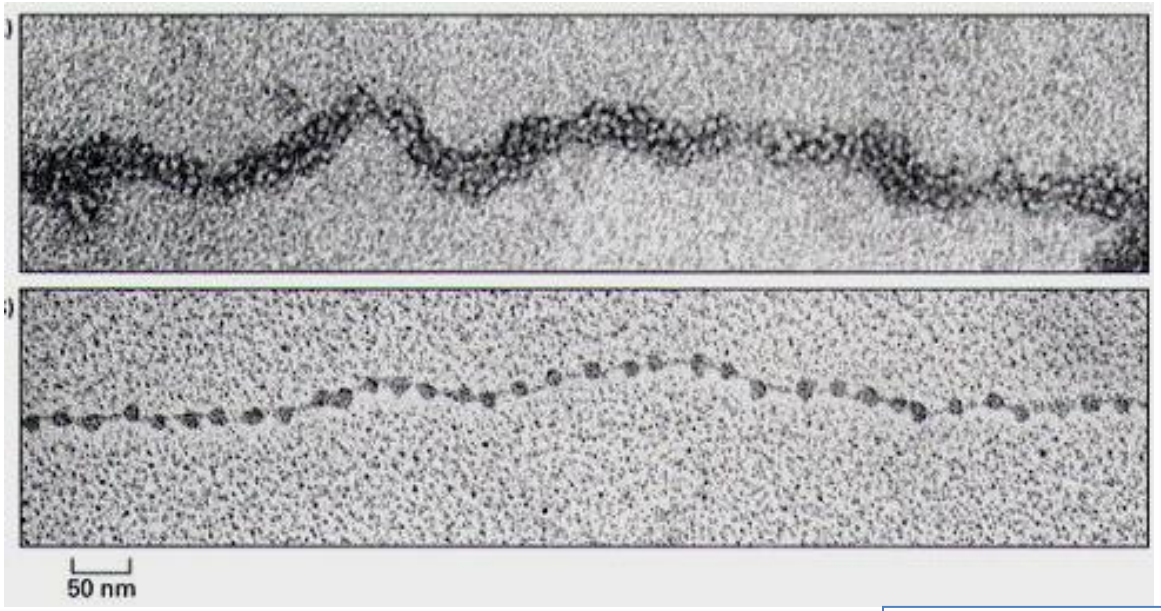
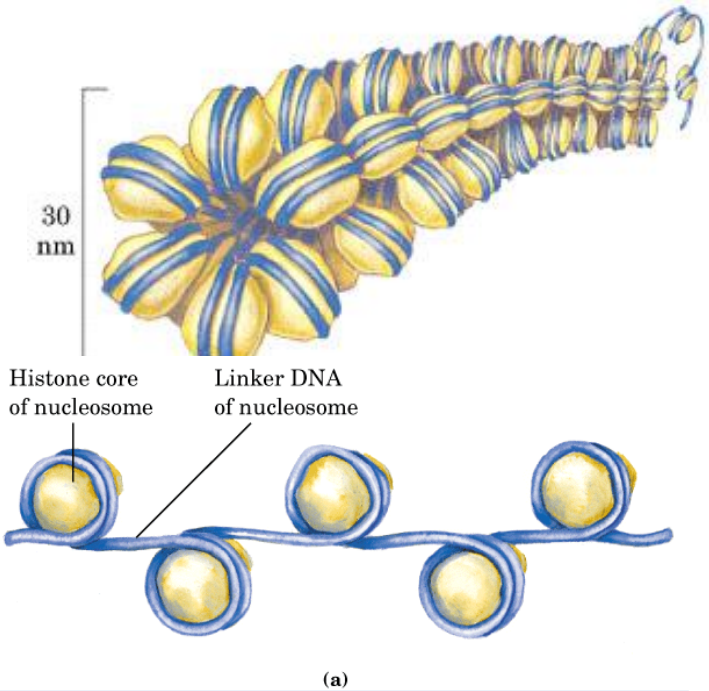
L1.4

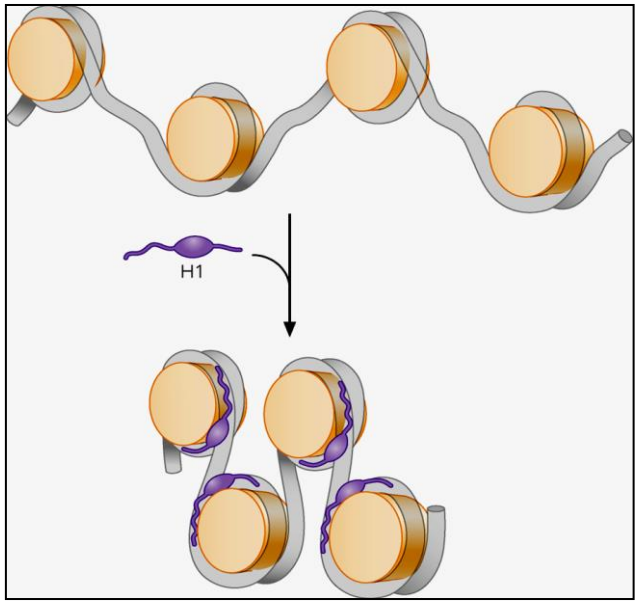
Chromatin Structure

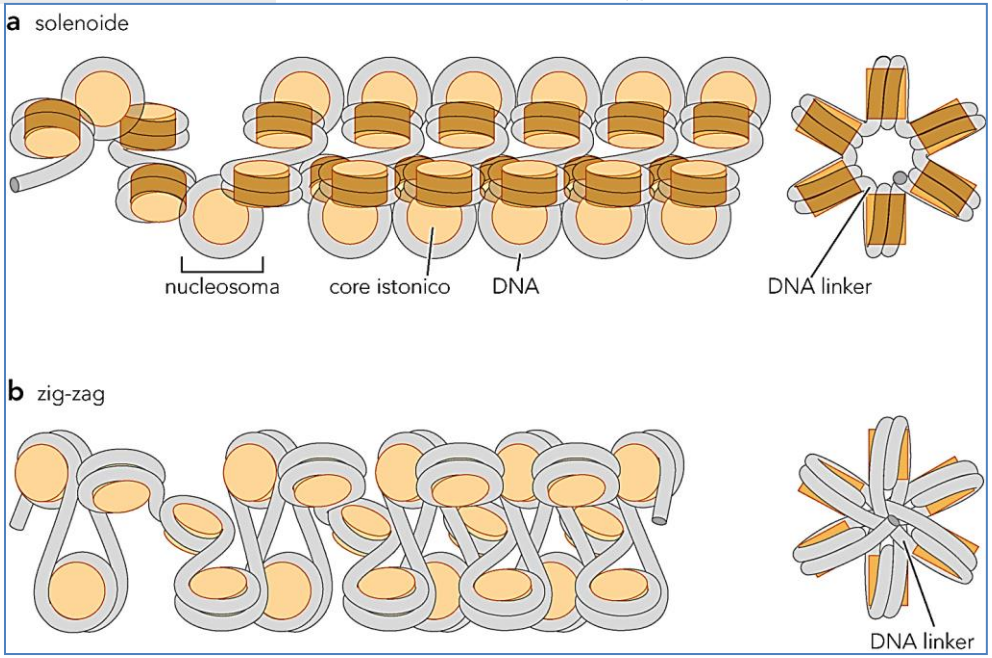
EC *vs* HC

A common, reductive view of chromatin



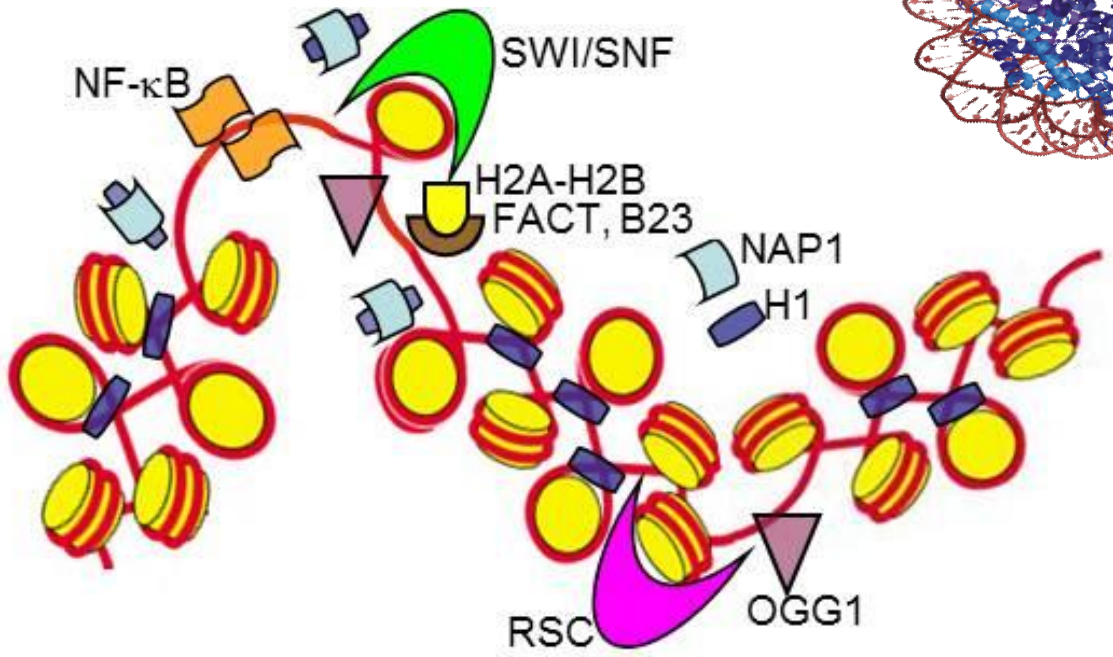


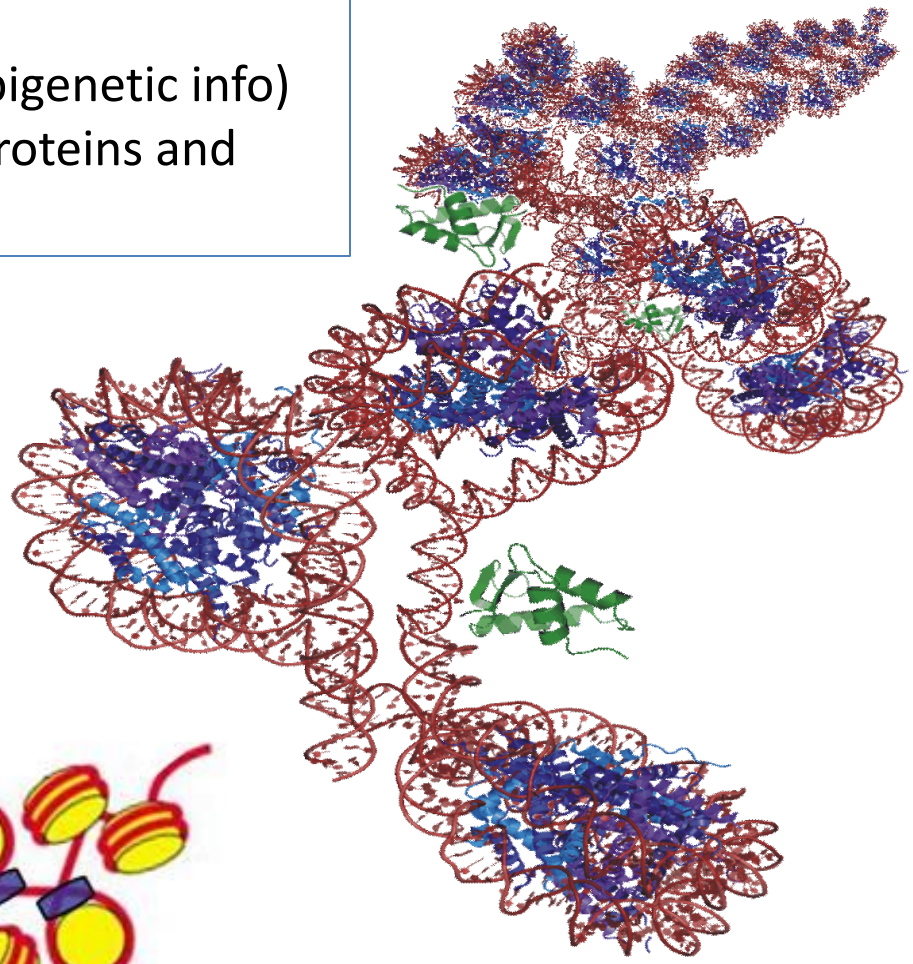




Chromatin is composed of

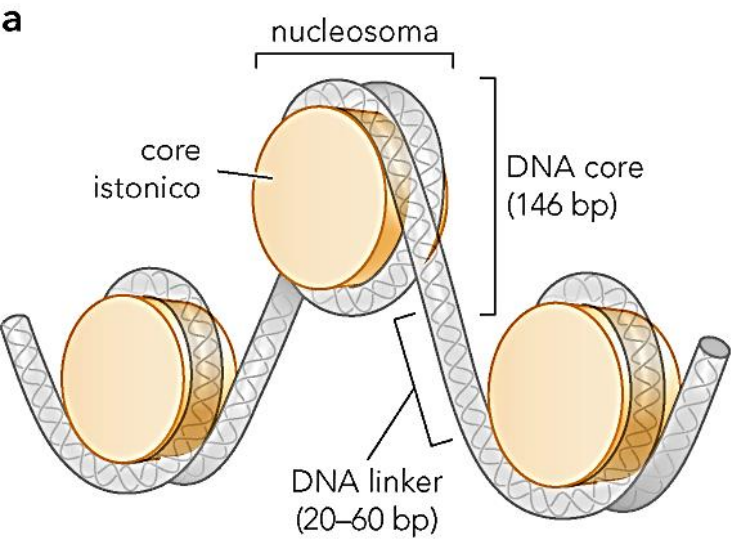
- DNA (genetic information)
- variously modified histone proteins (epigenetic info)
- DNA binding factors, enzymes, other proteins and noncoding RNAs (regulatory).

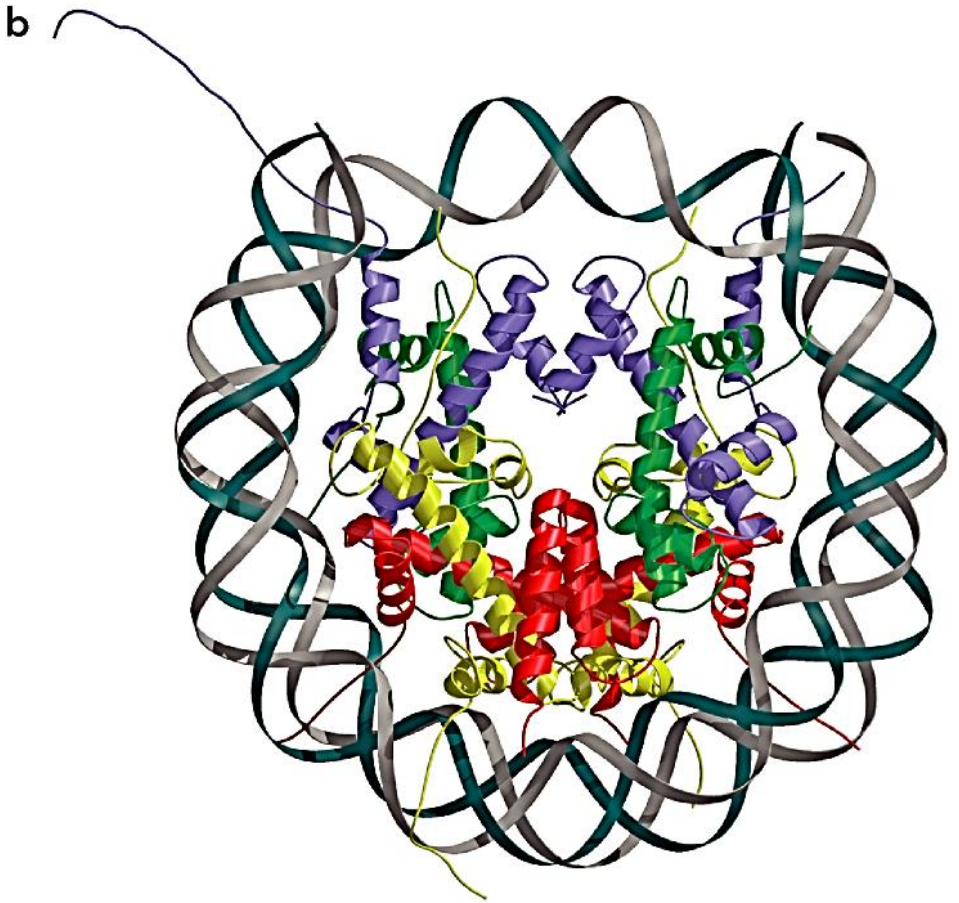




ENS, LYON

(Proteins in this cartoon are just examples)

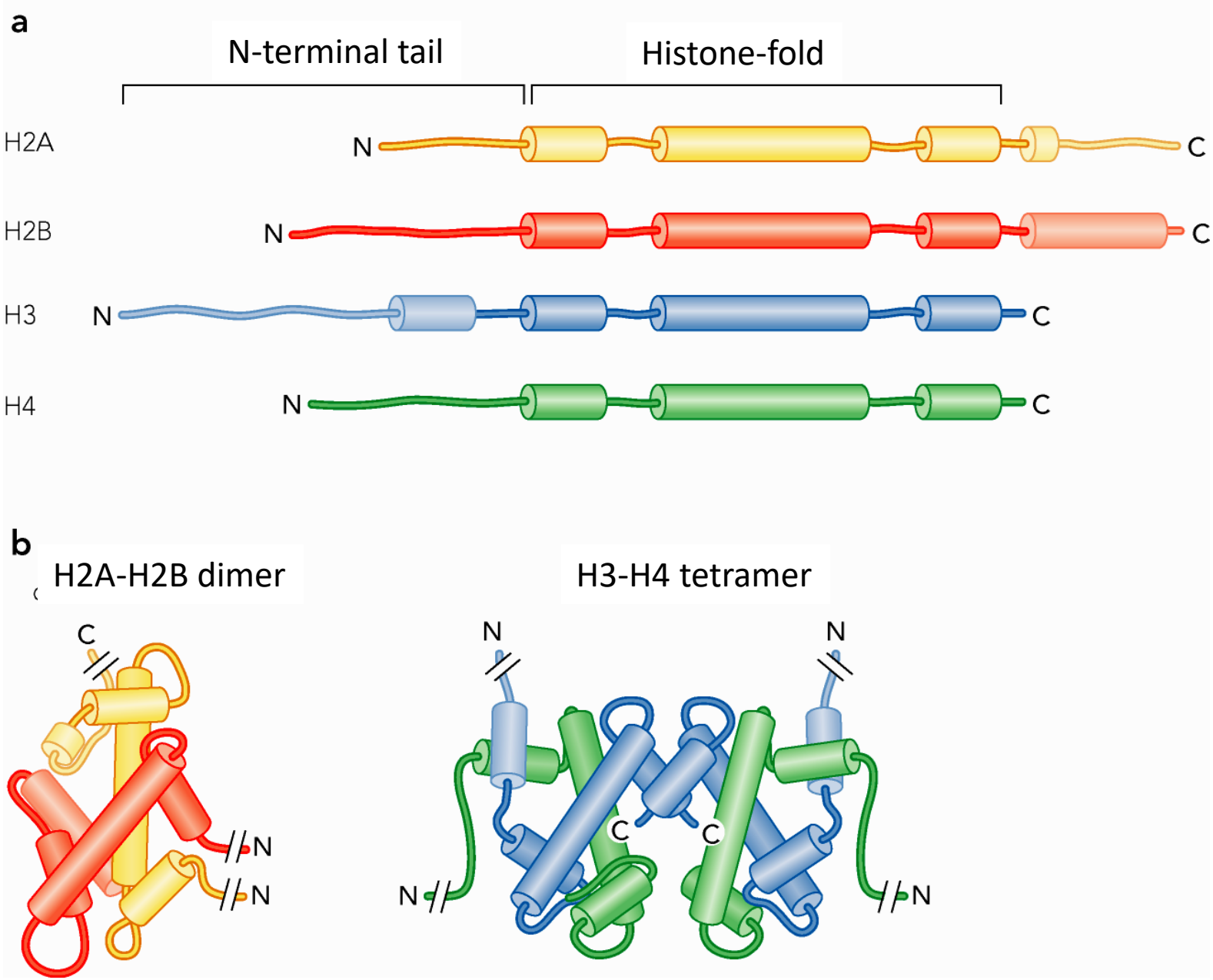


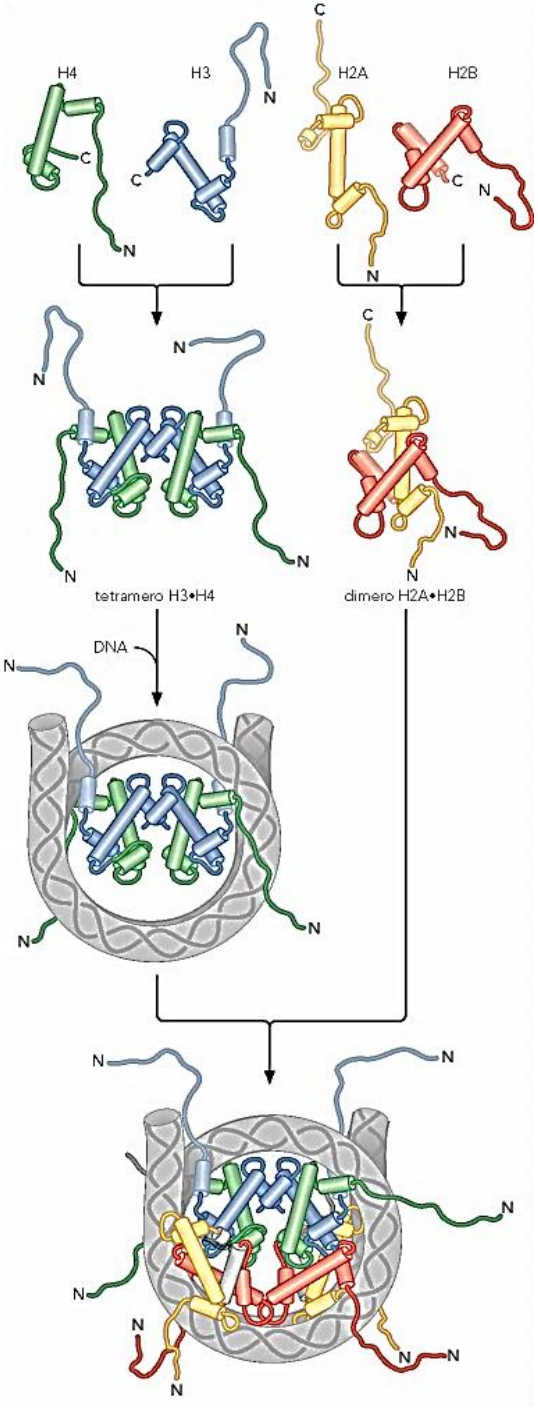


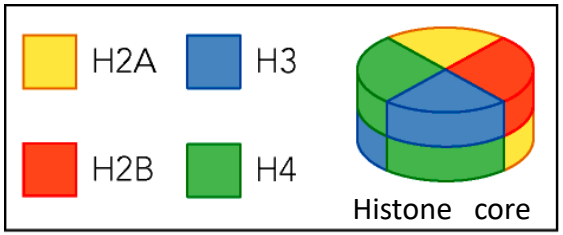
Nucleosome DNA structure

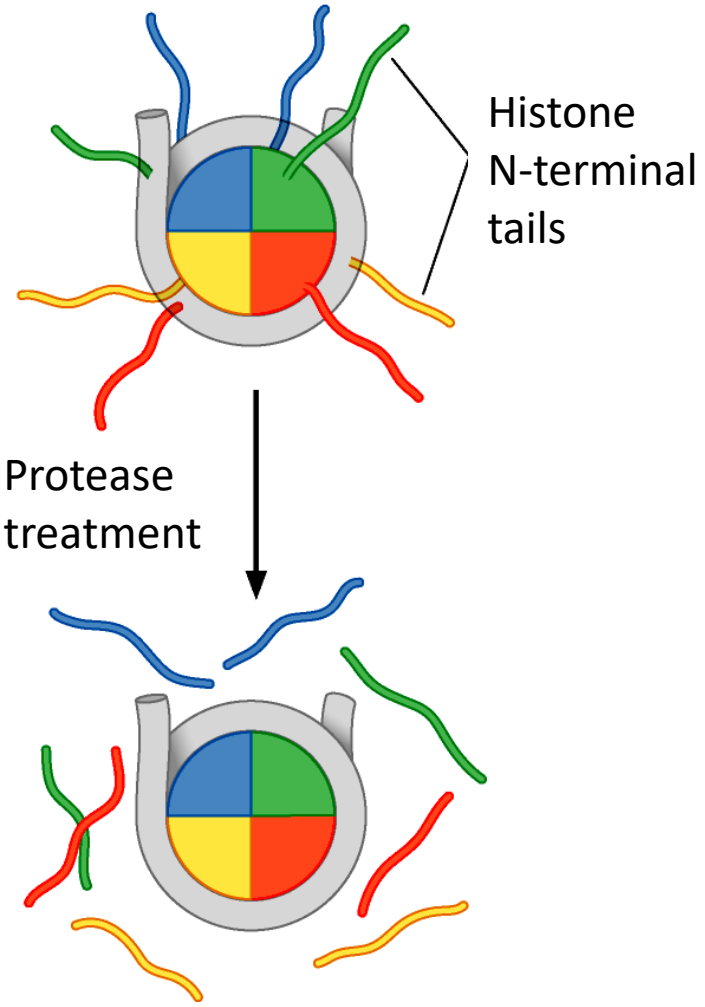
topology (supercoiling)

topoisomerases


a
N-terminal tail
Histone-fold
H2A
H2B
H3
H4
N
C
b
H2A-H2B dimer
H3-H4 tetramer
C
N
N
N
N
C
C
N
N


H4
H3
H2A
H2B
tetramero H3•H4
dimero H2A•H2B
DNA


H2A
H3
H2B
H4
Histone core


Histone
N-terminal
tails
Protease
treatment

Molecular organization of **heterochromatin** *versus* **euchromatin**

Features distinguishing HC from EC

- Accessibility
- Nucleosome positioning
- Post-translational Modifications to Histones (PTM)
- Histone-binding proteins and modifying enzymes
- Histone variants
- **DNA methylation**
- **Noncoding RNA**
- **Transcription factors**

} nucleosome

Heterochromatin

- Facultative
- Constitutive

Facultative contains genes that are activated at some phases, then silenced.

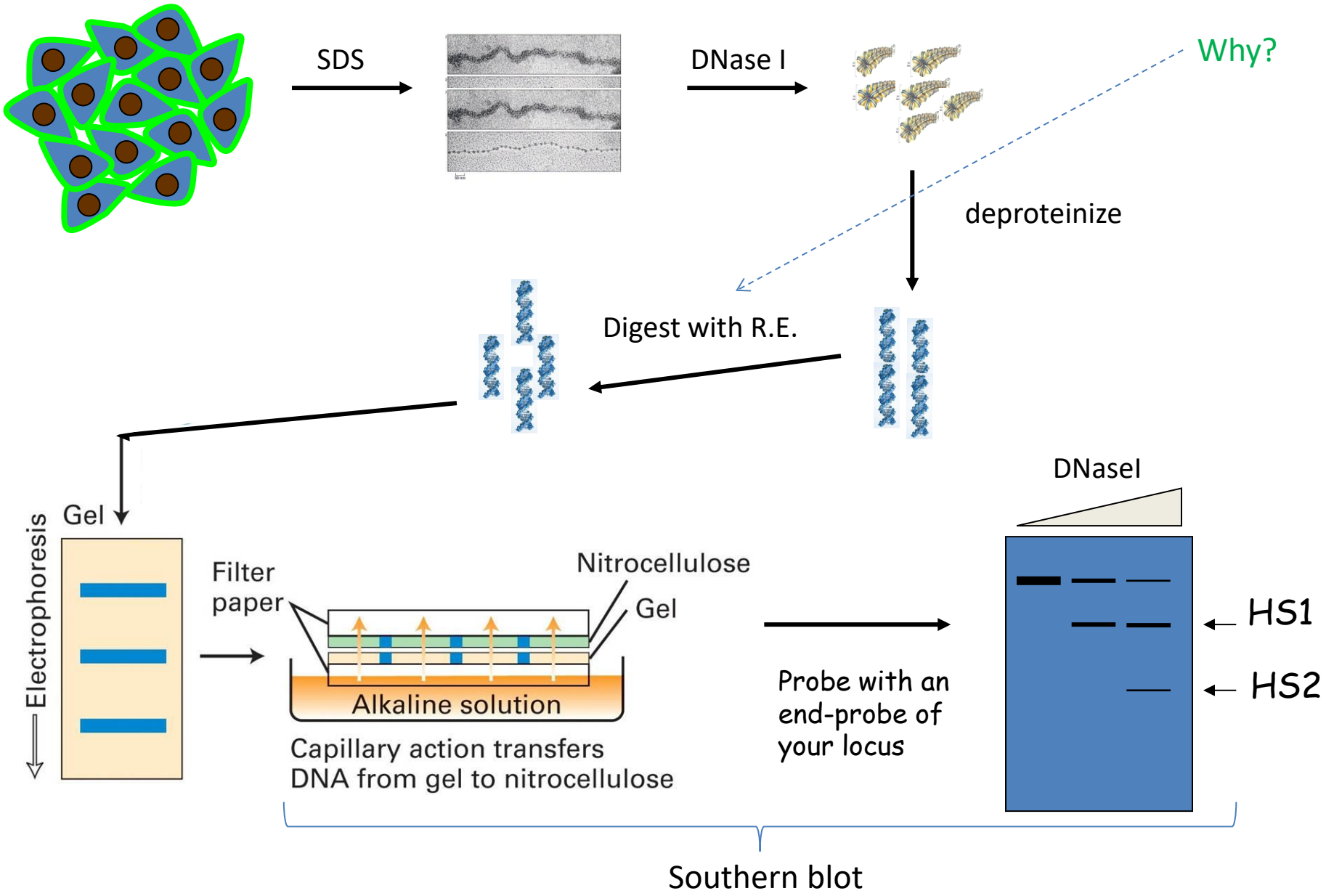
Constitutive at centromeres, telomeres

The leading fact here is represented by **repetitive sequences**

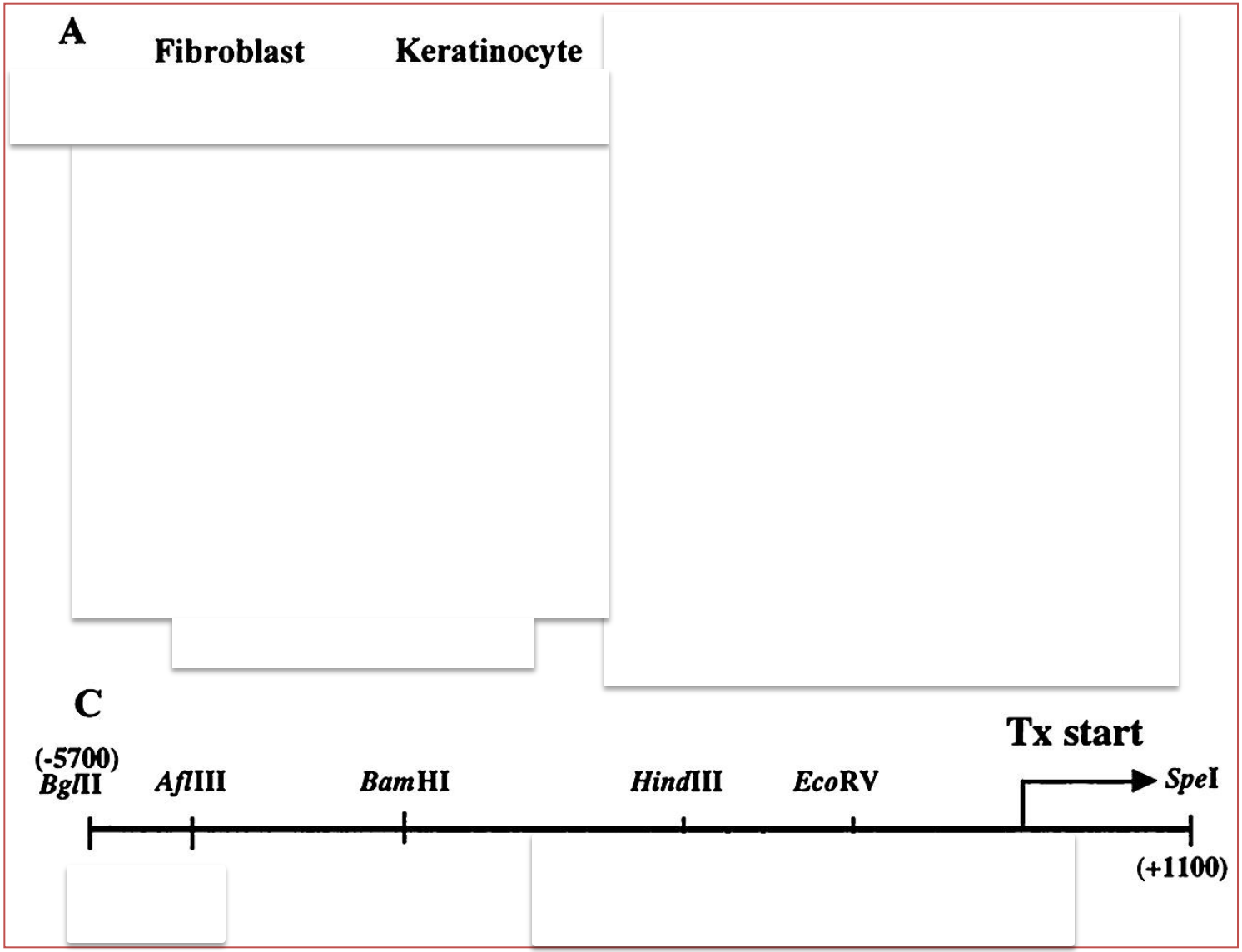
Are condensed (heterochromatic) and noncodensed (euchromatic) chromatin fragments (loci) distinguishable by simple biochemical assay ?

General principle: **enzyme accessibility**.

The classical assay to detect the gross organization of chromatin at a specified locus: the **DNase I Hypersensitivity** Assay

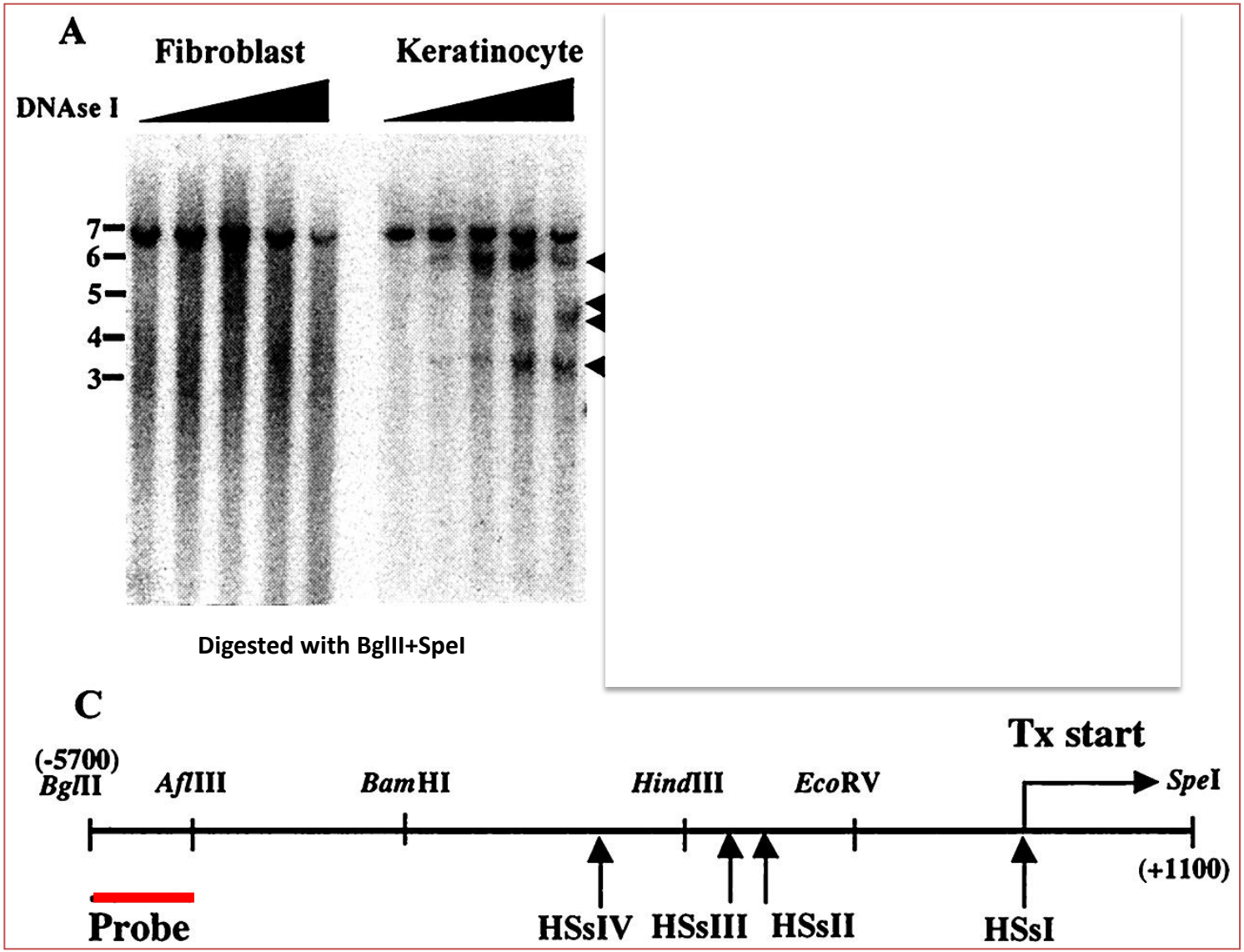


Keratin K14 gene, analysis of the 5’-flanking region



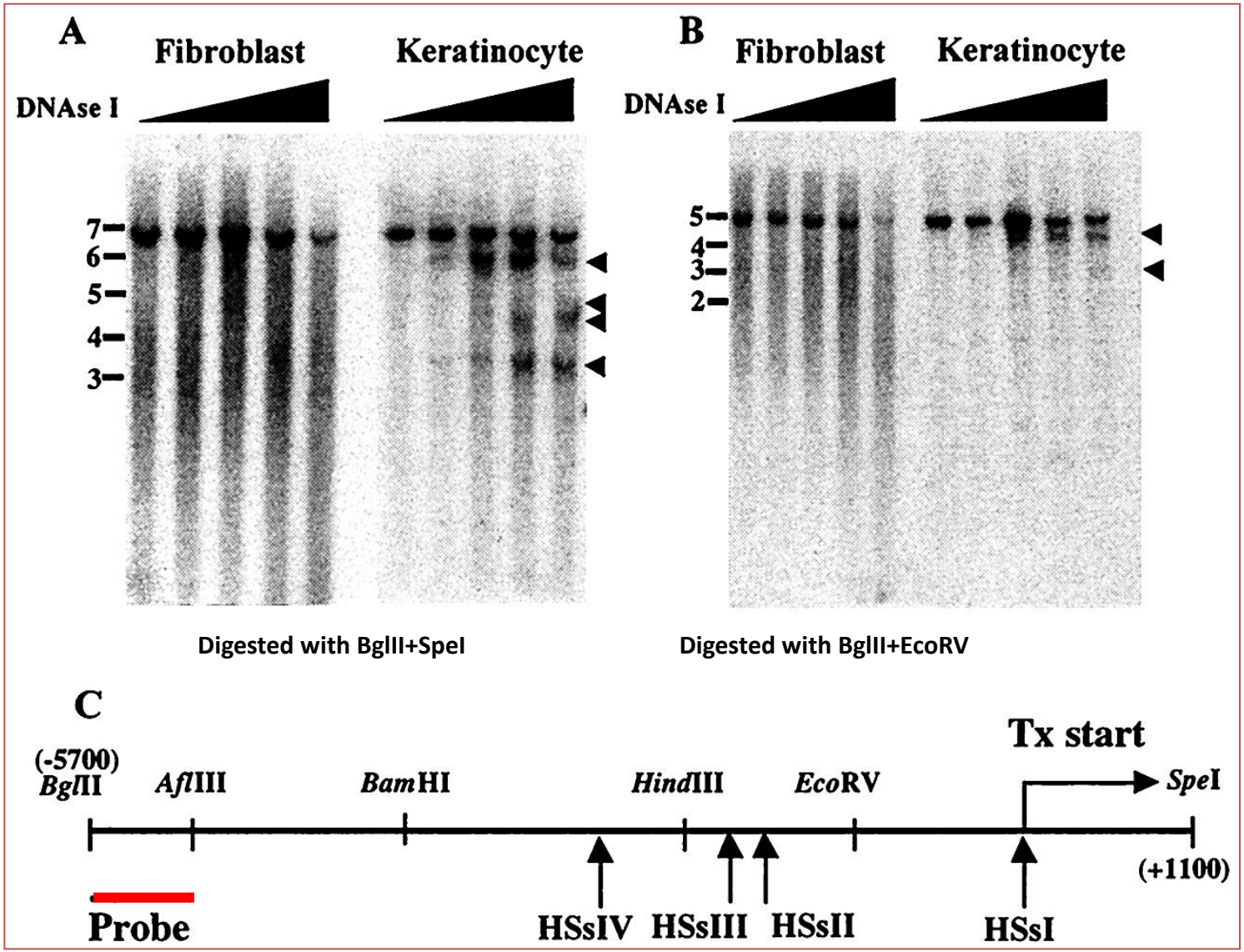
Tratto da: Sinha et al., (2000), Mol Cell Biol, 20: 2543-2555.

Keratin K14 gene, analysis of the 5’-flanking region



Tratto da: Sinha et al., (2000), Mol Cell Biol, 20: 2543-2555.

Keratin K14 gene, analysis of the 5'-flanking region



Tratto da: Sinha et al., (2000), Mol Cell Biol, 20: 2543-2555.

If accessibility is not the same in all parts of chromatin, may it depend (also) from different distribution of nucleosomes ?

Questions that are relevant to genome organization:

Do nucleosome display preferred positions in the genome ?
(are they just arrayed regularly spaced as they appear in pictures?)

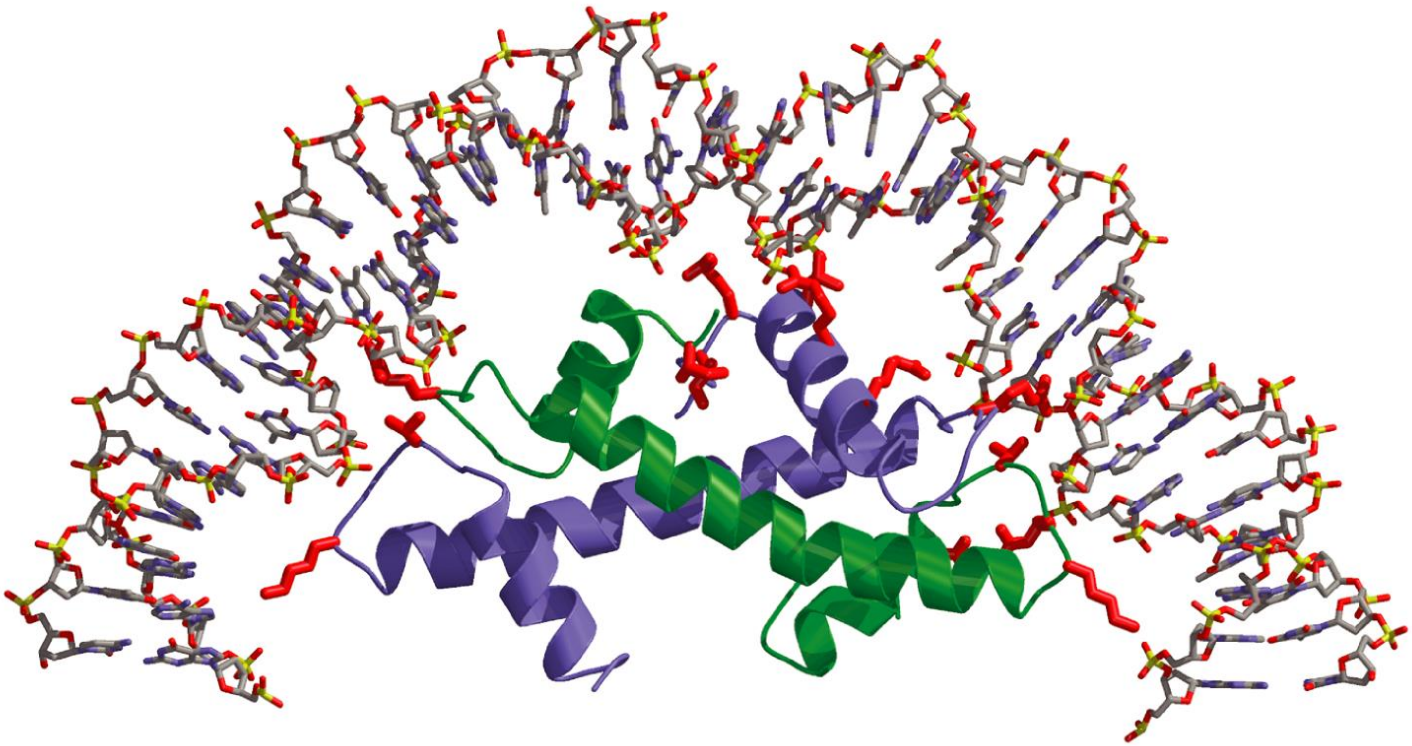
Do nucleosomes display any kind of **specificity** for nucleotide sequence?

Are nucleosome mobile ? Can they translate position over DNA ?
(are they stable on the same sequence they were assembled to ?)

Is nucleosome position **functional** to gene/genome regulation ?

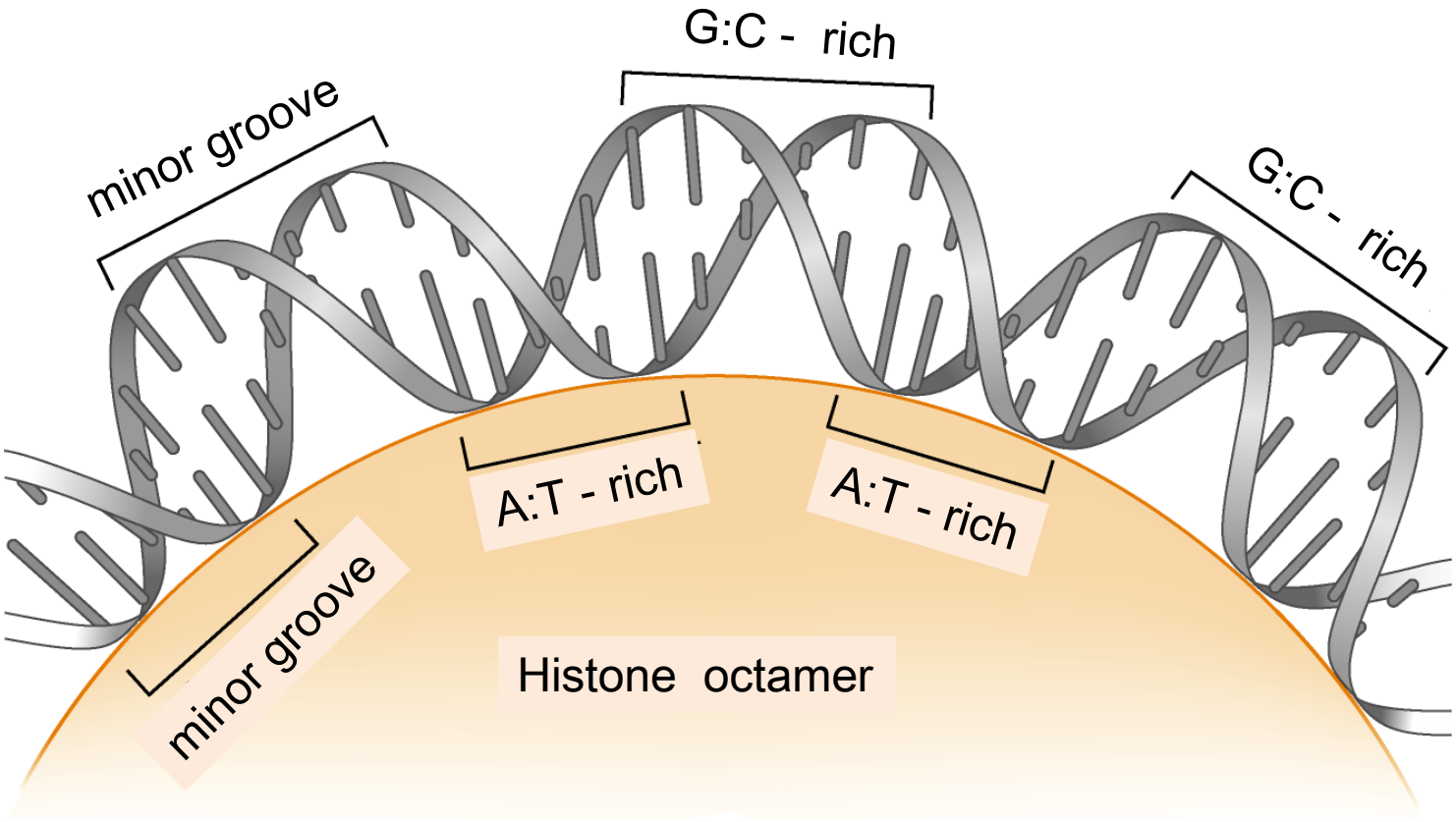
«Passive» positioning

Protein-DNA contacts are mainly due to basic aminoacids – backbone phosphates



«Passive» positioning

Weak sequence specificity: in general, A/T rich sequences are in touch, whereas C/G rich stay outside.
This is due to intrinsic bending of DNA axis in different sequence contexts.

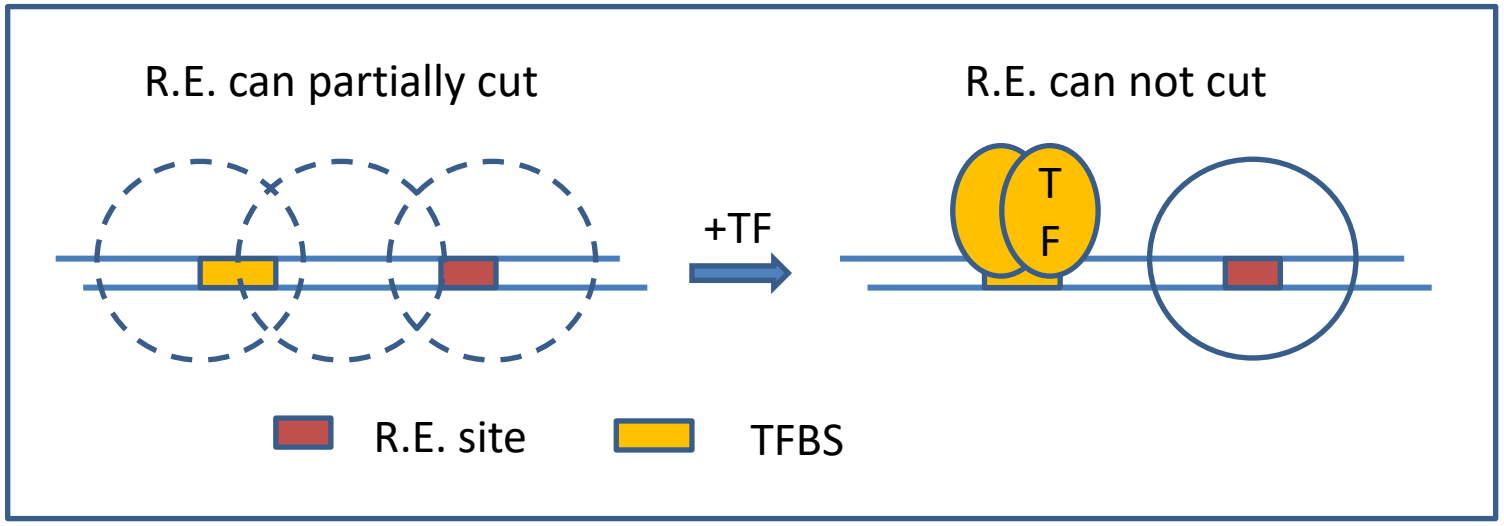


«Assisted» nucleosome positioning:

- ATP-dependent chromatin remodeler
- Presence of DNA-binding factors, active transcription and other factors

How can nucleosome position on a given locus be assessed ?

1. <u>In the test tube: accessibility to restriction enzymes</u>



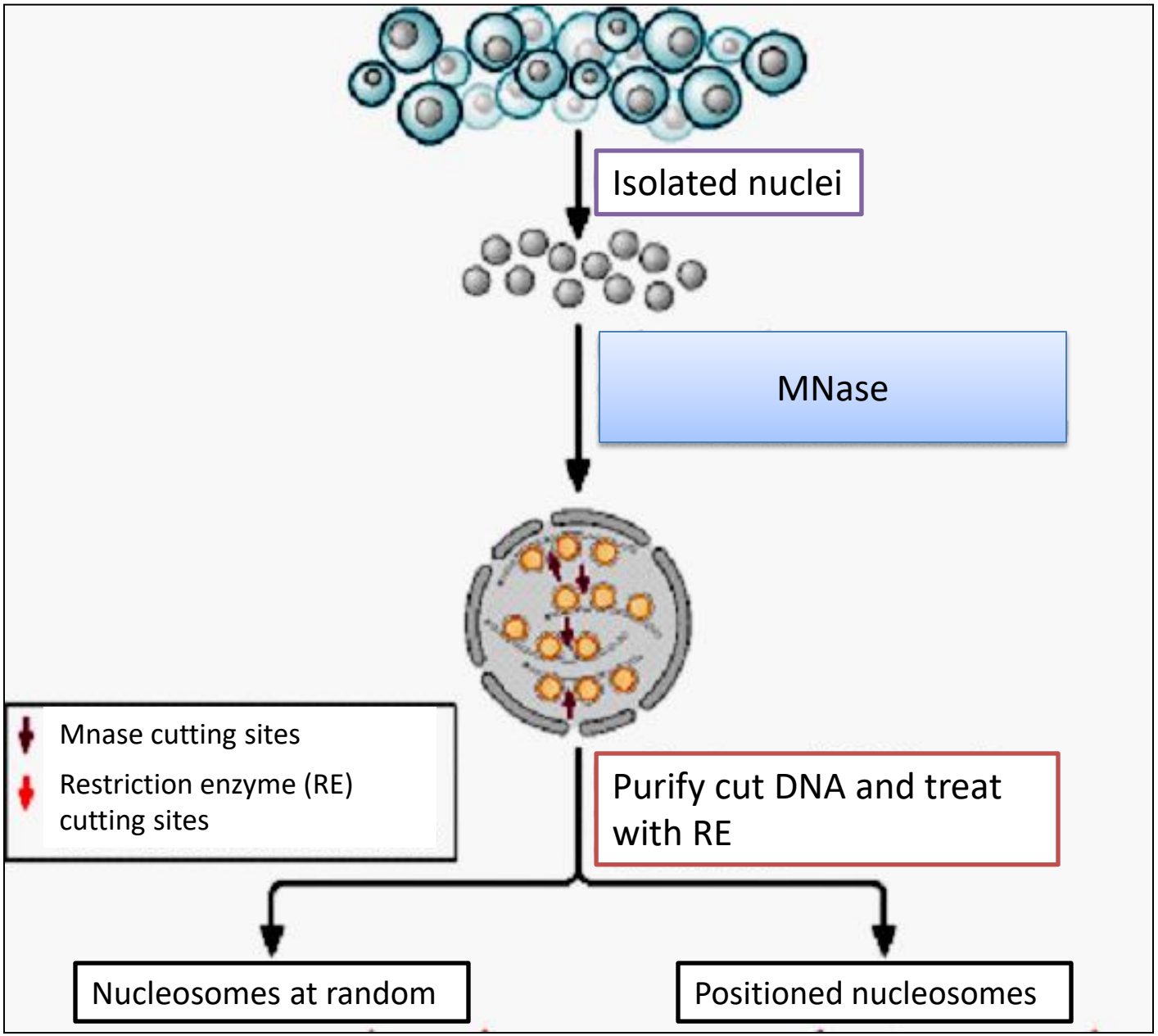
2. <u>In vivo: Micrococcus Nuclease digestion</u>

Micrococcal Nuclease I is a mixed endo-exonuclease.

MNase induces single-strand breaks and subsequently double stranded ones by cleaving the complementary strand in close proximity to the first break. MNase continues to digest the exposed DNA ends until it reaches an obstruction, such as a nucleosome or a very stably bound Transcription Factor.
In appropriate condition, then, MNase releases fragments of approximately one nucleosome length (~147 bp), which are typically selected for sequencing in MNase-Seq experiments.

Nucleosome positioning analysis

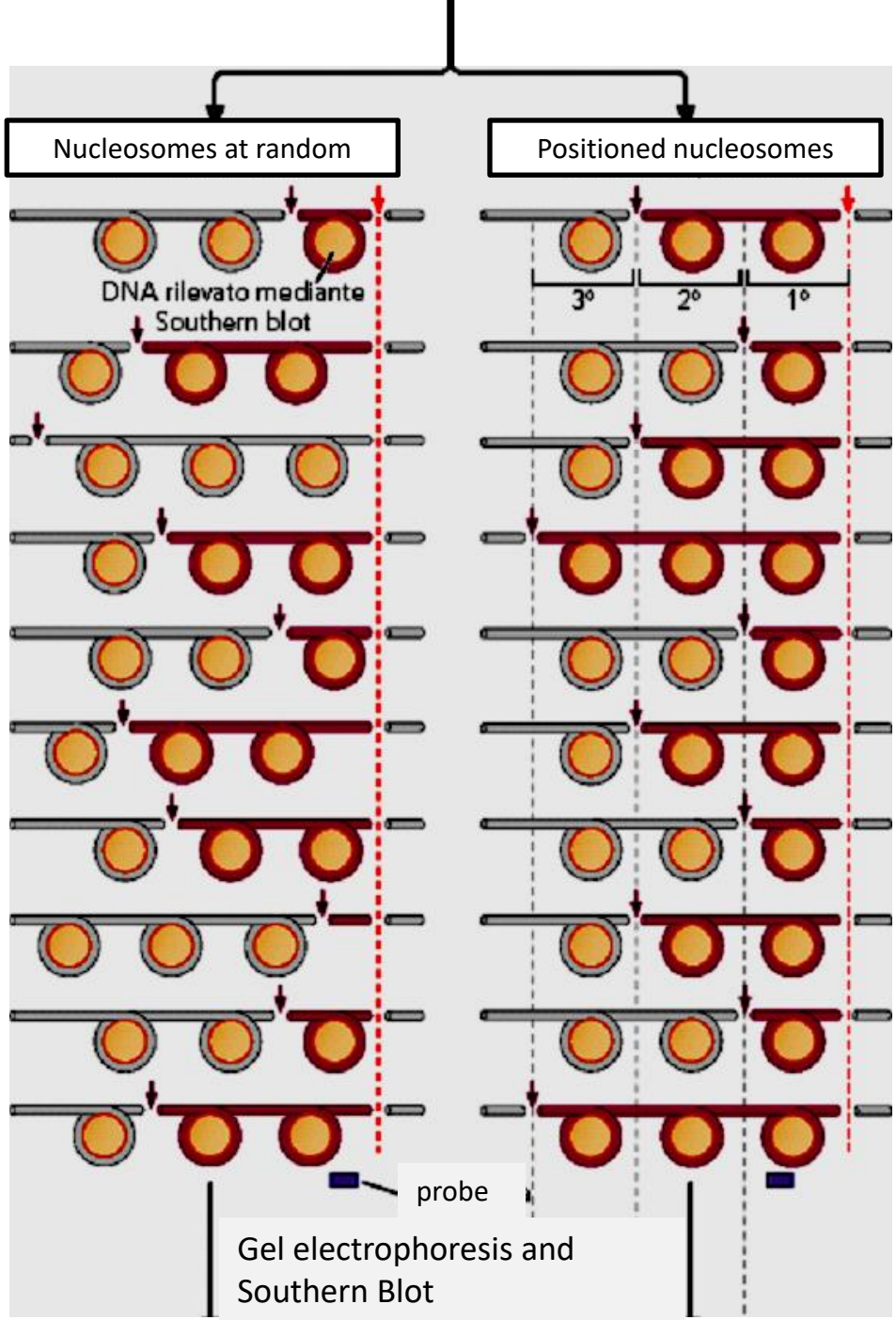
1st step: digestion

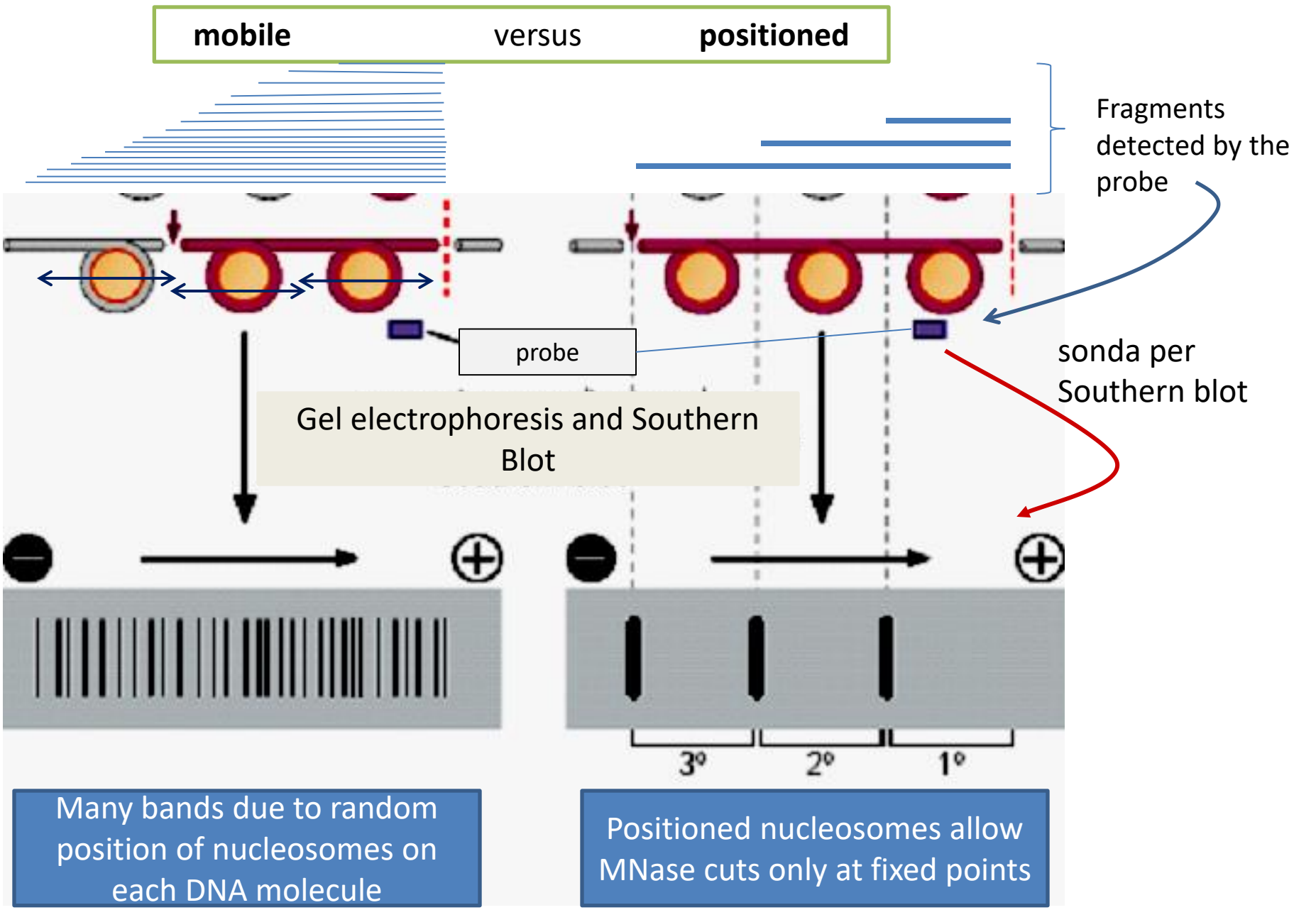


2 different cases:
mobile nucleosomes
versus
positioned nucleosomes

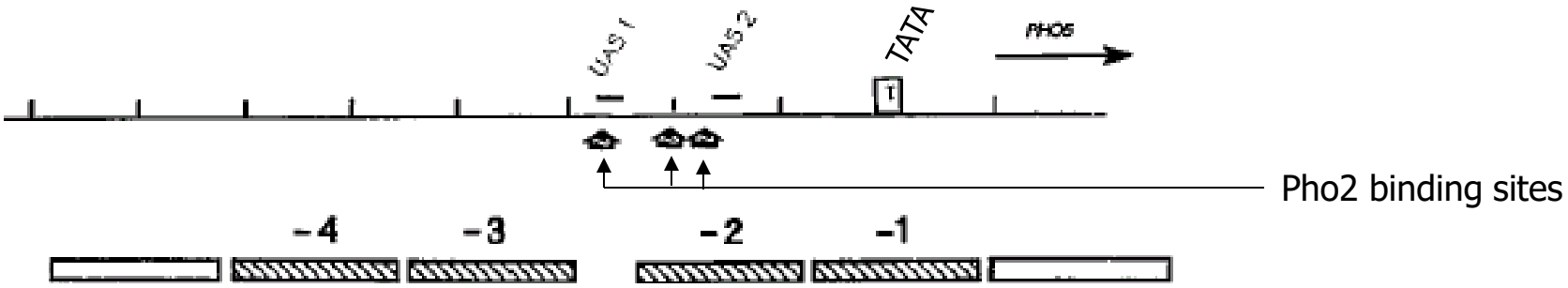
⬇ MNase cuts

⬇ R.E. cuts




mobile versus positioned
Fragments detected by the probe
probe
sonda per Southern blot
Gel electrophoresis and Southern Blot
3°
2°
1°
Many bands due to random position of nucleosomes on each DNA molecule
Positioned nucleosomes allow MNase cuts only at fixed points

Is nucleosome position functionally relevant ?



Pho4 is the P-sensitive inducer, whereas Pho2 is constitutive

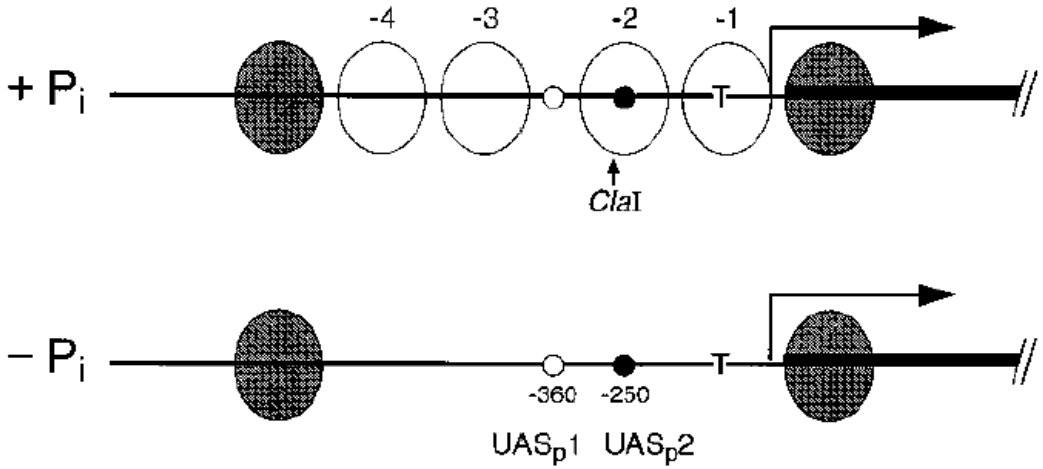


Figure 1. Chromatin Structure at the *PHO5* Promoter

Nucleosomes −1, −2, −3, and −4 are remodeled upon activation of the promoter by phosphate starvation conditions (Almer et al., 1986). The small circles mark UASp1 (open) and UASp2 (solid), which are Pho4-binding sites found by in vitro (Vogel et al., 1989) and in vivo (Venter et al., 1994) footprinting experiments. The positions are listed relative to the coding sequence (solid bar). T denotes the TATA box (Rudolph and Hinnen, 1987). The location of a ClaI site at −275 relative to the coding region is shown.

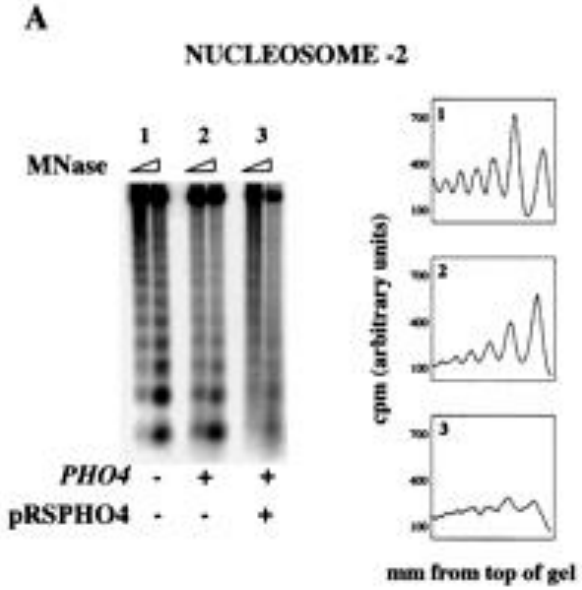
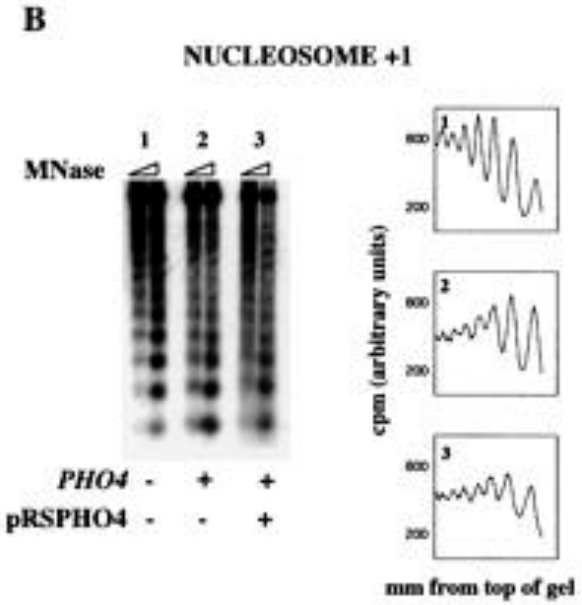


FIG. 5. In vivo chromatin remodeling of episomal *PHO5*. Spheroplasts from the indicated strains were treated with micrococcal nuclease (MNase), and the DNA was purified and Southern blotted. (A) The blot was probed with probe A (Fig. 2). Data from each sample were quantified, and the distance from the top of the gel was graphed against the signal density. (B) The blot shown in panel A was stripped and reprobed with probe B (Fig. 2).



Nucleosome - **histones**

post-translational modifications (PTMs)

Table 1. Different Classes of Modifications Identified on Histones

Chromatin Modifications	Residues Modified	Functions Regulated
Acetylation	**K**-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	**K**-me1 **K**-me2 **K**-me3	Transcription, Repair
Methylation (arginines)	**R**-me1 **R**-me2a **R**-me2s	Transcription
Phosphorylation	**S**-ph **T**-ph	Transcription, Repair, Condensation
Ubiquitylation	**K**-ub	Transcription, Repair
Sumoylation	**K**-su	Transcription
ADP ribosylation	**E**-ar	Transcription
Deimination	**R > Cit**	Transcription
Proline Isomerization	**P-cis > P-trans**	Transcription

Overview of different classes of modification identified on histones. The functions that have been associated with each modification are shown. Each modification is discussed in detail in the text under the heading of the function it regulates.

From: Kouzarides, T. (2007), Cell 128:693-705

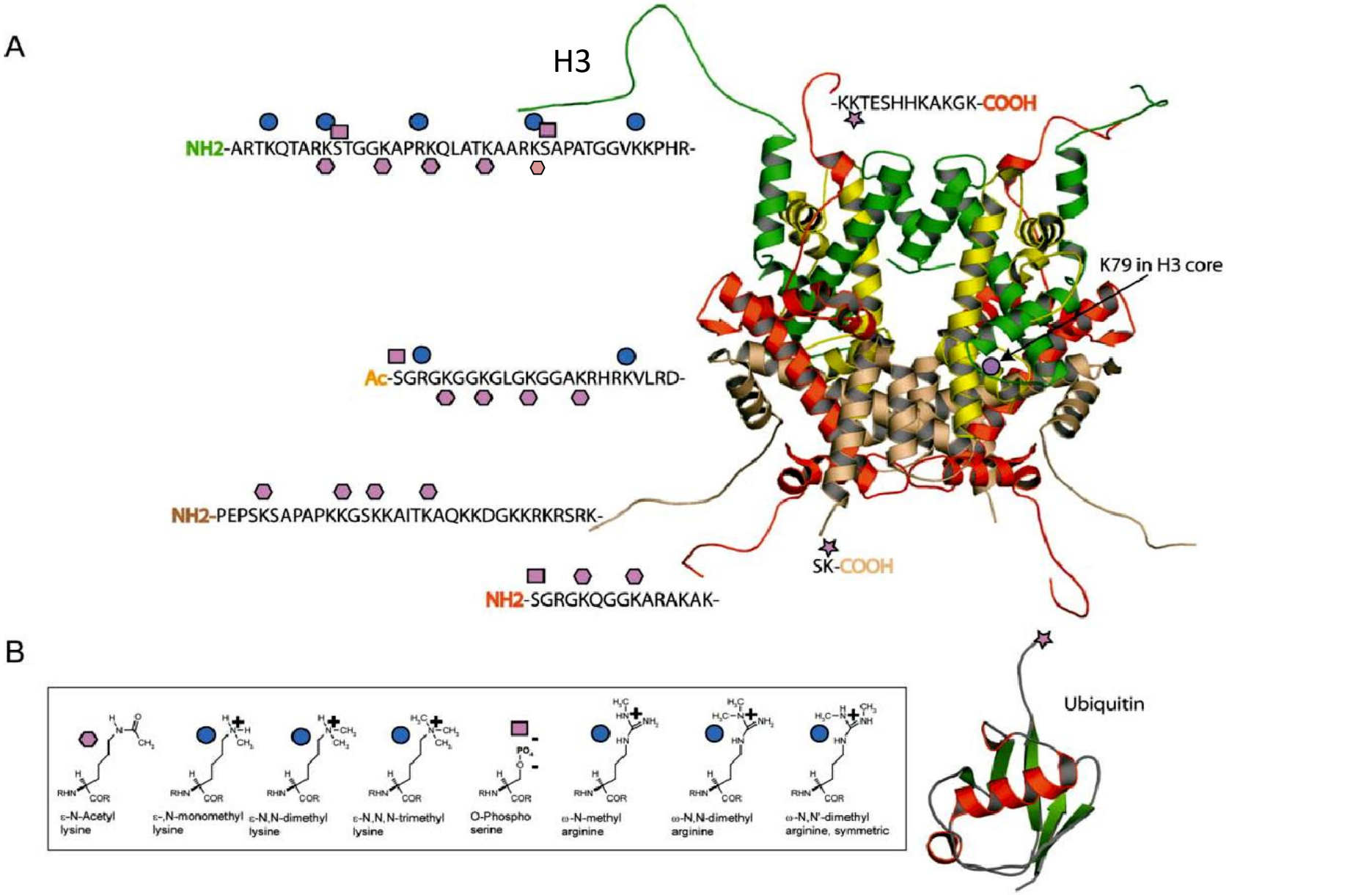
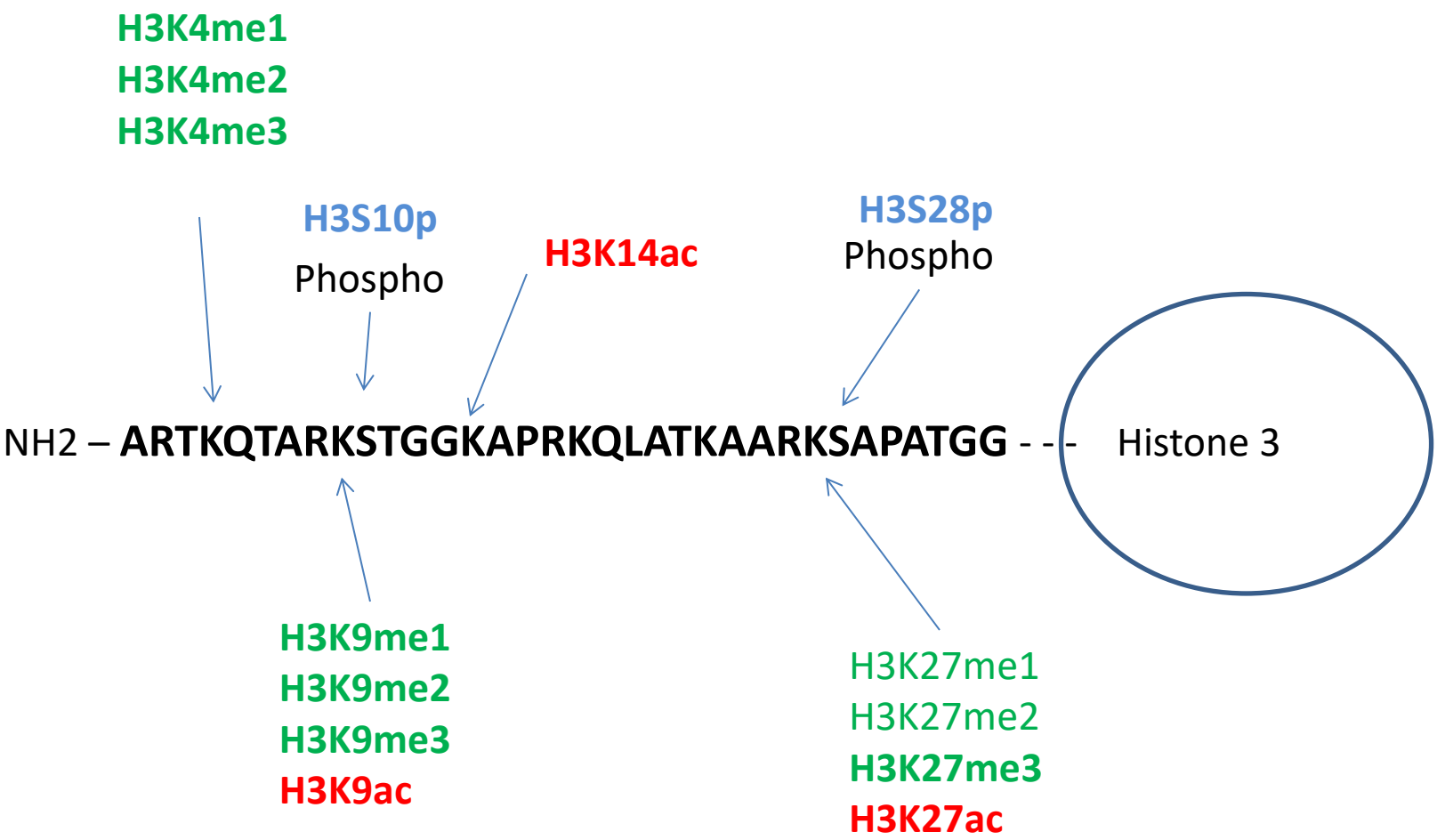


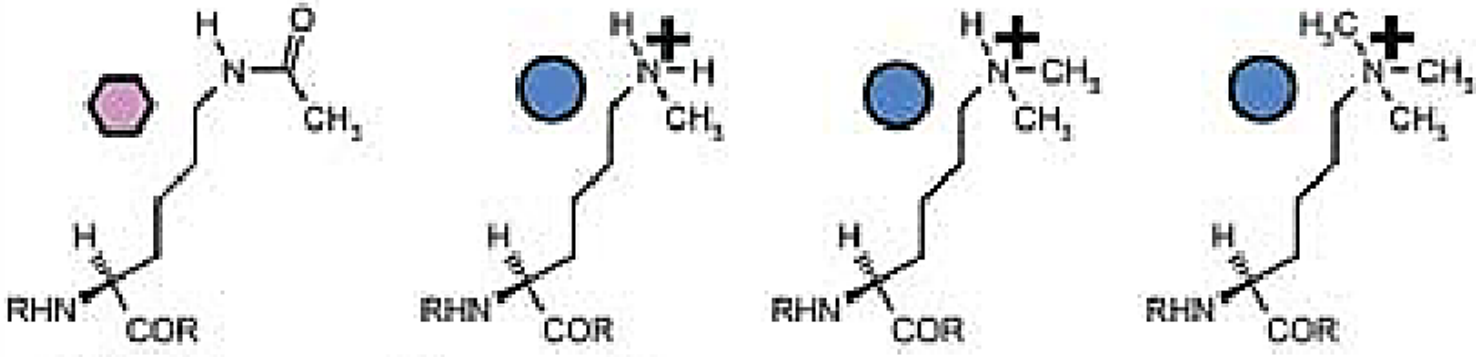
Figure 4. The Types of Posttranslational Modifications Observed on the Core Histones

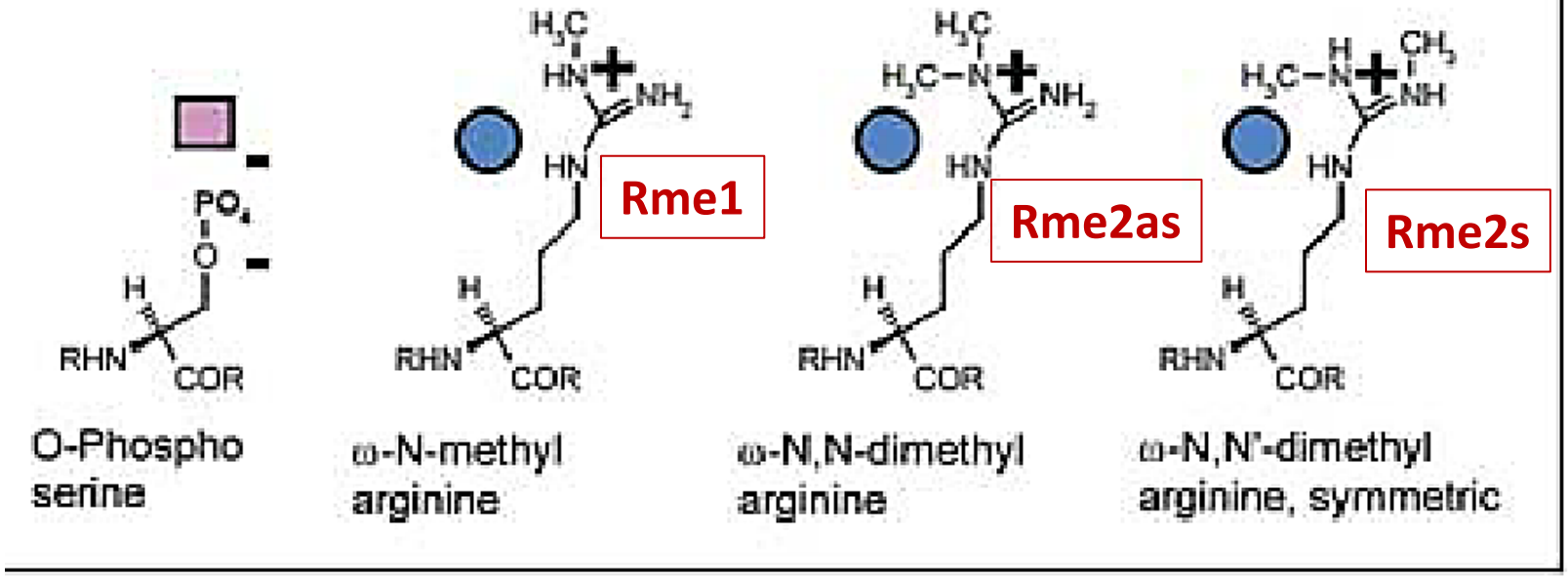
(A) The histone octamer portion of the nucleosome core particle is shown. The sites of modifications on marked. For clarity, the modifications are shown on one copy of each protein.

(B) The covalent modifications of the amino acids are shown.

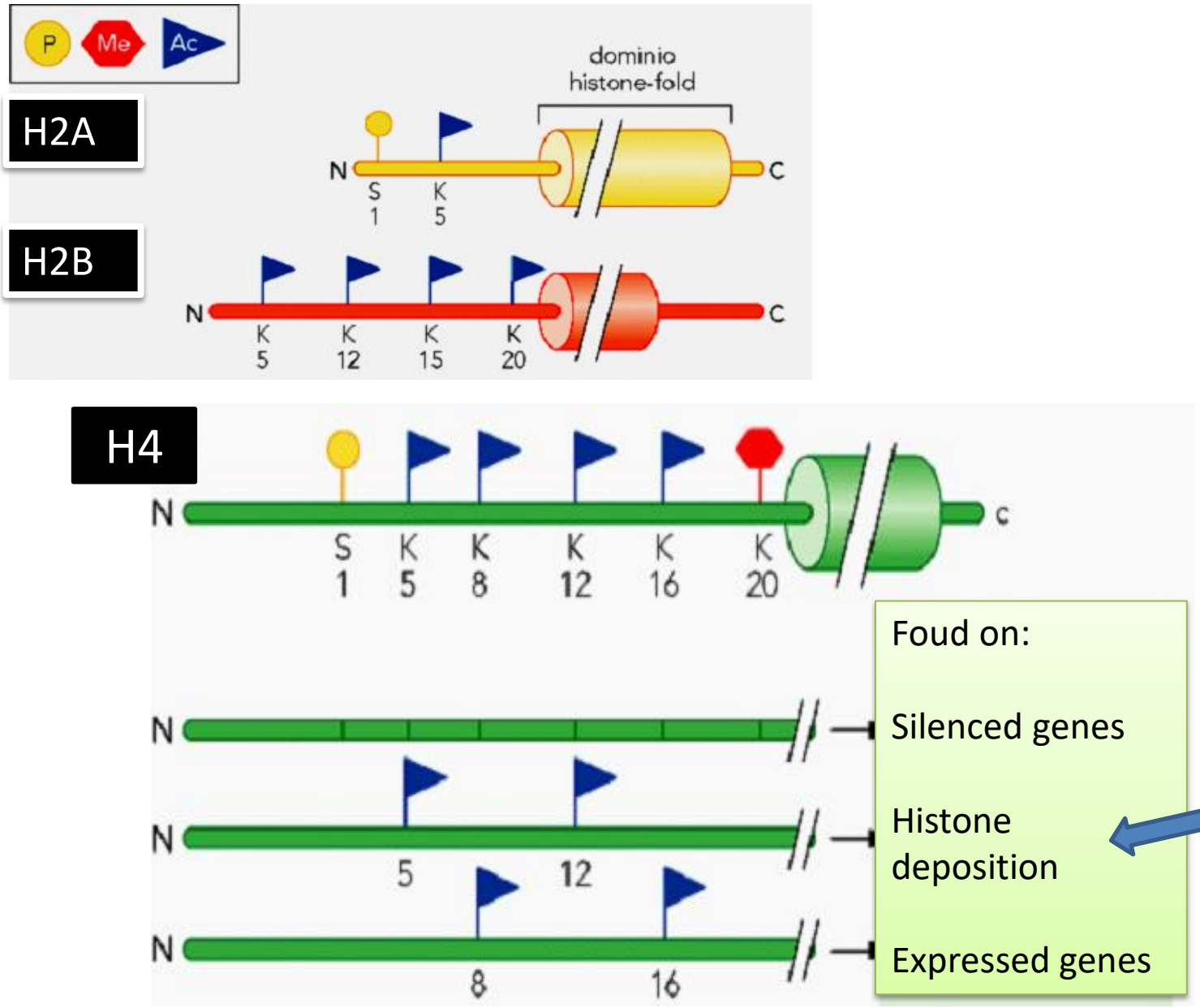
Nomenclature




ε-N-Acetyl lysine
ε-,N-monomethyl lysine
ε-N,N-dimethyl lysine
ε-N,N,N-trimethyl lysine

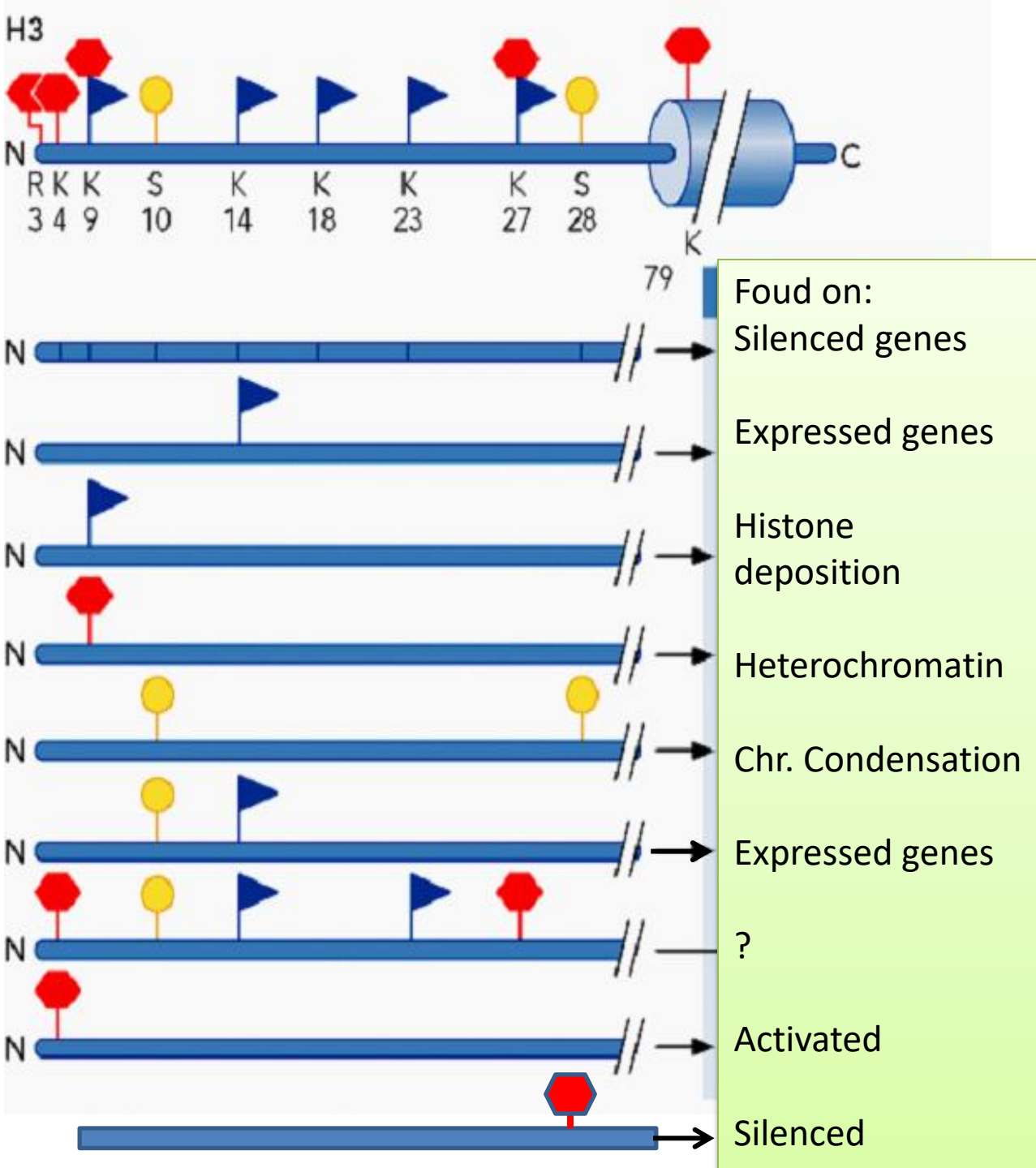

O-Phospho serine
ω-N-methyl arginine
Rme1
ω-N,N-dimethyl arginine
Rme2as
ω-N,N'-dimethyl arginine, symmetric
Rme2s

- the histone code



- the histone code

H3



This view was derived from single-gene studies.

PTMs are context-specific and quite dynamic

- Some PTMs always go together
- Other PTMs are mutually exclusive (e.g. Kac/Kme)
- Specific PTMs display effects on other PTMs (see next figure)

Question: are the two copies of each histone modified in the same way in the individual nucleosome?

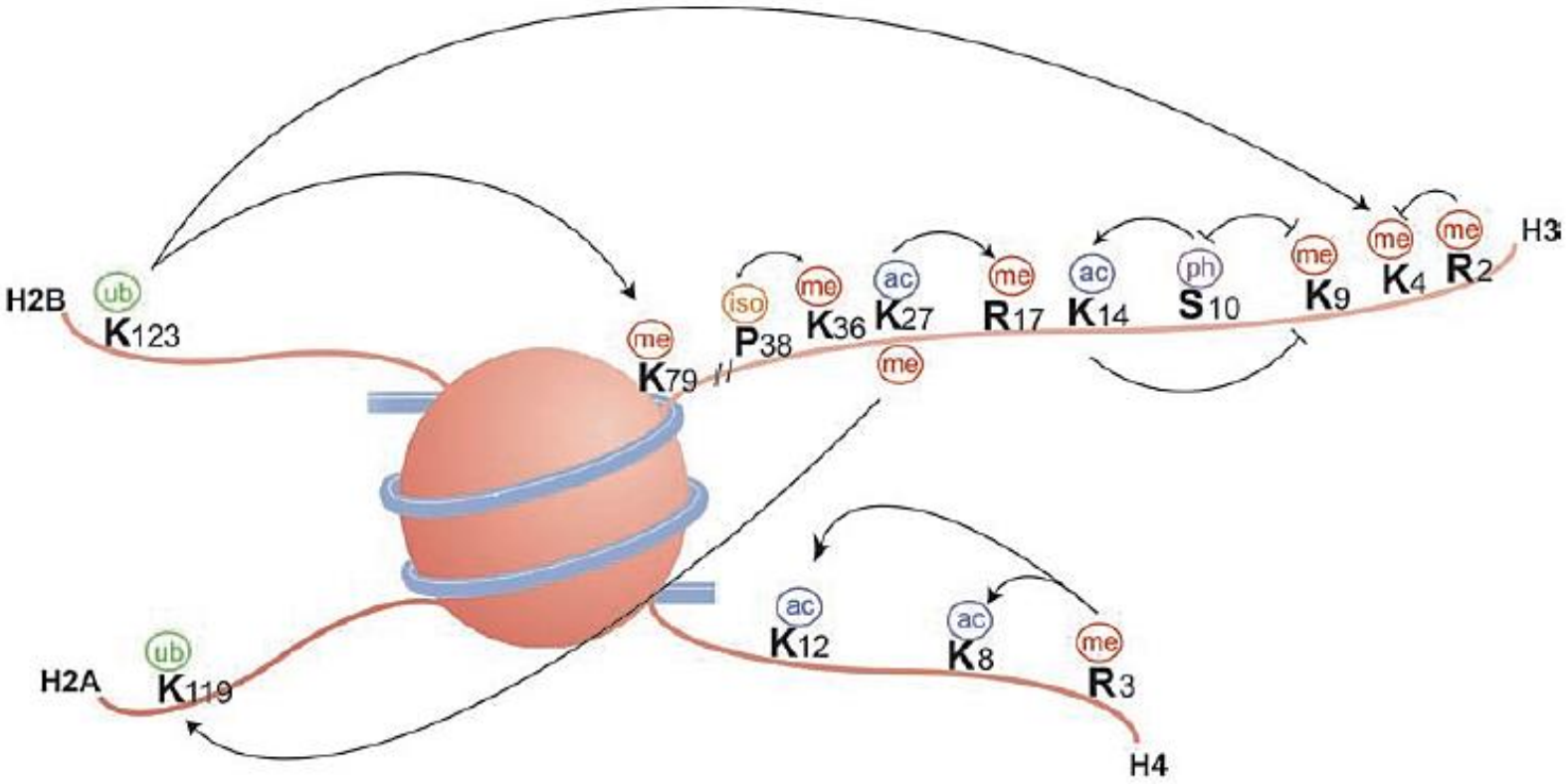
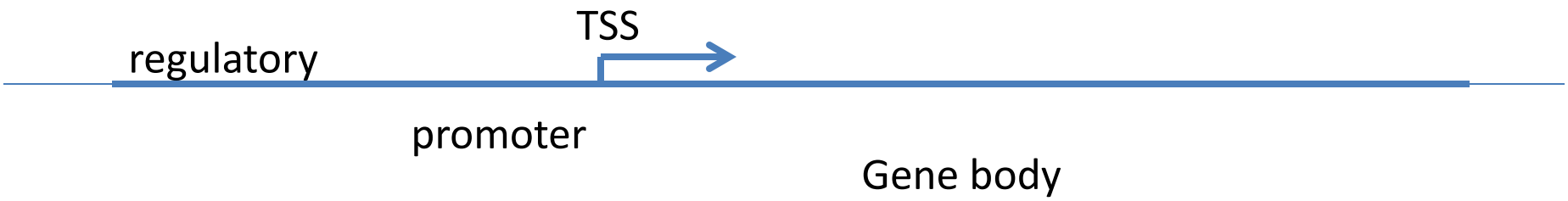


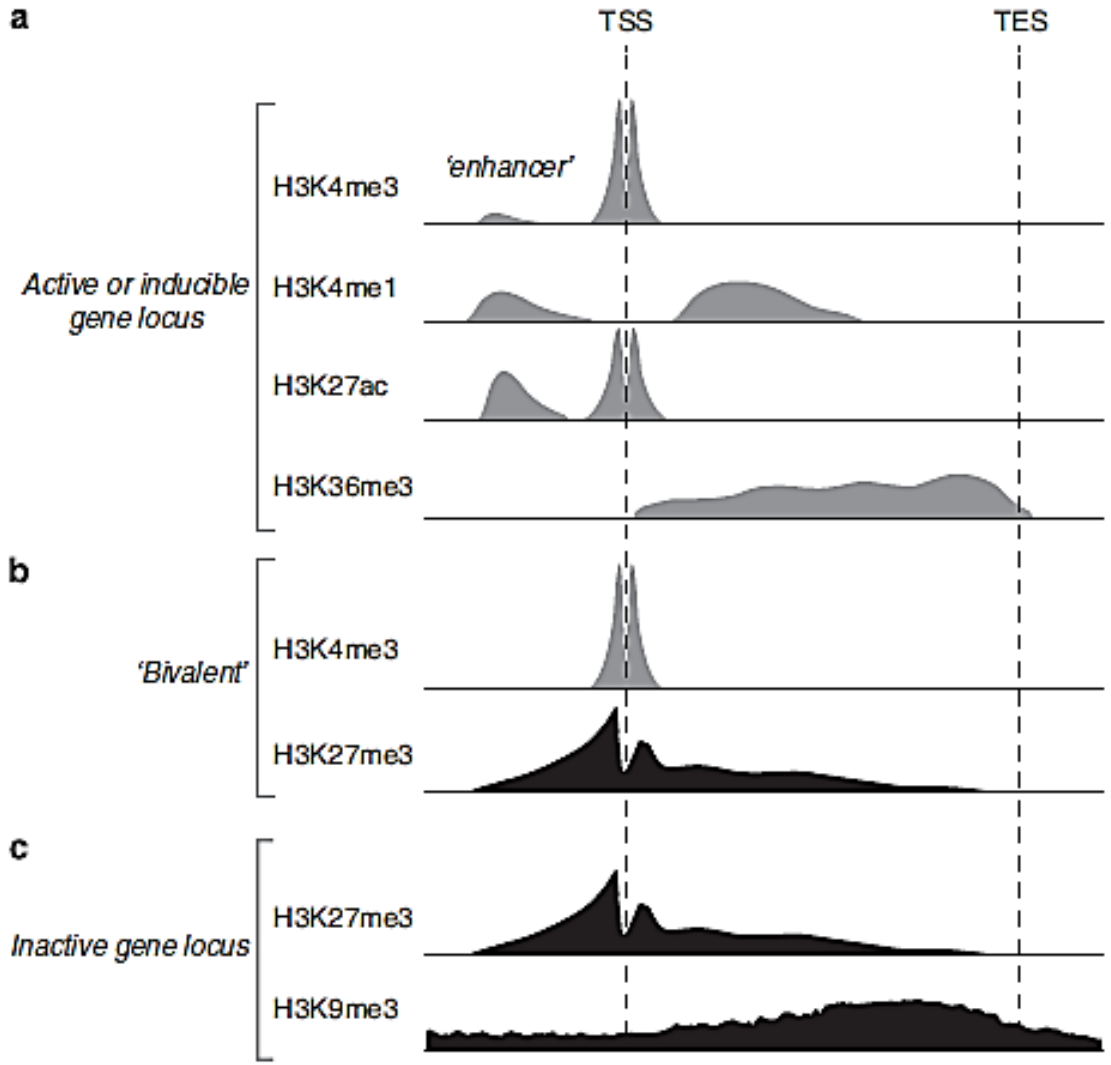
Figure 2 Histone modification cross-talk. Histone modifications can positively or negatively affect other modifications. A positive effect is indicated by an arrowhead and a negative effect is indicated by a flat head (updated from reference [53]).

Textbook: Bannister & Kouzarides, 2011

PTMs are not only «grossly» distributed to HC/EC

They show finely distinct distributions in functional parts of chromatin, for examples in genes:





Kimura et al, J Human Genet (2013) 58:439-445

Figure 3 Distribution of histone modifications. Distributions of six modifications with respect to genes are schematically illustrated. TSS, transcription start site; TES, transcription end site. H3K4me3 is enriched around TSSs. H3K4me1 is nriched around enhancers and more downstream. H3K27ac is enriched around active enhancers and TSSs. In undifferentiated stem cells, both H3K4me3 and H3K27me3 (active and inactive marks, respectively) are enriched around TSSs on many genes. H3K27me3 is enriched around inactive TSS in somatic cells. H3K9me3 is broadly distributed on inactive regions. H3K27me3 and H3K9me3 are usually not colocalized. TSSs are generally devoid of nucleosomes.

interacting proteins

Interacting proteins are:

Enzymes introducing modifications	PTM **writers**
Enzymes reversing modifications	PTM **erasers**
Proteins binding to specific PTM	PTM **readers**

PTM = Post-Translational Modification

histone-modifying enzymes:

HAT- histone acetyltransferases
HDAC – histone deacetylases
HMT – histone methyltransferases
histone demethylases
histone kinases
histone ribosyltrabsferases
ubiquitin-transferases

(ATP-dependent remodelling enzymes)

Modifications are dynamic and reversible

Enzymes that Modify Histones	Residues Modified
Acetyltransferase	
HAT1	H4 (K5, K12)
CBP/P300	H3 (K14, K18) H4 (K5, K8) H2A (K5) H2B (K12, K15)
PCAF/GCN5	H3 (K9, K14, K18)
TIP60	H4 (K5, K8, K12, K16) H3 K14
HB01 (ScESA1, SpMST1)	H4 (K5, K8, K12)
ScSAS3	H3 (K14, K23)
ScSAS2 (SpMST2)	H4 K16
ScRTT109	H3 K56

Other enzymes exist with reduced specificity

From: Kouzarides, T. (2007), Cell 128:693-705

Enzymes that Modify Histones	Residues Modified
Deacetylases	
SirT2 (ScSir2)	H4 K16

histone deacetylases (low specificity)

HDAC: two classes

Class III: NAD-dependent deacetylases (Sir-Sirtuins)

From: Kouzarides, T. (2007), Cell 128:693-705

writers

Lysine Methyltransferase	
SUV39H1	H3K9
SUV39H2	H3K9
G9a	H3K9
ESET/SETDB1	H3K9
EuHMTase/GLP	H3K9
CLL8	H3K9
SpClr4	H3K9
MLL1	H3K4
MLL2	H3K4
MLL3	H3K4
MLL4	H3K4
MLL5	H3K4
SET1A	H3K4
SET1B	H3K4
ASH1	H3K4
Sc/Sp SET1	H3K4
SET2 (Sc/Sp SET2)	H3K36
NSD1	H3K36
SYMD2	H3K36
DOT1	H3K79
Sc/Sp DOT1	H3K79
Pr-SET 7/8	H4K20
SUV4 20H1	H4K20
SUV420H2	H4K20
SpSet 9	H4K20
EZH2	H3K27
RIZ1	H3K9

erasers

Enzymes that Modify Histones	Residues Modified
Lysine Demethylases	
LSD1/BHC110	H3K4
JHDM1a	H3K36
JHDM1b	H3K36
JHDM2a	H3K9
JHDM2b	H3K9
JMJD2A/JHDM3A	H3K9, H3K36
JMJD2B	H3K9
JMJD2C/GASC1	H3K9, H3K36
JMJD2D	H3K9

From: Kouzarides, T. (2007), Cell 128:693-705

writers

Arginine Methlytransferases	
CARM1	H3 (R2, R17, R26)
PRMT4	H4R3
PRMT5	H3R8, H4R3
Serine/Thrionine Kinases	
Haspin	H3T3
MSK1	H3S28
MSK2	H3S28
CKII	H4S1
Mst1	H2BS14
Ubiquitilases	
Bmi/Ring1A	H2AK119
RNF20/RNF40	H2BK120
Proline Isomerases	
ScFPR4	H3P30, H3P38

From: Kouzarides, T. (2007), Cell 128:693-705

readers

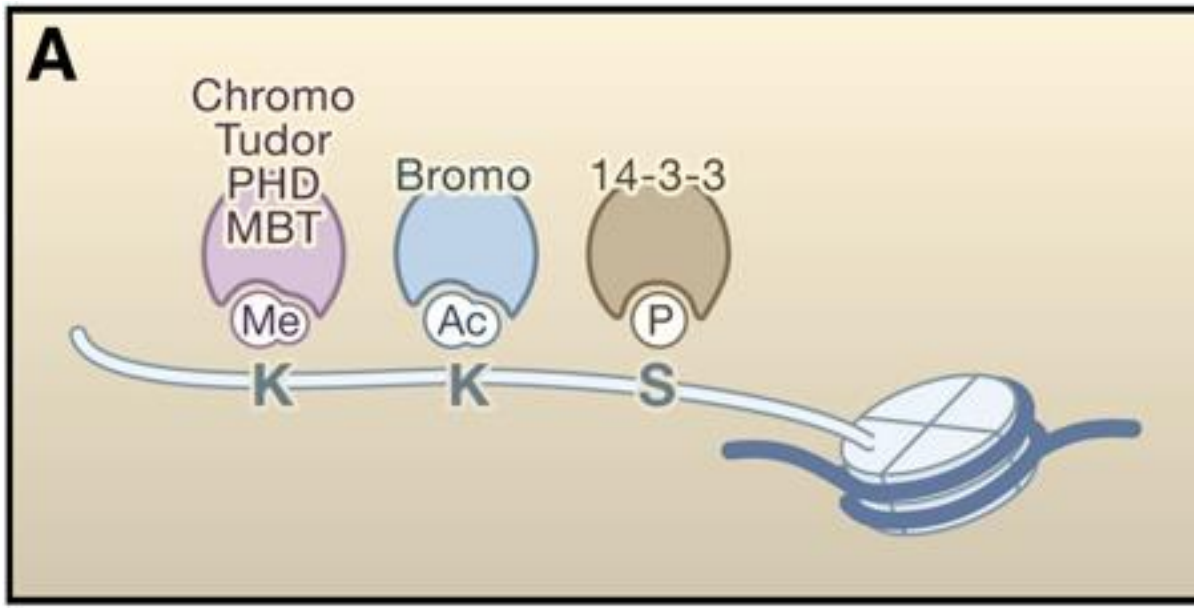


Figure 1. Recruitment of Proteins to Histones
(A) Domains used for the recognition of methylated lysines, acetylated lysines, or phosphorylated serines.

From: Kouzarides, T. (2007), Cell 128:693-705

readers

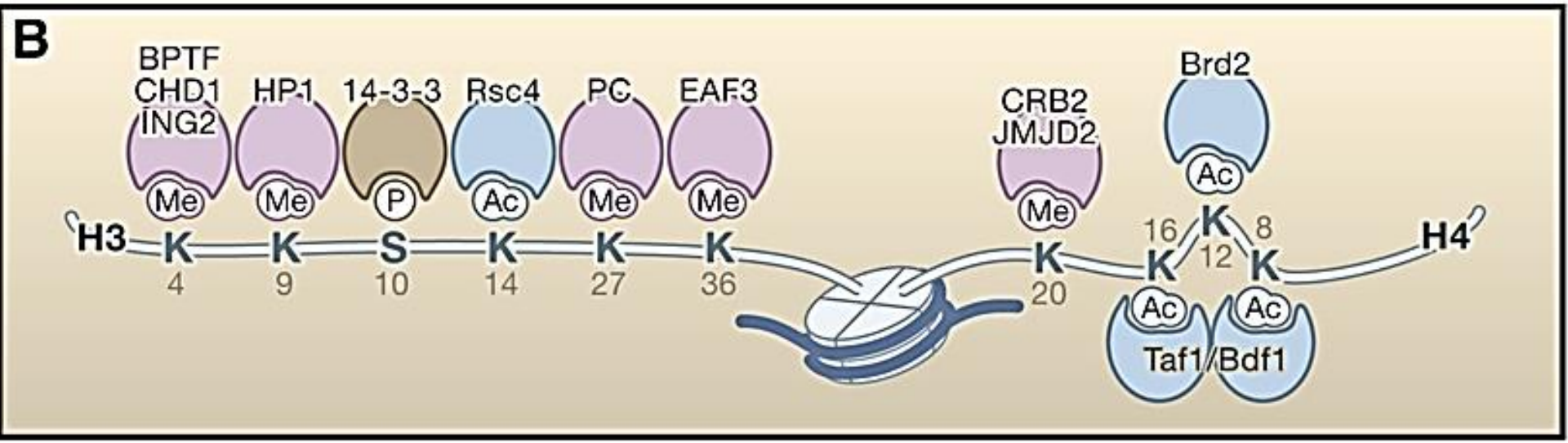
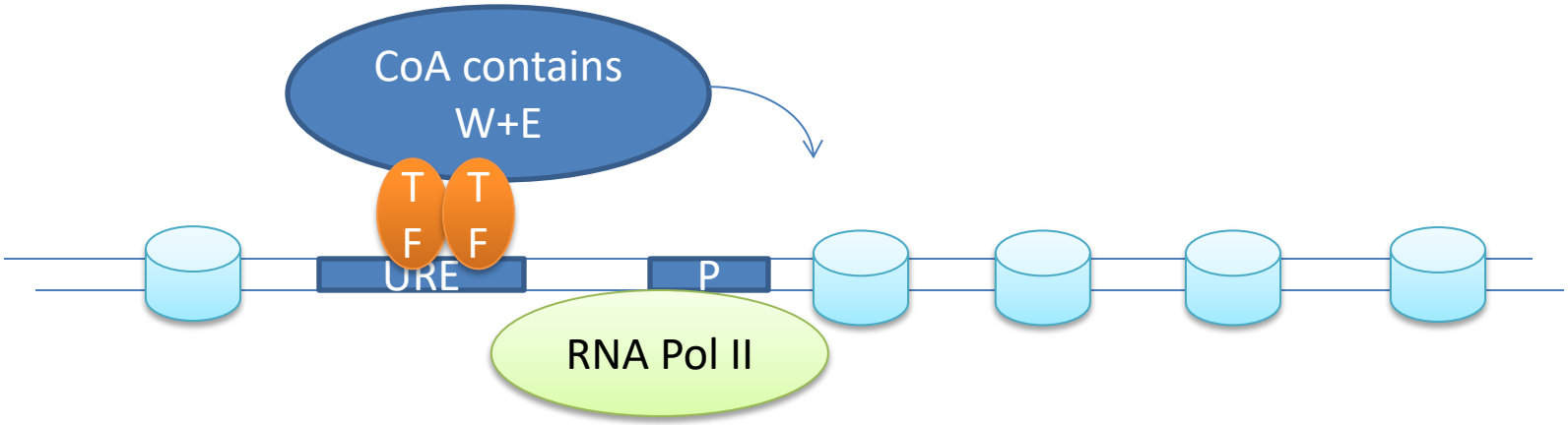


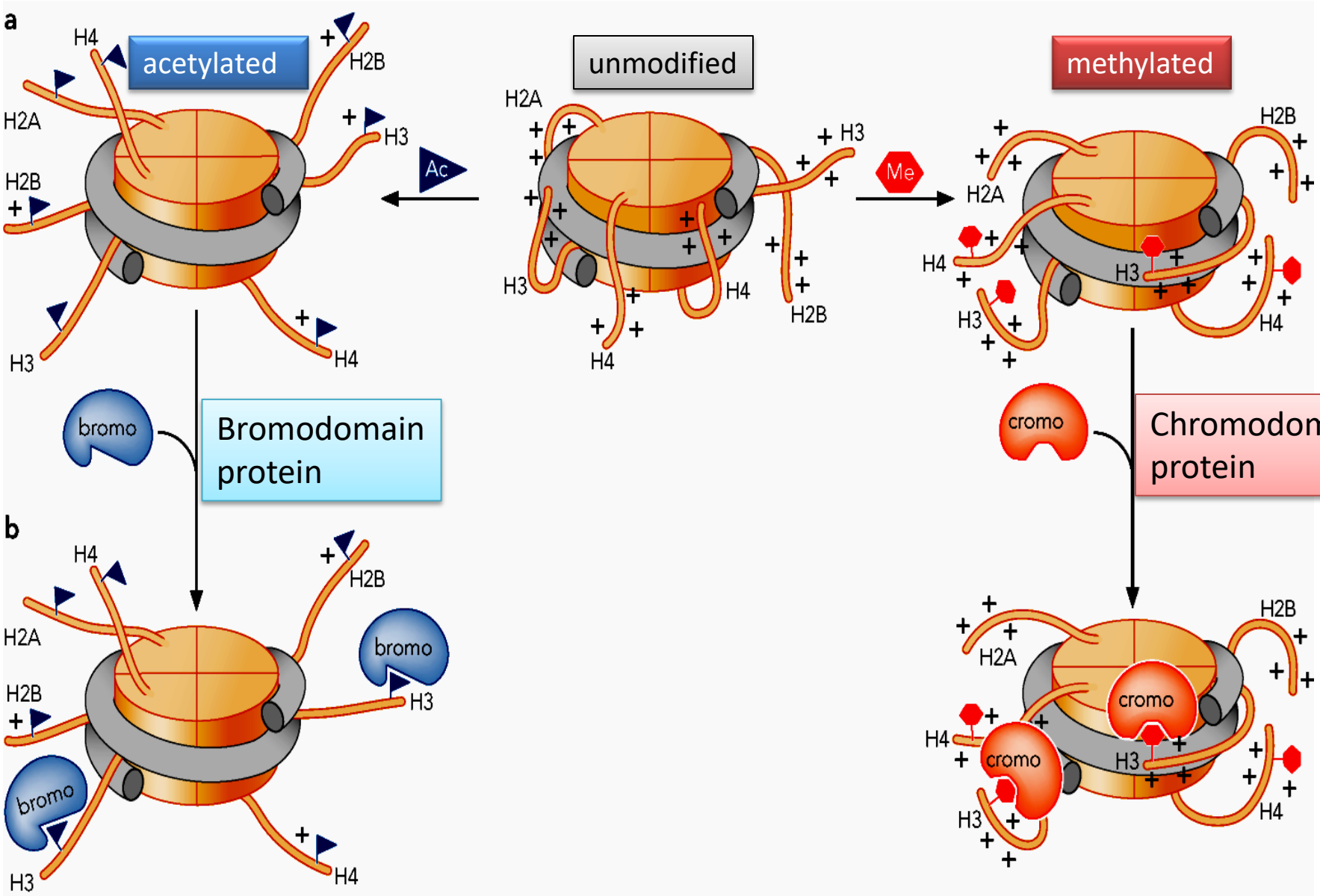
Figure 1. Recruitment of Proteins to Histones
(B) Proteins found that associate preferentially with modified versions of histone H3 and histone H4.

Causes and effects of PTMs

PTM are introduced in nucleosomes when Transcription Factors bind to regulatory elements through recruitment of Co-activatory complexes



PTMs influence nucleosome stability



RNA Pol II leads to specific histone methylation

RNA Pol II has a CTD made by repeated peptide.
(52 repeats of the sequence **Tyr-Ser-Pro-Thr-Ser-Pro-Ser**).

S2
Ⓟ in elongating

S5
Ⓟ in initiation

Binds Set2

Binds Set1

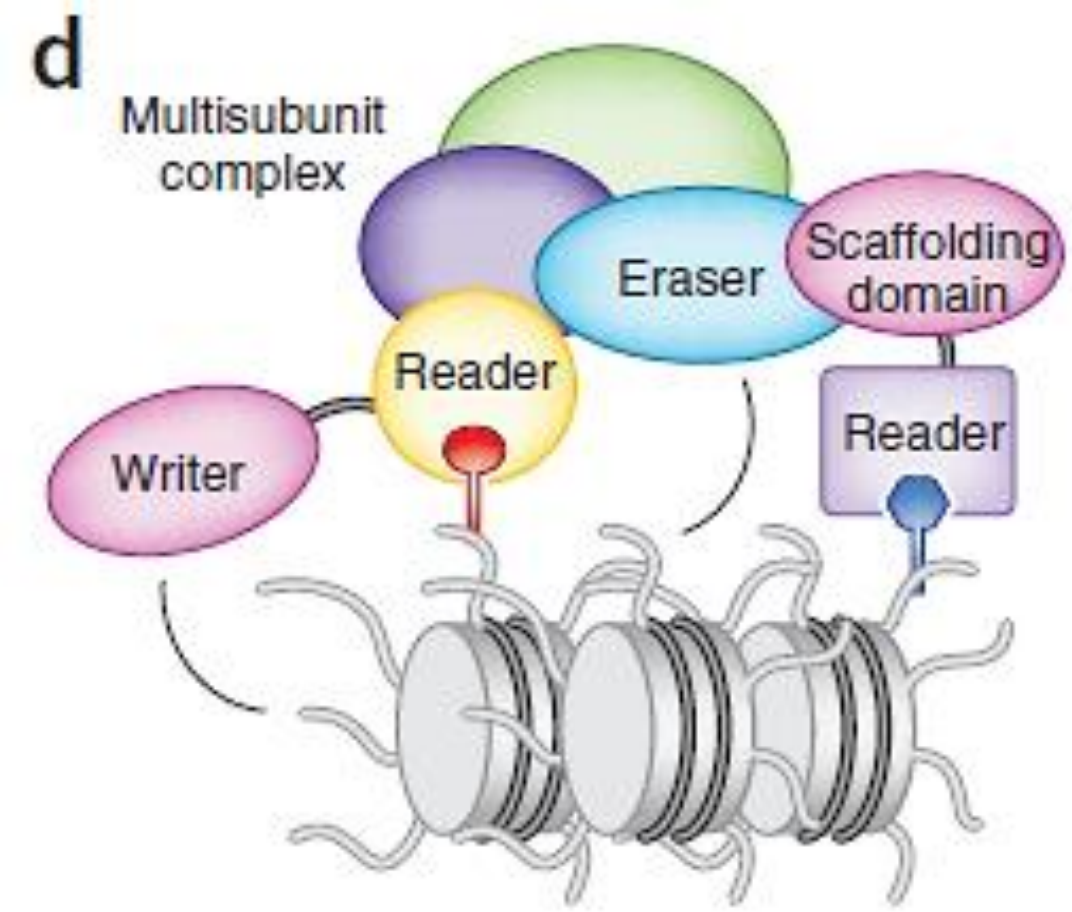
Methylates H3K36

H3K36me3 is found in the gene body (active genes)

Methylates H3K4

H3K4me3 is found around TSS of active genes.

In some cases, the PTM itself guides recruitment of multisubunit complexes, thus explaining «spreading» and maintenance of the chromatic status

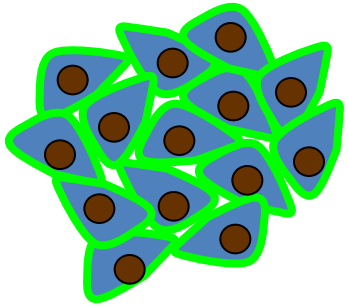


How are chromatin interacting proteins studied?

Same as for Histone PTMs:

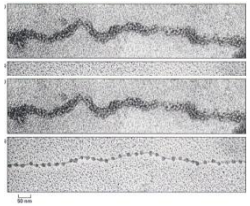
Chromatin Immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP)

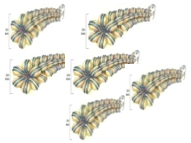


Cross-link with formaldehyde

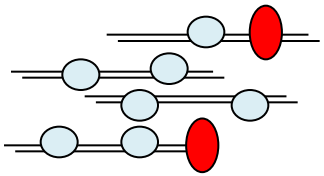
SDS



sonicate

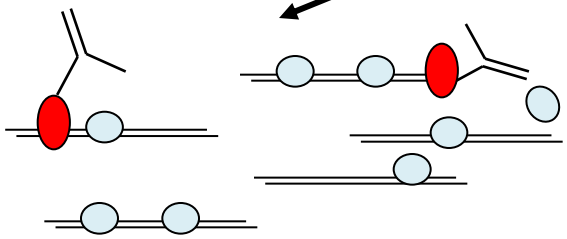


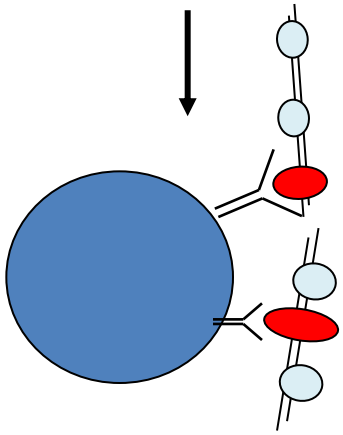
Anti-YFP



Nucleosome carrying your PTM

Other nucleosomes





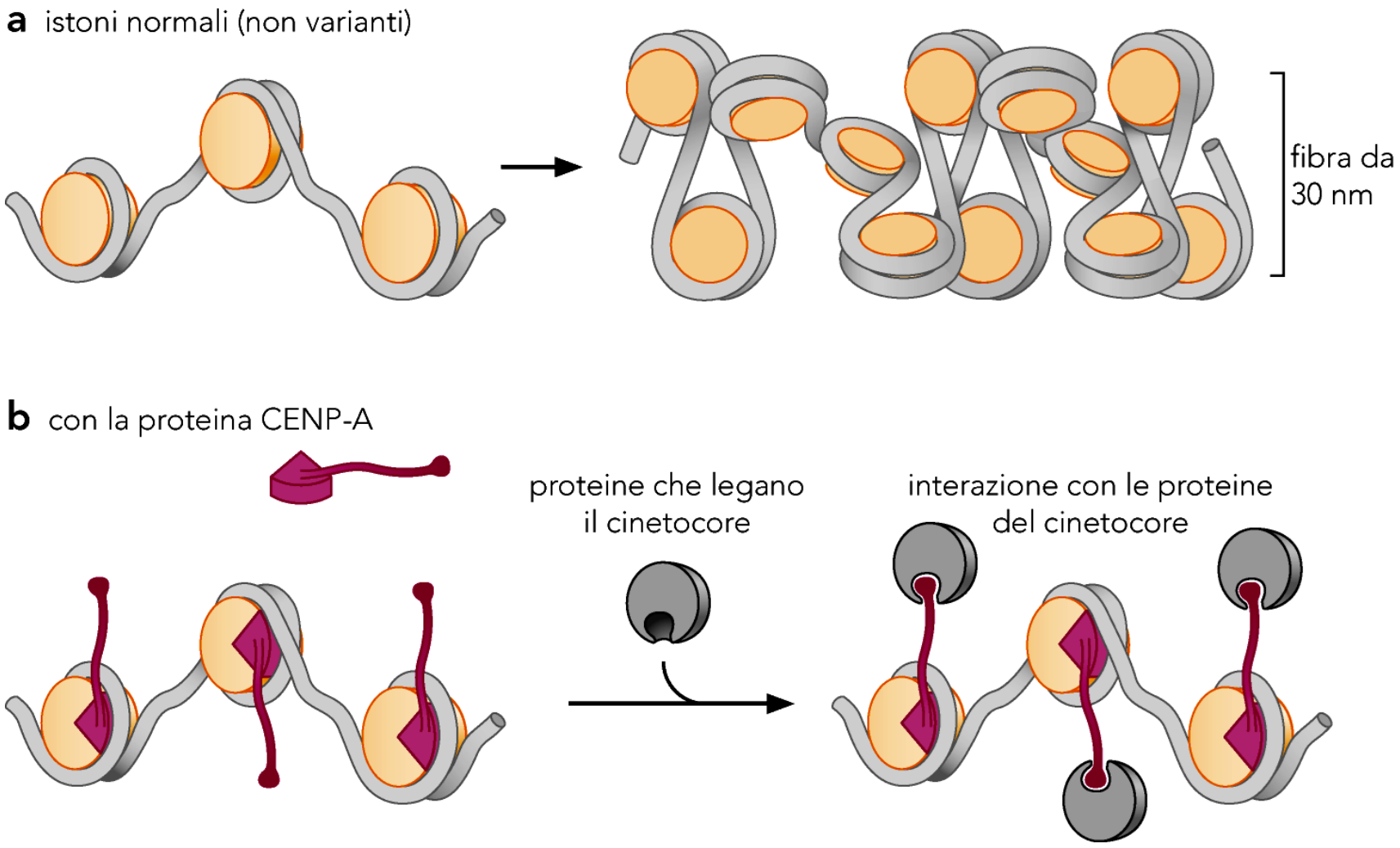
Elute, deproteinize

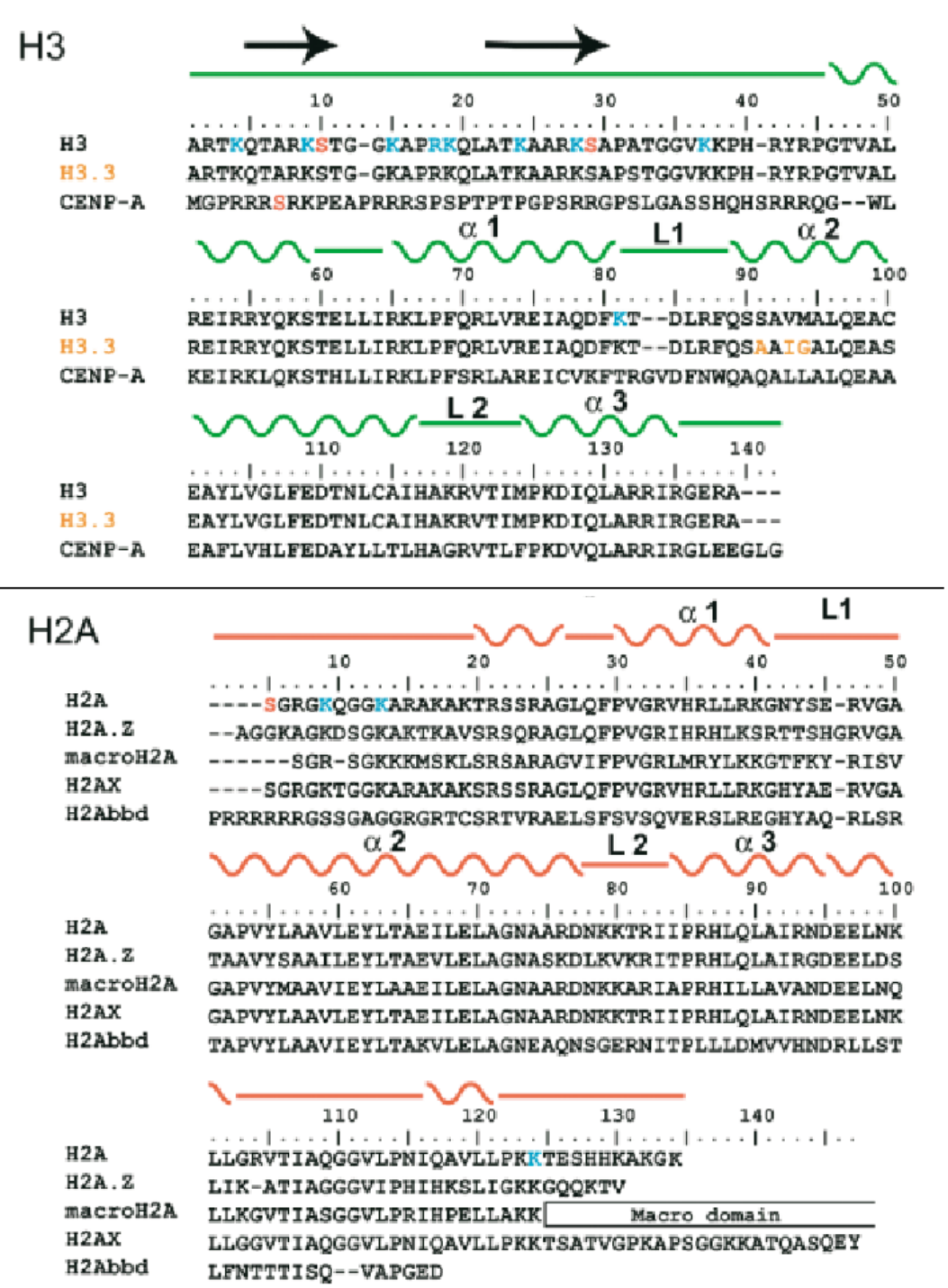
Single locus:
PCR with primers specific for the genome location you wish to interrogate

Genome-wide:
Ligate adapters, PCR-amplify then sequence fragments by NGS

histone isoform

isoform exchange





Sequence Alignment of Variants of Histones H3 and H2A with the Known Secondary Structures of H3 and H2A Depicted on Top.

Upper:
The sequences of the conserved H3.3 and CENP-A variants. H3.3 differs by only a few residues. The arrows above the H3 N-terminal tail indicate the sites that form strands upon binding to chromodomains.

Lower
The sequences of the conserved H2A.Z, macroH2A, H2AX, and H2ABbd variants of H2A. The sequence of H2ABbd is most divergent, while others are closely related with some changes in the turn regions connecting the helices.

CENP-A is an H3-like histone and is found only at centromeres over a stretch of 300-500 Kbp

CENP-A, unlike all other histones, is not replaced by protamines in sperm: chromatin status inheritance.

H3.3 is a variant of H3 showing only 4 aminoacids variation.

H3.3 is deposited in chromatin also outside S-phase

H3.3 replaces H3 carrying H3K9me in re-activated genes

H2A.Z in S. cerevisiae is incorporated near silenced regions and inhibits the spread of heterochromatin.

If interested, see review by Talbert 2010

Modification of DNA: 5-methyl-cytosine

a. **Cytonine-5-methylation** (+hydroxymethylation) is the only epigenetic modification concerning the DNA in higher organisms

b. It occurs mostly at **CpG** dinucleotide (also at CpNpG in some cases, e.g. in plants and *see later in next chapter*)

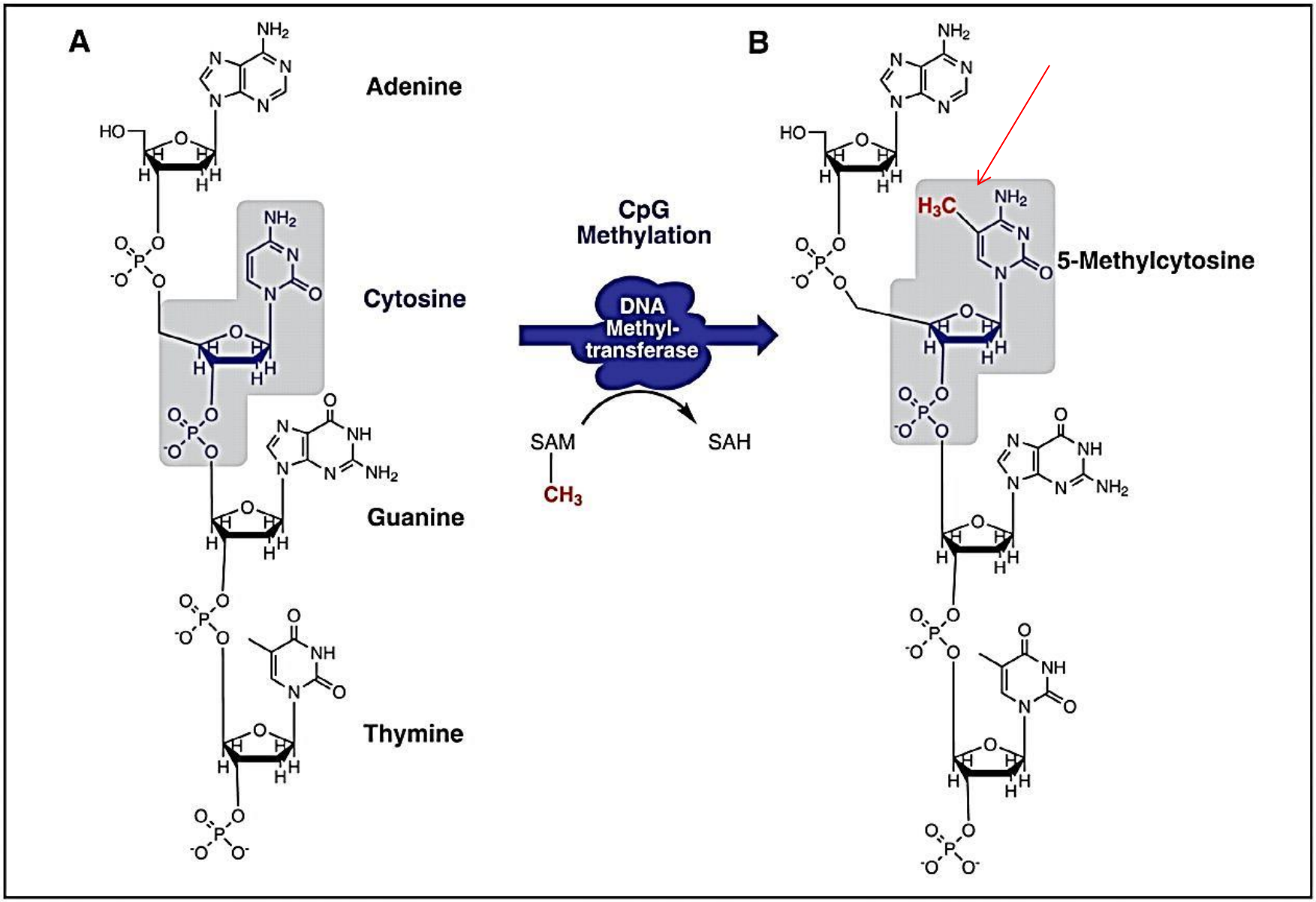
c. Methylation of CpG is observed at regulatory regions of **silenced** genes

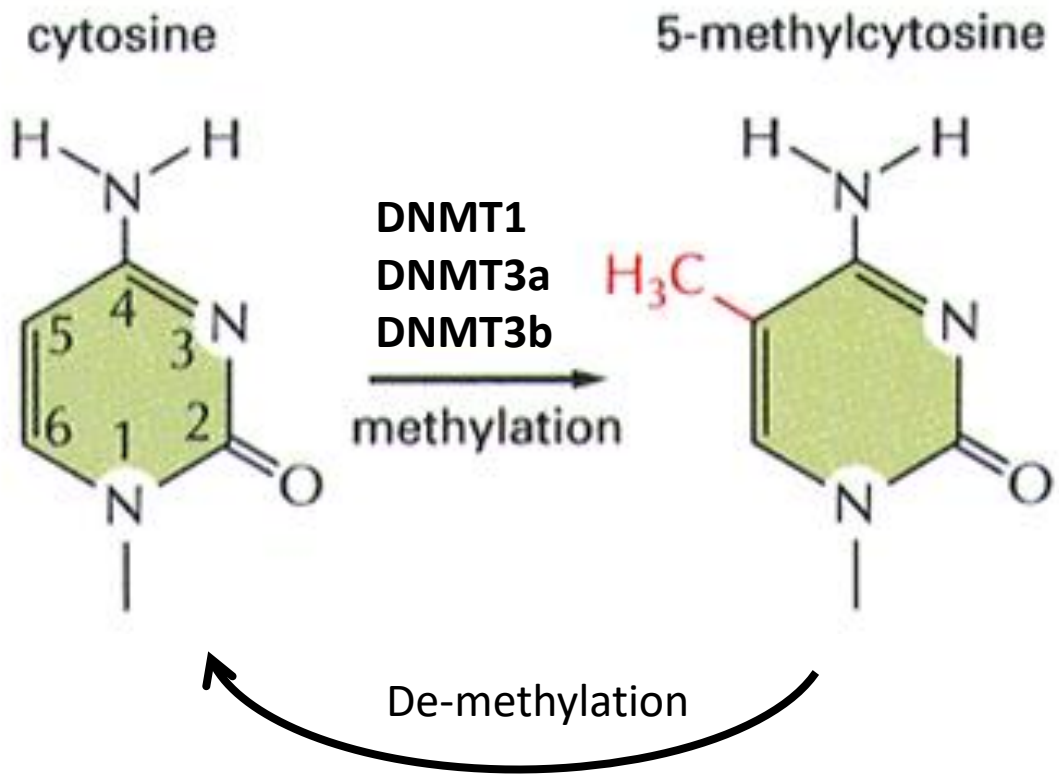
d. Hypermethylation is observed through the inactive X chromosome

e. Housekeeping genes (constitutively expressed) show unmethylated CpG islands at promoters

Important: Cytosine methylation is common in Mammals and Plants, but is not used in the same way in *S. cerevisiae* and *C. elegans*

CpG methylation





1. **Oxidative pathway (Tet enzymes)**
2. **Dilution (DNMT1 off during repeated cell division)**

TF interference

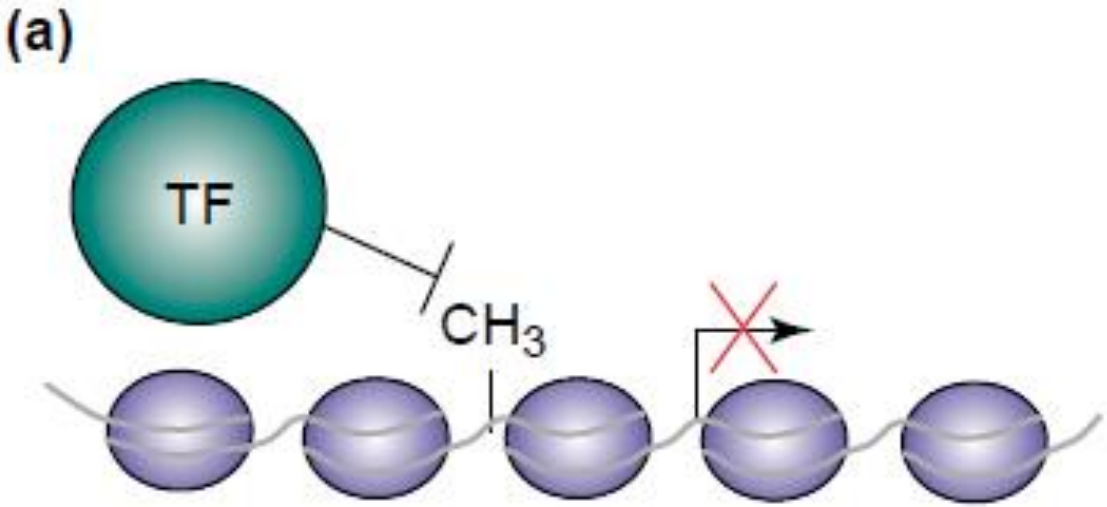


Figure 2. Mechanisms of DNA-methylation-mediated repression.
(**a**) DNA methylation in the cognate DNA-binding sequences of some (*not many, ndr*) transcription factors (TF) can result in inhibition of DNA binding. By blocking activators from binding targets sites, DNA methylation directly inhibits transcriptional activation

From Klose & Bird, 2006

Methyl-C binding proteins

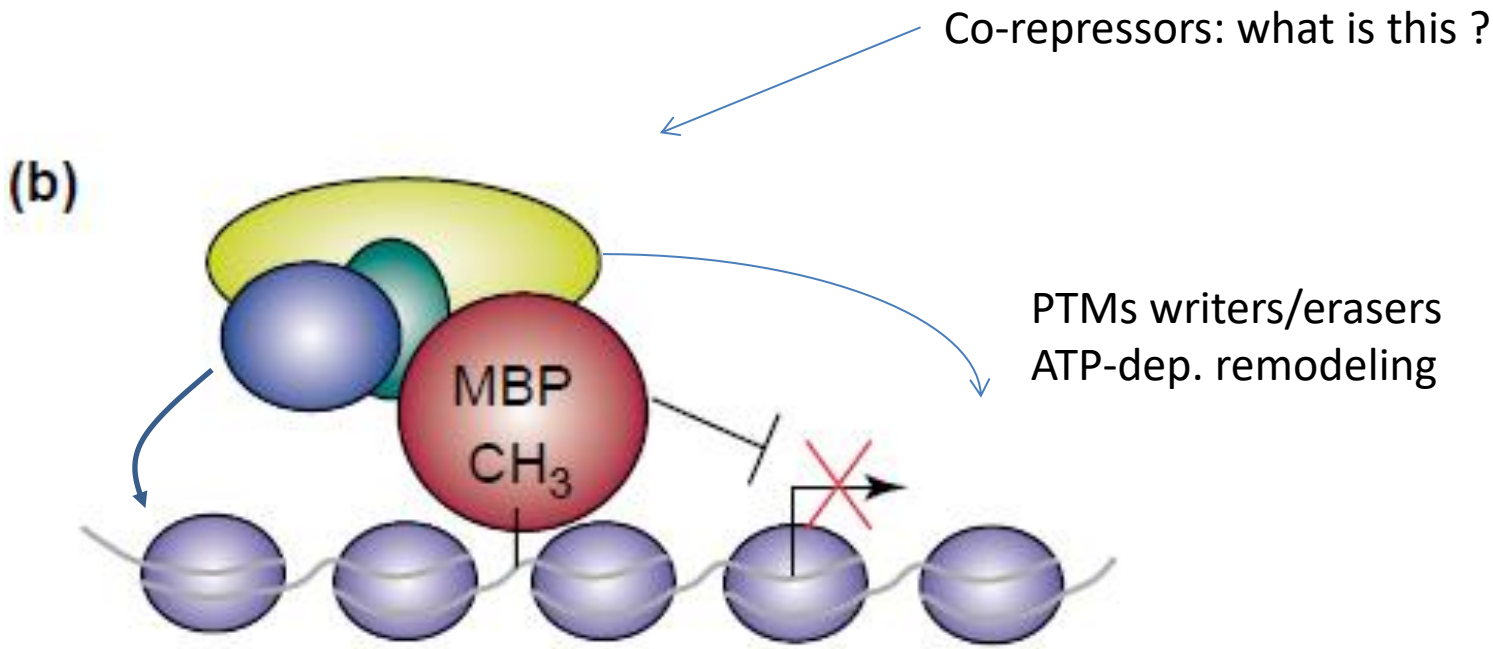


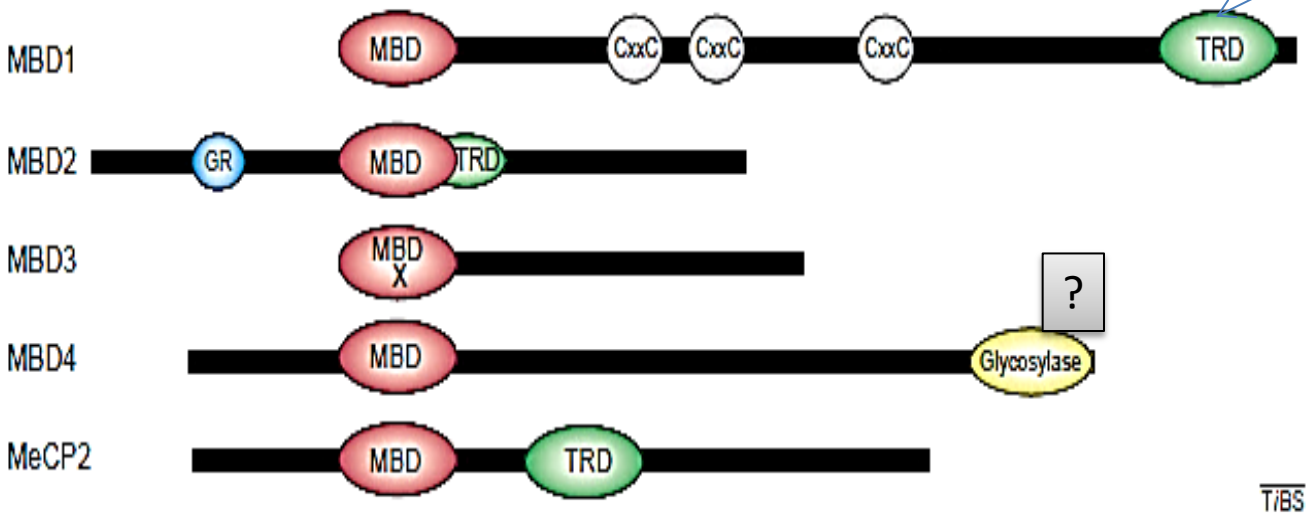
Figure 2. Mechanisms of DNA-methylation-mediated repression.
(**b**) Methyl-CpG-binding proteins (MBPs) directly recognize methylated DNA and recruit **co-repressor molecules** to silence transcription and to modify surrounding chromatin

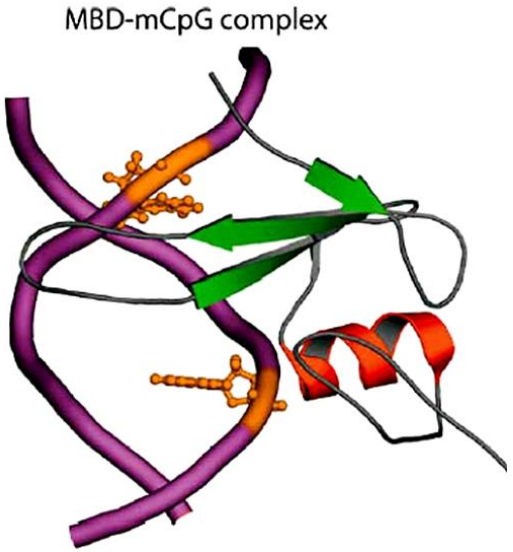
From Klose & Bird, 2006

Readers

Methyl DNA binding proteins have common and specific domains

How do we define a «transcriptional repressor domain ?»





A family of **methyl-CpG-binding proteins** (MBPs). Six mammalian MBPs have been characterized so far:

1. MBD1 uses its methyl-binding domain (MBD) to bind methylated DNA sequences. In addition, MBD1 contains three zinc-binding domains (CxxC), one of which binds specifically to non-methylated CpG dinucleotides, and a C-terminal transcriptional repression domain (TRD).
2. MBD2 possesses an MBD that overlaps with its TRD domain, and a GR repeat at its N terminus.
3. MBD3 contains a well-conserved MBD domain that does not recognize methylated DNA owing to crucial amino acid changes.
4. MBD4 binds methylated DNA through an MBD domain and has a C-terminal glycosylase domain that is important for is function in DNA repair.
5. MeCP2 is the founding member of the MBD protein family and contains a conserved MBD domain and an adjacent TRD domain.
6. The sixth protein, Kaiso, is unrelated and is not shown here.

From Klose & Bird, 2006, TiBS 31: 89-97.

How to measure CpG methylation

1. Methylation-sensitive restriction enzymes
    1. HpaII*/MspI (CCGG)
    2. SmaI*/XmaI (CCCGGG)
    3. McrBC – recognize 2 methylated (G/A)pC (50-1000bp apart)

2. Bisulfite conversion (Methyl-C is not modified , C is converted to U)
    1. Enzymatic analysis (e.g. Cobra-next slide)
    2. Cloning alleles and Sanger sequencing

Most common methods to detect CpC methylation are based on **bisulfite**, which converts C (but not methyl-C) to Uridine. Uridine in DNA is then replaced by T in the following PCR.

Then one can sequence frament(s) to detect changed bases.

In some cases, when a single CpG in a sequence has to be detected, we can use also bisulfite conversion followed by simple restriction analysis + PCR. Example: