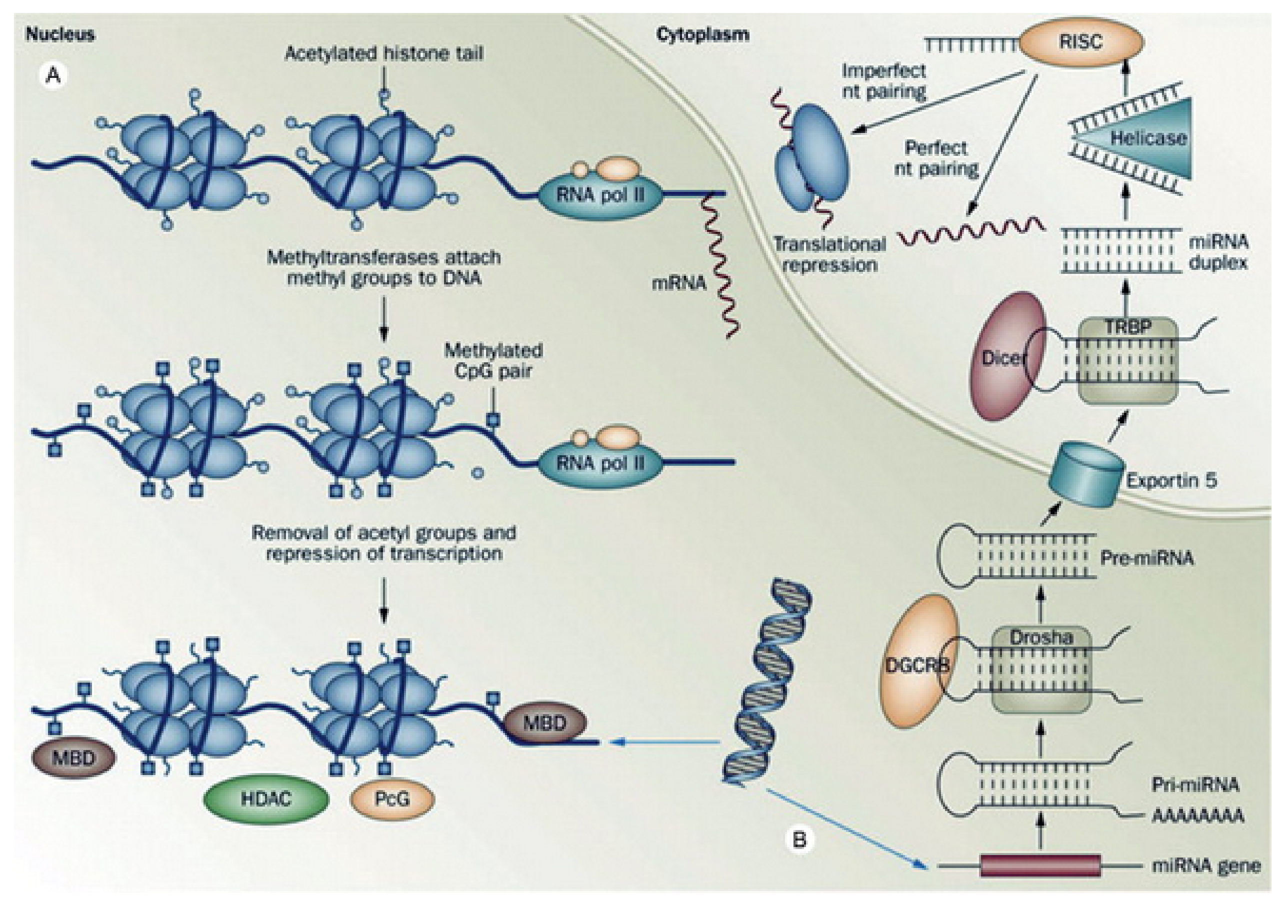
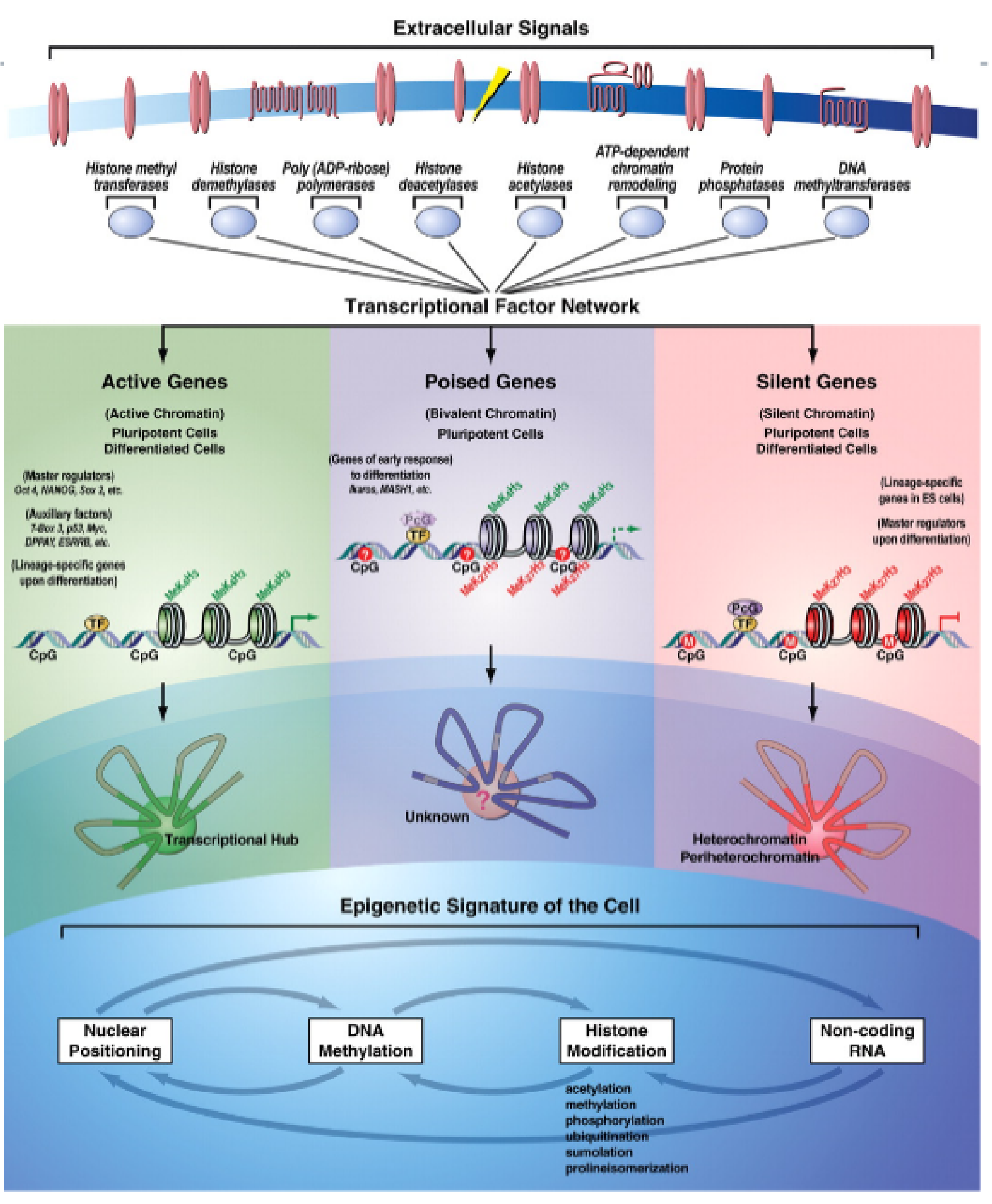
Control of gene expression in eukaryotes



Control of gene expression in eukaryotes



Part 1

Histone modifications:

the histone code

http://www.actrec.gov.in/histome/

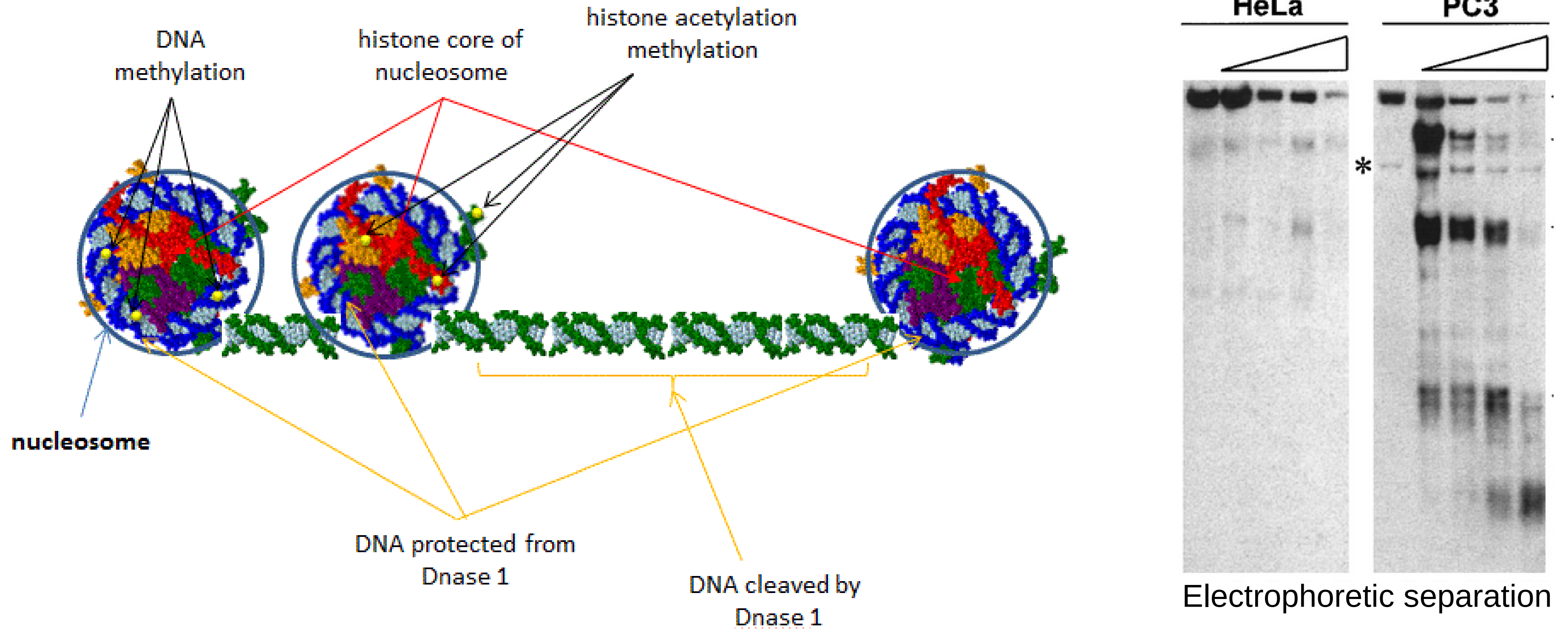
DNAse I assay

DNAse I sensitivity has been for long time used to monitor the accessibility of DNA and determine the chromatin status of genes. The assay basically relies on the fact that DNA tightly packs into histones cannot be easily reached by this endonuclease, while open DNA can.

DNAse I have no recognition sites, it completely degrades DNA.

Upon mild treatment, DNAse treated chromatin forms ladder-like pattern, not smears

=> DNA is always wrapped around histones, it (almost) never unwraps !



Electrophoretic separation

Chromatin status and activity of genes

Accessible chromatin is frequently interpreted as a sign of active transcription. But is it really like this?

- Is a rarely transcribed gene sensitive to DNAseI ? **Yes.**

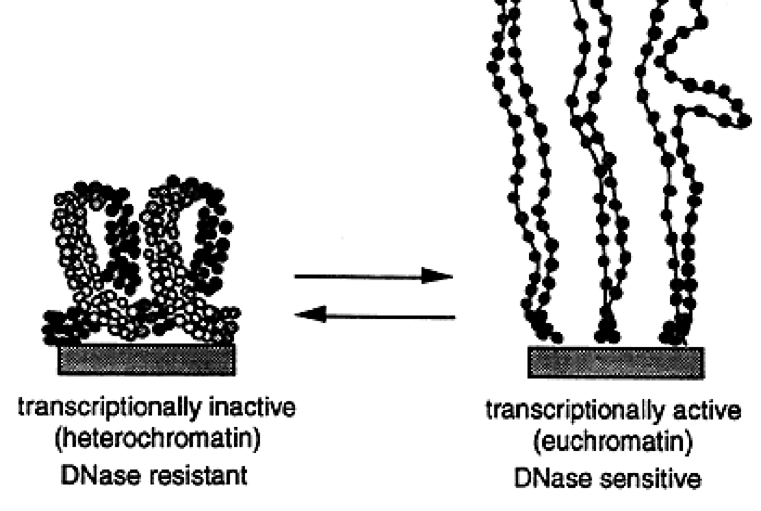
- Is a not yet "stimulated" gene sensitive to DNAseI ? **Yes.**

- After transcription, does a gene becomes insensitive to DNAse ? **No.**

- Before transcription, does a gene is sensitive to DNAse ? **Yes.**

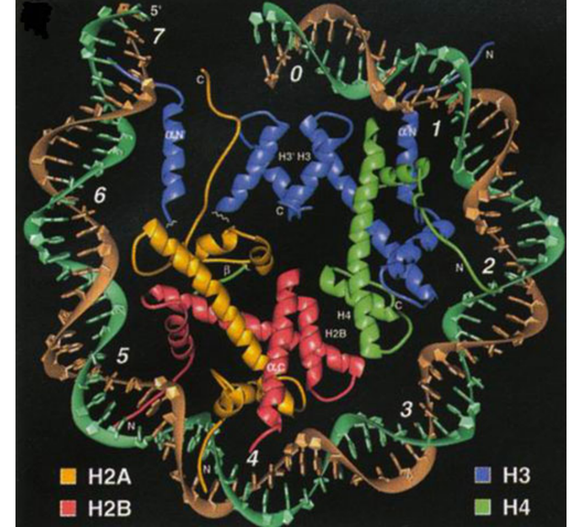
**In fact,
chromatin accessibility
primarily relates to
"usability" not "usage"
of the genes**

This further implies that chromatin accessibility relates to cell commitment, that is determined before the realization of the program.



Histone structure

The histone code hypothesis predicts that the post-translational modifications of histones, alone or in combination, function to direct specific and distinct DNA-templated programs.



(a) Nucleosome core particle

H3 / H4 / H2A / H2B / DNA

Unstructured histone tail

H3/H3 4HB*

H2A-H2B dimer structure

H2A α2 helix

Docking domain

H2B αC helix

H2A α1 helix

H2A L1 loop

H2B L2 loop

H2B/H4 4HB*

Histone octamer

Two copies of H2A form L1/L1 interface

* 4HB – four helix bundle

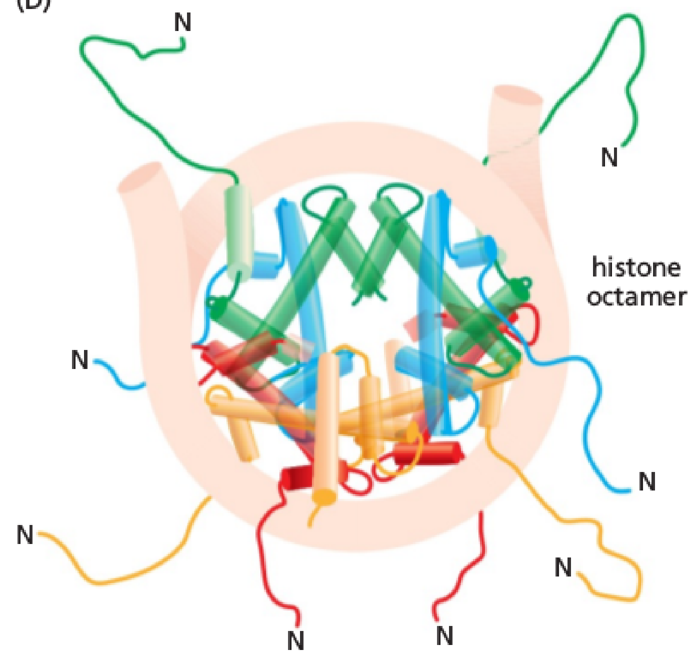
(b) Intact X-ray structure

Hexasome

H2A-H2B/(H3-H4)$_2$ opening

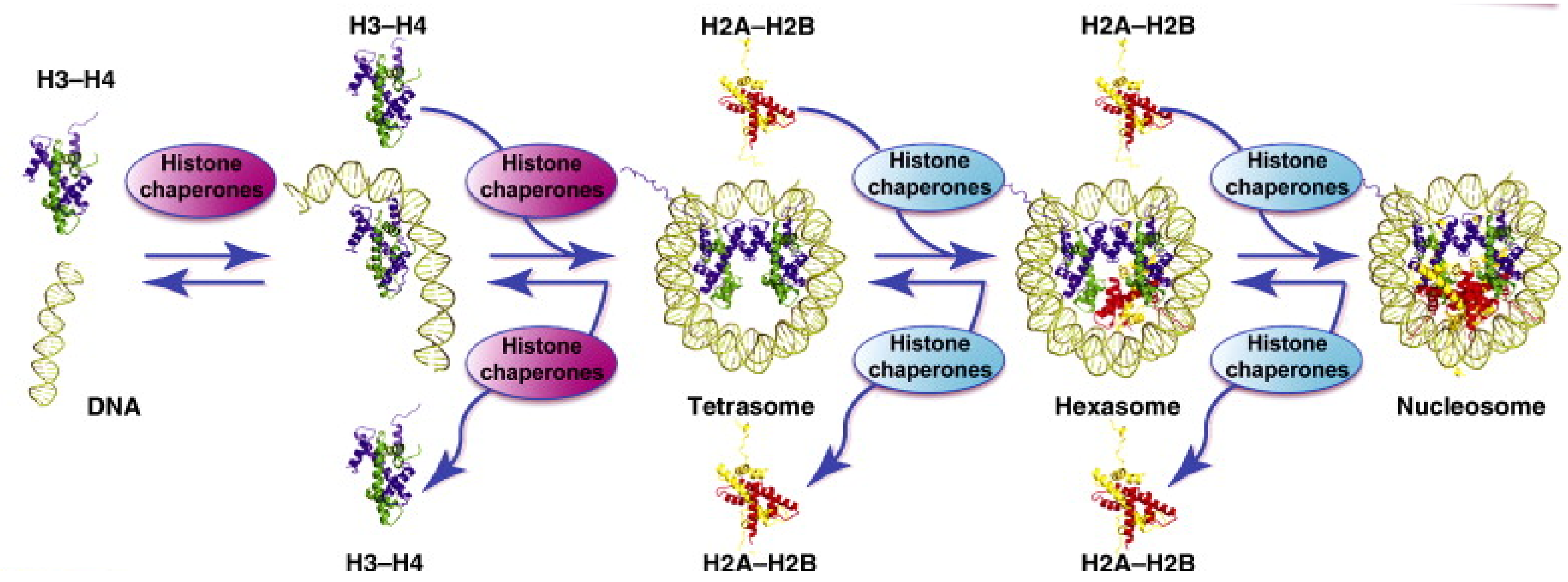
DNA unwrapping

Current Opinion in Structural Biology



Histone assembly

The assembly of histones needs DNA to occur. In solution, histone octamer exist only at 4°C and NaCl > 2M. This reflects the highly cooperative deposition and the affinity of binding of the different histones.



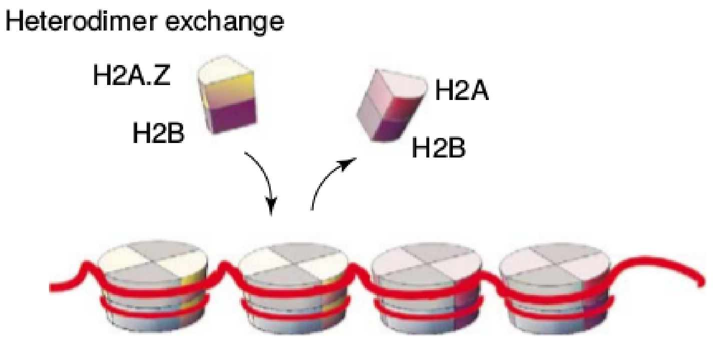
The exact sequence of assembly is:

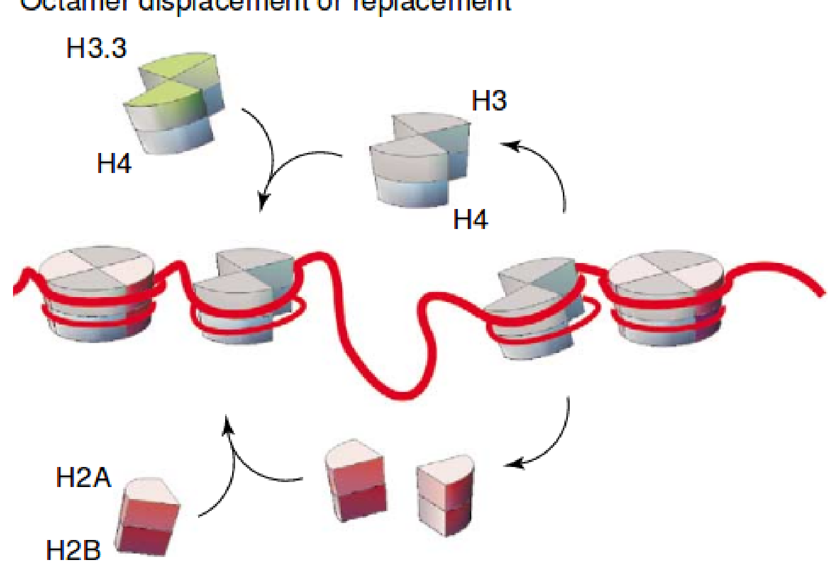
1. H3-H4 dimer contact DNA, then it forms a homo-tetramer.
2. a first H2A-H2B dimer binds, forming a hexasome.
3. the last H2A-H2B dimer binds, forming the final nucleosome

The assembly (and the disassembly) of histones requires the intervention of histone chaperones that modulate the deposition kinetic and order.

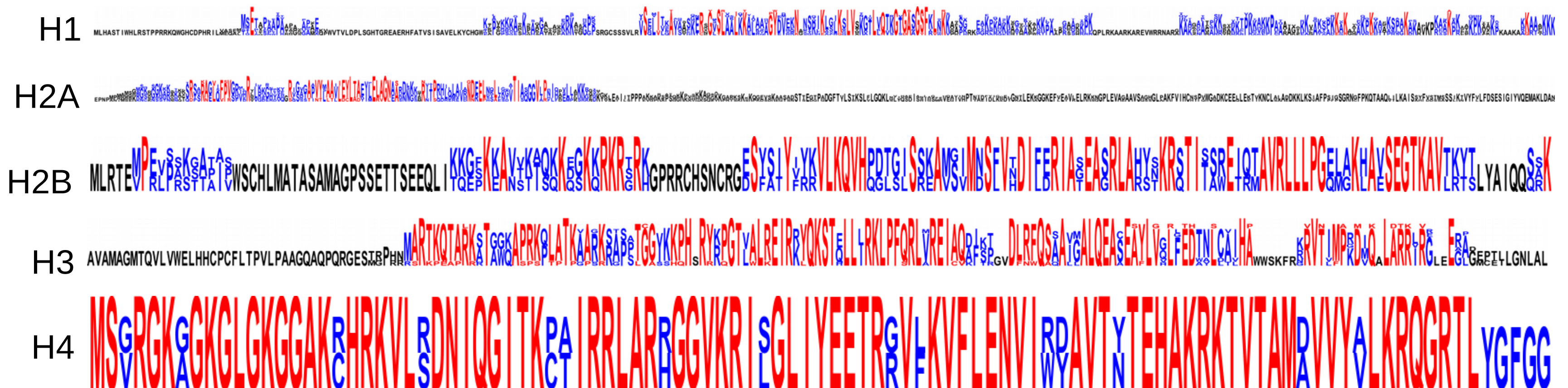
Histone variants

Canonical histones can be replaced by specific variants acting at different stage of chromatin life. They participate in active remodeling but are normally transient.



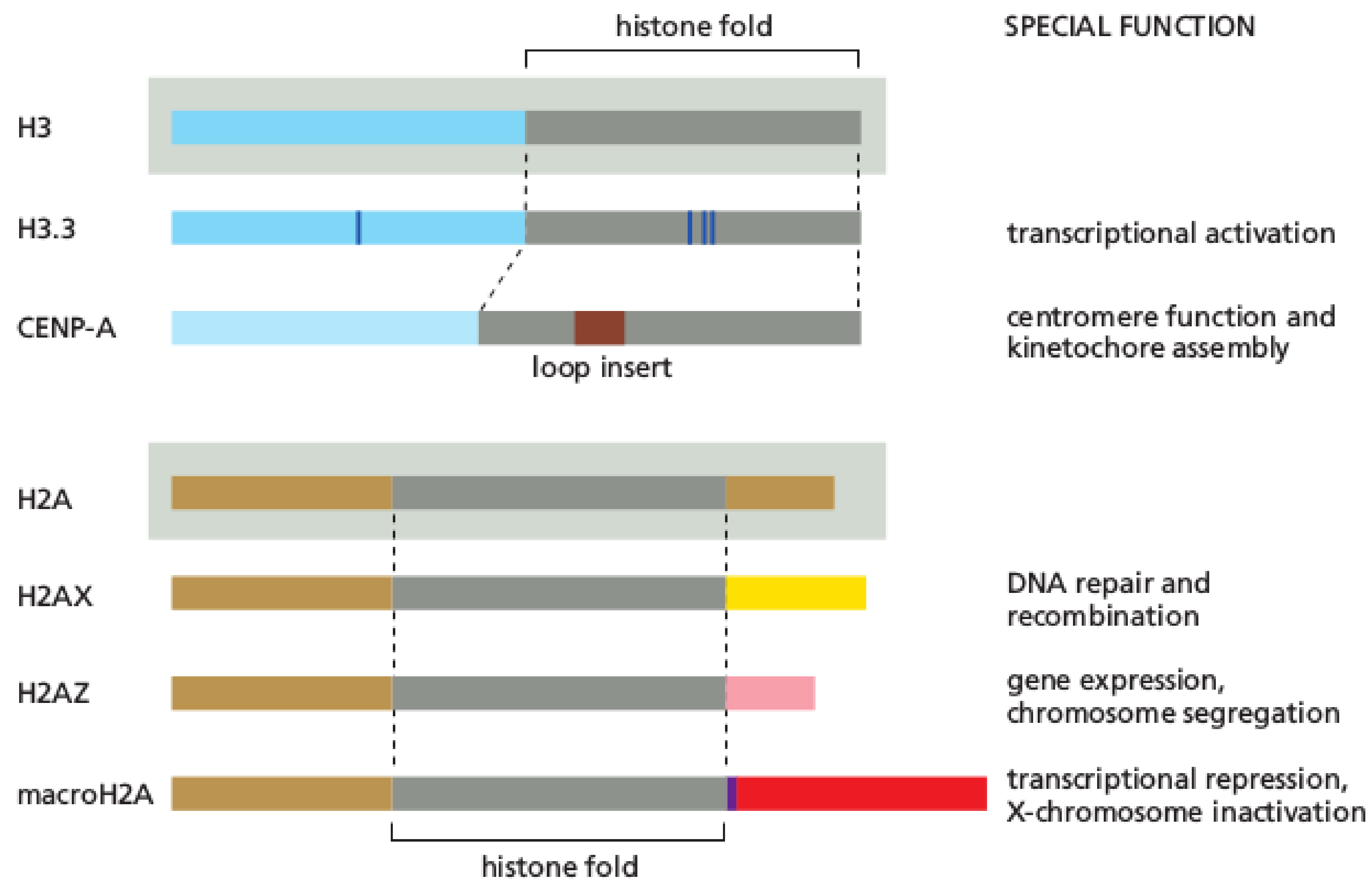


NCBI hosts a curated database of (aligned) histone isoforms: the Histone DB 2.0



www.ncbi.nlm.nih.gov/projects/HistoneDB2.0
Jin et al. 2005 TIBS 30(12):680–687

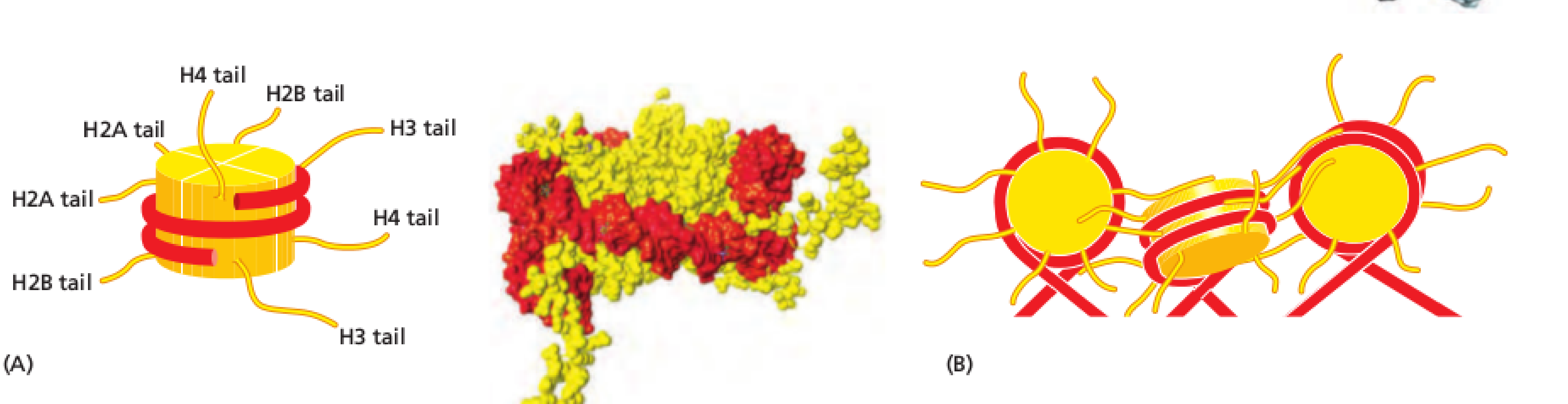
Roles of histone variants



Histone packaging and accessibility

Histones are arranged so that they can bind each other and form compact fibers, that constitute the largest portion of chromatin. Histone packaging actually relies on four main players:

- histone tails (particularly the H4 tail)
- the ancillary histone H1 (linker histone)
- the multitude of proteins that bind DNA
- modifications of DNA

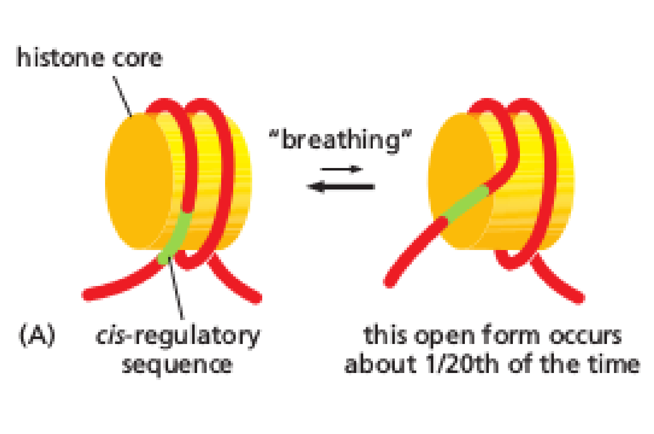


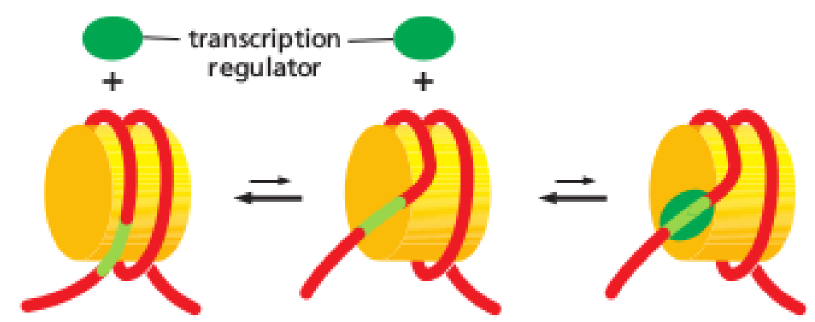
Histones tails are highly solvent exposed and always visible for binding proteins, even when they are involved in the formation of a compact fiber

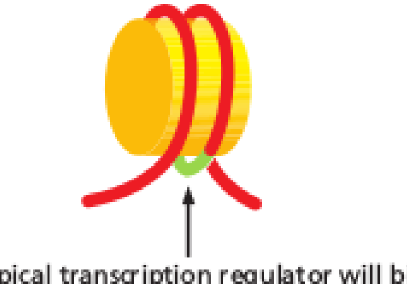
Affinity, cooperativity and dynamics of binding

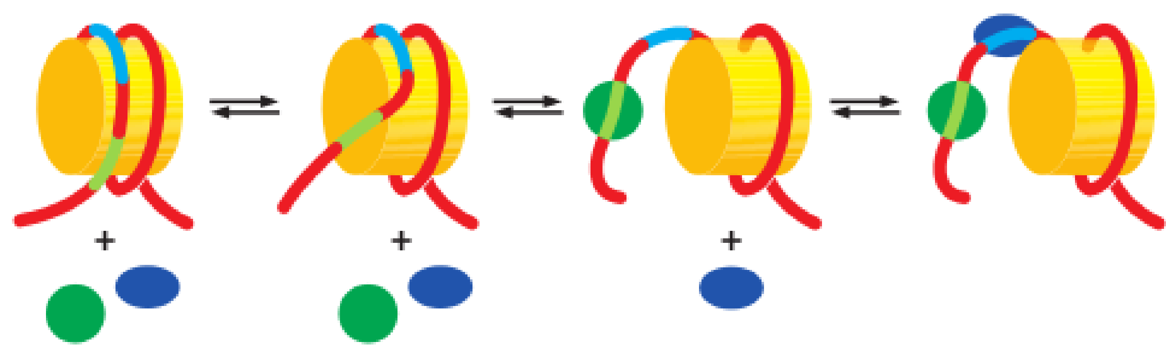
Proteins always bind DNA through weak interactions. In molecular dynamics terms, weakness correlates with a ratio of time of effective binding and time of unbounding.

This dynamic behavior of the interaction holds true both for histones and other binding proteins, meaning that a protein binding DNA with a high affinity displaces a protein with lower affinity and may expose stretches of DNA that were previously hidden.









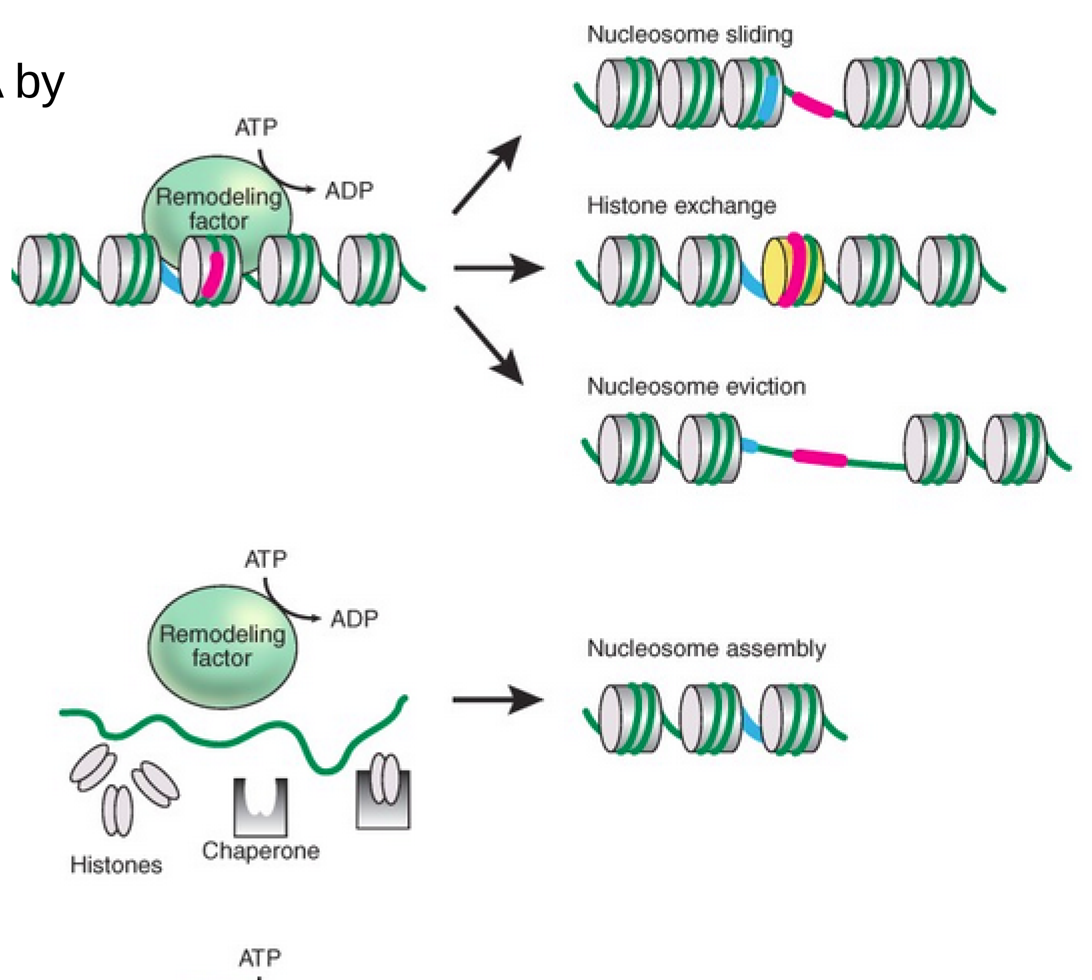
ATP-dependent nucleosome remodeling

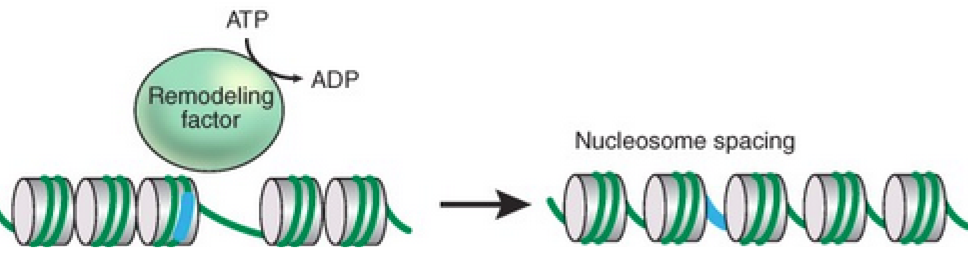
Remodeling is the process of modifying the accessibility of DNA by acting on the nucleosomes.

Energy (ATP) is required to remodel nucleosomes.

Remodeling examples are:

- histone exchange
- partial DNA/histone unwrap
- nucleosome excision
- nucleosome shift
- nucleosome spacing





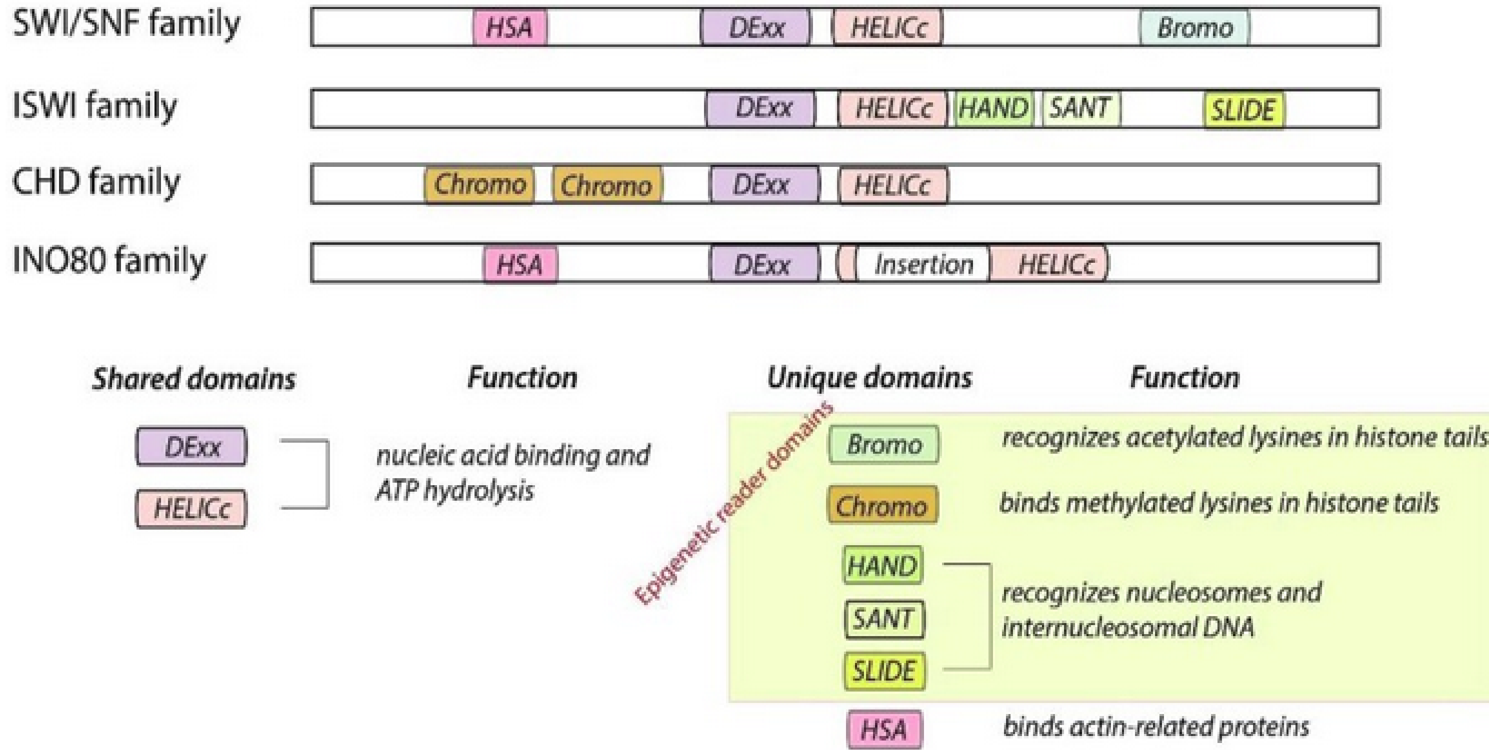
Becker and Workman 2013 Cold Spring Harb Perspect Biol. 5(9):a017905

Chromatin remodeling complexes

There are four subfamilies of ATP-dependent chromatin remodeling complexes:

1. SWI/SNF
2. INO80
3. ISWI
4. CHD

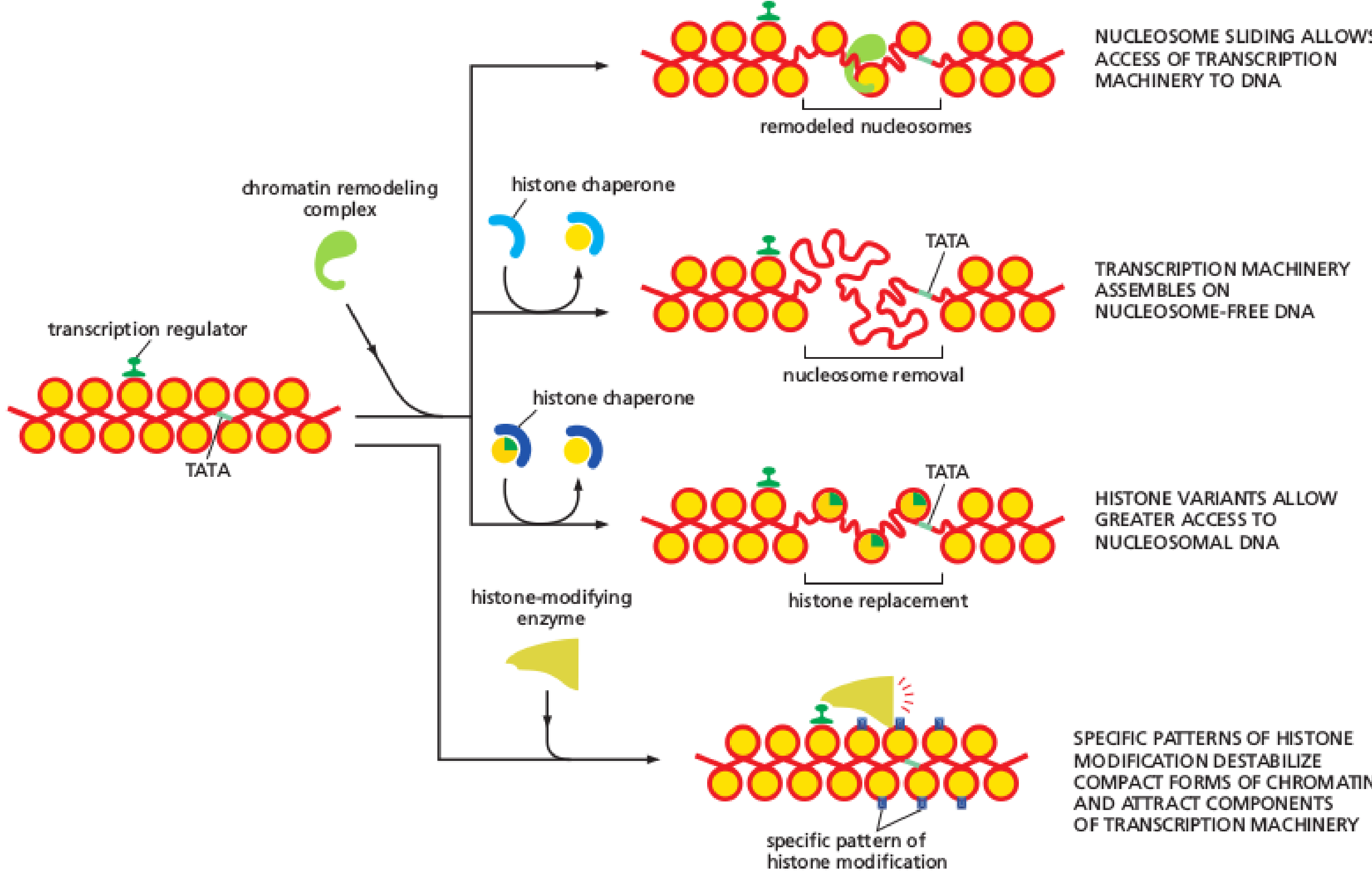
Each family is defined by **a characteristic ATPase** subunit that is related to the DExx superfamily of **DNA helicases**, but they also possess unique motifs that mediate binding to nucleosomes and individual complex subunits.



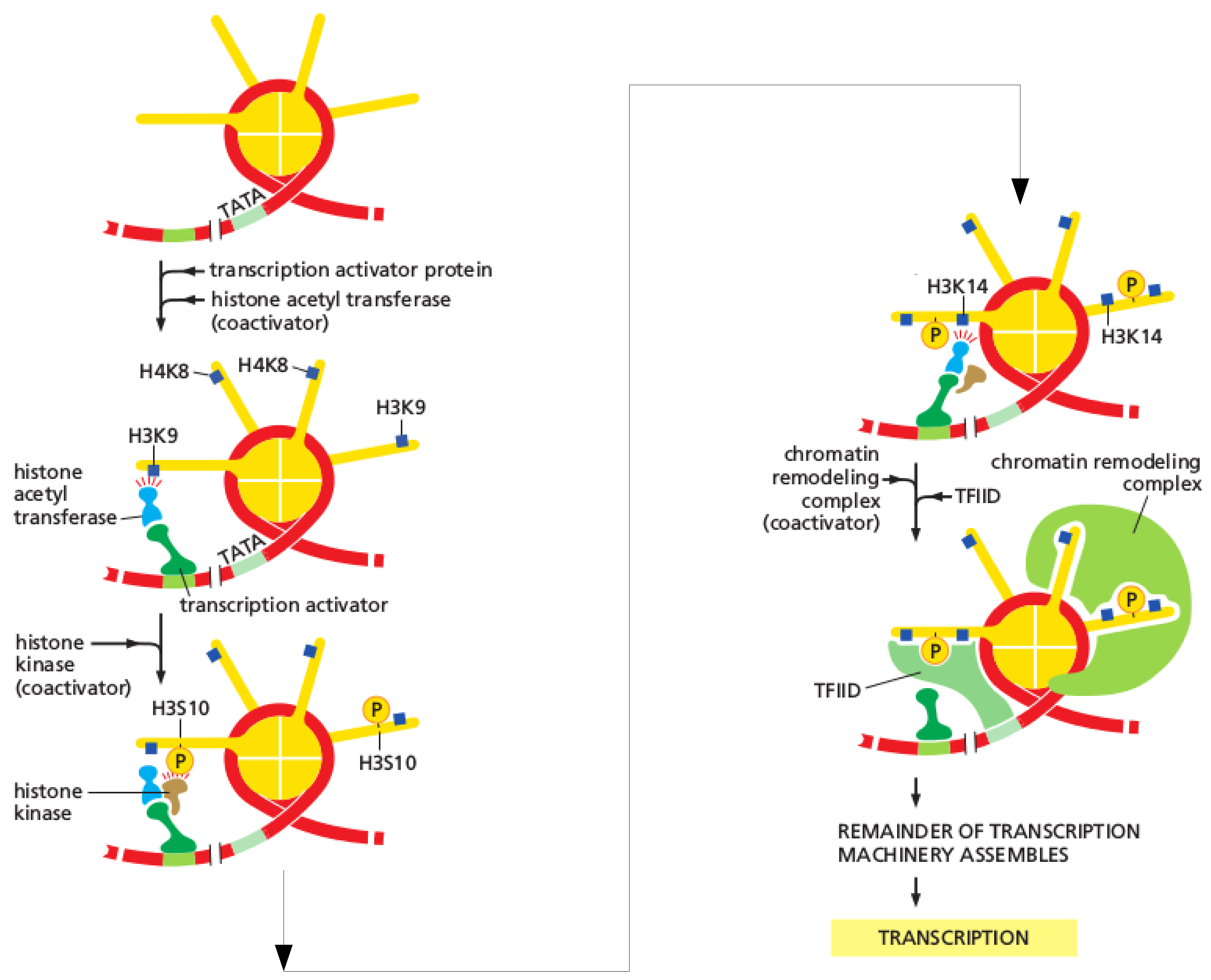
Längst and Manelyte 2015 Genes 6(2):299-324

Accessing histone-bound DNA: the transcription activators

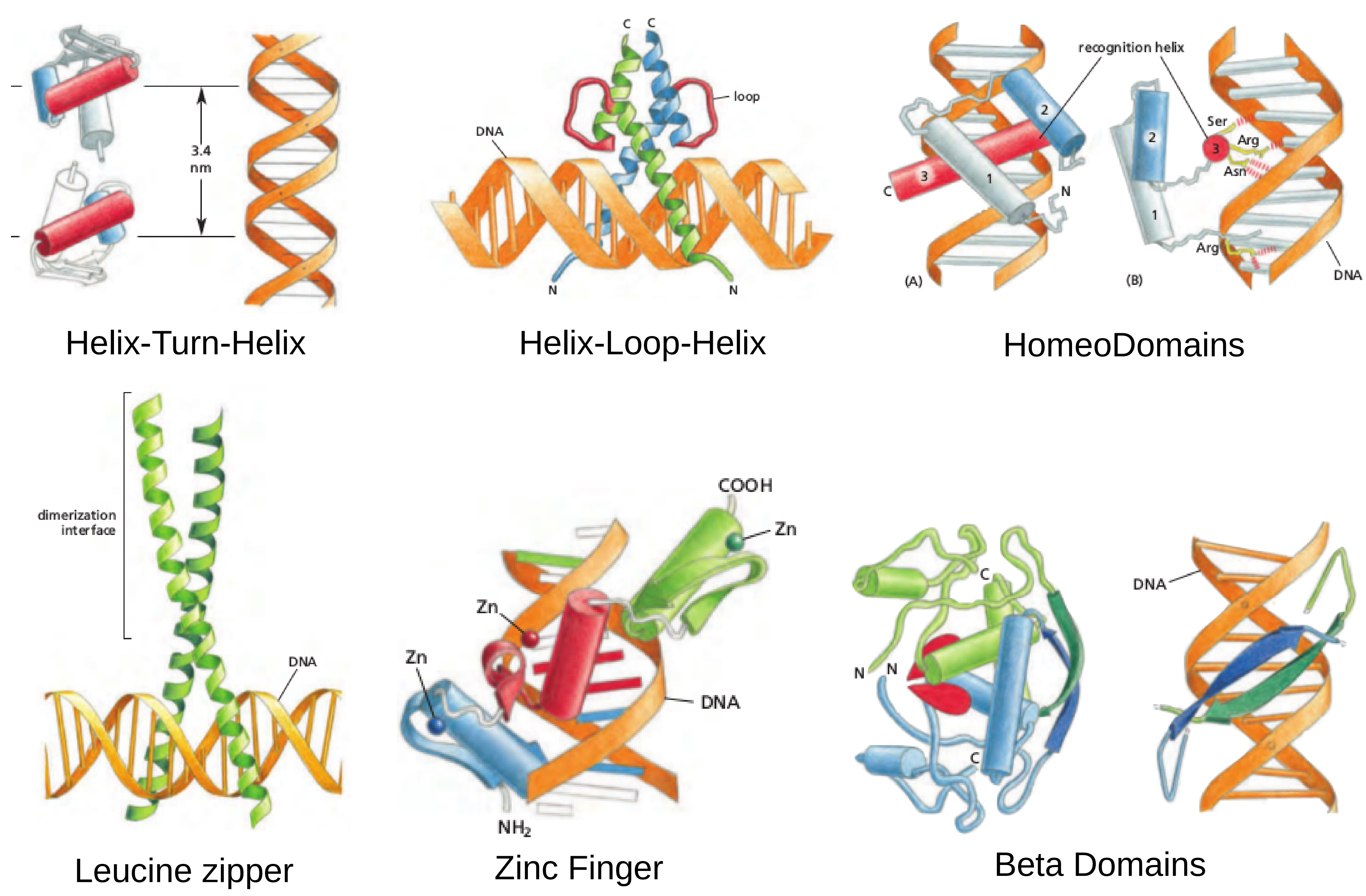
For transcription, chromatin must be opened and DNA read. But normally DNA is collapsed around the core of histones. There are several ways of reversibly “freeing” DNA from histones or loosening DNA/histone interaction. These alterations increase the accessibility of DNA and facilitate the binding of RNA polymerase and the general transcription factors.



Histone modifications: controlling DNA availability



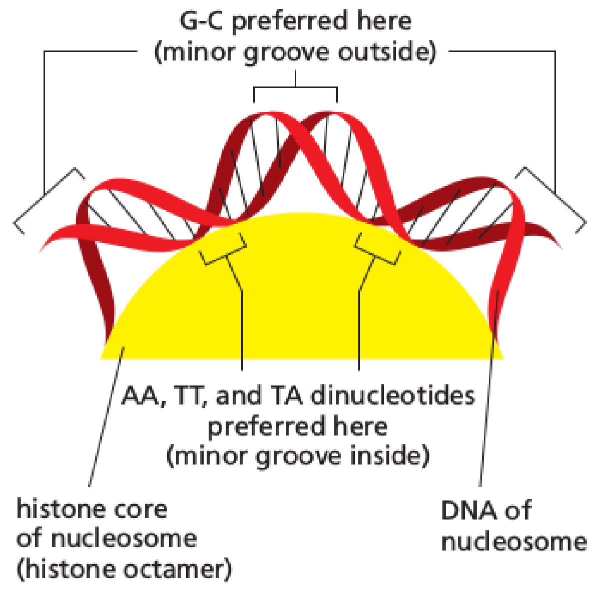
DNA binding motifs in proteins



Reading DNA without opening the helix

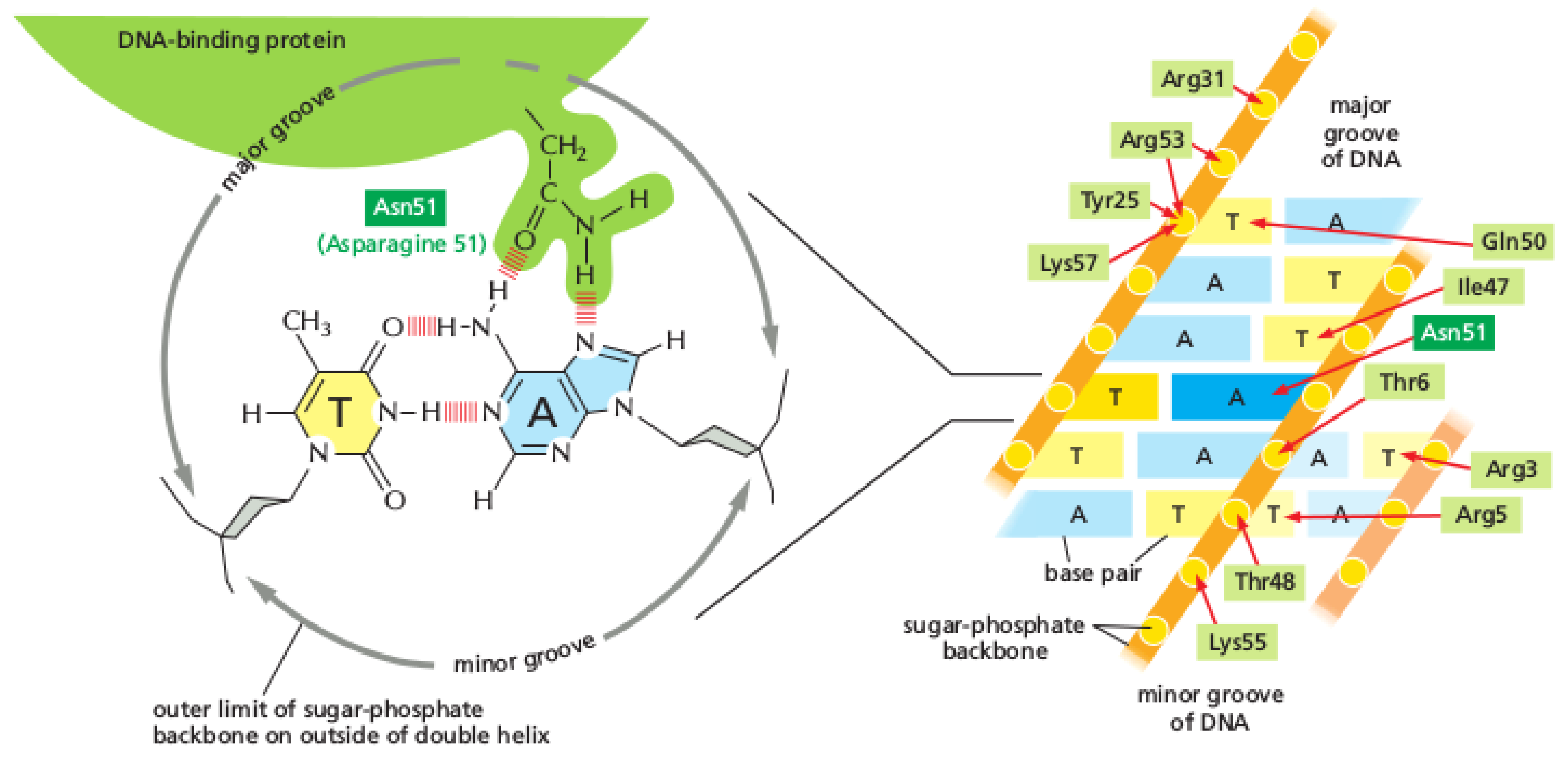
Different combinations of A/T/G/C base pairings design on the major and minor grooves of the DNA specific chemical patterns.

The geometric regularity of the double helix helps binders in sliding over DNA, searching for chemical patterns for which they are designed to interact.



Reading DNA without opening the helix

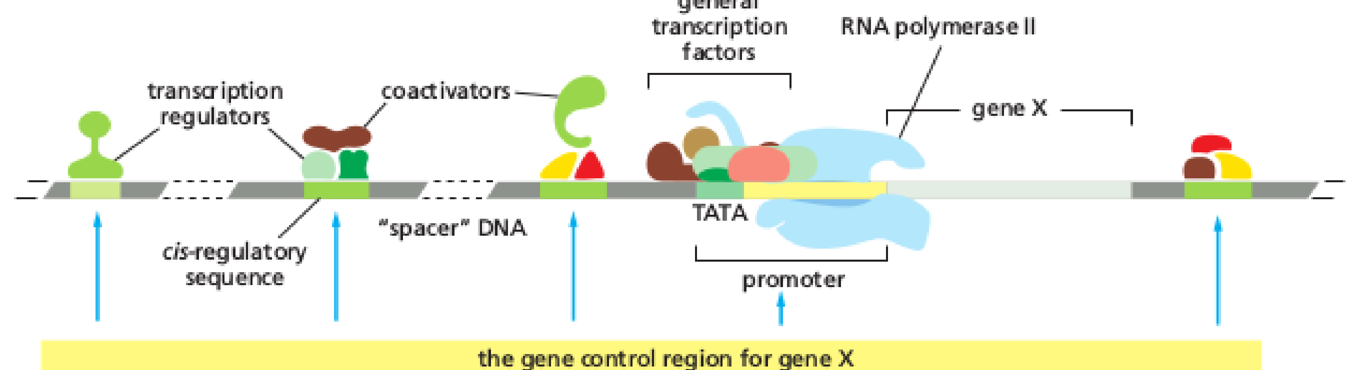
DNA binding proteins, such as transcription regulators, “read” the DNA without opening the double helix by using residues that make weak interactions with chemical patterns on the major or minor groove (or both, depending on the complexity of the protein) and with the phosphate groups of the sugar backbone of the nucleic acid. Only residues with specific chemical features and that are specifically exposed on the surface of DNA binding proteins can play a role in the interaction.



The DNA transcription complex of eukaryotes

The transcription machinery of eukaryotes is a complicated interaction network of several DNA/protein and protein/protein complexes controlled by the sequences and modifications present in the control region of a gene.

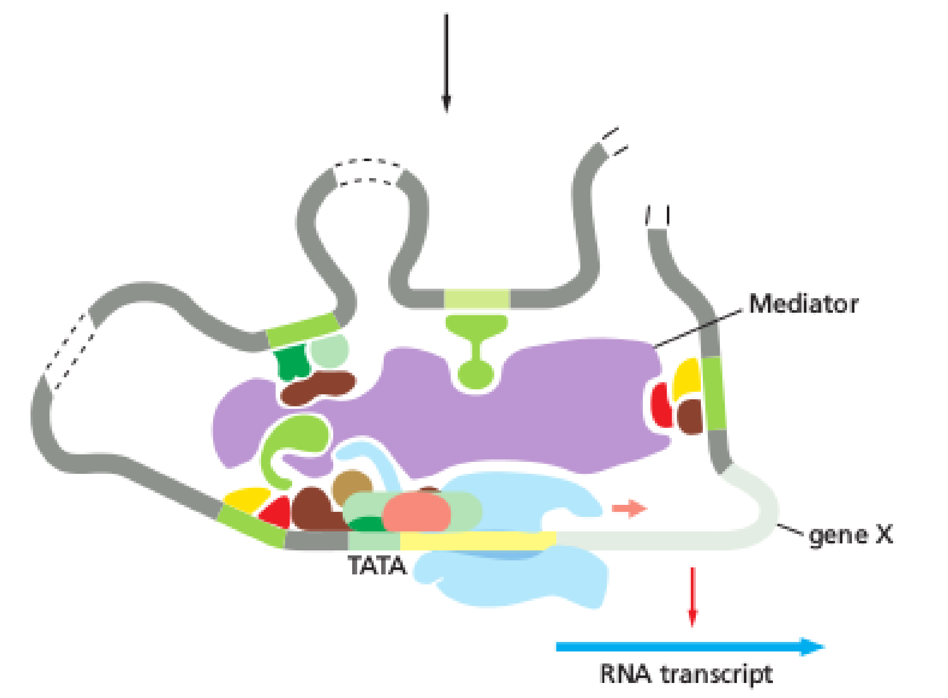
- Activators
- Adapters
- Regulators
- Transcription factors
- Polymerases



The transcription complex can be rationally described along a DNA sequence, according to the specific binding sites.

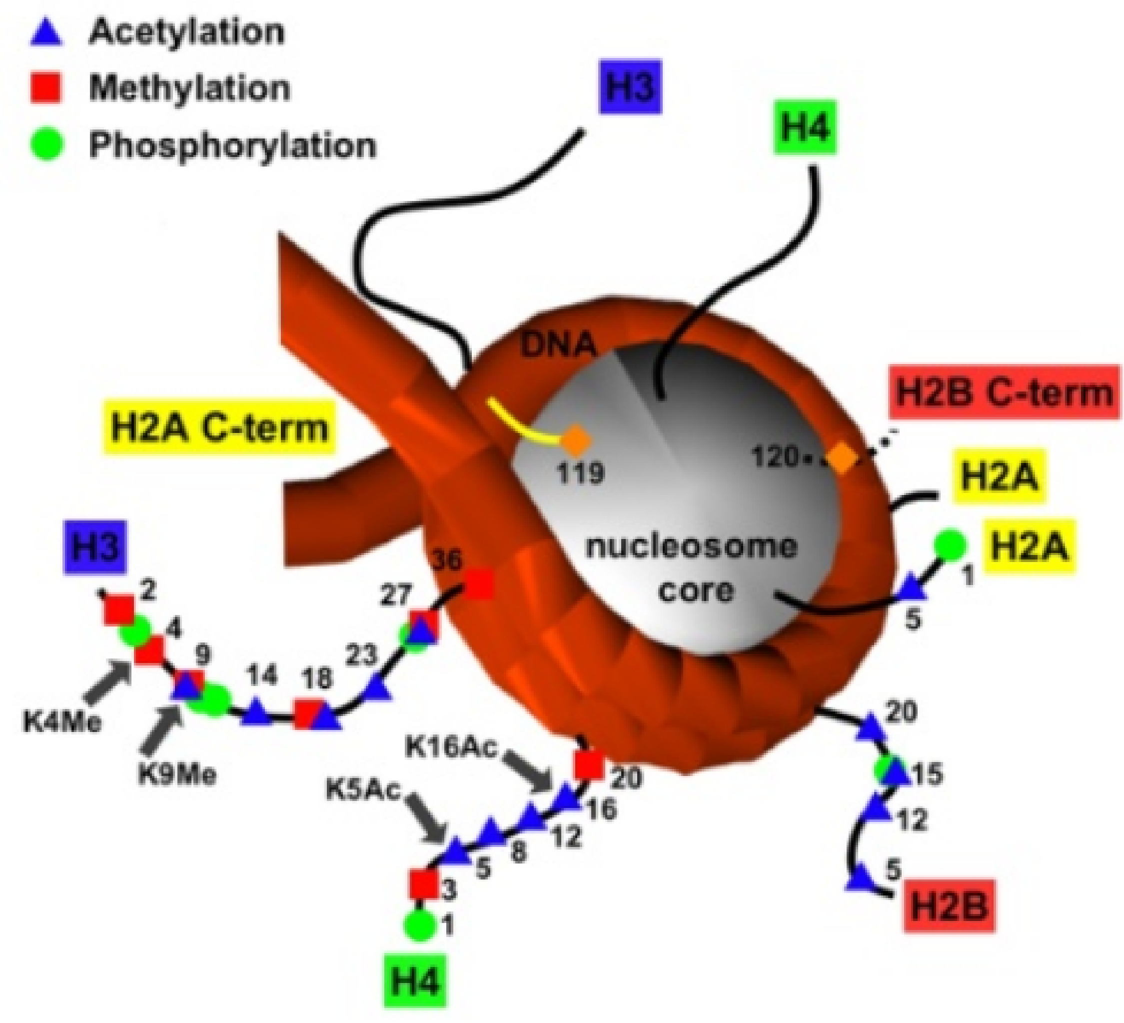
In reality, the full machinery self assembles in a packed three-dimensional structure where DNA/protein and protein/protein interactions occur.

The composition of such network can be different from gene to gene...

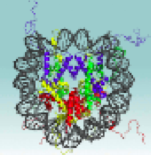


The histone code

The histone code hypothesis predicts that the post-translational modifications of histones, alone or in combination, function to direct specific and distinct DNA-templated programs.

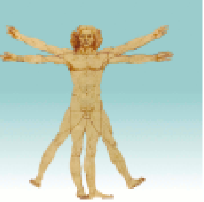


The histone code



*HI*stome: The *H*istone *I*nfobase

A database of human histones, their post-translational modifications and modifying enzymes



- Home
- How to use HIstome
- Lead-in
- Histones and histone variants
- Histone PTMs
- Histone modifying enzymes - I. The writers
- Histone modifying enzymes - II. The erasers
- Search *HI*stome
- Contact us

SGRGKQGGKARAKAKTRSSRAGLQ **H2AX** LPKKTSATVGPKAPSGGKKATQASQEY

H2AX "Linker"

SQ motif

SGRGKQGGKARAKAKTRSSRAGLQ **H2A** LPKKTESHHKAKGK

PEPAKSAPAPKKGSKKAVTKA **H2B** KAVTKYTSSK

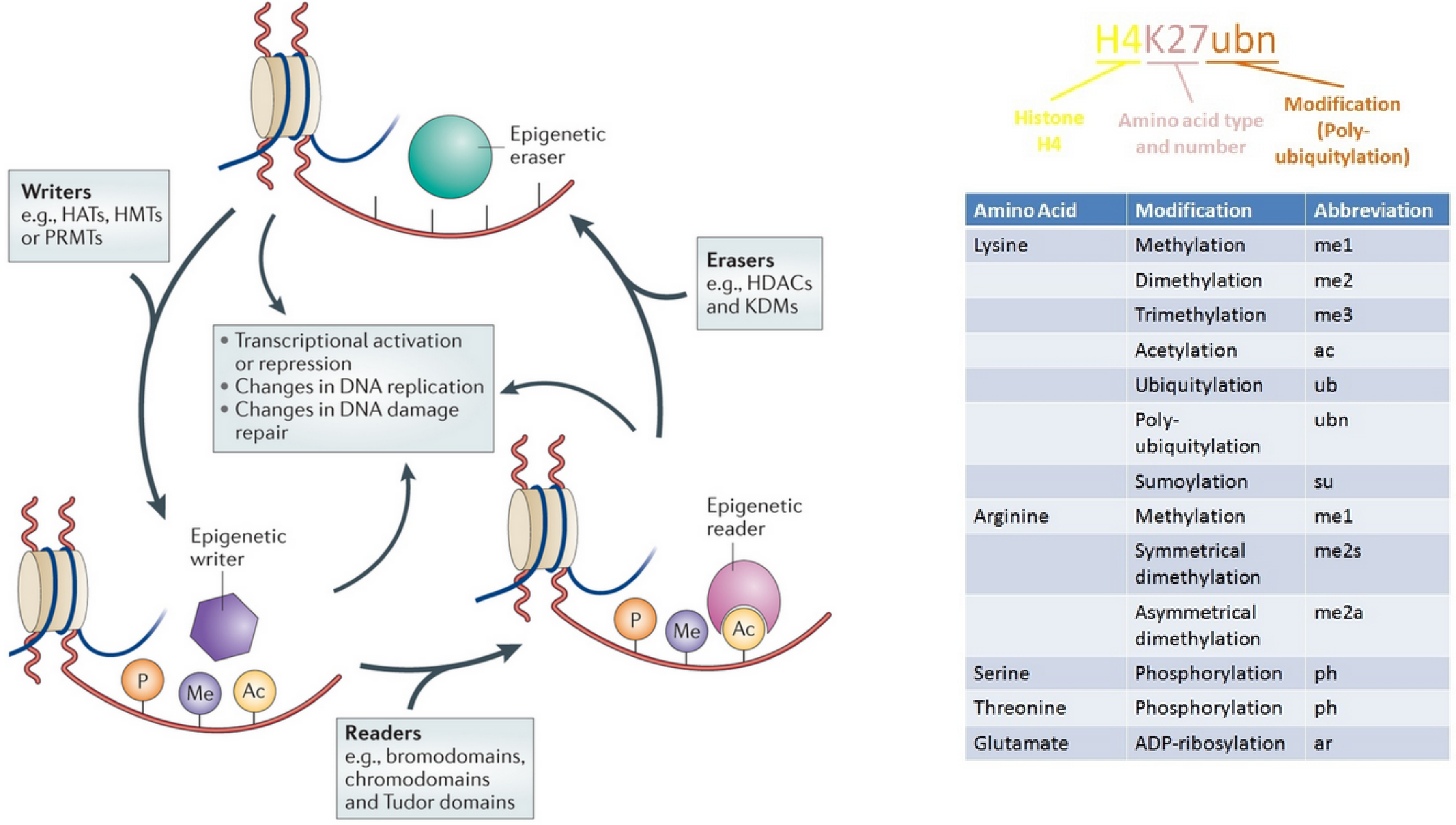
ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKP **H3** GERA

SGRGKGGKGLGKGGAKRHRKVLR **H4** YGFGG

- Ac Acetylation
- Bi Biotinylation
- Ci Citrullination
- Me Methylation
- P Phosphorylation
- PIso Proline Isomerization
- Pr Propionylation
- Rb ADP-ribosylation
- Ub Ubiquitylation

The histone code

A specific combination of signals is imposed on specific regions of chromatin by **writers** and then interpreted by **readers** (effector proteins) through dedicated binding domains. **Erasers** eventually switch off the signals. Upon recognizance, readers mediate the formation of complexes that **orchestrate distinct and consistent cellular processes**, such as replication, transcription, DNA repair and chromosome condensation.

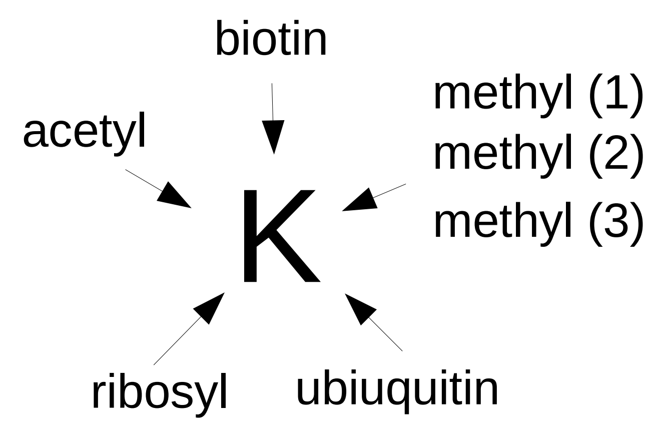


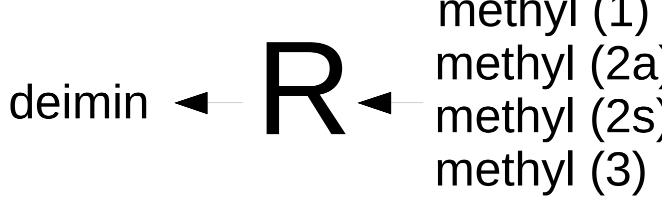
Amino Acid	Modification	Abbreviation
Lysine	Methylation	me1
	Dimethylation	me2
	Trimethylation	me3
	Acetylation	ac
	Ubiquitylation	ub
	Poly-ubiquitylation	ubn
	Sumoylation	su
Arginine	Methylation	me1
	Symmetrical dimethylation	me2s
	Asymmetrical dimethylation	me2a
Serine	Phosphorylation	ph
Threonine	Phosphorylation	ph
Glutamate	ADP-ribosylation	ar

Writers

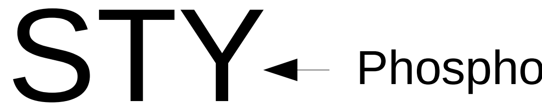
Various chemical groups can be used to modify histones, such as:

- Acetyl: Histone (K) acetyl transferases (HATs)
- Methyl: K/R Methyl Transferases (SET, PRMT)
- NH3-removal: protein R deiminases (PADs)
- Phosphate: S/T/Y kinase (AGC, CaMK,CMGC, PTK)
- Ribose: K ribosylase (PARP)
- Ubiquitin: K ubiquitinase
- Biotin: holocarboxylase synthetase





Enzymes driving the modifications have catalytic domains as well as reader domains to recognize specific PTM coded histone substrates.



=> PTMs guide other PTMs !!!

https://www.cellsignal.com/common/content/content.jsp?id=science-tables-histone

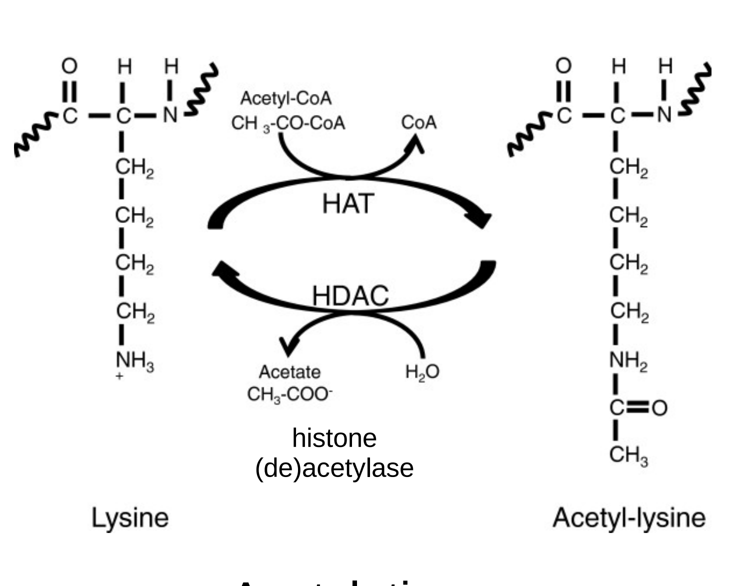
Erasers

Since histone PTM are always reversible (not necessarily with the same mechanisms), the various chemical groups can be removed to restore original residues:

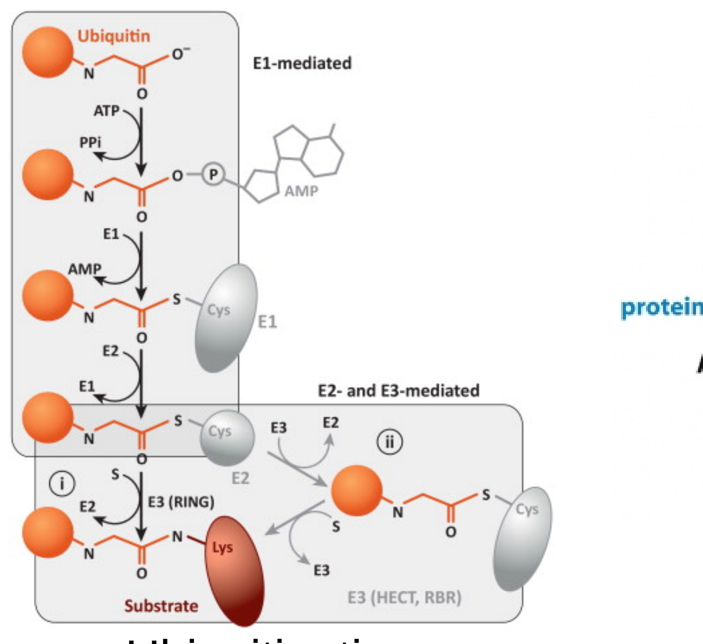
- **Acetyl**: Histone (K) deacetylases (HDACs, Sirtuins)
- **Methyl**: K/R demethylase (LSD,KDM,JMJD)
- **Phosphate**: S/T/Y kinase (PPP,PPM,PTP,)
- **Ribose**: Poly (ADP-ribose) glycohydrolase (PARG)
- **Ubiquitin**: deubiquitinases
- **Biotin**: biotinase

https://www.cellsignal.com/common/content/content.jsp?id=science-tables-histone

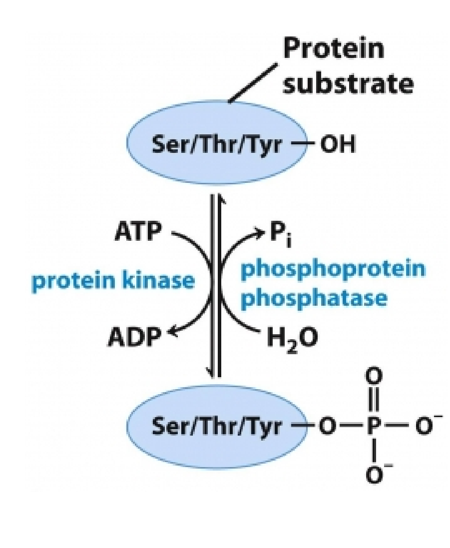
A wealth of reactions...



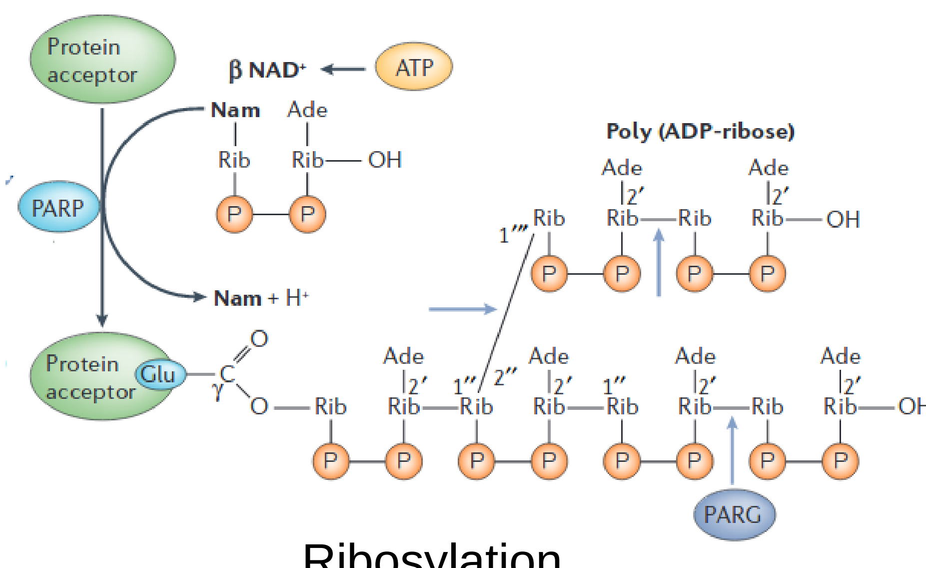
Acetylation



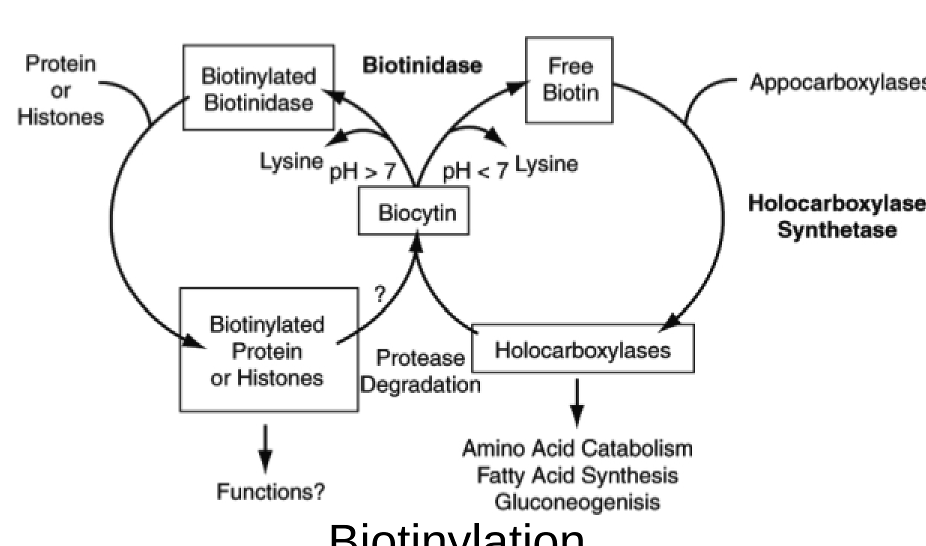
Ubiquitination



Phosporylation



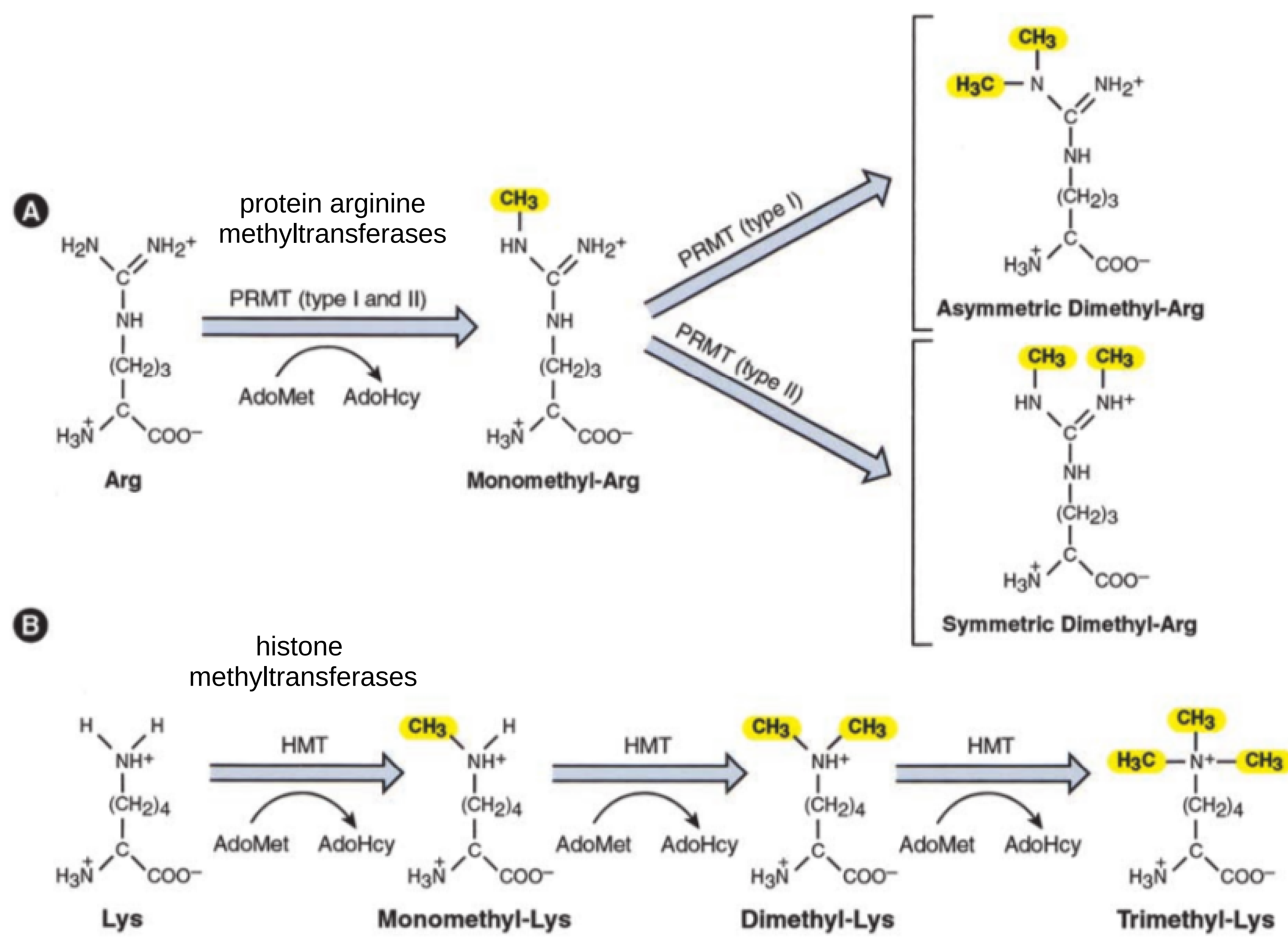
Ribosylation



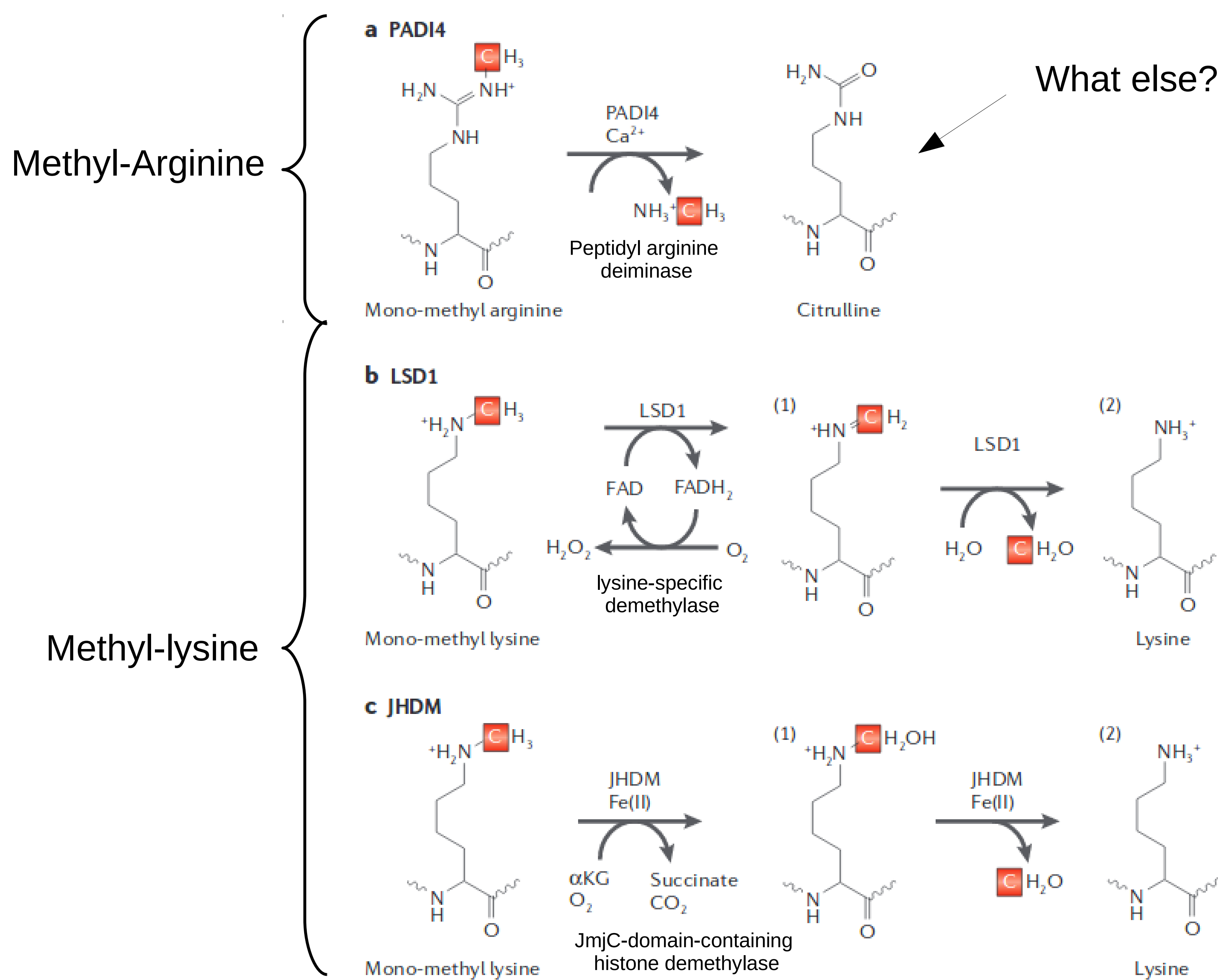
Biotinylation

matteo.ramazzotti@unifi.it

Methylation: different patterns on the same residue



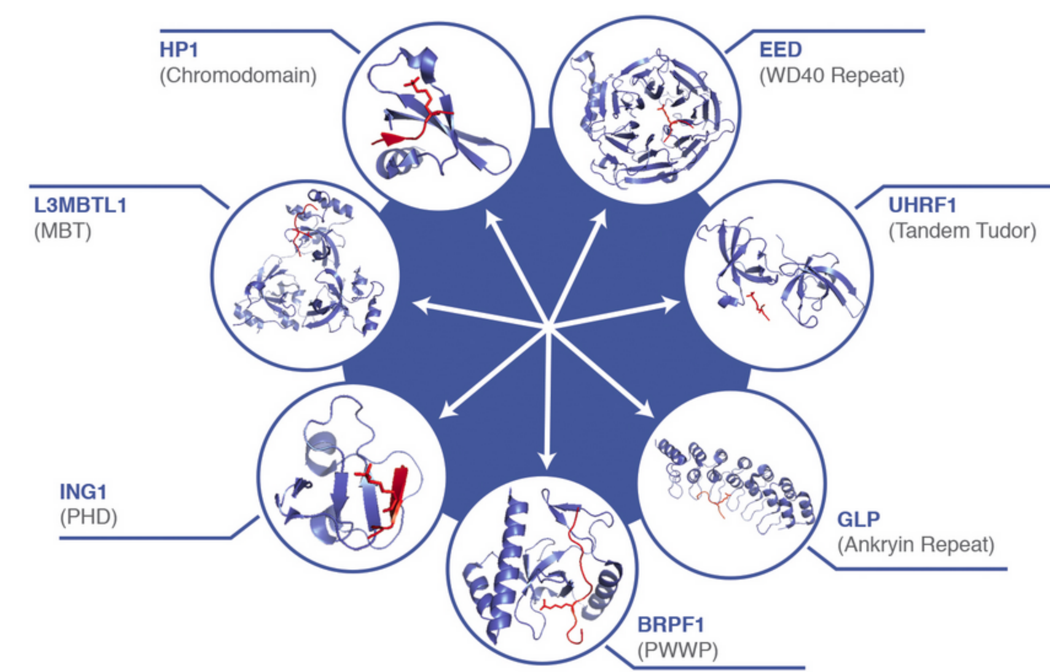
Demethylation: several ways

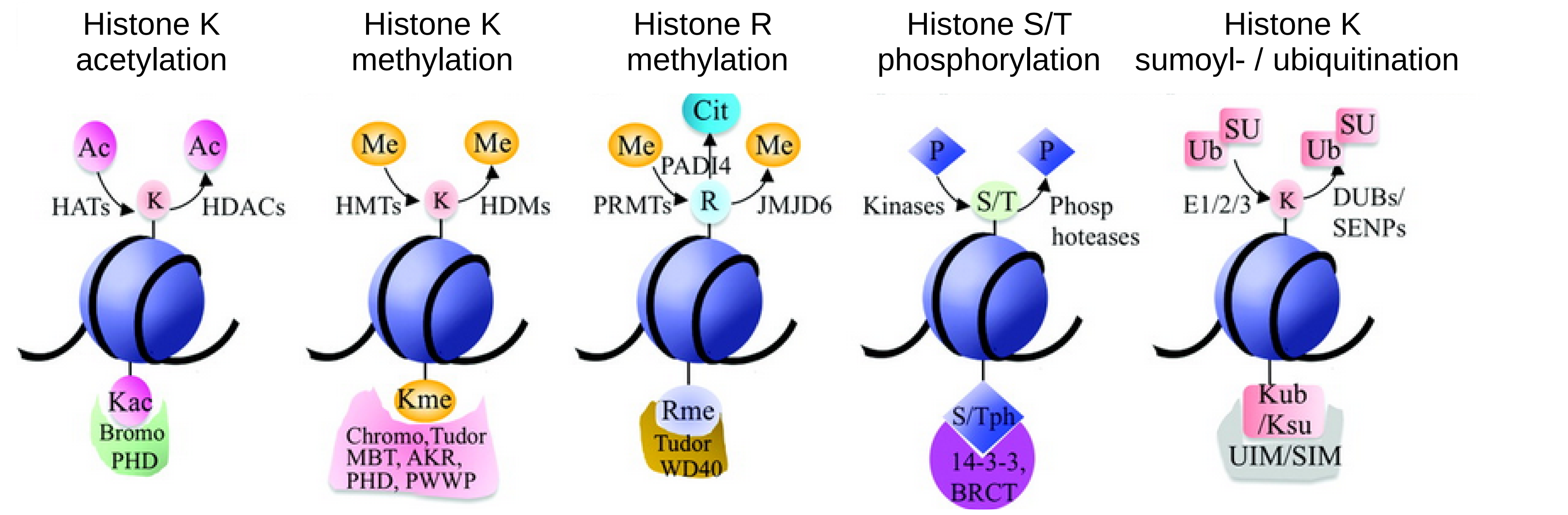


Readers and reader domains

Histone modifications provide docking sites for the recruitment of specific binding proteins which recognize modified histones via specialized domains.
Each domain has evolved with one specific target, so readers frequently harbor several of such modules.

Other domains contained in chromatin binding proteins then direct specific effector functions.





Reader domains

Reader domains are contained in writers, erasers and protein binders: in fact their unique property of binding DNA at specific sites make them the perfect cross-agents for protein that work on DNA "interpretation" and repair.

Domain	Structure	Proposed role	Complexes	Proteins
Bromo		Binds acetylated lysine	SAGA	Gcn5, Spt7
			NuA3	Yta7
Chromo		Binds methylated lysine	SAGA	Chd1
			NuA4	Esa1, Eaf3
Tudor		Binds methylated lysine and arginine	SAGA	Sgf29
SANT		Predicted to bind to histone tails	SAGA	Ada2
			NuA4	Eaf1, Eaf2

Domain	Structure	Proposed role	Complexes	Proteins
SWIRM		Predicted to regulate transcription through protein–protein interactions	SAGA	Ada2
WD40		Binds methylated lysine	SAGA	Spt8
			HatB	Hat2
			Elongator	Elp1, Elp2
PHD		Binds methylated lysine	NuA4	Yng2
			NuA3	Yng1, Nto1

Lee and Workman 2007 Nat Rev Mol Cel Bio. 8:284-295

Reading acetylations

Acetylation is largely the most investigated histone PTMs. **Lysine** is the target of acetylations and the overall role of Kac is the relaxation of the chromatin structure due to a neutralization of the positive charge of histones by the Ac moiety. In fact, acetylation **changes the overall charge.**

H2A 5,9 ac

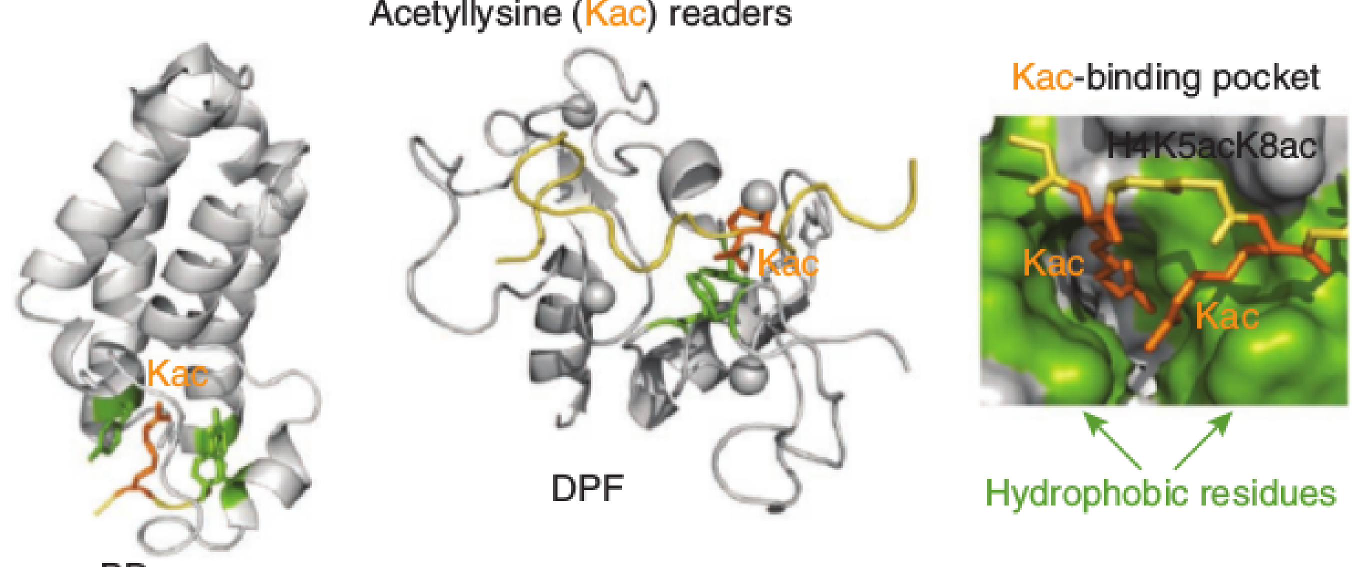
H2B K5,12,15,16,20,120 ac

H3K 4,9,14,18,23,27,36 ac

H4K 5,8,12,16,20,91 ac

} gene-activation marks

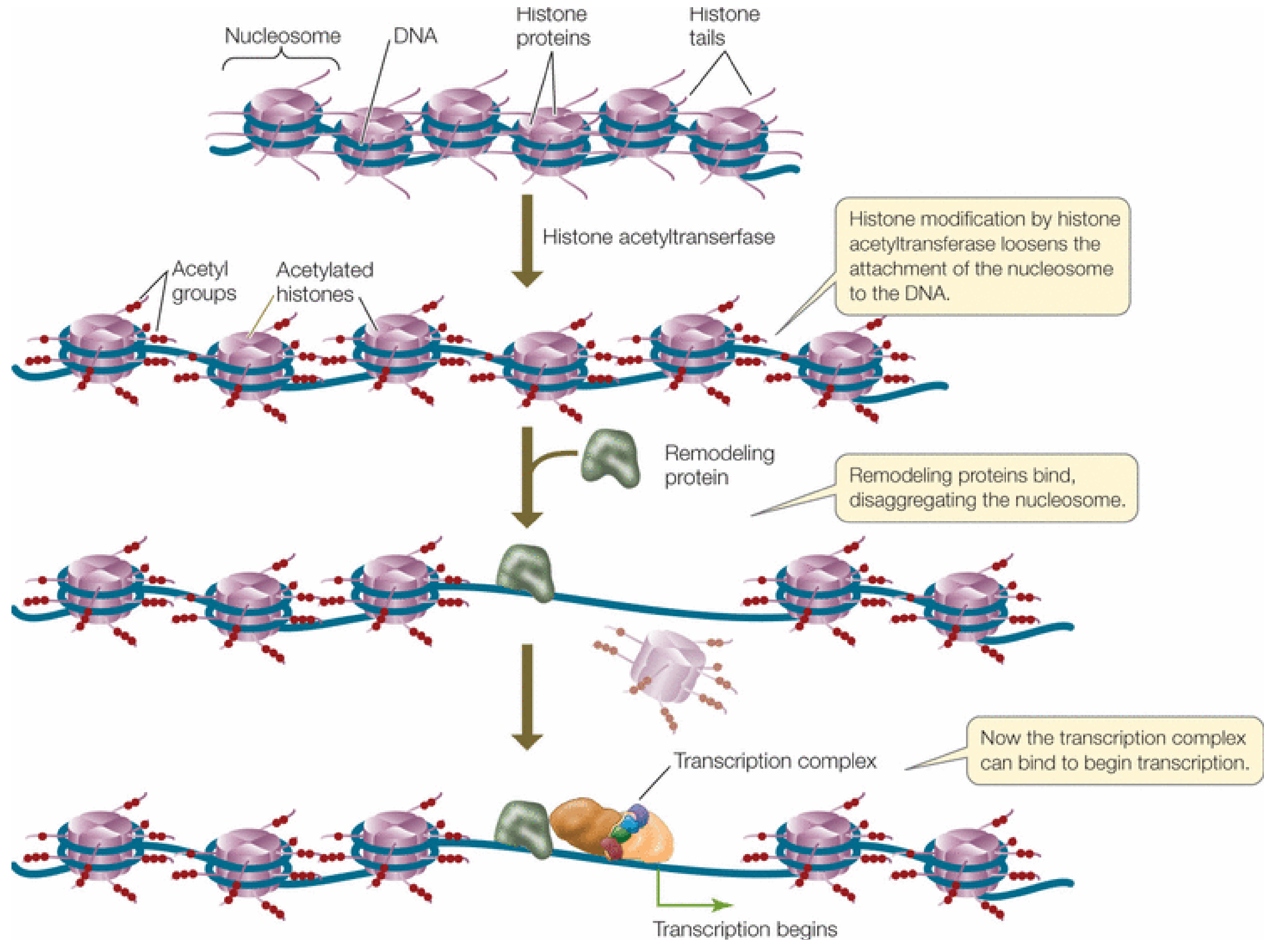
H3K56ac } gene-silencing mark



Musselman et al. 2012 Nat Struct Mol Biol. 19(12):1218-27

Histone acetylation and increase in transcription

A general tendency can be described: open (active) chromatin (euchromain) is hyper-acethylated. In fact, the loss of + charge of K loosen the DNA/histone assication. And in fact, closed (inactive) chromatin (heterochromain) is hypo-acetylated.



Lysine residues: a competition site

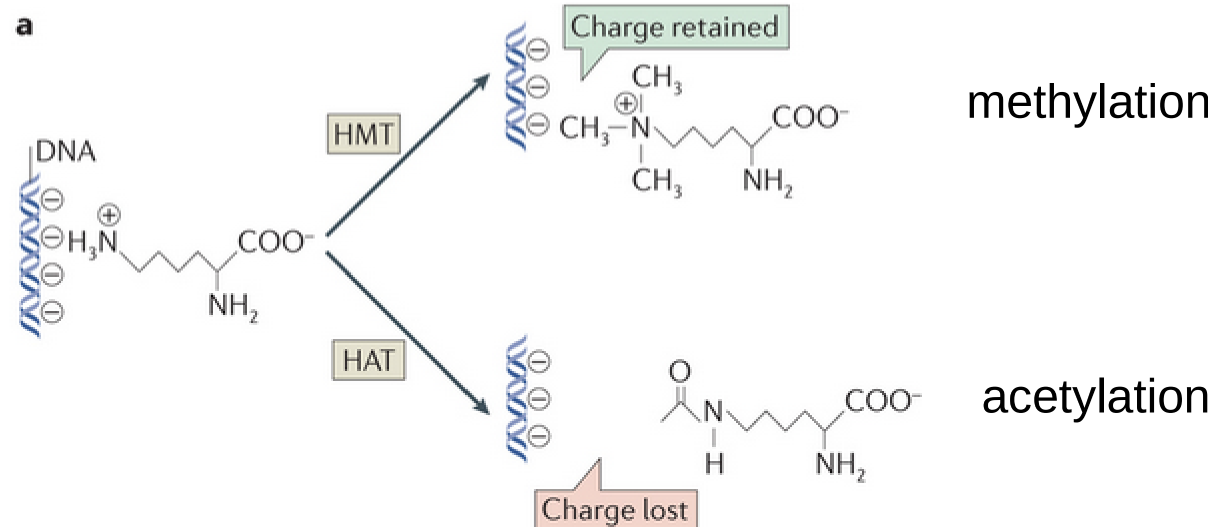
acetyl lysine

lysine

monomethyl lysine

dimethyl lysine

trimethyl lysine



Reading methylations

Methylation is perhaps the most versatile of all histone PTMs. Both **lysine** and **arginine** can be methylated, and each has three (four) possible methylation states. Unlike other modifications, methylation **does not change the overall charge**, although it does alter the hydrophobic character and size of the modified residue.

Lysine:

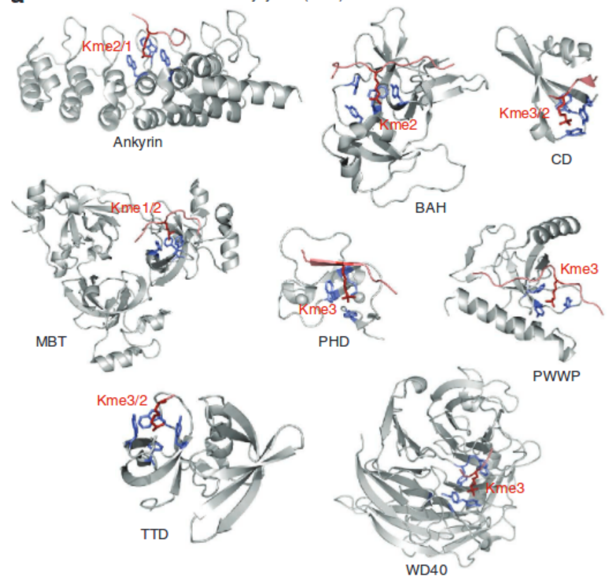
H3K4me, H3K36me } gene-activation marks

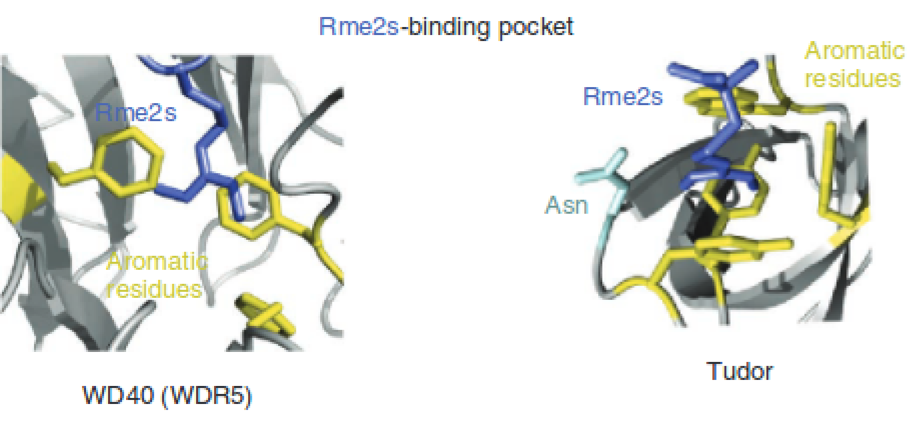
H3K9me, H3K27me } gene-silencing marks

Arginine

H3R17me2a, H4R3me2a, H3R2me2s } gene-activation marks

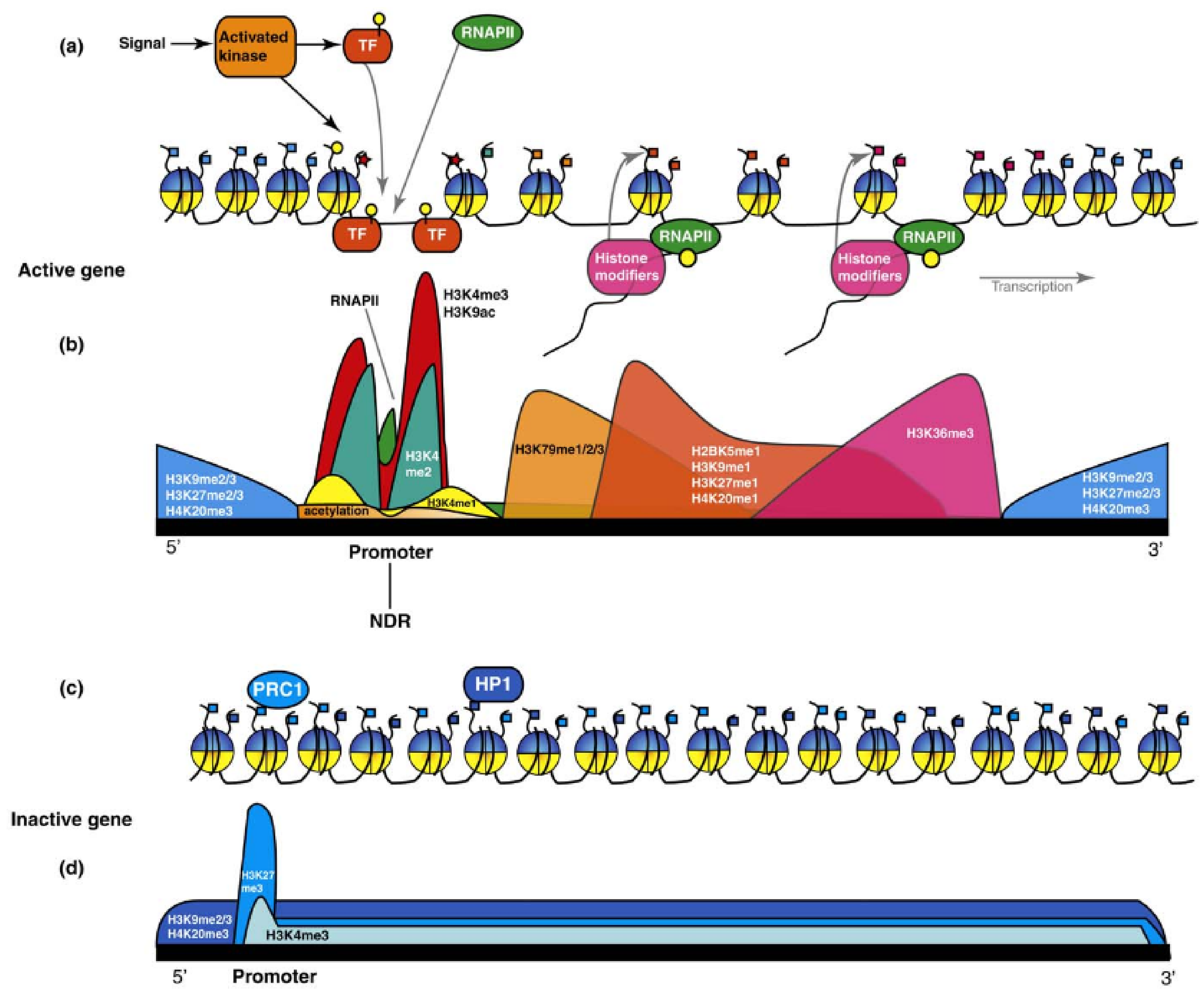
H4R3me2s } gene-silencing marks





Musselman et al. 2012 Nat Struct Mol Biol. 19(12):1218-27

Methylation and acetylation control DNA availability



Barth Imhof 2010 Trends Bioch Sci. 35(11): 618-626

The wave of chromatin condensation

A writer enzyme is recruited at a specific site by a transcription regulatory protein.

It creates a specific modification on one or more of the four nucleosomal histones.

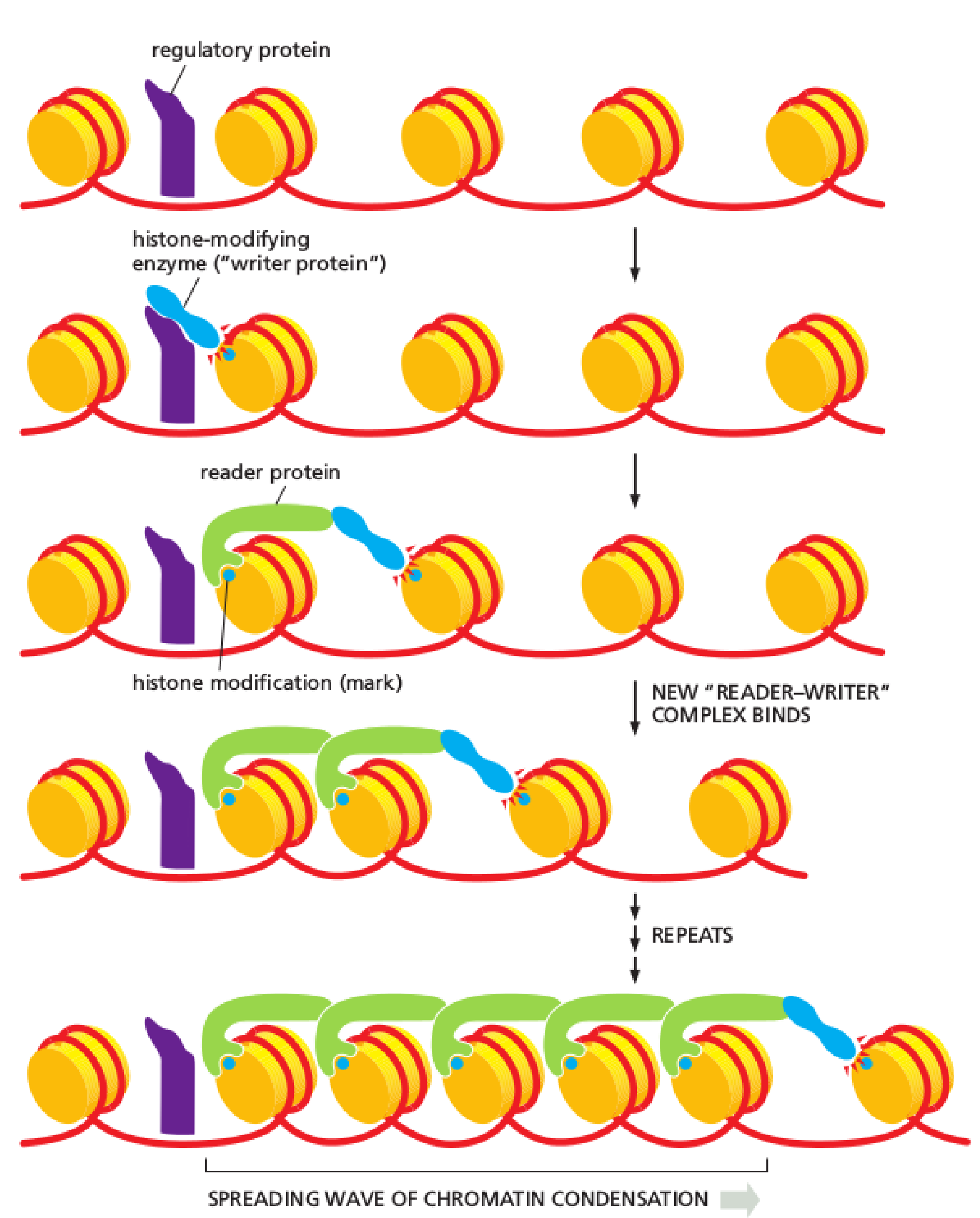
A reader protein reads the modification brought by the writer protein, binds and activate its binding activity for the writer.

The writer is moved to the next histone by ATP-dependent chromatin remodeling Complex, that repositions the histone.

Once there, it produces the next modification.

This mechanism can rapidly condense chromatin, leading to a complete inactivation of the whole region.

A reverse effect can be achieved if the reader recruits the eraser...



Silencing via H3K9 methylation and HP1 cascade

The classical mechanism of HP1 (heterochromain protein 1) mediated heterochromatin formation is as follows:

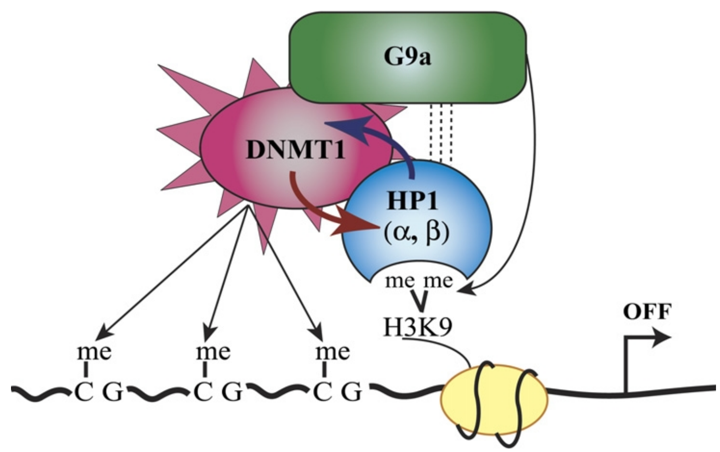
1. H3 gets methylated at K9, with Me2 or Me3 markers

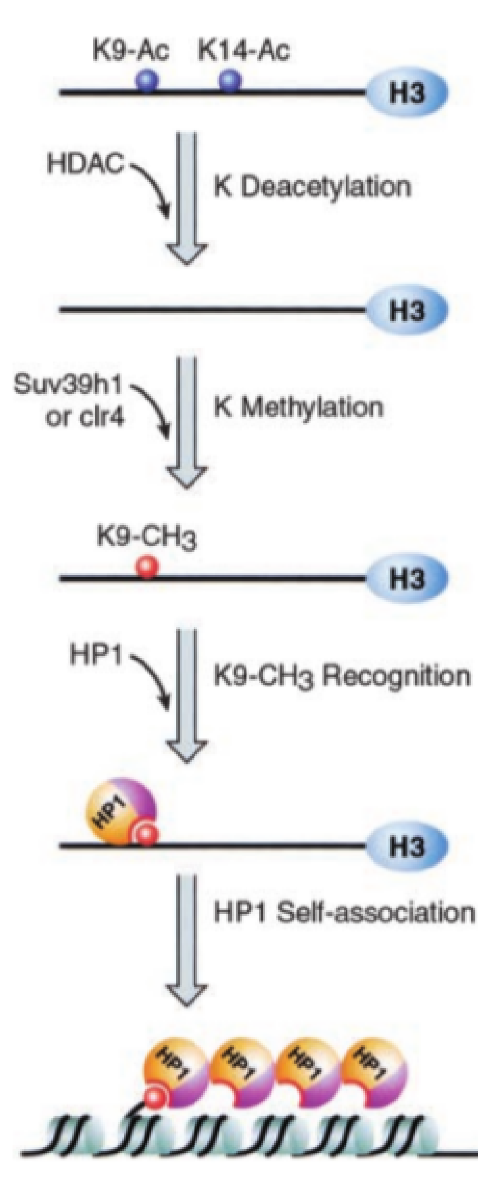
2. H3K9me2/3 recruits HP1

3. H3K9 methyltransferase SUV39H1 is recruited

4. A positive feedback loop is established

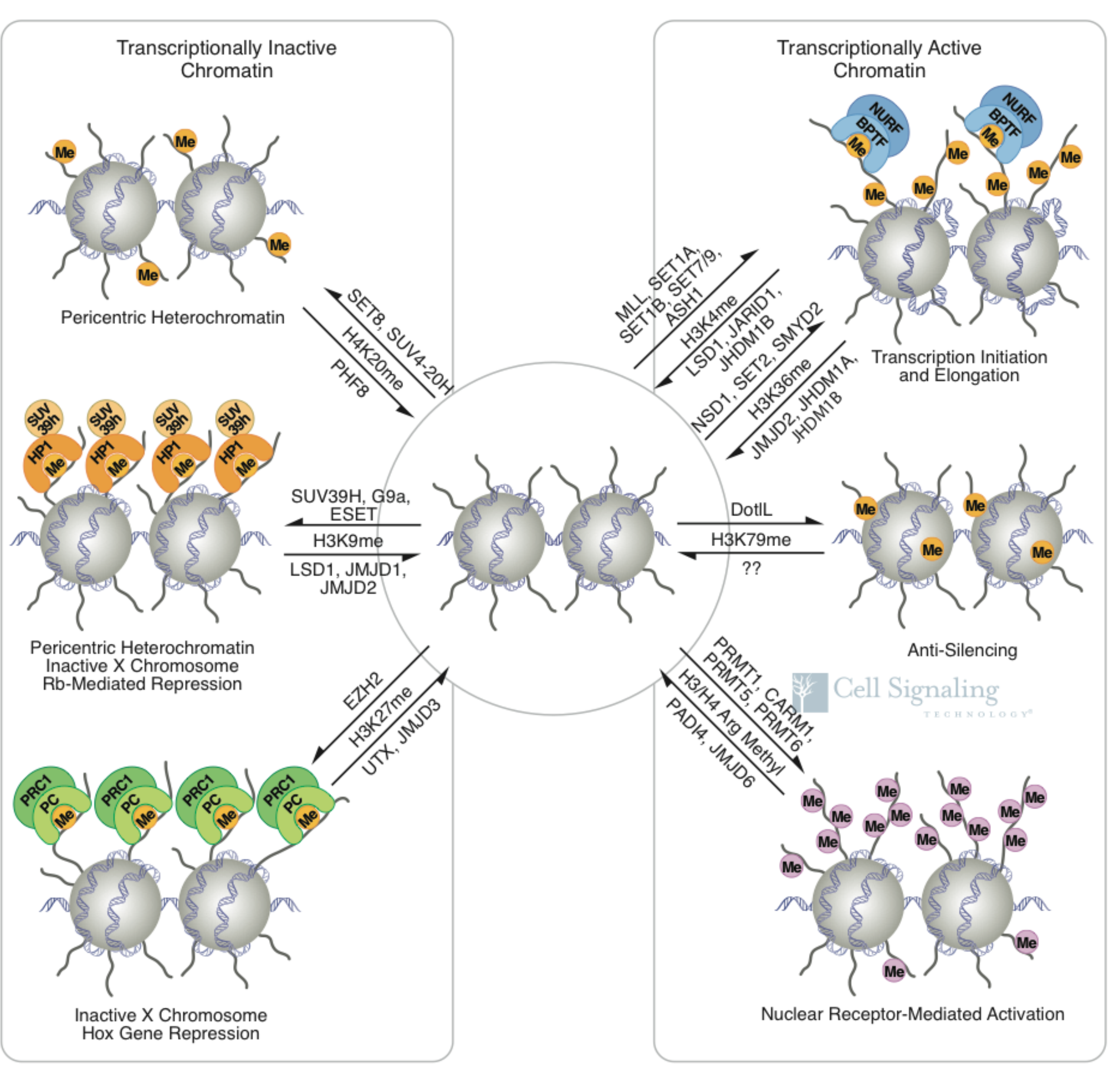
5. The propagation of the H3K9me2/3 mark and HP1 itself then serve to bind additional proteins (e.g. DNMT, methyl CpG BP) leading to heterochromatin formation and gene silencing.





Zhang 2001 Genes Dev. 15(18):2343-60

Several mechanisms cooperate



Reading phosphorylations

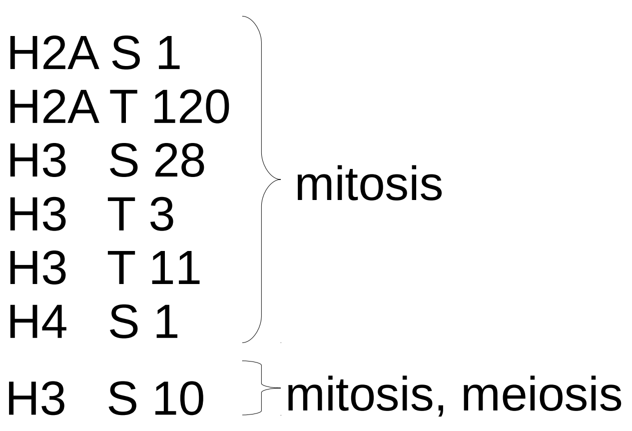
Phosphorylation of Serine and Threonine introduces a negative charge close to the DNA, so chromatin destabilization and opening would seem logical, and it is true. In addition, P-ylation serves as **a recruitment scaffold** for other binding proteins and relates to critical intermediate step in chromosome condensation during cell division, transcriptional regulation, and DNA damage repair.

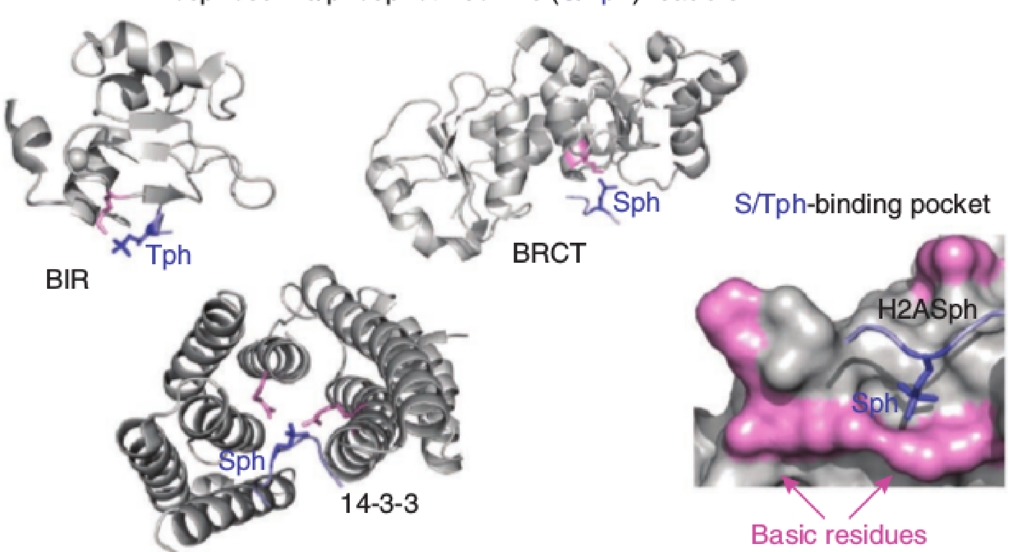
H2A T 142
H2A S 139
H2B S 14
H4 S 1
} DNA repair

H2AX S 139 } DNA ds breaks

H3 S 28
H3 S 10
} Immediate-early activation

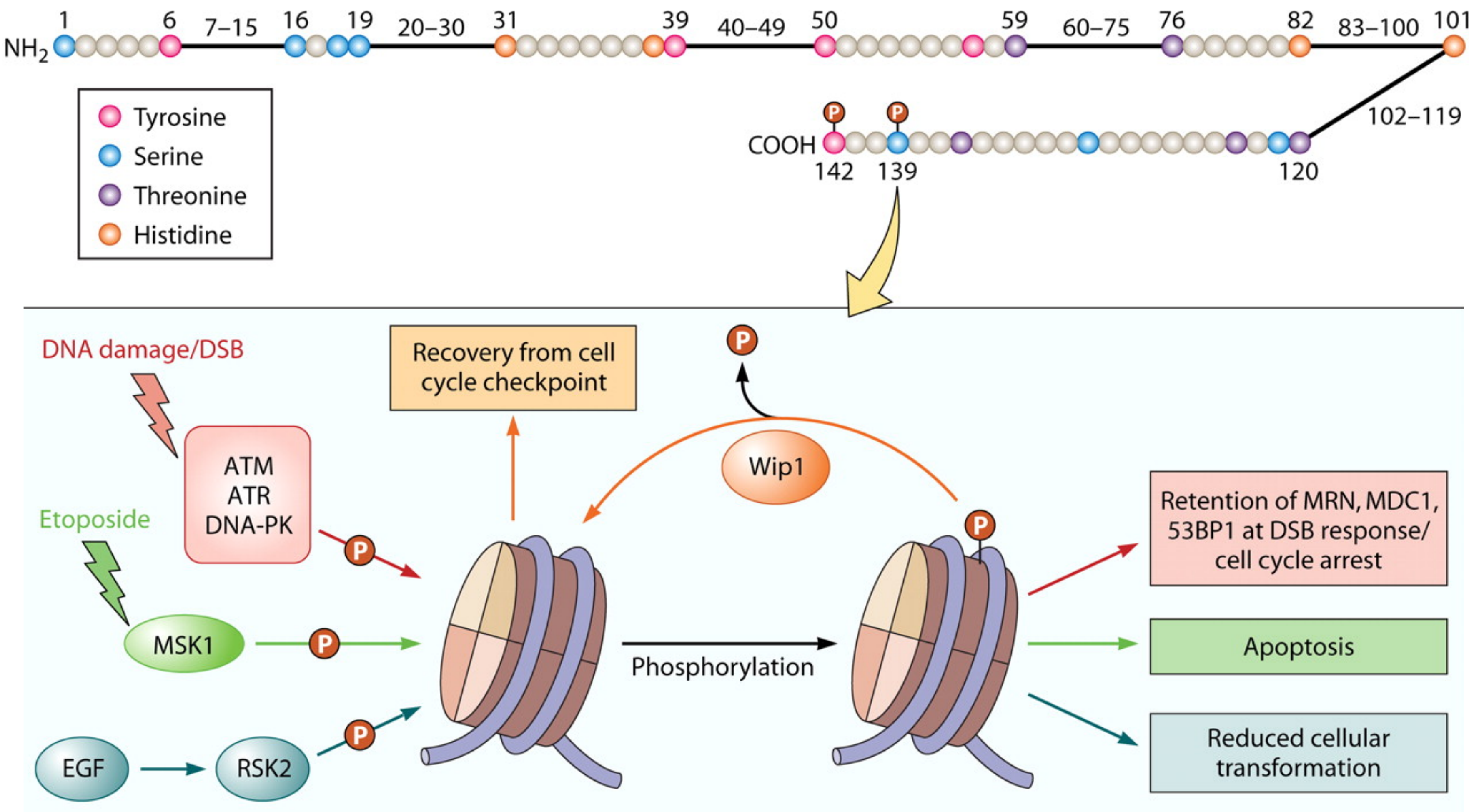
H1 S 27
H2B S 36
H3 S 10
H3 Y 41
} transcriptional activation





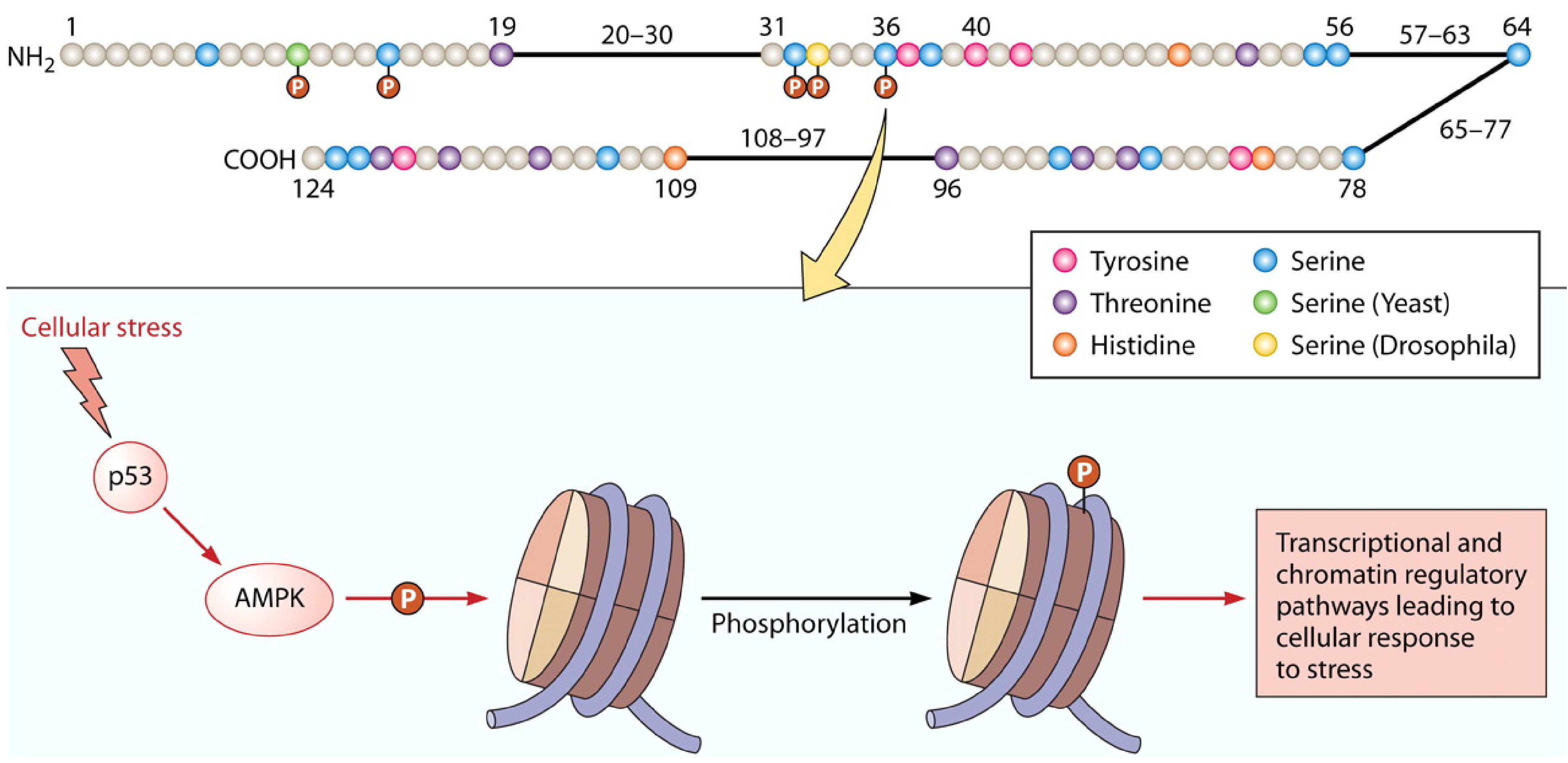
Musselman et al. 2012 Nat Struct Mol Biol. 19(12):1218-27

Role of H2AX S139 phosphorylation



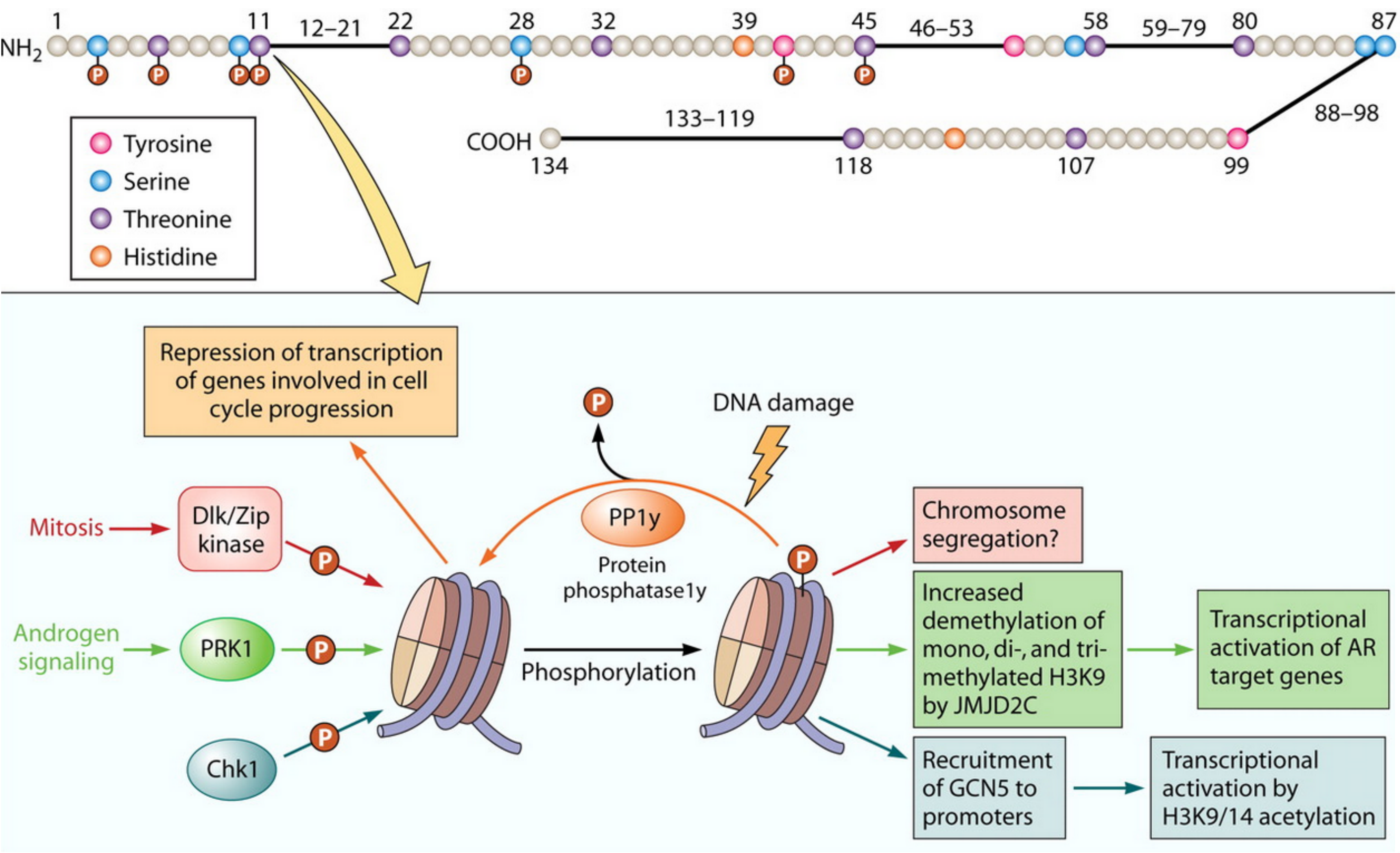
Banerjee and Chakravarti 2011 Mol Cell Biol. 31(24):485873

Role of H2B S36 phosphorylation



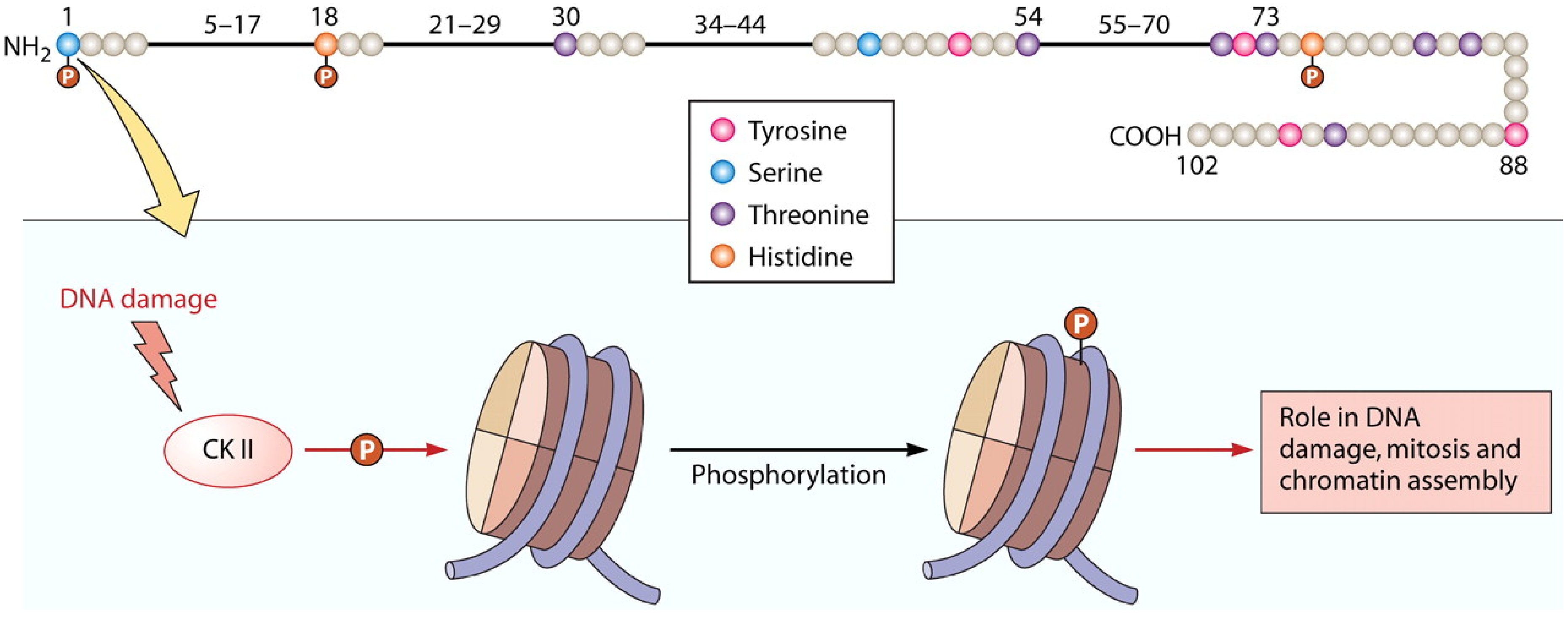
Banerjee and Chakravarti 2011 Mol Cell Biol. 31(24):485873

Role of H3 T11 phosphorylation



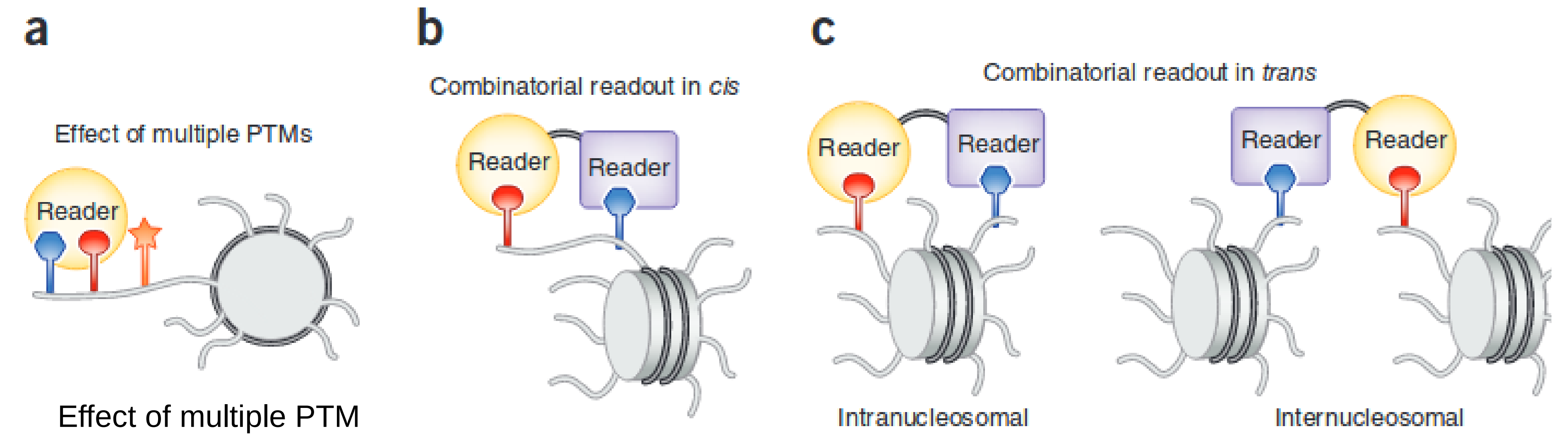
Banerjee and Chakravarti 2011 Mol Cell Biol. 31(24):485873

Role of H4 S1 phosphorylation

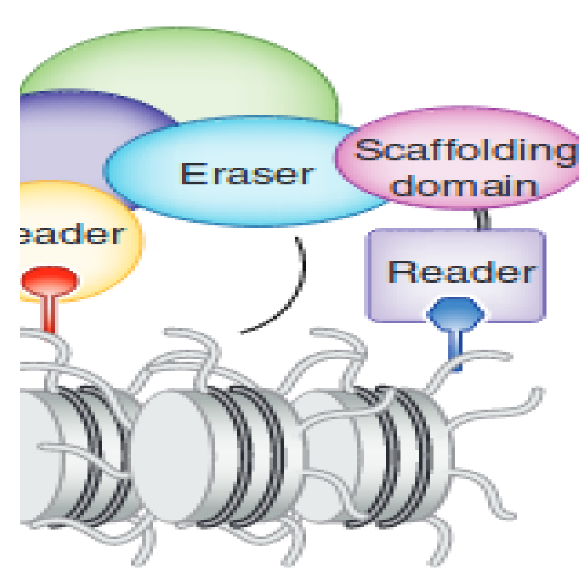


Banerjee and Chakravarti 2011 Mol Cell Biol. 31(24):485873

Histone code is a PTM combinatorial system



Effect of multiple PTM



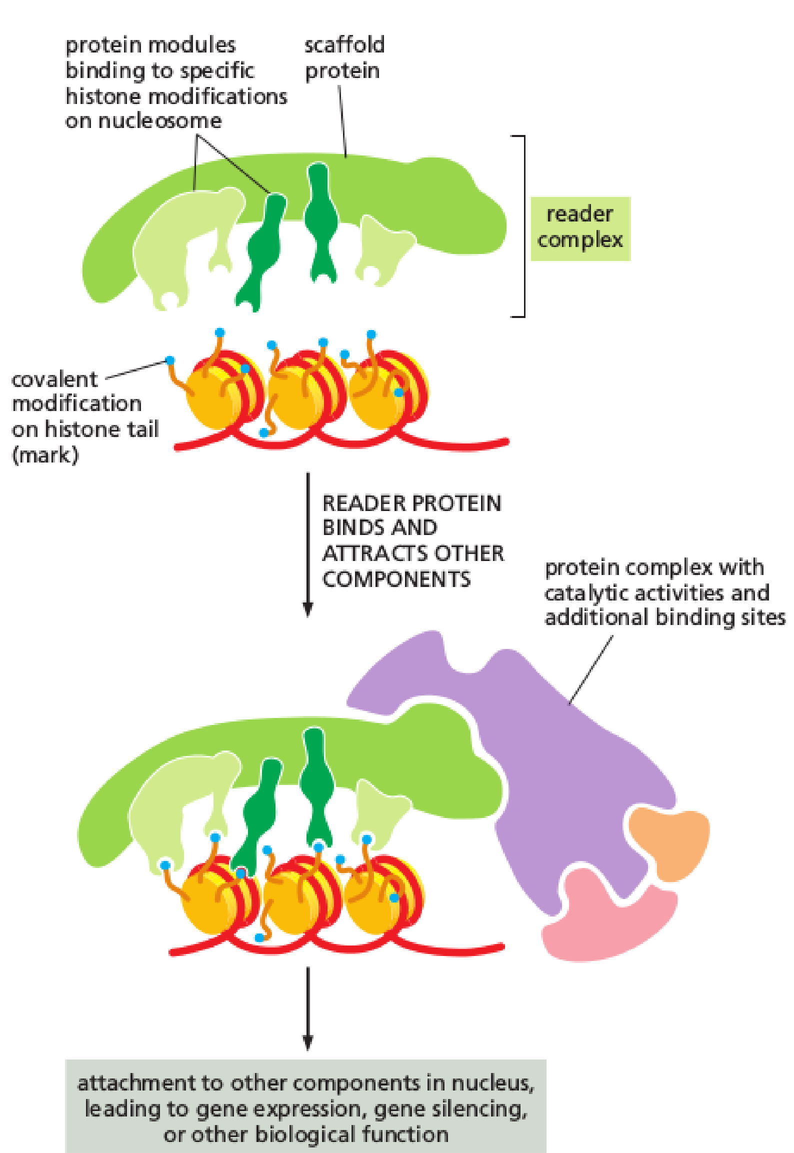
Musselman et al. 2012 Nat Struct Mol Biol. 19(12):1218-27

The detection and interpretation of PTM combinations

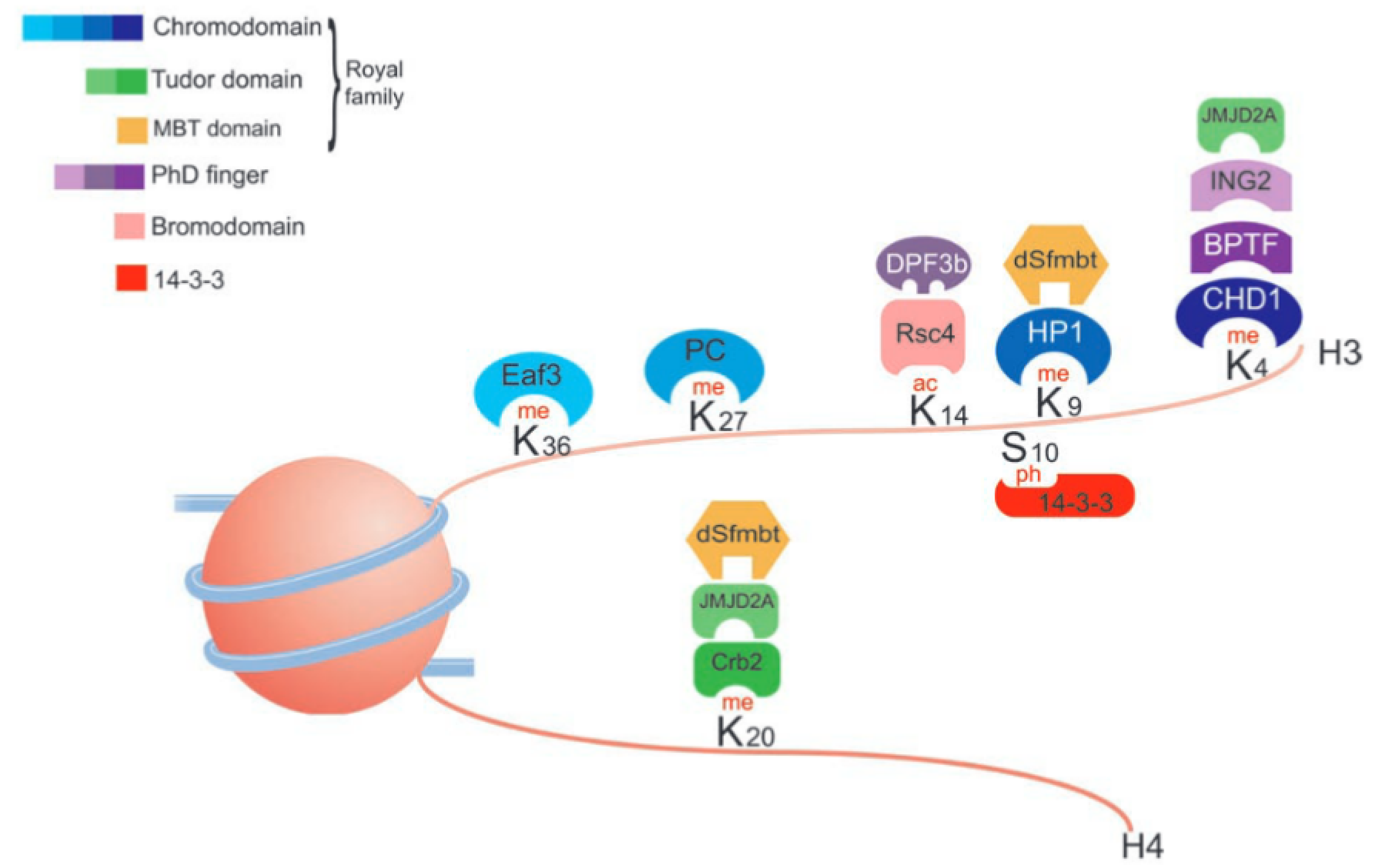
A large protein complex that contains a series of protein modules, each of which recognizes a specific histone mark: this is a good definition of a “reader complex”.

This complex will bind tightly only to a region of chromatin that contains several of the different histone marks that it recognizes.

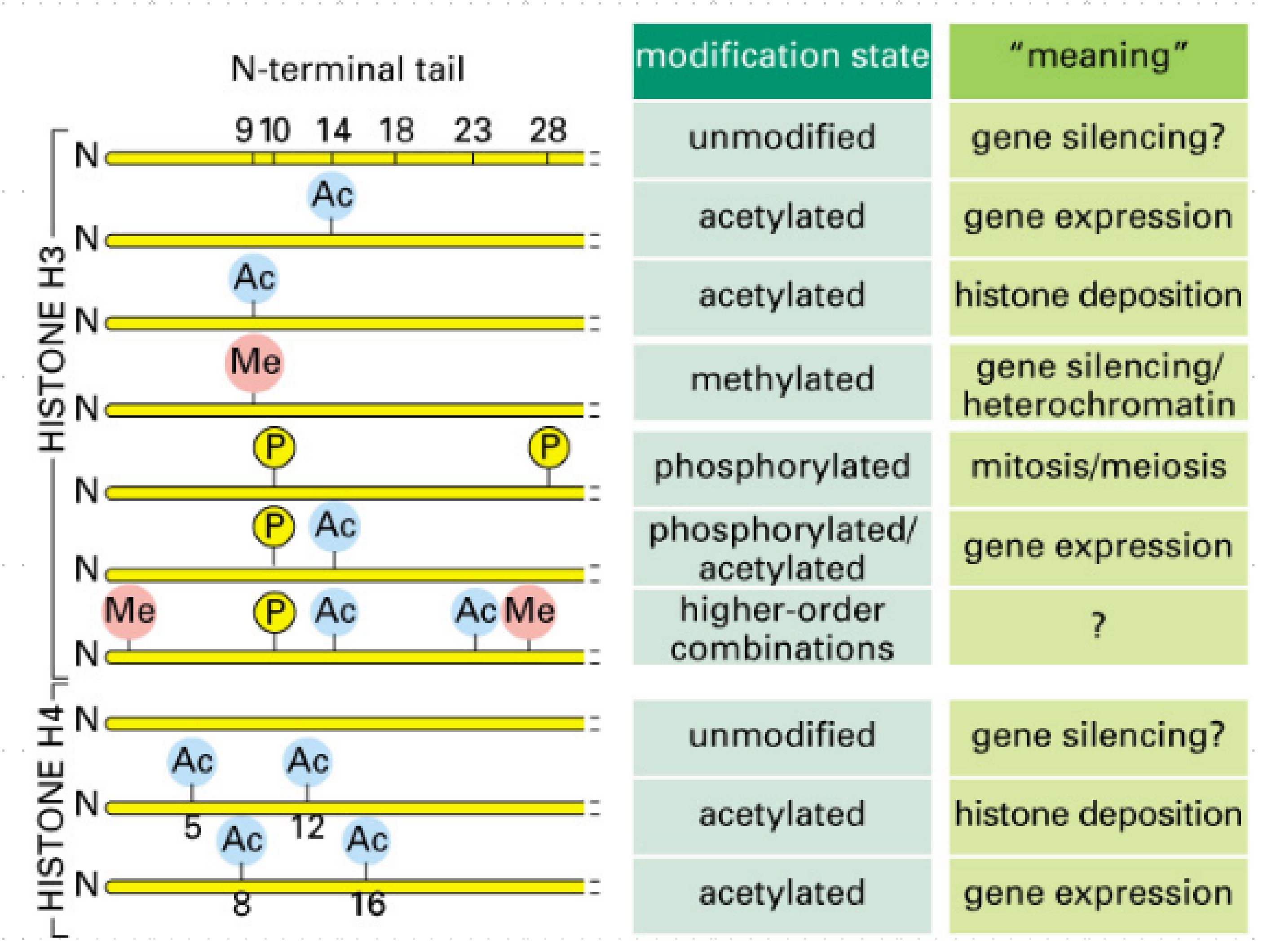
Therefore, only a specific combination of marks will cause the complex to bind to chromatin and attract the additional protein complexes needed to catalyze a biological function.



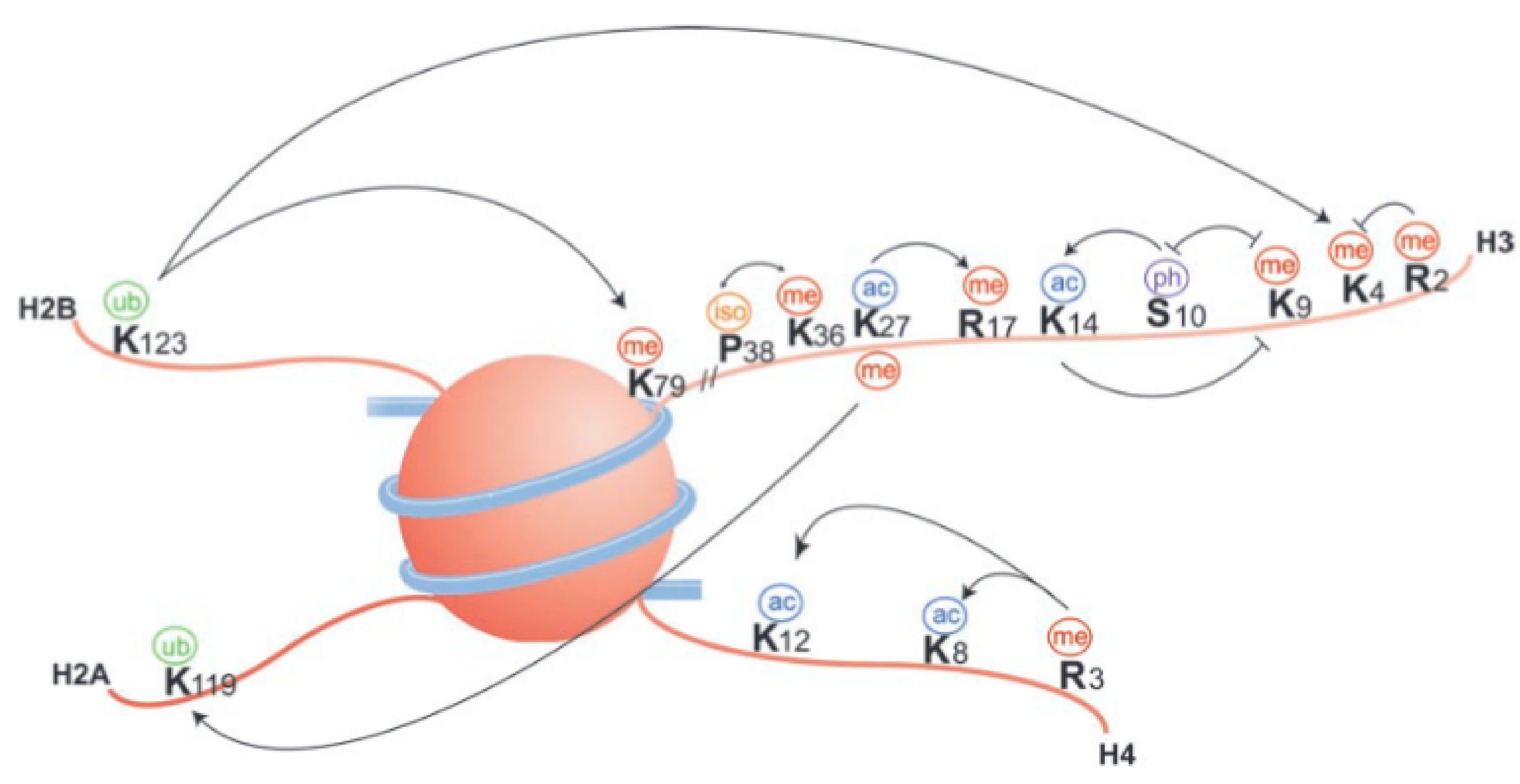
Histone code: an example



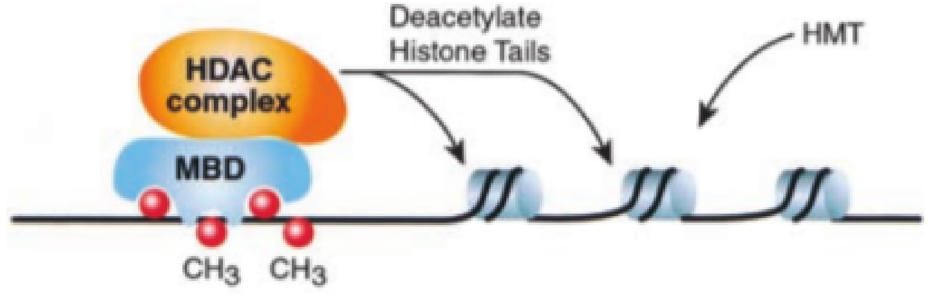
Deciphering the histone code



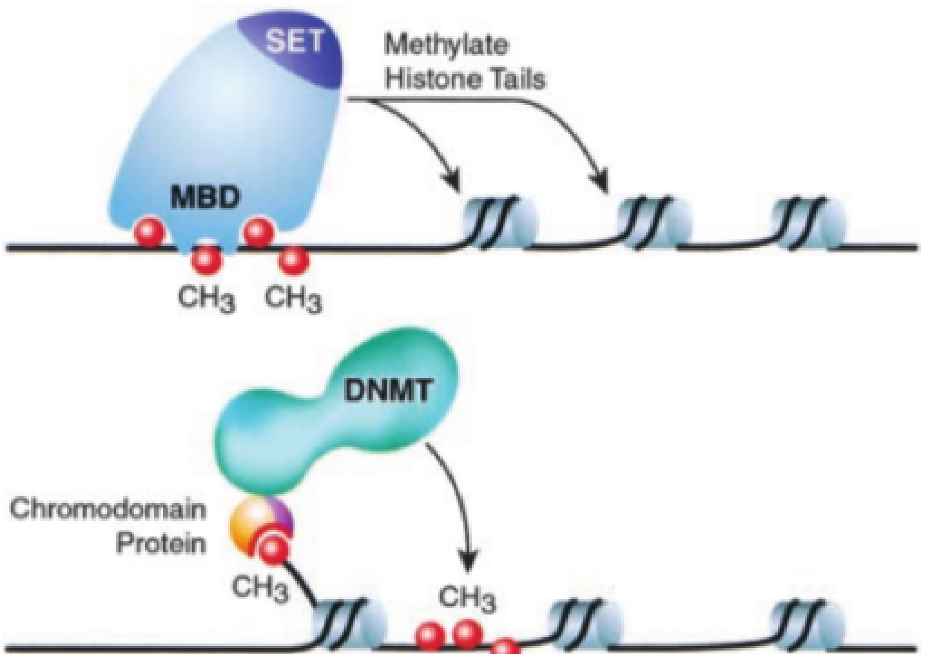
Histone PTM crosstalk



Relationship between histone methylation and DNA methylation



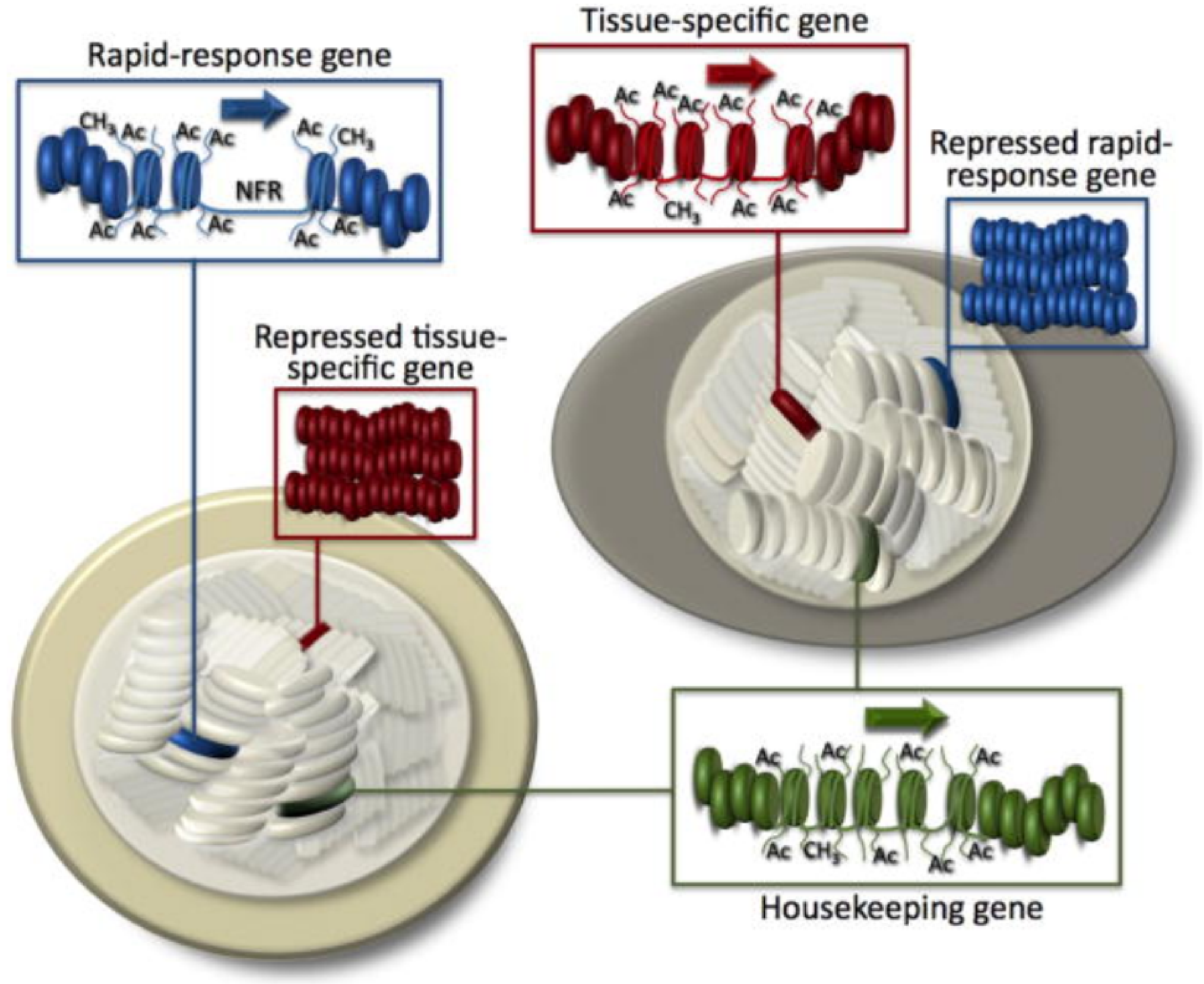
Methyl-CpG-binding proteins (MBD domain) recruit HDAC complex to de-acetylate histones so that their tails will be amenable for subsequent methylation by HMTs.



In chromatin domains where histones are hypoacetylated, the MBD domain-containing HMTs may bind directly and methylate the histones.

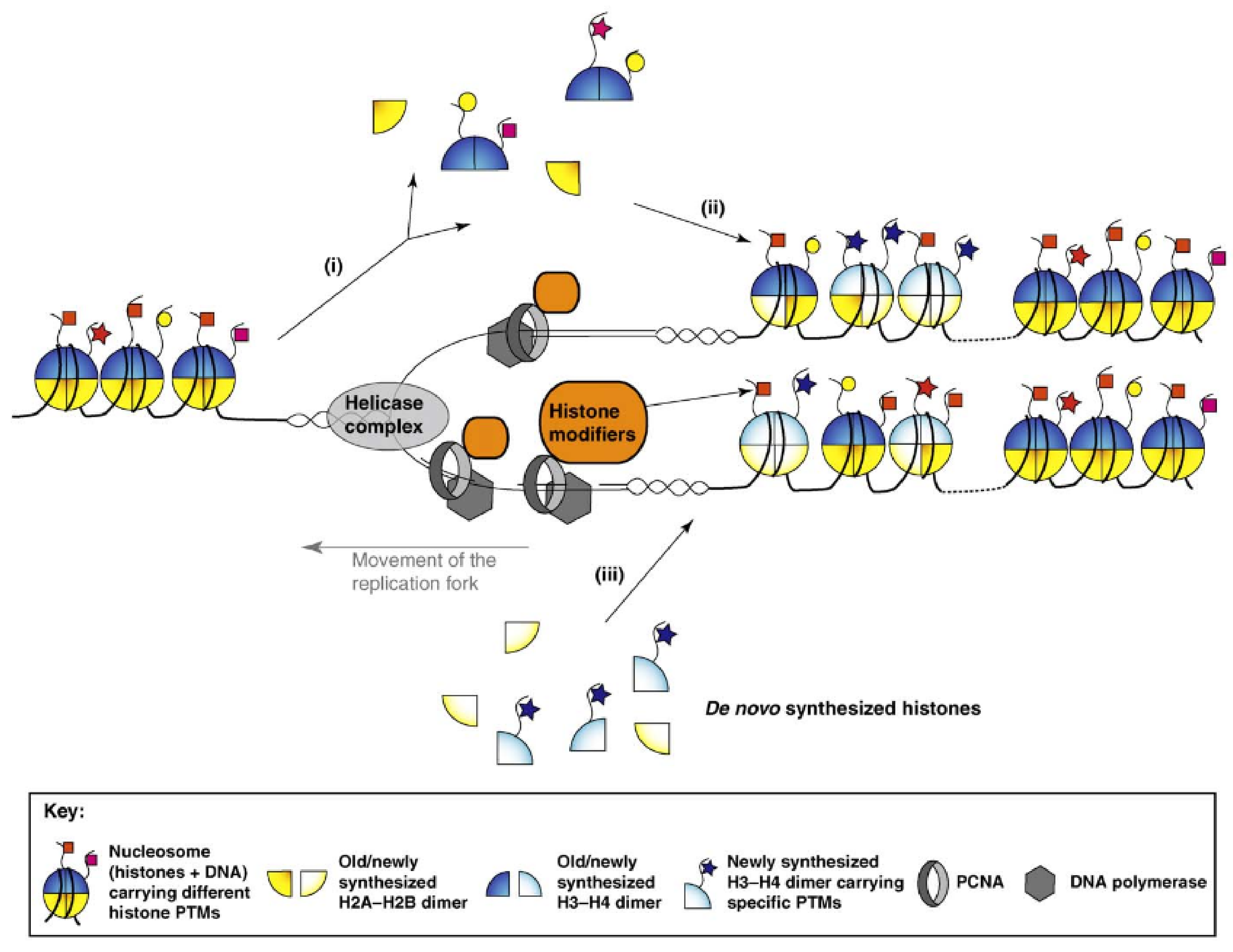
Methylated histone tails may recruit DNMTs to methylate DNA for long-term gene silencing.

Maintaining phenotype and commitment



Histones modification memory after DNA replication

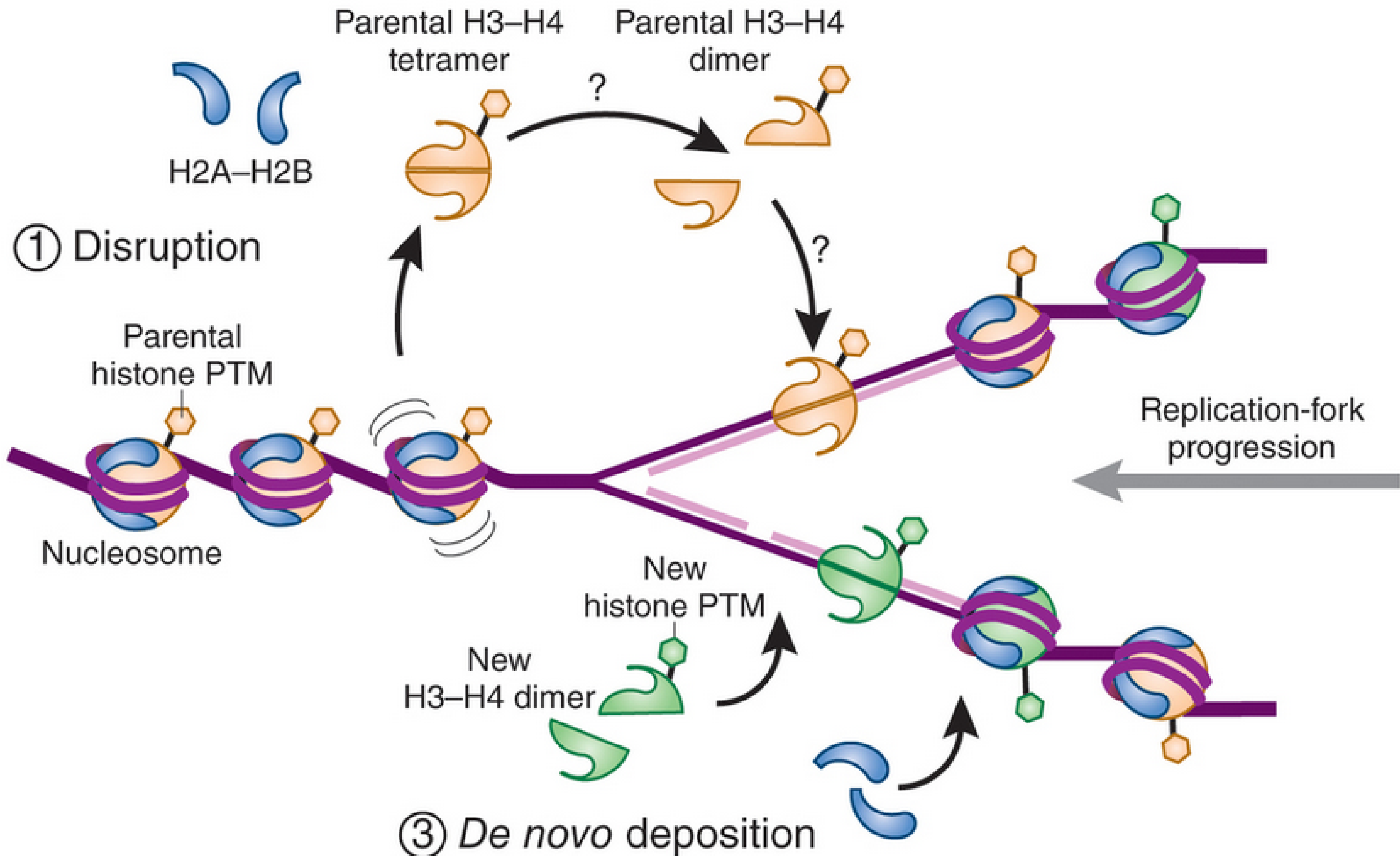
During DNA replication (modified) histones disassemble to favor accessibility, and during the S phase a mass of new histones is produced. New and old histones mix and assemble just after the replication fork, and this proximity ensure a track record of the previous histone modification status.



Barth Imhof 2010 Trends Bioch Sci. 35(11): 618-626

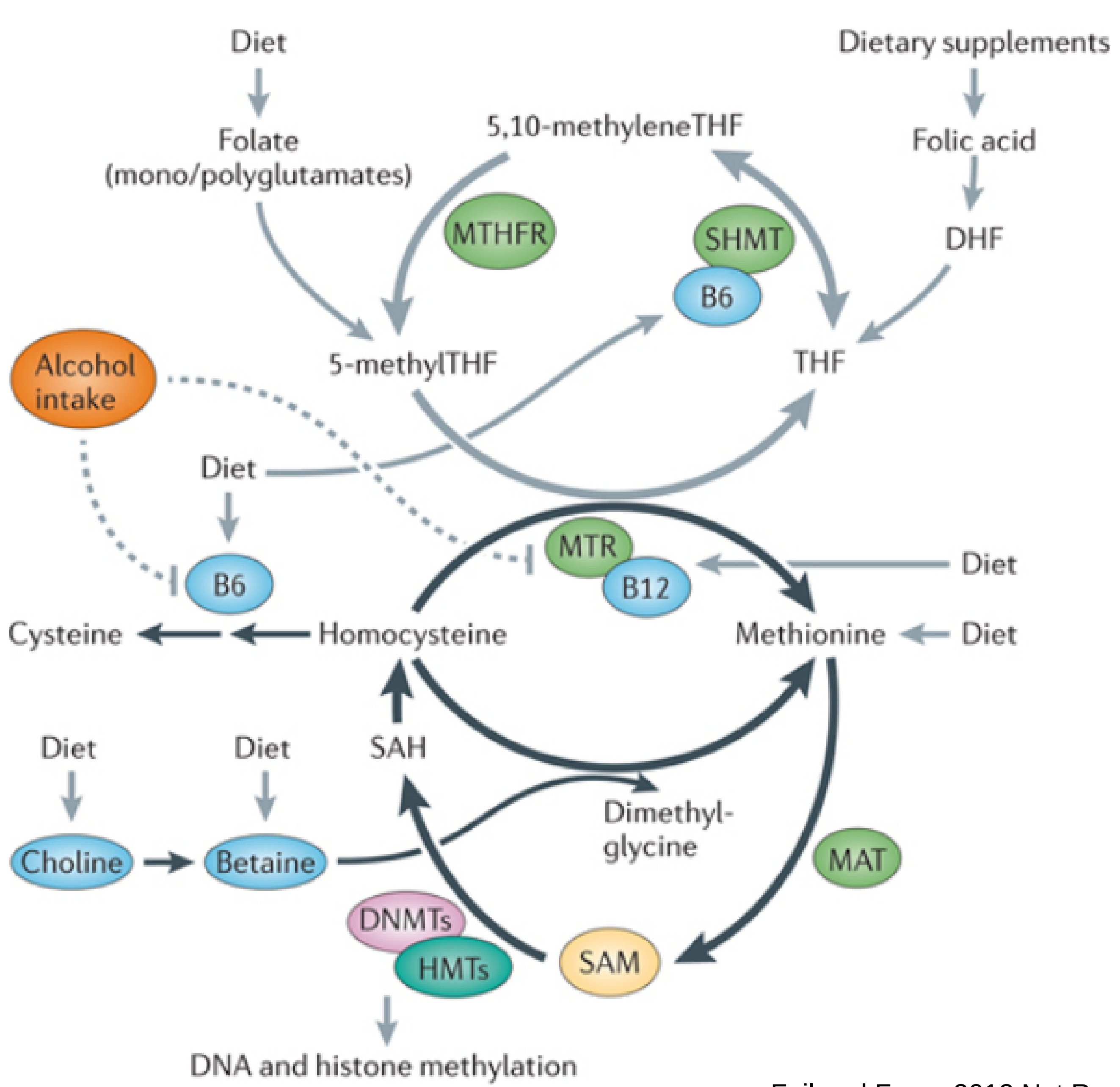
Histones and Okazaki fragments

The lagging strand is replicated by the Okazaki process. Okazaki fragment size is highly correlated with histone wrapping of DNA. Accordingly, replication, as well and PTM inheritance by daughter cells, seems to be mediated by chromatin assembly/disassembly.



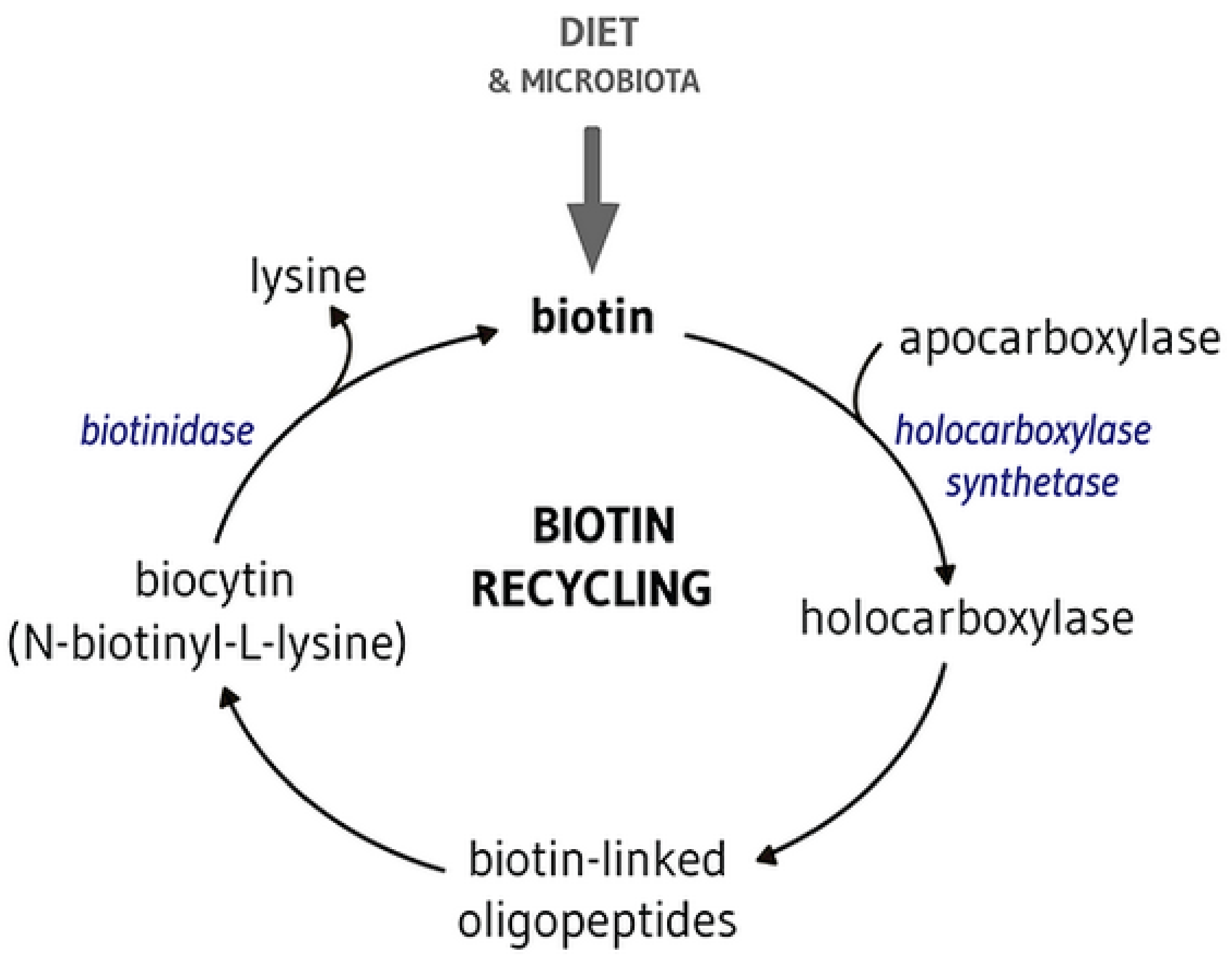
Clement and Almouzni 2015 Nat Struc Mol Biol 22:587–589

Diet and histone modifications



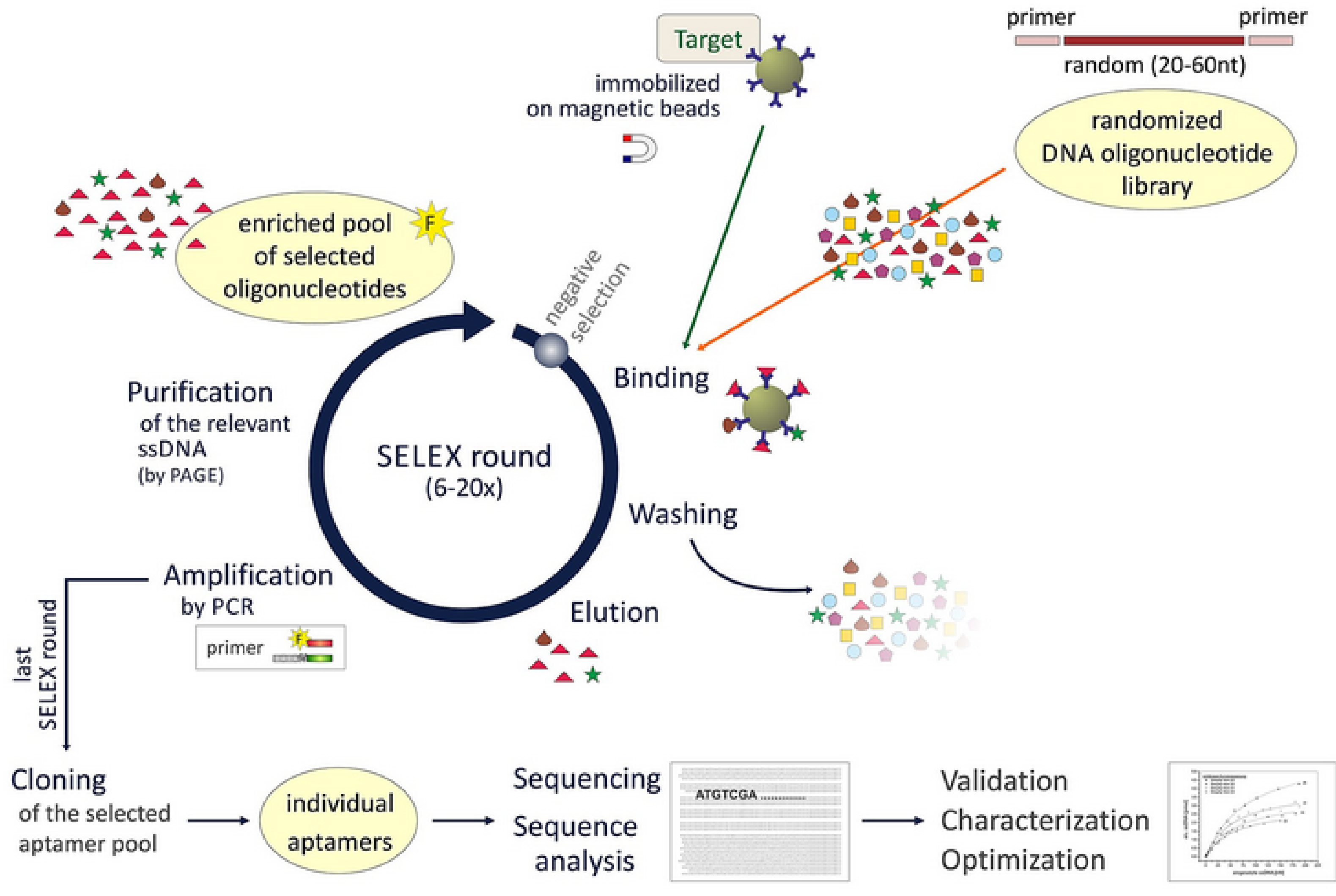
Feil and Fraga 2012 Nat Rev Genetics 13:97-109

Diet and histone modifications



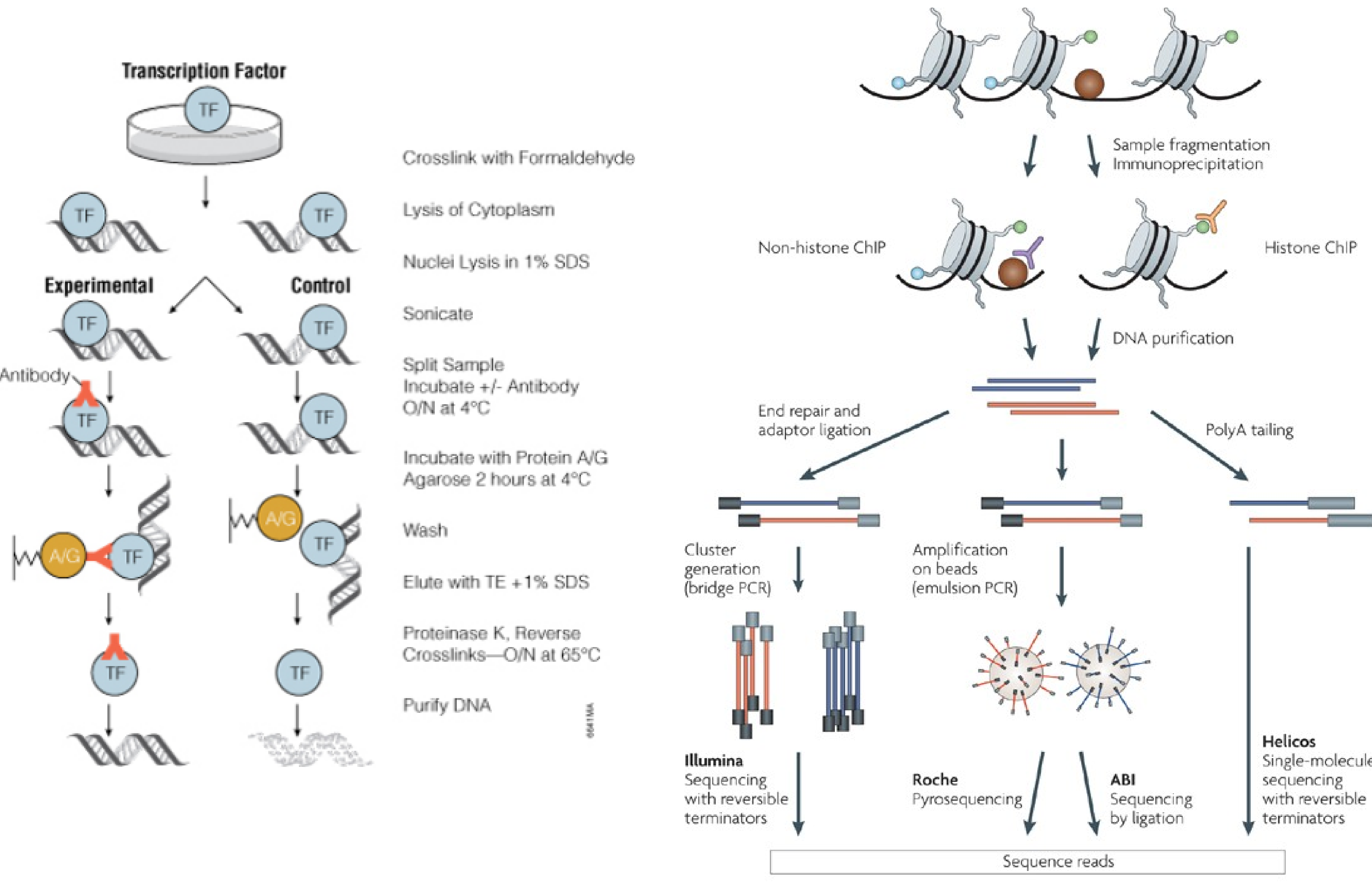
SELEX: identification of binding sites of TFs

The SELEX (Systematic Evolution of Ligands by EXponential enrichment) method has revolutionized the investigation of DNA binding sites of protein binding factors and domains.

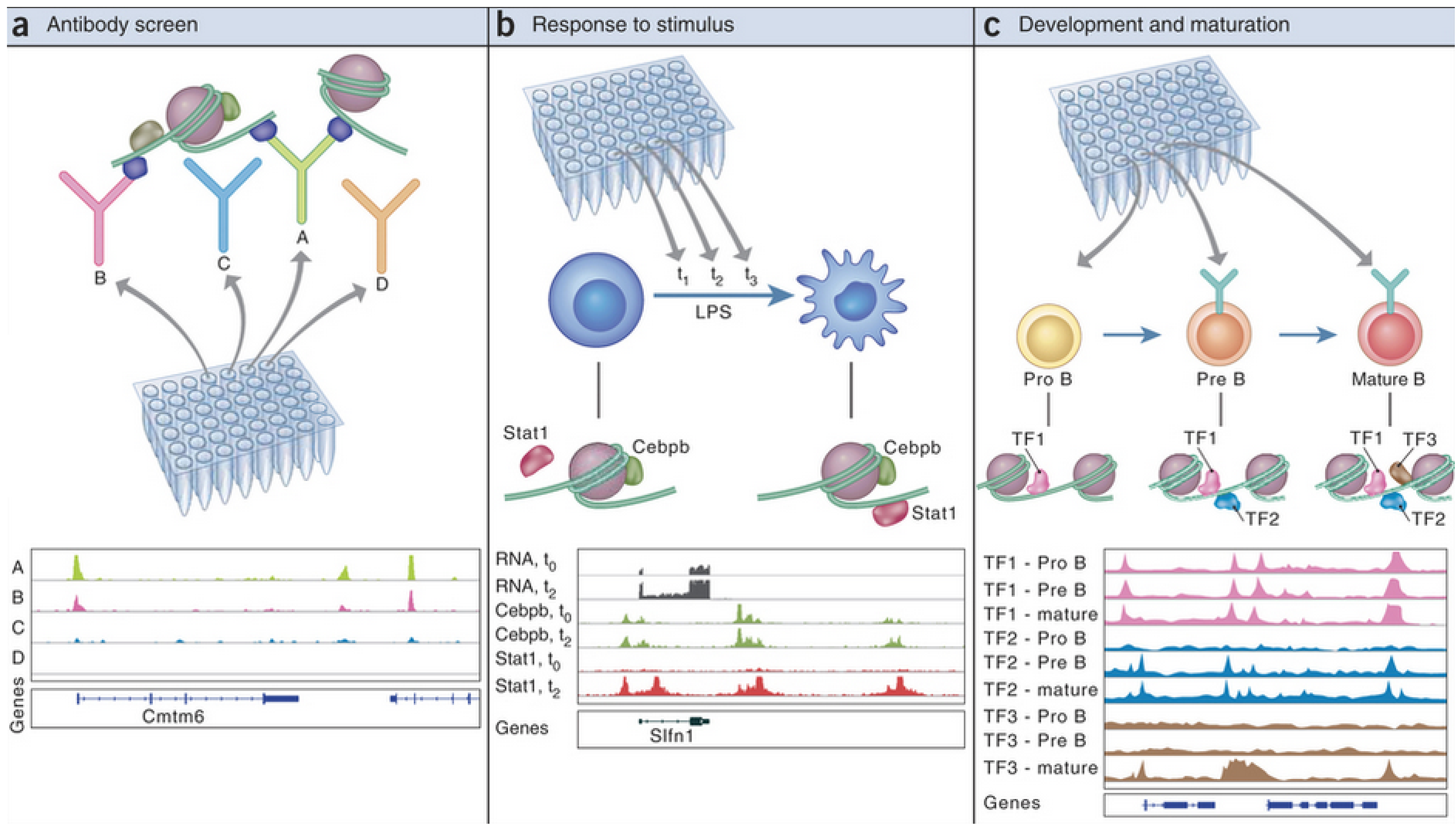


Stoltenburg et al. 2015 PLoS ONE 10(7): e0134403

ChIP and Chip-seq



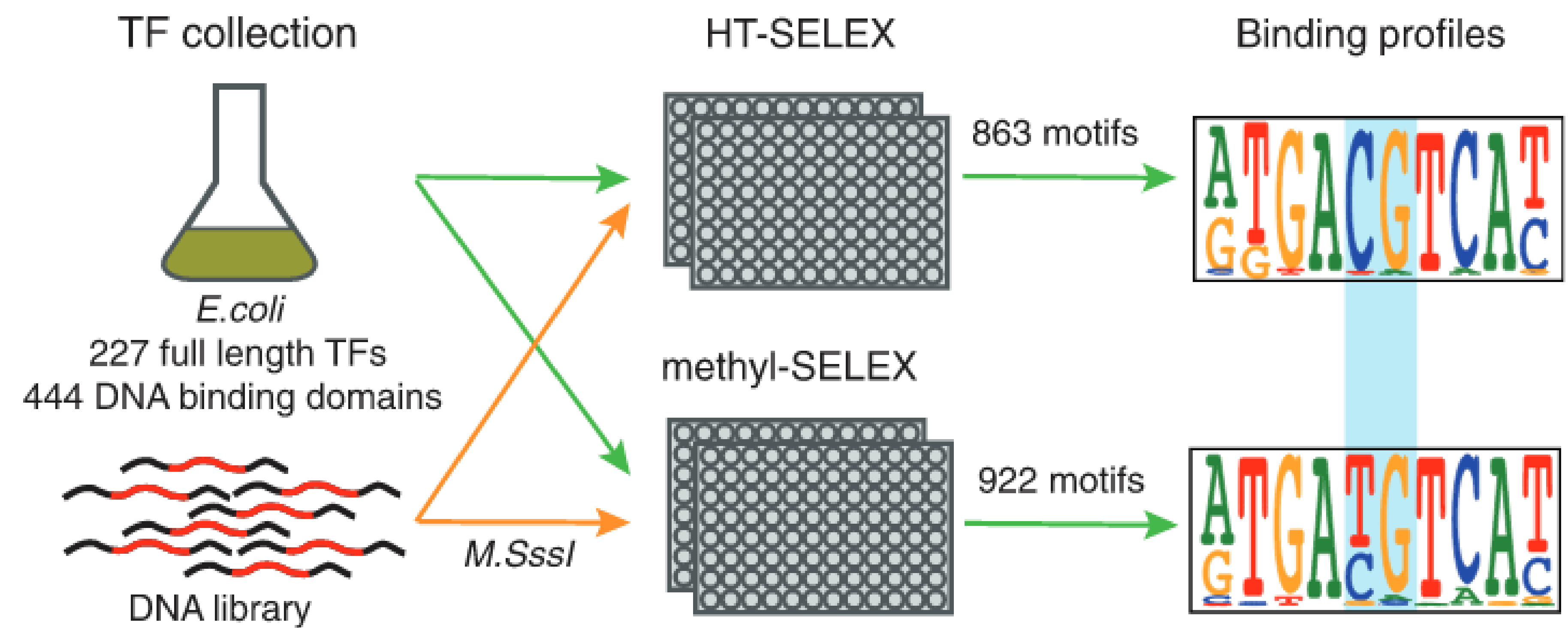
A typical Chip-seq experiment



HT- / Methyl / Bi sulfite SELEX

The SELEX method has recently been modified to monitor the effect of epigenetic changes on binding sites. The input DNA fragment libraries have been prepared with methylated fragments that can be subject to bisulfite treatments (C→U→T via demethylation).

Differential analysis found that epigenetic modification such as DNA-methylation can both inhibit or enhance the binding of specific factors.



Yin et al. 2017 Science 356, 502 (2017)

Array-based identification of protein binding sites

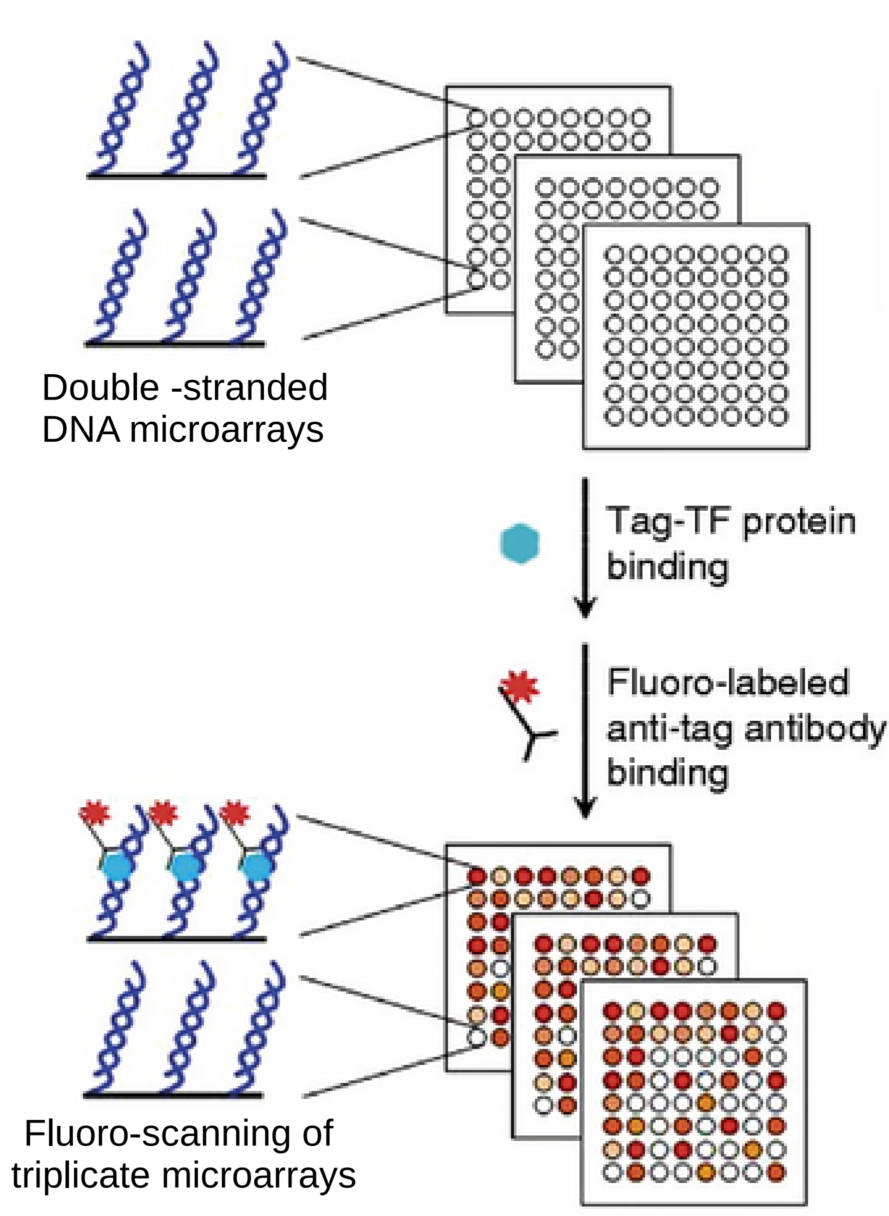
DNA Microarrays also have had a fundamental role in the high-throughput identification of binding sites.

Thousands (in this case, > 40.000) random DNA fragments of 20-40 bp are conjugated on the chip surface with appropriate linkers (spacers).

Tagged binding proteins are layed on the chip and incubated in physiological conditions to allow binding.

A subsequent anti-tag fluorescent antibody is used to highlight spots with bound proteins.

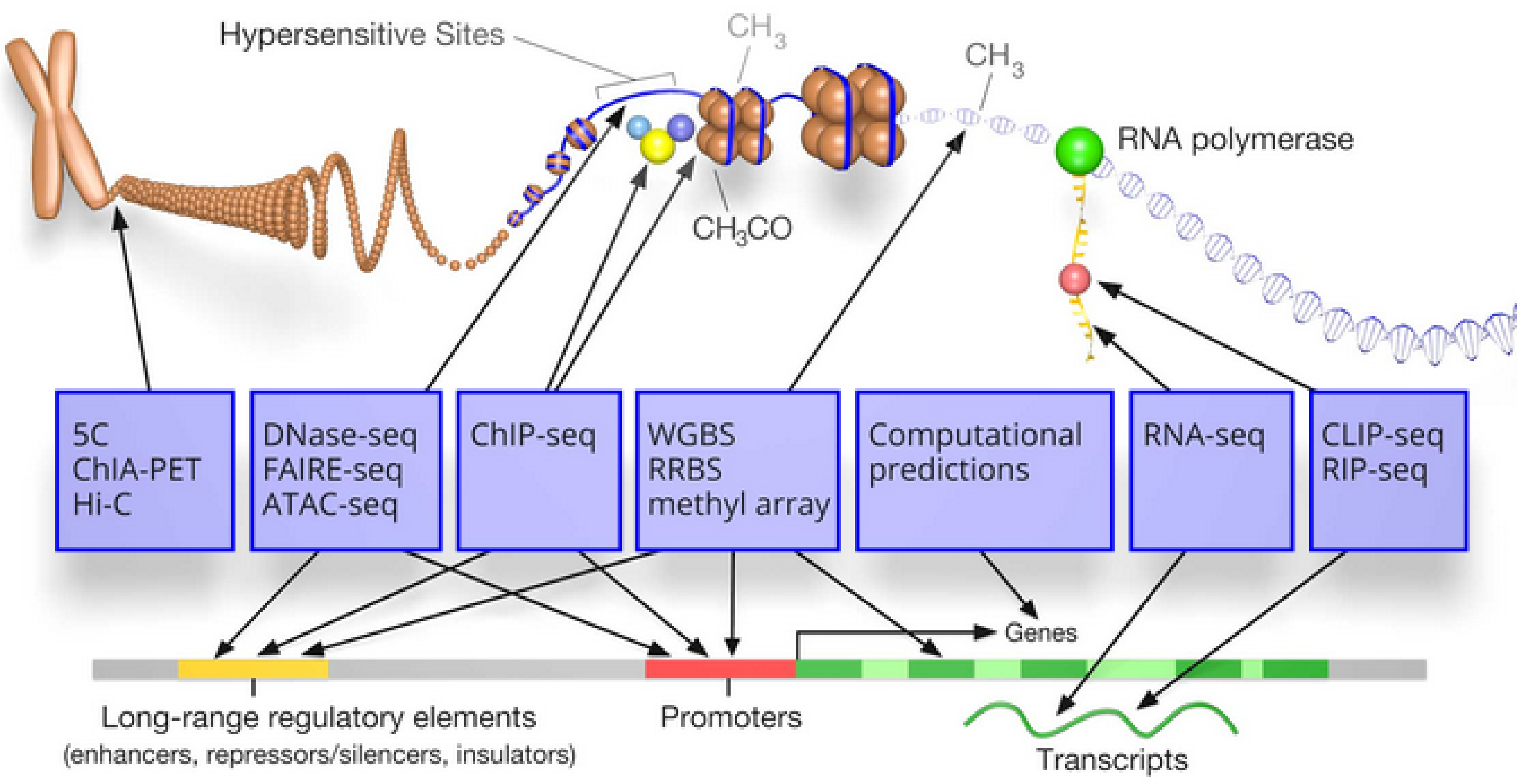
Given that the sequence of DNA in different spots is known, the binding site(s) can be identified and further characterized.



Wang et al. 2011 J Endocrinol 210 15-27

The ENCODE project

The goal of ENCODE (ENcyclopedia Of Dna Elements) is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.



www.encodeproject.org

The AnimalTFDB