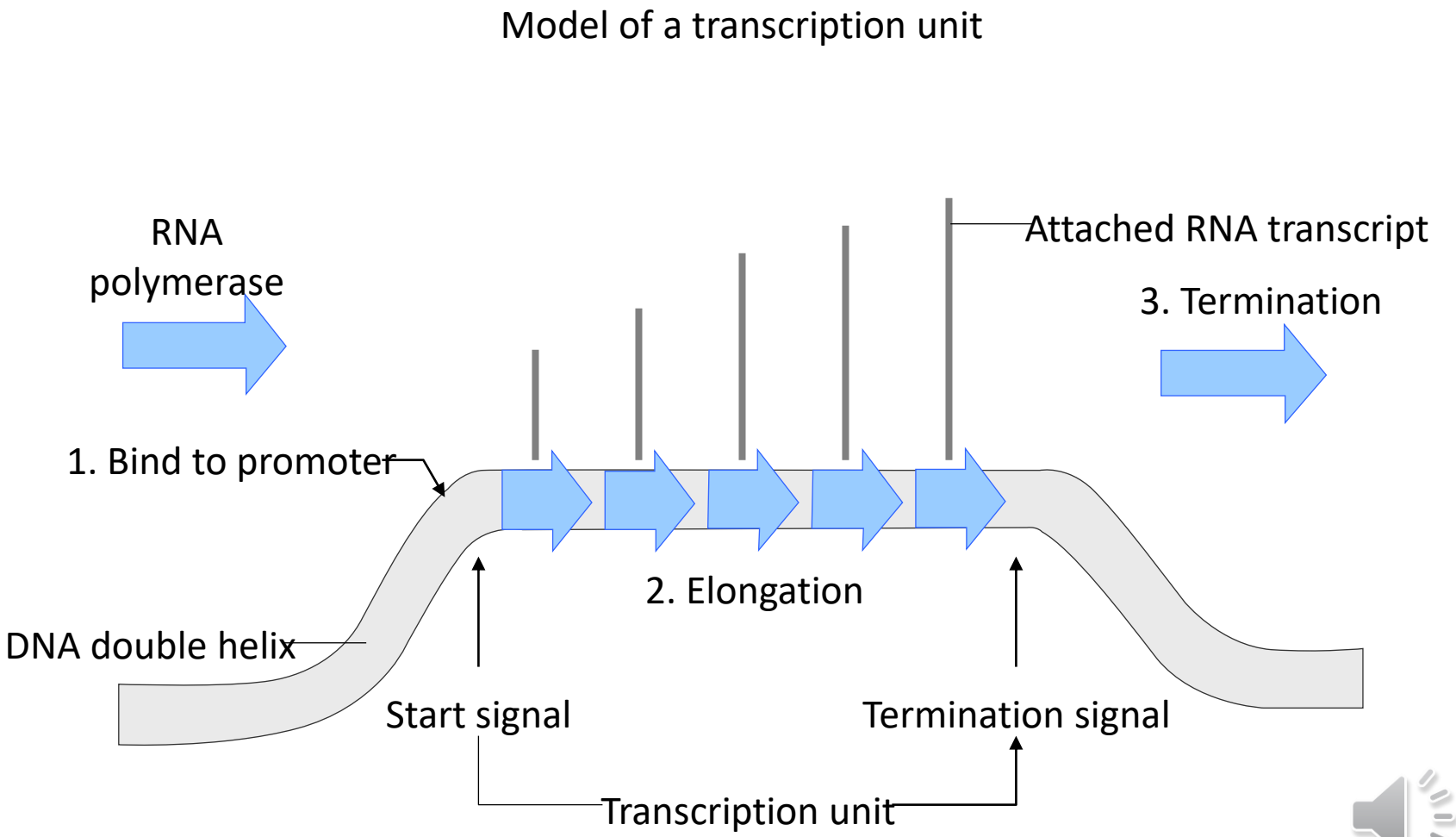
6. Regulation of gene expression in eukaryotes nuclear receptors (cell signaling)

Regulation is mediated by

1) interactions of regulatory proteins

with regulatory sequences on DNA
2) ncRNA



Levels of gene expression control

0. Chromatin
1. Where and how often is a given gene transcribed (transcriptional control)

2. How the primary transcript is spliced (post-transcriptional-spliced control)

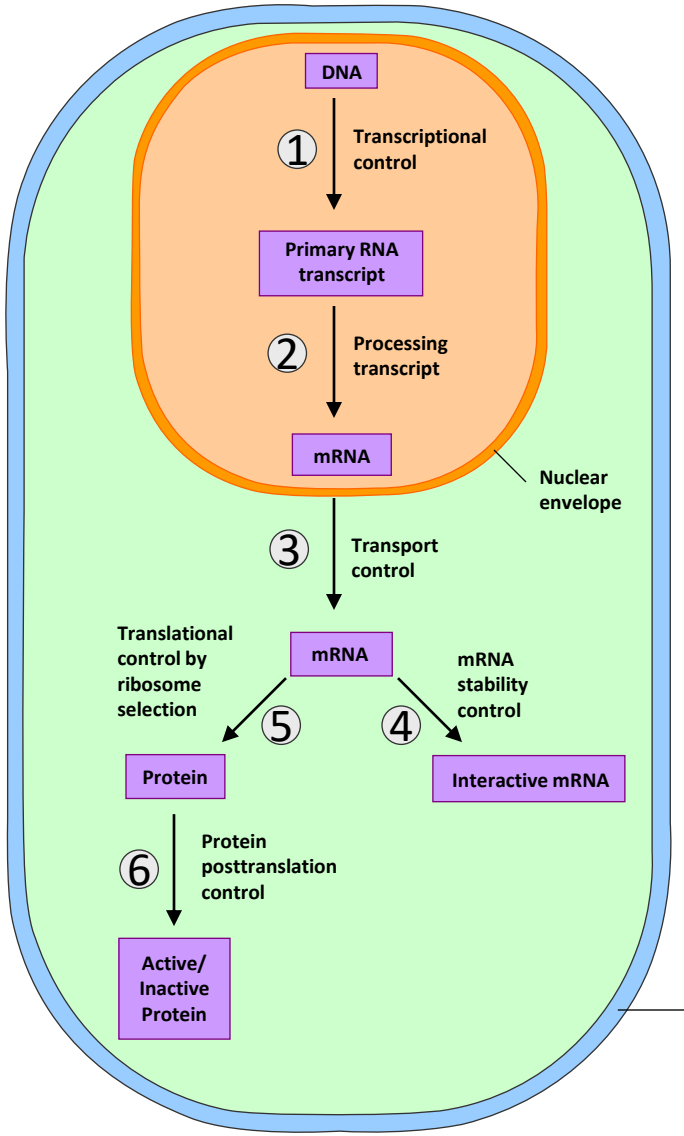
3. Selection of RNAs to be transported from the nucleus to the cytoplasm (control of RNA transport)

4. Selection of mRNAs to be translated on ribosomes (translational control)

5. Selective destabilization of certain mRNAs in the cytoplasm (mRNA degradation)

6. Selective activation, inactivation and compartmentalization of specific proteins after they have been synthesized (protein activity control - post-translational control, transport)

Six steps of information transfer in eukaryotes
that constitute potential regulatory points of gene expression



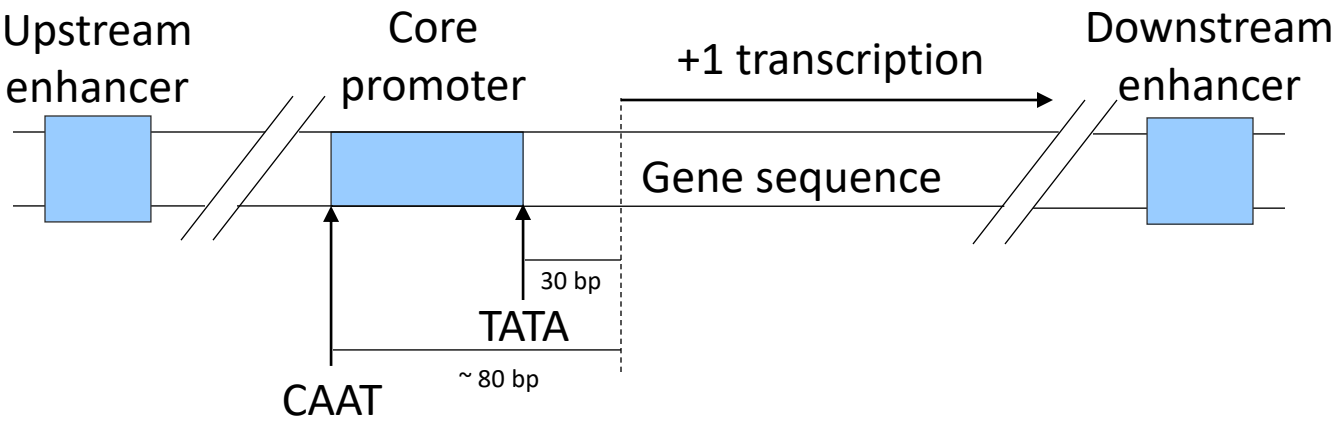
DNA segments that can modulate transcription by binding gene regulatory proteins

Regulation is mediated by

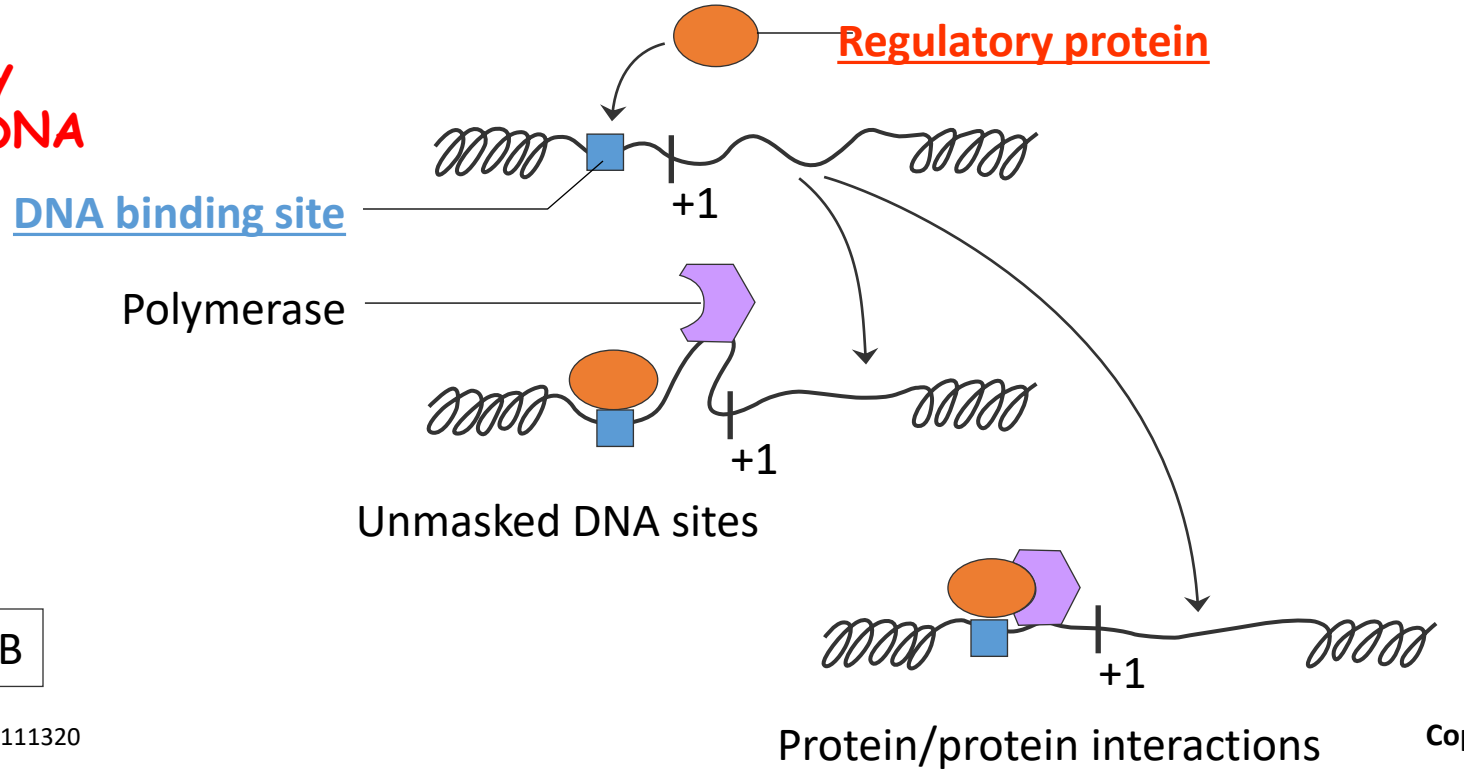
1) interactions of regulatory proteins

with regulatory sequences on DNA

2) ncRNA



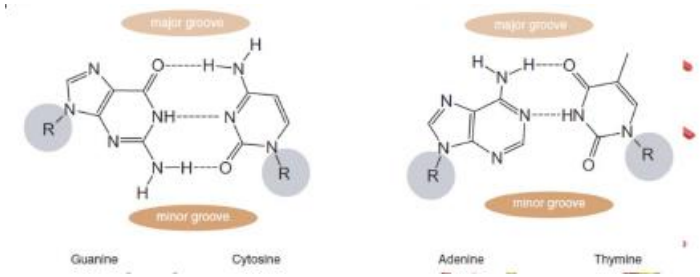
A



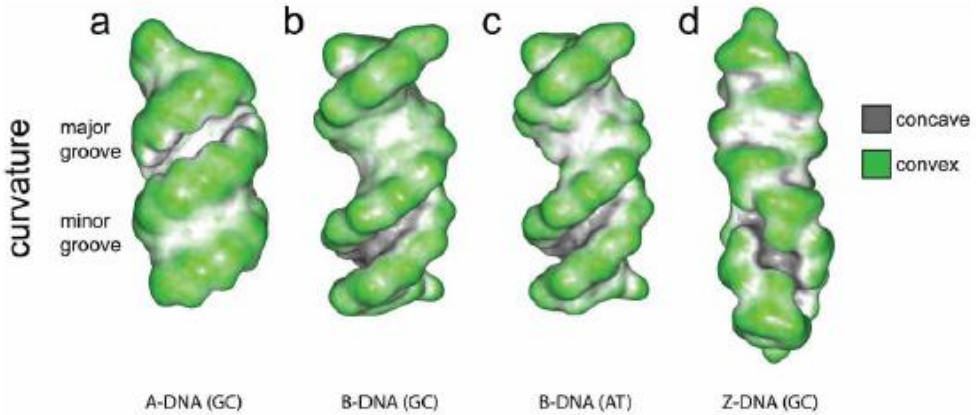
B

5111320

Protein-DNA interactions



- proteins interact with sugar phosphate skeleton (phosphate) or through grooves with bases
- **Sequences not sequence specific** (skeleton - histones; structurally specific - HMG proteins) or sequentially specific (skeleton + grooves - combination: BglII (AGATCT) and BamHI (GGATCC) contact the same bases and
They "read" the curvature of the surroundings
The shape and charge specificity of DNA determines the types of DNA binding domain DNA…)

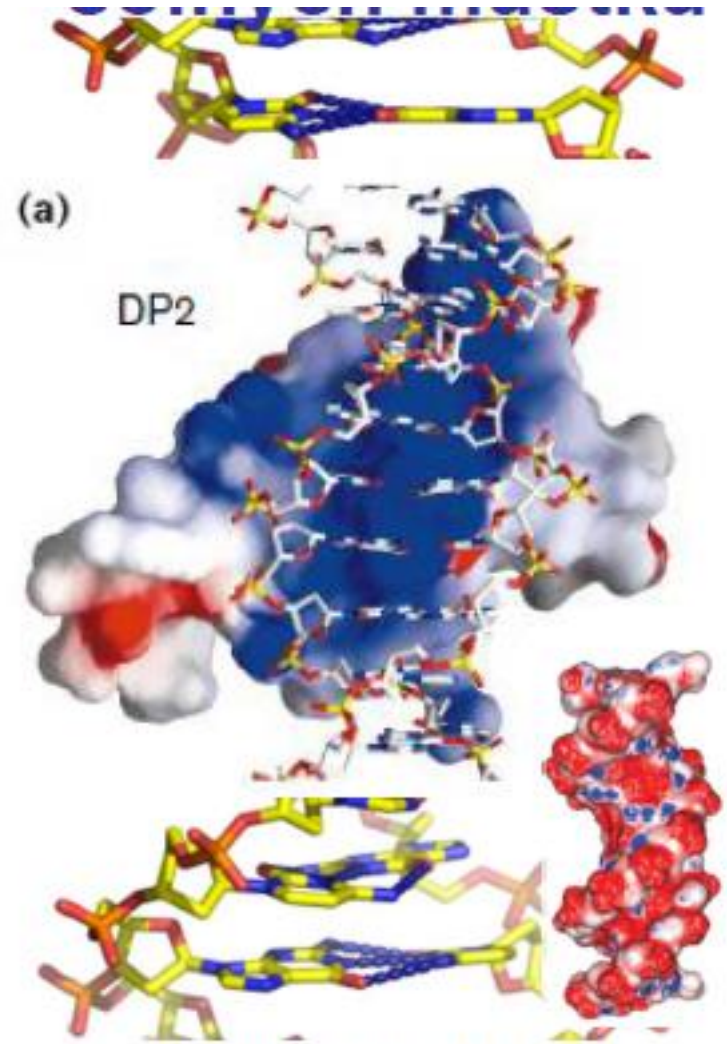


„shape readout"
zakřivení kostry - souvisí se sekvencí a prostředím

Rohs et al, Annu Rev Bioch, 2010

Types of interactions

- **salt bridges** - between phosphates and + charged AK side chains (Lys, Arg, His)
- **hydrogen bonds** - between phosphates, sugars, bases in NK and peptide bond or hydrophilic AK side chains
- **stacking** - between aromatic amino acids (Trp, Tyr, Phe, His) and bases
- **hydrophobic interactions** - between bases in NK and non-polar side chains of AK



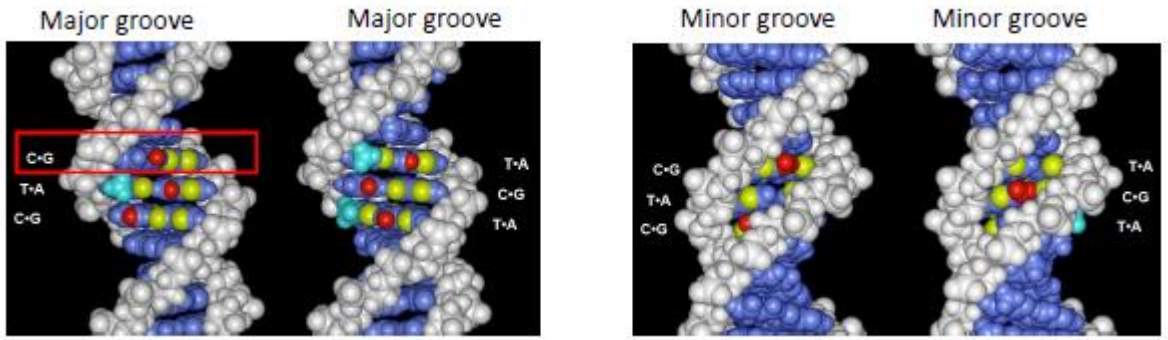
Gajiwala & Burley, COiSB, 2000

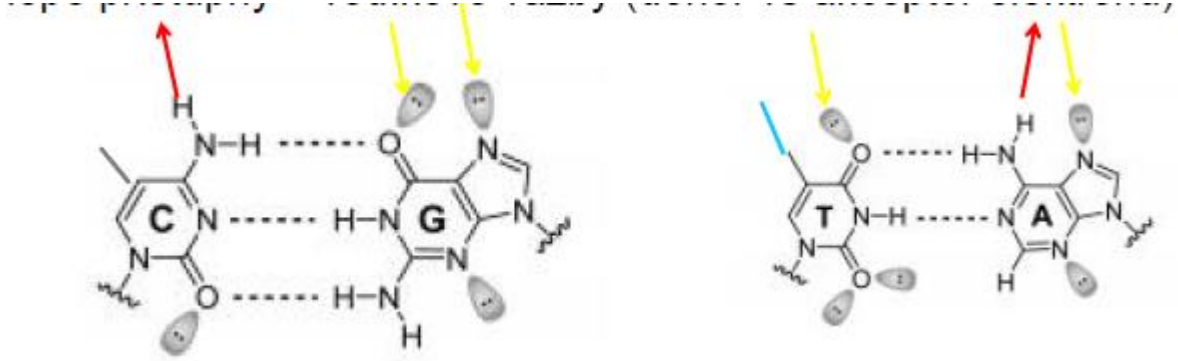
Phosphates can interact with salt bridges

- Arg and Lys - saline
- salt bridges
- (positive charges Arg and Lys
- creates a bond with the negative
- phosphate group charge)
- Electrostatic charge / surface indicates protein binding capabilities

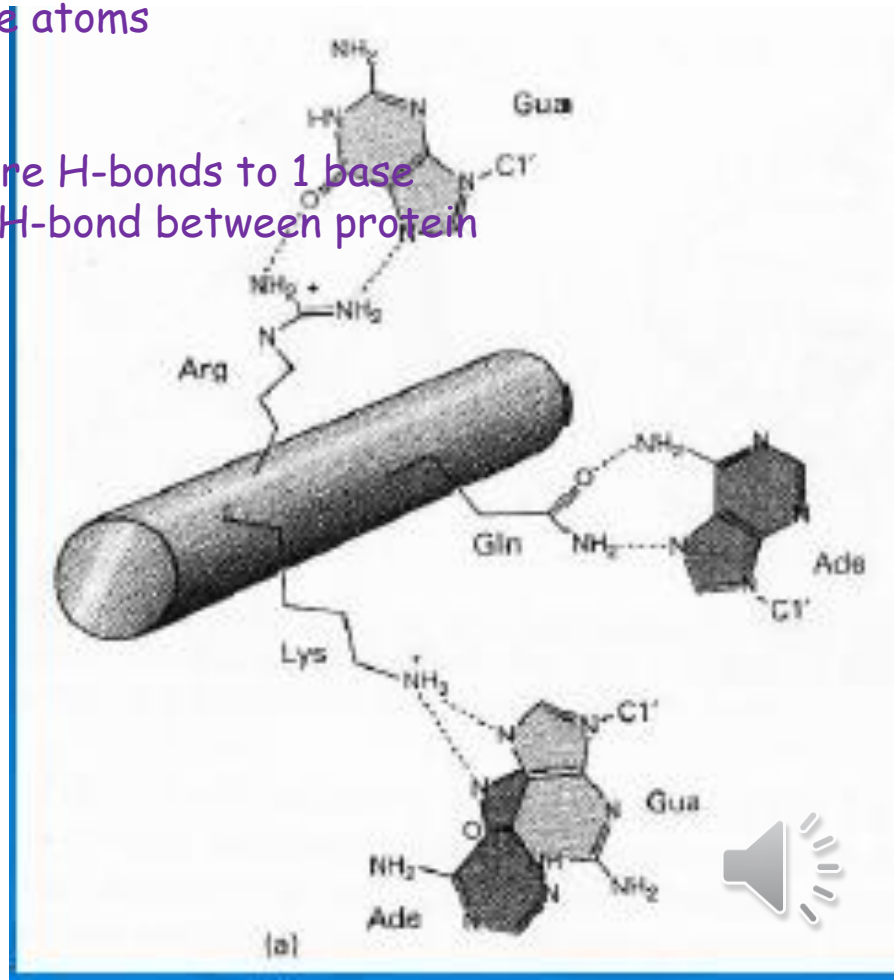
Hydrogen bonds

- sequence-specific protein contacts the base ("direct"
- readout) - through a large or small groove - a large groove is
- more accessible **- hydrogen bonds (donor vs electron acceptor**



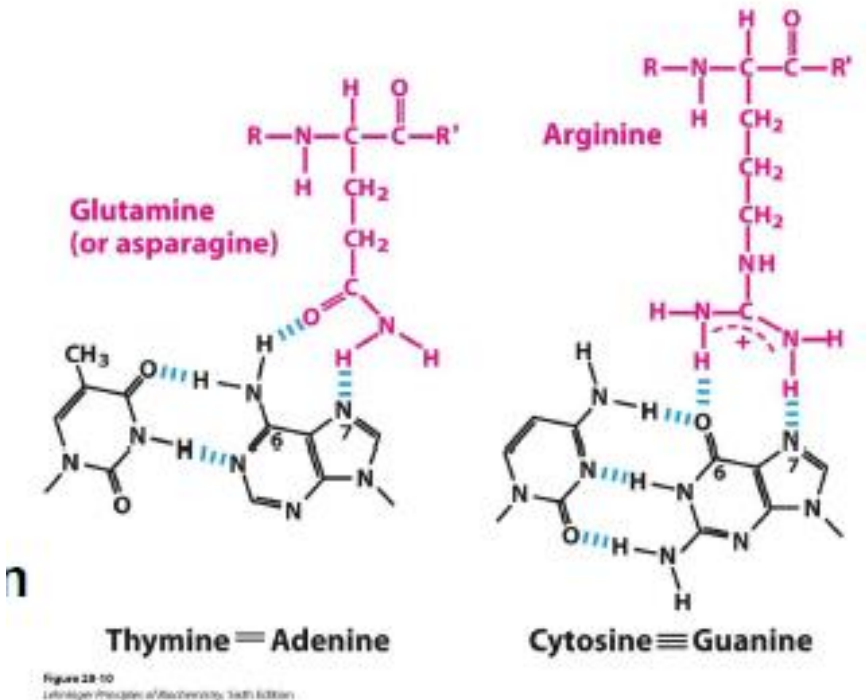


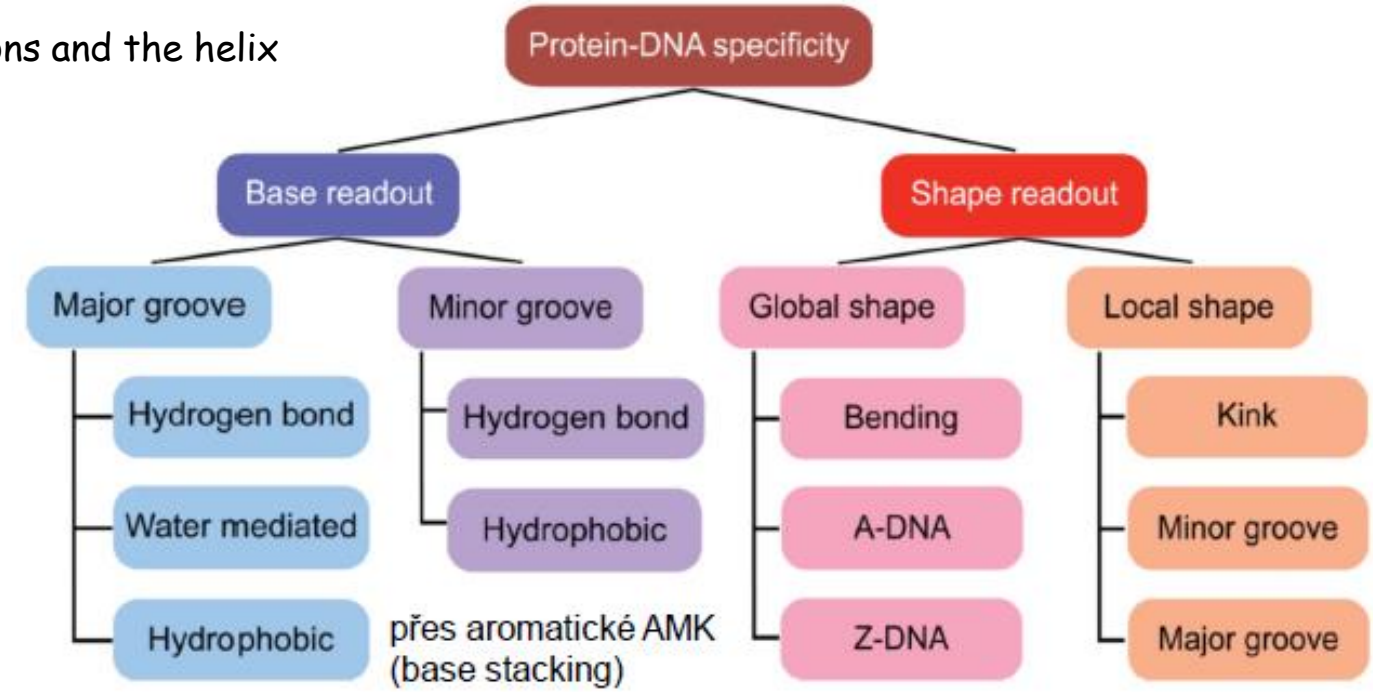
related to recognition and helium - pre interaction and helix with DNA ex. many direct interactions between AK side chains and NA bases.
Uncommon interactions:
O6 or N7 guan atoms…
Side chains Arg, Lys, Gln, Asn, Ser
Less often:
N6 or N7 adenine atoms
Exceptionally:
Pyrimidines
often occurs: more H-bonds to 1 base
water-mediated H-bond between protein and NA



Binding of proteins to DNA via hydrogen bonds

- The large groove has the size corresponding to the dimensions and the helix
- and has exposed H-linking groups
- Ade residues C-6 (NH2) and N-7 may form specific ones
- hydrogen bonds with Gln and Asn
- Gua can form specific hydrogen bonds with Arg
- Strong binding, sequence specific - affinity nM - uM
- Weak binding, structural specific - affinity uM - mM

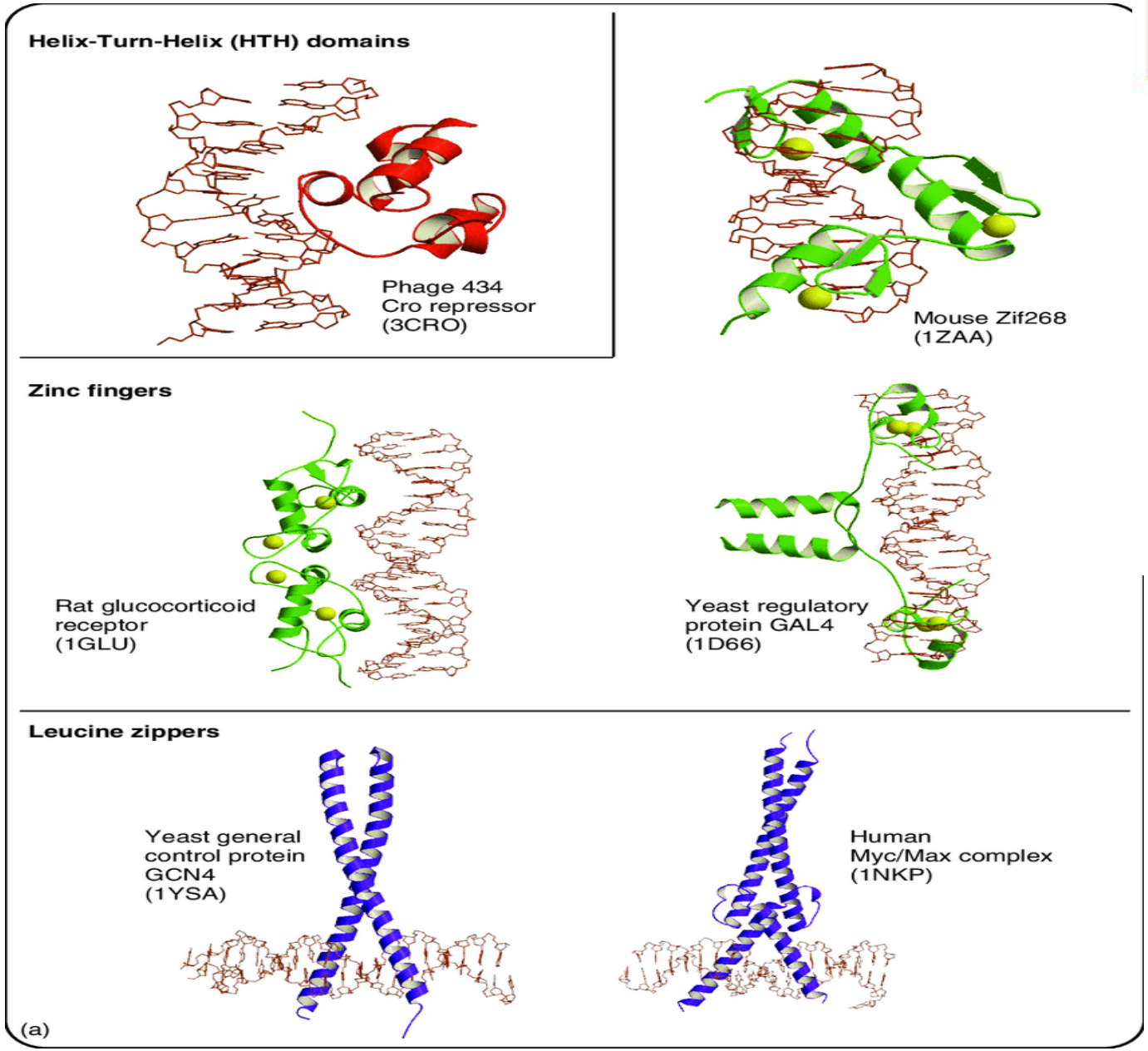




- více jak 70 SCOP superrodin (strukturních motivů)
- dle sekundárních struktur – α-šroubovice (17), β-listy (7), smíšené α/β motivy (48)

Rohs et al, Annu Rev Bioch, 2010

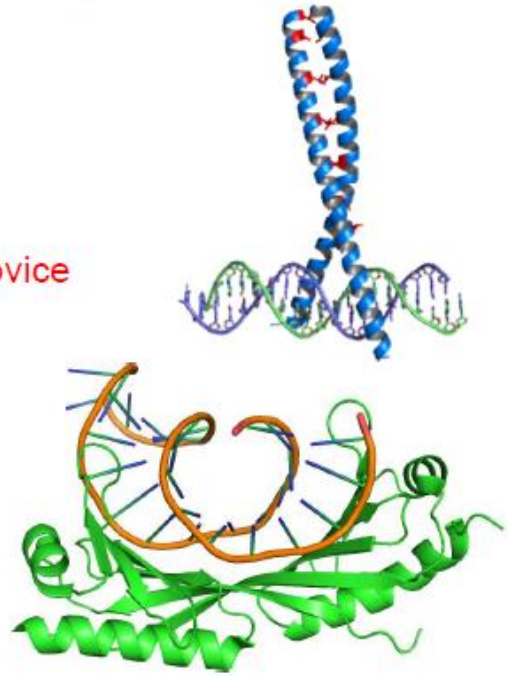
Protein motifs interacting with DNA



- **Zipper typ**
  - Leucinový zip
  - Helix-loop-helix
- **Helix-otáčka-helix**
  - HTH
  - Winged helix
  - TALE
- **Zinkový prst**
  - ββα zinc-finger
  - Hormon-receptor
  - Loop-sheet-helix
  - Gal4
- **Histon, HMG-box**
- **β-sheet motivy**

α-šroubovice

β-listy



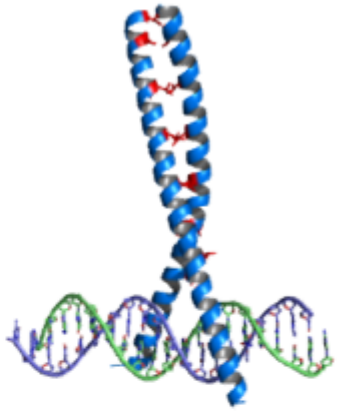
Luscombe et al, Genome Biology, 2000

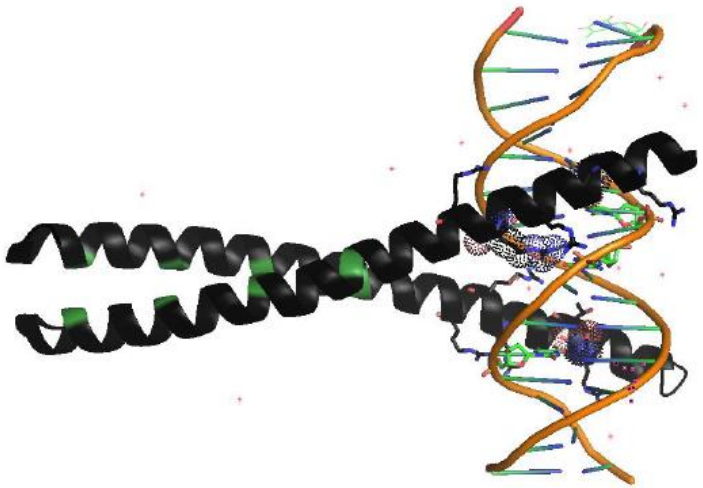
Rooman, Marianne and Wintjens, René (March 2015) Protein–DNA Interactions. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0001348.pub3

Motivy- Transkripční faktory - GCN4 a AP1 a c-Myc

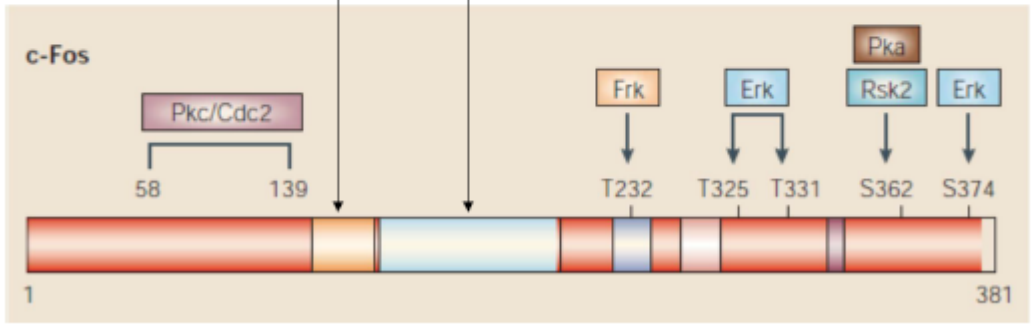
- **Zipper typ** (dle způsobu dimerizace)
  - **Leucinový zip** (tzv. bZIP = basic)

  (transcr. fact. yGCN4, c-Jun/c-Fos=AP-1)
    - 2 α-helixy (2 x 60 AMK)
    - coiled-coil (>30AMK, Leu, C-term)
    - bazická část (N-terminus, navazuje na CC)
    - bazická šroubovice vázána do VŽ

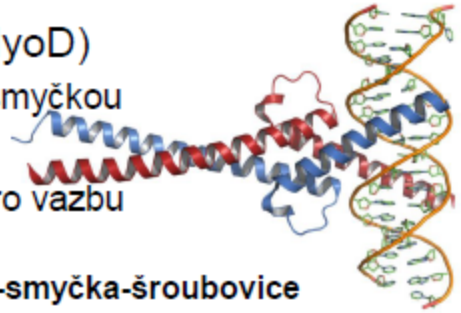




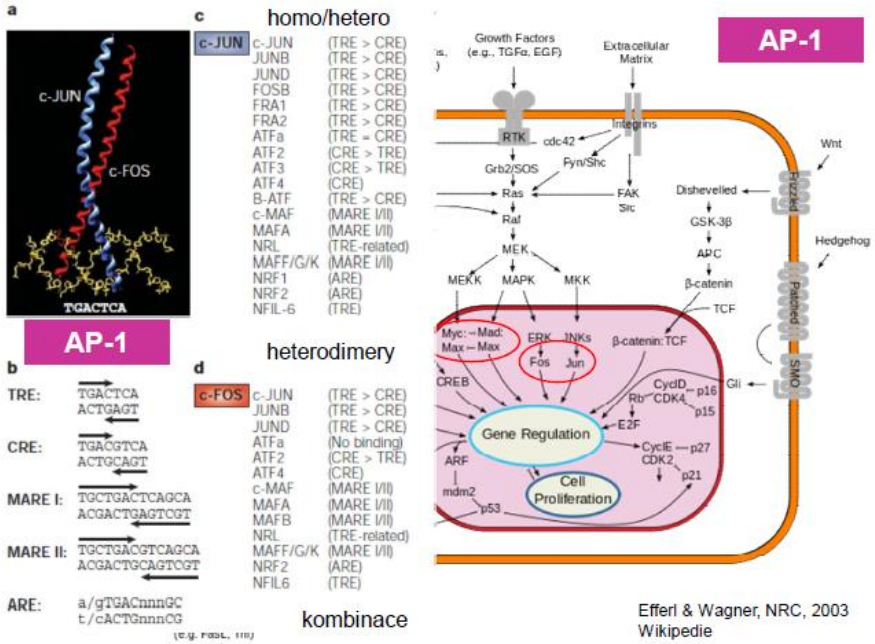
Interakce bazických AMK: Arg(232+240)=PO_4, Arg(243)=Gua
Konsensus sekvence: TGACTCA
GCN4 – regulace genů pro syntézu AMK



- **Helix-loop-helix** (c-Myc/Max, MyoD)
  - CC a bazické části jsou odděleny smyčkou
  - bazická šroubovice vázána do VŽ
  - smyčka poskytuje větší flexibilitu pro vazbu



šroubovice-smyčka-šroubovice

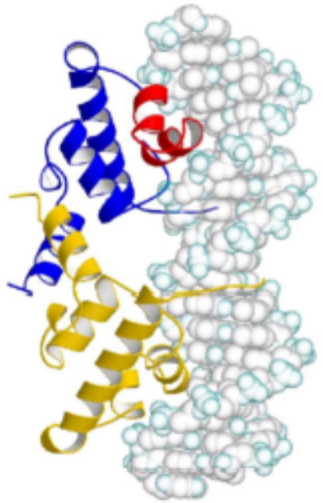


Efferl & Wagner, NRC, 2003
Wikipedie

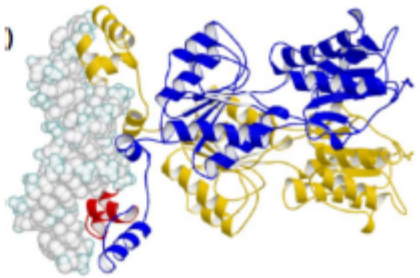
Helix-otáčka-helix
- HTH
- Winged helix
- TALE

Helix-turn-helix motiv (HTH)

- Obsahuje ~ 20 AMK ve dvou šroubovicích vzájemně kolmých
  - α-helix pro vazbu na DNA („recognition“) - β-obrátka – druhá šroubovice
  - Sekvenčně-specifická vazba prostřednictvím „recognition“ šroubovice a velkého žlábku
  - nejčastější motiv u prokaryot - homodimery vážou palindrom. sekvence
  - HTH motiv se obvykle vyskytuje ve svazku 3-6 šroubovic (stabilizovaných hydrofobním jádrem)
  - motiv může být buď součástí hlavního proteinu (Cro) nebo z něj může pouze vybíhat (LacI)



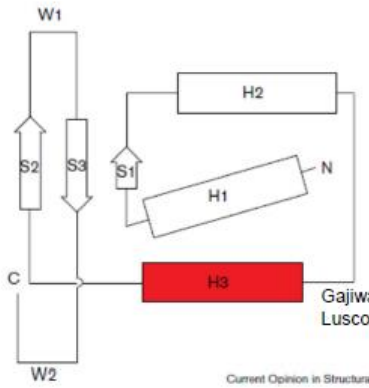
1. Cro and Repressor (1lmb)



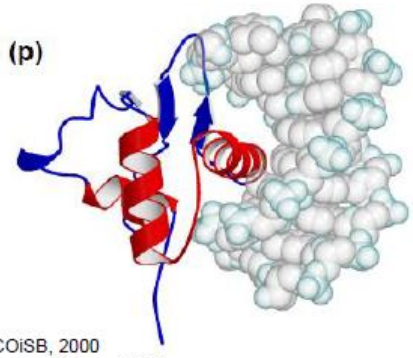
Luscombe et al, Genome Biology, 2000

3. LacI repressor (1wet)

- „winged“ HTH obsahuje „recognition“ šroubovici **(H3)** a β-listy, které poskytují další kontakty s DNA



Current Opinion in Structural Biology



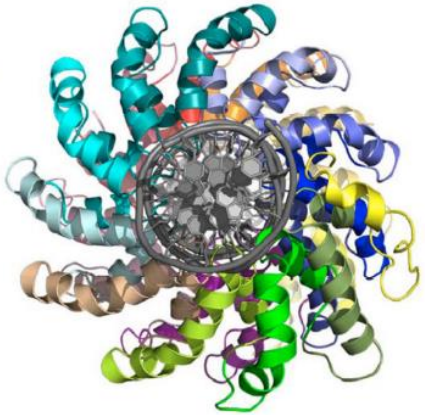
Gajiwala & Burley, COiSB, 2000
Luscombe et al, Genome Biology, 2000

16. Ets domain (1bc8)

Méně často křídlo ve VŽ a cukr-fosfátová kostra se šroubovicí (hRFX1)

Transcription activator-like effectors (TALE)
Patogenní bakterie injikují do rostlinných buněk ovlivňují transkripci rostlinných promotorů

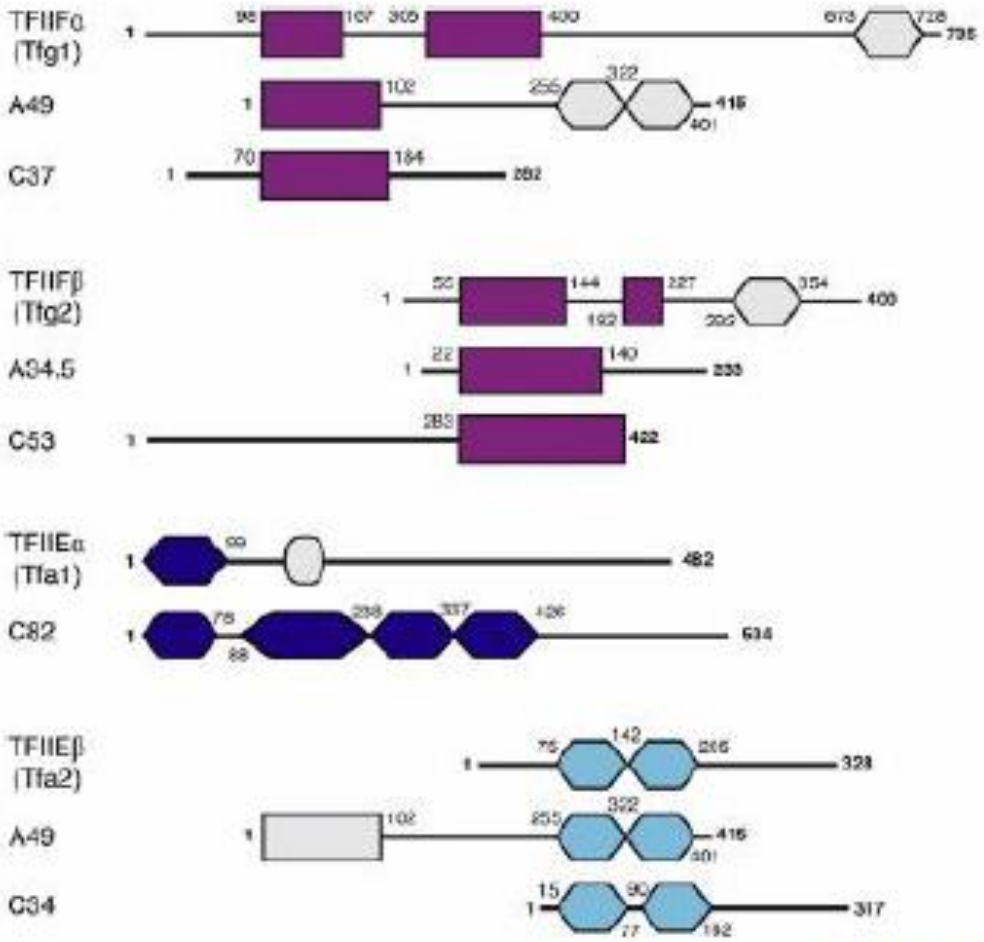
PthXo1
23 repetic obtáčí DNA ve VŽ

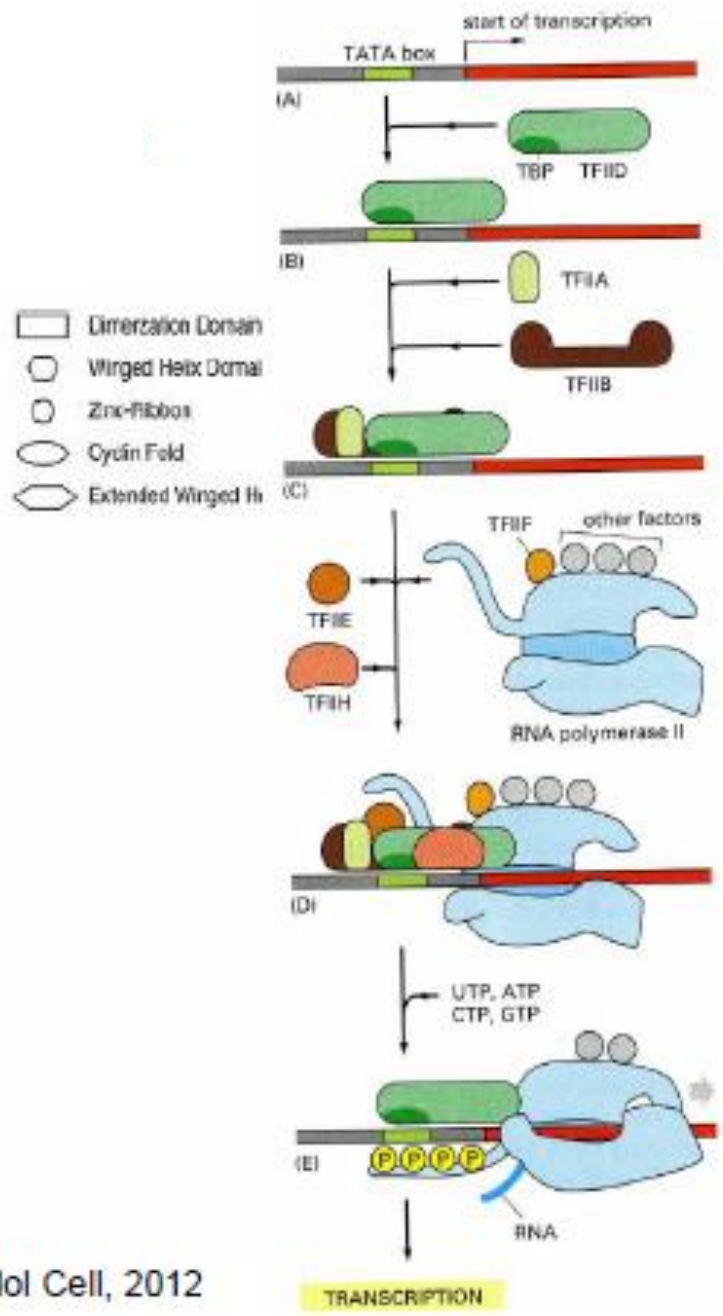


TALEN technologie

Mak et al, Science, 2012

General transcription factors-motive HTH

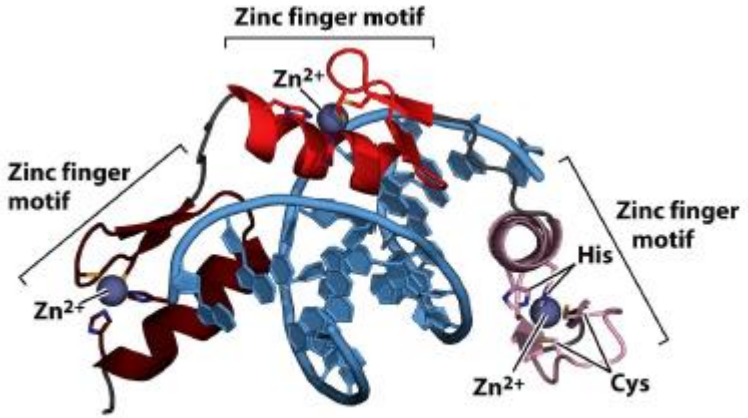




Vanini & Cramer, Mol Cell, 2012

Zinc finger

- ββα zinc-finger
- Hormon-receptor
- Loop-sheet-helix
- Gal4

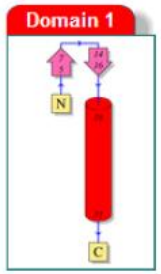


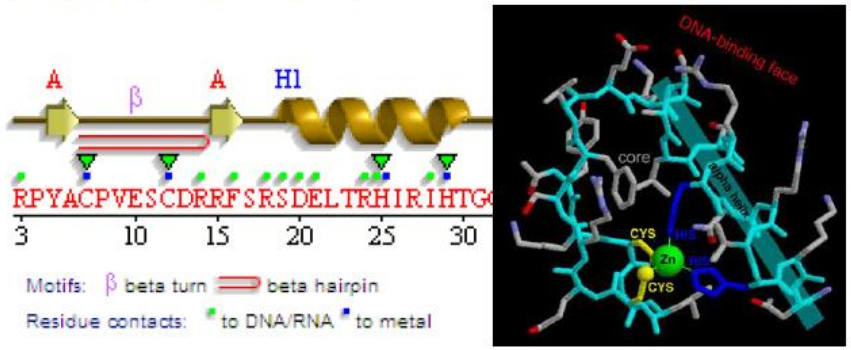
Zinc-finger/Zinkový prst

- cca 30 AMK ve dvou krátkých antiparalelních β-listech a α-šroubovici
- smyčka („hairpin“) stabilizovaná („crosslinked“) Zn^{2+} - koordinovaný 4xCys nebo 2xCys + 2xHis (tetraedrická struktura)

C2H2 motiv:
Cys-X_{2-4}-Cys-X_3-Phe-X_5-Leu-X_2-His-X_3-His

PDB grafika

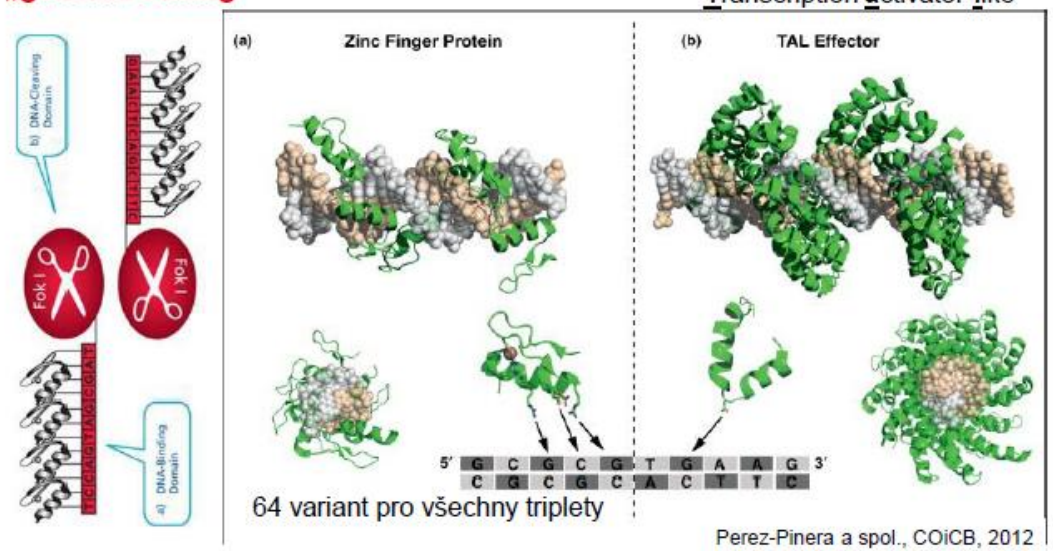




- Dobře charakterizované DNA-proteinové kontakty –je známá specifita ZFs pro všech 64 možných kombinací 3 sousedních bp
- Lze pro specifickou sekvenci DNA poskládat ZFs – nová technologie „zinc nuclease“ pro genové manipulace

„genome editing“

Transcription activator-like

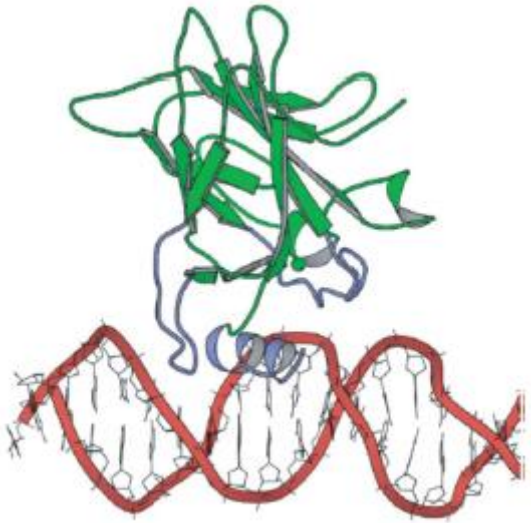


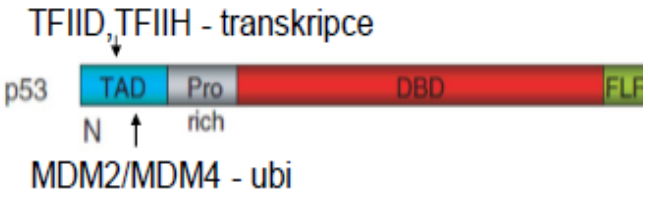
Perez-Pinera a spol., COiCB, 2012

DNA-interacting p53 protein

Loop-sheet-helix

- loops coming out of main core domain - protrudes b-sheet and a-helix
- 3 Cys and 1His coordinate Zn helix in a large groove and loop in a small groove
- Activation of transcription through acidic TA domain
- core / DNA-binding domain p53 - transcription factor important for cell cycle regulation, apoptosis and repair of the damaged DNA (tumor suppressor)

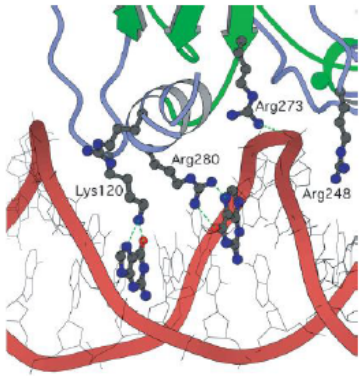


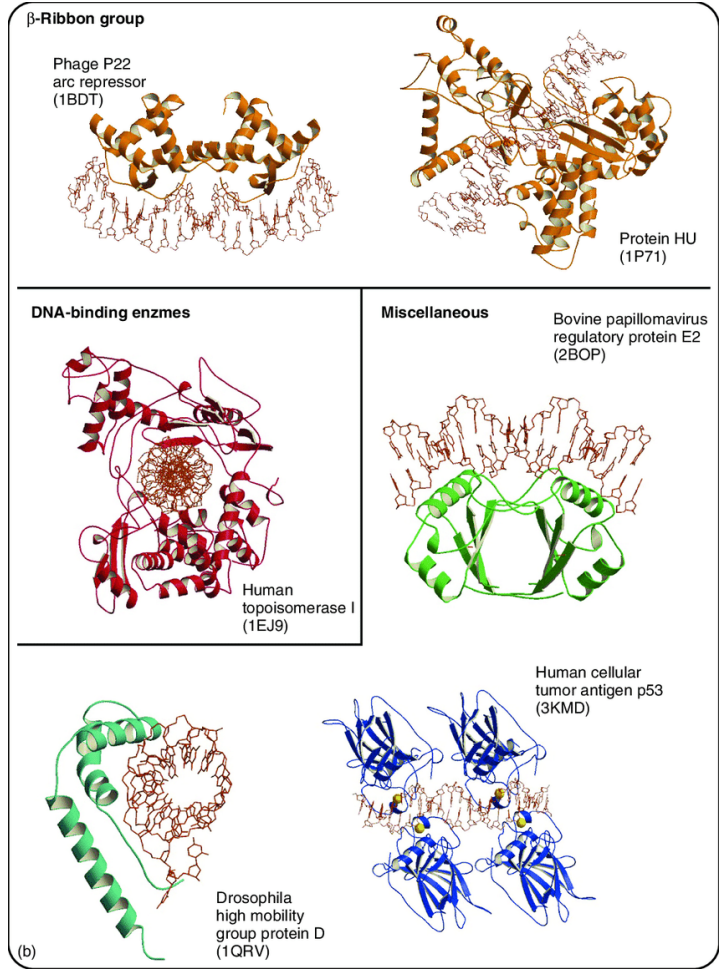


Loop-sheet-helix

- core/DNA-vazebná doména p53 – transkripční faktor důležitý pro regulaci buněčného cyklu, apoptozy a opravy poškozené DNA (nádorový supresor)
  - Konsensus sekvence PuPuPuC(A/T)(T/A)GPyPyPy (v promotorech p21, PUMA)
  - 95% "nádorových" mutací je v „core" doméně (R273H)
  - Regulace/aktivace modifkací C-koncové domény

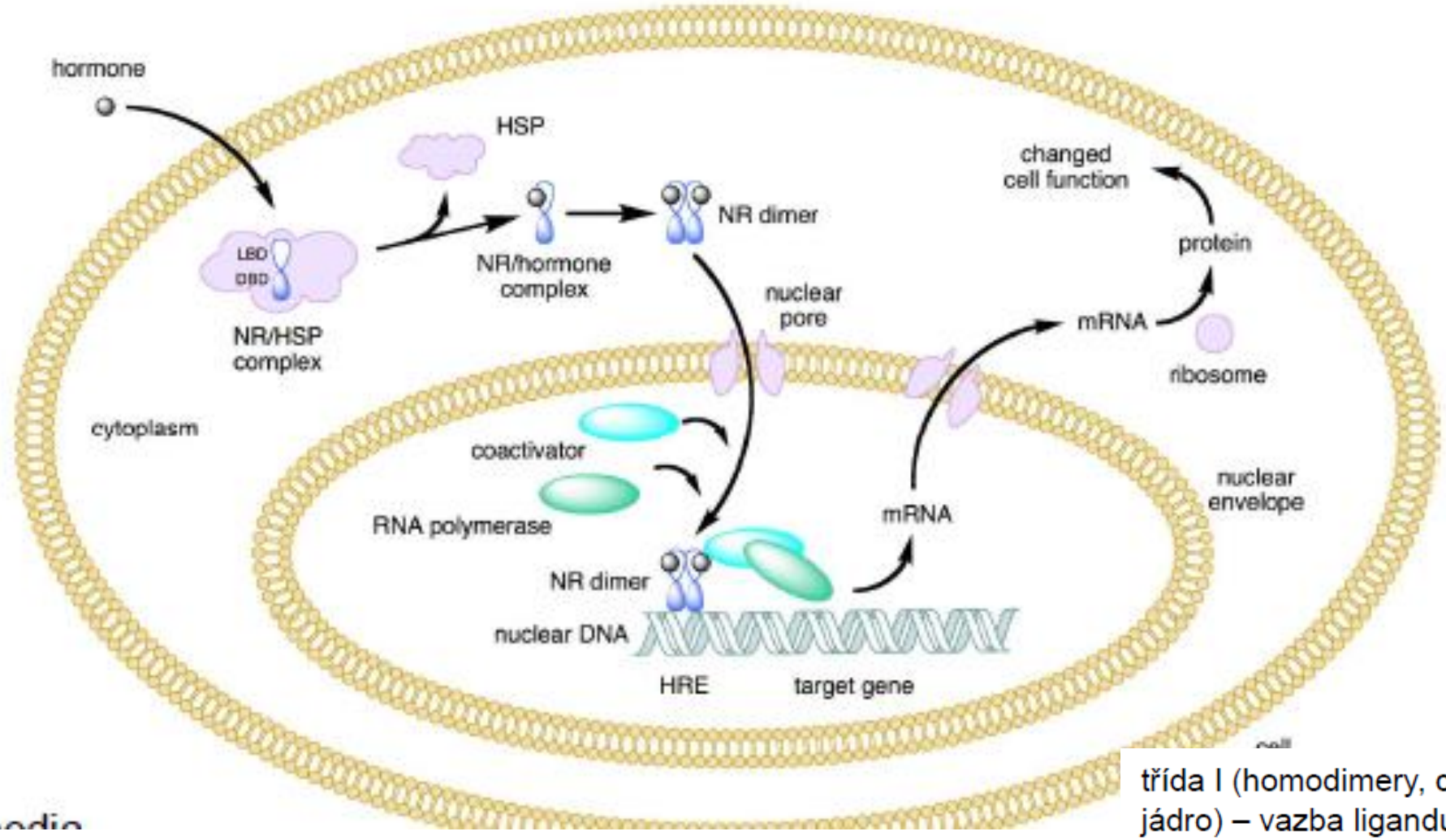
Protein se váže jako tetramer (C-koncová doména)





Rooman, Marianne and Wintjens, René (March 2015) Protein–DNA Interactions. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0001348.pub3

Motives receptors for hormons, loop-sheet-helix, GAL4

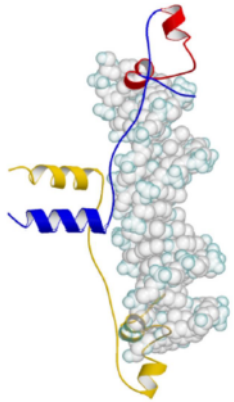


Gal4

pční faktor reguluje v kvasinkách metabolismus
(kvasinkový dvou-hybridní systém)

šroubovice
s koordinuje 2 Zn (2 Cys
ené 2 Zn)
elix ve velkém žlábku a 2.
takt s cukr-fosfátovou
trou
erizuje přes krátký CC
ment

l.: Nature, 1992



20. Gal4-type (1d66)

třída I (homodimery, cytoplasma) a třída II (heterodimery, jádro) – vazba ligandu také moduluje vazbu ko-aktivátorů (dalších transkripčních faktorů nebo chromatinových remodelátorů)

Regulation at the transcriptional level

Basic regulation of transcription (common to all genes)

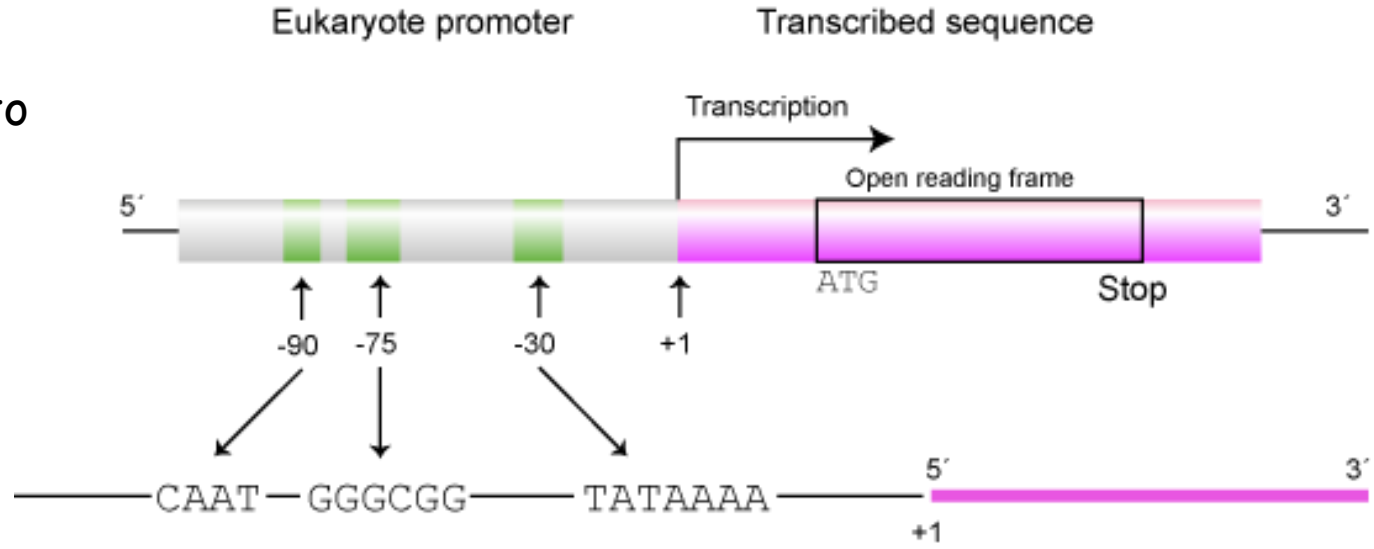
Regulation by components of the " basal transcription complex " (RNA polymerase binding to the TATA box, TATA binding proteins and other "basal" transcription factors binding to the RNA polymerase or in the promoter region)

Genes regulated only in this way:

Constitutively expressed genes

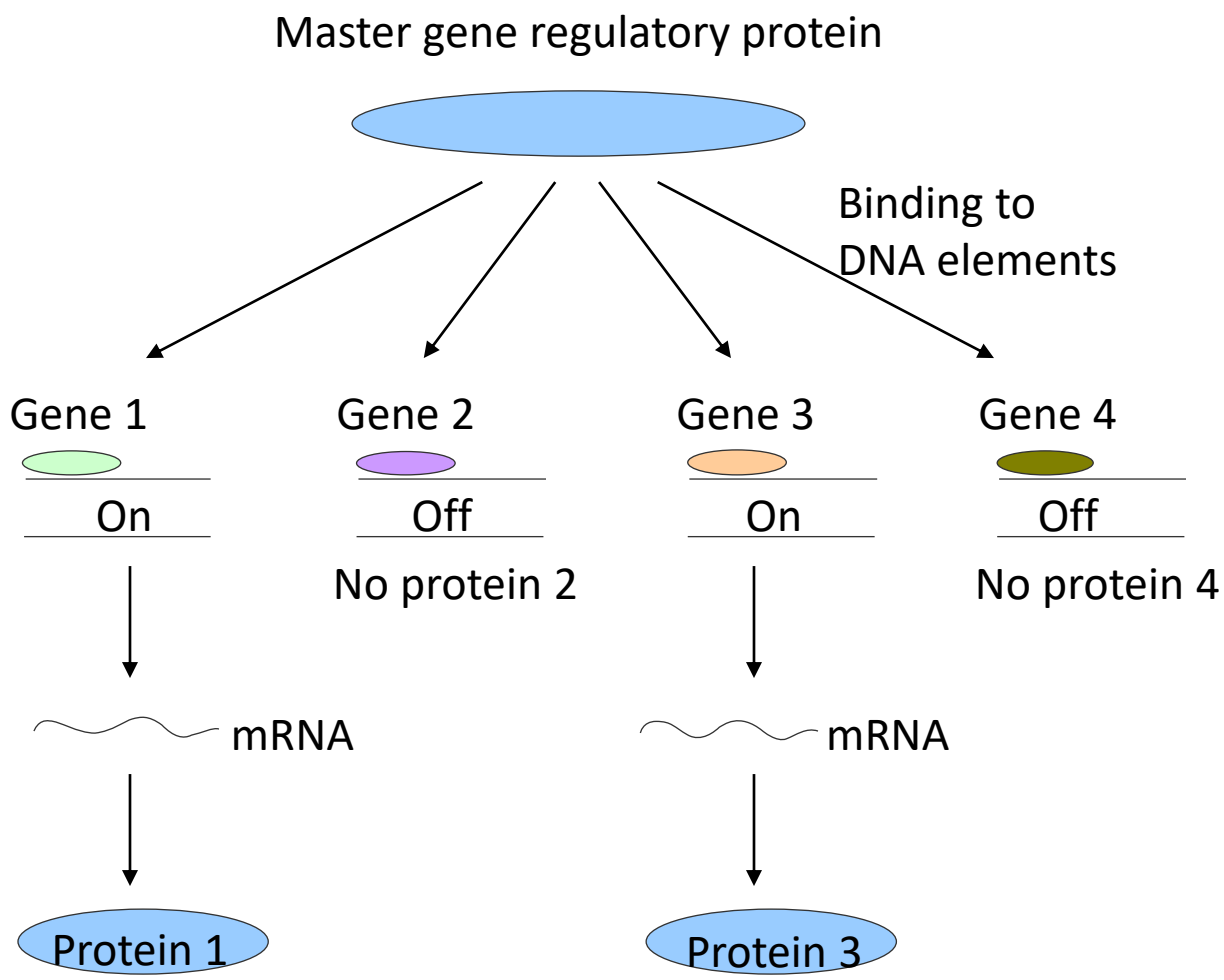
Specific effects on gene expression:

Through regulatory sequences in DNA and specific transcription factors.



basal transcription complex

Scheme of the activity of a master gene



transcription factors

Necessary to initiate transcription

They usually induce transcription, exceptionally they ca
inhibit it

Their various combinations bind to the promoter before
the RNA polymerase is attached

General transcription factors

in all or most cell types

necessary to induce transcription

- **basal TF-low** activity, minimal cell requirements

most common: TFIIA, TFIIB, TFIID (includes a subuni
called TATA binding protein (TBP) - binds specifically t
the TATA box sequence), TFIIE, TFIIF and TFIIH)

- constitutive TF - increase the basal activity of the cel

according to the cell type, the basic requirements of the cell (present (and active) in the cell at all times

- **general transcription factors**, Sp1, NF1, *CCAAT*)
- **special transcription factors**

They apply to inducible transcription

- only in cells of certain tissues and certain situations

(example p53)

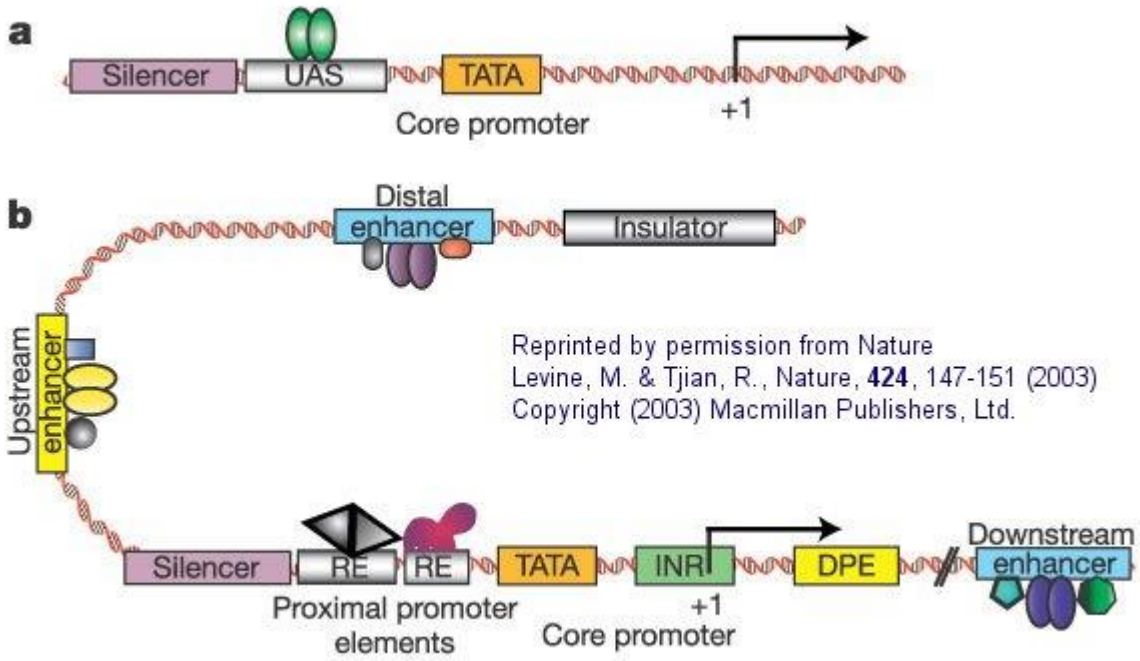


Figure 1 Comparison of a simple eukaryotic promoter and extensively diversified metazoan regulatory modules. **a**, **Simple eukaryotic transcriptional unit. A simple core promoter (TATA), upstream activator sequence (UAS) and silencer element spaced within 100–200 bp of the TATA box that is typically found in unicellular eukaryotes.** **b**, Complex metazoan transcriptional control modules. A complex arrangement of multiple clustered enhancer modules interspersed with silencer and insulator elements which can be located 10–50 kb either upstream or downstream of a composite core promoter containing TATA box (TATA), Initiator sequences (INR), and downstream promoter elements (DPE).

Animal Genetics - Gene exoression. [online]. [cit. 2014-08-15]. Dostupné z: http://web2.mendelu.cz/af_291_projekty2/vseo/stranka.php?kod=307

https://employees.csbsju.edu/hjakubowski/classes/ch331/bind/eukarypromNat0703e.htm

Terminology

Enhancers - regulatory sequences in DNA that bind transactivators

Transactivators bind coactivators

Silencers - regulatory sequences that bind the corepressor

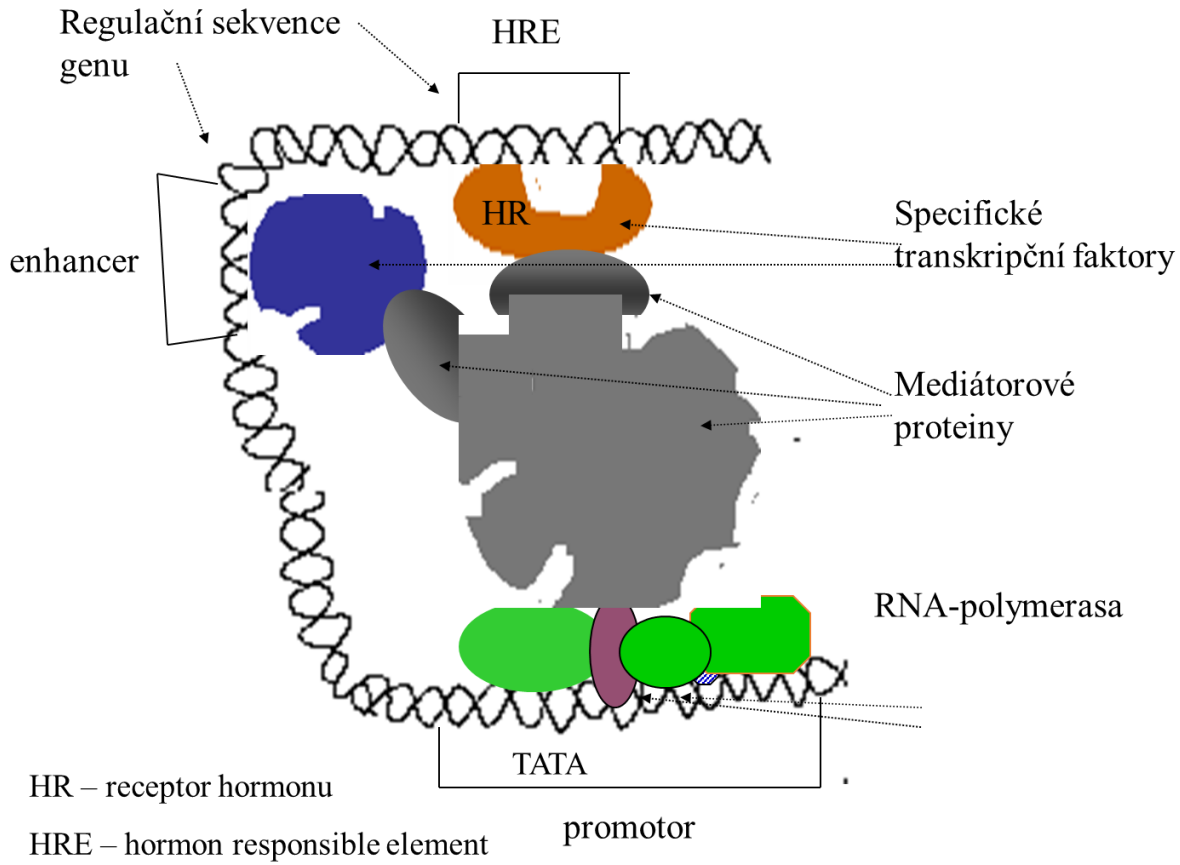
Hormones bind to the intracellular receptor, which binds to the hormone response element

These terms are still used. The terms are gradually being replaced:

regulatory sequences in DNA (enhancer, silencer, hormone response element)

specific transcription factors (different from basal transcription factors)

mediator proteins -coactivators



•**Transcription factors** are proteins that help turn specific genes "on" or "off" by binding to nearby DNA.
•Transcription factors that are **activators** boost a gene's transcription. **Repressors** decrease transcription.
•Groups of transcription factor binding sites called **enhancers** and **silencers** can turn a gene on/off in specific parts of the body.
•Transcription factors allow cells to perform logic operations and combine different sources of information to "decide" whether to express a gene.

➢Components of the eukaryotic promoter:

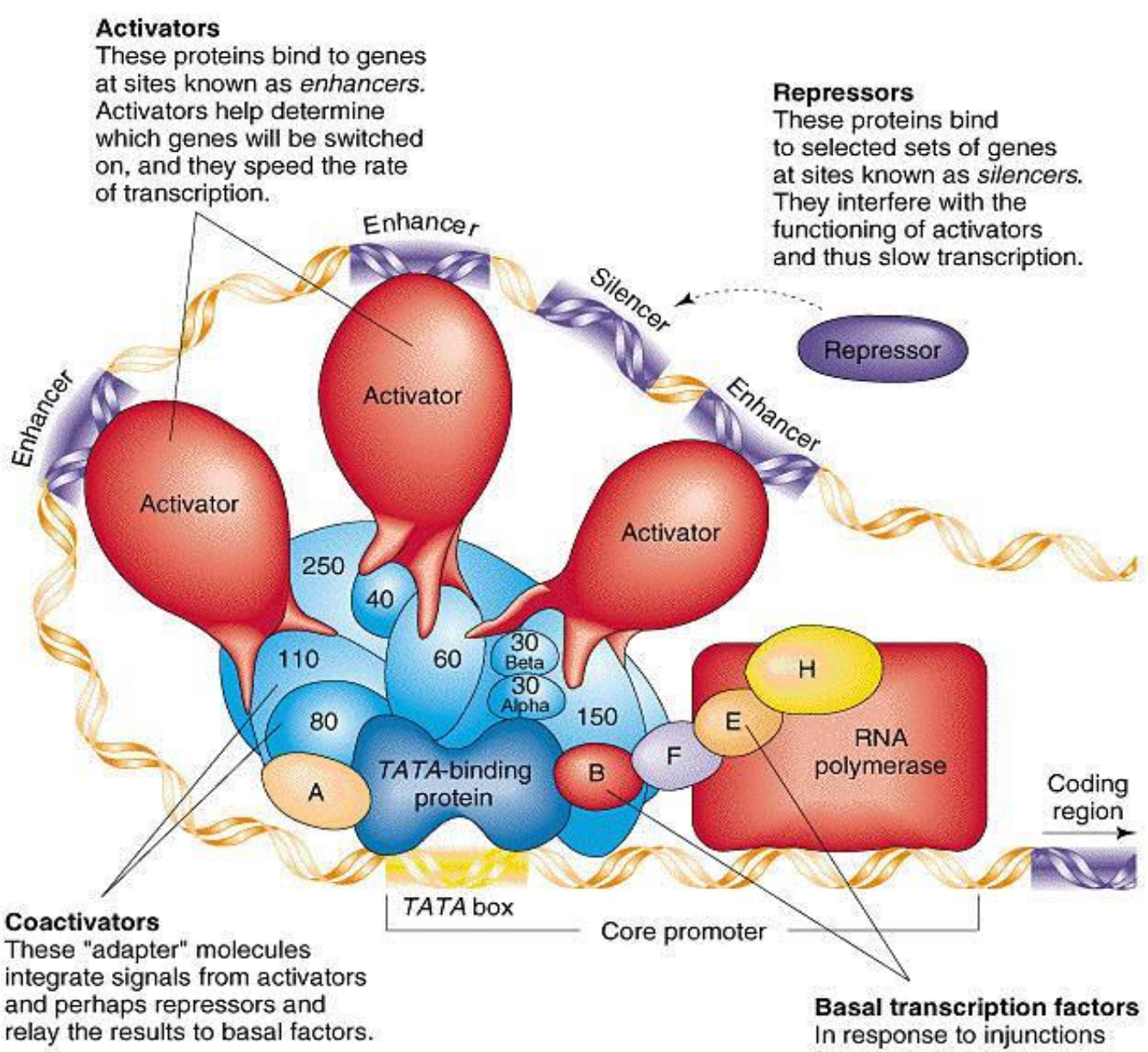
- Basal transcription factors

(basal, constitutive)

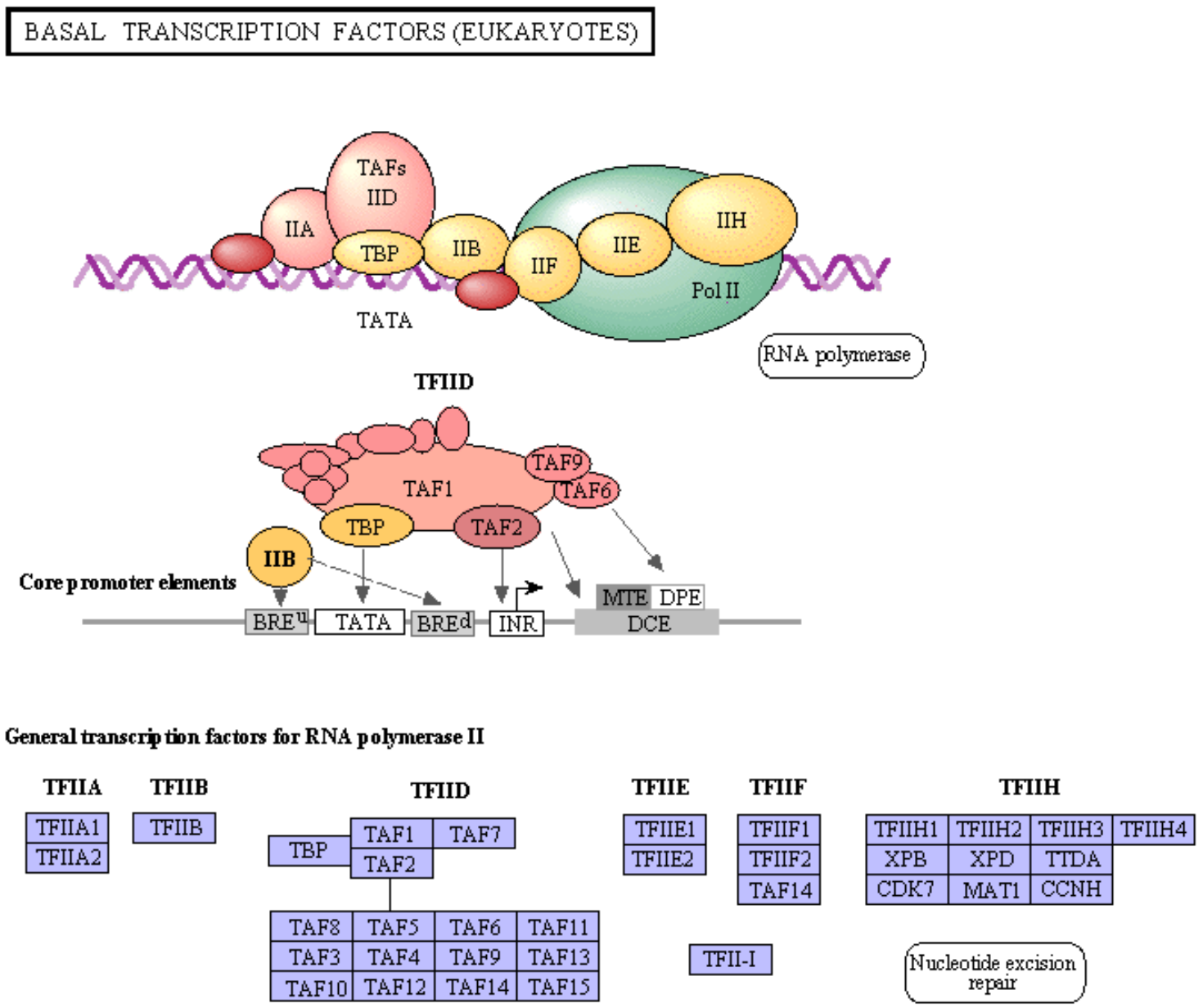
Special transcription factors

1.Constitutive - present (and active) in the cell at all times - general transcription factors, Sp1, NF1, CCAAT

2.Conditionally active - their activation required



http://www.cbs.dtu.dk/dtucourse/cookbooks/dave/Lekt03bkg.html



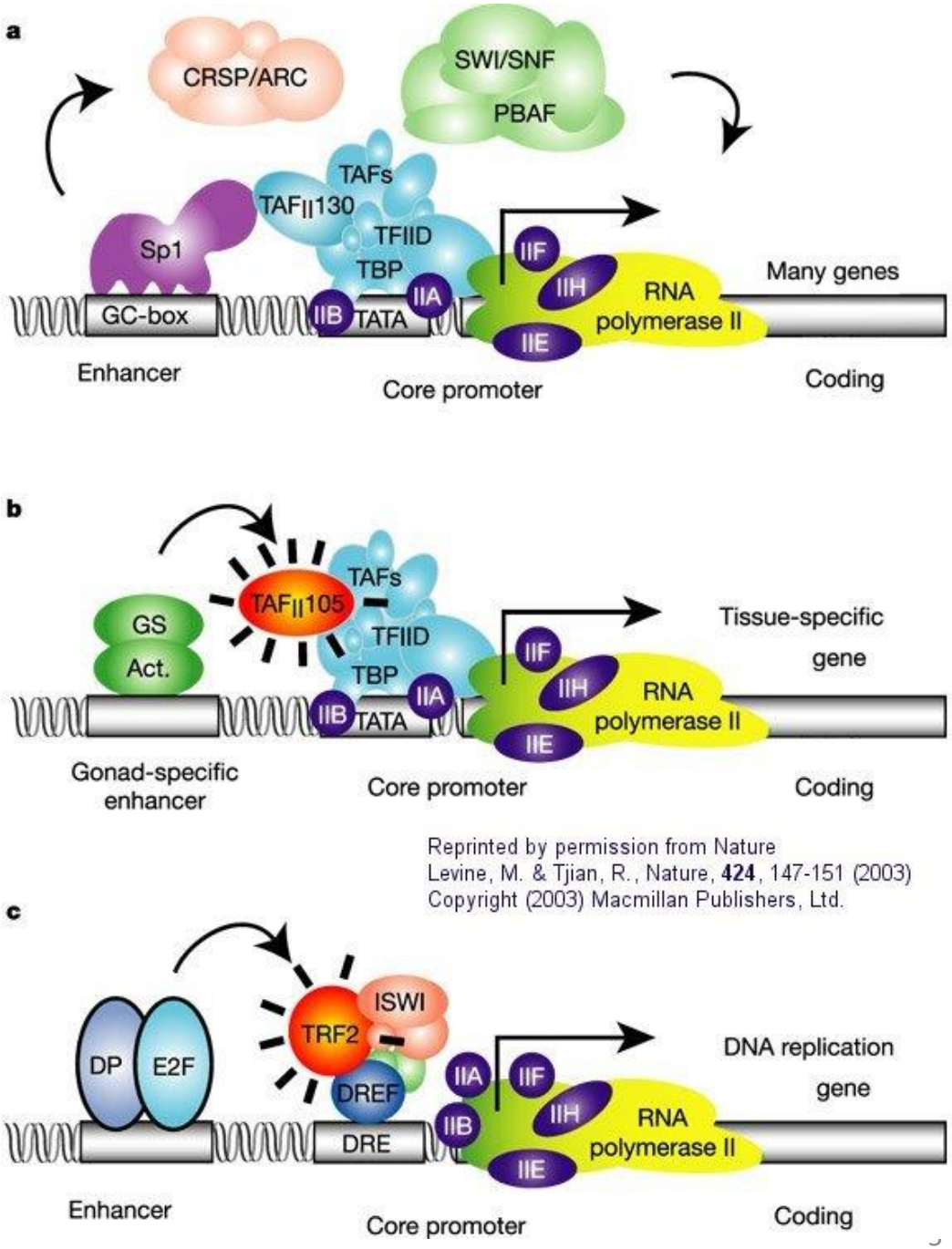
Mechanical division

Transcription factors of the general transcription complex (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH - are ubiquitous and react with the promoter (often TATA box) of structural genes, important in the development of vertebrates and invertebrates

Upstream transcription factors (UTF) - upstream - towards the 5´ part, proteins that bind to the regulatory part of the RNA polymerase I promoter at position -110 to -180, the presence is not necessary to initiate transcription, but multiplies its efficiency (it can also repressive)

Inducible transcription factors - same as UTF, but need to be activated or inhibited

https://employees.csbsju.edu/hjakubowski/classes/ch331/bind/olbindtransciption.html



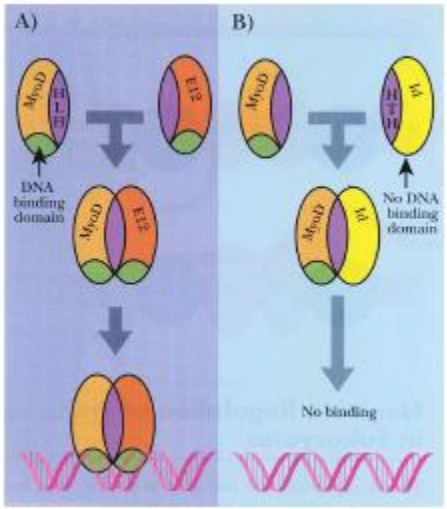
Reprinted by permission from Nature
Levine, M. & Tjian, R., Nature, **424**, 147-151 (2003)
Copyright (2003) Macmillan Publishers, Ltd.

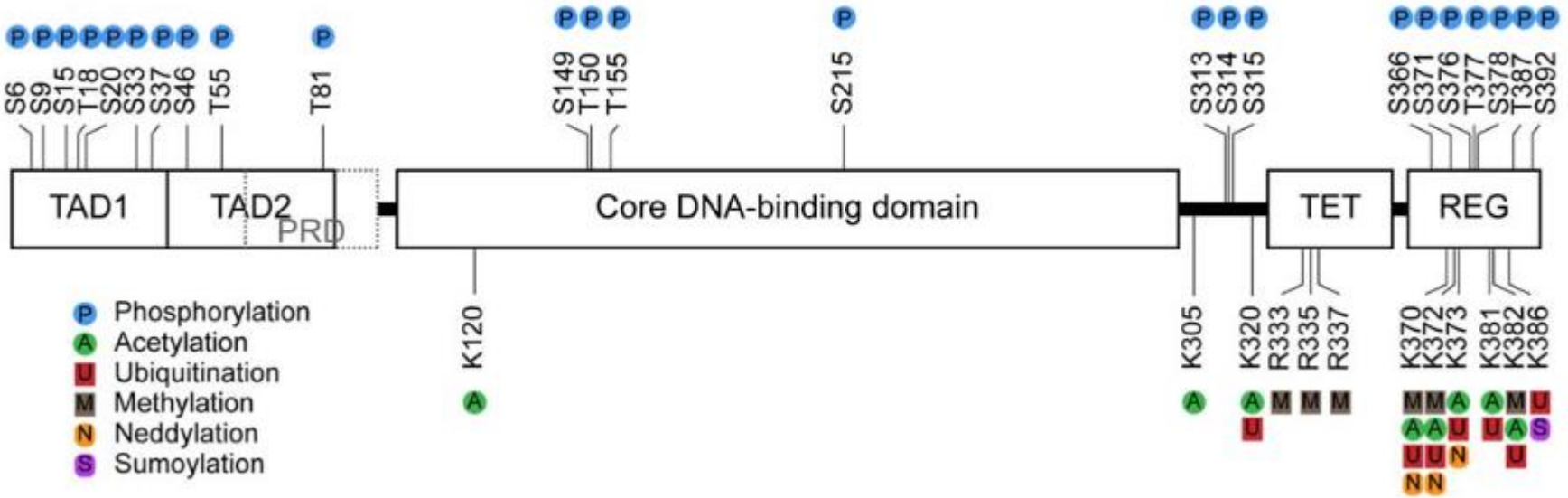
A. "The eukaryotic transcription apparatus can be divided into three sets, which include the RNA polymerase II core complex and related general transcription factors (TFIIA, -B, -D, -E, -F and -H), multi-subunit cofactors (mediator, CRSP, TRAP, ARC / DRIP, etc.) and various chromatin modifying or remodeling complexes (SWI / SNF, PBAF, ACF, NURF and RSF).
B. Metazoa organisms have developed multiple gene-selective and tissue-specific TFIID-like assemblies using alternative TAFs (TBP [TATA Binding Protein] -related factors such as ovarian-specific TAF105), as well as TRF (TBP-[TATA Binding Protein-associated factors] -related factors, such as is TRF2 in Drosophila and mice), which mediate the formation of specialized RNA polymerase initiation complexes that direct the transcription of tissue-specific and gene-selective expression programs. "(Natural link in the picture above.)""

Methods of activation of transcription factors:

- ligand-induced conformation change (signal, eg hormone)
- conformational change after removal of the inhibitory protein
- conformational transition induced by phosphorylation
- phosphorylation by protein kinase
- phosphatase depho
- stabilization of the transcription facto

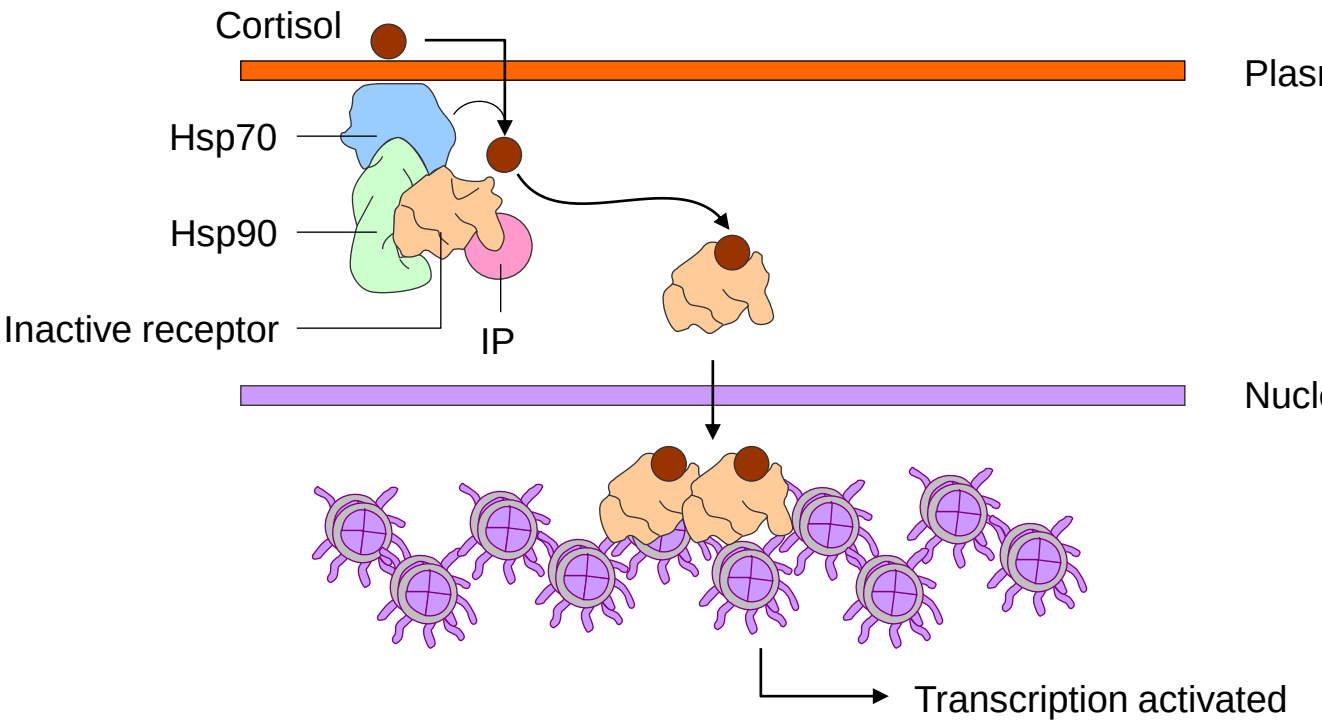
- transkripční faktor indukující expresi genů potřebných pro svalové buňky
- obsahuje doménu šroubovice-smyčka-šroubovice (HLH) pro vazbu na DNA a dimerizaci
- funguje jako heterodimer složený z tkáňově specifického proteinu HLH (MyoD) a obecného proteinu HLH (proteinu E)
- při heterodimerizaci s proteinem, který postrádá DNA vazebnou doménu (Id) k vazbě komplexu na DNA nemůže dojít (inhibice svalové diferenciace)





Gene regulation by members of the nuclear receptor superfamily

A. Glucocorticoid receptor

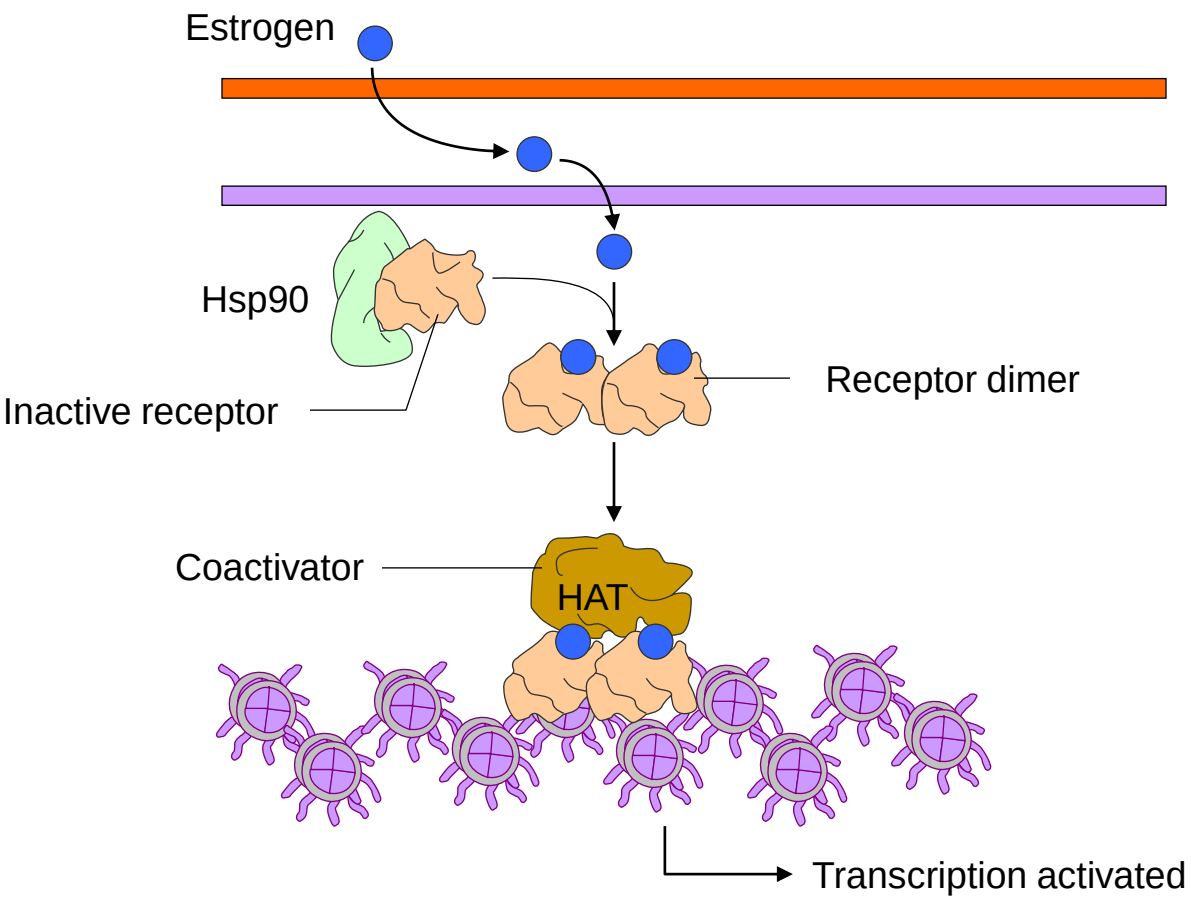


The glucocorticoid receptor (GR, or GCR), also known as NR3C1 (nuclear receptors of subfamily 3, group C, member 1), is a receptor to which cortisol and other glucocorticoids bind. GR is expressed in almost every cell in the body and regulates genes that control development, metabolism and the immune response. Because the receptor gene is expressed in several forms, it has many different (pleiotropic) effects in different parts of the body. When glucocorticoids bind to GR, its primary mechanism of action is the regulation of gene transcription. The unbound receptor resides in the cytosol of the cell. After the receptor is bound to the glucocorticoid, the receptor-glucocorticoid complex can proceed in either of two ways. The activated GR complex regulates the expression of anti-inflammatory proteins in the nucleus or suppresses the expression of pro-inflammatory proteins in the cytosol (by preventing the translocation of other transcription factors from the cytosol to the nucleus). In humans, the GR protein is encoded by the NR3C1 gene, which is located on chromosome 5 (5q31). Glucocorticoid receptor - https://en.qaz.wiki/wiki/Glucocorticoid_receptor

5111362

Gene regulation by members of the nuclear receptor superfamily

B. Estrogen receptor



HAT – histone acetyltransferase

.[2] Estrogen receptors (ERs) are steroid receptors present in the cell nucleus [1] of vertebrates to which estrogen binds. Humans and other mammals have two types of estrogen receptors, the estrogen receptor α (ERα, also ESR1) and the estrogen receptor β (ERβ, also ESR2). Both receptors can form homodimers as well as common heterodimers. However, the GPER receptor, which is a special G protein-coupled receptor, also responds to estrogen. [2] All of these types of receptors also occur in other vertebrates, including fish. [2]

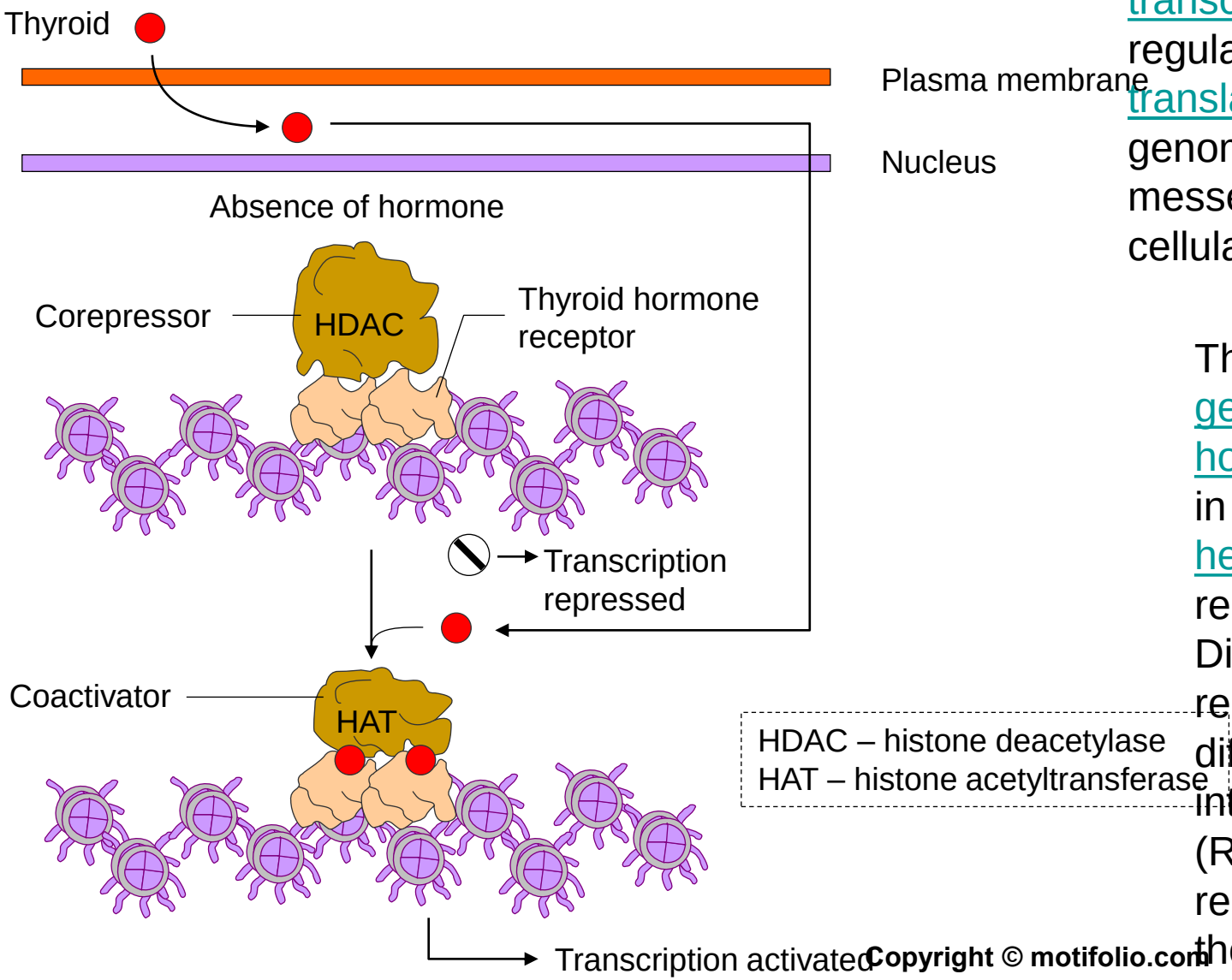
Estrogen receptors allow the detection of estrogen at specific sites in the vertebrate body. At rest, they are usually found in the cytosol, while upon binding to the ligand (estrogen), they are activated, dimerized, and enter the cell nucleus. There it binds to DNA sequences known as estrogen responsive units (EREs). The binding is also affected by other co-regulators (coactivators and corepressors). [3]

Due to its receptors, estrogen controls reproduction, both the development of the reproductive system and reproductive behavior. The best known, however, is the influence on the development of female (female) genitals. Furthermore has several functions not related to reproduction, e.g. affects bone density and strength, blood lipid levels, fat storage, and management of water with salts, as well as some higher brain functions (memory effect). However, it probably also affects the development of parts of the male reproductive system, such as sperm maturation. [2]

Co

Gene regulation by members of the nuclear receptor superfamily

C. Thyroid receptor



HDAC – histone deacetylase
HAT – histone acetyltransferase

5111364

The **thyroid hormone receptor** (**TR**)[1] is a type of nuclear receptor that is activated by binding thyroid hormone.[2] TRs act as transcription factors, ultimately affecting the regulation of gene transcription and translation. These receptors also have non-genomic effects that lead to second messenger activation, and corresponding cellular response.[3]

Thyroid hormone receptors regulate gene expression by binding to hormone response elements (HREs) in DNA either as monomers, heterodimers with other nuclear receptors, or homodimers.[4] Dimerizing with different nuclear receptors leads to the regulation of different genes. THR commonly interacts with the retinoid X receptor (RXR), a nuclear retinoic acid receptor.[9] TR/RXR heterodimers are the most transcriptionally active form of TR.[10]

Regulation after transcription

Alternative splicing, miRNAs and siRNAs, translation initiation factors, & protein modifications.

- Even after a gene has been transcribed, gene expression can still be regulated at various stages.
- Some transcripts can undergo **alternative splicing**, making different mRNAs and proteins from the same RNA transcript.
- Some mRNAs are targeted by **microRNAs**, small regulator RNAs that can cause an mRNA to be chopped up or block translation.
- A protein's activity may be regulated after translation, for example, through removal of amino acids or addition of chemical groups.

Regulation of RNA level

rare in bacteria, common in higher organisms

RNA processing- alternative splicing

mRNA stability - for many genes, RNA interference affects life span or translation rate

translation - regulatory proteins bind to mRNA and / or the ribosome and affect the translation rate

control options:

mRNA degradation rate control (mRNA stability)

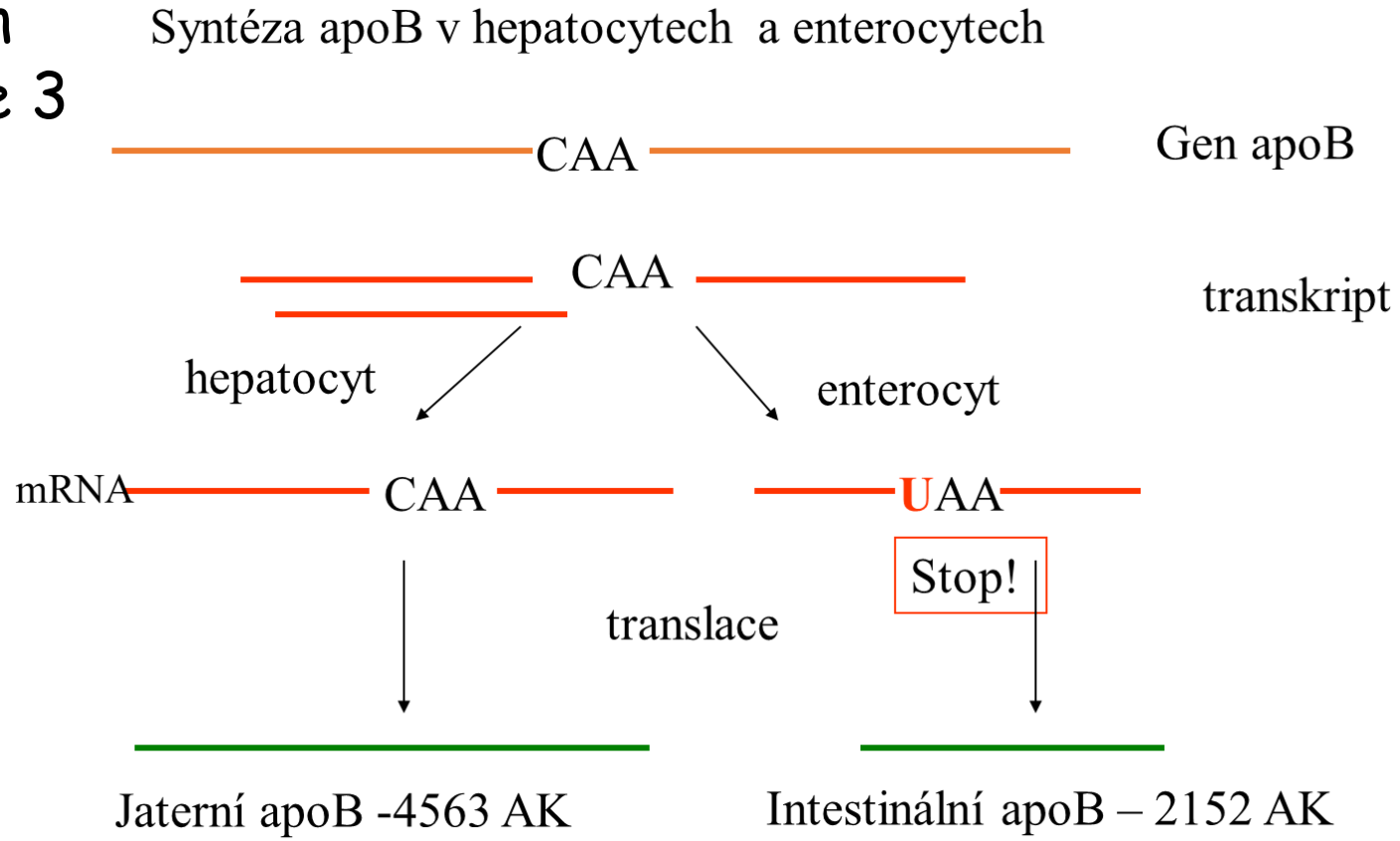
converting the non-translatable mRNA into a form that can be translated

translational control by regulatory proteins

binding of antisense RNA to mRNA

Regulation of gene expression by transcriptional modification

Alternative splicing and variation of thepolyadenylation site at the 3 'end causes a single gene to produce different proteins



RNA editing

In some cases, the RNA may be edited after transcription.

The primary transcript (hnRNA) is identical, after transcription there is a base exchange or nucleotide addition (deletion)

Syntéza apoB v hepatocytech a enterocytech (je součástí chylomikronů a VLDL)

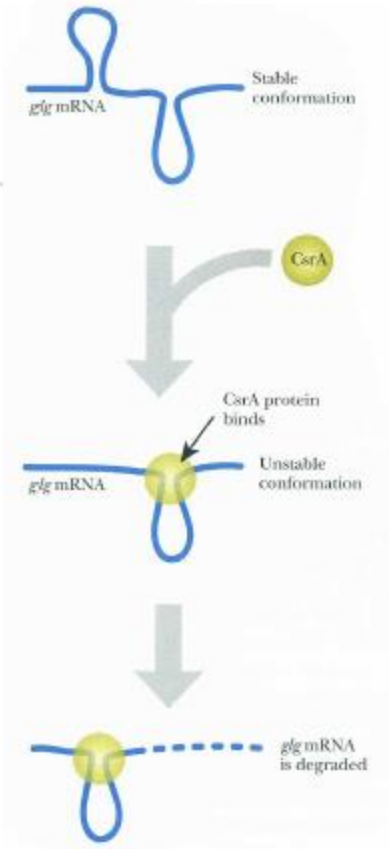
Gen apoB produkuje v játrech protein obsahující 4563 AK

Tentýž gen v enterocytech produkuje apoB obsahující jen 2152 AK

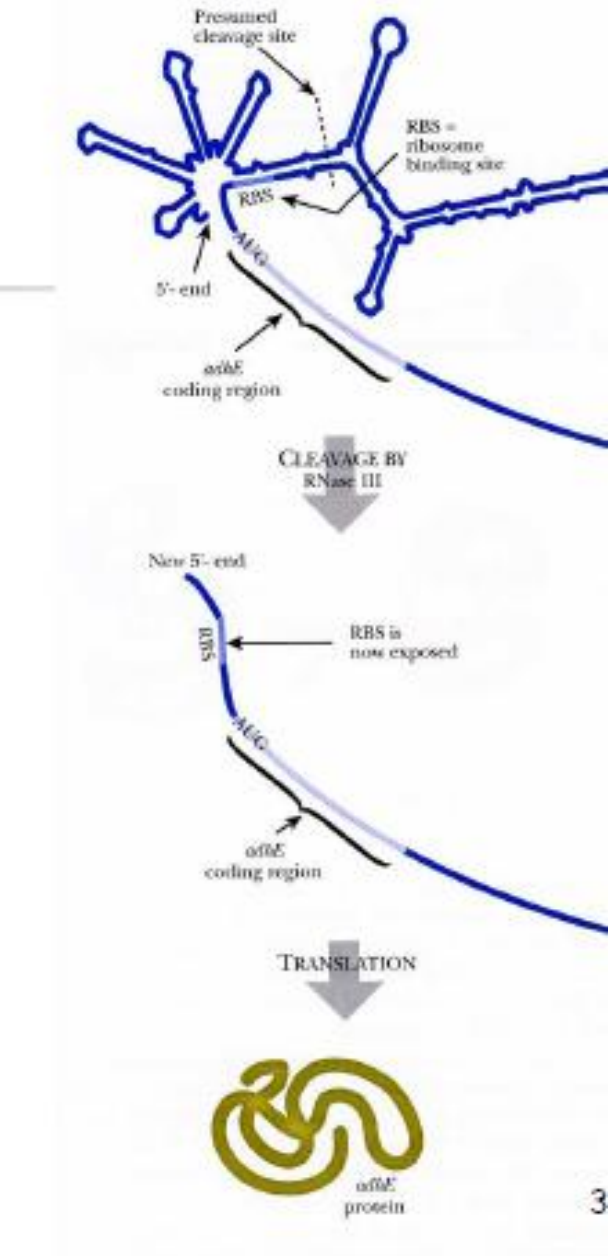
Konverze C(cytosin) na U (uracil) deaminací v RNA transkriptu generuje stop-kodón v intestinální mRNA. Tak protein produkovaný v enterocytu má pouze 48 % délky proteinu hepatálního

RNA level regulation

- RNA stability
- mRNA has a short half-life, upon degradation it undergoes ribonuclease degradation
- sensitivity to RNases depends on the secondary structure
- this may be affected by regulatory signals that induce the binding of regulatory proteins to RNA



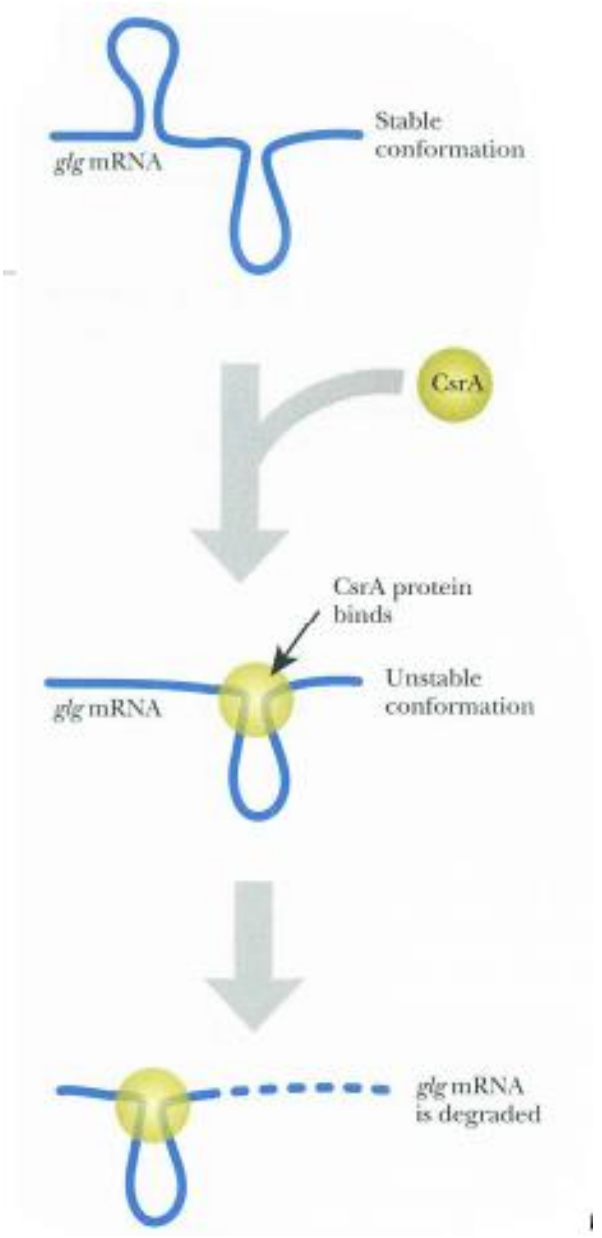
- Translation regulation
- the ribosome binding site (RBS) on the mRNA may be hidden by the secondary structure
- cleavage of a portion of the mRNA by RNase III restores RBS accessibility



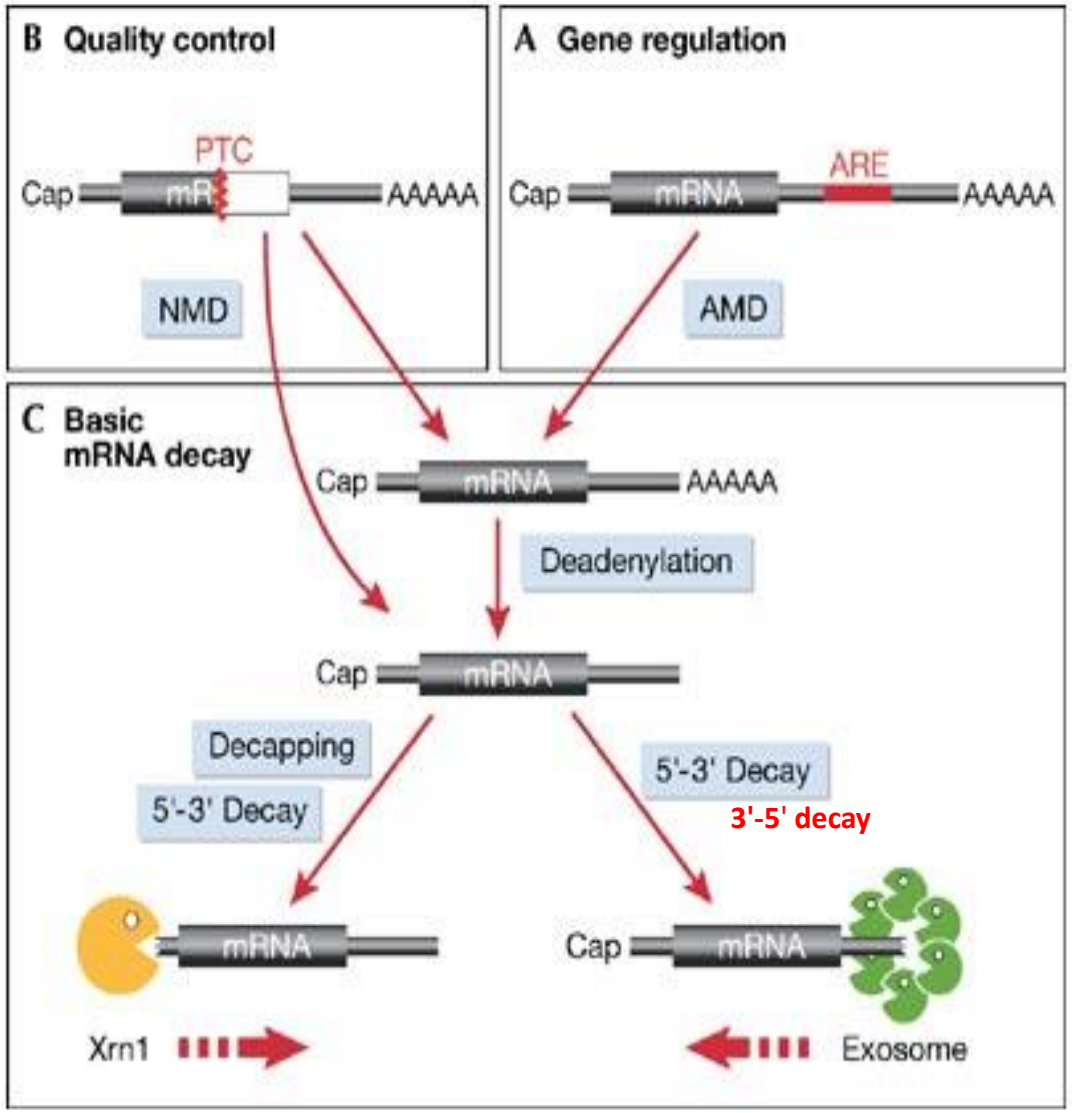
RNA stability

- mRNA has a short half-life, it is readily degraded by ribonucleases
- mRNA secondary structure is a key component in RNAse sensitivity

- mRNA secondary structure can be altered by protein binding – regulation signals



General scheme of messenger RNA decay pathways.



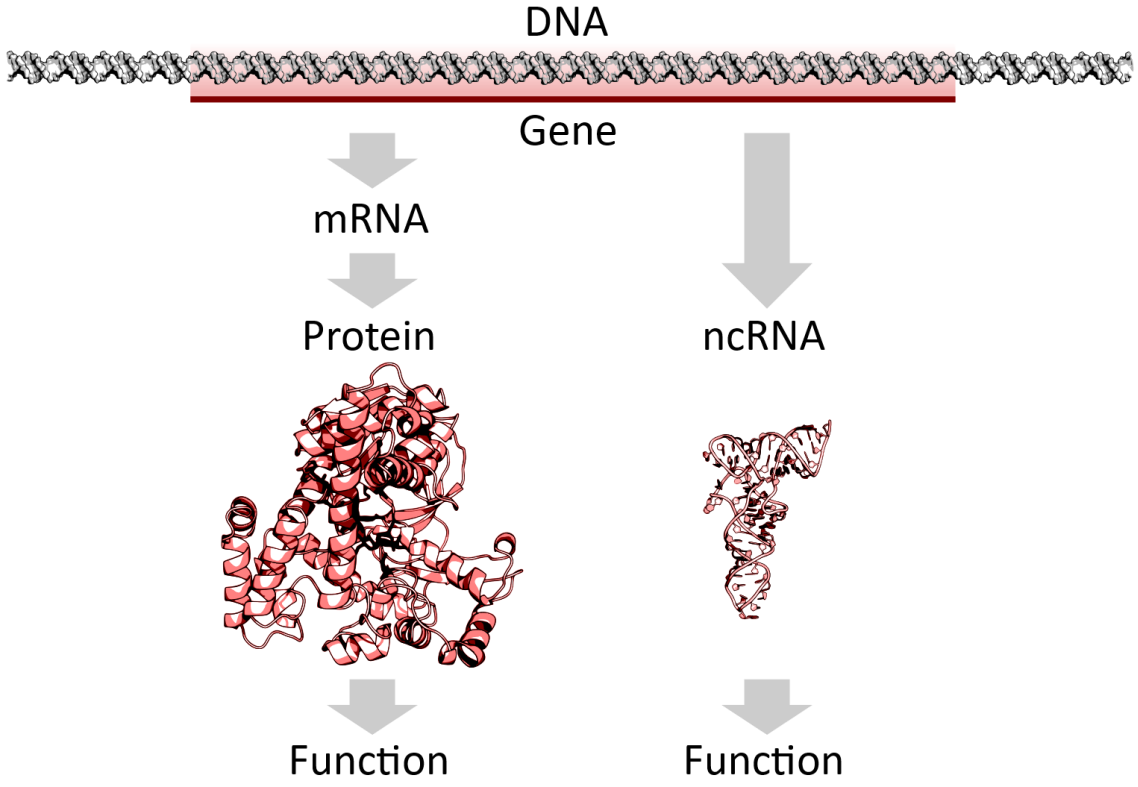
(A) mRNAs containing an AU-rich element (ARE) in their 3' UTR undergo rapid ARE-mediated mRNA decay (AMD) in resting cells. **Concealing ARE sequence from AMD induces gene expression. (B)** Quality control mechanisms. mRNAs that contain a premature termination codon (PTC) are recognized and specifically degraded by the nonsense-mediated mRNA decay (NMD) pathway. **(C)** The basic mRNA decay machinery in the cytoplasm initially removes the poly(A) tail through the activity of deadenylating enzymes. Subsequently, the mRNA can be further degraded from the 3' end by a complex of 3'–5' exonucleases known as the exosome. Alternatively, the mRNA is decapped at the 5' end, and the 5'–3' exonuclease Xrn1 proceeds to degrade the body of the mRNA.

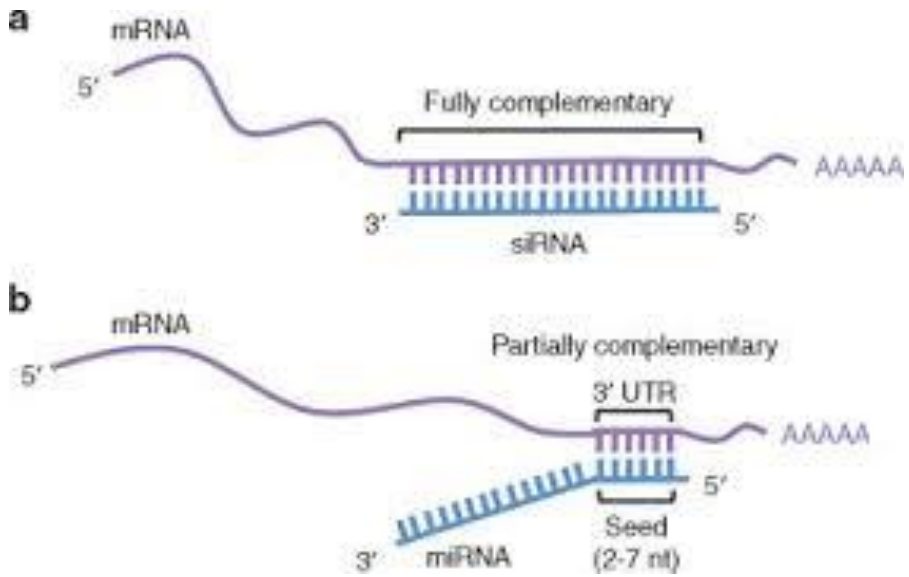
TEST

SUMMARY TABLE 18.1 **Regulating Gene Expression in Bacteria and Eukaryotes**

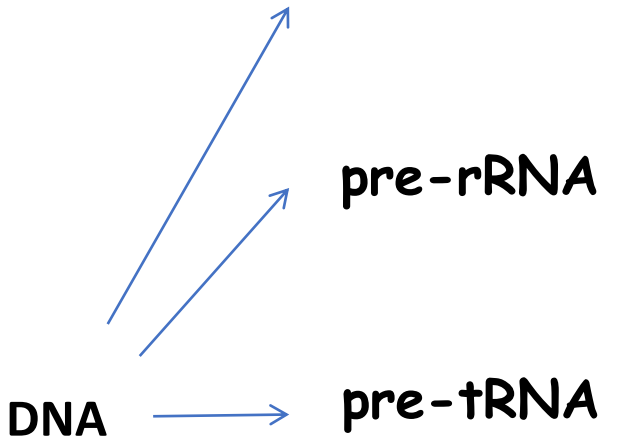
Level of Regulation	Bacteria	Eukaryotes
Chromatin remodeling	• Limited packaging of DNA • Remodeling not a major issue in regulating gene expression.	• Extensive packaging of DNA • Chromatin must be opened for transcription to begin.
Transcription	• Positive and negative control by regulatory proteins that act at sites close to the promoter • Sigma interacts with promoter.	• Positive and negative control by regulatory proteins that act at sites close to and far from promoter • Large basal transcription complex interacts with promoter. • Mediator complex required.
RNA processing	• None documented	• Extensive processing: alternative splicing of introns addition of 5′ cap and 3′ tail
mRNA stability	• Some RNA interference documented	• For many genes, RNA interference limits life span or translation rate.
Translation	• Regulatory proteins bind to mRNAs and/or ribosome and affect translation rate.	• Regulatory proteins bind to mRNAs and/or ribosome and affect translation rate.
Post-translational modification	• Folding by chaperone proteins • Chemical modification (e.g., phosphorylation) may change activity.	• Folding by chaperone proteins • Chemical modification (glycosylation, phosphorylation) • Ubiquination targets proteins for destruction by proteasome

Regulatory mechanisms mediated by transcription factors by RNA





Functional types of RNA



hnRNA	mRNA	Translation; encodes a sequence of proteins
pre-rRNA	rRNA	Translation; part of the ribosome
pre-tRNA	tRNA	Translation; transfer / amino acids activation

Small non-coding RNAs

Long non-coding RNAs

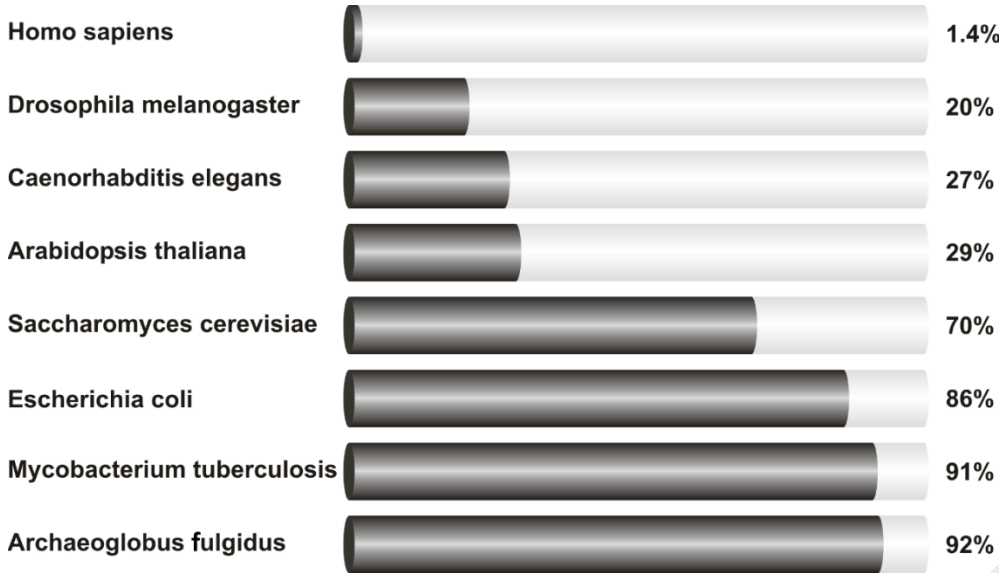
snRNA, snoRNA, ..Splicing / modification of RNA

siRNA, miRNA, … Regulation of gene expression

World of noncoding RNAs

encoding genes represent less than 2% of the total genome sequence
vs.
at least 90% of the human genome is actively transcribed
the more complex organism, the more it comprises non-coding RNAs

The percentage of protein-coding genes sequences in several eukaryotic and bacterial genomes.



Recent evidence suggests that the non-coding RNAs (ncRNAs) may play major biological roles in cellular development, physiology and pathologies.
NcRNAs could be grouped into two major classes based on the transcript size: small ncRNAs and long ncRNAs.

Small non-coding RNAs

miRNA
siRNA
piRNA
snoRNA
PARS
tiRNA

microRNA (miRNA)
Piwi-interacting RNA (piRNA)
small interfering RNA (siRNA)
small nucleolar RNA (snoRNAs)
tRNA-derived small RNA (tsRNA)
small rDNA-derived RNA (srRNA)
small nuclear RNA,
also commonly referred to as U-RNA

Long non-coding RNAs

lincRNA
TERRAs
T-UCR

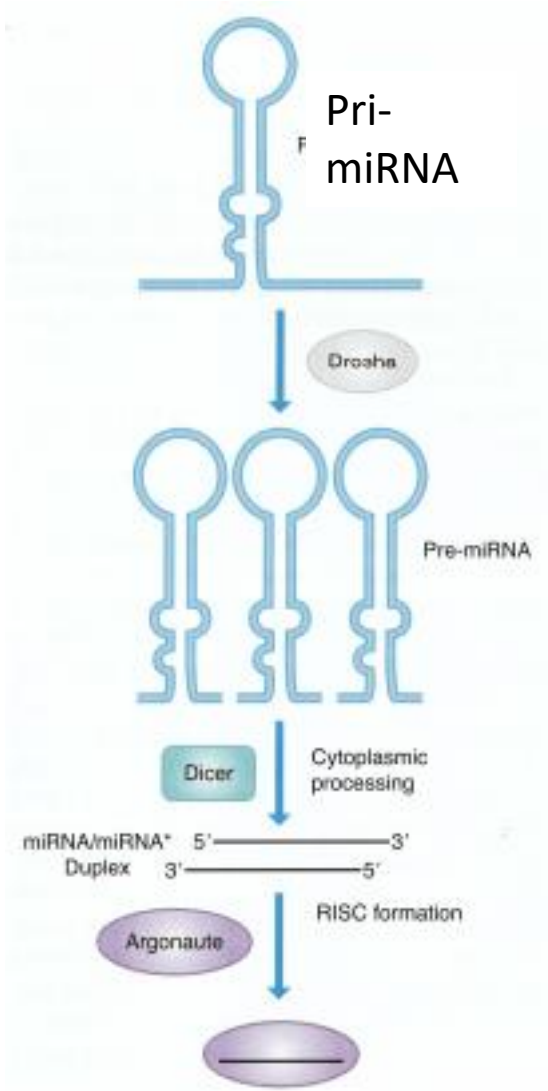
Table 1 Types of recently discovered human non-coding RNAs

	Class	Symbol	Characteristic	Disease / biological function associations
Small non-coding RNAs	MicroRNAs	miRNAs	18–25 nt; account 1–2% of the human genome; control the 50% of protein-coding genes; guide suppression of translation; Drosha and Dicer dependent small ncRNAs	initiation of various disorders including many, if not all, cancers / regulation of proliferation, differentiation, and apoptosis involved in human development
	Small interfering RNAs	siRNAs	19–23 nt; made by Dicer processing; guide sequence specific degradation of target mRNA	great potential in diseases treatment / posttranscriptional gene silencing mainly through RISC degradation mechanism; defence against pathogenic nucleic acids
	Piwi-interacting RNAs	piRNAs	26–30 nt; bind Piwi proteins; Dicer independent; exist in genome clusters; principally restricted to the germline and somatic cells bordering the germline	relationship between piRNAs and diseases has not yet been discovered / involved in germ cell development, stem self-renewal, and retrotransposon silencing
	Small nucleolar RNAs	snoRNAs	60–300 nt; enriched in the nucleolus; in vertebrate are excised from pre-mRNA introns; bind snoRNP proteins	association with development of some cancers / important function in the maturation of other non-coding RNAs, above all, rRNAs and snRNAs; miRNA-like snoRNAs regulate mRNAs
	Promoter-associated small RNAs	PASRs	20–200 nt; modified 5′ (capped) ends; coincide with the transcriptional start sites of protein- and non-coding genes; made from transcription of short capped transcripts	relationship with diseases has not yet been discovered / involved in the regulation of the transcription of protein-coding genes by targeting epigenetic silencing complexes
	Transcription initiation RNAs	tiRNAs	~ 18 nt ; have the highest density just downstream of transcriptional start sites; show patterns of positional conservation; preferentially located in GC-rich promoters	
	Centromere repeat associated small interacting RNAs	crasiRNAs	34–42 nt; processed from long dsRNAs	relationship between crasiRNAs and diseases has not yet been discovered / involved in the recruitment of heterochromatin and/or centromeric proteins
	Telomere-specific small RNAs	tel-sRNAs	~ 24 nt; Dicer independent; 2′-O-methylated at the 3′ terminus; evolutionarily conserved from protozoa to mammals; have not been described in human up to now	relationship between tel-sRNAs and diseases has not yet been discovered / epigenetic regulation
	Pyknons		subset of patterns of variable length; form mosaics in untranslated and protein-coding regions; more frequently in 3′ UTR	expected association with cancer biology / possible link with posttranscriptional silencing of genes, mainly involved in cell communication, regulation of transcription, signaling, transport,
Long non-coding RNAs	Long intergenic non-coding RNAs	lincRNAs	ranging from several hundreds to tens of thousands nts; lie within the genomic intervals between two genes; transcriptional cis-regulation of neighbouring genes	involved in tumorigenesis and cancer metastasis / involved in diverse biological processes such as dosage compensation and/or imprinting
	Long intronic non-coding RNAs		lie within the introns; evolutionary conserved; tissue and subcellular expression specified	aberrantly expressed in human cancers / possible link with posttranscriptional gene silencing
	Telomere-associated ncRNAs	TERRAs	100 bp - >9 kb; conserved among eukaryotes; synthesized from C-rich strand; polyadenylated; form inter-molecular G-quadruplex structure with single-stranded telomeric DNA	possible impact on telomere-associated diseases including many cancers / negative regulation of telomere length and activity through inhibition of telomerase
	Long non-coding RNAs with dual functions		both protein-coding and functionally regulatory RNA capacity	deregulation has been described in breast and ovarian tumors / modulate gene expression through diverse mechanisms
	Pseudogene RNAs		gene copies that have lost the ability to code for a protein; potential to regulate their protein-coding cousin; made through retrotransposition; tissue specific	often deregulated during tumorigenesis and cancer progression / regulation of tumor suppression and oncogenes by acting as microRNA decoys
	Transcribed-ultraconserved regions	T-UCRs	longer than 200 bp; absolutely conserved between orthologous regions of human, rat, and mouse; located in both intra- and intergenic regions	expression is often altered in some cancers; possible involvement in tumorigenesis / antisense inhibitors for protein-coding genes or other ncRNAs

RNA interference - RNAi

- **sequence-specific gene silencing mechanism triggered by double stranded RNA, on the post-transcriptional level or transcriptional level**
- inhibitory elements are small RNA molecules **(miRNAs, siRNAs….)**
- **miRNAs** generated by cleavage of **larger pre-miRNA molecules**
- nucleases **Drosha and DICER**, which are compiled into multiprotein complex **RISC** (RNA-induced silencing complex) with **proteins Argonaut**
- RNA interference is a process by which **noncoding RNA molecules** interfere (pair) with target regions of mRNA, resulting in **prevention of gene expression** of these mRNAs.
- For short, this proces is also called RNAi. We rank him among **posttransriptional mechanisms** of gene expression.

- **Most eucaryotic organisms is capable of RNA interference**, the process was first studied in the C. elegans.



RNA interference

RISC has helicase activity, thanks to which miRNA is loosened; only one chain remains associated with the complex

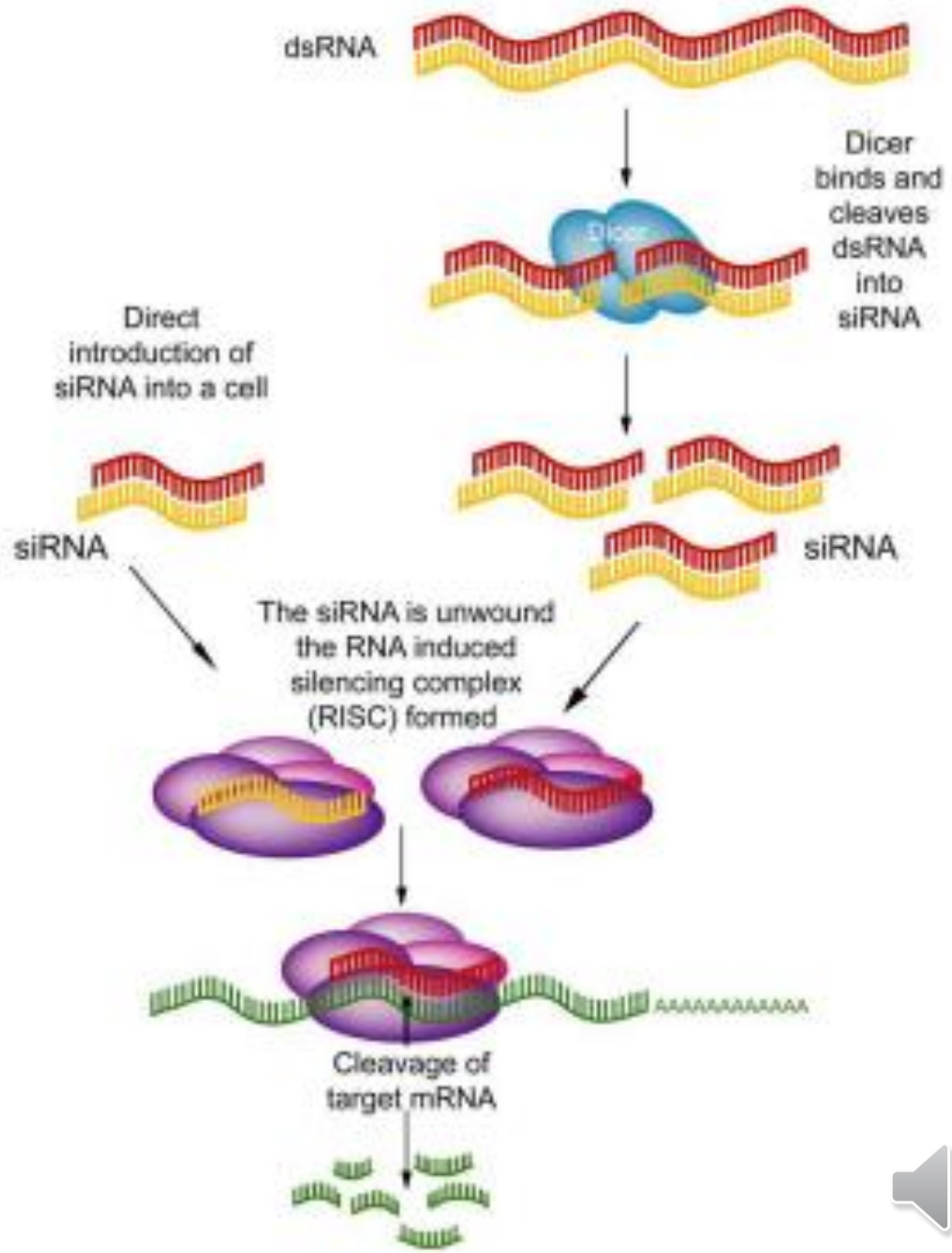
that **allows sequence-specific binding of the whole complex to the target complementary mRNA**

nuclease activity of RISC complex **cleaves the mRNA** - its degradation occurs

Originally protecting cells against viruses

common in eukaryotic cells

useful for targeted inactivation of genes: research of gene functions



Discovery of RNA interference (1998)

- silencing of gene expression with dsRNA

The Nobel Prize in Physiology or Medicine 2006

"for their discovery of RNA interference - gene silencing by double-stranded RNA"

Photo: L. Cicero

Andrew Z. Fire

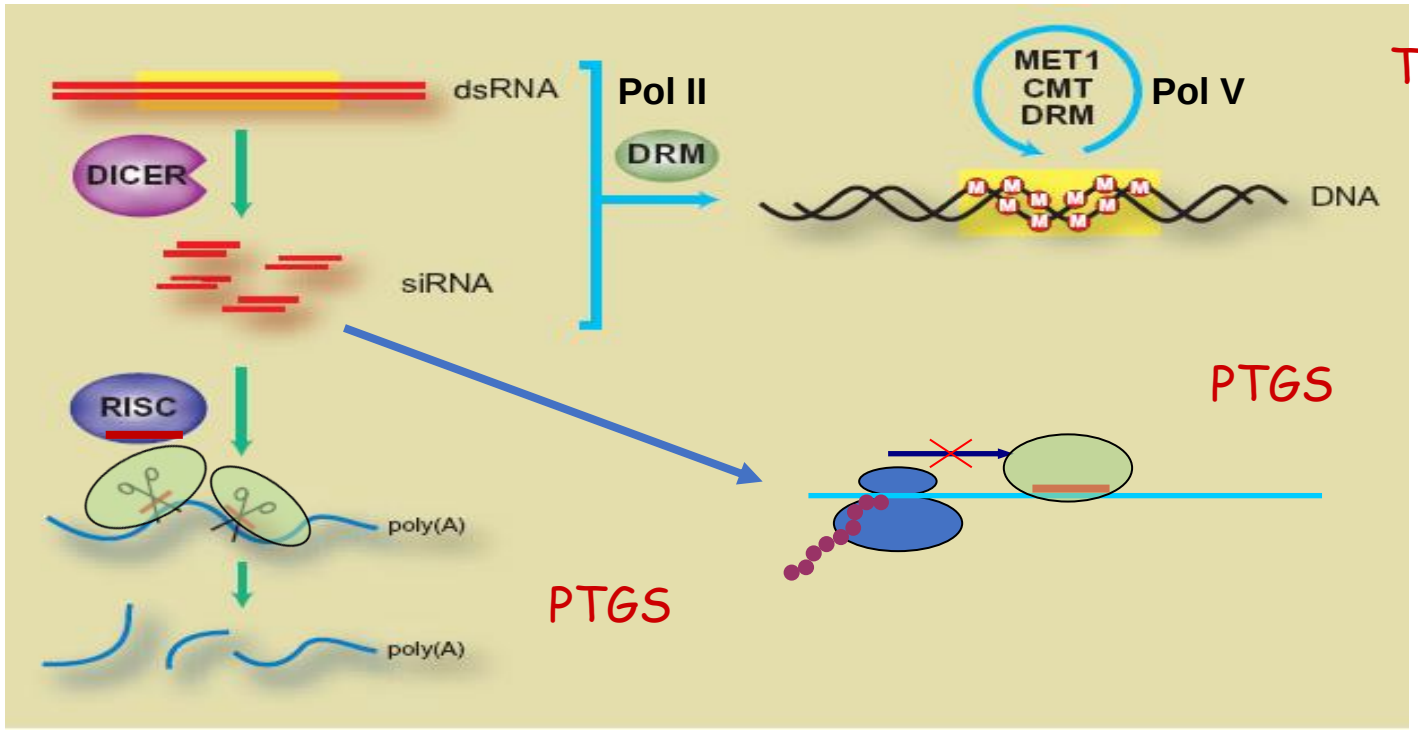
Photo: J. Mottern

Craig C. Mello

Cenorhabditis elegans

Mechanism of action of small RNA-

depends on the length of sRNA, biogenesis (precursor), ...



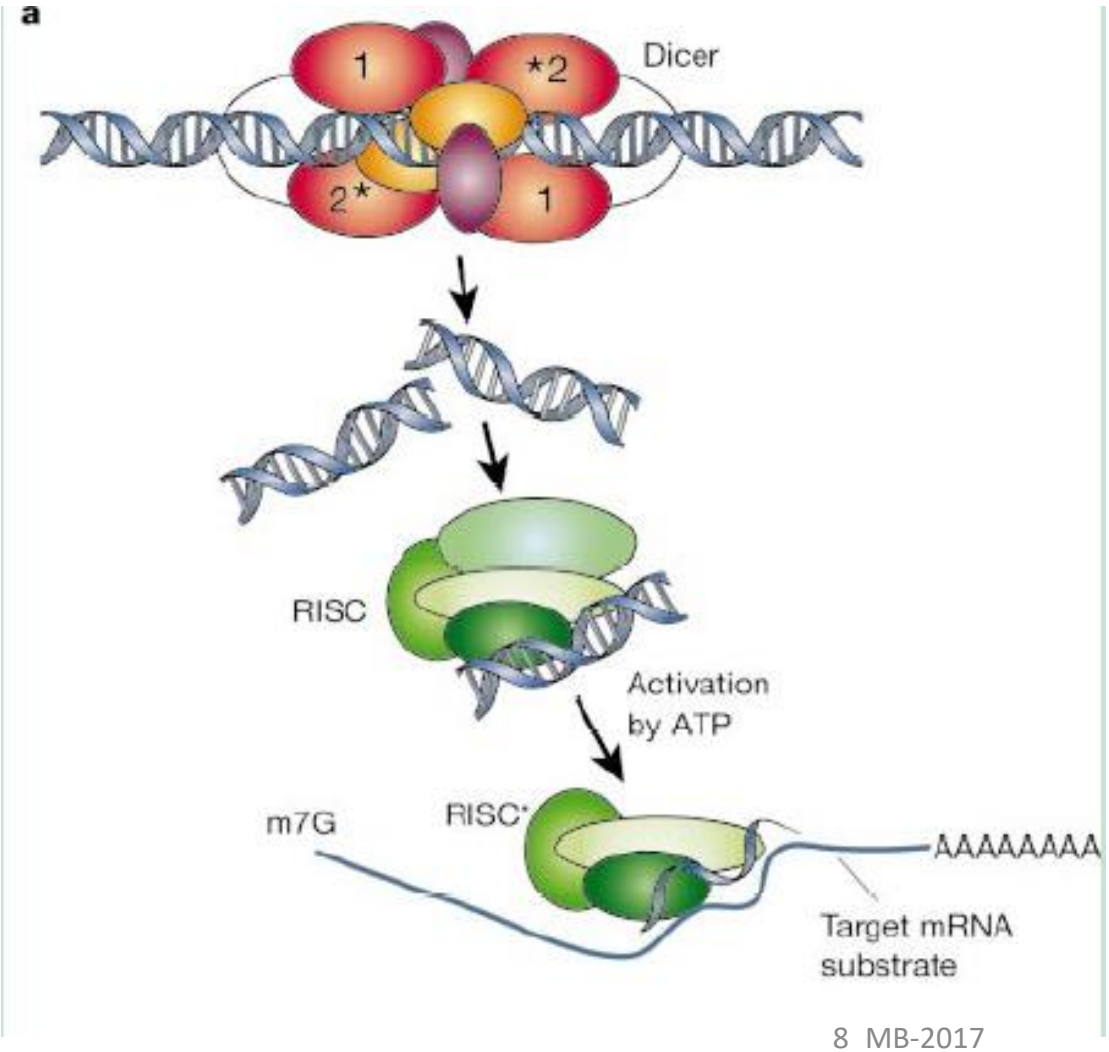
PTGS (posttranscriptional gene silencing):

- specific transcription degradation or translation blocking

TGS (transcriptional gene silencing):

- methylation of cytosines in the promoter (RdDM), heterochromatinization, inhibition of transcription factor binding

Basic mechanism of RNAi



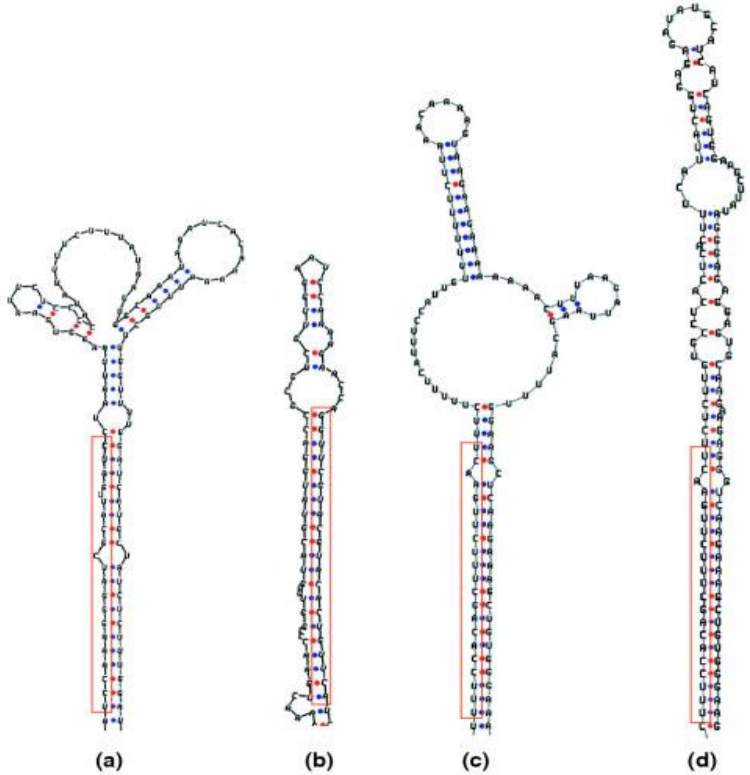
dsRNA in cell is cleaved by RNase DICER into short dsRNA fragments – sRNA

Argonaute with a single strand (from sRNA) mediates recognision of complementary sequences, which should be silenced (TGS, PTGS)

Small RNA in plants/animals

- 3’ end of sRNA methylated (HEN1) - protection
- **miRNA (micro**) – from transcipts of RNA Pol II (pre-miRNA)
  – hunderds MIR genes (*in trans*)

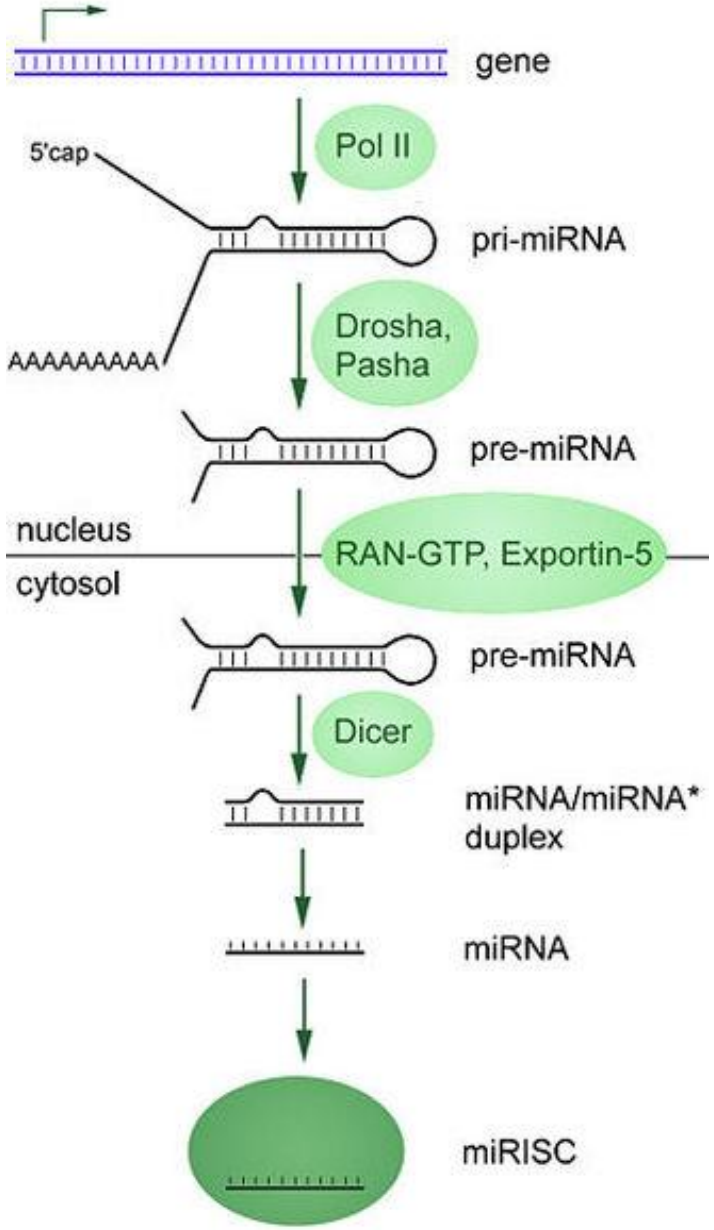
Pol II
DROSHA (Rnaze III),
PASHA (RNA binding protein),
DICER



- **siRNA (small interfering)** – from dsRNA of various origin (both internal and external – thousands types (both *in cis* and *in trans*)

….. (+ piRNA in animals)

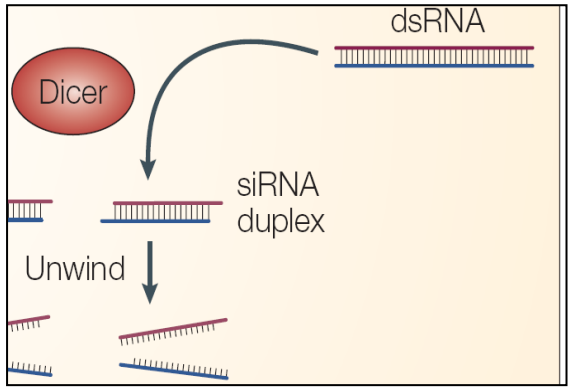
miRNA biogenesis

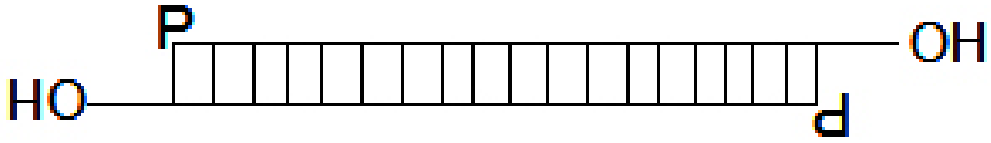


siRNA

• **siRNA (small interfering)** from dsRNA of various origin (both internal and external – thousands types (both *in cis* and *in trans*)

….. (+ piRNA in animals)

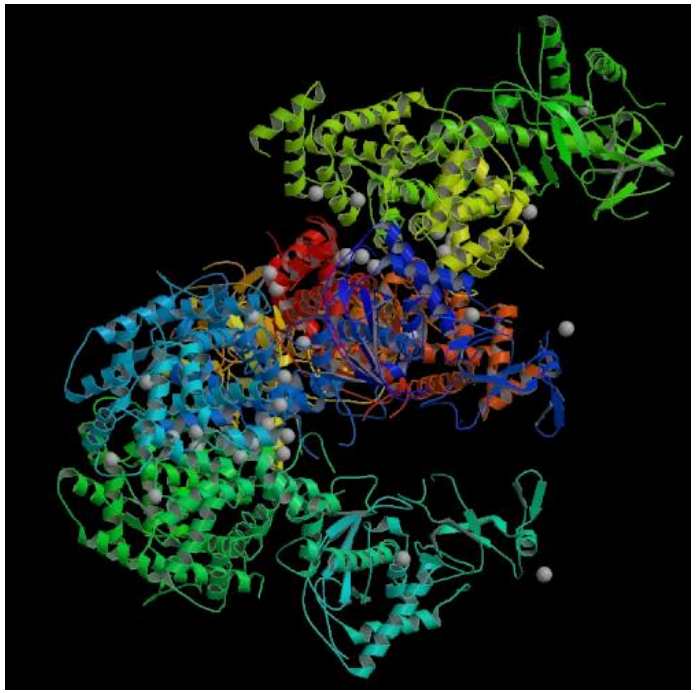




Schematic representation of a siRNA molecule: a ~19-21basepair RNA core duplex that is followed by a 2 nucleotide 3' overhang on each strand. OH: 3' hydroxyl; P: 5' phosphate.

DICER

Dicer, also known as **endoribonuclease Dicer** or **helicase with RNase motif**, is an enzyme that in humans is encoded by the *DICER1* gene. Being part of the RNase III family, Dicer cleaves double-stranded RNA (dsRNA) and pre-microRNA (pre-miRNA) into short double-stranded RNA fragments called small interfering RNA and microRNA, respectively. These fragments are approximately 20-25 base pairs long with a two-base overhang on the 3' end. Dicer facilitates the activation of the RNA-induced silencing complex (RISC), which is essential for RNA interference. RISC has a catalytic component argonaute, which is an endonuclease capable of degrading messenger RNA (mRNA).



Argonaute

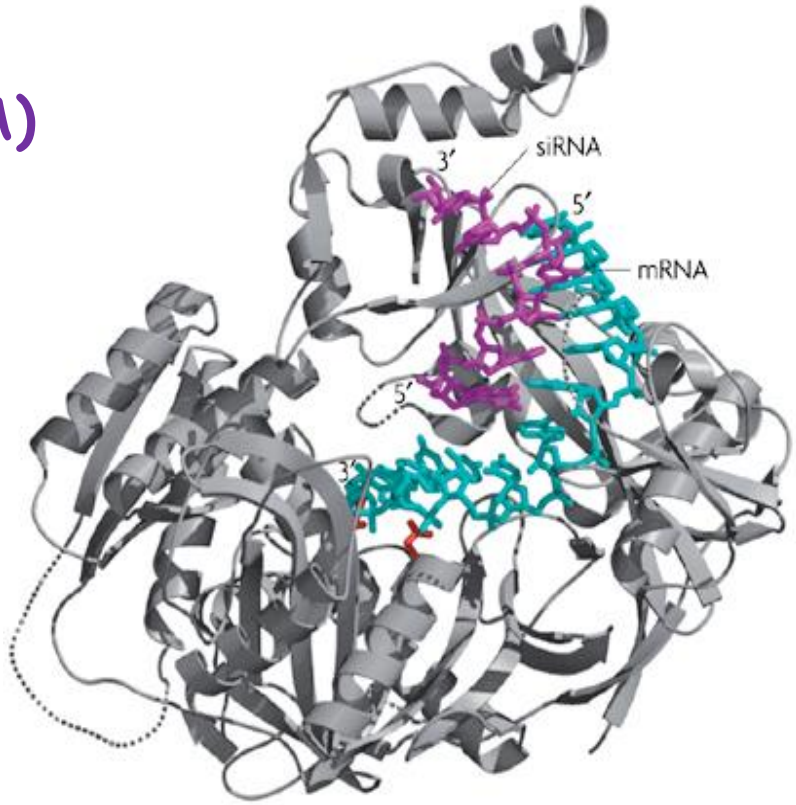
RNA binding protein (20-26 nt RNA)
- strand selection (5' nt, participation of HSP90)

- 10 genes in *Arabidopsis*

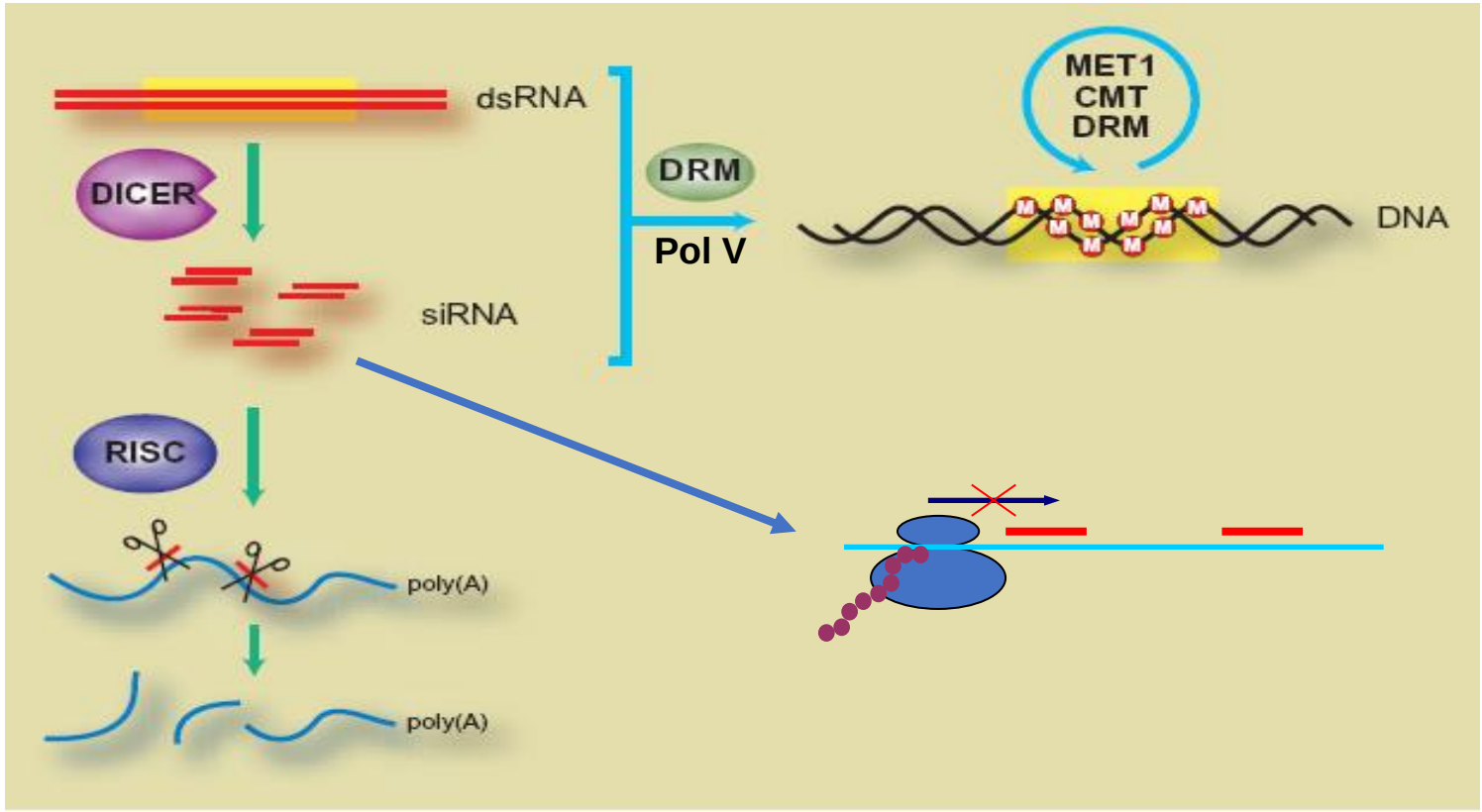
- main component of RISC (RNA induced silencing complex)

- block of translation or slicer (RNAse H-like endonuclease - PIWI doména)

- role in TGS (RdDM) (RNA directed DNA methylation)



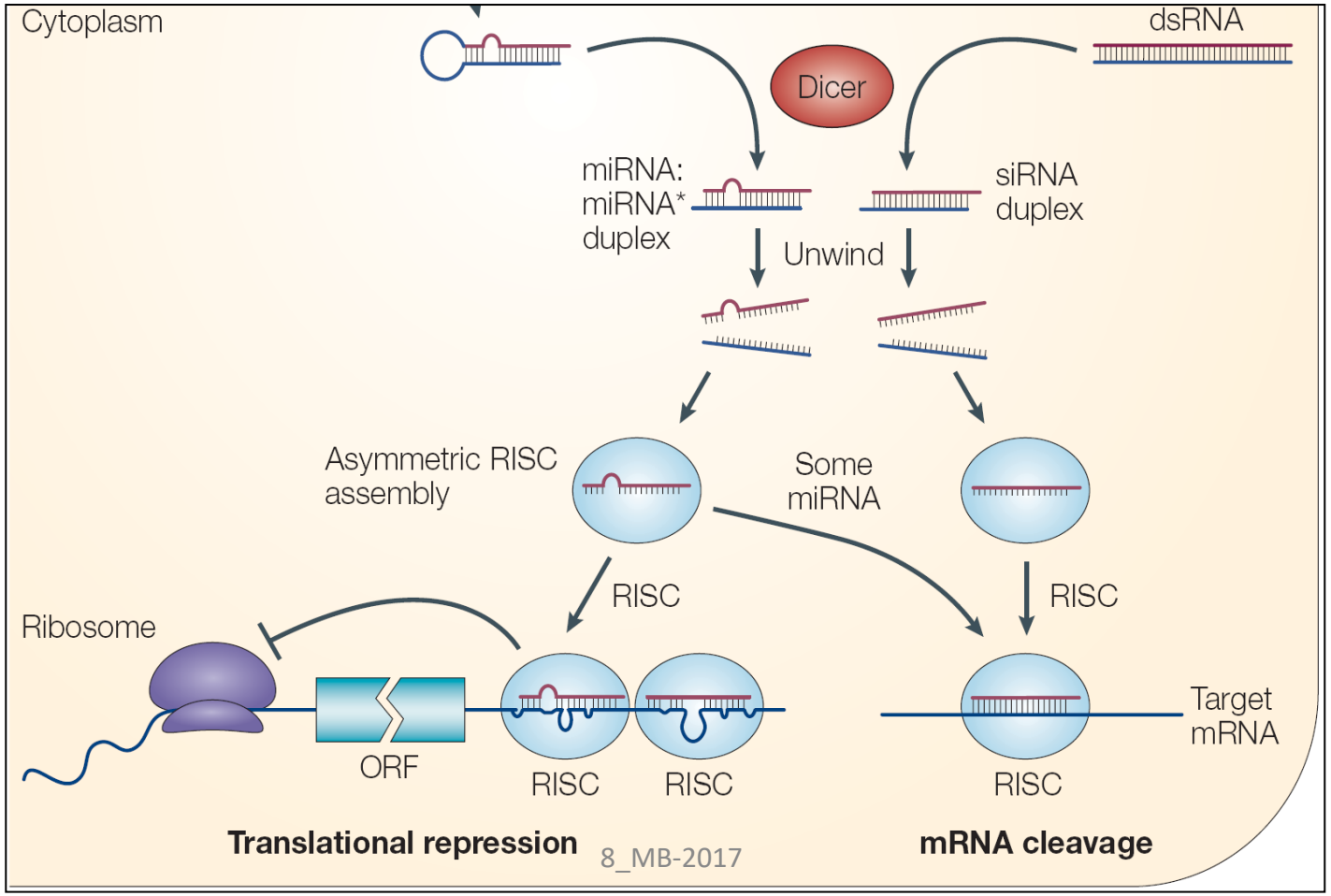
Mechanism of small RNA action - overview



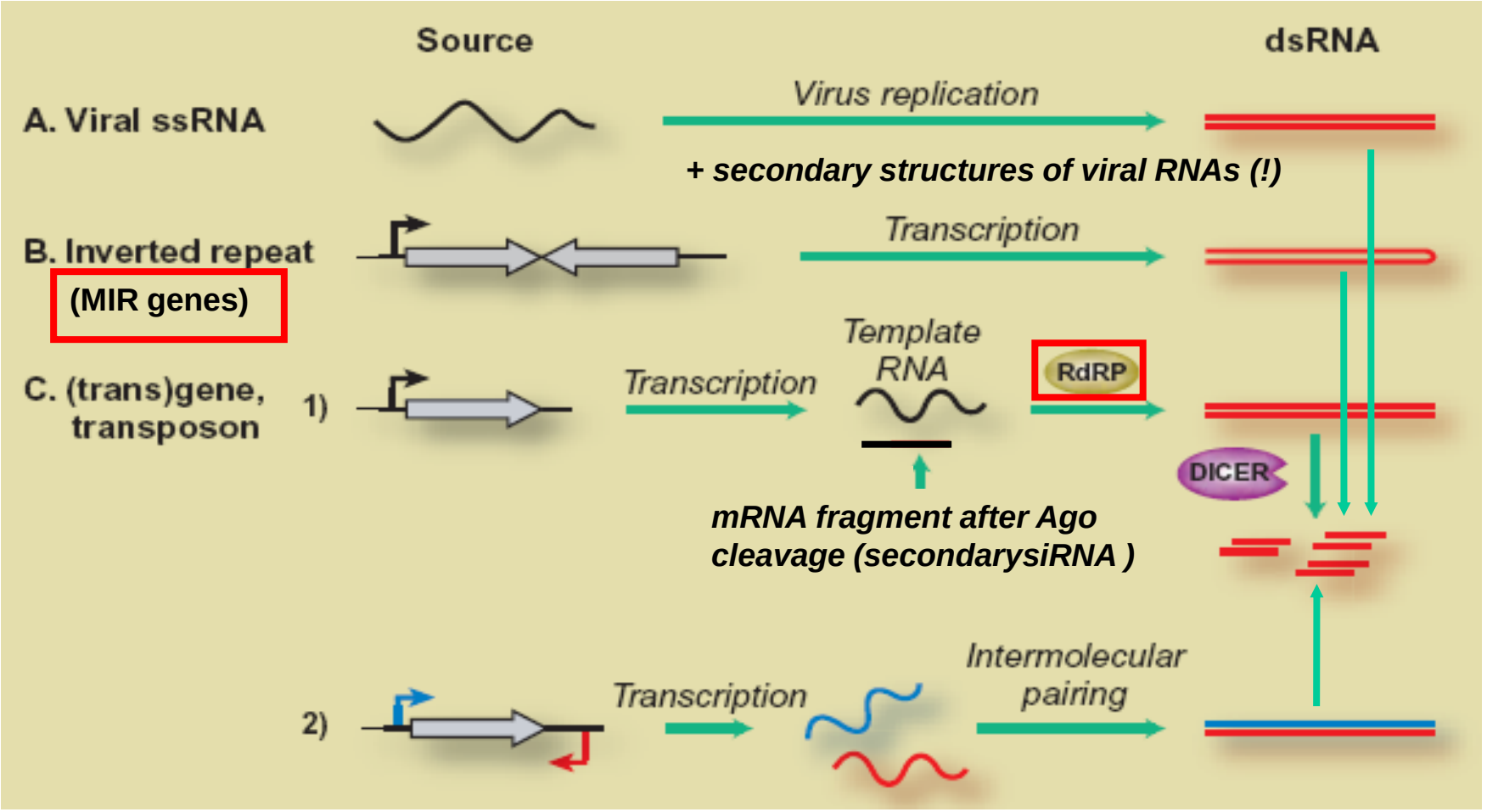
PTGS (21-22 nt): - specific cleavage of transcript
- - block of translation

TGS (24 nt): - methylation of promoter, heterochromatin formation
- preventing interaction of transcription factors

sRNA mode of action also depends on complementarity



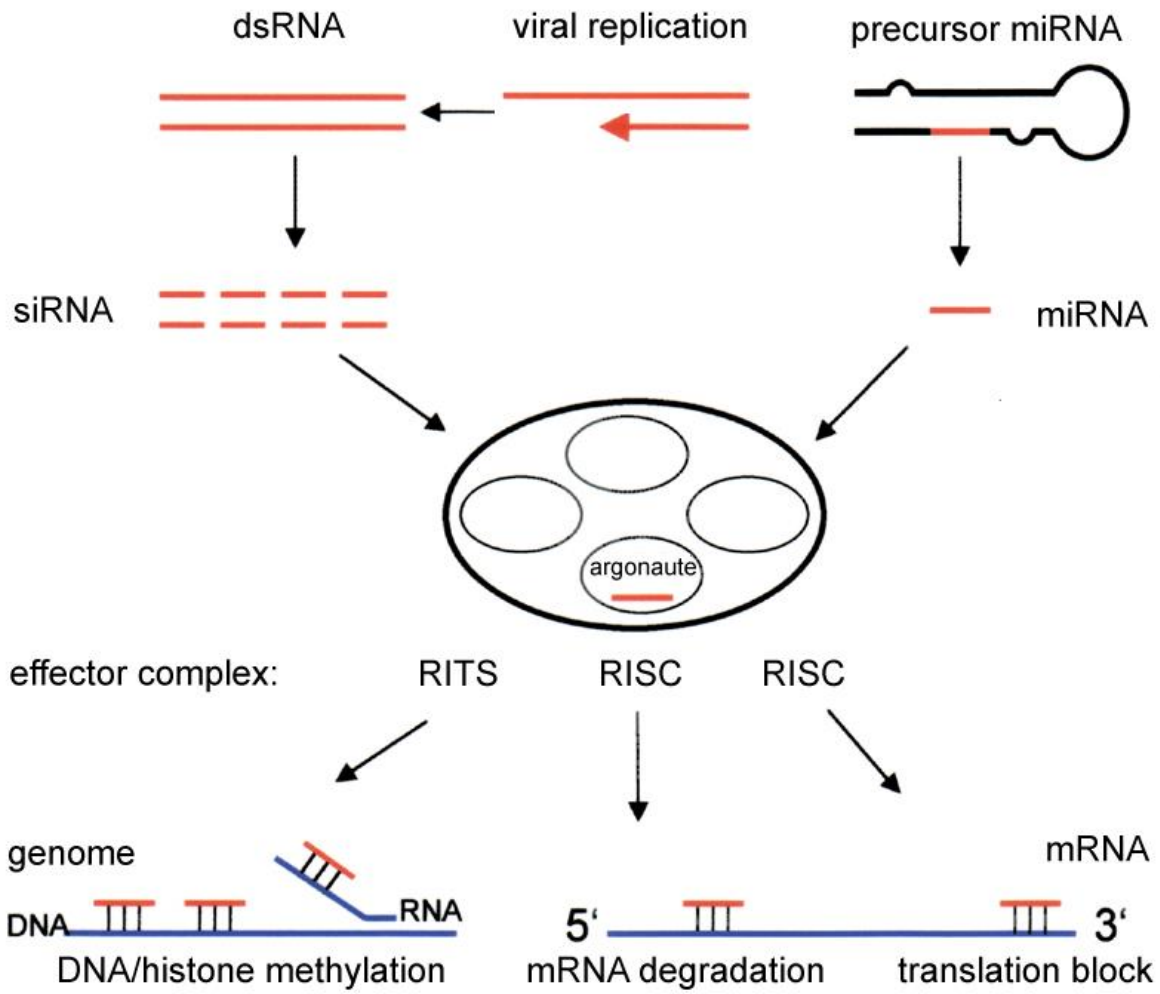
dsRNA formation



- RdRP = RNA-dependent RNA Polymerase – synthesis of compl. RNA strand
  templates: - transcripts cleaved by RISC
  - impaired mRNAs (without polyA or cap)
  - transcripts of RNA polymerase IV

Overview of RNA interference

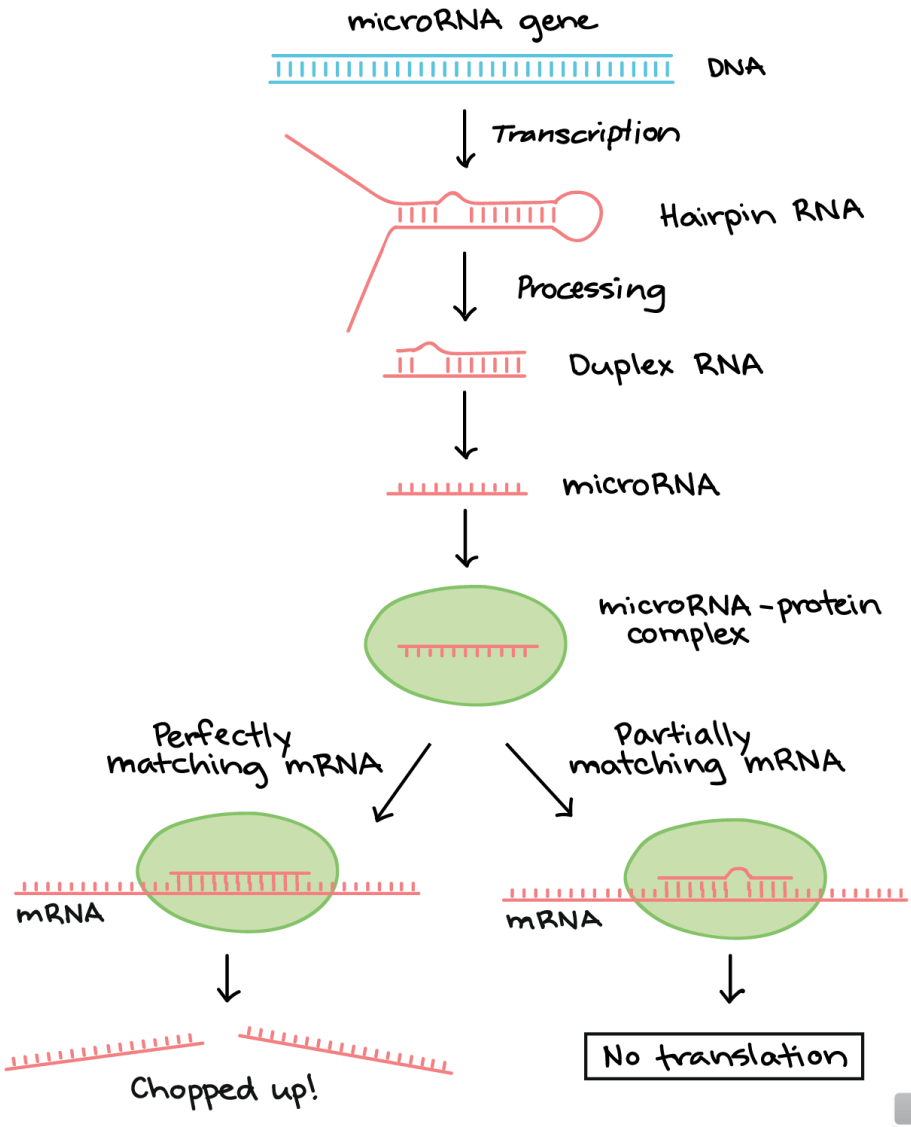
Overview of RNA interference. The dicer enzymes produce siRNA from double-stranded RNA and mature miRNA from precursor miRNA. miRNA or siRNA is bound to an **argonaute enzyme** and an effector complex is formed, either a **RISC (RNA-induced silencing complex)** or **RITS (RNA-induced transcriptional silencing**) complex. RITS affects the rate of transcription by histone and DNA methylation, whereas RISC degrades mRNA to prevent it from being translated.



Matzke MA, Matzke AJM – This figure is adapted from one by Matzke MA, Matzke AJM (2004) Planting the Seeds of a New Paradigm. PLoS Biol 2(5): e133 doi:10.1371/journal.pbio.0020133.

microRNA

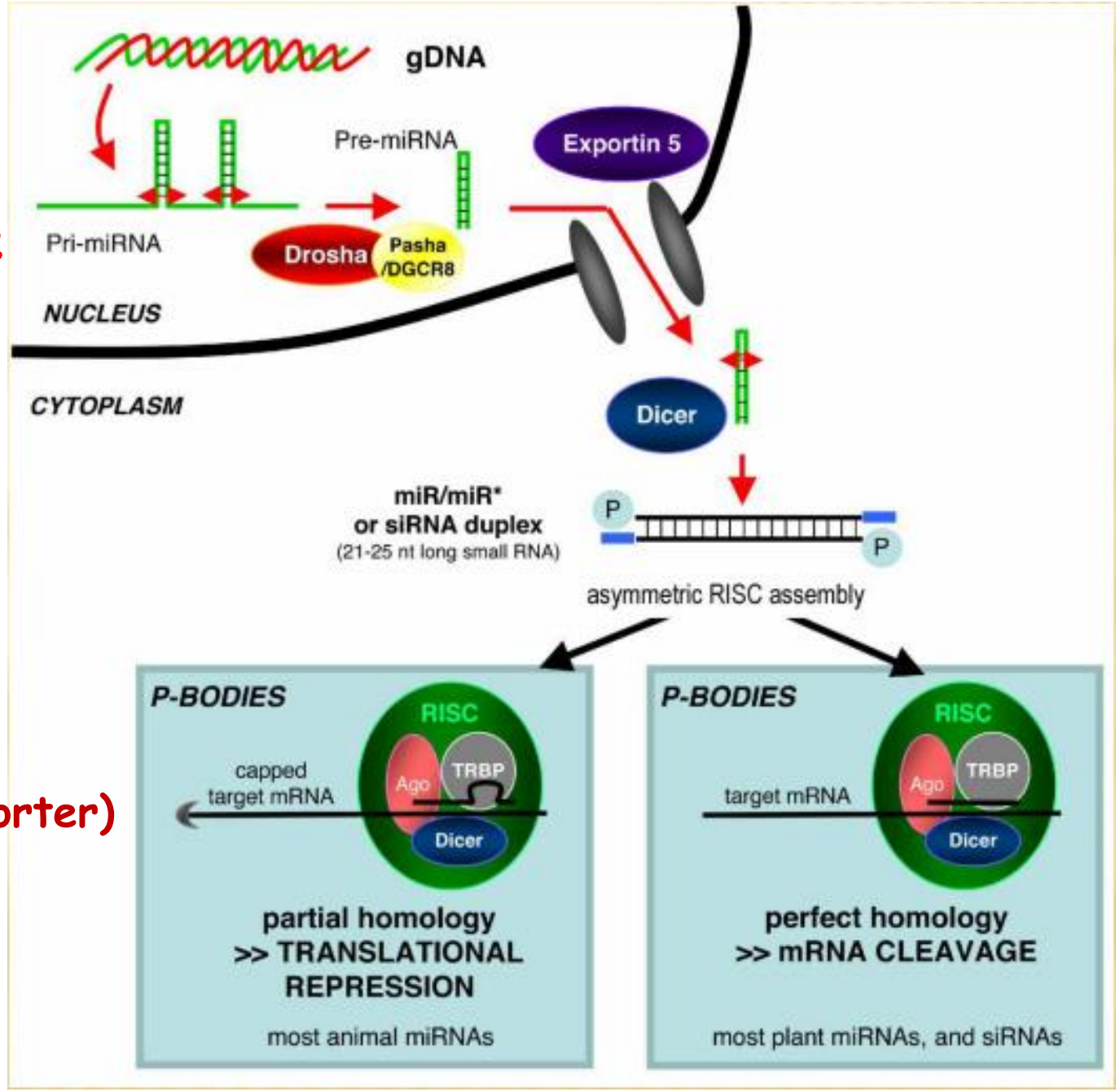
small non-coding RNA of 18-25 nucleotides in size
negative regulatory expression genes that degrade target mRNA or block its translation
microRNAs arise from primary pri-miRNA transcripts that are relatively large (even several kb)
pri-miRNAs are treated in the nucleus with Drosha RNAase and protein
Pasha binding dsRNA to pre-miRNA about 70 nucleotides long with imperfect hair structure
pre-miRNAs are exported to the cytoplasm by Exportin 5 and digested with Dicer nuclease to final 22 kb miRNA duplexes
The miRNA binds to the RISC, one fiber degrades and the other mediates the degradation or translation inhibition of the respective mRNA



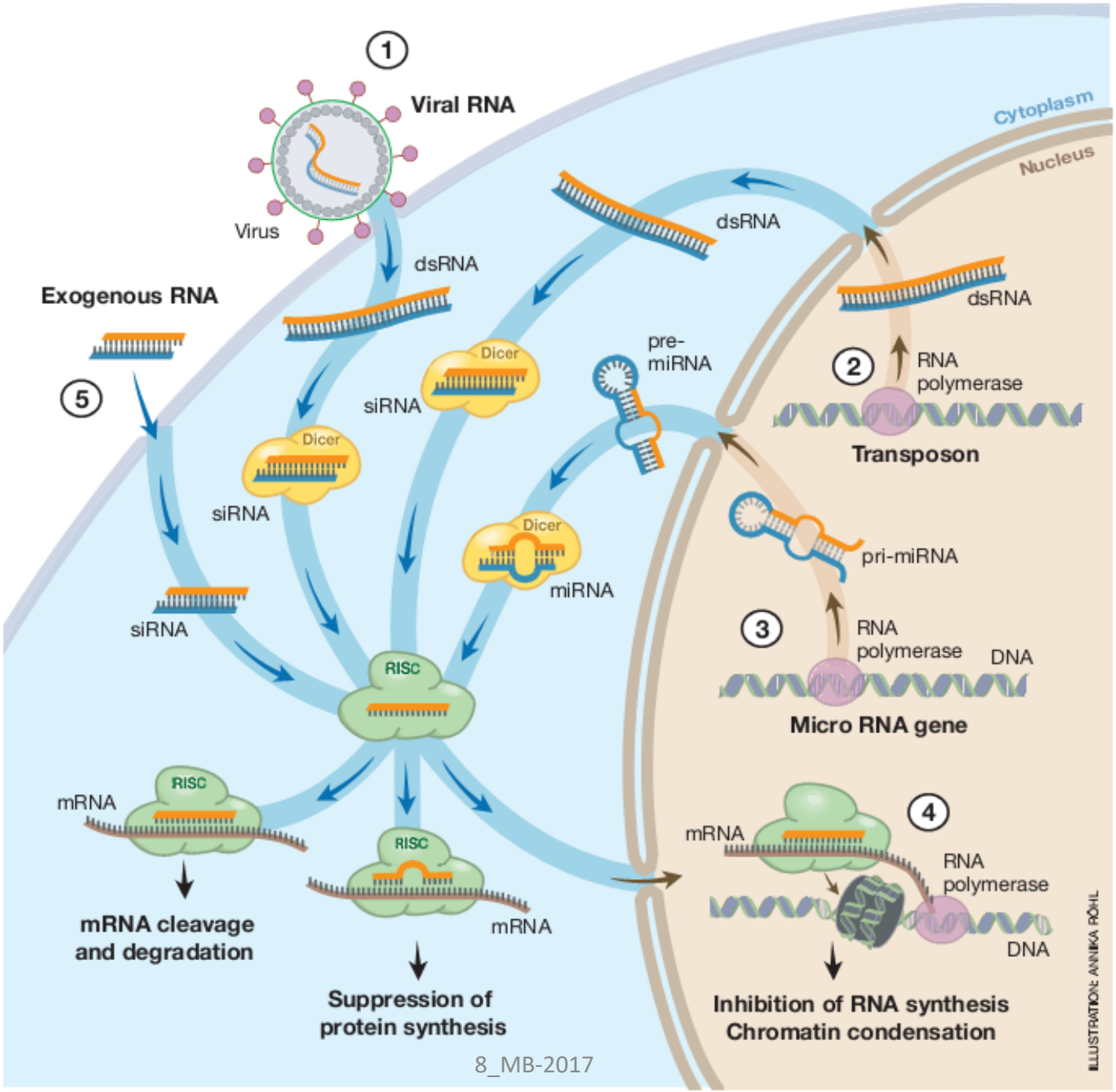
RNA interference

based on enzyme degradation or translation inhibition of specific mRNA

Drosha (RnasaIII)
Pasha (protein)
Exportin 5 (transporter)
Dicer (RNasaIII)
RISC (multiprotein complex)



TEST

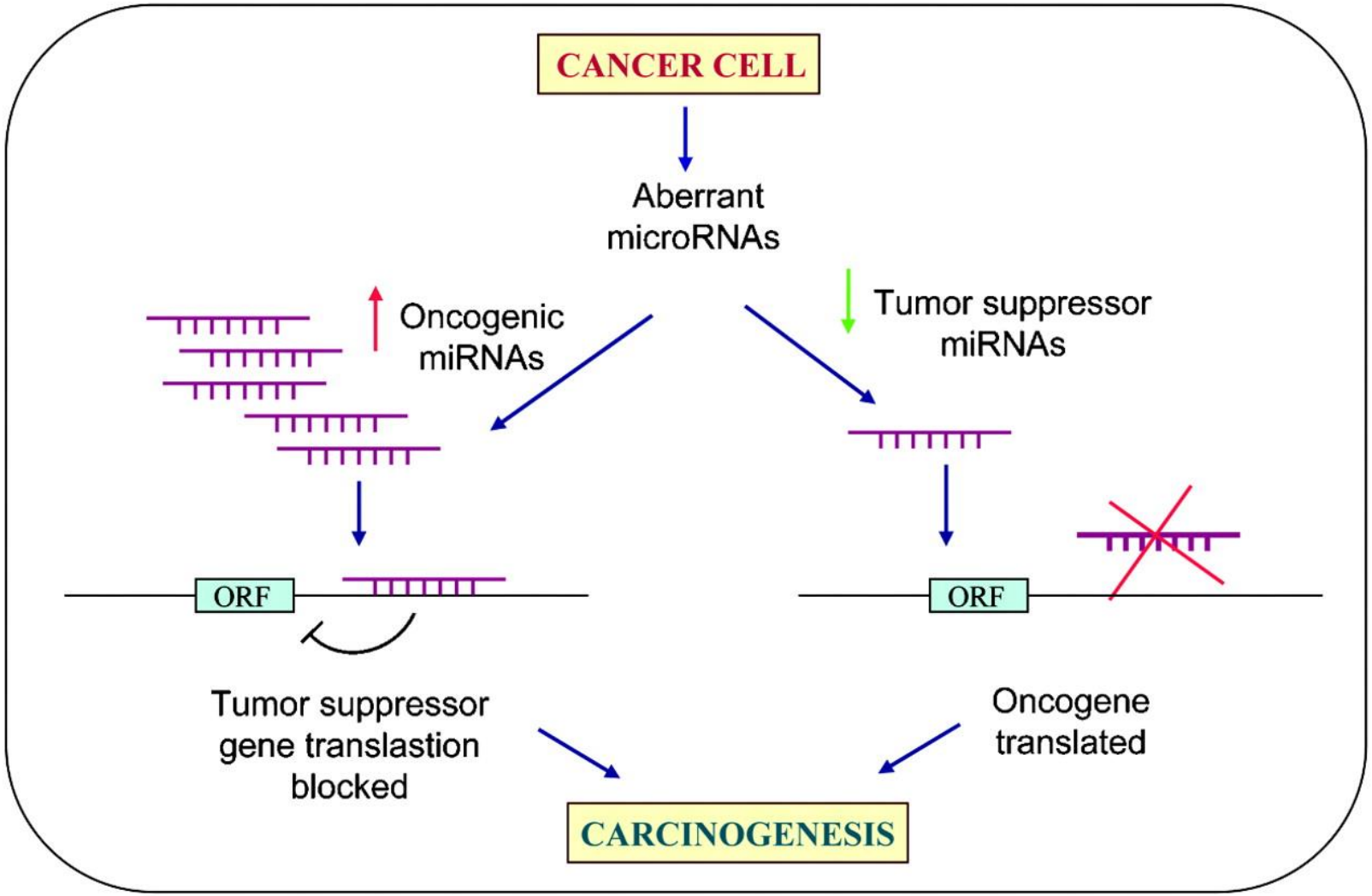

1
Viral RNA
Virus
Cytoplasm
Nucleus
dsRNA
Exogenous RNA
5
dsRNA
dsRNA
2
RNA polymerase
Transposon
Dicer
siRNA
pre-miRNA
Dicer
siRNA
pri-miRNA
3
RNA polymerase
DNA
Micro RNA gene
Dicer
miRNA
siRNA
RISC
RISC
mRNA
RISC
mRNA
mRNA cleavage and degradation
Suppression of protein synthesis
4
mRNA
RNA polymerase
DNA
Inhibition of RNA synthesis
Chromatin condensation
ILLUSTRATION: ANNIKA RÖHL

siRNA and miRNA utilisation:

1) gene analysis
2) gene therapies
3) anti-viral vaccines
4) transgenic organisms that have transiently inhibited selected genes

- iRNA usage does not fall under GMO
- Yet usage of cassettes producing iRNA does!

MicroRNAs as tumor suppressors or oncogenes



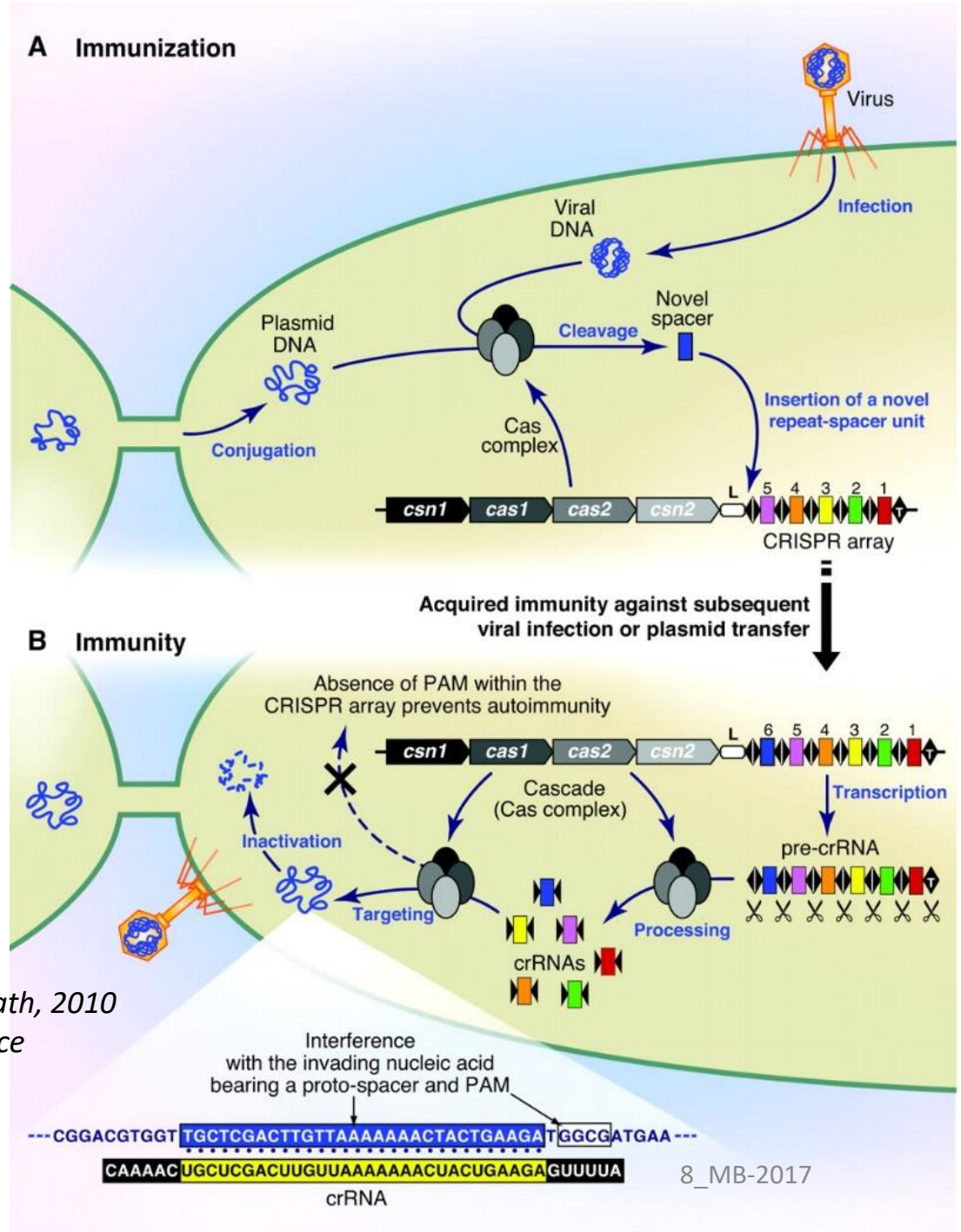
CRISPR system

In 2008, it was described RNAi analogous system designed to the degradation of viral NA

- **It uses internal "virus" sequences inserted in the inverted repeats (CRISPR)**
- **CRISPR = clusters of regularly interspaced short palindromic repeats**
- After transcription of this sequence leads to their progressive cleavage by **Cas proteins**
- **The resulting products interfere with the nucleic acid of the entering virus**

- Each of repeats followed by short segments called **Spacer DNA**, obtained during previous meetings with relevant bacterial viruses or plasmids.

Brouns et al. (2008): Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes, Science 321, 960-964

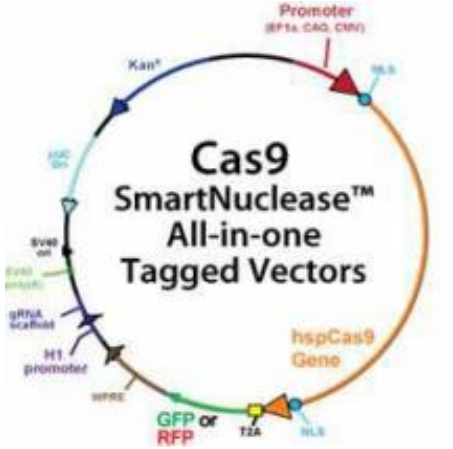


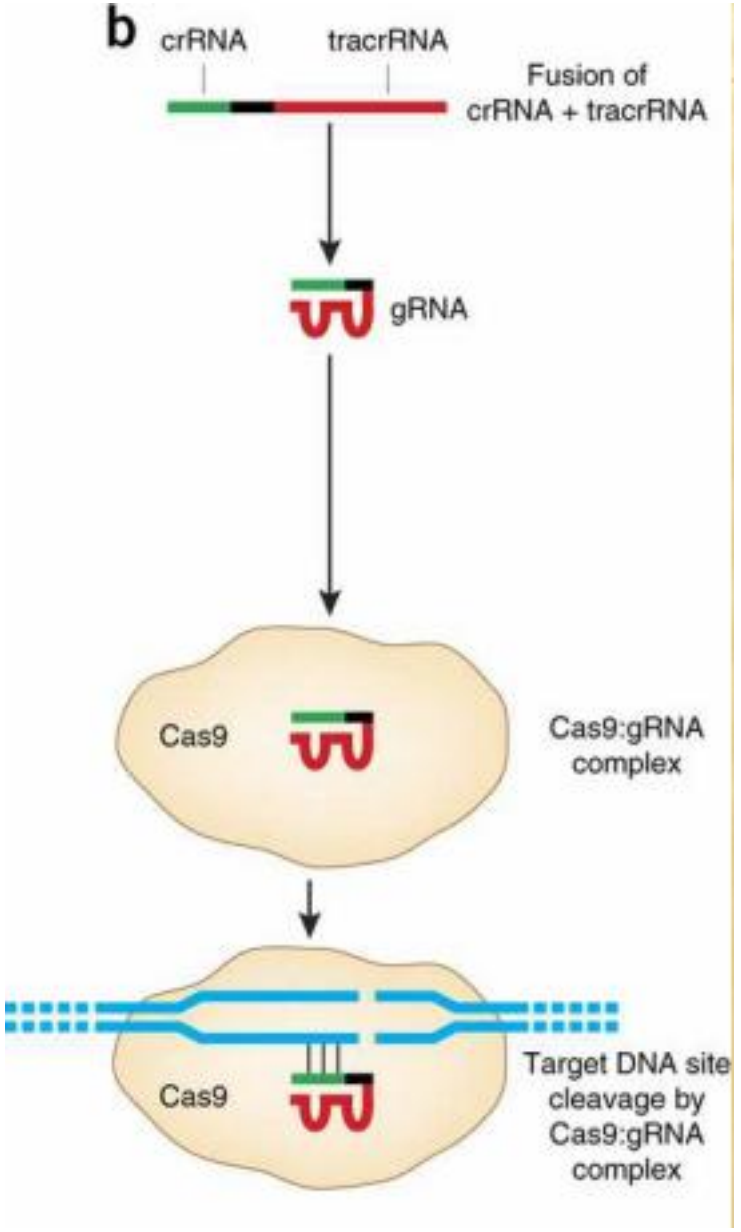
Horvath, 2010
Science

Overview of the CRISPR/Cas mechanism of action. (**A**) Immunization process: After insertion of exogenous DNA from viruses or plasmids, a Cas complex recognizes foreign DNA and integrates a novel repeat-spacer unit at the leader end of the CRISPR locus. (**B**) Immunity process: The CRISPR repeat-spacer array is transcribed into a pre-crRNA that is processed into mature crRNAs, which are subsequently used as a guide by a Cas complex to interfere with the corresponding invading nucleic acid. Repeats are represented as diamonds, spacers as rectangles, and the CRISPR leader is labeled L.

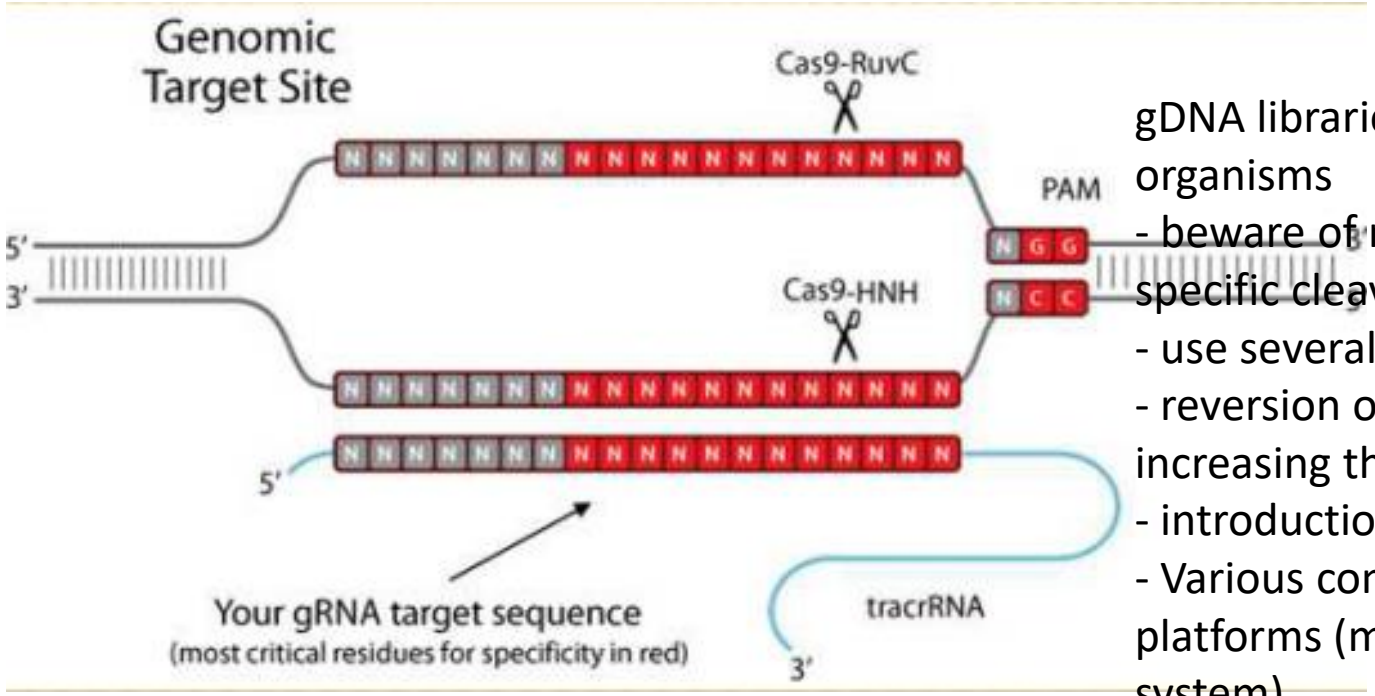
CRISPR/Cas9

- the whole system is modified for targeted mutagenesis
- – vector - gRNA = crRNA tracr + RNA
-
- part gRNA and 20nt complementary section to the target site in the genomic DNA
- – + Coexpression of Cas9 nuclease (even the same vector)

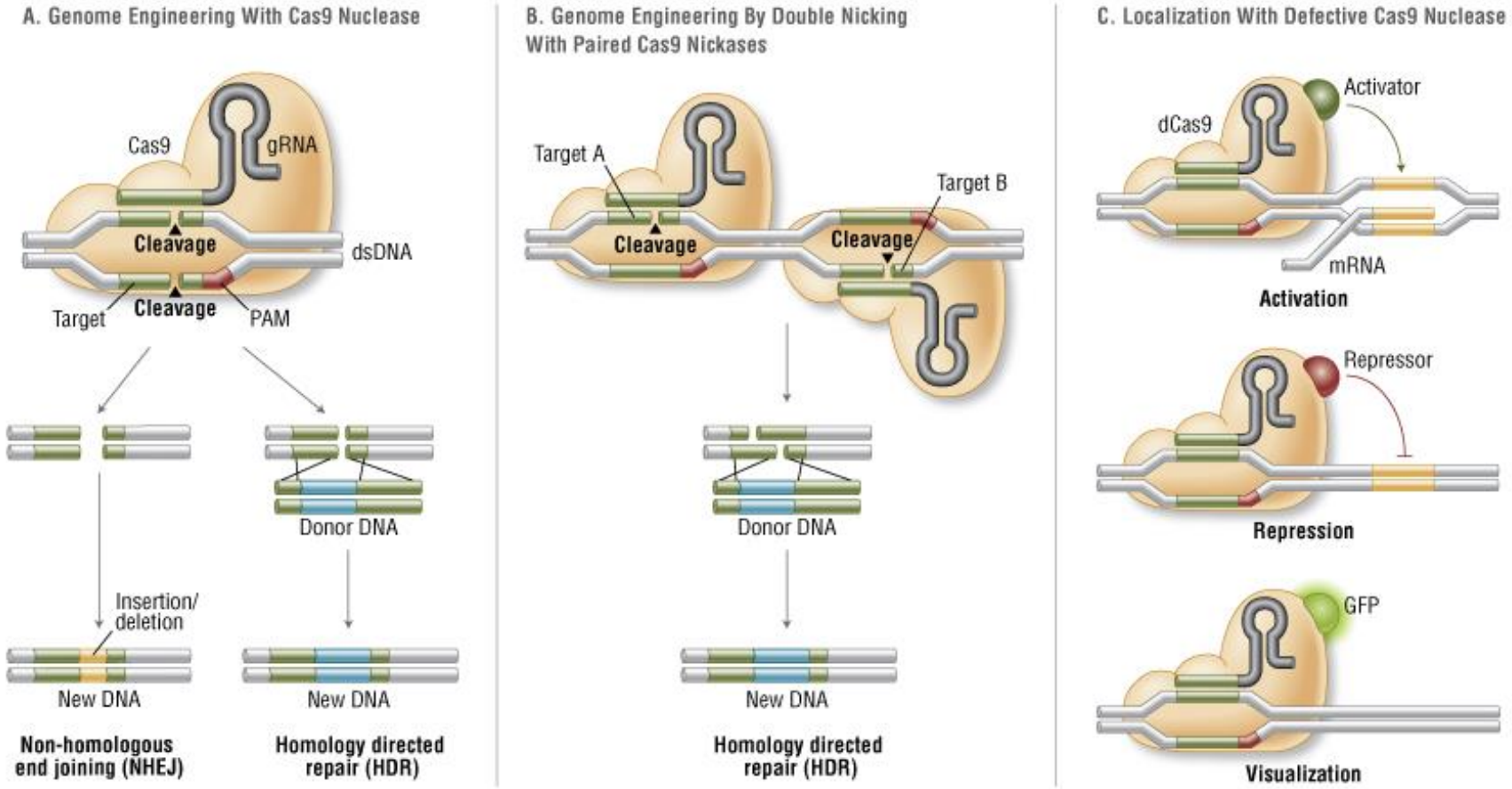




- PAM – protospacer adjacent motif
- sequence in the vicinity of gDNA
- required for efficient cleavage by Cas9 nuclease
- the original system "NGG" (but the development of systems with other sequences)
- according to the system target sequence must be in the N 20 -GG

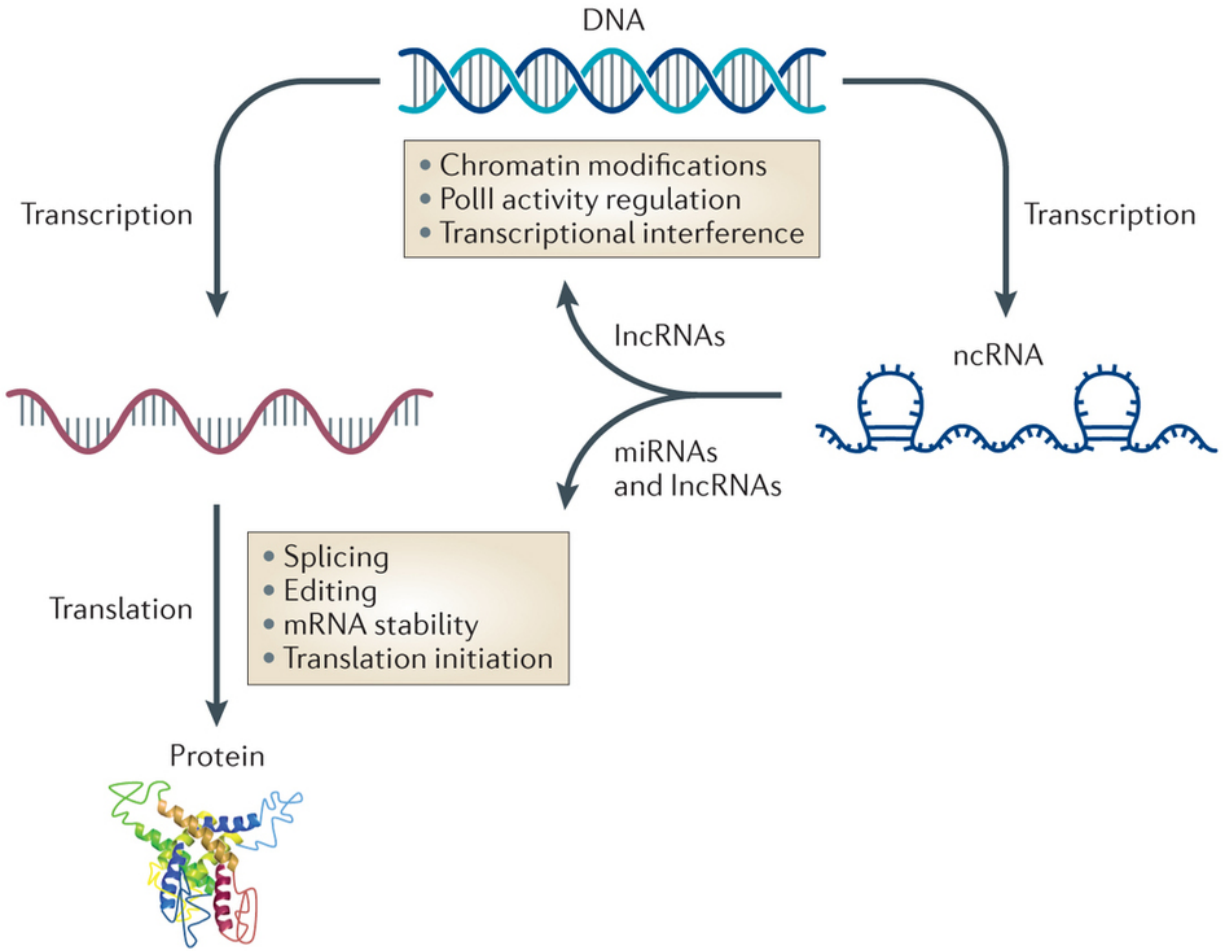


gDNA libraries available for various organisms
- beware of non-specific binding - non-specific cleavage
- use several different gDNA
- reversion of the phenotype by increasing the expression
- introduction of mutation
- Various companies - different platforms (modifications of the original system)
- available software can be used to design



A. Wild-type Cas9 nuclease site specifically cleaves double-stranded DNA activating double-strand break repair machinery. In the absence of a homologous repair template non-homologous end joining can result in indels disrupting the target sequence. Alternatively, precise mutations and knock-ins can be made by providing a homologous repair template and exploiting the homology directed repair pathway.
B. Mutated Cas9 makes a site specific single-strand nick. Two sgRNA can be used to introduce a staggered double-stranded break which can then undergo homology directed repair.
C. Nuclease-deficient Cas9 can be fused with various effector domains allowing specific localization. For example, transcriptional activators, repressors, and fluorescent proteins.

lncRNA - long non-coding RNA

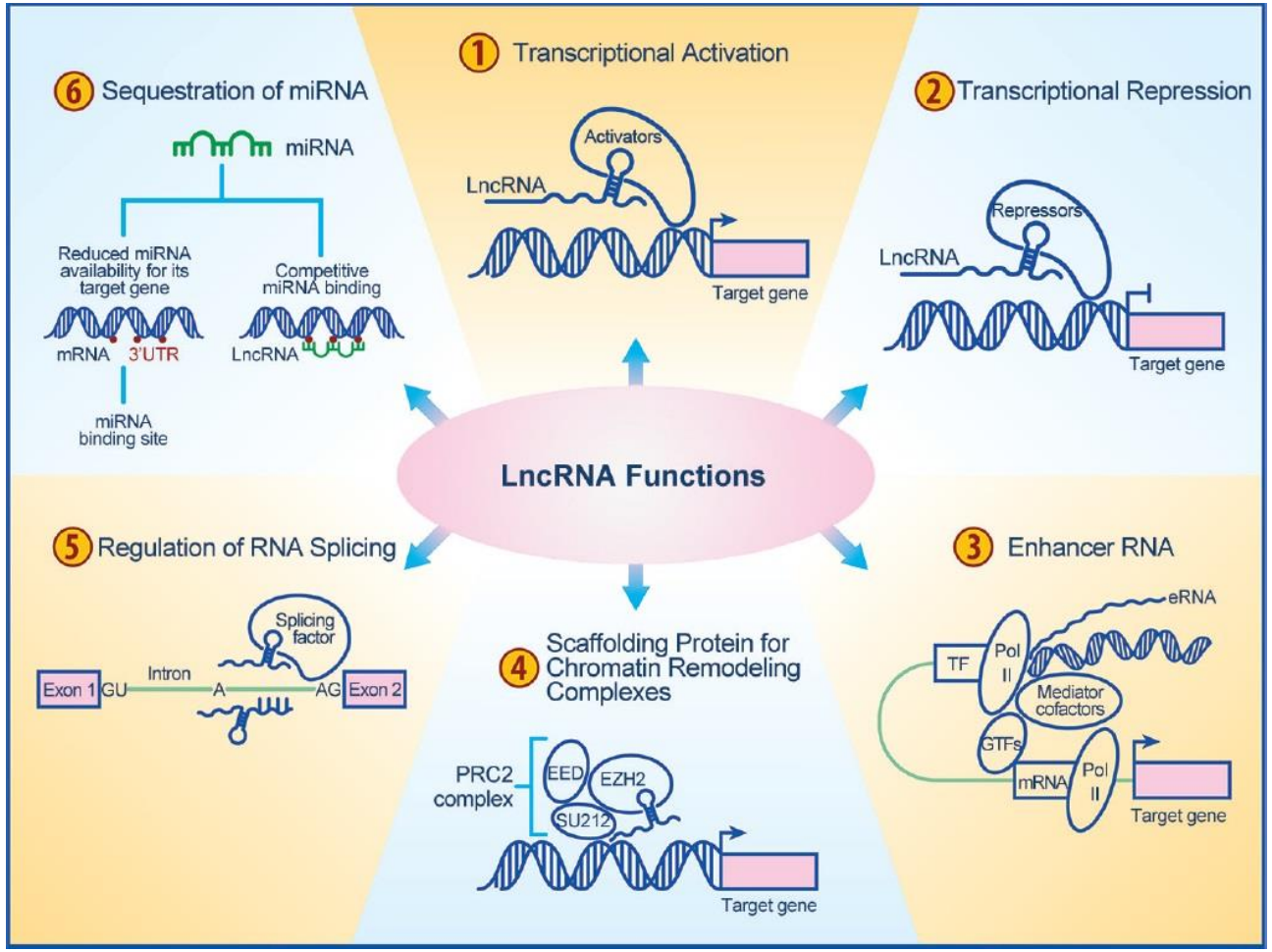


Nature Reviews | Drug Discovery

Long non-coding RNAs (long ncRNAs, lncRNA) are non-protein coding transcripts longer than 200 nucleotides.[1] This somewhat arbitrary limit distinguishes long ncRNAs from small regulatory RNAs such as microRNAs (miRNAs), short interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), and other short RNAs.[2]

2009

Long ncRNAs in the regulation of gene transcription



Long ncRNAs in gene-specific transcription

In eukaryotes, RNA transcription is a tightly regulated process. NcRNAs can target different aspects of this process, targeting transcriptional activators or repressors, different components of the transcription reaction including RNA polymerase (RNAP) II and even the DNA duplex to regulate gene transcription and expression (Goodrich 2006). In combination these ncRNAs may comprise a regulatory network that, including transcription factors, finely control gene expression in complex eukaryotes.