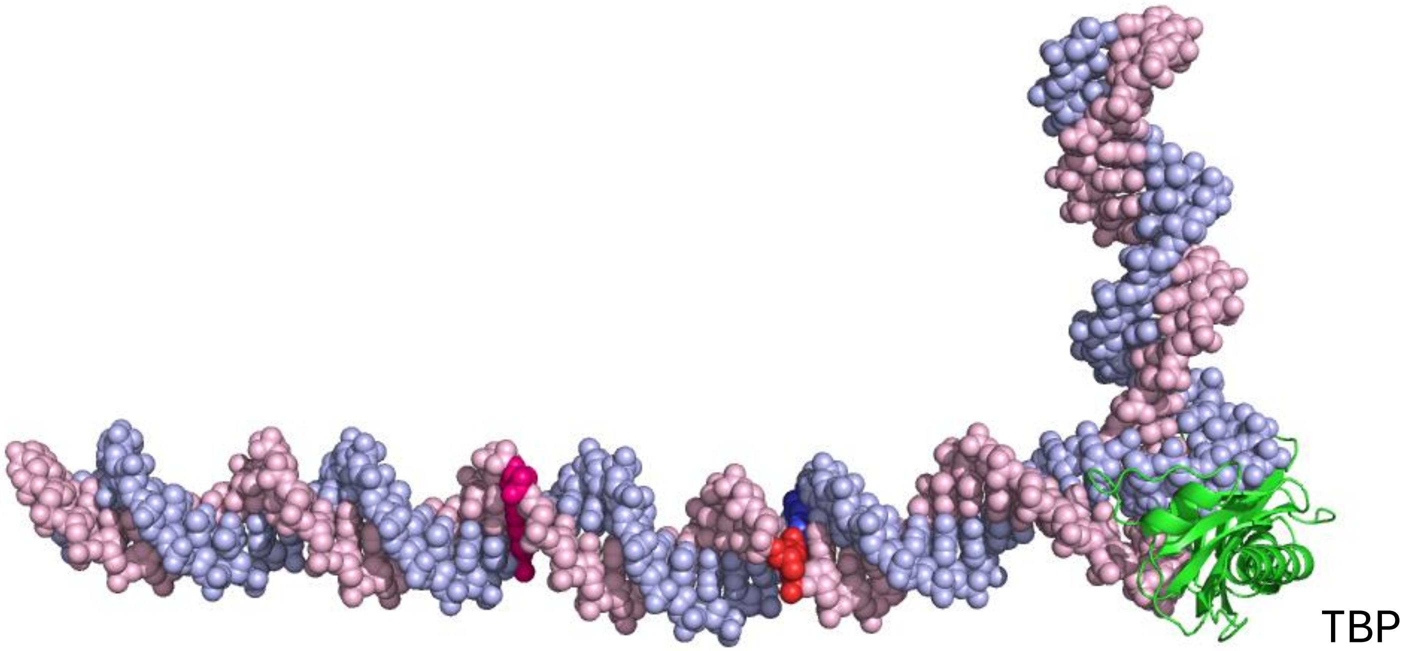
TBP is conserved from archaea to humans.

Yeast has one copy of TBP while Drosophila has three.

In *Drosophila*, TBP is used for nearly all Pol II genes. TBP related factor 1 (TRF1) is used for transcription of Pol III genes while TRF2 and TRF3 are used at selected Pol II genes. TRF3 may be very important in differentiated cells.

Function of core promoter elements in PIC assembly: TATA-containing promoters

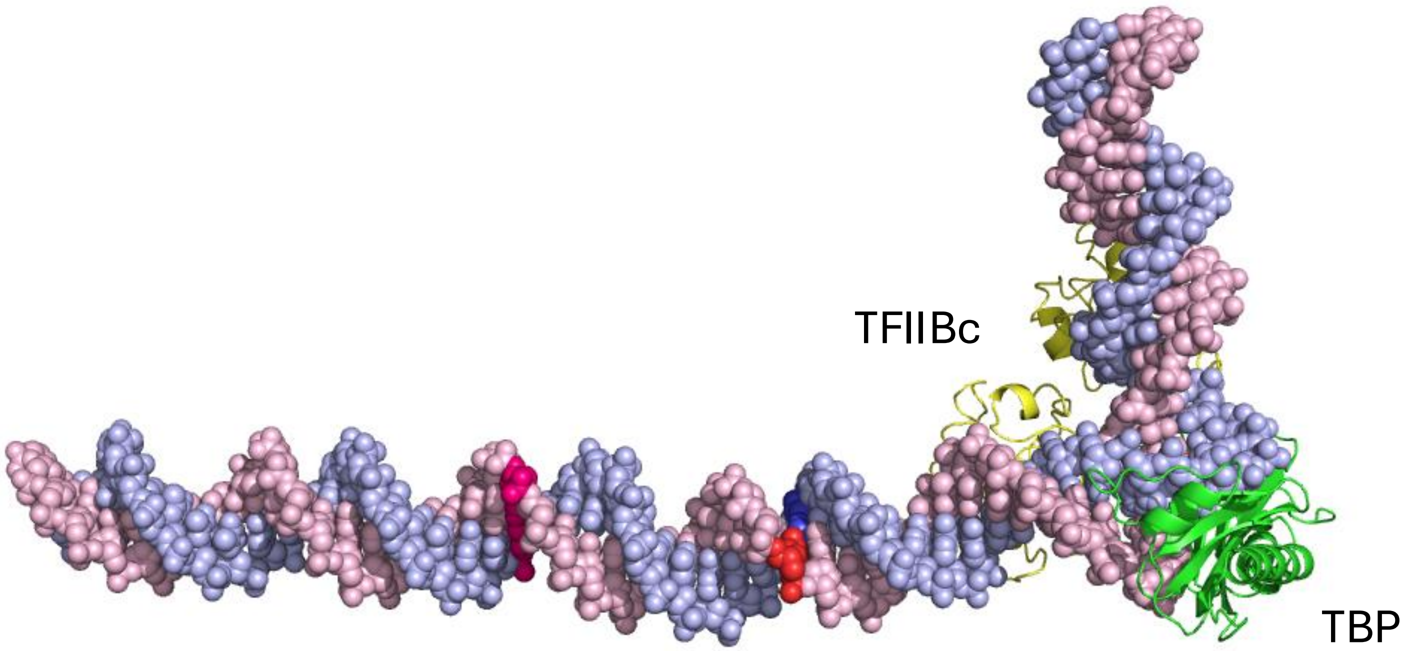
TBP-DNA complex



Kim Y et al, 1993; Kim J et al 1993

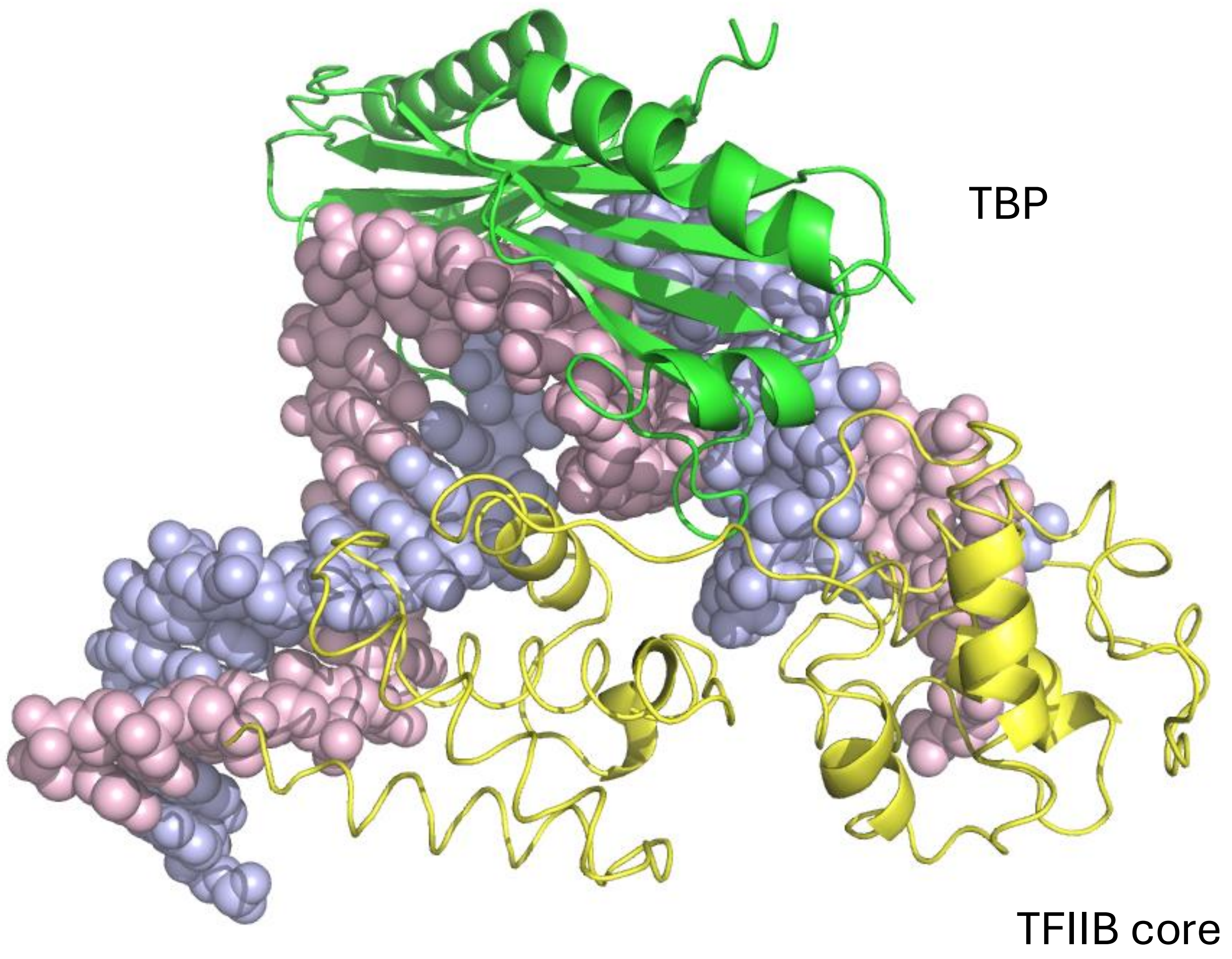
Function of core promoter elements in PIC assembly: TATA-containing promoters

TBP-TFIIB-DNA complex



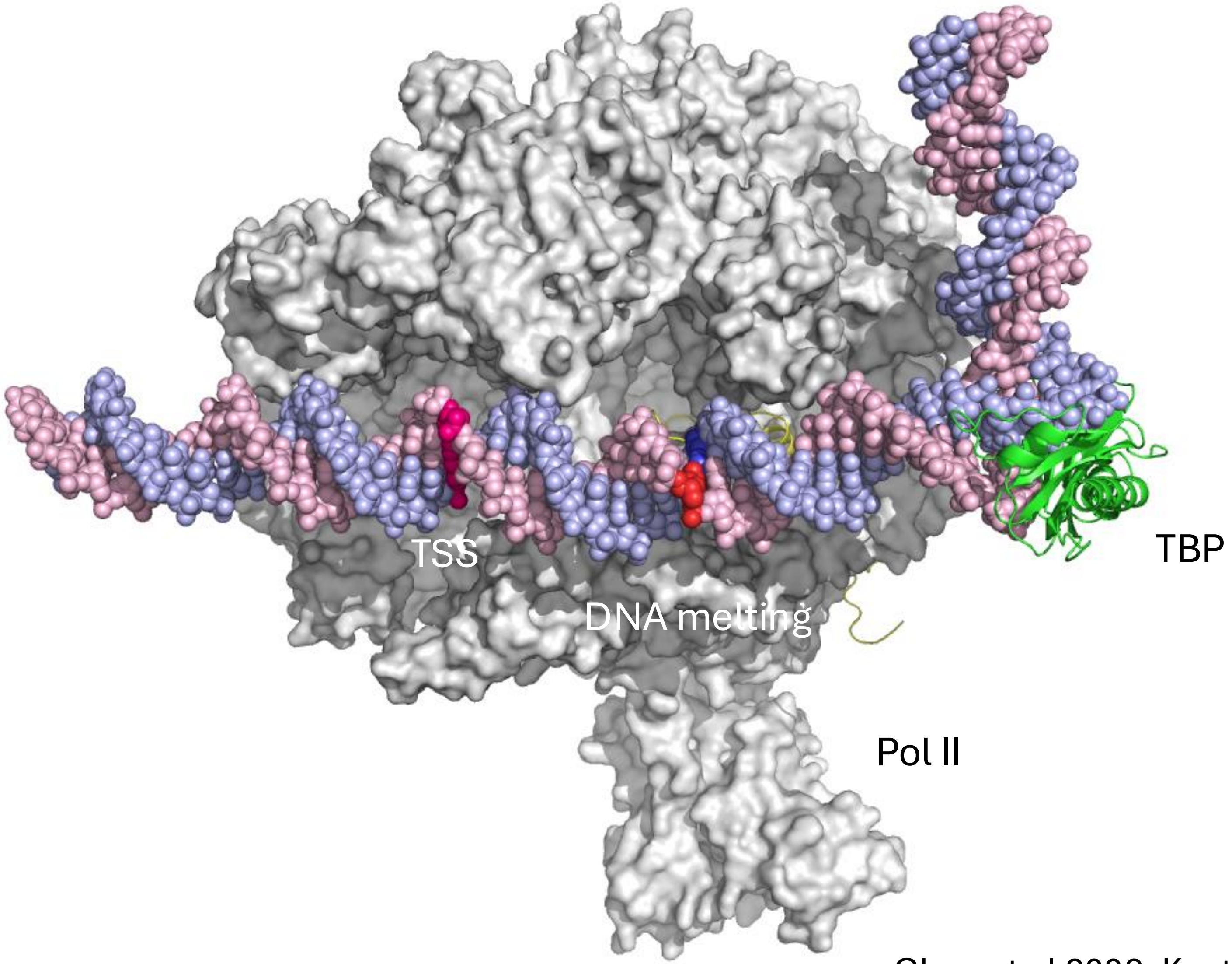
Nikolov et al 1995

The TFIIB core domain binds TBP, DNA upstream and downstream of TATA and Pol II



Function of core promoter elements in PIC assembly: TATA-containing promoters

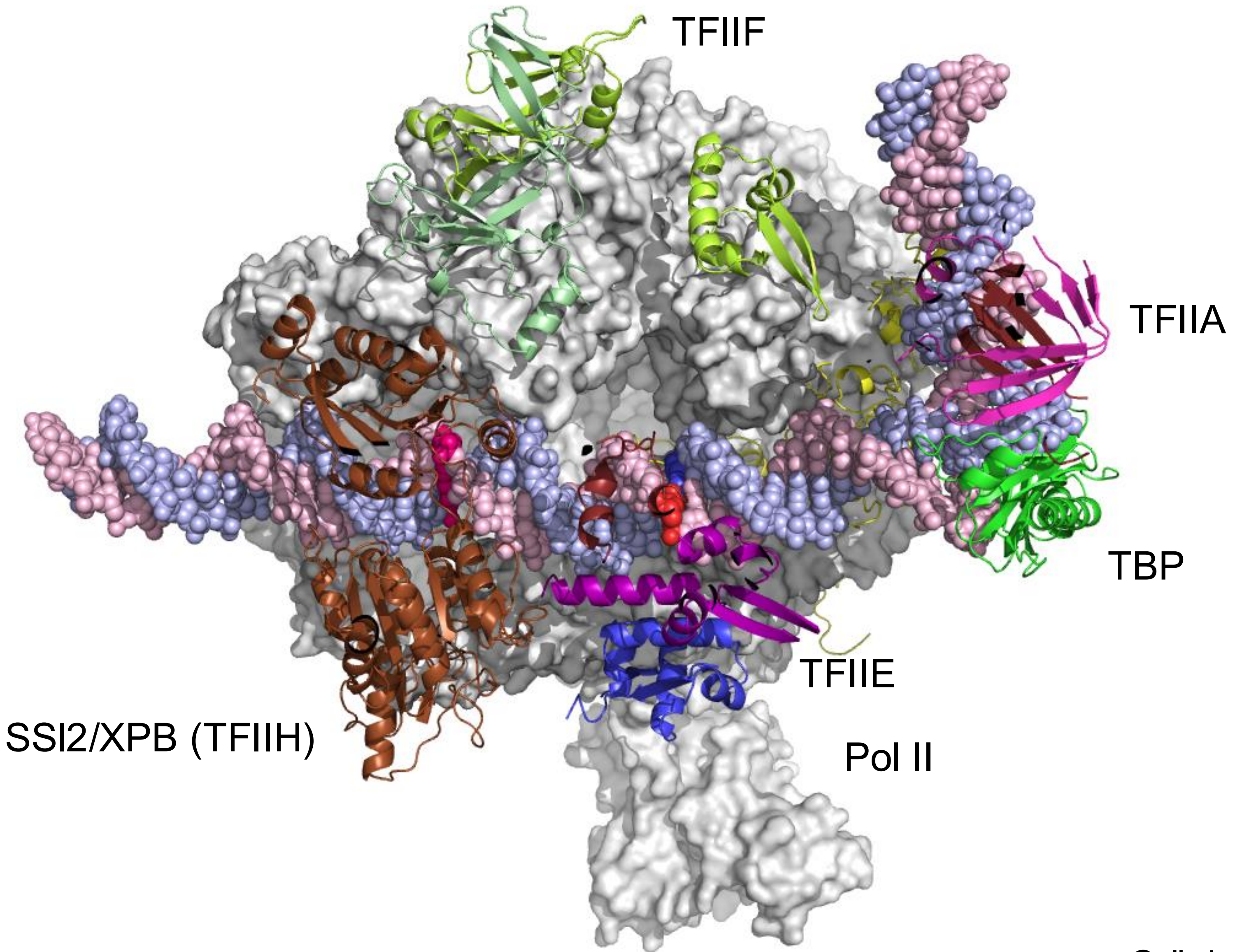
TBP-TFIIB-Pol II-DNA complex



Chen et al 2009; Kostrewa et al 2009

Complete PIC

PIC assembled from many weak protein-protein and protein-DNA interactions



Grünberg et al 2012

After PIC formation, how does Pol II unwind DNA and select a transcription start site?

ATP
Ssl2/XPB (TFIIH)

Pol II
GTFs
Promoter → **PIC** → **Open Complex** → **Start site scanning** → **Initiation** →→→ **RNA**

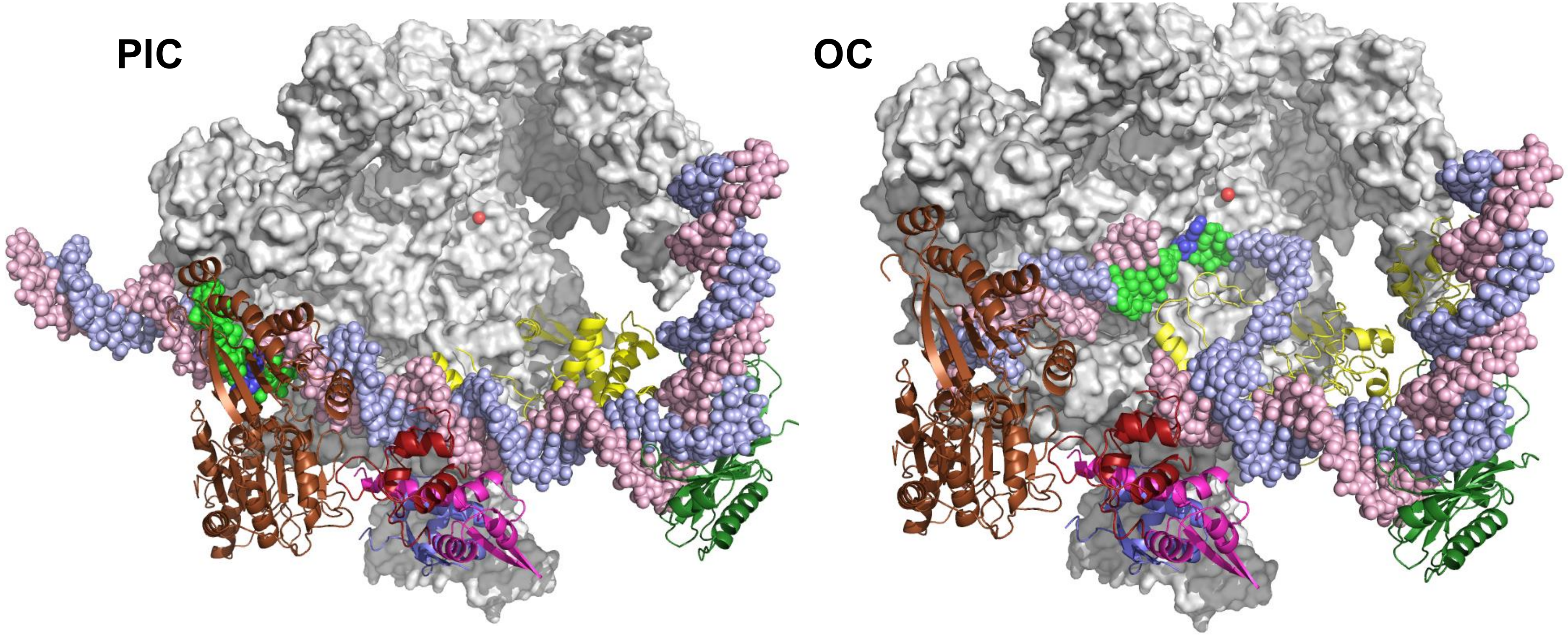
DNA strand unwinding

The ATPase TFIIH subunit Ssl2/XPB and ATP are used to unwind ~10 bp DNA and promote insertion of the template strand into the Pol II cleft

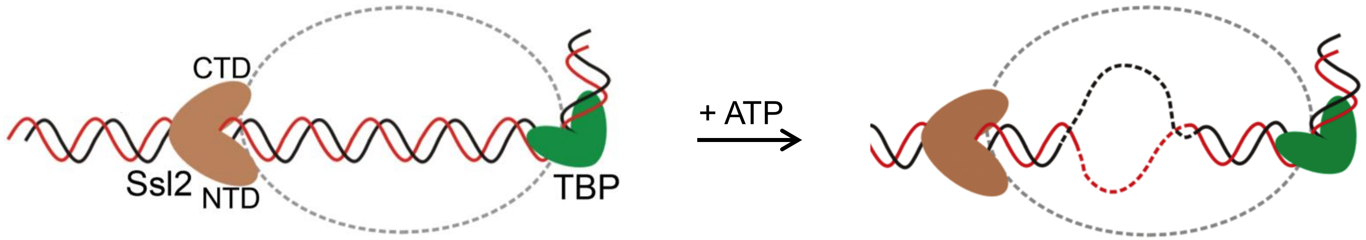
Next: Pol II scans for the transcription start site

The yeast TSS can be 20-100 bp distant from the initial site of DNA opening

Model for ATP-dependent promoter opening

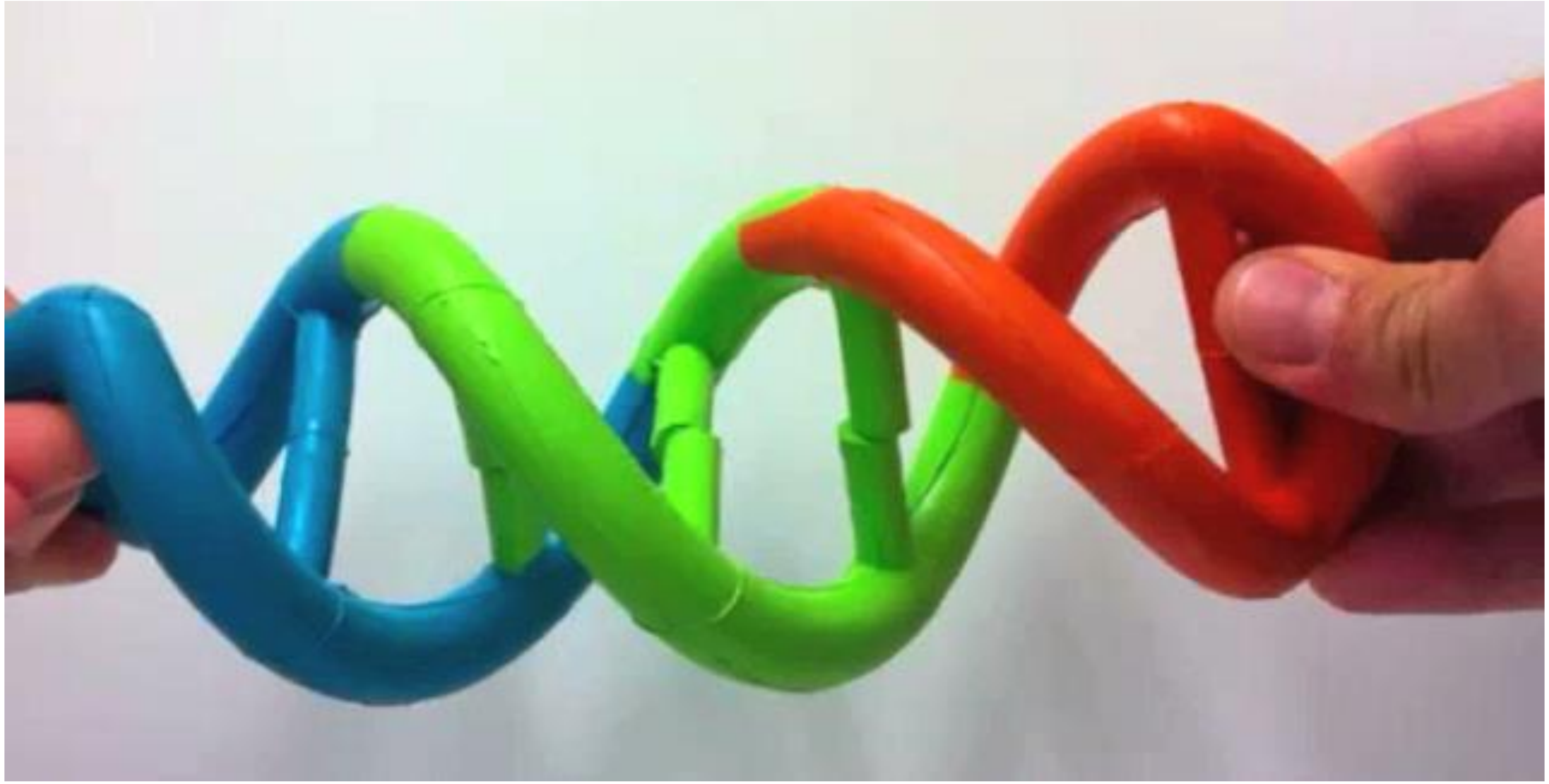


Comparison of the PIC and open complex (OC) models suggests a new role for Ssl2 in DNA strand opening



Sebastian Grünberg

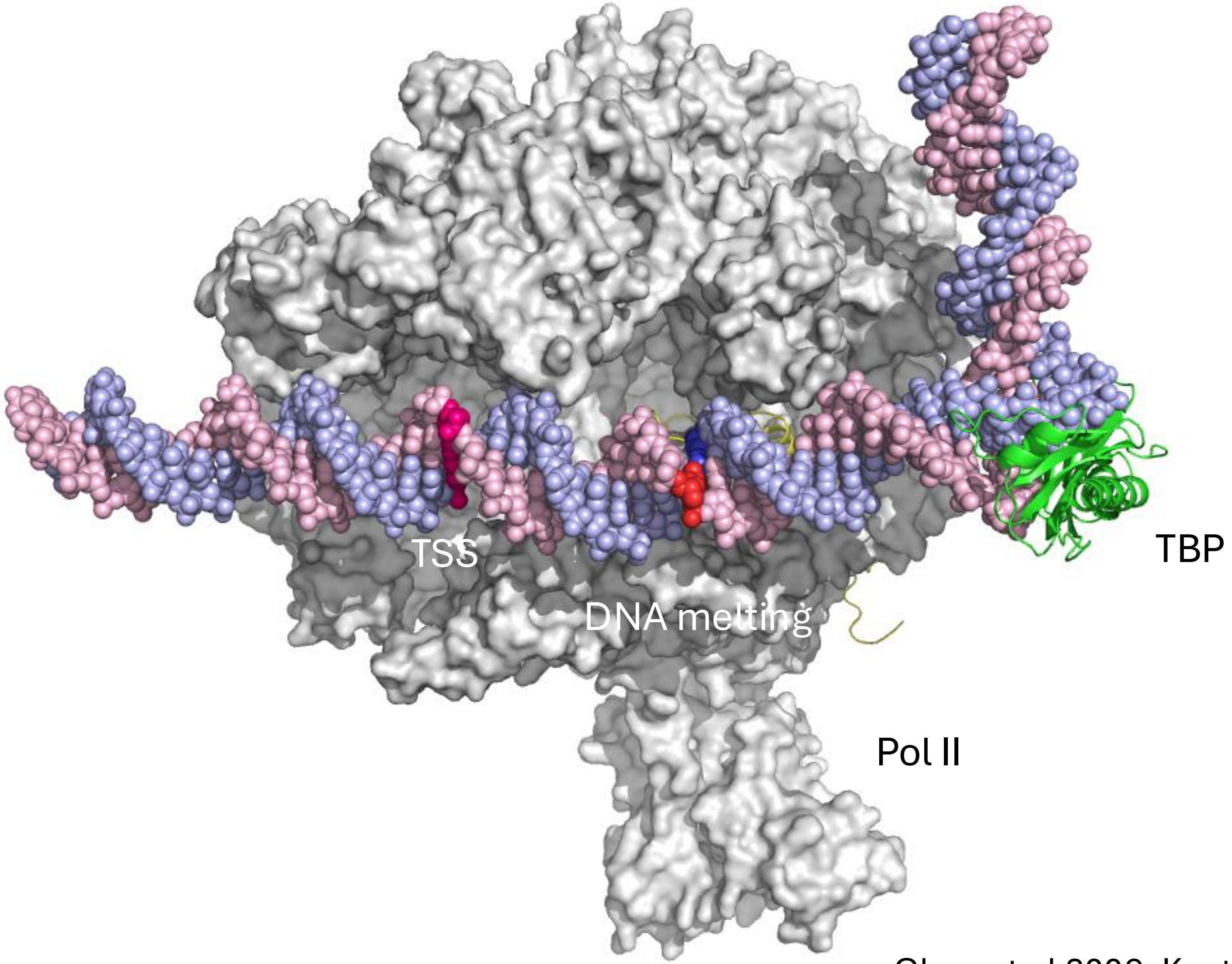
Right handed threading of DNA leads to unwinding



Sebastian Grünberg

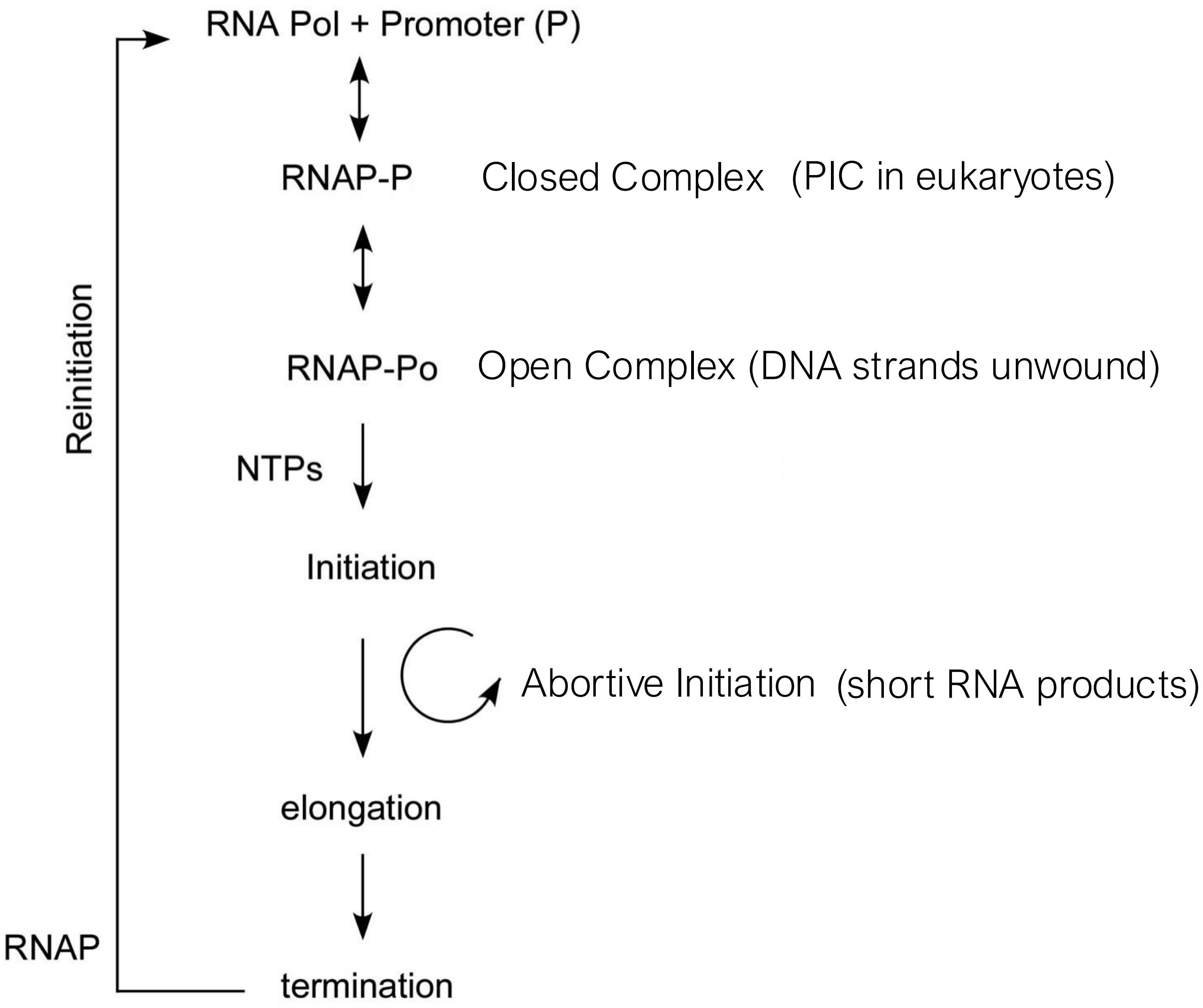
Function of core promoter elements in PIC assembly: TATA-containing promoters

TBP-TFIIB-Pol II-DNA complex



Chen et al 2009; Kostrewa et al 2009

Universal Pathway for Transcription by RNA polymerases



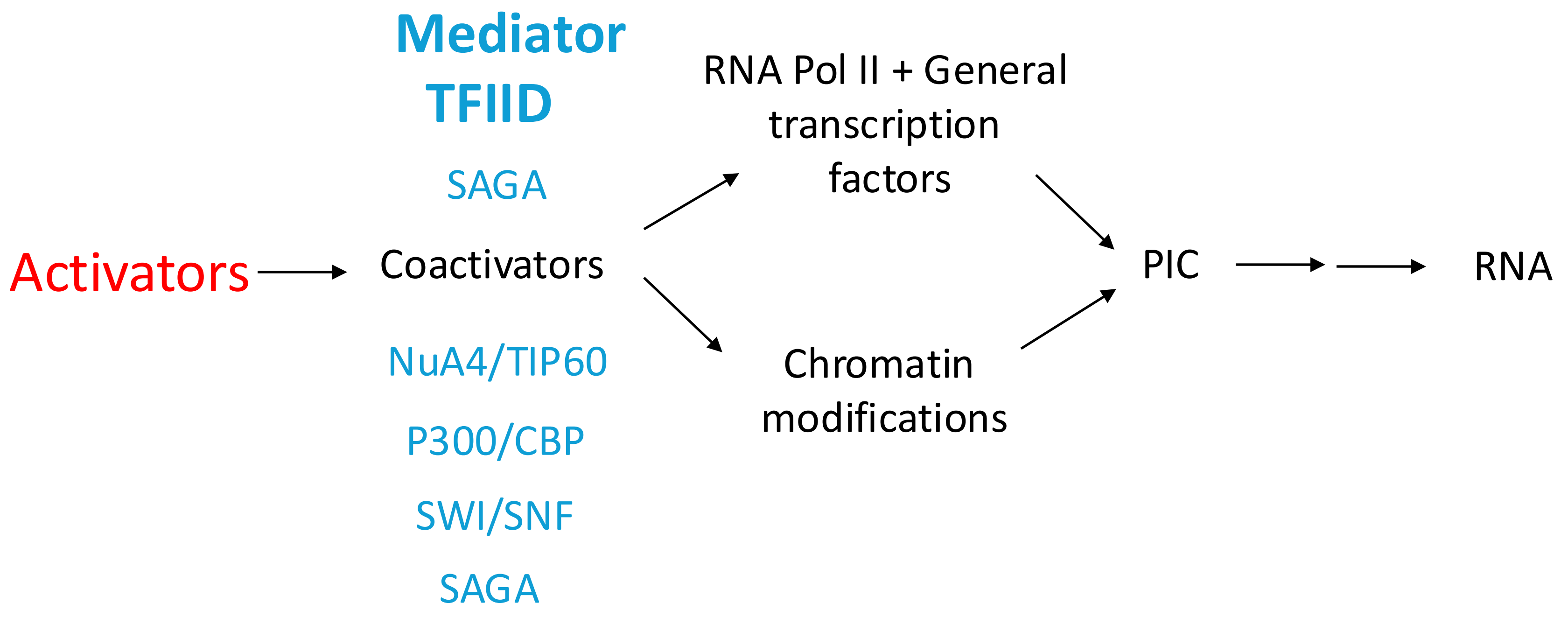
The Pol II CTD – a unique regulatory domain only found on Pol II

The largest Pol II subunit, Rpb1 has a unique 7 amino acid repeated sequence at the C-terminal termed the C-terminal domain (CTD)

(YSPTSPS) 26-52 repeats
1 2 3 4 5 6 7

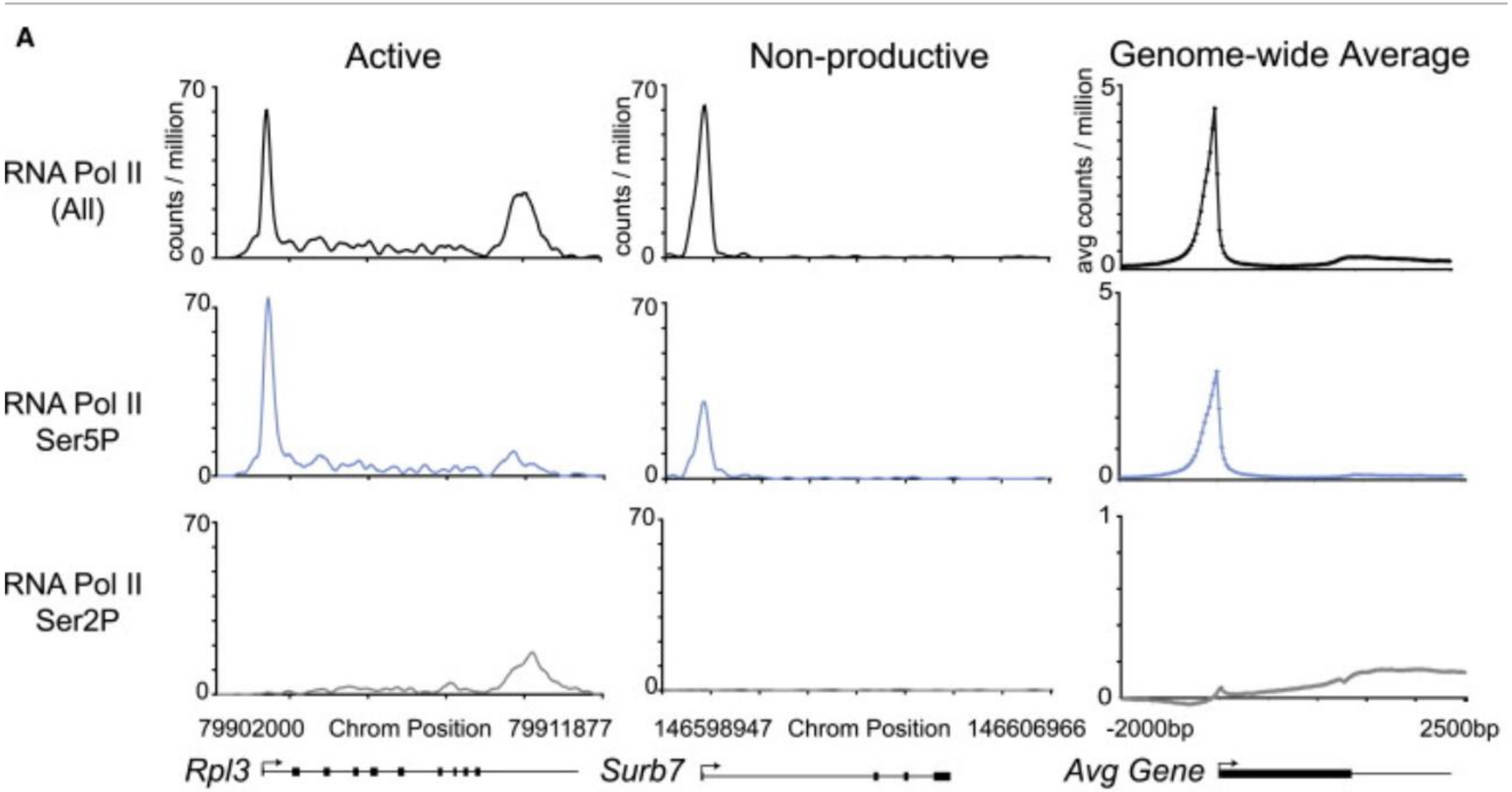
Ser 2, 5, and 7, Thr4 and Tyr1 are phosphorylated by specific kinases and dephosphorylated by specific Phosphatases at different points in the transcription cycle.

A pathway for transcription activation

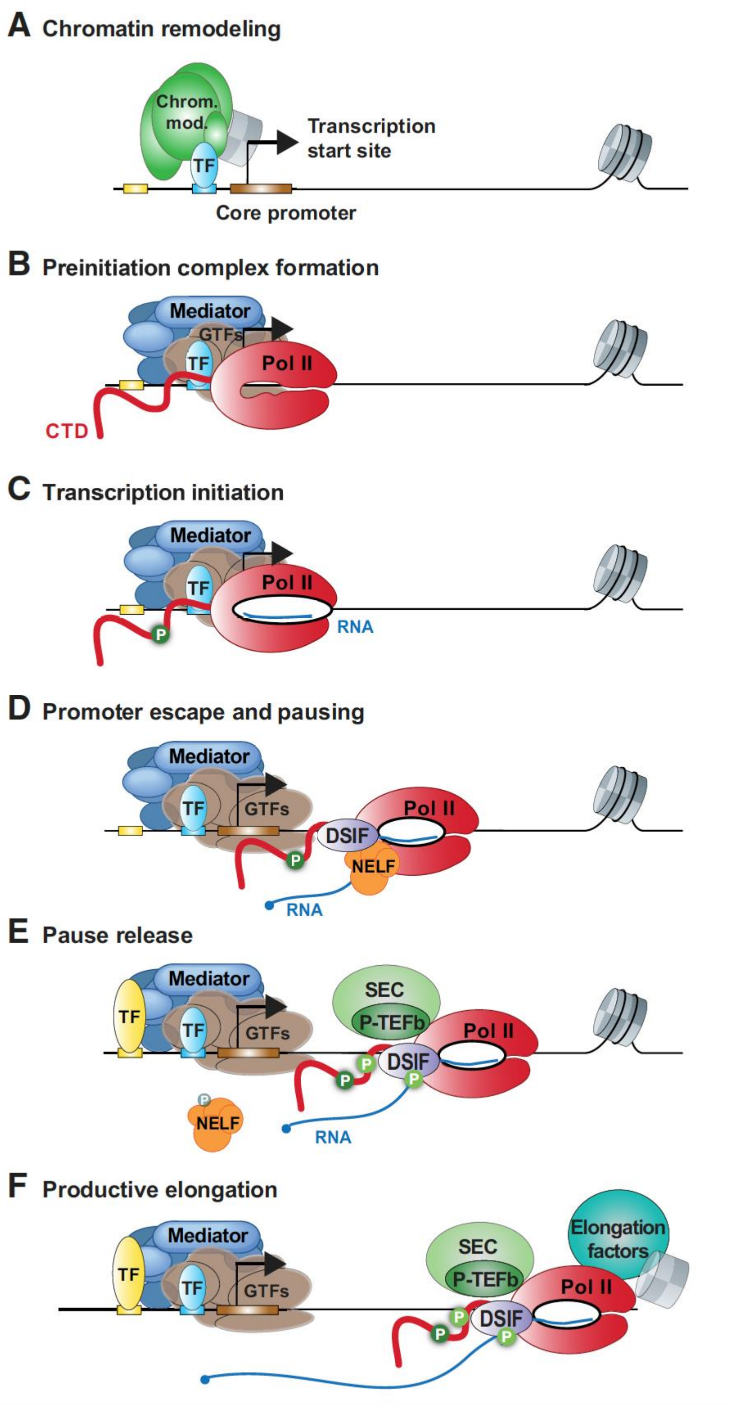


Transcription: Promoter proximal pausing

Pausing is detectable by ChIP of Pol II



Cell, Volume 141, Issue 3, 30 April 2010, Pages 432-445



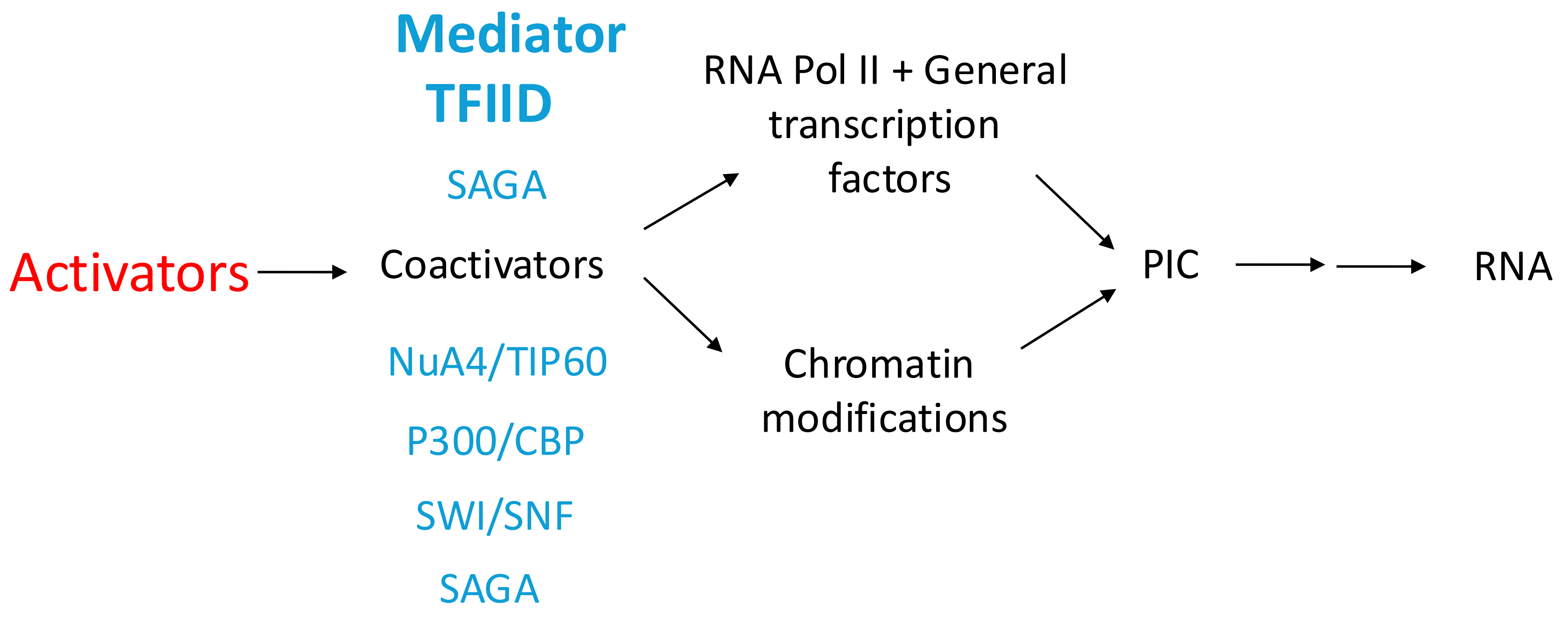
TFIIH associated kinase Cdk7 phosphorylates the CTD at Ser5 & 7

Ser 5P disrupts mediator interactions facilitating promoter escape. Ser5P is recognized by capping enzymes, so modifying 5' end of RNA.

During early elongation SPT4/5 (DSIF) binds to emerging RNA chain. Ser5P and DSIF allow NELF to engage and lead to the polymerase to stop transcribing after ~30-50 nt.

P-TEFb can release pause by phosphorylating CTD and DSIF (Spt5). P-TEFb is highly regulated and is controlled by many factors, including Brd4 which binds acetylated histones.

A pathway for transcription activation



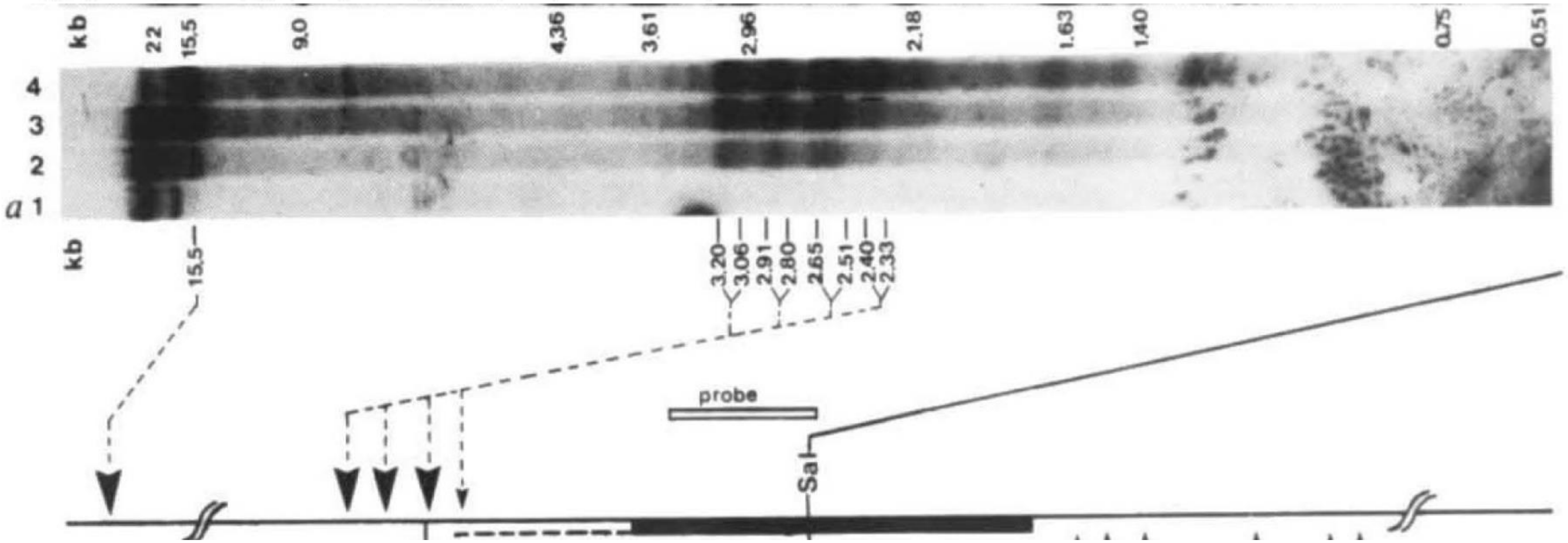
Chromatin remodeling at promoters

A major challenge in understanding chromatin function *in vivo* was the inability to assay nucleosome organization on specific genes in large eukaryotic genomes.

Significant advances were made by studying highly expressed genes, such as Globin and Heatshock genes.

The 5′ ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I

Carl Wu



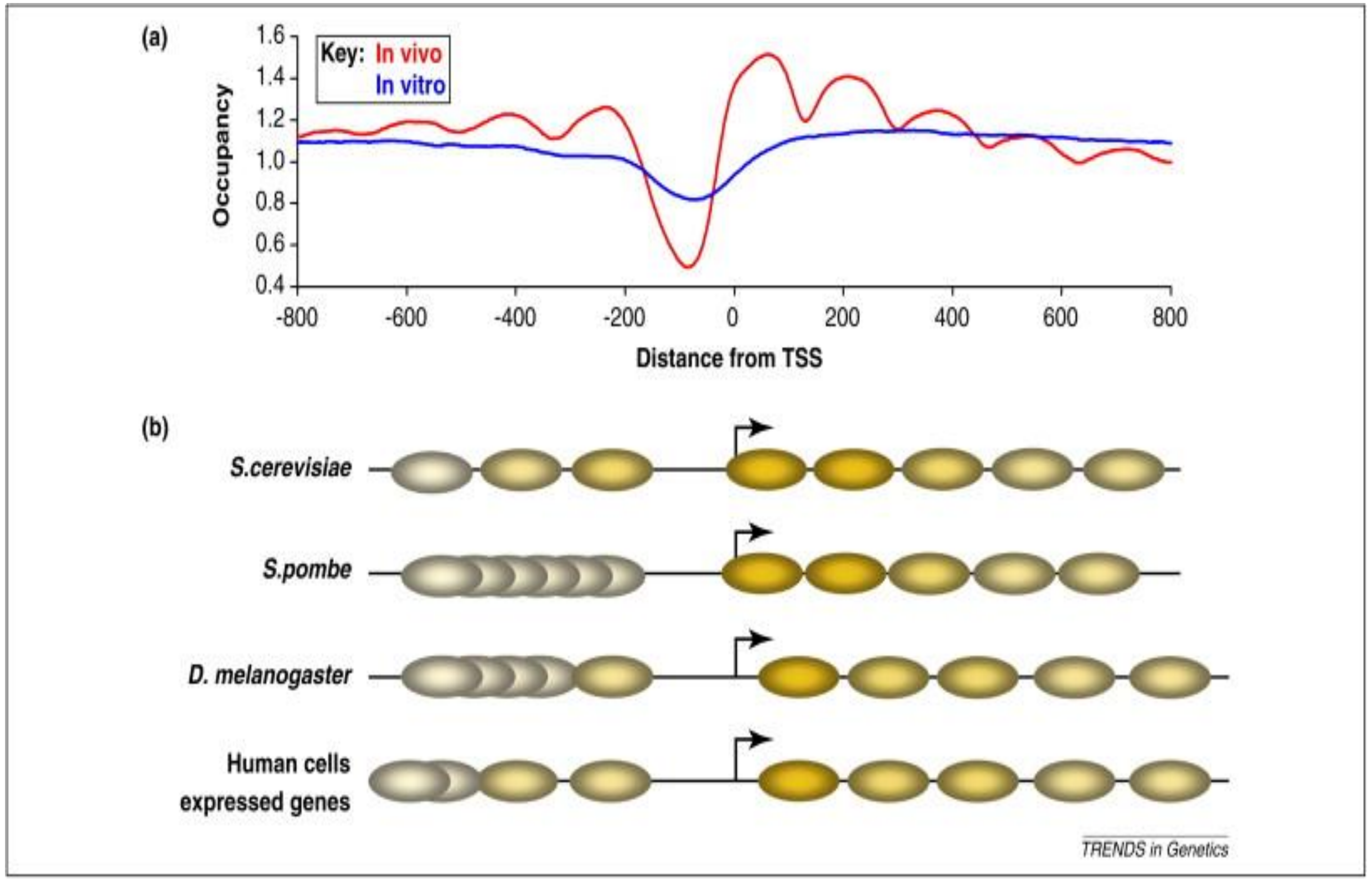
Cell, Vol. 20, 451–460, June 1980, Copyright © 1980 by MIT

Tissue-Specific DNA Cleavages in the Globin Chromatin Domain Introduced by DNAase I

Jurg Stalder*†, Alf Larsen†, James D. Engel‡,
Maureen Dolan‡, Mark Groudine† and
Harold Weintraub†

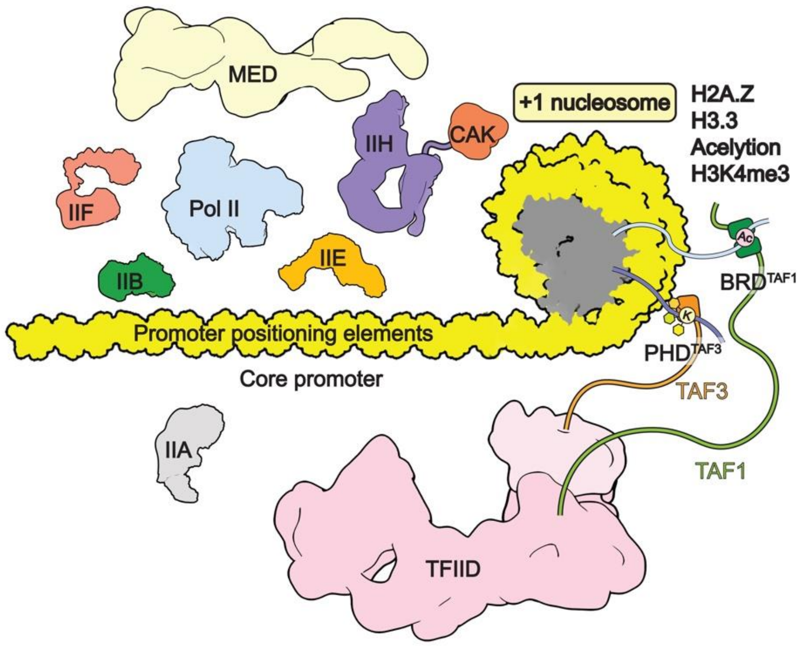
*The most significant observation of our work is that many of the specific cleavages introduced by DNAseI **are tissue specific.** This is most dramatically seen in the case of the B-globin genes where the DNAase sites differ even between adult and embryonic red cells.*

Nucleosomes are organized at the 5′ end of genes

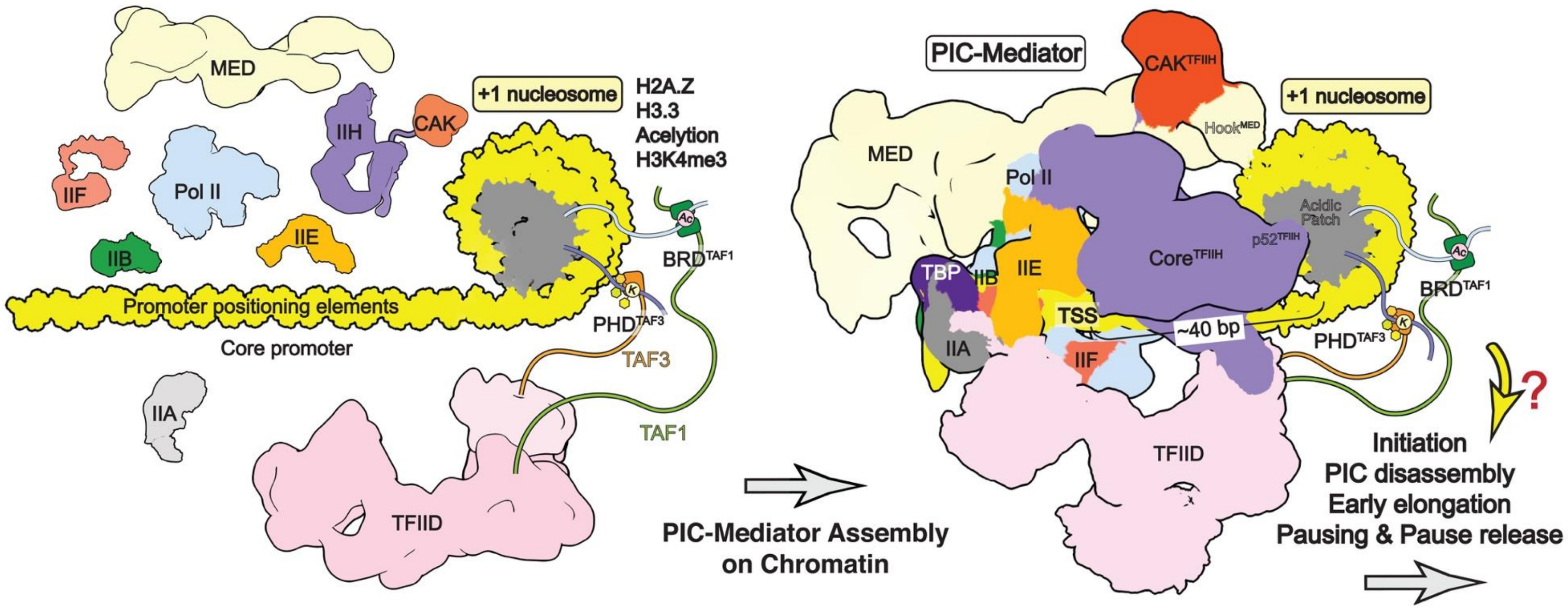


{Bai and Morozov, 2010, Trends Genet, 26, 476-83}

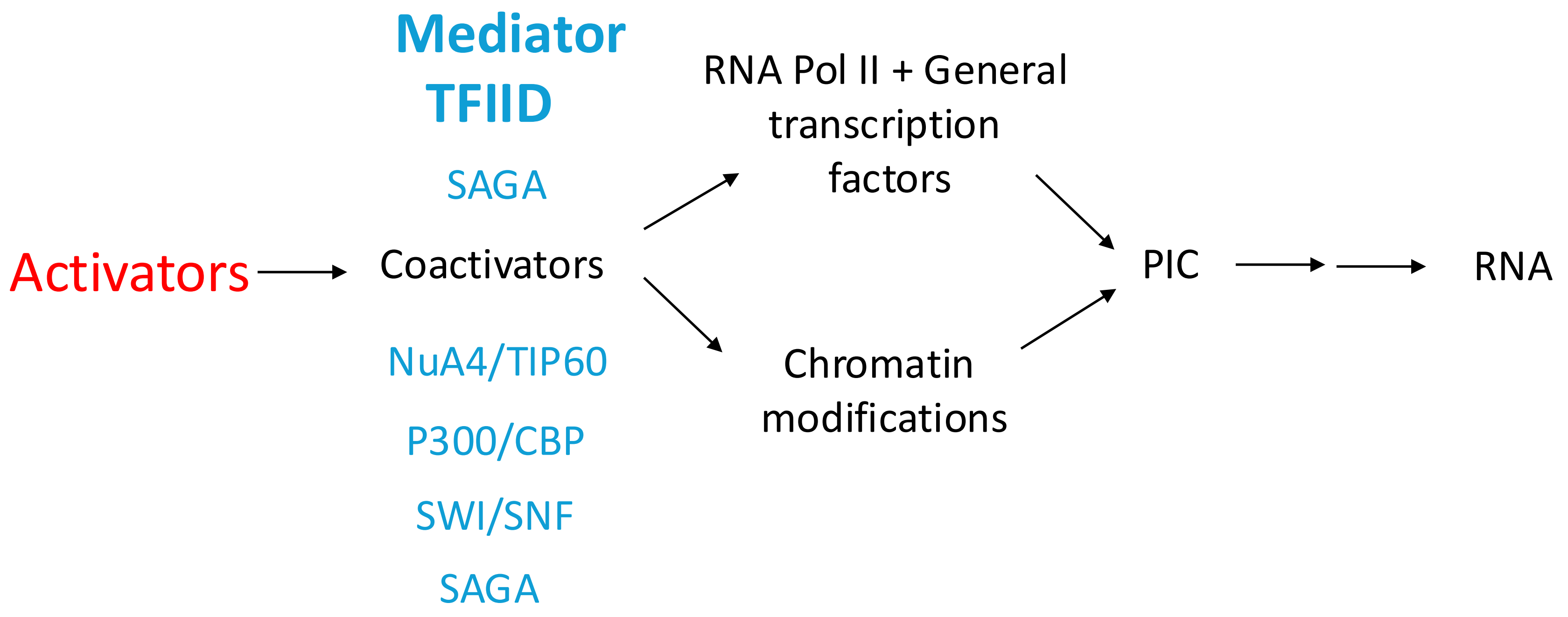
The structure of Mediator, TFIID and RNA polII bound to a template with a +1 nucleosome



The structure of Mediator, TFIID and RNA polII bound to a template with a +1 nucleosome

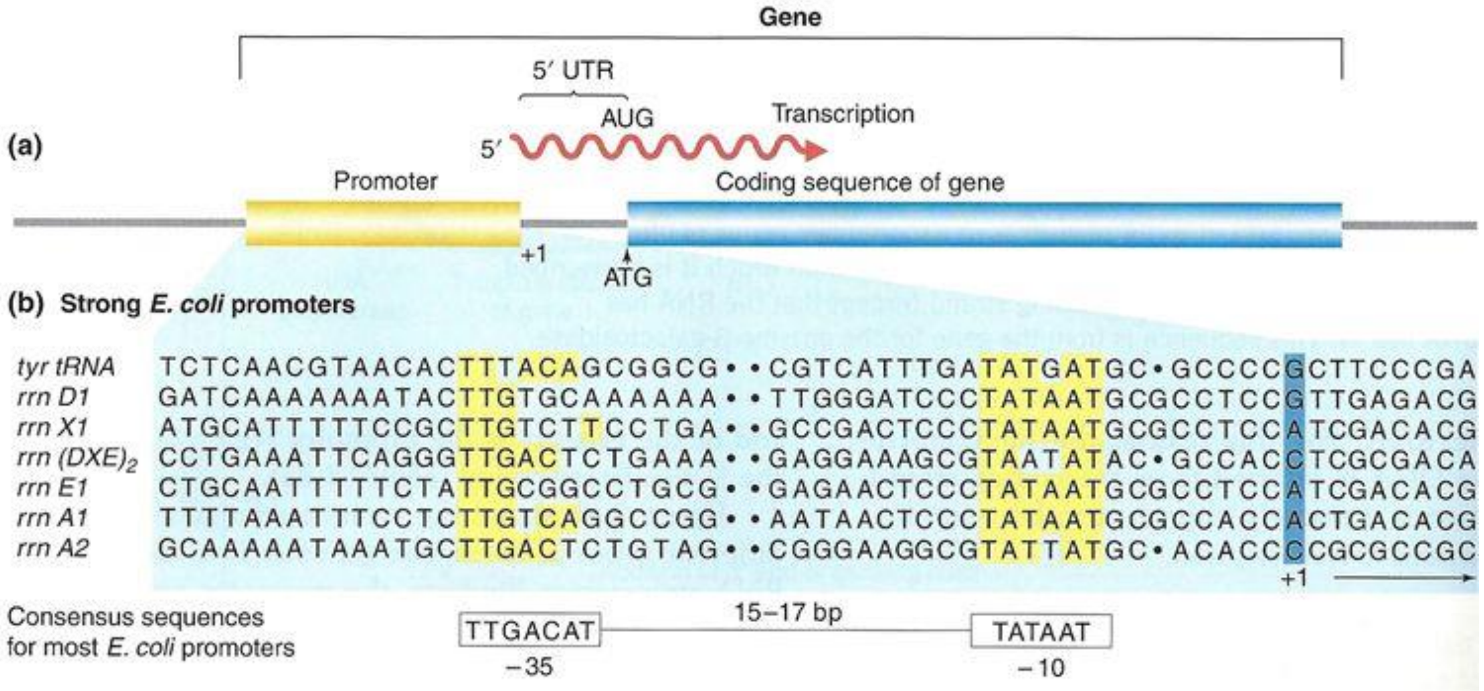


A pathway for transcription activation

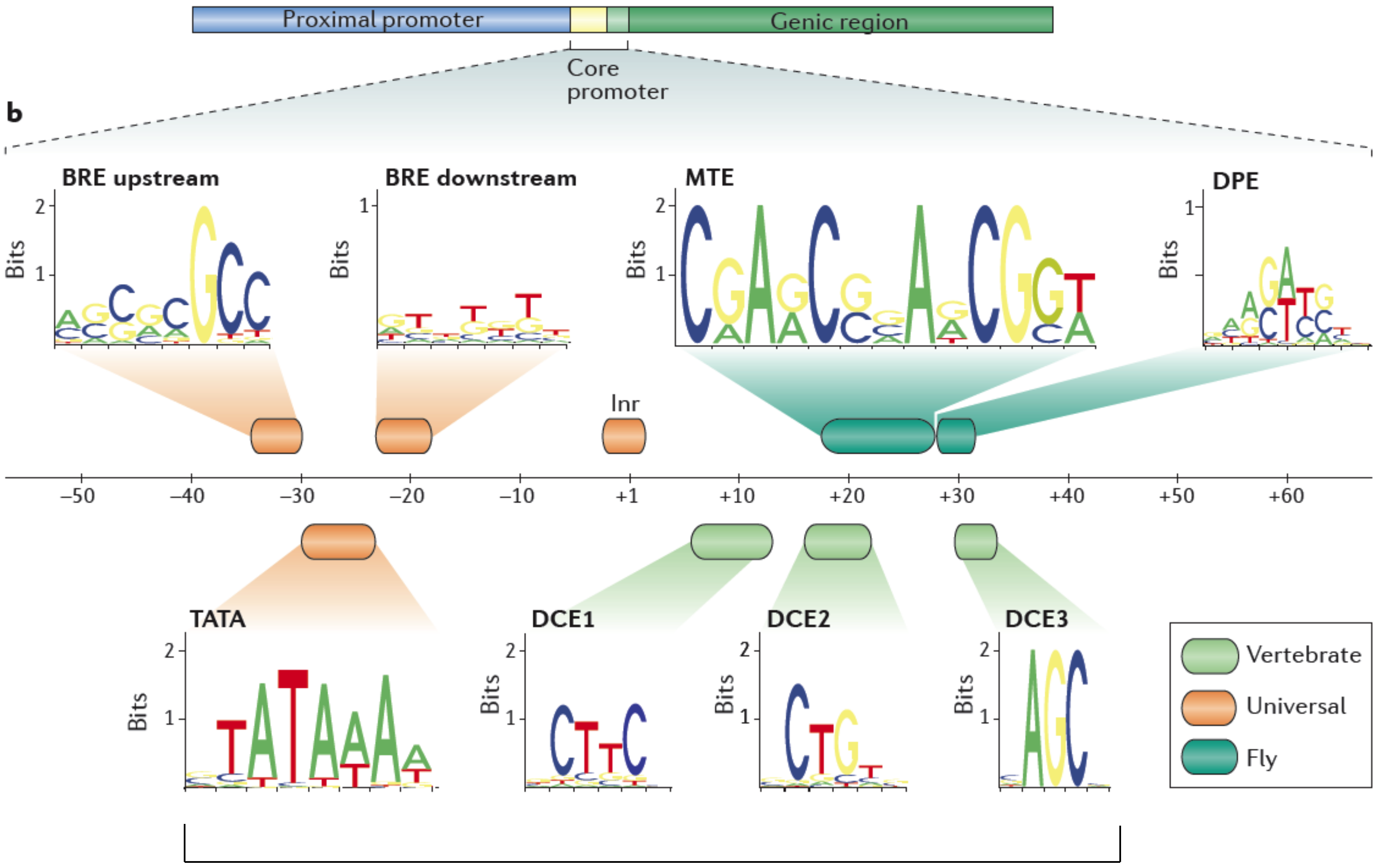


Promoter = DNA sequence coding for transcription initiation

E. Coli consensus promoter



Promoter motifs that serve as protein recognition sites in PIC formation

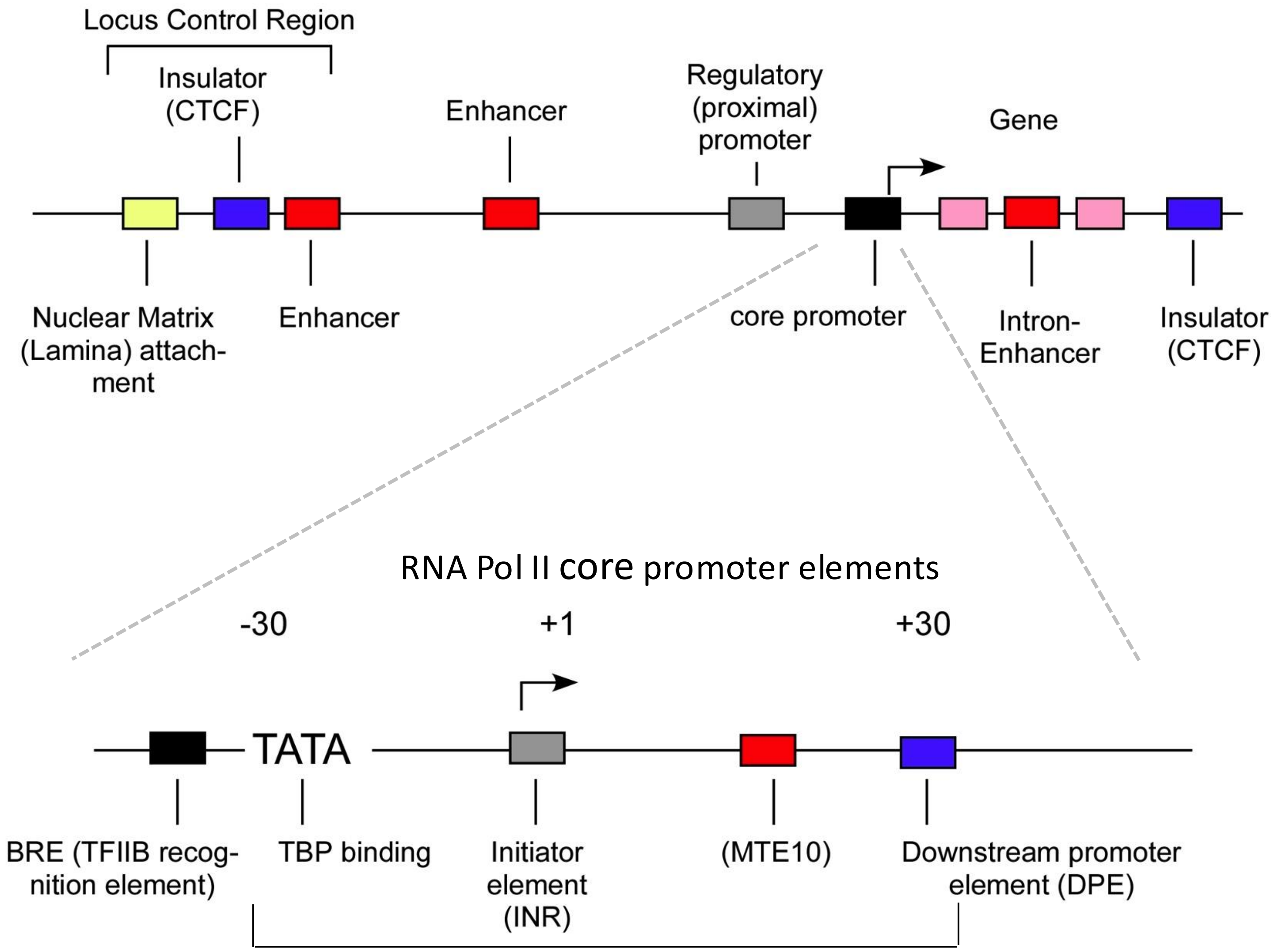


Binding sites for subunits of the TFIID coactivator complex

INR: varies from TCA(G/T)TC(CT) to pyrimidine-purine (C/T)-(A/G)

Lenhard et al 2012

Classical description of an RNA Polymerase II promoter (pre genome-wide studies - oversimplified)



Binding sites for subunits of the TFIID coactivator complex

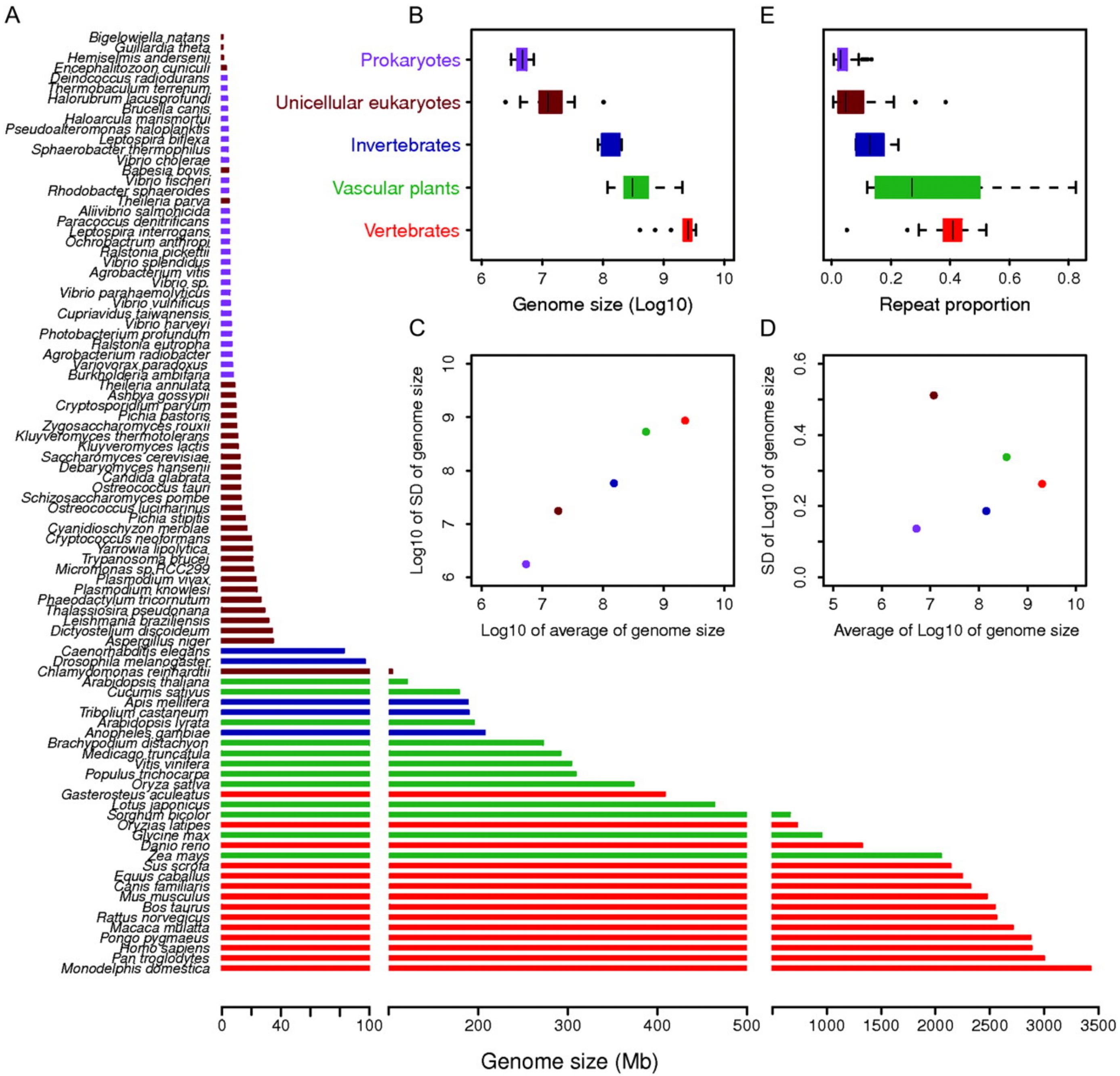
Very few if any natural promoters contain all these motifs

Understanding gene-specific transcription factor function

- How is DNA binding specificity determined?

- How is binding of a transcription factor related to gene regulation?

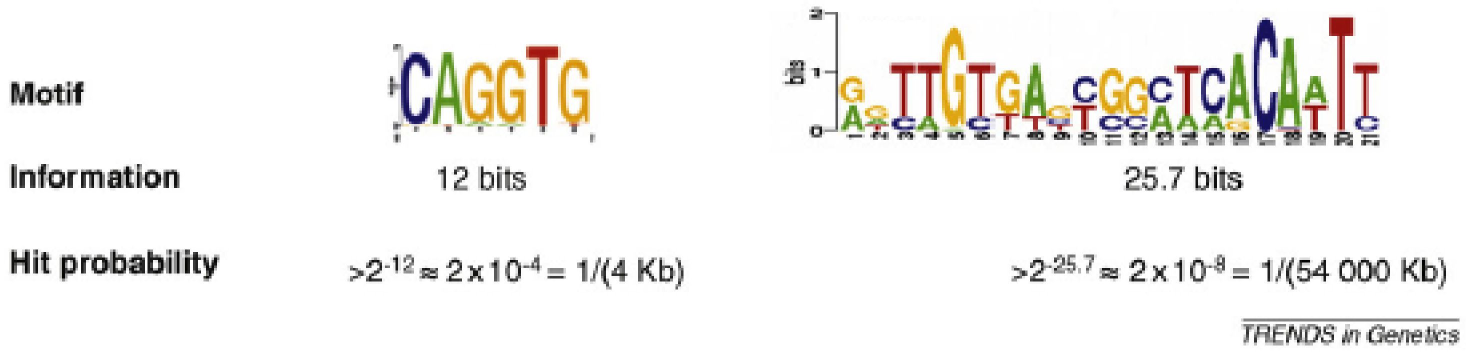
- What are the mechanisms that transcription factors use to stimulate transcription?



A
B
E
C
D
Prokaryotes
Unicellular eukaryotes
Invertebrates
Vascular plants
Vertebrates
Genome size (Log10)
Repeat proportion
Log10 of SD of genome size
Log10 of average of genome size
SD of Log10 of genome size
Average of Log10 of genome size
Genome size (Mb)
Bigelowiella natans
Guillardia theta
Hemiselmis andersenii
Encephalitozoon cuniculi
Deinococcus radiodurans
Thermobaculum terrenum
Halorubrum lacusprofundi
Brucella canis
Haloarcula marismortui
Pseudoalteromonas haloplanktis
Leptospira biflexa
Sphaerobacter thermophilus
Vibrio cholerae
Babesia bovis
Vibrio fischeri
Rhodobacter sphaeroides
Theileria parva
Aliivibrio salmonicida
Paracoccus denitrificans
Leptospira interrogans
Ochrobactrum anthropi
Ralstonia pickettii
Vibrio splendidus
Agrobacterium vitis
Vibrio sp.
Vibrio parahaemolyticus
Vibrio vulnificus
Cupriavidus taiwanensis
Vibrio harveyi
Photobacterium profundum
Ralstonia eutropha
Agrobacterium radiobacter
Variovorax paradoxus
Burkholderia ambifaria
Theileria annulata
Ashbya gossypii
Cryptosporidium parvum
Pichia pastoris
Zygosaccharomyces rouxii
Kluyveromyces thermotolerans
Kluyveromyces lactis
Saccharomyces cerevisiae
Debaryomyces hansenii
Candida glabrata
Ostreococcus tauri
Schizosaccharomyces pombe
Ostreococcus lucimarinus
Pichia stipitis
Cyanidioschyzon merolae
Cryptococcus neoformans
Yarrowia lipolytica
Trypanosoma brucei
Micromonas sp.RCC299
Plasmodium vivax
Plasmodium knowlesi
Phaeodactylum tricornutum
Thalassiosira pseudonana
Leishmania braziliensis
Dictyostelium discoideum
Aspergillus niger
Caenorhabditis elegans
Drosophila melanogaster
Chlamydomonas reinhardtii
Arabidopsis thaliana
Cucumis sativus
Apis mellifera
Tribolium castaneum
Arabidopsis lyrata
Anopheles gambiae
Brachypodium distachyon
Medicago truncatula
Vitis vinifera
Populus trichocarpa
Oryza sativa
Gasterosteus aculeatus
Lotus japonicus
Sorghum bicolor
Oryzias latipes
Glycine max
Danio rerio
Zea mays
Sus scrofa
Equus caballus
Canis familiaris
Mus musculus
Bos taurus
Rattus norvegicus
Macaca mulatta
Pongo pygmaeus
Homo sapiens
Pan troglodytes
Monodelphis domestica

Motif length	Possible combinations
1	4
2	16
3	64
4	256
5	1,024
6	4,096
7	16,384
8	65,536
9	262,144
10	1,048,576
11	4,194,304
12	16,777,216
13	67,108,864
14	268,435,456
15	1,073,741,824
16	4,294,967,296
17	17,179,869,184
18	68,719,476,736
19	274,877,906,944
20	1,099,511,627,776
21	4,398,046,511,104
22	17,592,186,044,416
23	70,368,744,177,664
24	281,474,976,710,656
25	1,125,899,906,842,620
26	4,503,599,627,370,500
27	18,014,398,509,482,000
28	72,057,594,037,927,900
29	288,230,376,151,712,000
30	1,152,921,504,606,850,000

Are the DNA-binding strategies different for bacterial and eukaryotic transcription factors?



The sequence logos for low- and high-information motifs and the probability of a spurious hit to the motif in a random genetic background

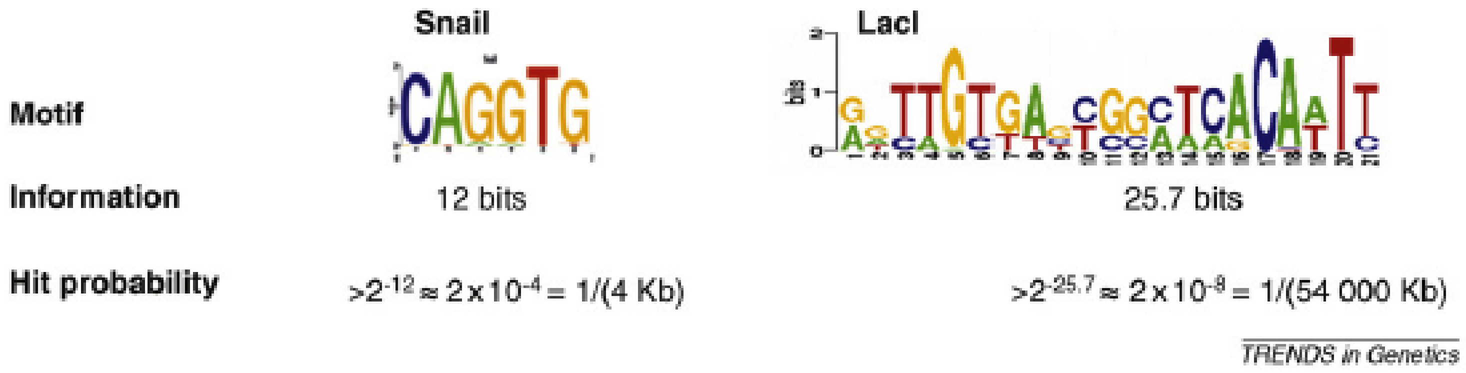
Information content for a binding motif normalized for base composition

$$I = \sum_{i=1}^{L} \sum_{b \in \{A,C,G,T\}} p_i(b) \log_2(p_i(b)/q(b))$$

L is the length of the motif, pi(b) is the frequency of base b at position I and q(b) is background frequency

Wunderlich and Mirny (2009)

Are the DNA-binding strategies different for bacterial and eukaryotic transcription factors?



The sequence logos for low- and high-information motifs and the probability of a spurious hit to the motif in a random genetic background

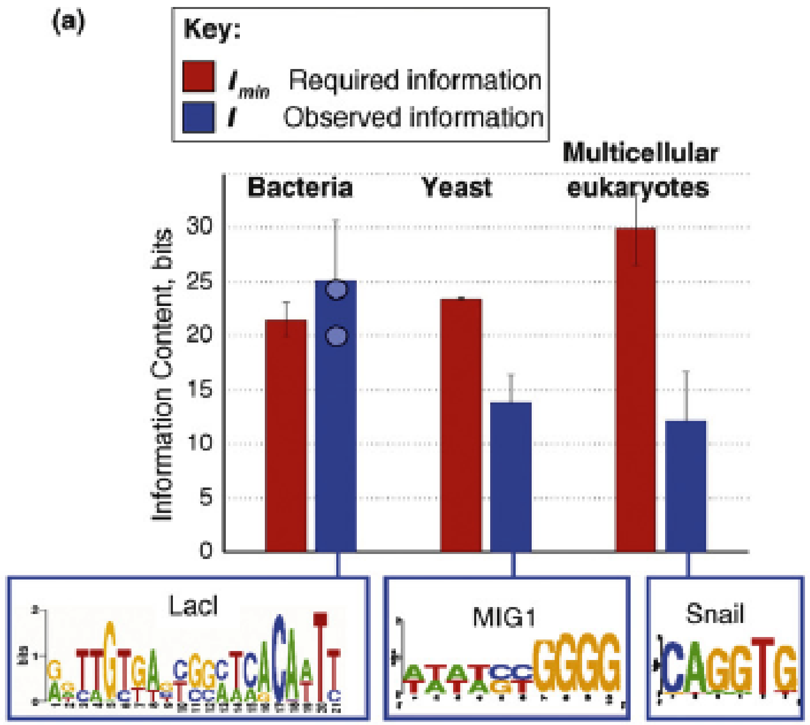
Information content for a binding motif normalized for base composition

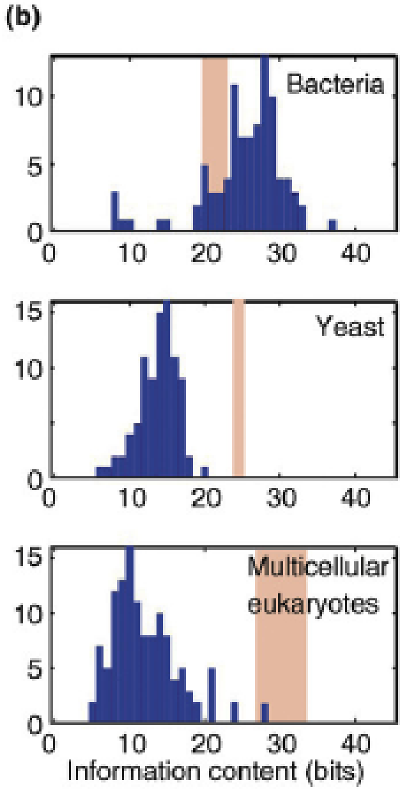
$$I = \sum_{i=1}^{L} \sum_{b \in \{A,C,G,T\}} p_i(b)\log_2(p_i(b)/q(b))$$

L is the length of the motif, pi(b) is the frequency of base b at position I and q(b) is background frequency

Wunderlich and Mirny (2009)

Properties of DNA binding motifs for bacteria, yeast and multi cellular eukaryotes





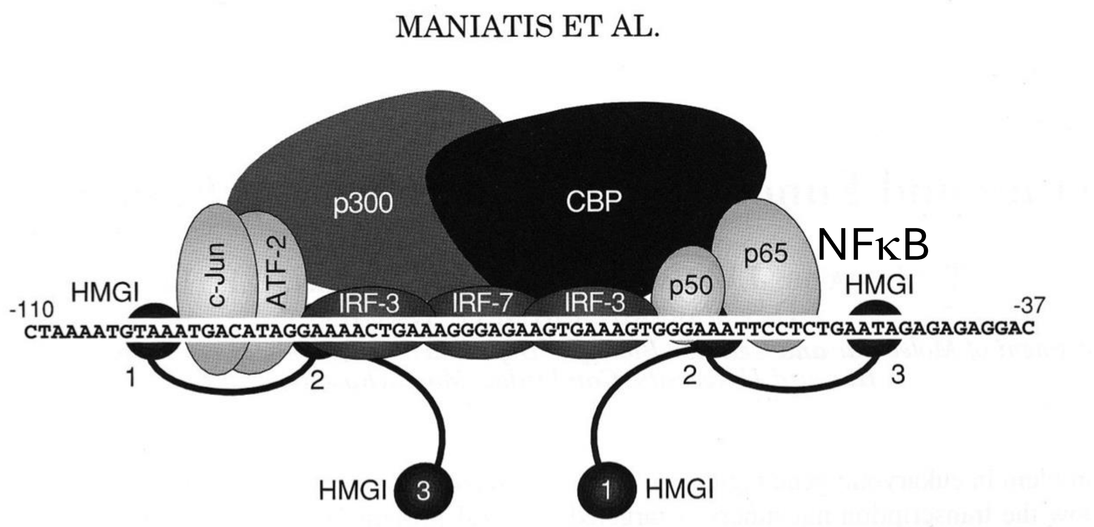
(c)

Genome size (N, bps)	$1–9 \times 10^6$	1.2×10^7	$10^8–10^{10}$
Expected number of spurious hits (h)	<1	~100*	$\sim 10^3–10^5$**
Spacing between hits (s, bps)	$>10^7$	10 000	4000
Minimal number of sites per cluster of 1000 bp ($n_{cluster}$)	1	5	7–9 (1 TF), 12–20 (3 TFs)

*Assuming 80% chromatization
**Assuming 90% chromatization

Wunderlich and Mirny (2009)

Interferon-β enhanceosome – an example of how clustered binding sites generates specificity



1998 View of the IFN-β enhanceosome (Maniatis, CSHL Symp)

IFN-β is activated by virus infection. Interferon binds cell surface receptors where it activates signaling pathways to inhibit virus replication.

NFkB (p50-p65) is in an inactive form in the cytoplasm (complexes with NFκb-IkB). Cytokines, viruses, phorbol esters, UV, etc activate signaling pathways that degrade IκB leading to the active form of NFκB and transport into the nucleus.

Upon Virus infection, only a small subset of NFκB–regulated genes are induced. This is because NFκB must assemble with other factors to form active enhancer complexes.

IFR 3 and 7 are homo or heterodimers in an inactive form before virus infection. Phosphorylation of IRFs after viral infection leads to a form active in DNA binding and association with p300/CBP

Alignment of the INF-β enhancer

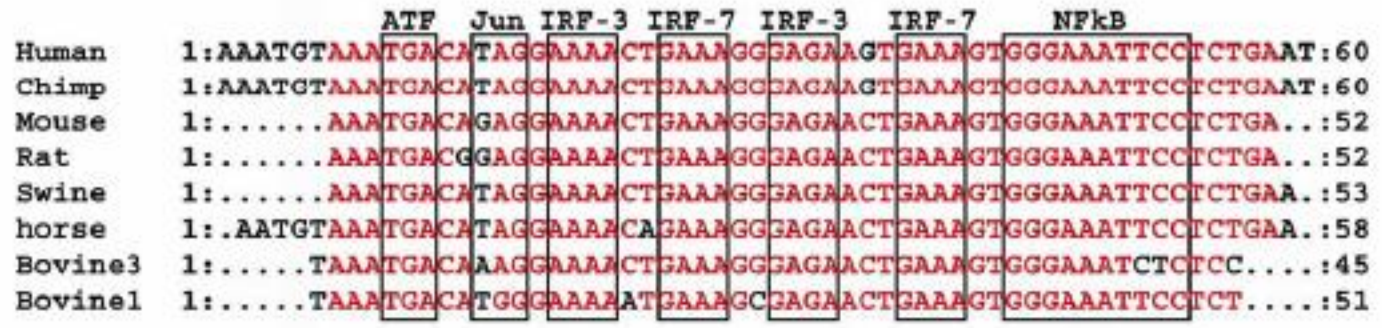
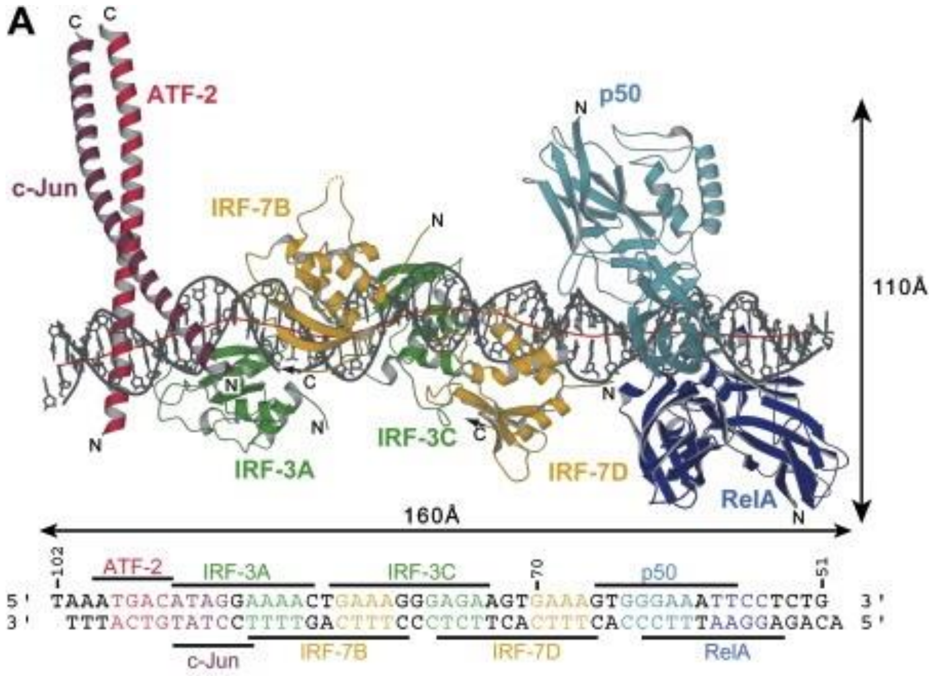


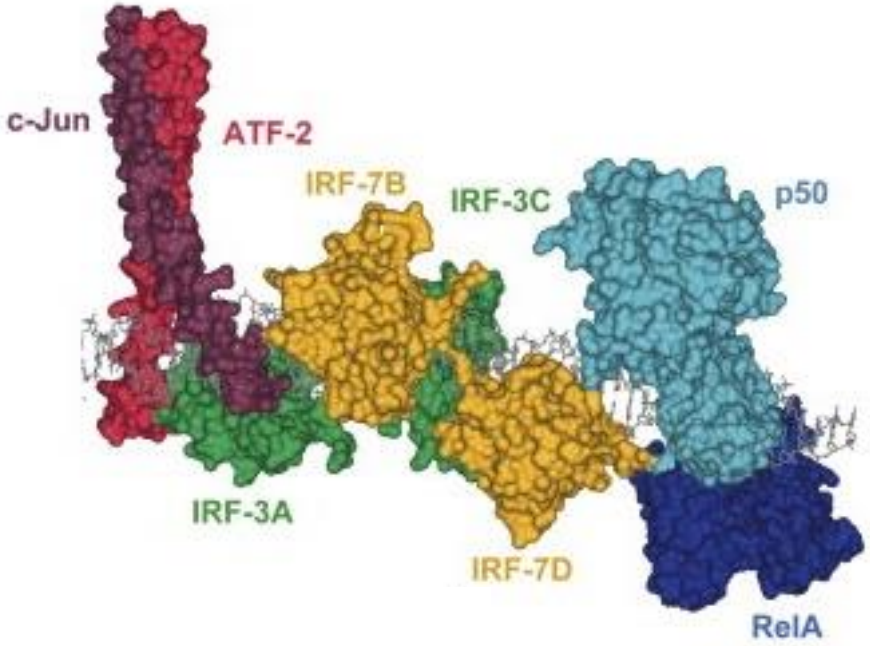
Figure S6: Sequence alignment of the INF-β enhancer from different species. Conserved residues are shown in red. The core binding sites for each protein are boxed.

Nearly every base pair in the enhancer is conserved

Panne et al (Harrison & Maniatis) 2007

Overall Structure of the IFN-β enhanceosome

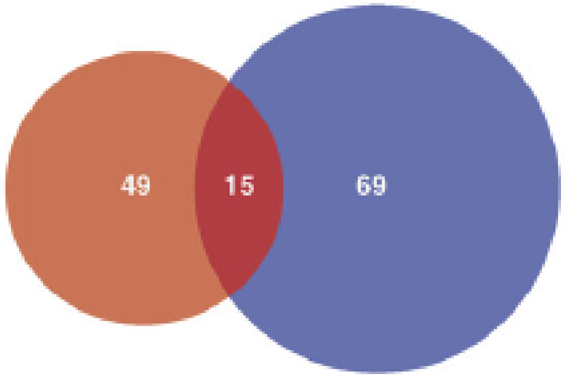




Overlapping binding sites generate an all or none binding mechanism

Panne et al (Harrison & Maniatis) 2007

Very few DNA-binding domain families are shared between bacteria and eukaryotes



Blue = eukaryotic
Brown = bacterial

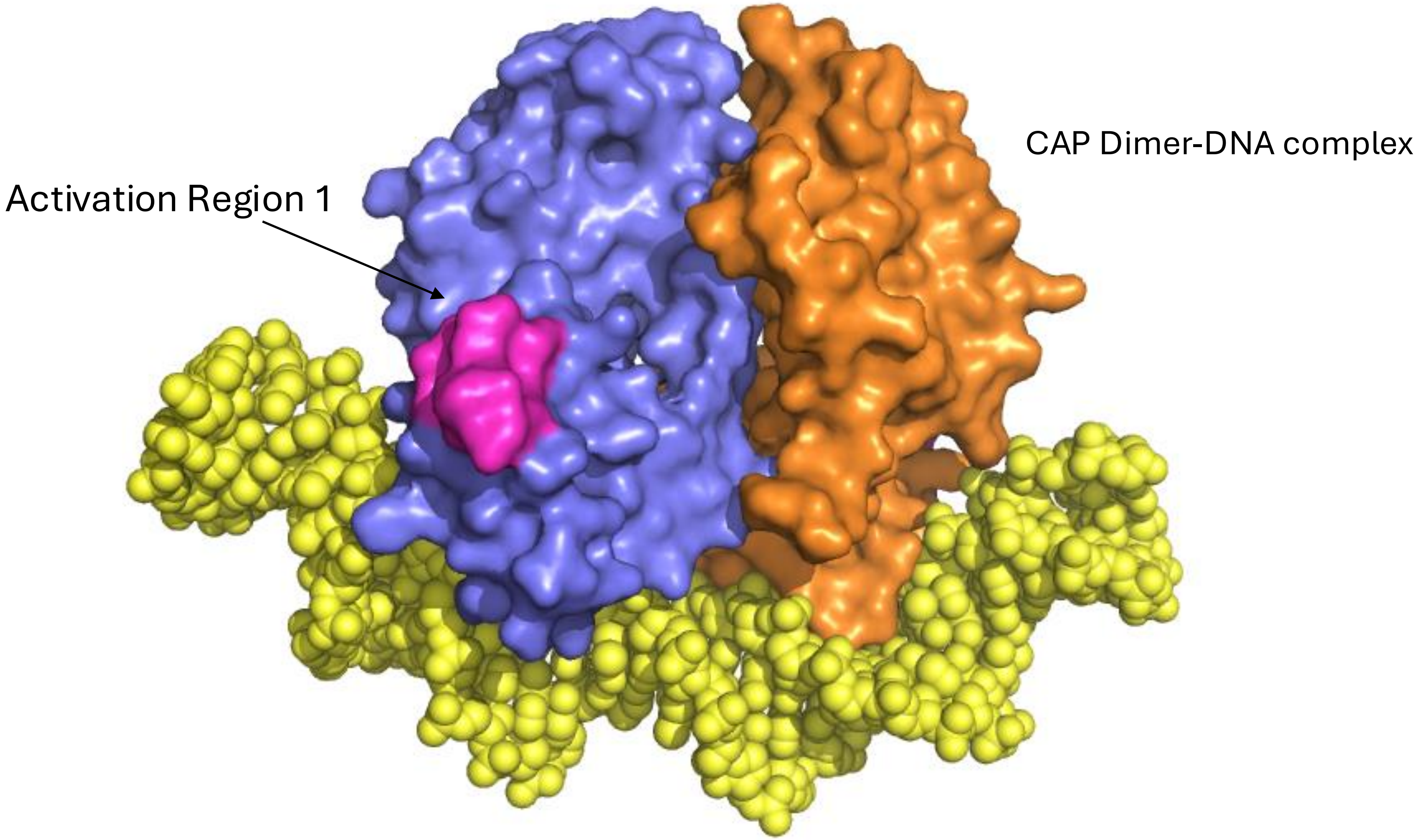
The lack of DNA-binding information content in eukaryotes fits with much higher levels of transcription factors compared to bacteria

Table S4. The number of TF copies per cell

Transcription Factor	*Organism*	*TF copies per cell*	*Source*
LacI tetramer	*E. coli*	10	[9]
LacI dimers	*E. coli*	3	[10]
Yeast transcription factors	*S. cerevisiae*	~2000	[7, 8]
p53	*H. sapiens*	10^5	www.bioNumbers.org and [11]

Wunderlich and Mirny (2009)

Eukaryotic transcription factors typically do not have easily identifiable "activating regions"

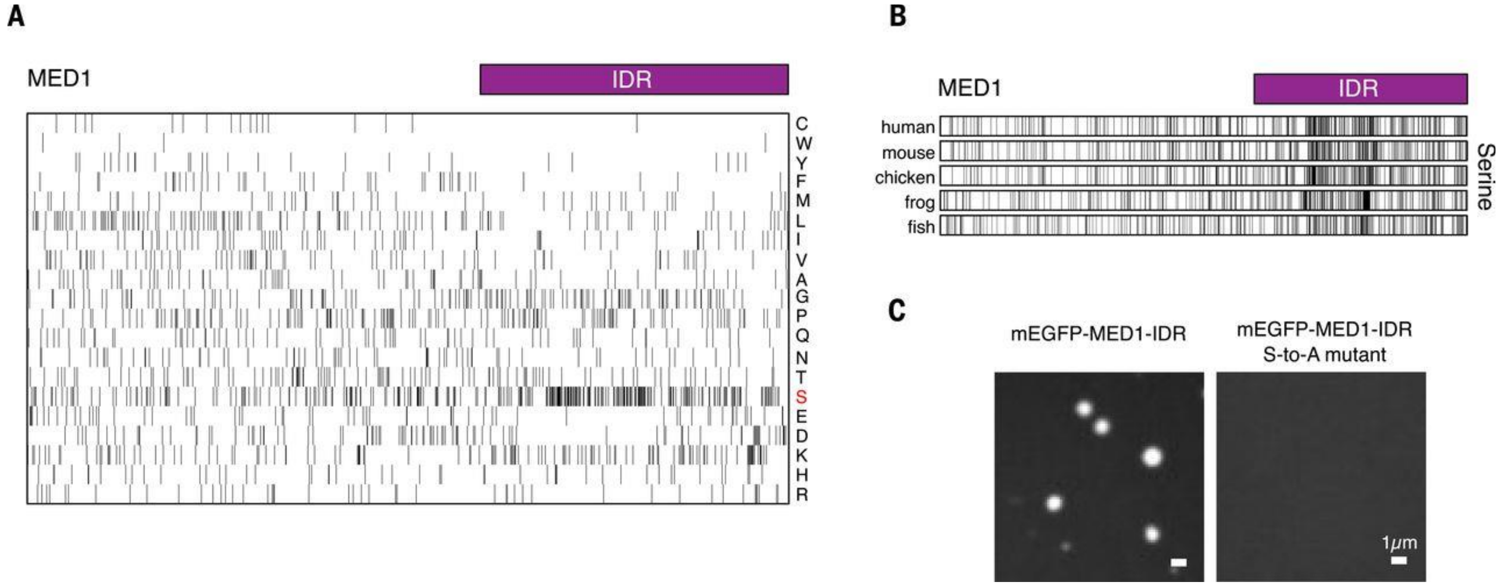


Ebright et al

Eukaryotic transcription factors often contain simple, repetitive sequences of amino acids that are predicted to be “intrinsically disordered”

IDR’s can act as activator domains…

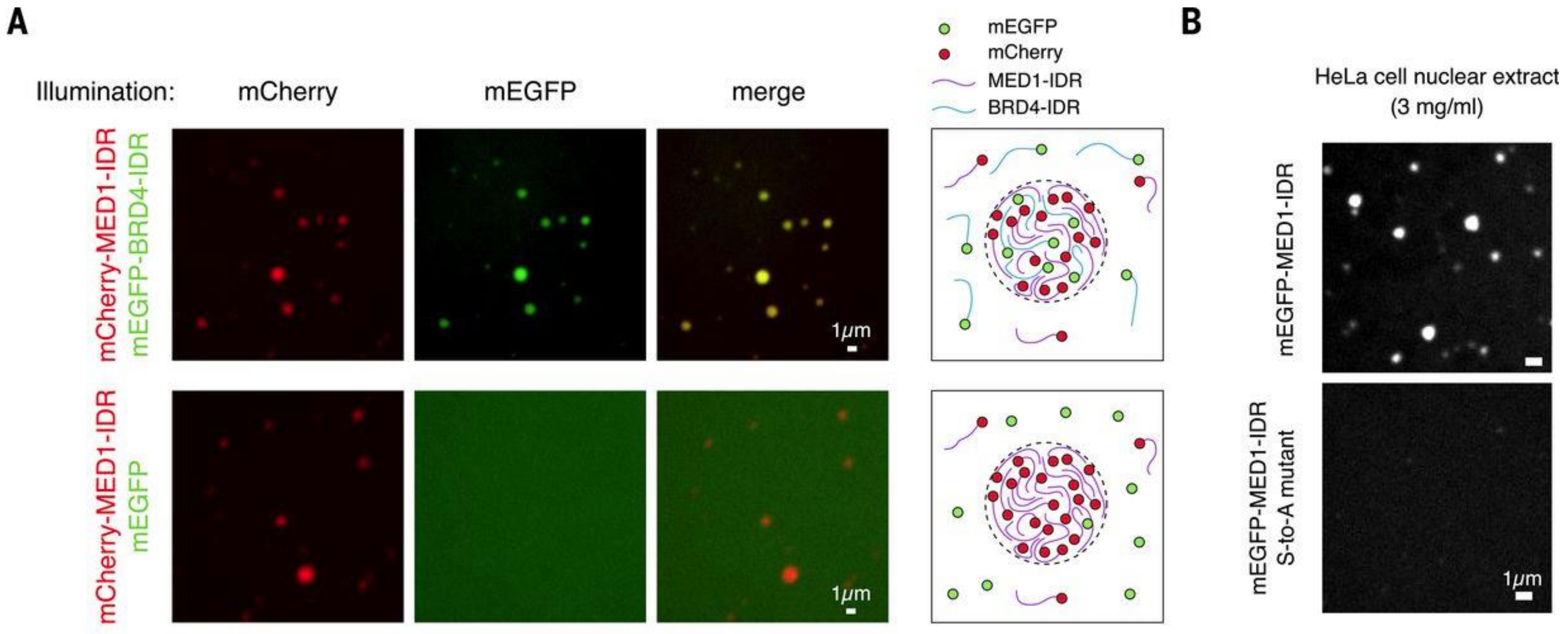
Phase separation



Coactivator condensation at super-enhancers links phase separation and gene control

Science 27 Jul 2018:

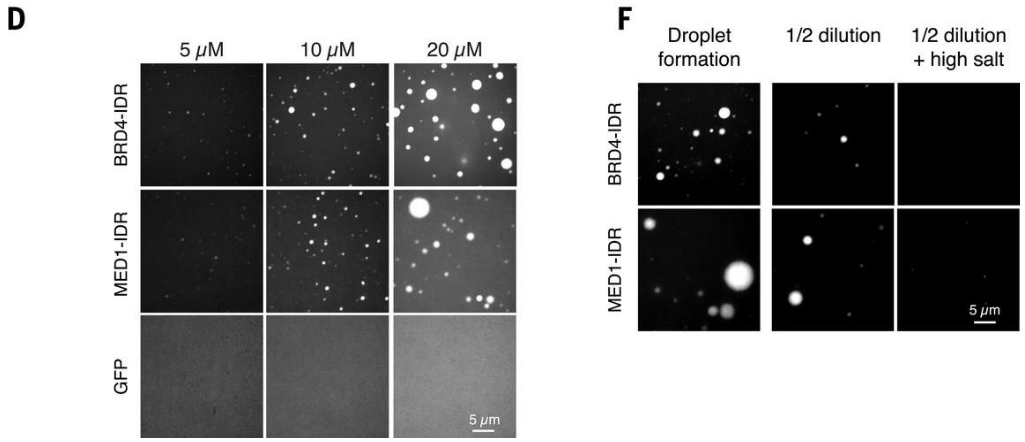
Phase separation



Coactivator condensation at super-enhancers links phase separation and gene control *Science* 27 Jul 2018:

Phase separation

Phase separation observed in vitro typically requires high concentration of proteins



Coactivator condensation at super-enhancers links phase separation and gene control *Science* 27 Jul 2018:

Phase separation

Phase-separated “droplets” may not exist in cells as we see them in vitro

Transcription associated proteins frequently contain IDRs

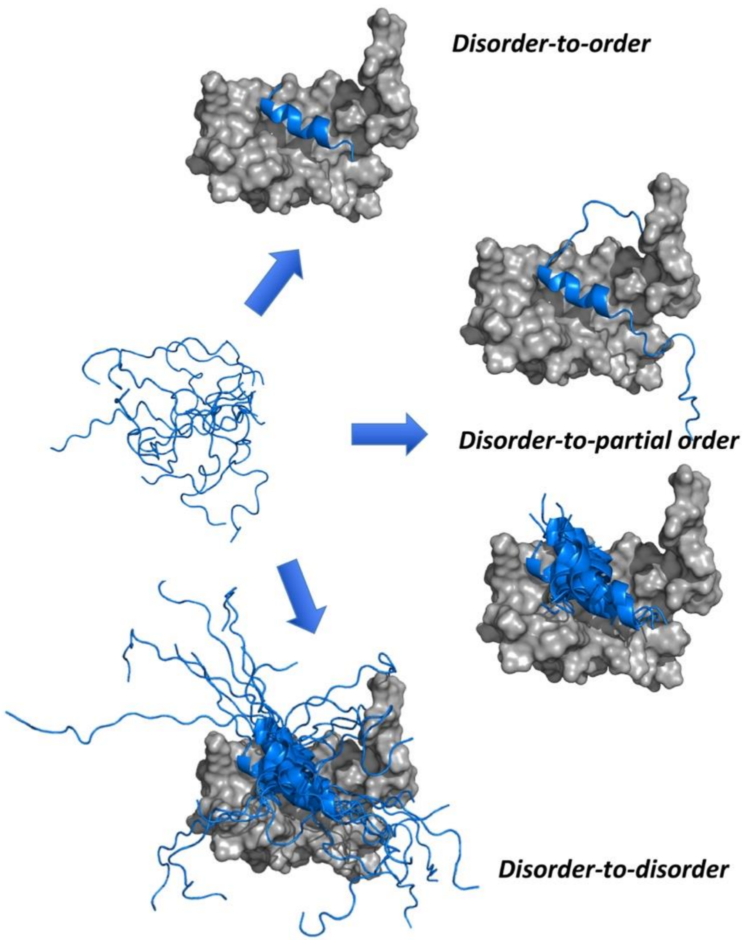
IDRs can act as “activation” domains on transcription factors.

IDRs appear able to participate in “fuzzy” multi-valent interactions.

IDR function

IDRs can adopt many different conformations.

This allows for a broad spectrum of binding partners.



Pioneer Factors

Genome-wide mapping of transcription factors showed that only a few % of consensus binding sites are occupied – why?

DNA is tightly packaged in chromatin where much of the DNA surface faces the nucleosome. It's estimated that ~80% of chromosomal DNA is inaccessible to most transcription factors.

In higher eukaryotes – many gene regulatory regions do not intrinsically disfavor nucleosome binding – so how do factors bind?

1) Cooperative binding by transcription factors

1) Some genes are activated by sequential binding of factors – initial binding cant be explained by cooperativity