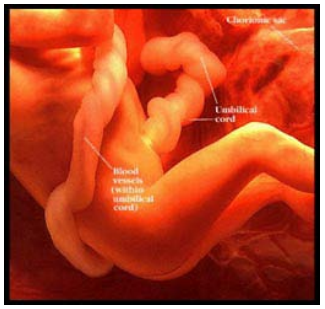

Tyr-Cys-Arg-Ala-Asn-Gl

Gene Expression in Eukaryotes

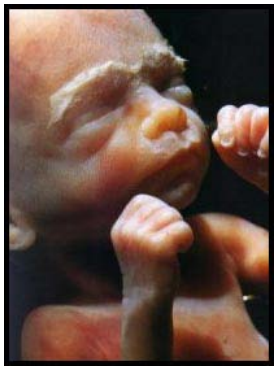
真核生物基因表达调控

复旦大学分子医学教育部重点实验室

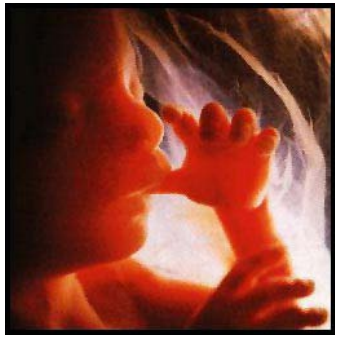
马 端



25-28w

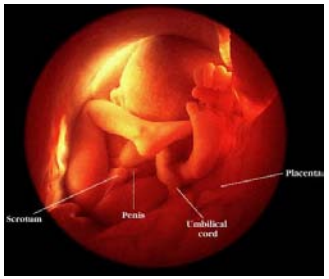


24w

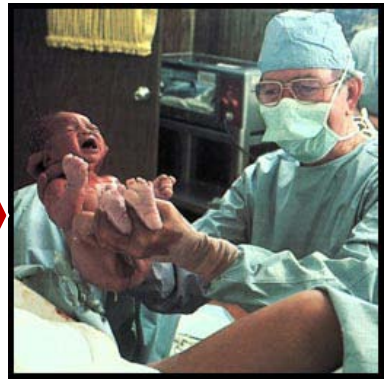


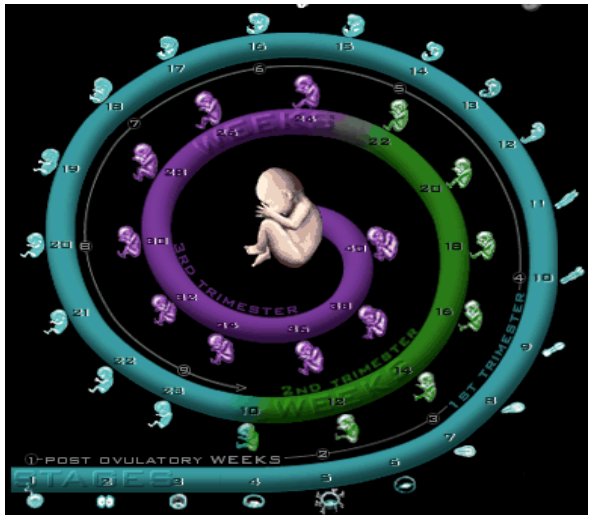
13-16w

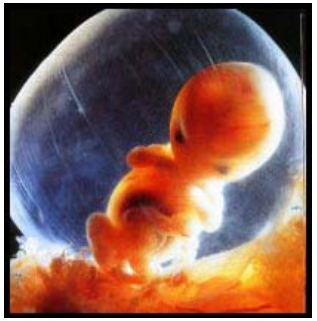
9-12w



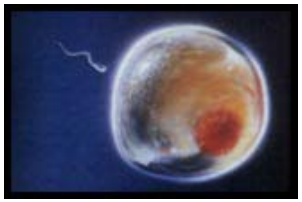
29-32w



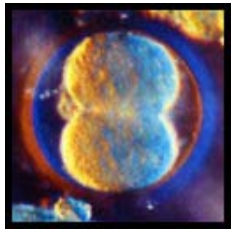




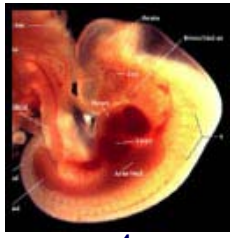
8w



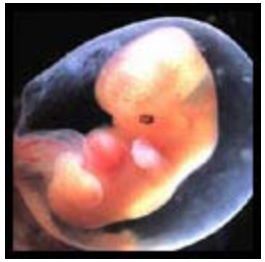
Fertilization



zygote

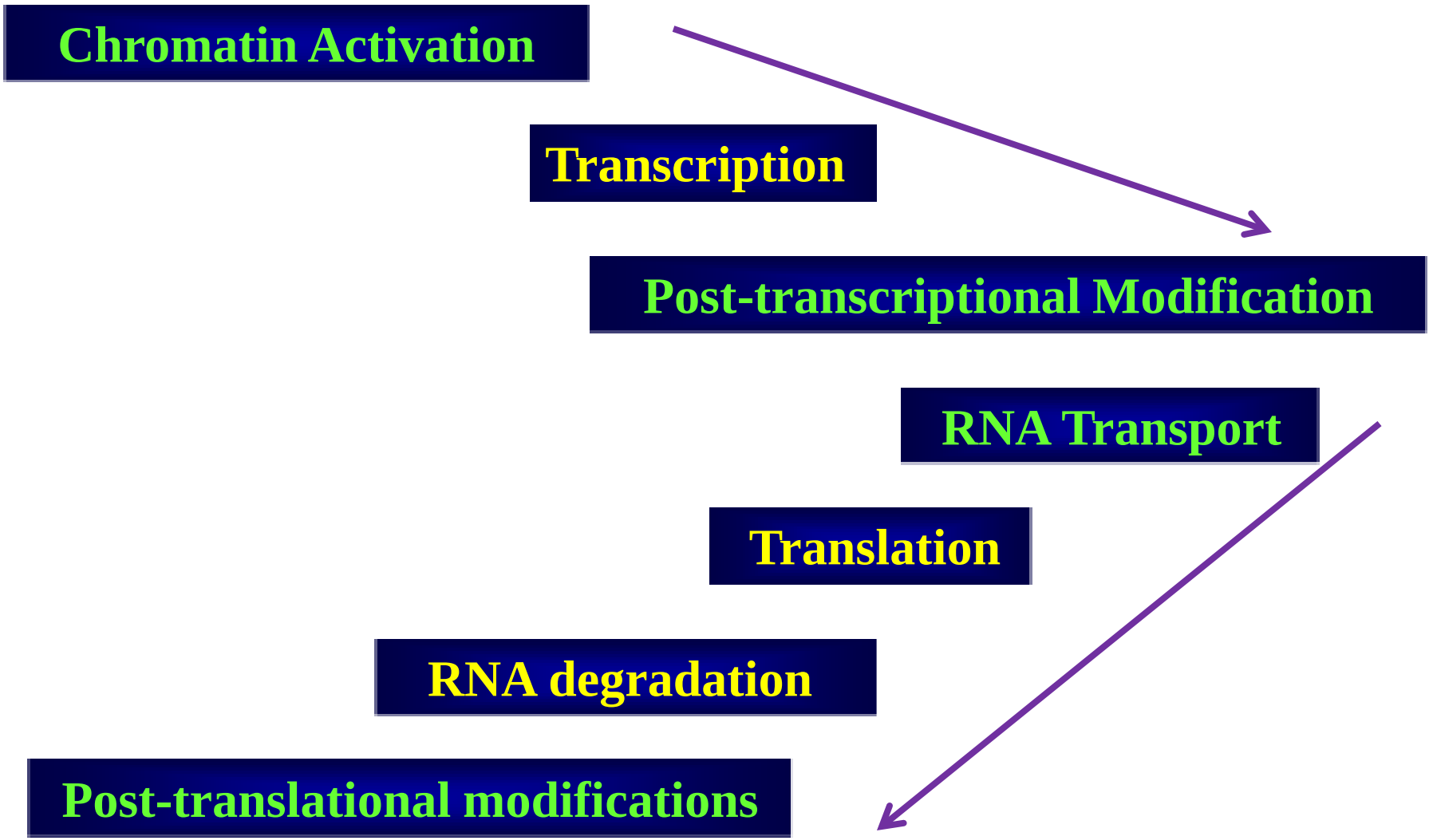


4w



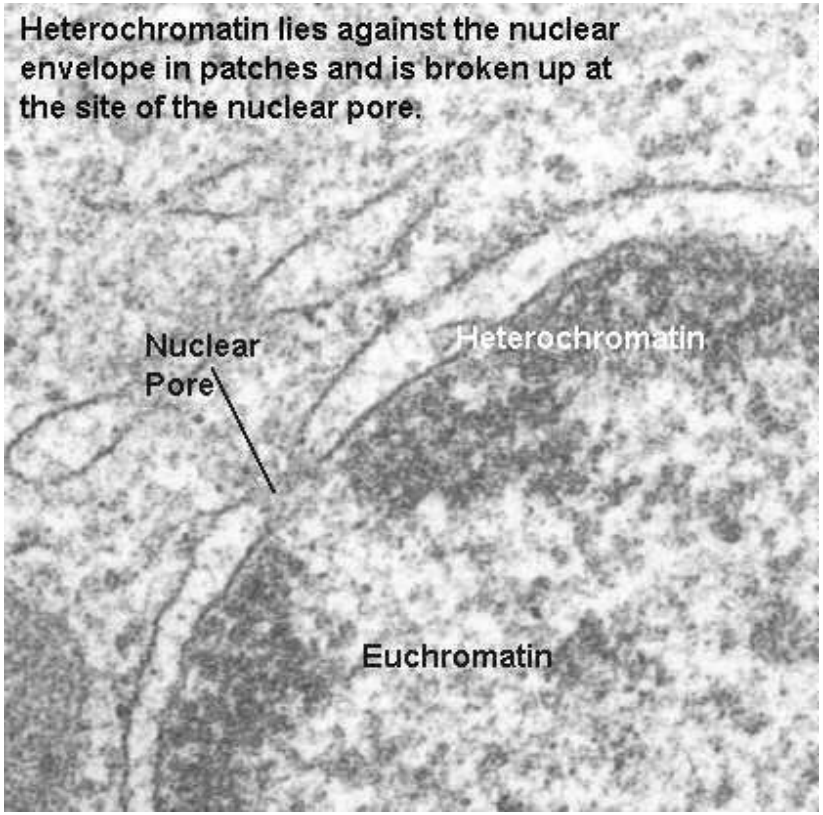
7w

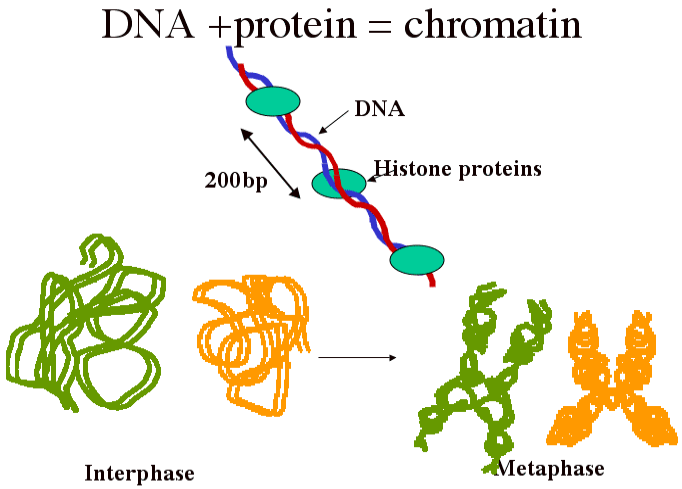
Regulation Points of Gene Expression

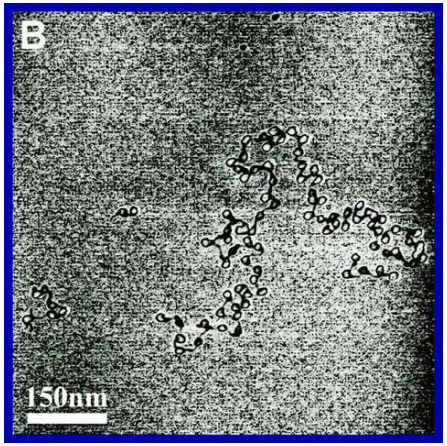


Chromatin Activation

染色质活化

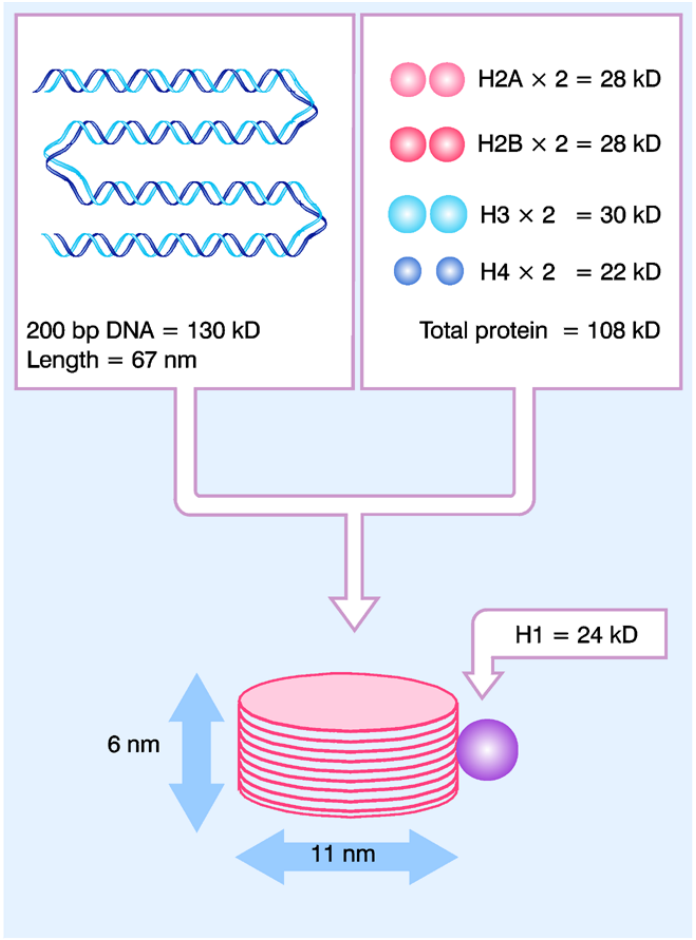


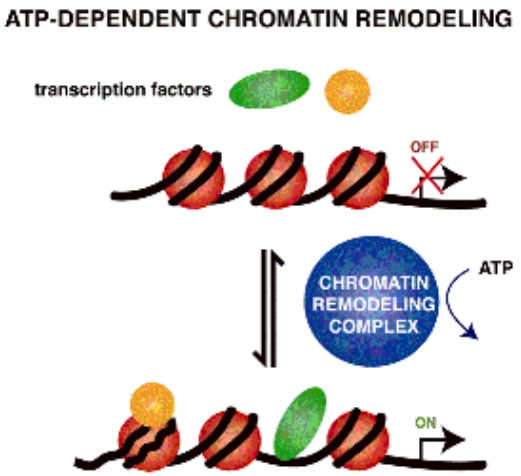


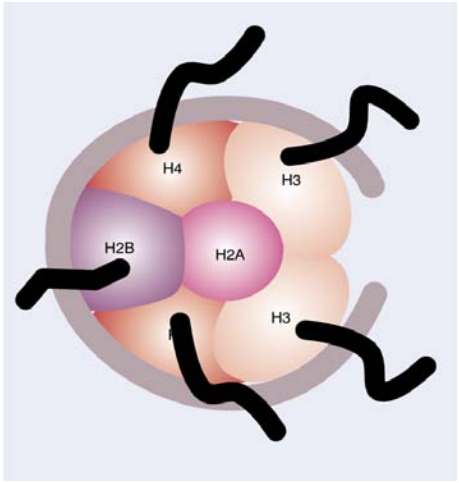


Chromatin-remodeling Complexes

染色质重塑复合物



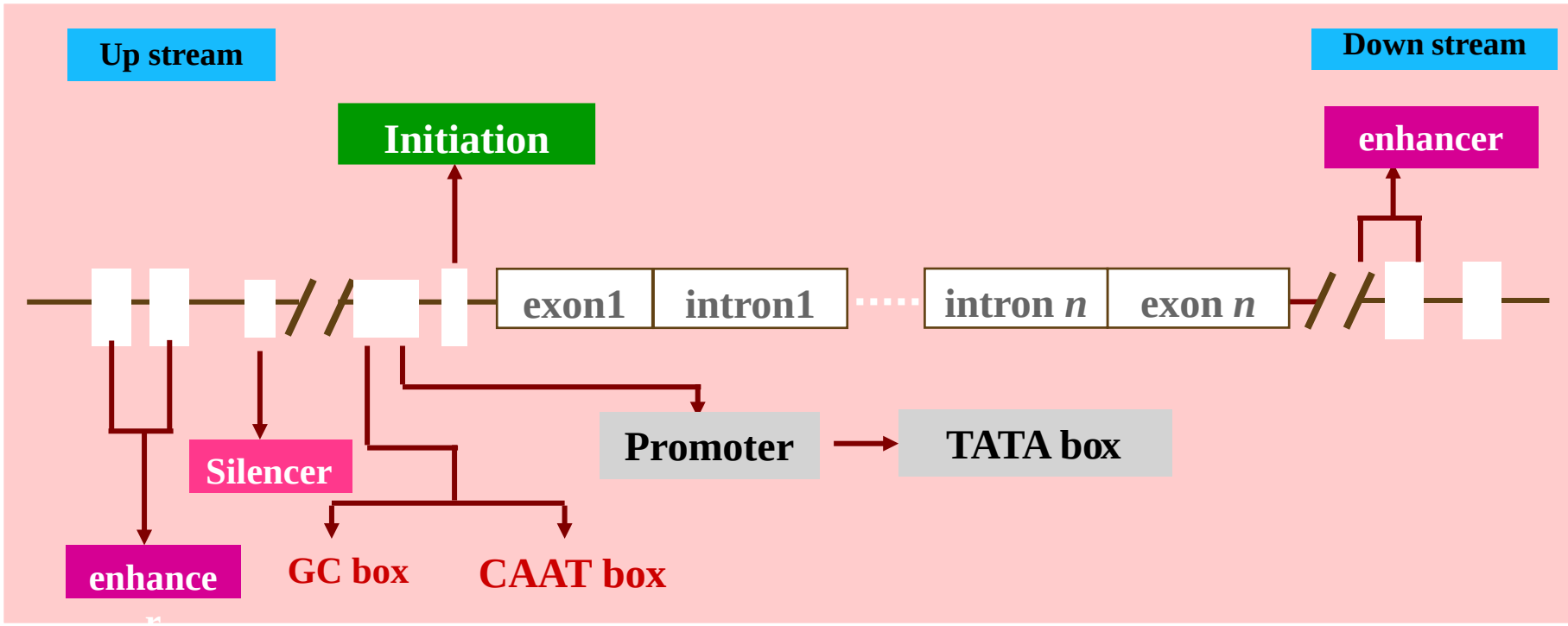




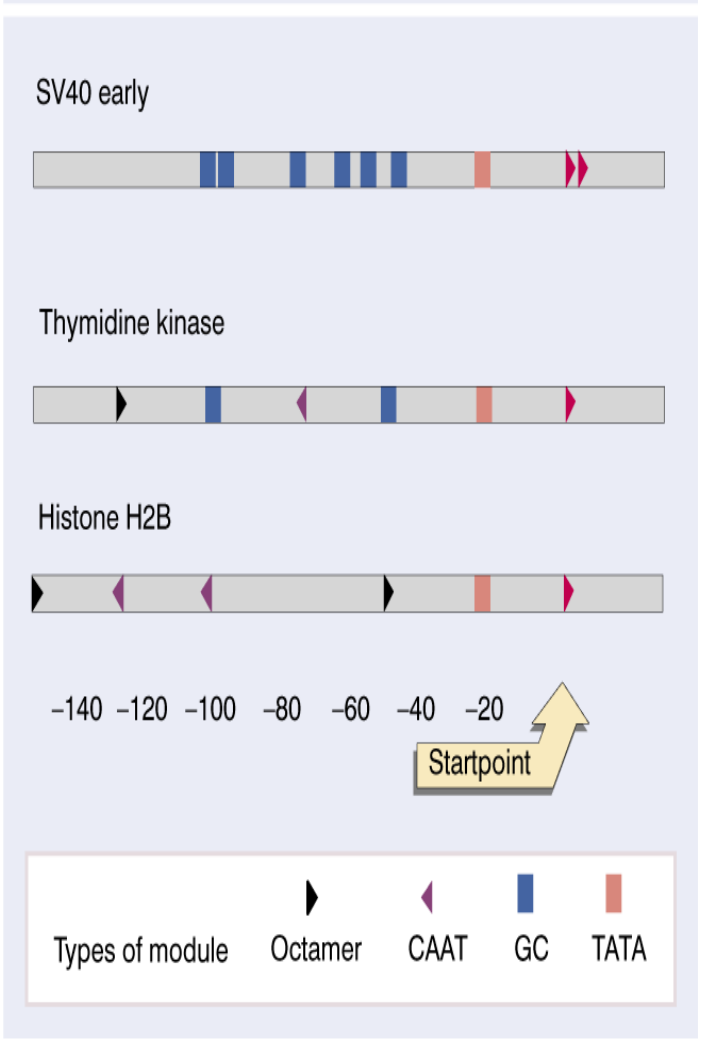
Structure of nucleosome（核小体）

Control of Transcription Initiation

转录起始调控



Some Cis-elements in Promotor



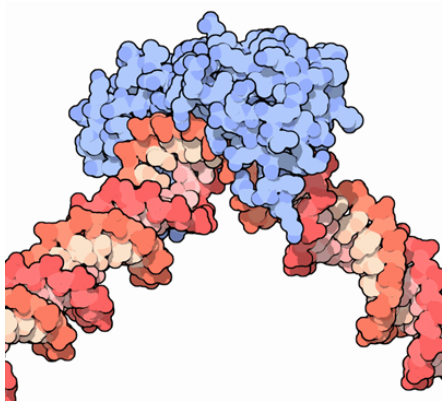
TATA box (TATAAAA) is at -25 and is involved in positioning the enzyme for correct initiation.

GC box is at -90 contains the sequence GGGCGG and is recognized by the factor SP1.

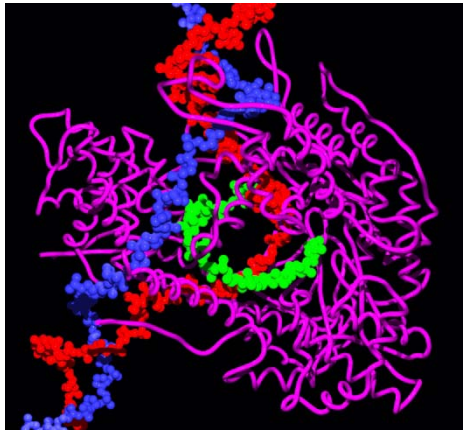
CAAT box (CCAATCT) is at –75 and is recognized by a large group of transcription factors and plays a strong role in determining the efficiency of the promoter.

Transcription Factors（转录因子）

Transcription factor (sequence-specific DNA binding factor) is a protein that binds to specific sequences of DNA and thereby controls the transfer (or transcription) of genetic information from DNA to RNA.



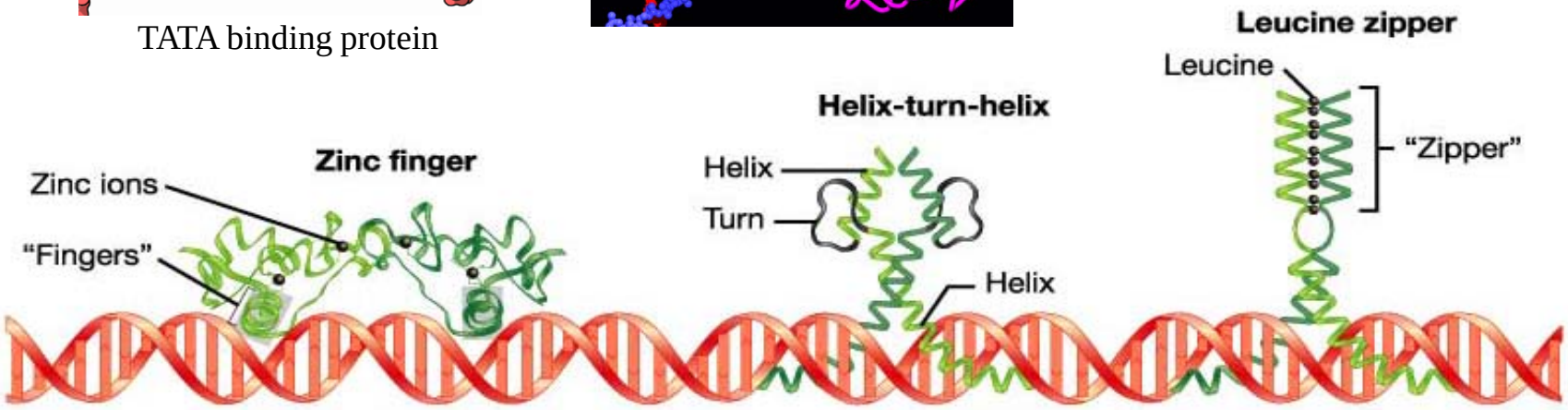
TATA binding protein



General Transcription Factor

T7 RNA polymerase

Specific Transcription Factor

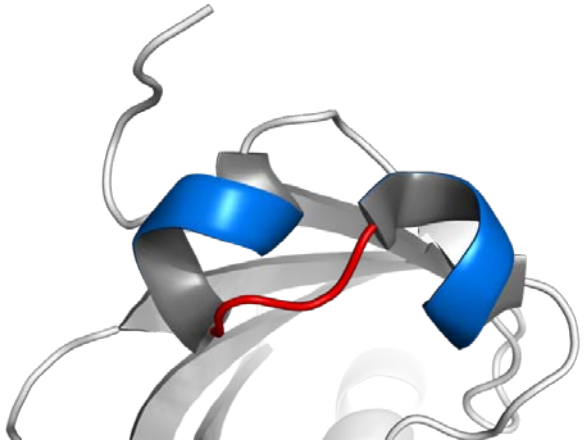


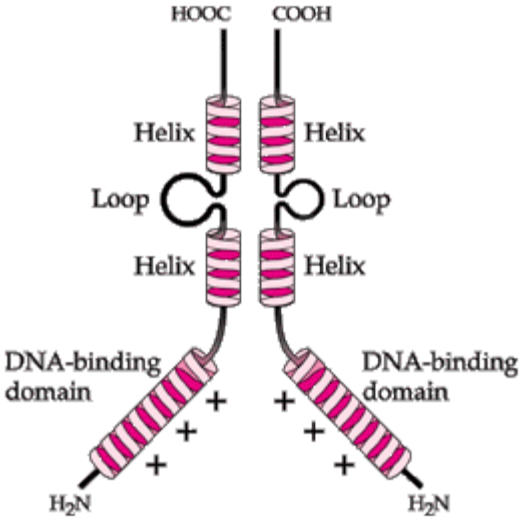
Structure of Transcription Factors

- **DNA-binding domain(DBD):** attach to specific sequences of DNA.
- **Trans-activating domain** (**TAD**)**:** contain binding sites for other proteins.
  - **glutamine-rich activation domains**
  - **proline-rich activation domains**
  - **acidic activation domains**
- **Signal sensing domain** (**SSD**) (ligand binding domain) which senses external signals and in response transmit these signals to the rest of the transcription complex resulting in up or down regulation of gene expression.
- **Protein-protein interaction domain**

Basic helix-loop-helix (bHLH)

碱性螺旋－环－螺旋基序





bHLH structural motif (模体，基序）: two α helices connected by a loop

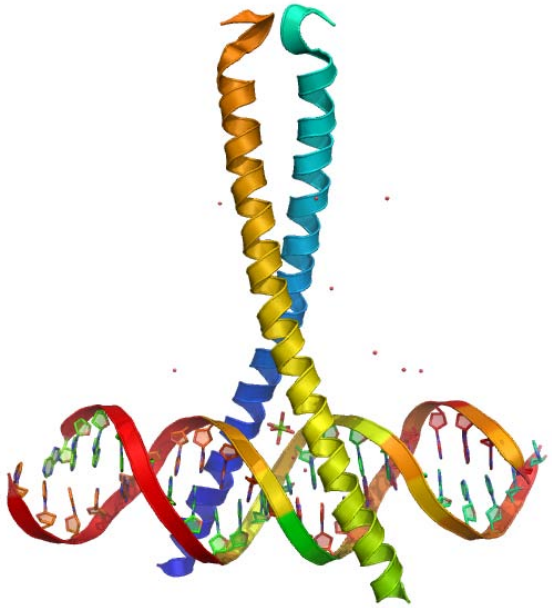
Structural motif: a pattern in a protein structure formed by the spatial arrangement of amino acids.

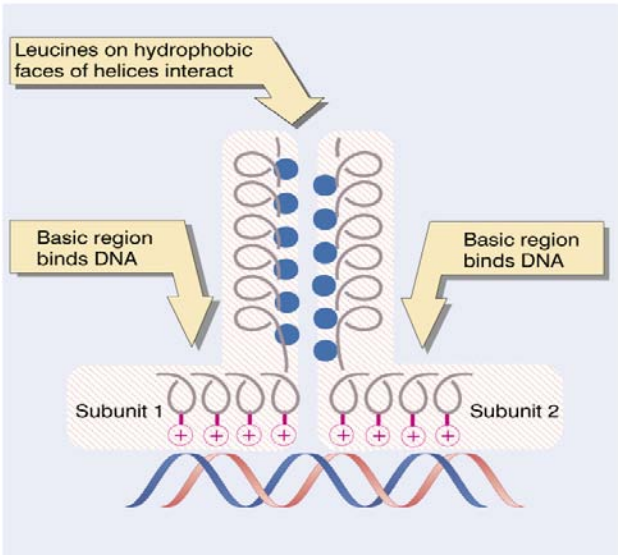
In general, transcription factors including this domain are dimeric, each with one helix containing basic amino acid residues that facilitate DNA binding.

bHLH proteins typically bind to a consensus sequence called an E-box which is CACGTG , however some bHLH transcription factors bind to different sequences, which are often similar to the E-box.

Basic Leucine Zipper Domain (bZIP domain)

碱性亮氨酸拉链结构域



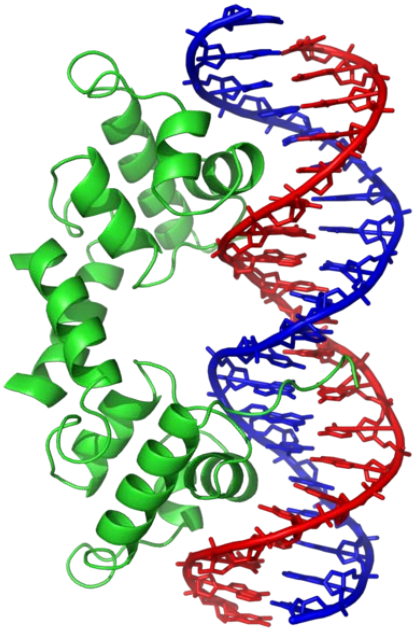


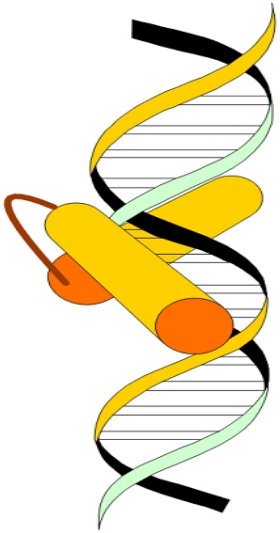
Protein domain: an autonomously folding functional module of a protein.

One part of **bZIP domain** contains a region that mediates sequence specific DNA binding properties and the leucine zipper that is required for the dimerization of two DNA binding regions. The DNA binding region comprises a number of basic aminoacids such as arginine and lysine.

helix-turn-helix (HTH)

螺旋－回折－螺旋

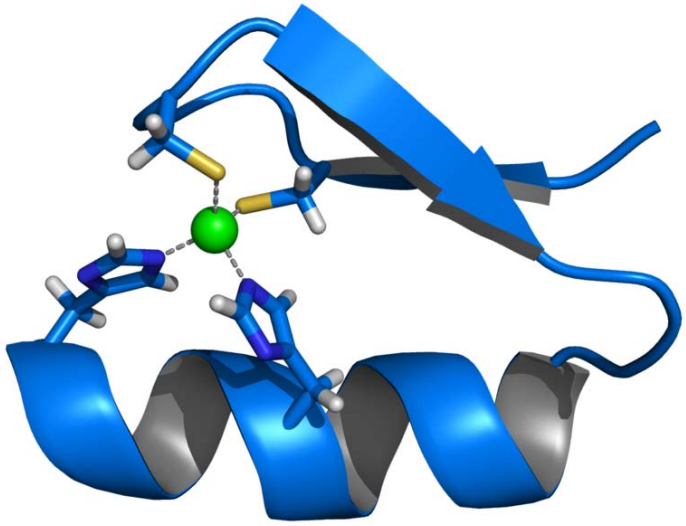


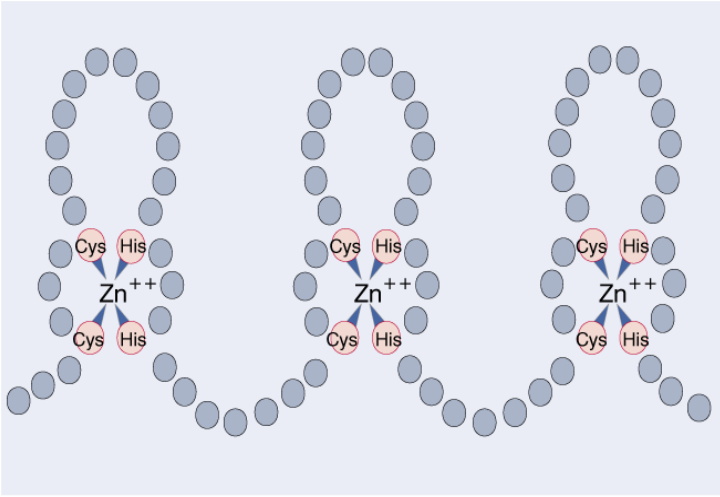


HTH is composed of two α helices joined by a short strand of amino acids. Recognition and binding to DNA is done by the two α helices. One α helices binds to the major groove of DNA through a series of hydrogen bonds and various Van der Waals interactions with exposed bases. The other α helix stabilizes the interaction between protein and DNA, but does not play a particularly strong role in its recognition.

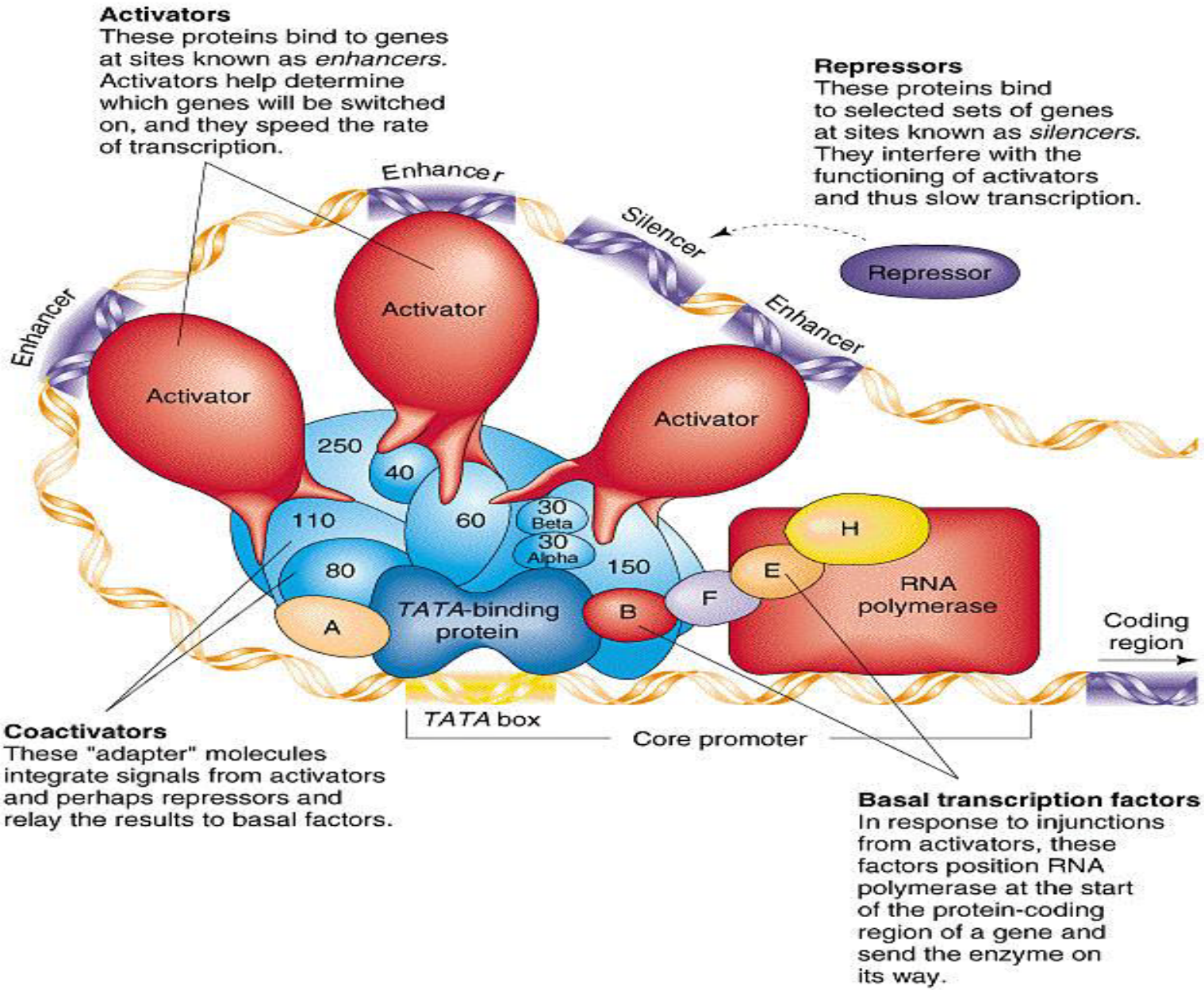
Zinc Finger

锌 指



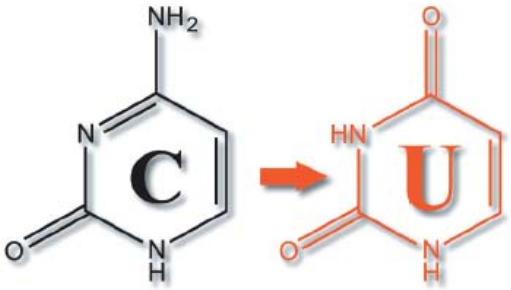


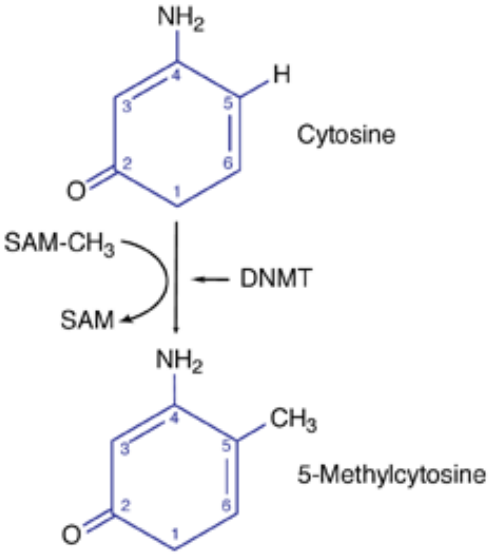
A zinc finger consists of two antiparallel β strands, and an α helix. The zinc ion is crucial for the stability. Transcription factors with zinc fingers interact with the major groove along the double helix of DNA in which case the zinc fingers are arranged around the DNA strand in such a way that the α-helix of each finger contacts the DNA, forming an almost continuous stretch of α-helices around the DNA molecule.


Activators
These proteins bind to genes at sites known as enhancers. Activators help determine which genes will be switched on, and they speed the rate of transcription.
Repressors
These proteins bind to selected sets of genes at sites known as silencers. They interfere with the functioning of activators and thus slow transcription.
Enhancer
Silencer
Repressor
Enhancer
Enhancer
Activator
Activator
Activator
250
40
110
60
30 Beta
30 Alpha
150
80
A
TATA-binding protein
B
F
E
H
RNA polymerase
Coding region
TATA box
Core promoter
Coactivators
These "adapter" molecules integrate signals from activators and perhaps repressors and relay the results to basal factors.
Basal transcription factors
In response to injunctions from activators, these factors position RNA polymerase at the start of the protein-coding region of a gene and send the enzyme on its way.

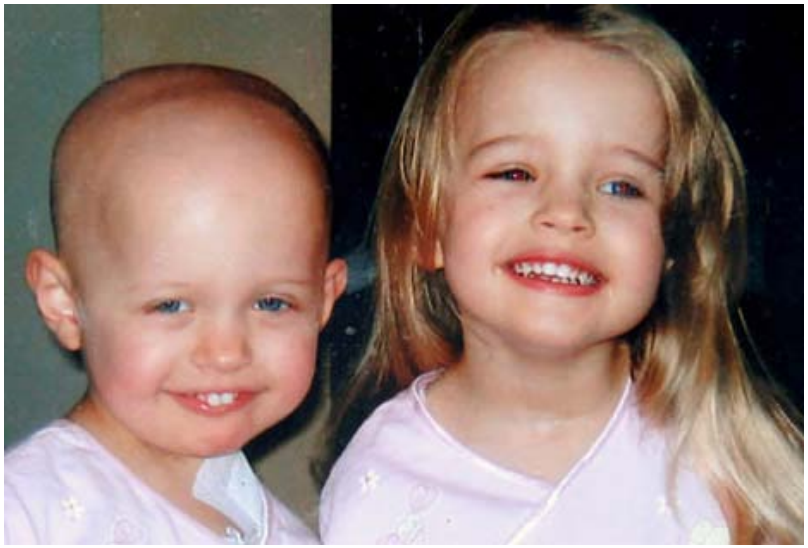
DNA Methylation

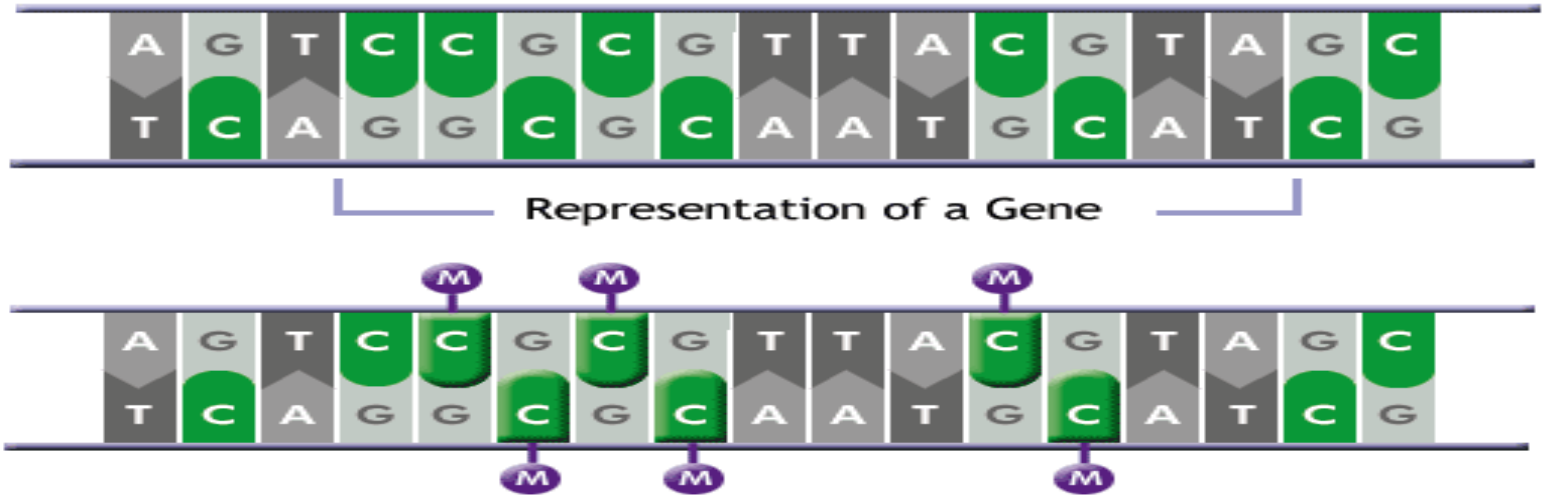
DNA甲基化



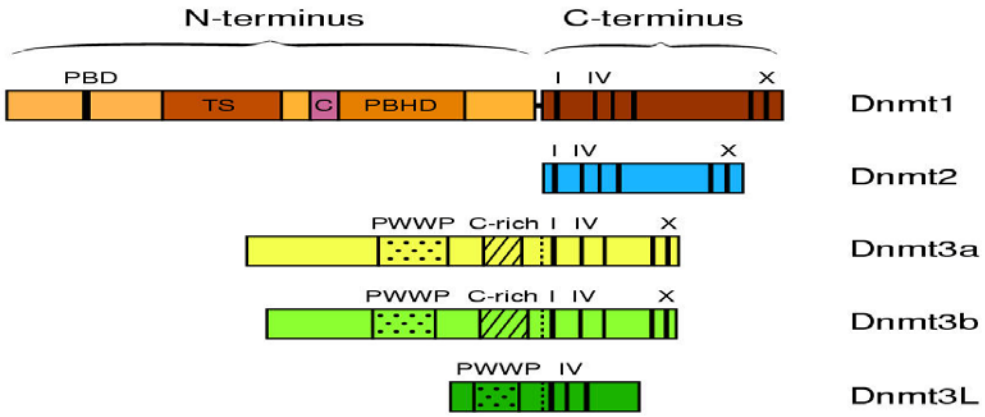


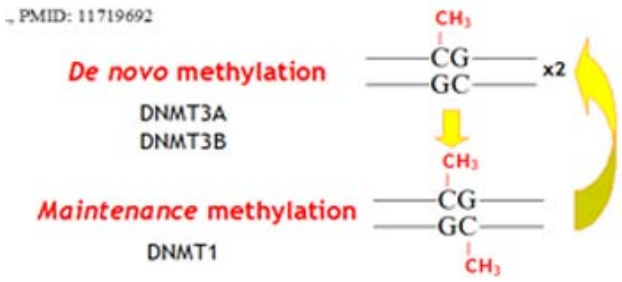
DNA methylation is the covalent addition of a methyl group (CH_3) to cytosine within the context of CpG dinucleotide (CpG island).



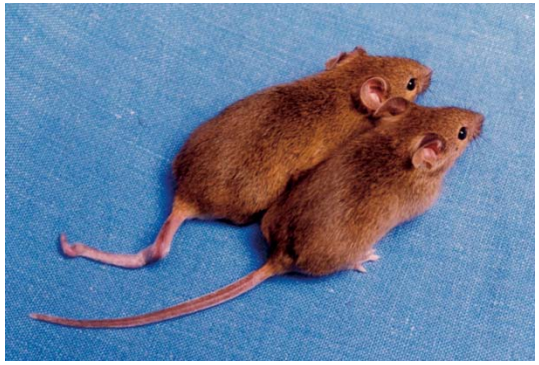


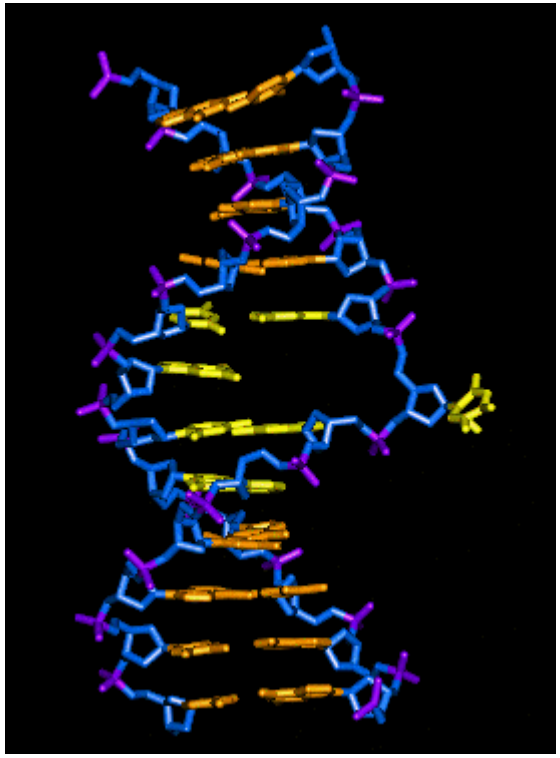
DNA Methylation is a modification of one of DNA`s four bases: cytosine.
This is represented with the addition of the "M" symbol for the methyl group.

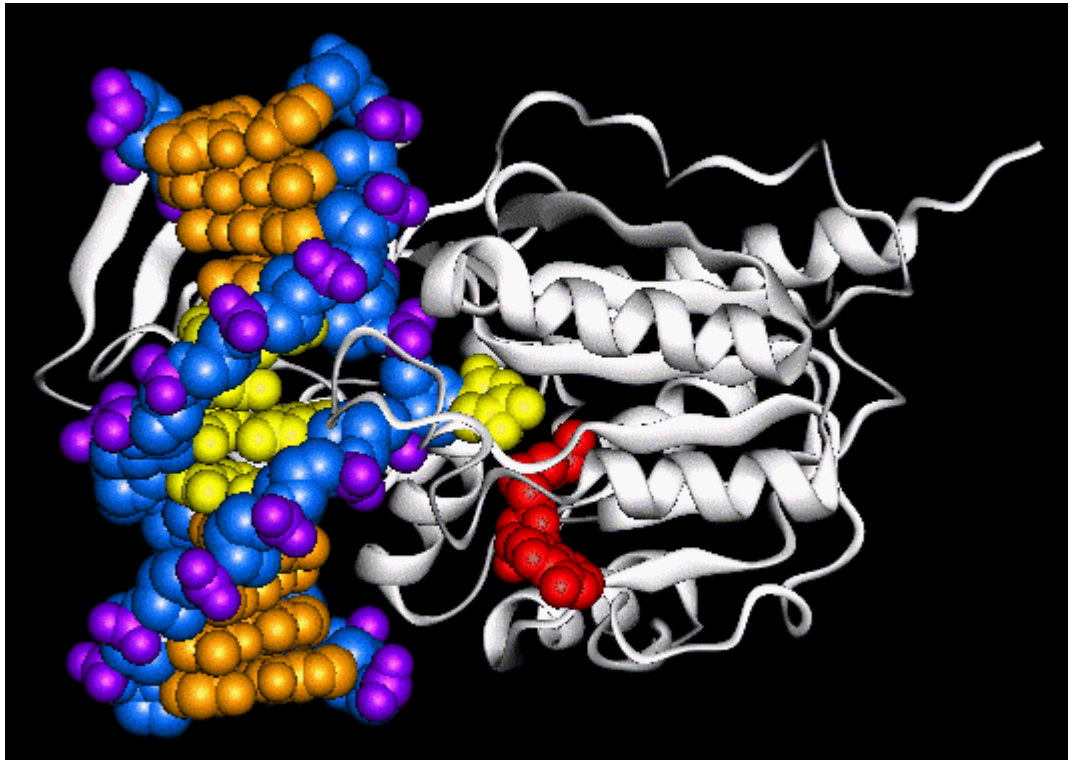




DNA Methyltransferases，DNA甲基转移酶

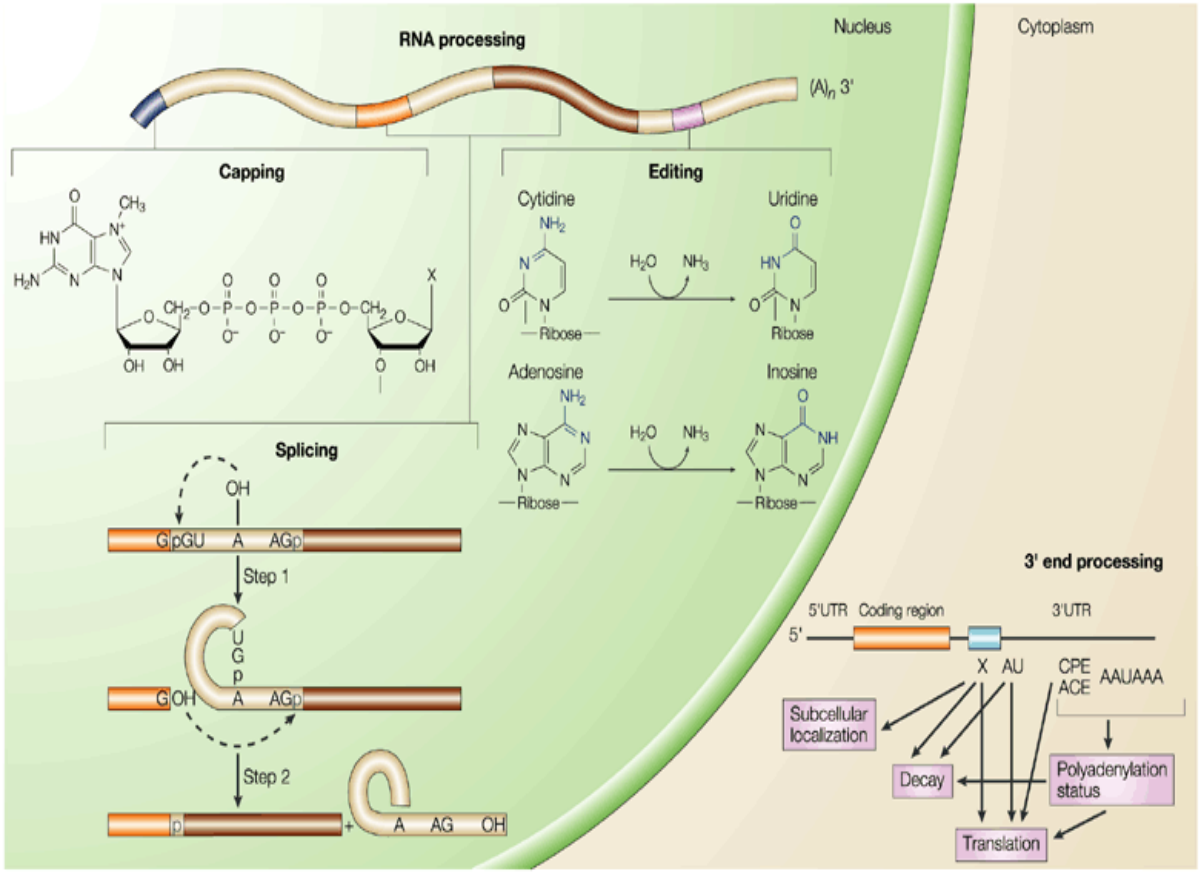






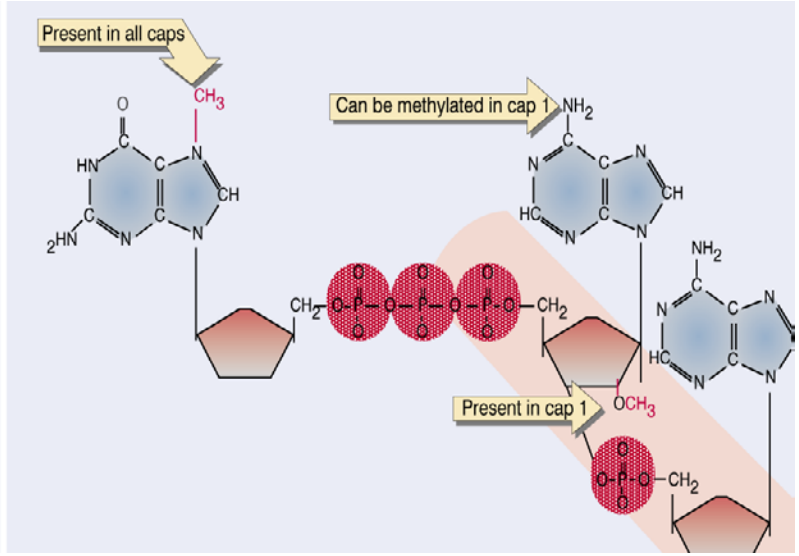
The flipping out of the Cytosine by the methyltransferase from the DNA duplex.

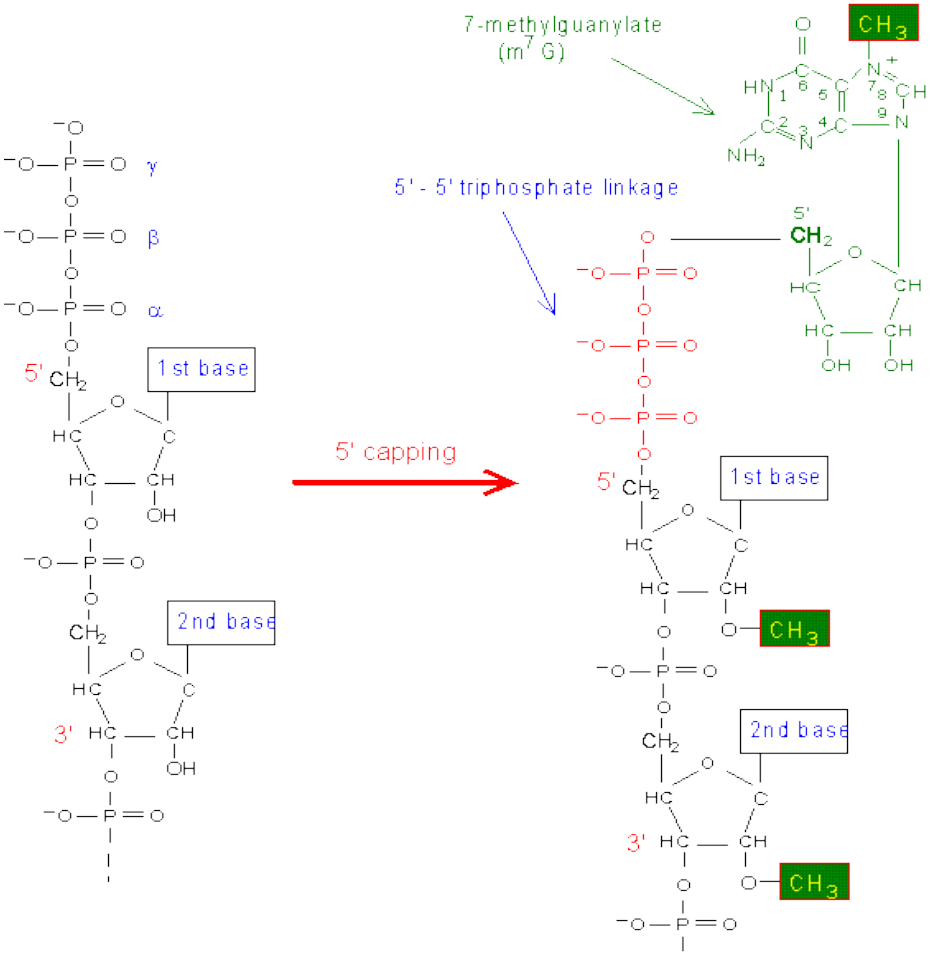
Post-transcriptional Modification



Post-transcriptional modification is a process in eukaryotic cellsby which primary transcript RNA is converted into mature RNA.

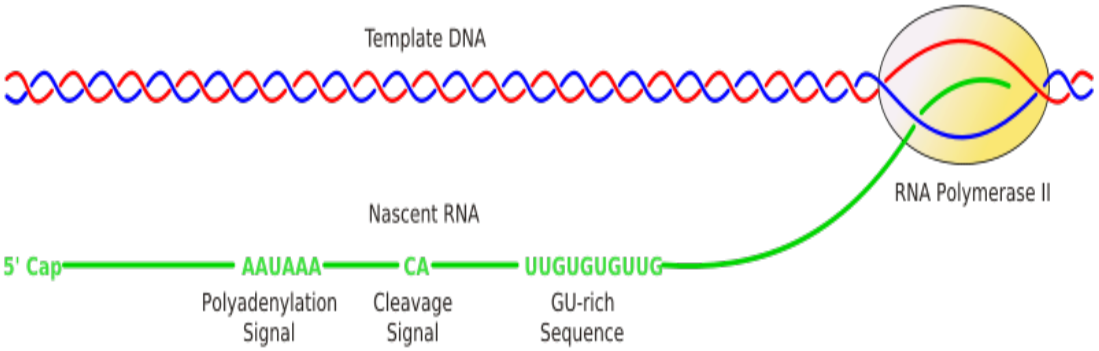
Capping

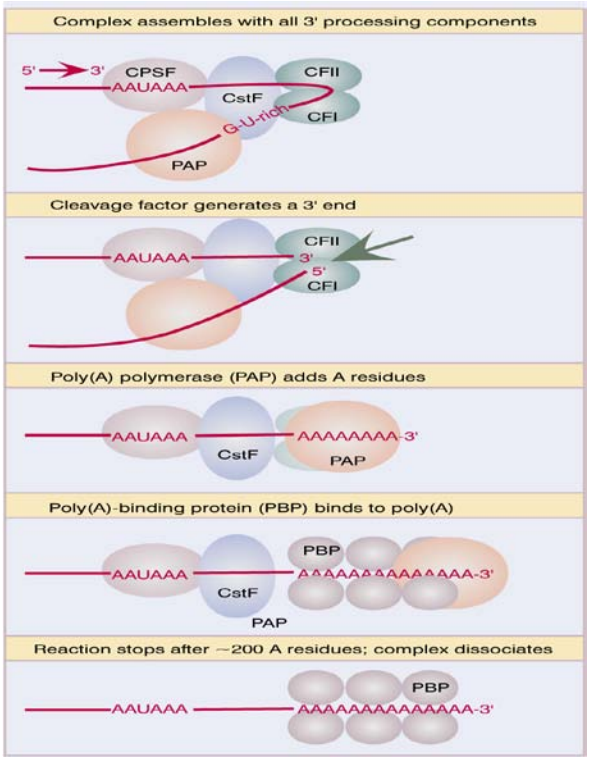




Capping of the pre-mRNA involves the addition of **7-methylguanosine (m^7G)** to the 5' end which is called a **cap 0 structure**. The ribose of the adjacent nucleotide may also be methylated to give a **cap 1**. Methylation of nucleotides downstream of the RNA molecule produce **cap 2**, **cap 3** structures and so on. The cap protects the 5' end of the primary RNA transcript from attack by ribonucleases.

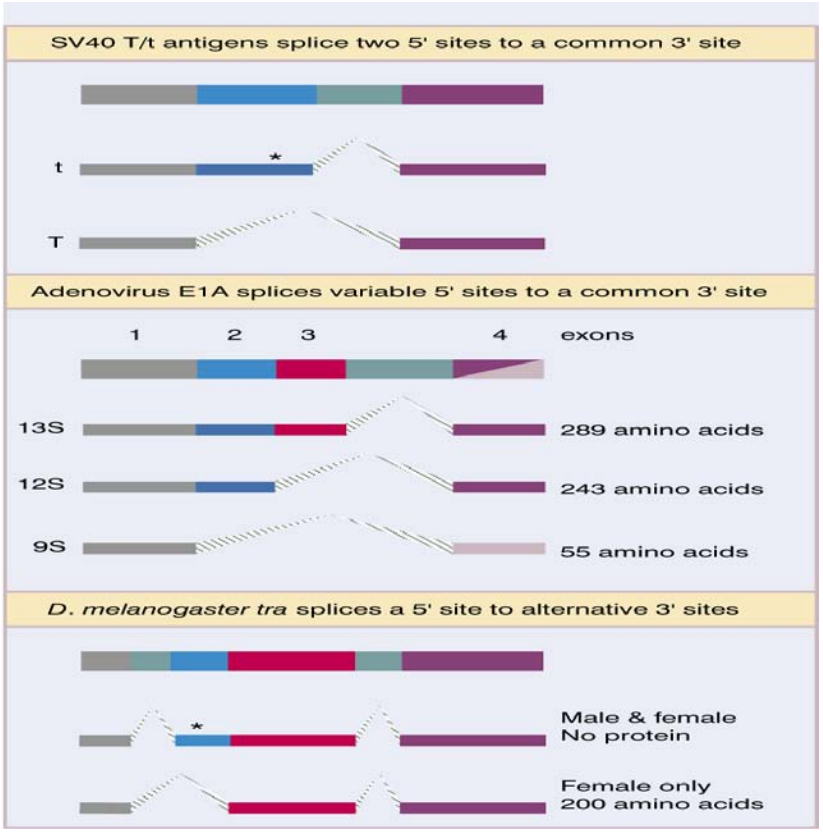
Cleavage and Polyadenylation





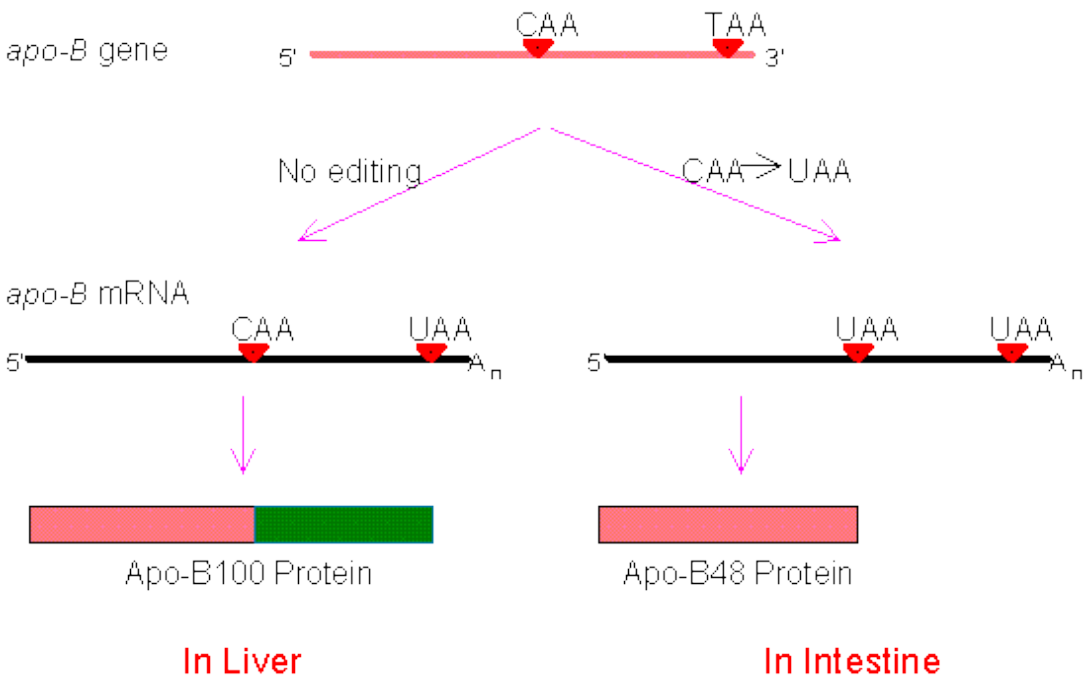
The pre-mRNA processing at the 3' end of the RNA involves cleavage and the addition of about 200 adenine residues to form a poly(A) tail.
Cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CStF) are two factors bind to the sequence elements. A protein complex forms which contains additional cleavage factors and the enzyme Polyadenylate Polymerase (PAP). This complex cleaves the RNA between the polyadenylation sequence and the GU-rich sequence at the cleavage site. Poly(A) polymerase then adds about 200 adenine units to the new 3' end of the RNA molecule using as a precursor. As the poly(A) tails is synthesised, it binds multiple copies of poly(A) binding protein, which protects the 3'end from ribonuclease digestion.

Alternative RNA Splicing



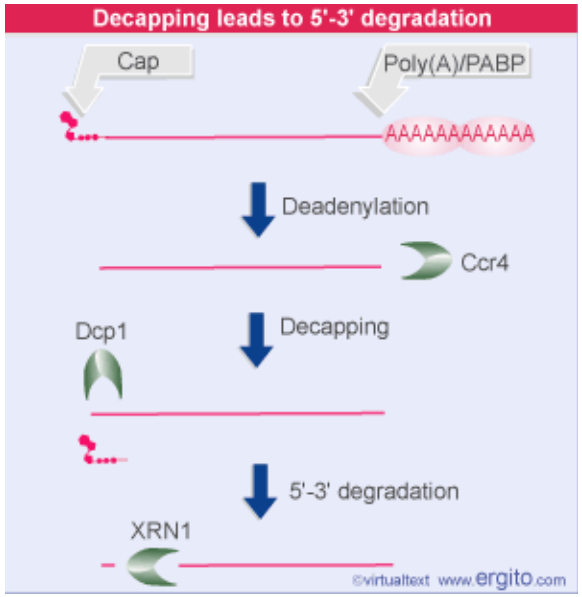
In many cases, the splicing process can create a range of unique proteins by varying the exon composition of the same messenger RNA. This phenomenon is then called alternative splicing. Exons can be extended or skipped, or introns can be retained.

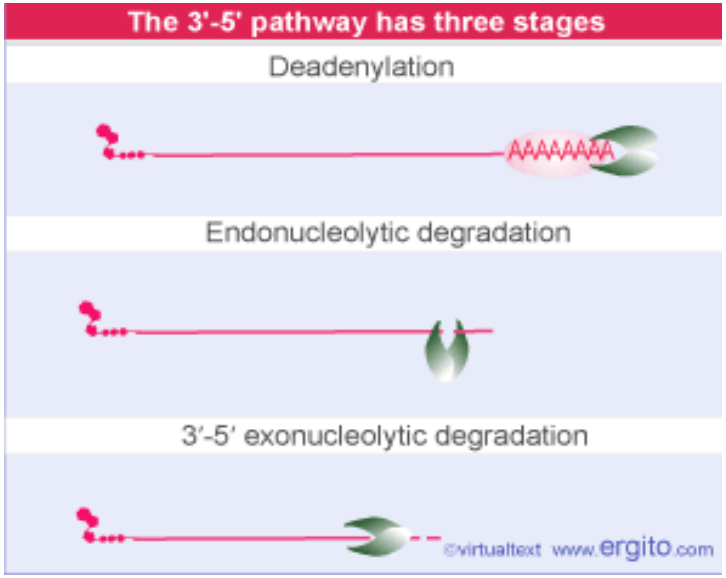
RNA Editing

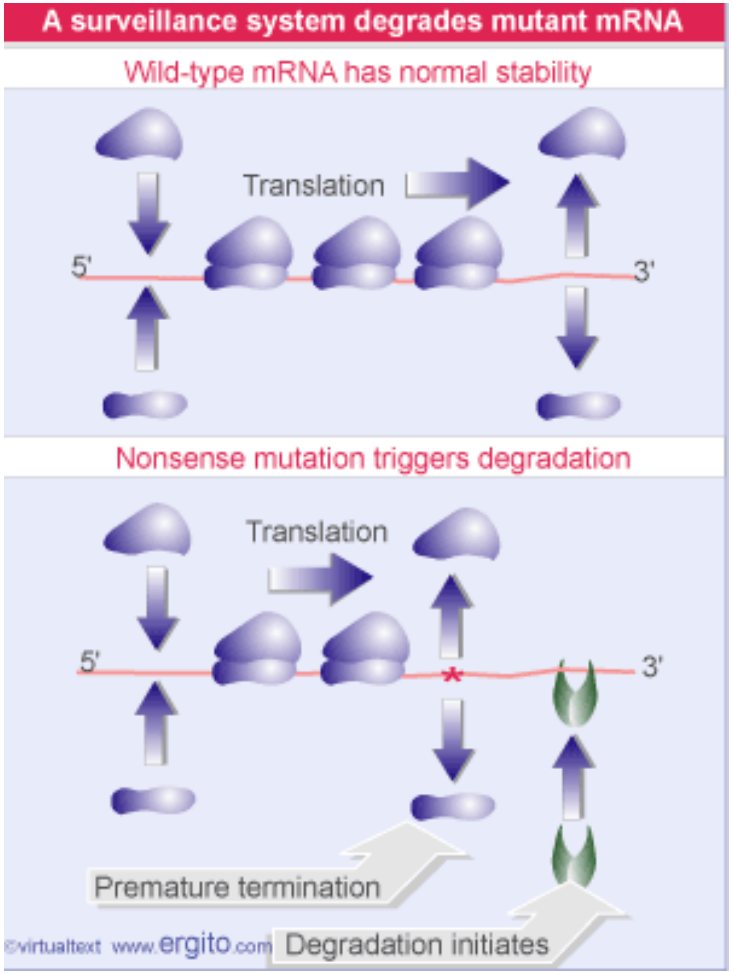


RNA editing describes those molecular processes in which the information content in an RNA molecule is altered through a chemical change in the base makeup. RNA editing in mRNAs effectively alters the amino acid sequence of the encoded protein.

mRNA Degradation

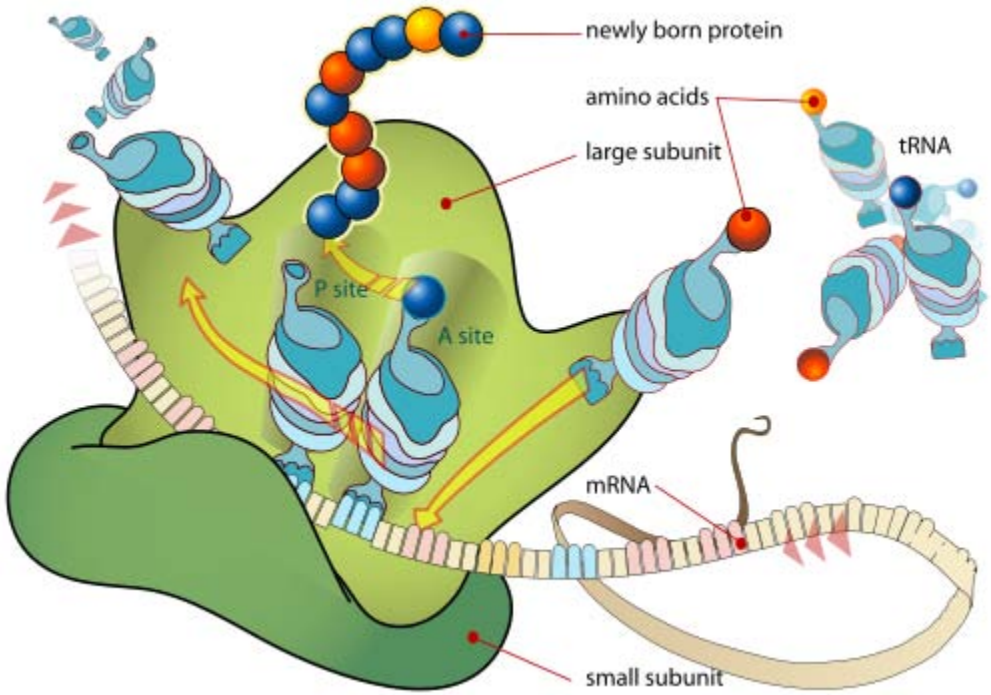






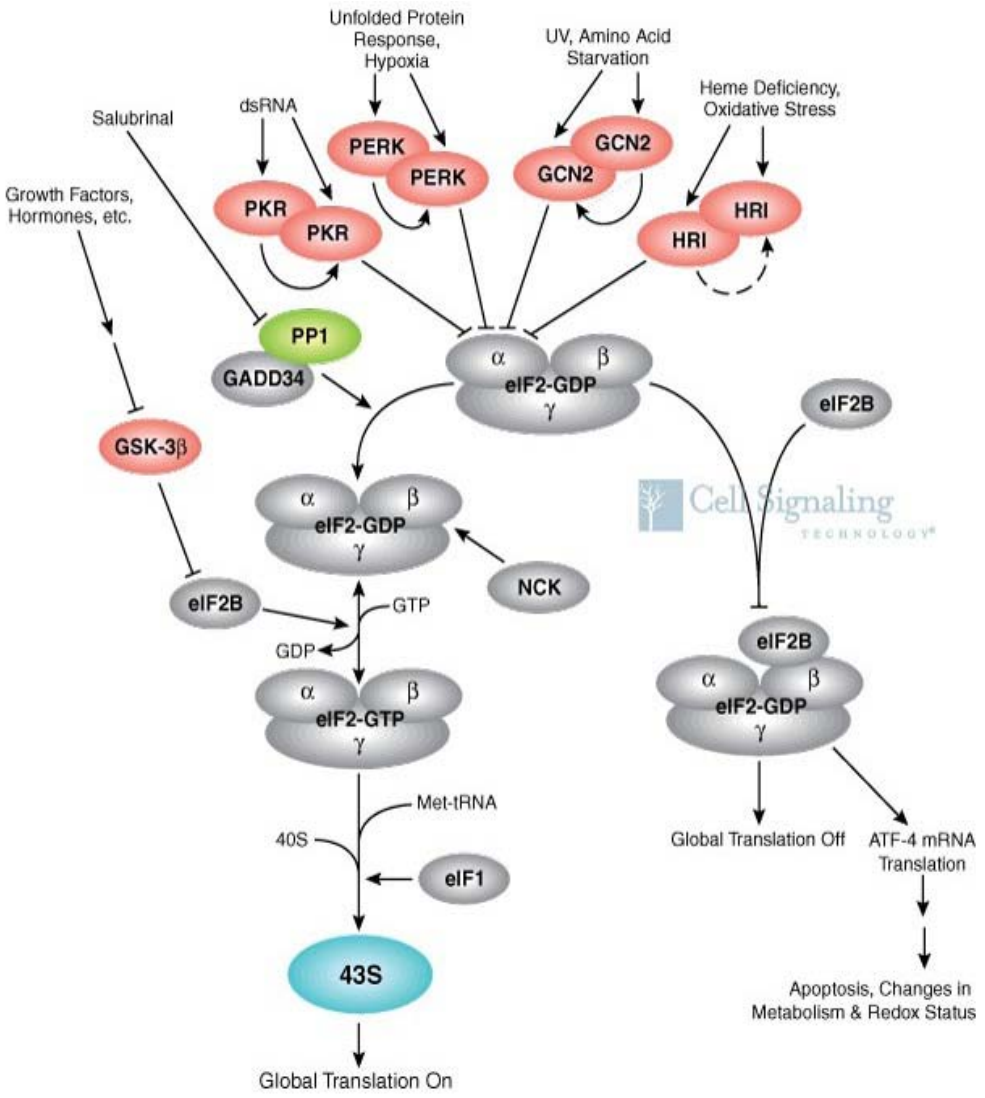
Translation Control

翻 译 调 控



Phosphorylation of eIF2 Regulates Synthesis of Protein

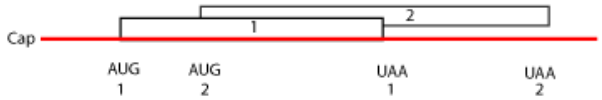
eIF2 (Eukaryotic Initiation Factor 2) is an eukaryotic initiation factor. It is required in the initiation of translation.



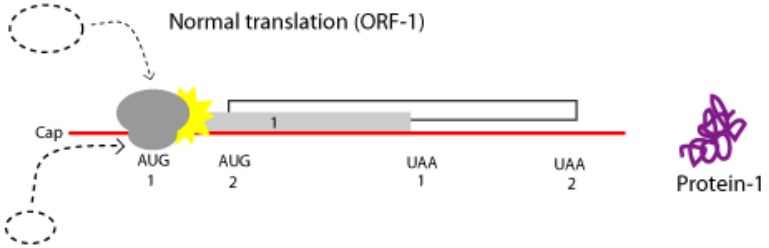
Protein phosphorylation plays an important role in the control of eIF2. eIF2 binds GTP and Met-tRNA and transfers the Met-tRNA to the 40S subunit to form the 43S pre-initiation complex. Later in the cycle, prior to elongation, the bound GTP is hydrolyzed, releasing eIF2-GDP. For eIF2 to promote another round of initiation, GDP must be exchanged for GTP, a reaction catalyzed by eIF2B. This phosphorylation stabilizes the eIF2-GDP-eIF2B complex, inhibiting the turnover of eIF2B.

Leaky Scanning

易遗漏扫描

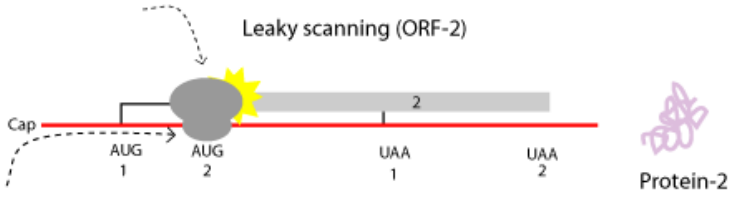


Leaky scanning is a way to translate two overlapping open reading frames (ORF) on the same mRNA independently of each other.

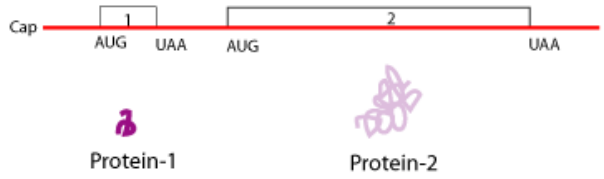


When the small ribosome subunit reaches the first AUG codon it may recruit the large subunit and initiate translation of the first ORF.

However, the nucleotides surrounding the AUG codon may disturb this process and the small subunit will then continue to the next AUG codon and initiate translation of ORF-2.



In this way two independent proteins can be expressed from a relatively small genomic region.



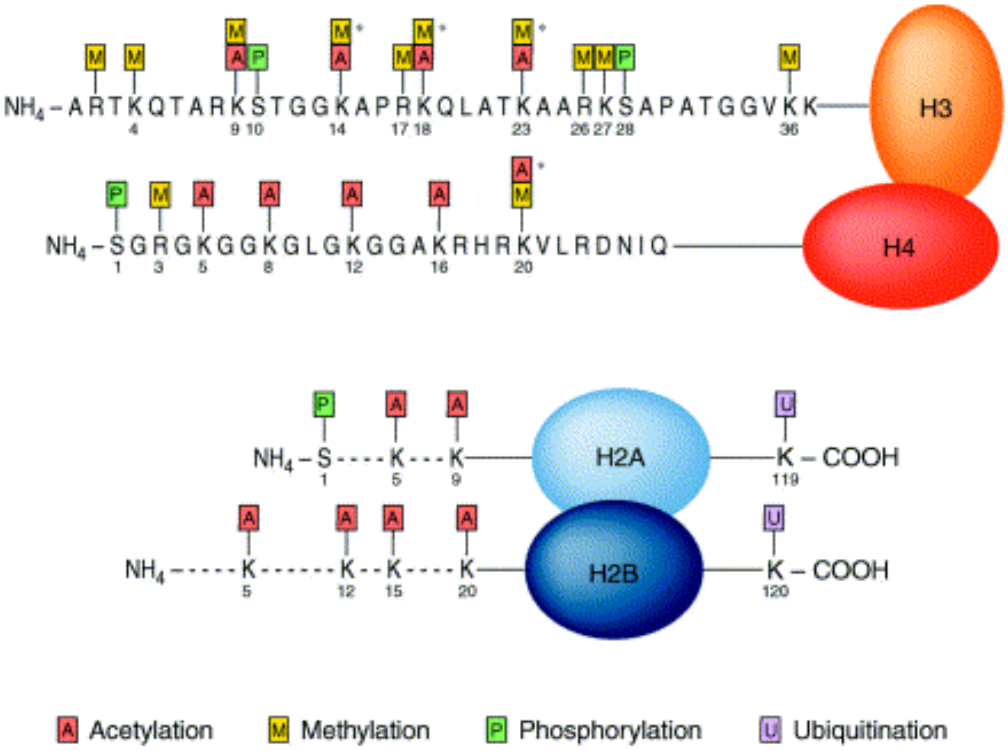
Leaky scanning is also used to express a protein encoded downstream of the first ORF on the mRNA.

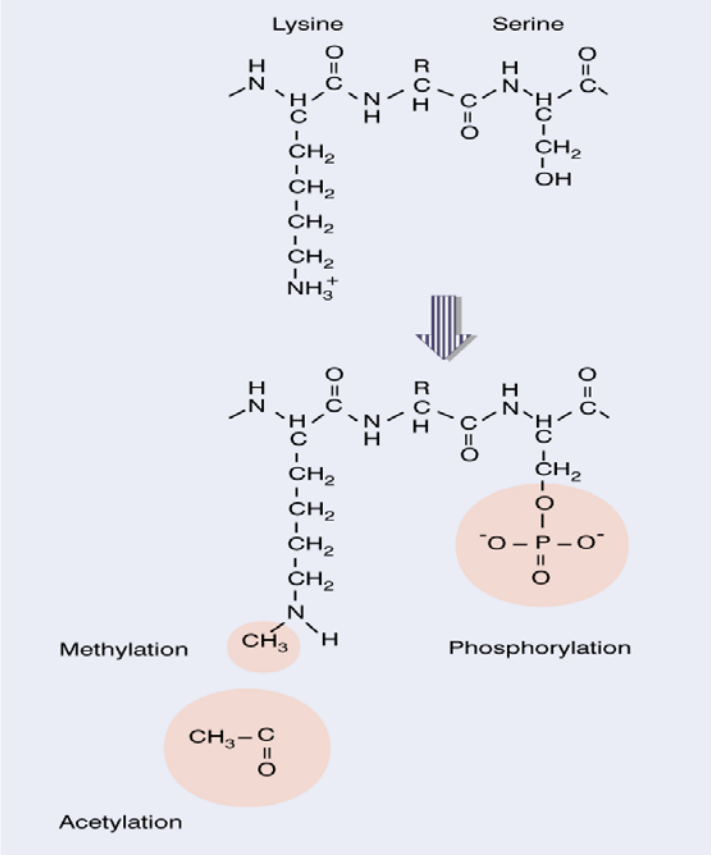
Posttranslational Modification

Posttranslational modification (PTM) is the chemical modification of a protein after its translation.

After translation, the posttranslational modification of amino acids extends the range of functions of the protein by attaching to it other biochemical functional groups such as acetate, phosphate, various lipids and carbohydrates, by changing the chemical nature of an amino acid or by making structural changes, like the formation of disulfide bridges.

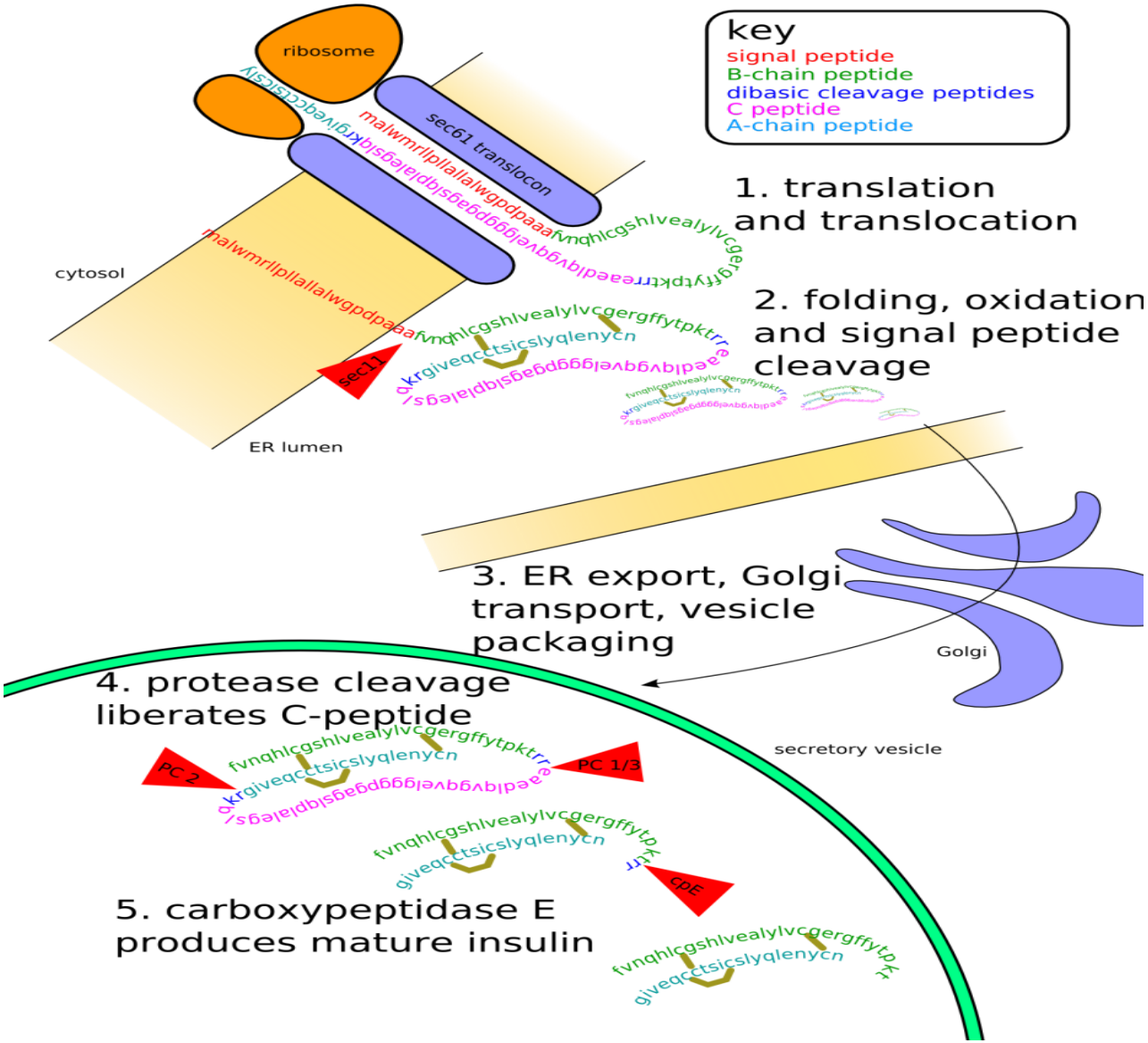
Post-translational Modifications of Histones





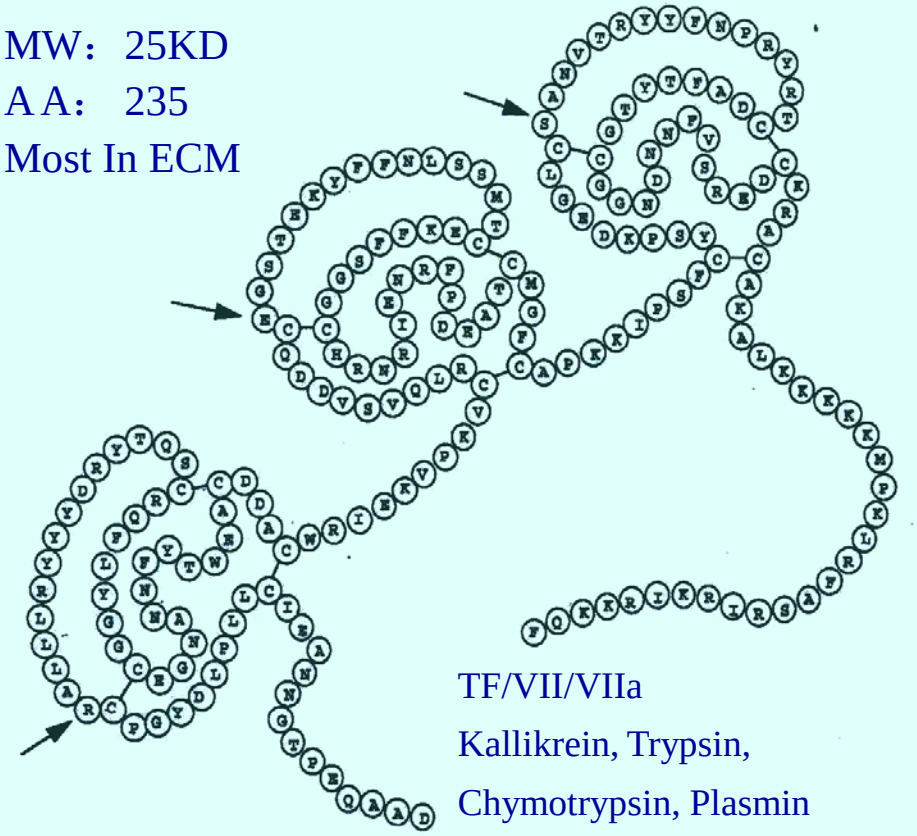
Acetylation: 乙酰化；Methylation: 甲基化；Phospharylation:磷酸化；Ubiquitination：泛素化

Modification of Primary Structure of Insulin



Structure and Function of TFPI-2

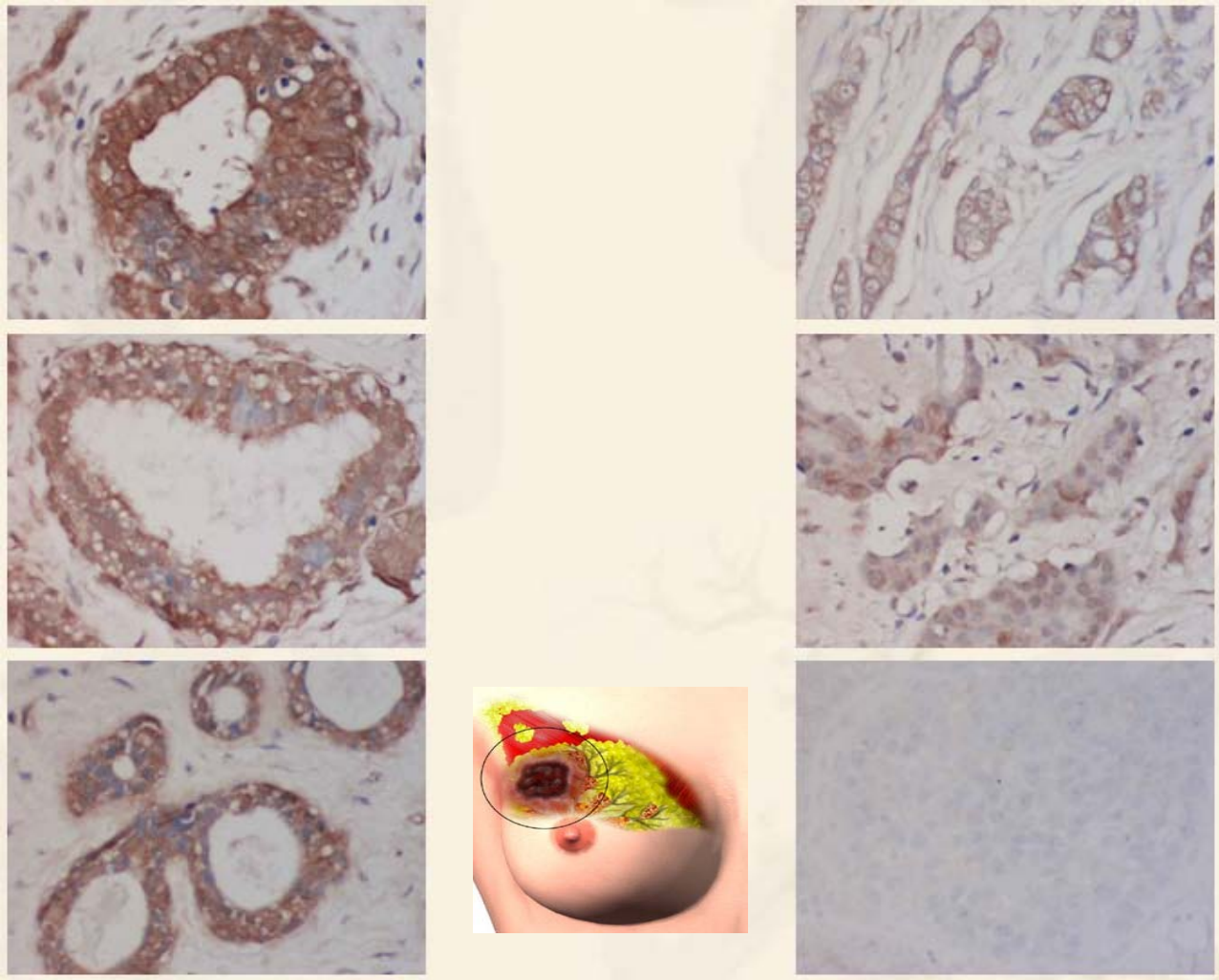
MW：25KD
AA： 235
Most In ECM

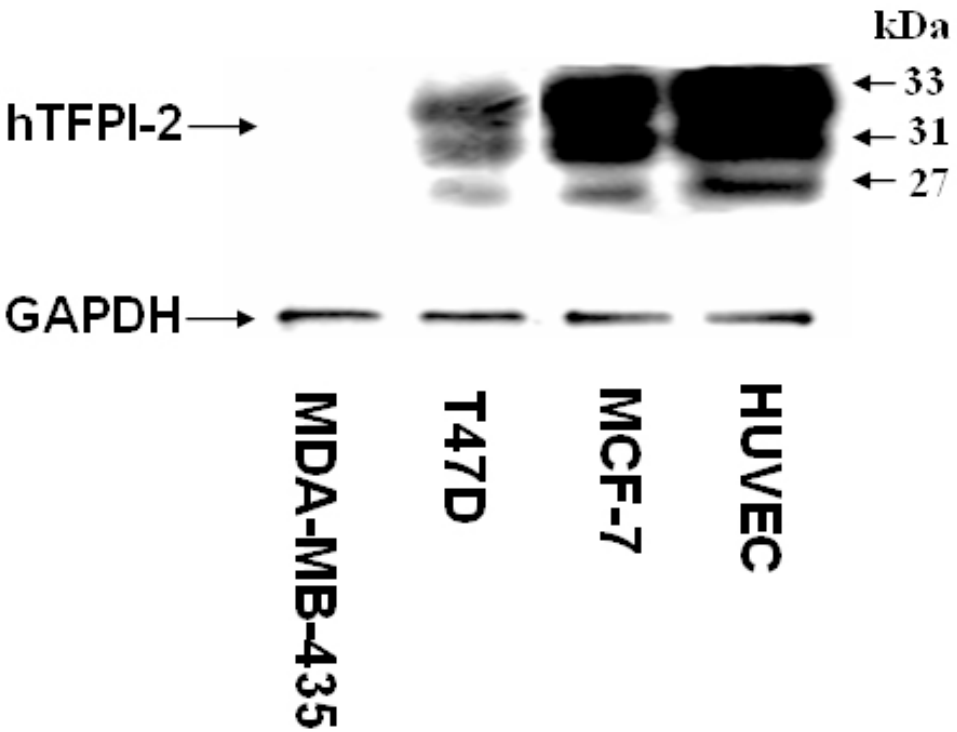


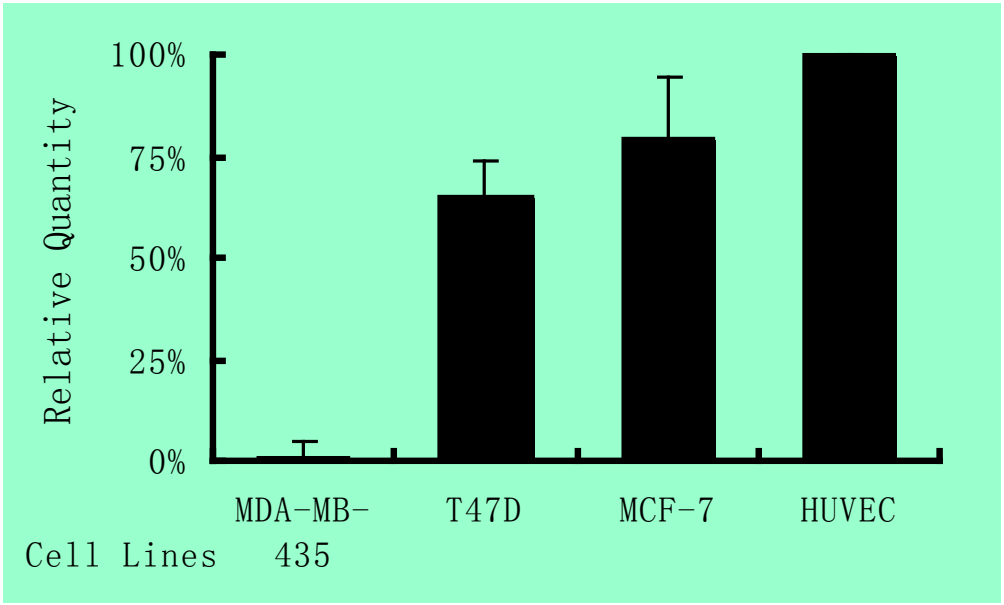
TF/VII/VIIa
Kallikrein, Trypsin,
Chymotrypsin, Plasmin

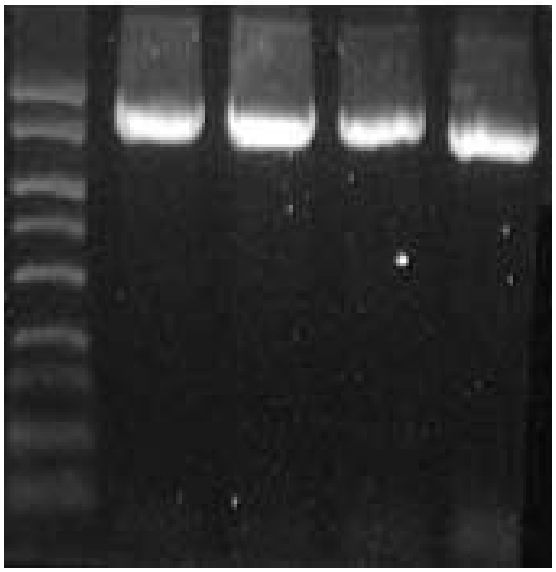
- **Physiological role**
  - Unclear
- **Inhibits many tumors invasion**
  - **Gliomas, Fibrosarcoma, Melanoma, Lung cancer, Prostate cancer, etc.**
- **Inhibits atherosclerosis plaques rupture.**
- **Target diseases**
  - Malignant Tumor metastasis
  - Unstable ungina
  - Acute pancreatitis

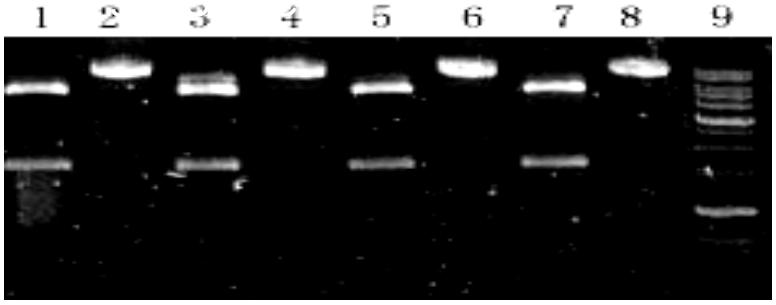
Expression of TFPI-2 in Breast Cancer Sample




kDa
hTFPI-2
33
31
27
GAPDH
MDA-MB-435
T47D
MCF-7
HUVEC


Relative Quantity
100%
75%
50%
25%
0%
MDA-MB-435
T47D
MCF-7
HUVEC
Cell Lines





435 435 231 231 M7 M7 EC EC M

HUVEC: T -839 C;

MCF-7: G –1196 C;

C –735 G;

C –702 T;

G –467 C

MDA-MB-231:

A –1222 G;

T -876 --;

C –463 T

MDA-MB-435:

A –1100 G;

A –1077 C;

A –1072 T;

A –1059 G;

T –876 --;

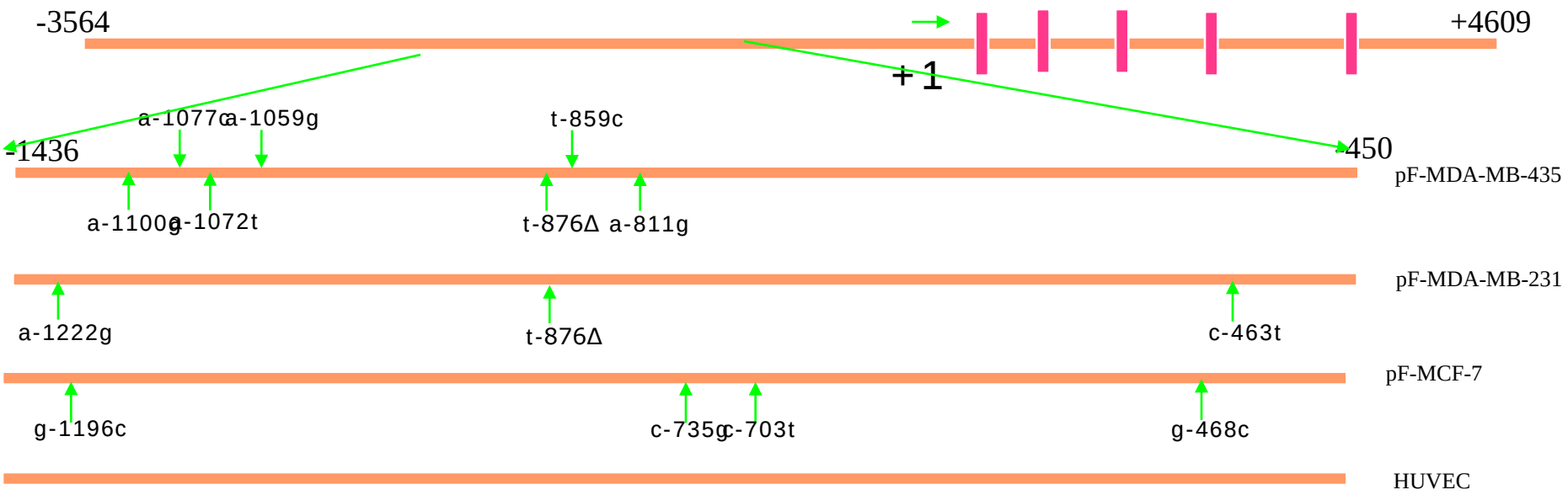
T –859 C;

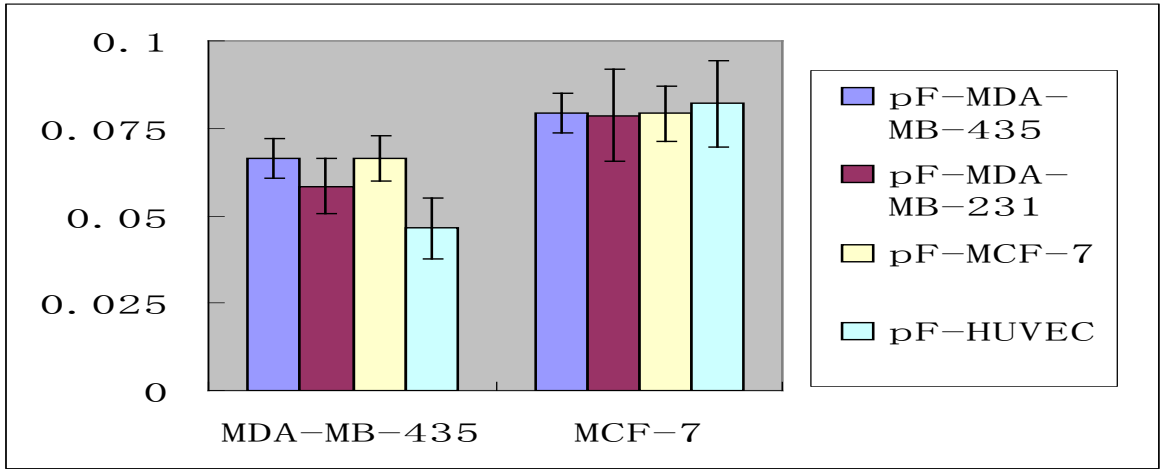
A –811 G;

A 13 G

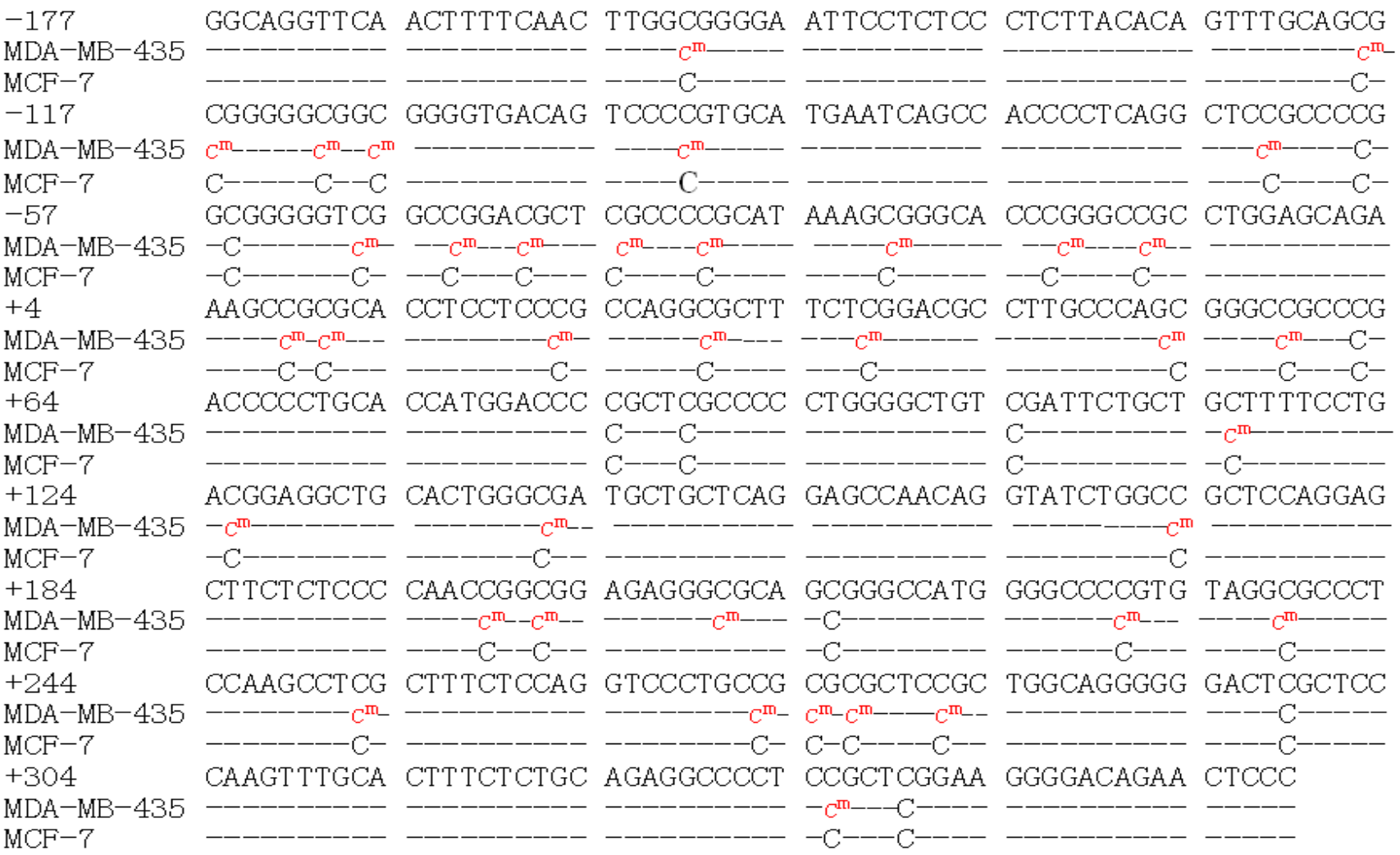
```
-1344  AGTAATAAGA CATTTGTGAT ATATTAGAAT ATTACTCAAT GATTTAAAGG AACAGACTAT Normal
-1344  AGTAATAAGA CATTTGTGAT ATATTAGAAT ATTACTCAAT GATTTAAAGG AACAGACTAT HUVEC
-1344  AGTAATAAGA CATTTGTGAT ATATTAGAAT ATTACTCAAT GATTTAAAGG AACAGACTAT MCF-7
-1344  AGTAATAAGA CATTTGTGAT ATATTAGAAT ATTACTCAAT GATTTAAAGG AACAGACTAT  231
-1344  AGTAATAAGA CATTTGTGAT ATATTAGAAT ATTACTCAAT GATTTAAAGG AACAGACTAT  435

-1284  TGATACAGAC AGCAACATAT TTATACATTG TGGGATATTC ATACGTTTTA TATTACTCAA Normal
-1284  TGATACAGAC AGCAACATAT TTATACATTG TGGGATATTC ATACGTTTTA TATTACTCAA HUVEC
-1284  TGATACACAC AGCAACATAT TTATACATTG TGGGATATTC ATACGTTTTA TATTACTCAA MCF-7
-1284  TGATACAGAC AGCAACATAT TTATACATTG TGGGATATTC ATACGTTTTA TATTACTCAA  231
-1284  TGATACAGAC AGCAACATAT TTATACATTG TGGGATATTC ATACGTTTTA TATTACTCAA  435

-1224  TGATTTAAAG GAACAGACTA TTGATACAGA CAGCAACATA TTTATACATT GTGGGATATT Normal
-1224  TGATTTAAAG GAACAGACTA TTGATACAGA CAGCAACATA TTTATACATT GTGGGATATT HUVEC
-1224  TGATTTAAAG GAACAGACTA TTGATACACA CAGCAACATA TTTATACATT GTGGGATATT MCF-7
-1224  TGGTTTAAAG GAACAGACTA TTGATACAGA CAGCAACATA TTTATACATT GTGGGATATT  231
-1224  TGATTTAAAG GAACAGACTA TTGATACAGA CAGCAACATA TTTATACATT GTGGGATATT  435

-1164  CATACGTTTT GGTAAATTCA TACCTTGGAC TACGCAGGAA TTAAAAGAAA CAGACTATTG Normal
-1164  CATACGTTTT GGTAAATTCA TACCTTGGAC TACGCAGGAA TTAAAAGAAA CAGACTATTG HUVEC
-1164  CATACGTTTT GGTAAATTCA TACCTTGGAC TACGCAGGAA TTAAAAGAAA CAGACTATTG MCF-7
-1164  CATACGTTTT GGTAAATTCA TACCTTGGAC TACGCAGGAA TTAAAAGAAA CAGACTATTG  231
-1164  CATACGTTTT GGTAAATTCA TACCTTGGAC TACGCAGGAA TTAAAAGAAA CAGACTATTG  435

-1104  ATGCAGTGAC CTGGGCGTAT CTCAAAAACA GGAAGTGTGA GAAAGAAATA CTACACAGTG Normal
-1104  ATGCAGTGAC CTGGGCGTAT CTCAAAAACA GGAAGTGTGA GAAAGAAATA CTACACAGTG HUVEC
-1104  ATGCAGTGAC CTGGGCGTAT CTCAAAAACA GGAAGTGTGA GAAAGAAATA CTACACAGTG MCF-7
-1104  ATGCAGTGAC CTGGGCGTAT CTCAAAAACA GGAAGTGTGA GAAAGAAATA CTACACAGTG  231
-1104  ATGCGGTGAC CTGGGCGTAT CTCAAAACCA GGTAGTGTGA GAAAGGAATA CTACACAGTG  435

 -924  CTGCAGGAGC TTTCTGGAGT GATGGAAATG TTCTAACTCT TTTTTTTTTC TTCTTCTTCT Normal
 -924  CTGCAGGAGC TTTCTGGAGT GATGGAAATG TTCTAACTCT TTTTTTTTTC TTCTTCTTCT HUVEC
 -924  CTGCAGGAGC TTTCTGGAGT GATGGAAATG TTCTAACTCT TTTTTTTTTC TTCTTCTTCT MCF-7
 -924  CTGCAGGAGC TTTCTGGAGT GATGGAAATG TTCTAACTCT TTTTTTTTOC TTCTTCTTCT  231
 -924  CTGCAGGAGC TTTCTGGAGT GATGGAAATG TTCTAACTCT TTTTTTTTOC TTCTTCTTCT  435

 -864  TCTTTTCTTT CGTTTCGAGA CGGAGTTCCA CTCTGCCGCC CAGACTGGAG TGCAGTGGCG Normal
 -924  TCTTTTTCTT CGTTTCGAGA CGGAGCTCCA CTCTGCCGCC CAGACTGGAG TGCAGTGGCG HUVEC
 -924  TCTTTTTCTT CGTTTCGAGA CGGAGTTCCA CTCTGCCGCC CAGACTGGAG TGCAGTGGCG MCF-7
 -924  TCTTTTTCTT CGTTTCGAGA CGGAGTTCCA CTCTGCCGCC CAGACTGGAG TGCAGTGGCG  231
 -924  TCTTTCTCTT CGTTTCGAGA CGGAGTTCCA CTCTGCCGCC CAGACTGGAG TGCGGTGGCG  435

 -744  CCCGAGGGGC TGGGACTACA GGTGCCCGCC ACCACGCCCG GCTAATTTTT TGTGTTTTTA Normal
 -744  CCCGAGGGGC TGGGACTACA GGTGCCCGCC ACCACGCCCG GCTAATTTTT TGTGTTTTTA HUVEC
 -744  CCCGAGGGGG TGGGACTACA GGTGCCCGCC ACCACGCCCG GTTAATTTTT TGTGTTTTTA MCF-7
 -744  CCCGAGGGGC TGGGACTACA GGTGCCCGCC ACCACGCCCG GCTAATTTTT TGTGTTTTTA  231
 -744  CCCGAGGGGC TGGGACTACA GGTGCCCGCC ACCACGCCCG GCTAATTTTT TGTGTTTTTA  435

 -504  TACGATATAT TATAAGCCTG TATTAAATGT AAATTAGAAC TCGATTGAAA TCTGTGTGTA Normal
 -504  TACGATATAT TATAAGCCTG TATTAAATGT AAATTAGAAC TCGATTGAAA TCTGTGTGTA HUVEC
 -504  TACGATATAT TATAAGCCTG TATTAAATGT AAATTACAAC TCGATTGAAA TCTGTGTGTA MCF-7
 -504  TACGATATAT TATAAGCCTG TATTAAATGT AAATTAGAAC TTGATTGAAA TCTGTGTGTA  231
 -504  TACGATATAT TATAAGCCTG TATTAAATGT AAATTAGAAC TCGATTGAAA TCTGTGTGTA  435

  -24  GCGGGCACCC GGGCCGCCTG GAGCAGAAAG CCGCGCACCT CCTCCCGCCA GGCGCTTTCT Normal
  -24  GCGGGCACCC GGGCCGCCTG GAGCAGAAAG CCGCGCACCT CCTCCCGCCA GGCGCTTTCT HUVEC
  -24  GCGGGCACCC GGGCCGCCTG GAGCAGAAAG CCGCGCACCT CCTCCCGCCA GGCGCTTTCT MCF-7
  -24  GCGGGCACCC GGGCCGCCTG GAGCAGAAAG CCGCGCACCT CCTCCCGCCA GGCGCTTTCT  231
  -24  GCGGGCACCC GGGCCGCCTG GAGCAGAAAG CCGCGCGCCT CCTCCCGCCA GGCGCTTTCT  435
```

Variation of DNA sequence in promoter region has no significant effect on it's activity



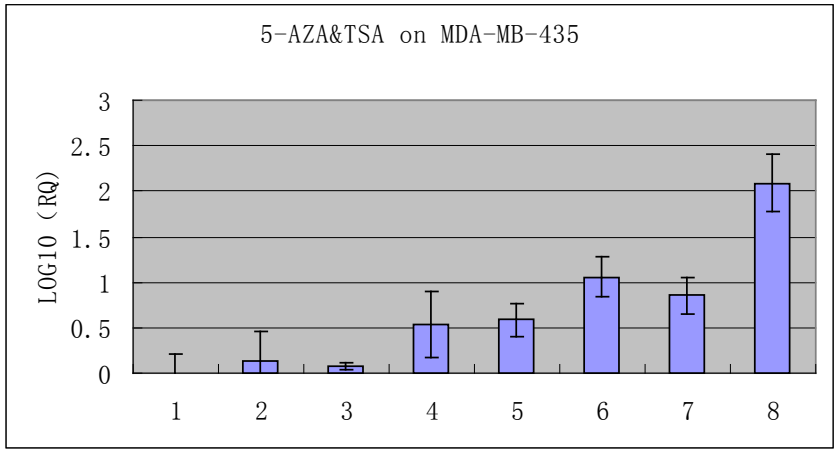


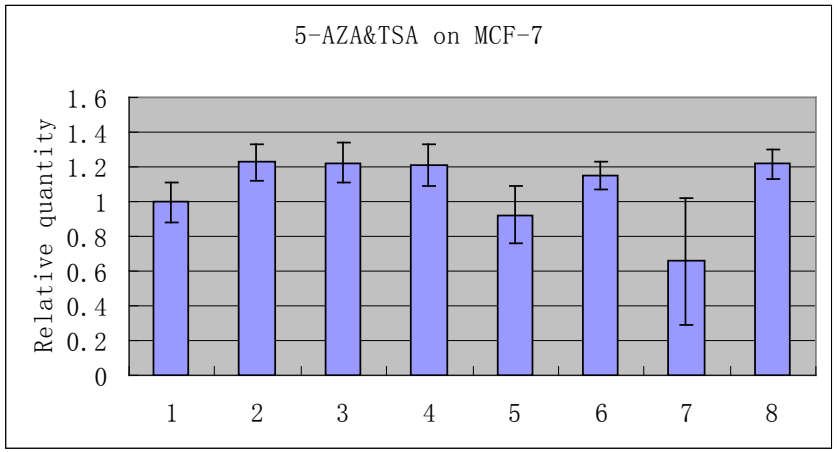
Methylation Difference in CpG Island

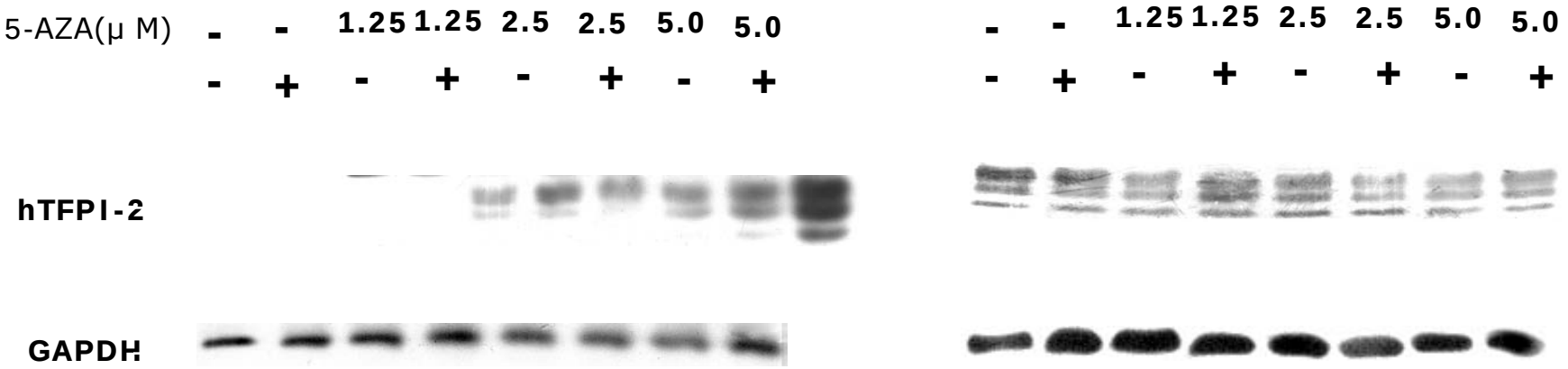


bisulfite sequencing PCR

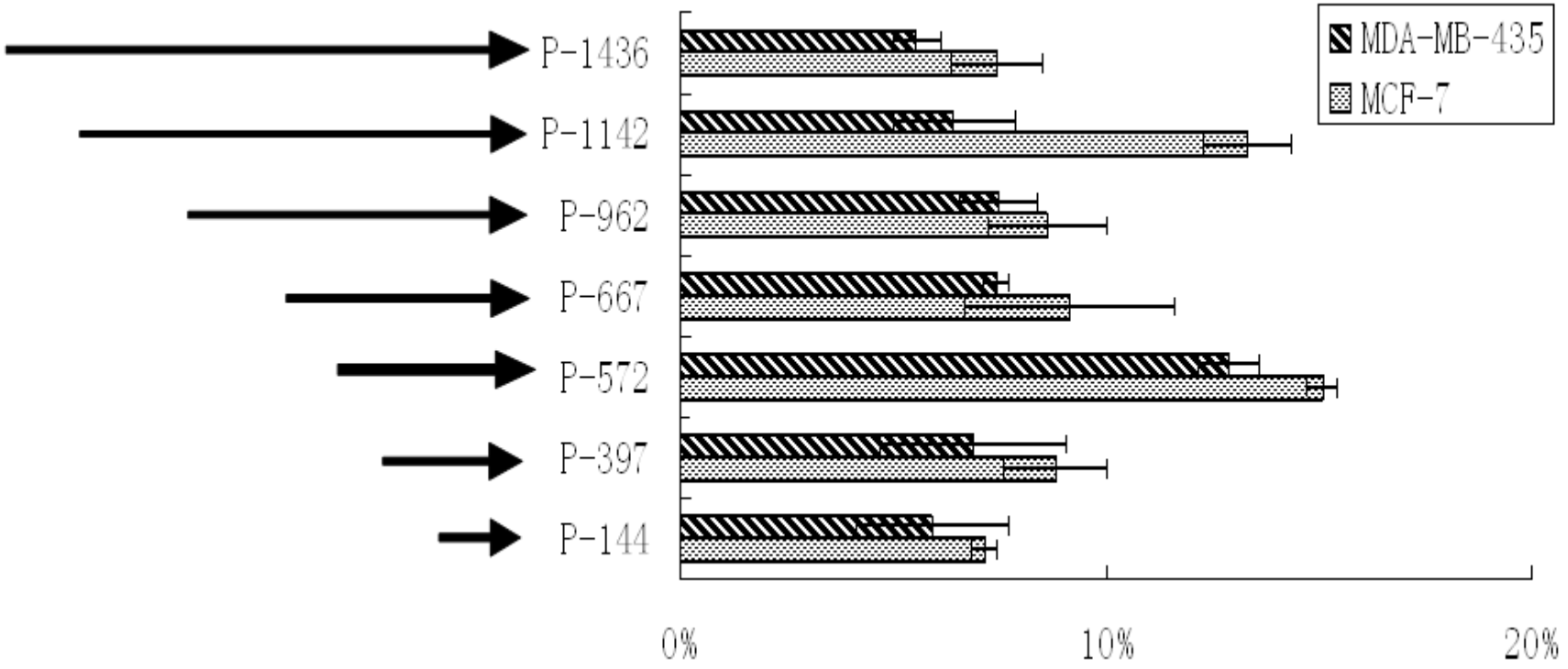
Effect of 5-AZA(DNMT inhibitor), TSA(HDAC inhibitor) on the mRNA and protein expression of TFPI-2



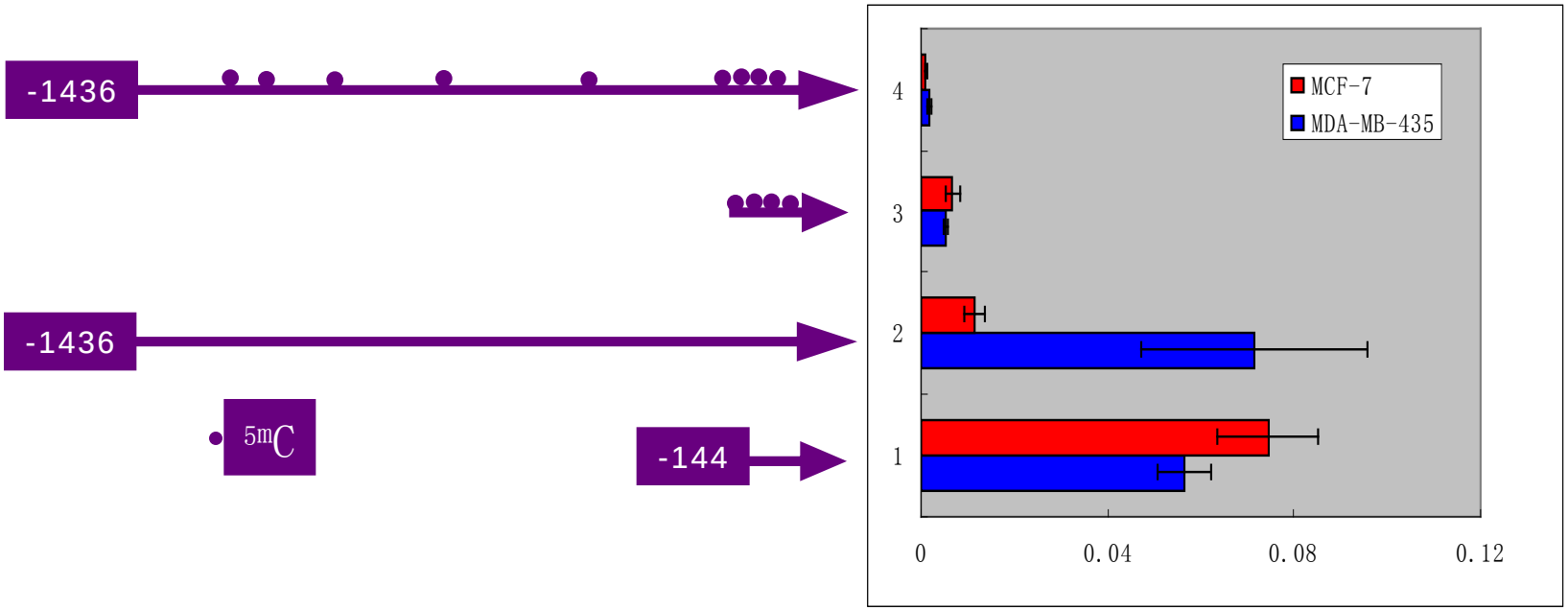




P-144 has nearly same luciferase activity as P-1436

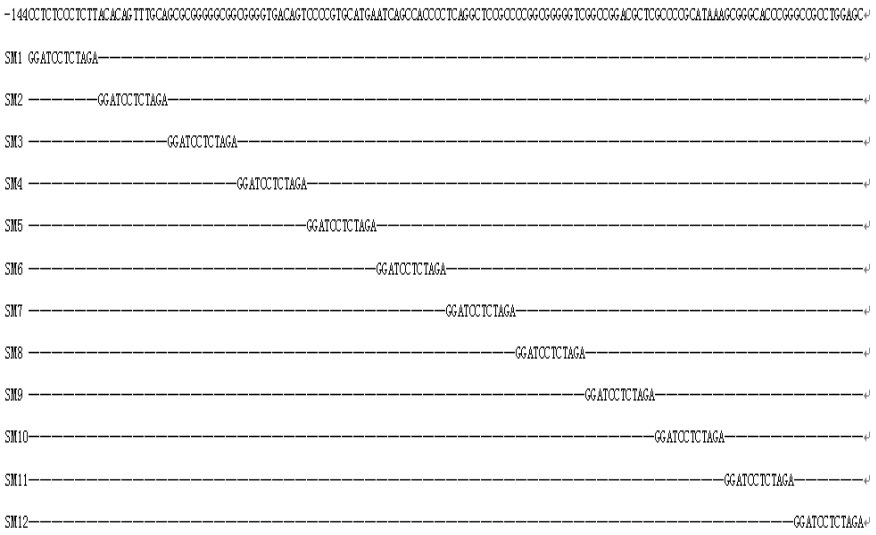


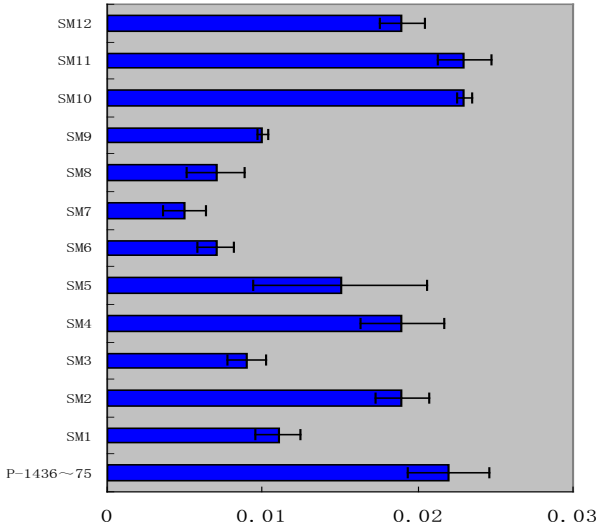
Methylation Represses TFPI-2 Promoter Activity In Vitro



Scanning Mutation

-84 ~ - 36 was essential

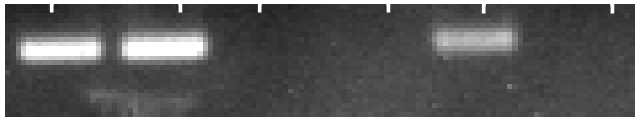




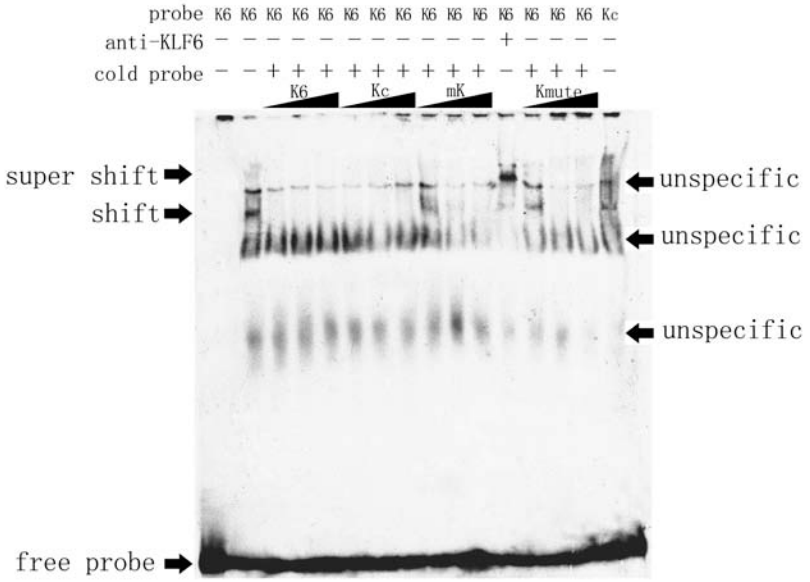
TFs	Opt.	Position.	S.	C.m.	M.m	Sequence
AP1	0.95	-89 to -79	(+)	0.884	0.971	catgaATCAgc
KLF6	0.87	-86 to -64	(+)	0.923	0.875	gaatcagCCACccctcaggctcc
KLF6	0.87	-72 to -50	(+)	1	0.908	tcaggctCCGCcccggcgggggt
SP1	0.88	-70 to -56	(-)	1	0.942	gccggGGCGgagcct
AP2	0.89	-63 to -51	(-)	1	0.895	ccCCCGccggggc

Searching Transcription Factor Binding Sites

Binding of KLF6 Was BLoked by DNA Methylation



A: MCF-7 B: MDA-MB-435

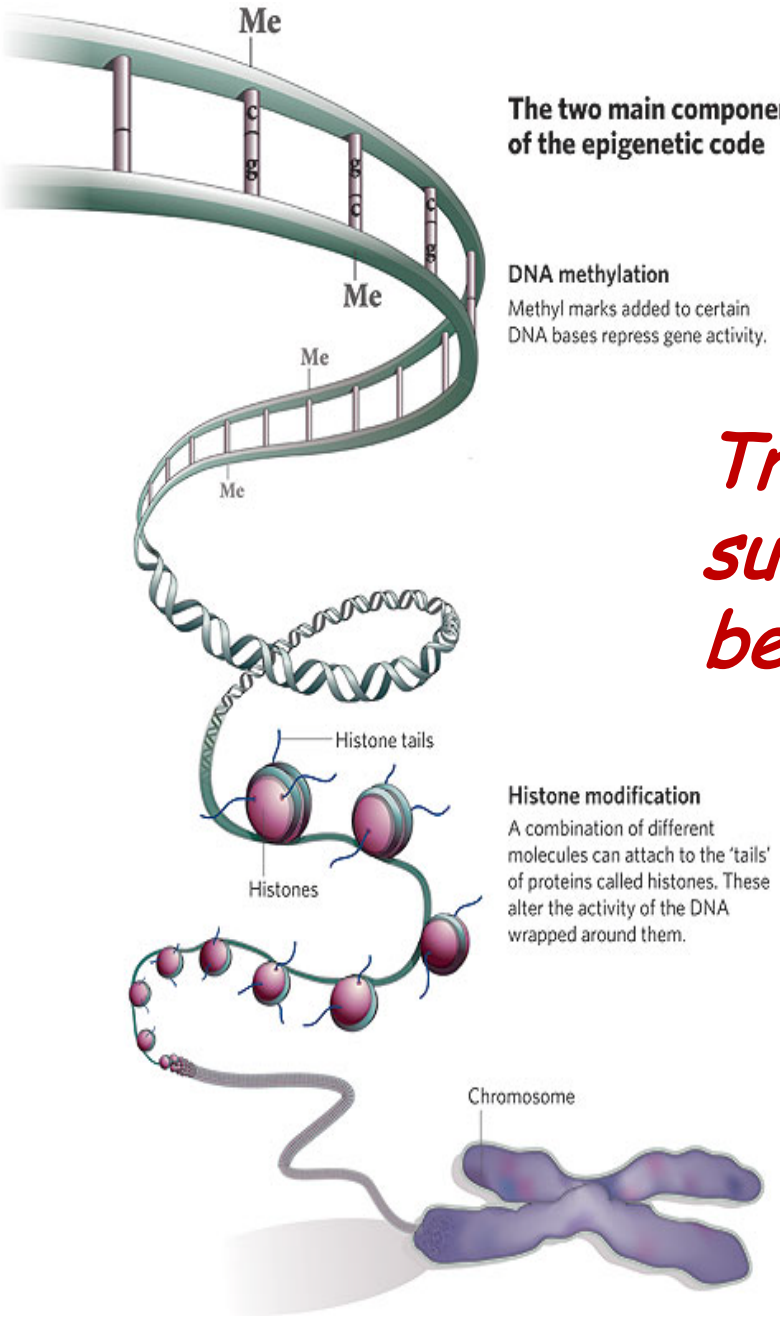


-86 KLF6 SP1 -42

5’-GAATCAGCCACCCCTCAGGCTCC^{m}GCCCCGGCGGGGGTCGGCCGGA-3';
3’-CTTAGTCGGTGGGGAGTCCGAGGC^{m}GGGGCCGCCCCCAGCCGGCCT -5'

Summary

- Chromatin Activation
- Control of Transcription Initiation: cis-elements, transcription factor
- Post-transcription modification: capping, polyadenylation, alternative mRNA splicing, RNE editing, mRNA degradation
- Post-translational Control: protein modification
- An example of Epigenetic regulation



The two main components of the epigenetic code

DNA methylation
Methyl marks added to certain DNA bases repress gene activity.

Histone modification
A combination of different molecules can attach to the 'tails' of proteins called histones. These alter the activity of the DNA wrapped around them.

Try not to become a man of success but rather try to become a man of value.

-- A. Einstein