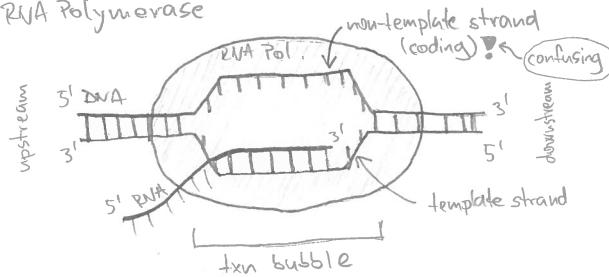


MITx 7.28.2x MOLECULAR BIOLOGY: TRANSCRIPTION AND TRANSPOSITION

- transcription determines cell type (DNA contains all information, but less than 1/5 of it is being transcribed by a cell at any given time)
- key questions for studies of TXN
  1. how does RNA Pol identify genes (specificity)
  2. how frequently is a gene transcribed (efficiency)

## - RNA Polymerase



- TXN requires - dsDNA template

- template strand directs rNTPs addition
- non-template strand is not involved in TXN
- 4 rNTPs (rATP, rUTP, rCTP, rGTP)
- works exactly the same as in DNA replication (except rNTPs have hydroxyl on 2' carbon)

	RNA Pol	DNA Pol
only add to 3' of polymer	✓	✓
require DNA template to direct base addition	✓	✓
use single active site to add any of the four rNTPs/dNTPs	✓	✓
NTPs used	rNTPs	dNTPs
requires a primer	X	✓
fidelity	$\frac{\sim 1 \text{ mistake}}{10^4 \text{ bases}}$	fewer mistakes
rate	60-100 nt/sec	1000 nt/sec
processivity	$\frac{> 4000 \text{ nt}}{\text{binding}}$	lower
copies DNA strands	unidirectional, one strand	both strands

DNA: permanent molecule } cell can tolerate mistakes  
RNA: temporary molecule } in RNA much better than mistakes in DNA

can be MUCH higher: there are RNA Pols that synthesise 10s, 20s of kb per binding, some do even over 1Mb per binding

↑  
intrinsic to RNA Pol - no sliding clamp, etc.

- reverse transcriptase (RT) - can use both RNA and DNA as a template

↑  
synthesizes DNA using either RNA or DNA template (e.g. telomerase)

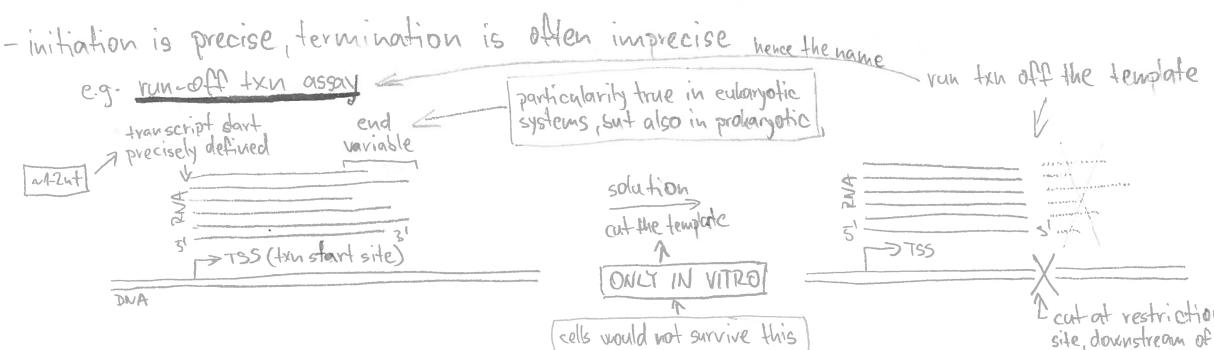
- requires primer (telomerase extends end of chromosome, which serves as a primer)
- product is DNA and therefore uses dNTPs, not rNTPs
- low fidelity

# ASSAYS FOR TXN

## INCORPORATION ASSAY

- labelled rNTP  $\rightarrow$  typically rUTP (unique to RNA)
- process: same as for DNA
- set up template
- add RNA Pol
- add rNTPs (at least one labelled)
- separate RNA product from labelled rNTPs
  - filter binding
  - gel electrophoresis
- RNA folds on itself (RNA is ss, while DNA can be ds and therefore not have this problem)
  - $\Rightarrow$  always use denaturing gel
  - agarose: over few hundred nt-sized product
  - Acrylamide: smaller product
- heat beforehand & add urea**

Q: Should not txn end at exact termination site?



- advantages [relatively fast quantitative]

most often the key regulatory sequences that are controlling txm are located immediately proximal to the TSS

- disadvantages [not good at defining TSS (critical piece of information)]  
[one gene at a time]

can clone single suspected TSS to a plasmid, but it is more time consuming than primer extension assay

if there is more than one TSS in substrate DNA, it would be hard & time consuming to separate individual products of different TSSs

## PRIMER EXTENSION ASSAY

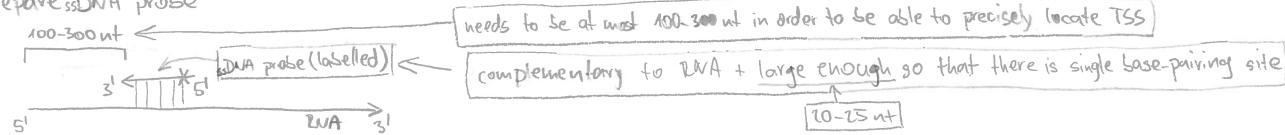
- primarily used for mapping 5' ends of transcripts

- both *in vivo* & *in vitro*

- process

- isolate RNA (*in vivo*: separate mRNA from ribosomal RNA and tRNA; *in vitro*: just take all RNA synthesized by the reaction)

- prepare ssDNA probe

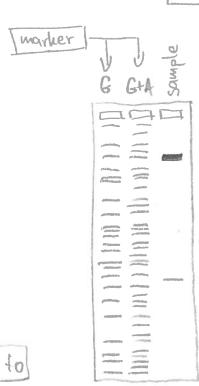
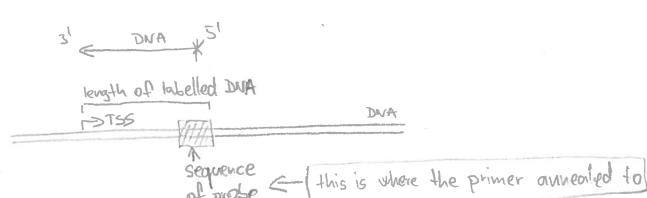


- anneal primer to RNA

- extend primer w/ reverse transcriptase



- purify & separate on denaturing acrylamide gel



GC-rich section of RNA has tendency to fold on itself and create a hairpin. Polymerase can not replicate through pause

nowadays, reverse transcriptases from thermophilic organisms are used, which can work at ~60-78°C, preventing the problem (such high temperature would melt RNA that has folded on itself)

- advantages - precise 5'-end information

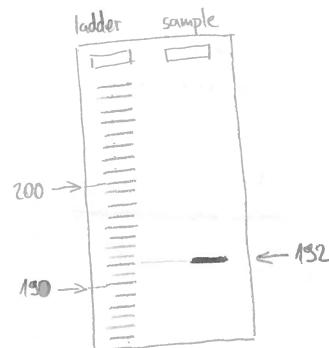
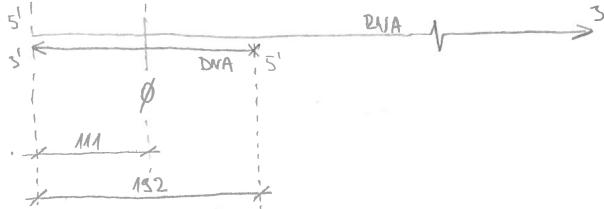
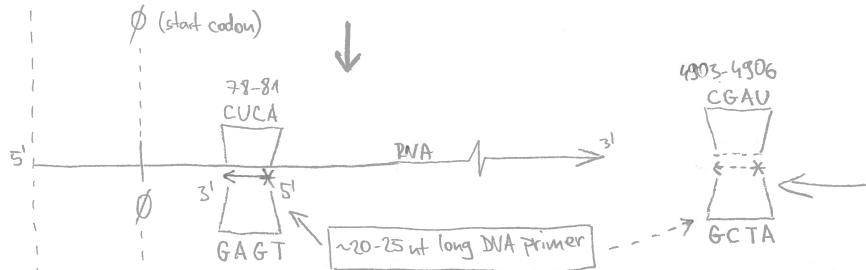
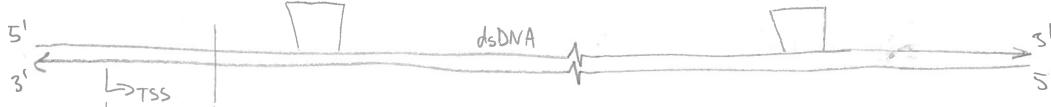
- disadvantages [one gene at a time]

[low to medium sensitivity (*in vivo*)]

it is still DENATURING gel, these pictures show products before they are separated from one another

Q

all these are ~20-25 nt long  
in reality



## REVERSE TRANSCRIPTASE PCR

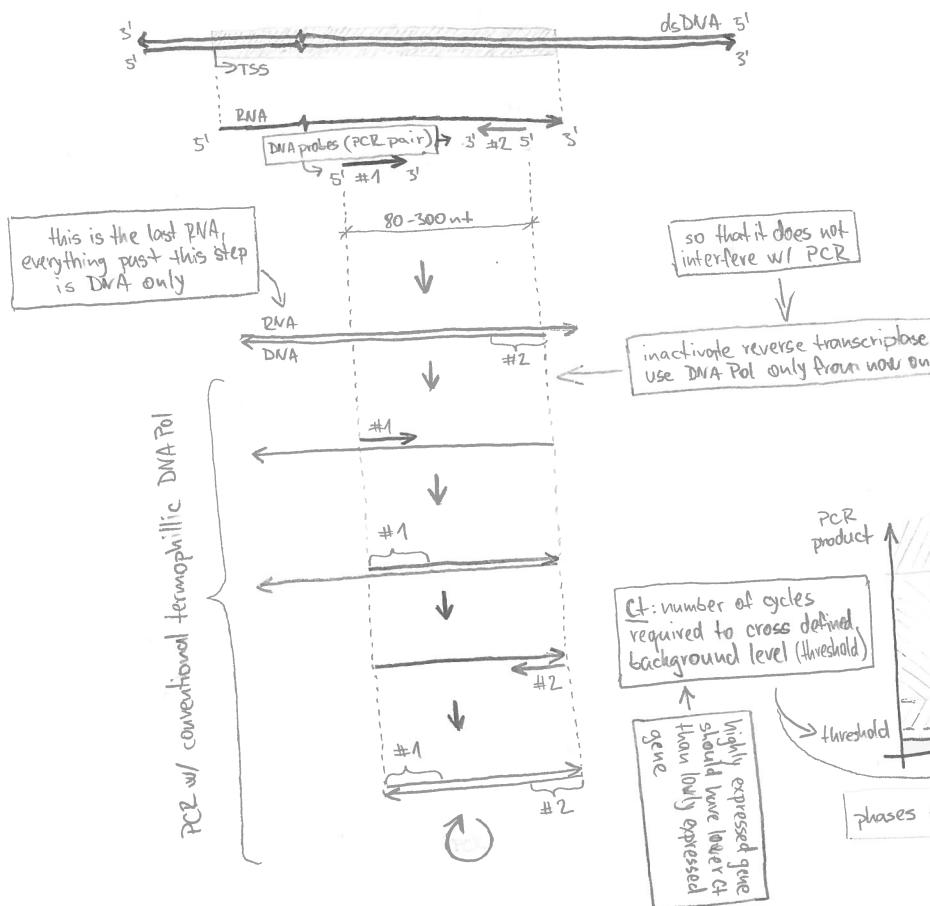
- most common way to look at gene expression
- process - isolate RNA (treat by DNase to digest DNA)
  - make 2 DNA probes (PCR pair)
  - ~10-25 nt long (to ensure unique base-pairing)
  - target region at the 3' end of the gene (5' end is made first)

- RNA / non-template aka. coding DNA strand
- #2
- hybridize (anneal) DNA probes w/ RNA (only one of them will base-pair w/ RNA)
  - extend annealed primer w/ reverse transcriptase (RT)
  - PCR w/ (conventional) thermophilic DNA Pol
  - separate product on NON-DENATURING gel
  - or monitor by real-time PCR (RT RT-PCR)

this is to ensure that transcription has been completed

there are cases, where RNA Pol starts, go about 30 nt into the transcript and then just sit there and wait

reverse transcriptase  
real-time



exponential phase: reagents are not limiting,  
near 100% efficiency

linear phase: products continue to accumulate,  
but reagents become limiting and efficiency  
begins to fall

plateau phase: accumulation of products  
ceases as the reaction is exhausted  
for number of different reasons

- WAL157  
1:23
- real-time PCR: observe a dye incorporated into dsDNA product to become progressively more intense
  - can be even used to "count" number of molecules present (based on a standard added to the sample)
  - usually used to do relative comparisons
  - ! - always include no-template control (when doing PCR), particularly when looking for low abundance RNAs
  - e.g. primers can get contaminated by just tiny amount of sample and ever since give small amount of signal even when there is no template (or add that small amount of false signal to real one)

real-time PCR is always performed using expression of a housekeeping gene (e.g. actin), measuring relative expression of gene of interest to expression of housekeeping gene

advantages & disadvantages

- A - PCR is extremely sensitive way of looking for any nucleic acid (DNA / RNA)  
- some people claim one-molecule precision
- D One gene at a time  
[no information about 5' end (not useful for mapping TSS)]

housekeeping gene: a gene that is essential for basic functions of a cell and therefore it is expressed by cells of all types & at same levels regardless of (outside) conditions (e.g. stress)

## RNA Seq (in vivo only)

- process - isolate RNA

- isolate mRNA from rRNA/tRNA

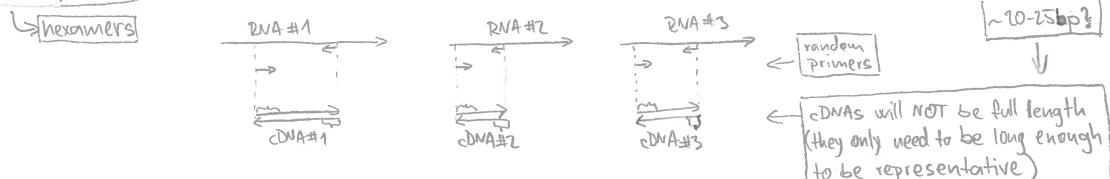
- eukaryotes: enrich for poly(A) RNA (eukaryotes polyadenylate mRNA as part of process of producing mature mRNA)

→ isolate poly(A) RNA

- prokaryotes: there are methods to eliminate RNAs w/o a 5' tri-phosphate (most rRNA/tRNA is processed by cutting off their 5' tri-phosphate)

- make cDNA (complementary/copy DNA)

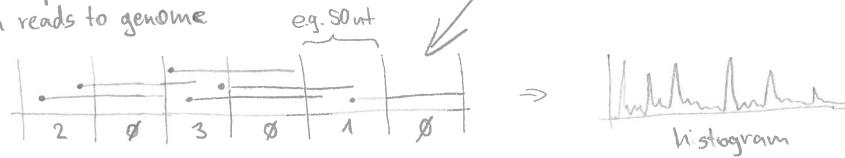
- made using random primers and reverse transcriptase



- shear cDNA into smaller pieces (~300-500 bp)

- deep sequence

- align reads to genome



need to have sequenced disassembled genome first in order to be able to align reads to it

e.g. treating cells by a drug can influence half-life of mRNA instead of expression rate - it is VERY IMPORTANT to keep in mind that RNA Seq measures only steady state (always, including relative comparisons) !

- advantages - information on all transcribed genes

- disadvantages - not intrinsically quantitative (can do relative quantitation) ←

measures steady state RNA levels (i.e. amount of mRNA, as opposed to speed of transcription) ←

- includes rate of txn + rate of turnover! ←

RNA half-life

can compare e.g. two different cells, but can not tell how much mRNA is there in a cell in absolute amounts

e.g. two genes transcribed at the same rate, but RNA from one has half-life of 30 sec and RNA from the other has half-life of 24 hours, this assay will only tell that there is much more mRNA for the second gene, but will not say why.

This may make a false impression that the second gene (24h half-life) is making much more RNA than the first one (30 sec. half-life) !

↑ steady state ↑  
↓ txn rate ↓

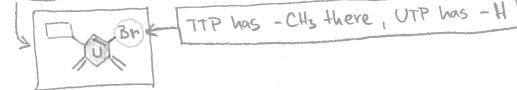
## GRO Seq (Global Run-On Seq)

- process - isolate nuclei or permeabilize cells

- goal is to label any RNA Pol engaged in txn at the time of isolating nuclei/permeabilizing cells

- depletes any nucleotides present (they are very soluble & will get diluted tremendously)

- add rNTP including BrUTP (Bromo-UTP) ←



- there are anti-bodies that specifically bind BrUTP  
- RNA PIs will start incorporate BrUTPs into RNA (there will still be some residual UTPs incorporated as well, but just a few BrUTPs is enough for anti-bodies to recognize the Br-marked RNA)

- isolate RNA

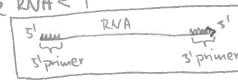
- use anti-bodies to immuno-precipitate the BrUTP-containing RNA

- ligate different primers onto 5' and 3' ends of the RNA ←

- make double-stranded copy DNA w/ RT

- deep sequence

- advantages - information on ongoing txm (strand specific) → discovered by GRO Seq



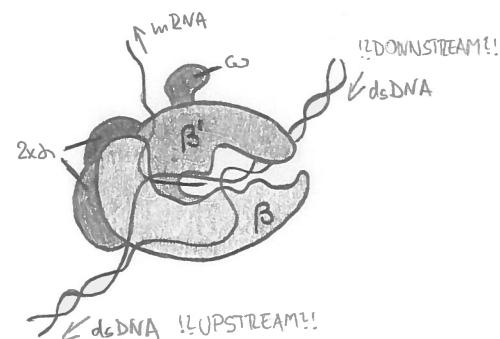
- in eukaryotes, RNA PIs are often fully loaded at TSSs, waiting to do txm

- for many eukaryotic promoters, there are two RNA PIs, each one going opposite direction (mechanism of recruiting RNA PIs turns out NOT to be so unidirectional); this was not discovered before, because RNA going "wrong" direction is rapidly degraded

# MECHANISM OF BACTERIAL TRANSCRIPTION

## RNA POLYMERASE

- 5 subunits:  $\beta$  (156 kDa),  $\beta'$  (151 kDa), 2x  $\alpha$  (2x 37 kDa),  $\omega$  (10 kDa)
- active site at interface between  $\beta$  and  $\beta'$  subunits ("pincers of a lobster claw")
- 2x  $\alpha$ : "hinge of a lobster claw"
- $\omega$ : at the bottom of  $\beta'$ , not clear what is its purpose, but it is conserved through essentially all organisms
- channels
  - downstream DNA channel (between  $\beta$  and  $\beta'$  subunits)
    - guides dsDNA into active site
    - template channel
      - guides template DNA strand into active site
      - non-template channel
        - guides non-template DNA strand away from active site
        - RNA exit channel (mRNA, upstream DNA)  $\leftarrow$  blocked by  $\sigma_{34}$  initially
        - removes RNA from the template strand (helicase-like)
      - NTP channel
        - gives access to NTPs at active site



- Core RNAP can initiate on nicked dsDNA

[ DNA ends ]

can NOT initiate on e.g. intact dsDNA circle

- non-sequence-specific

- 6th subunit of RNA Pol confers sequence specificity

- multiple of these per cell ( $\approx 4-7$  different):  $\sigma$  (sigma) factors

- they recognize different classes of promoters

[ also ends of dsDNA ]

RNA Pol	nicked dsDNA	dsDNA+gene	$\leftarrow$ dsDNA $\Rightarrow$ closed circular dsDNA
Core RNAP	+++	-	$\leftarrow$ Holoenzyme $\Rightarrow \sigma +$ Core RNAP
Holoenzyme	-	++++	

-  $\sigma$  increases specific transcription  $10^3$ -fold

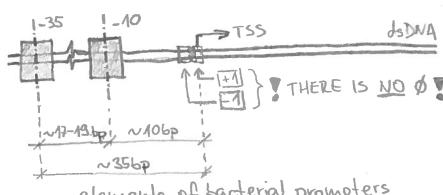
[ decreases non-specific initiation ( $\geq$  1000)  $10^4$ -fold ]

- recognises promoter sequences

[ elongated protein that wraps itself around  $\beta$  ]

[  $\Sigma = 10^3$ -fold preference shift for specific txm ]

## BACTERIAL PROMOTERS



$\leftarrow$  [-10] named based on original canonical promoters from E.Coli  
[-35]

their positions may slightly vary for different promoters,  
but their names are always -10 & -35

- elements of bacterial promoters

$\sigma_2 \rightarrow$  -10 element  
 $\sigma_2 \rightarrow$  transcription start site (TSS) } present in ALL promoters

$\sigma_4 \rightarrow$  -35 element (helix-turn-helix motif) } present in MOST promoters

$\sigma_4 \rightarrow$  -35 - -10 spacing ( $\sim 17-19$  bp)

$\sigma$ -CTD  $\rightarrow$  up element (further "left" from -35) (helix-turn-helix motif)

$\sigma_3 \rightarrow$  extended -10 element (extends -10 element to the "left") } almost always in promoters lacking -35

$\sigma_1 \rightarrow$  discriminator element (downstream / "right" of -10 element)

$\leftarrow$  elements are recognised by  $\sigma$  factor ( $\sigma$  can be divided into 4 domains:  $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ )

$\sigma_2$  binds non-template strand ssDNA (not recognizing sequence, but rather unwound DNA at the region)

up element is bound by C-terminal domain of  $\alpha$  subunit of Core RNAP

- combined these interactions localise RNAP to the promoter

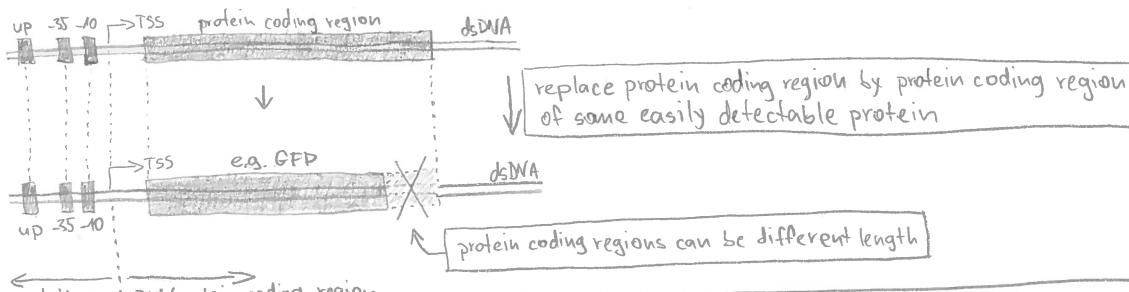
} present in SOME promoters

## MAPPING BACTERIAL PROMOTERS

- 1st map 5' end of the transcript - most bacterial promoters (with VERY few exceptions) are located within ~50bp of TSS
- mutational map sequence upstream of 5'-end (point mutants, small deletions, multiple mutations, etc)
  - conventionally focus at regions -10 and -35 upstream of TSS (but also try to move around for less conventional cases)
  - look at regions upstream of -35 (looking for up-element)
- assay mutants w/ assays already described or Promoter Fusion Assay

### Promoter Fusion Assay

- fuses promoter to easily detectable gene/protein (studying txn indirectly, looking for resulting protein instead of mRNA)
  - e.g. GFP (green fluorescent protein), LacZ (beta galactosidase - easy to assay), luciferase, etc.



mutations  
RNA/protein coding region  
MUST stay the same for  
all samples that are going  
to be compared to one  
another !!!

This is because txm is NOT directly measured. What is being measured is amount of resulting protein, which may differ solely because of regulation of mRNA → protein process. This regulation is governed by sequence downstream of TSS. In order for samples to be comparable, they MUST undergo exactly the same regulation of mRNA → protein process and therefore sequence downstream of TSS must be the same.



## ASSAYS FOR EVENTS OF INITIATION

### Closed complex formation

- DNA binding assay → EMSA ("gel shift")
- 
- template association assay
- 
- protection ("footprinting") assay

- ChIP - does not really have needed resolution (RNAP at promoter vs. just left the promoter)

### Open complex formation

- DNA unwinding assay ( $\text{KMnO}_4$  ssDNA is encircled by active site, need small molecule that can get there)
- 
- nuclease is big and would have hard time to get to ssDNA

- Helicase assay - need to completely separate DNA strands, which never happens in txn

$\text{UTC} \rightarrow \text{STC}$  transition (unstable to stable ternary complex transition)

$\text{UTC}$ : 8-10nt long RNAs

ratio  $\frac{\text{UTC RNAs}}{\text{STC RNAs}}$

$\text{STC}$ : full length RNAs

- incorporation assay + gel electrophoresis

- measure ratio of abortive RNA transcripts (8-10nt long) to full length RNA transcripts

- need to do SINGLE round of initiation

- heparin: polysulphated polysaccharide (widely used as blood clotting prevention agent)

  ↳ mimics ssDNA (polyphosphate polysaccharide)

- inhibits binding of  $\text{RNAP}^{\text{Holo}}$  to promoters (including OCF) - probably acts mostly at CCF  $\rightarrow$  OCF transition

- does not affect RNAP that is engaged in txn, or  $\text{RNAP}^{\text{Holo}}$  at OCF

  OCF is committed step, heparin can NOT affect anymore

- process - bind  $\text{RNAP}^{\text{Holo}}$ , allow OCF (does not require rNTPs)

- add rNTPs (start reaction) + heparin (does not affect already engaged RNAPs, but prevents formation of new OCFs)
  - heparin acts as "challenger" in template challenge assay (can not add 1000x excess of template, as it is done in template challenge assay, because there are no primers here  $\Rightarrow$  can not have labelled/unlabelled primers)
    - adding 1000x excess of unlabelled template  $\rightarrow$  productive interaction (just can not see unlabelled product)
    - adding heparin  $\Rightarrow$  non-productive interaction (there is no product)

- allow synthesis to occur

separate products on denaturing gel



- need to pro-rate signal strengths - rNTPs being incorporated are labelled  $\Rightarrow$  single 1kb long RNA will have much stronger signal than single 8-10nt long RNA (e.g. label only rCTPs and divide signal strength by number of Cs in the sequence)
- compute ratio of UTC/STC products

both signals will be approx. same strength (in absolute terms), because STC product is usually  $\sim 100$ x longer than UTC products and there is  $\sim 100$ x more of UTC products than STC products

- do the same under different conditions (presence/absence of activators/inhibitors)

## - 2. Elongation of txin (bacterial mostly)

- usually σ is released shortly after elongation (STC) starts  $\Rightarrow$  less σ than RNAP is needed (σ not needed for elongation).
- σ<sub>70</sub> has strong interactions w/ promoter } these are probably driving the release of σ, but σ can be still found attached
- σ<sub>70</sub> is removed from RNA exit channel } to RNAP up to ~200nt downstream (can not be bound to promoter at that point)

Q: really? DNA forming ~200bp-long loop?

### activities of elongating RNAP

- RNA synth
- DNA Helicase
- (DNA melting) - not elongation, initiation only
- RNA Helicase - separates synthesized RNA from template strand
- DNA Rewindase - probably coupled w/ RNA Helicase activity

same as in DNA Pol

### transcriptional proofreading

- RNAP intrinsically makes 1 mistake per  $10^4$  nt synthesized
- incorrect incorporation slows down RNA synthesis (3'OH is positioned incorrectly)

① stays in NTP channel for a while (more constricted space than with DNA Pol), thus it can not be immediately degraded by pyrophosphatase

- removing incorrect nucleotide

#### pyrophosphoryl editing

- reverse of RNA synthesis
- PP<sub>i</sub> (pyrophosphate) attacks bond between the newly added (incorrect) rNMP and the previous one
- this kind of reversing can be in fact happening all the time (still reverse reaction is not energetically favourable), but pausing RNA synthesis on adding incorrect nucleotide rises chances of reverse reaction to happen
- can occur ONLY on the last nucleotide incorporated
- RNAP makes 1 mistake per  $10^4$  nt synthesized INCLUDING pyrophosphoryl editing

### hydrolytic editing

- process - reverse translocate (move a few bases backwards)

- RNA end moved into rNTP channel
- further stalls the reaction (3'OH is in rNTP channel  $\Rightarrow$  completely misplaced)
- driven by thermal energy (will not happen at 0°C, but will at 37°C)

#### cleave displaced RNA by GreB (E.Coli) / TFIIS (eukaryotes)

- GreB / TFIIS binds RNAP and reaches into rNTP channel to cleave displaced RNA
- it is not known, but there must be a signal for stalled RNAP to recruit GreB / TFIIS

originally described as stimulator of elongation  
(after adding TFIIS, RNA synth will proceed faster, because of shorter stalls)

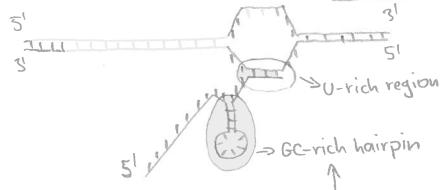
## - 3. Termination of bacterial txin

most genes terminate this way

### intrinsic termination (intrinsic $\Rightarrow$ encoded in the gene)

- requires GC-rich hairpin to form in RNA (stalls RNA synth)
- U-rich region in RNA (7-9nt) 3' of the GC-rich hairpin (makes it easier to separate DNA from template strand)

on each side - can not be much bigger, because of space constraints



there is no specific consensus sequence, it just needs to be able to form reasonably strong hairpin of 7-9nt on each side

it is unknown, but there must be some signal, since A-rich region would be ignored

! RNAP poorly interacts w/ dsDNA  $\Rightarrow$  it is released from the dsDNA !

RNAP w/o σ interacts w/:  
- ends of dsDNA  
- nicked dsDNA  
NOT w/ intact dsDNA

even w/ σ, it would require promoters

- the exact mechanism is unclear, but GC-rich hairpin does not seem to be an obstacle for releasing σ-synthesized RNA from RNAP, once RNAP has been released from dsDNA

it is rapidly released

### Rho-dependent termination

- requires Rho 5'→3' Helicase - it binds ~40nt C-rich sequence (Rho UTilisation sites, a.k.a. RUT sites)
- no specific consensus sequence (just ~60-70% + content of C)

- can only bind RNA after it exits RNAP
- once bound, Rho translocates on RNA towards the engaged RNAP
- once it reaches RNAP, it literally pulls RNA out of RNAP (and its active site)
- after removing RNA, DNA strands reanneal and RNAP gets released

- RUT site is in the 3' untranslated region (so that it does not interfere w/ the sequence encoding protein)
- this is LESS precise than intrinsic termination, because exact termination point depends on:

- how fast/soon Rho binds the RUT site
- how fast Rho manages to pull RNA out of RNAP's active site

## 4. Regulation of bacterial txn

- how is rate of initiation controlled

- differences in promoter sequence (increase/decrease CCF)

- consensus match [ strong: increase CCF  
weak: decrease CCF ]

- alternative sigma factors

$\sigma$ factor	( $\sigma$ gene)	function	-35	spacing	-10
$\sigma^{70}$	rpoD	general	TIGACA	17-19 bp	TATAAT
$\sigma^{32}$	rpoH	heat shock	CCCTTGAA	13-15 bp	CCCAGTNT
$\sigma^{54}$	rpoN	nitrogen starvation	CTGGNA	6 bp	TTGCA

- rpoH and rpoN genes need to be turned on by some signal  
both use  $\sigma^{70}$  for txm initiation

-35 is just a name, it is not necessarily at -35 position

listed below are the consensus sequences, which are rarely exactly present in genome - it is sufficient for promoter sequence to be close enough; exact matches have very strong affinity for  $\sigma$  factors; some positions (underlined) are more important than others

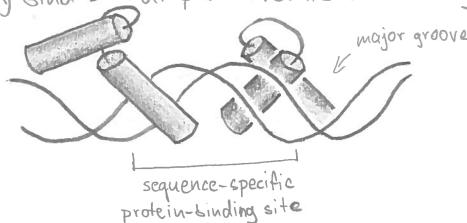
AT-rich -10 sequence makes it easier to melt DNA strands

heat shock -10 sequence does not have to be so much AT-rich, because it is much easier to melt DNA at higher temperatures

surprising: not that AT-rich as  $\sigma^{70}$

binding of seq-specific txm factors

- typically bind DNA at palindromic sites using helix-turn-helix motif



## transcriptional regulators in bacteria

unbound promoter

$$k_1 \downarrow \uparrow k_1$$

closed complex

$$k_2 \downarrow$$

open complex

$$k_3 \downarrow$$

unstable ternary complex

$$k_4 \downarrow$$

stable ternary complex

- rate-limiting step:  $k_1$  (often)

- activators/repressors will concentrate on rate-limiting step (speeding up/slowing down other steps would have little to no effect)

- repressors reduce rate-limiting step

create NEW rate limiting step

bind spacing DNA (between -10 and -35 promoters): RNAP must interact w/ both -10 and -35 promoters and protein binding DNA in between them effectively blocks any access for RNAP<sup>holo</sup> (binding -10 or -35 promoter directly would block all genes using given  $\sigma$  factor - they all have same promoters, but different spacing DNA sequences)

bind upstream of -35 promoter and hold RNAP in place (prevents  $k_4$ )

- enables "lock-draw... wait for it" ("loose" on release of repressor) ←

- activators increase the rate-limiting step of the reaction (often creates NEW rate-limiting step)

typically (in bacteria) the promoters of genes that are activated by DNA binding activators are weakened (e.g. poor match to -35 region)

- in human-speak: if the rate-limiting step is poor match of -35 region sequence, preventing RNAP<sup>holo</sup> from recognizing it, activator (e.g. CAP) can bind upstream of -35 sequence and facilitate CCF ( $k_1$ )

- such activator can act also as repressor ←

coactivating protein

"lock-draw... wait for it"

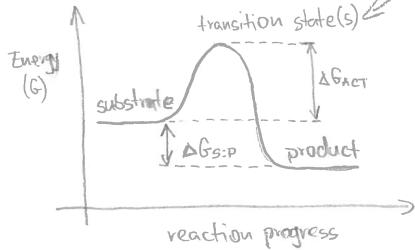
promoters (-10 & -35) are too far apart (shifting promoters will also cause them to "rotate around DNA axis", which may result them facing wrong side of DNA)

- MerR activator: binds between -10 & -35 and causes DNA to overtwist, effectively bringing these promoters closer to each other (correct distance now), but also rotates them to face the same direction (correct now, again)

- RNAP<sup>holo</sup> can bind ( $k_1$ ), because -35 is ok, but OCF is prevented by wrong positioning of -10 → MerR allows OCF ( $k_1$ )

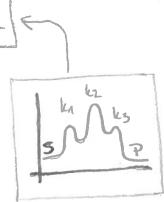
## BIOLOGICAL RATES & REACTIONS

- reaction coordinate diagram



there can be many intermediate transition states, each of them having its own  $\Delta G_{ACT}$  and  $k$

$\Delta G_{s:P}$  [  $< 0 \Rightarrow$  favourable (e.g.  $ATP \rightarrow ADP + Pi$  has  $\Delta G_{s:P} = -7.3 \text{ kcal/mol}$ ) ]  
 $> 0 \Rightarrow$  unfavourable ]



$\Delta G_{ACT}$  - activation energy [ inversely related ]  
 $k$  - rate of reaction [  $\Delta G_{ACT} \downarrow \Rightarrow k \uparrow$  ] [  $\Delta G_{ACT} \uparrow \Rightarrow k \downarrow$  ]

(e.g.  $ATP \rightarrow ADP + Pi$  has  $\Delta G_{ACT} = 32 \text{ kcal/mol}$ ,  $k \approx 2 \cdot 10^{-11} [\text{ATPs hydrolyzed/second}]$ )

↓ 1 mol of ATP hydrolyzed in  $\sim 1600$  years

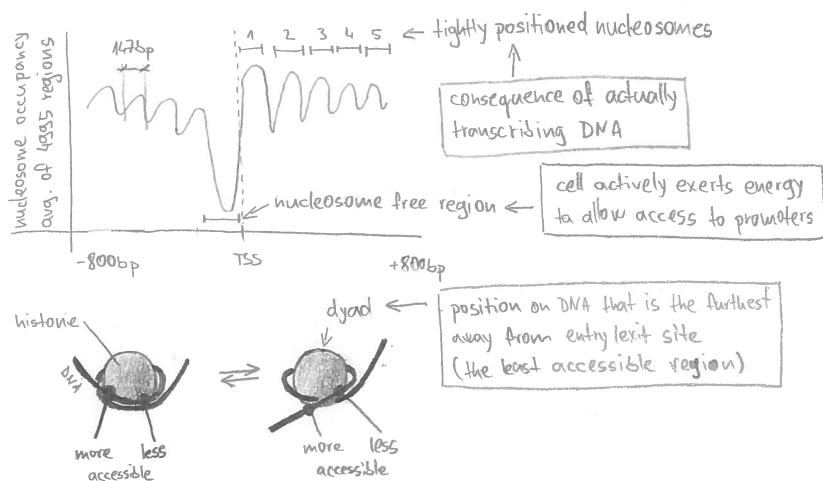
- enzymes [ lower  $\Delta G_{ACT}$  (catalysis)  
do NOT change  $\Delta G_{s:P}$  ]

## EUKARYOTIC TRANSCRIPTION

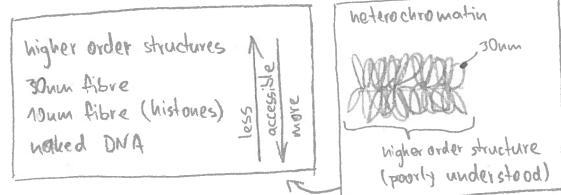
the same is true for other processes as well

- nucleosomes/chromatin are in general inhibitor of euk. transcription (as opposed to bacteria, where most of transcription regulation is negative, in eukaryotes most of transcription regulation is positive, helping to overcome general inhibitor of nucleosome)

### NUCLEOSOMES AND DNA ACCESSIBILITY



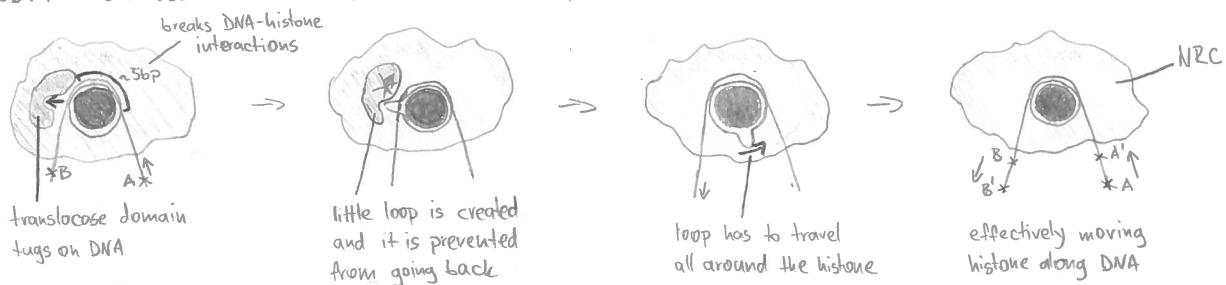
RNA Pol I	$\rightarrow$	rRNA
RNA Pol II	$\rightarrow$	mRNA
RNA Pol III	$\rightarrow$	small RNAs



### REGULATORS OF NUCLEOSOME FUNCTION/LOCATION

#### NUCLEOSOME REMODELLING COMPLEXES (very large: typically 4-8 subunits)

- use ATP binding/hydrolysis to slide histone octamers on DNA (protein binding region will be put on nucleosome-free region)
- remove/exchange histone octamers from one DNA to another (- - -)
- exchange histone subunits (change function of histone)
- NOT modify histones present in nucleosome (e.g. modify histone subunit tails)
- primary properties of nucleosome remodelling complexes used to manipulate nucleosome location
  - bind to histone octamer (disk-shaped pocket inside NRC, that encapsulates whole histone)
  - dsDNA Translocases (Translocase: e.g. RuvB - helicase-like protein responsible for manipulating Holliday junctions, hydrolysing ATP)



- there is lot of different NRCs w/ different properties

#### HISTONE VARIANTS

histone subunits

- canonical: H2A, H2B, H3, H4 (each histone has 2 of each)
- variants: particularly of H2A and H3

protein structure that builds on a chromatid during cell division and allows it to attach to a spindle fiber on a chromosome

- CENP-A: has version of H3 that interacts w/ proteins that build kinetochore
  - different N-terminus & C-terminus that interacts w/ kinetochore
- specific NRCs exchange canonical H3 for CENP-A H3 (only at centromere)
- H3.3: different H3 (differences particularly at N- and C- terminus)
  - found preferentially at sites of transcription
- H2A.X: has version of H2A
  - found at sites of ds-breaks (used to detect ds-breaks by many assays)
  - gets rapidly phosphorylated
    - there are antibodies for phosphorylated H2A.X
- H2A.Z: has version H2A
  - frequently found at boundaries of nucleosome-free regions

at centromere region of chromosome

we do not understand what puts it on centromere

## HISTONE MODIFYING ENZYMES

- enzymes that add molecules to specific histones

large: e.g. ubiquitin (molecule almost the same size as histone itself - ~100 amino acids long)

small (much more common): e.g. acetyl group, methyl group



- typically targets the unstructured (N-termini) of the histone

positively charged tails sticking out of histone, potentially interacting w/ DNA

especially acetyl group alters charge of histone tails  
typical modifications happen on Lysines & Arginines (positively charged residues)

- modifications are dynamic (i.e. they can be put on and off)

- most common modifications [acetylation of lysine (observed also on serine, threonine)]

[methylation of lysine or arginine (observed also on lysine)]

- also observed ubiquitination (lysine)

[phosphorylation (serine, threonine, tyrosine, histidine)]

lots of them

"WRITERS"

- enzymes [acetylation [addition: histone acetyltransferase (HAT)]

[removal: histone deacetylase (HDac)]

[methylation [addition: histone methyltransferase (HMT)]

[removal: histone demethylase (HDM)]

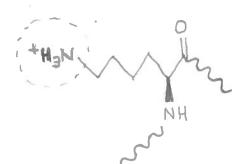
HDac tend to be position N-specific

! specific position !

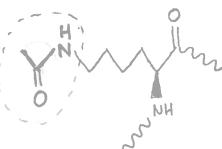
- many of them, each specific for histone and site (e.g. H3K4 methyltransferase methylates only lysine④ on H3)

may be even more than one for single site on single histone: redundant function

- examples acetylation of lysine



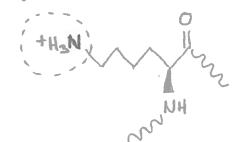
acetylated



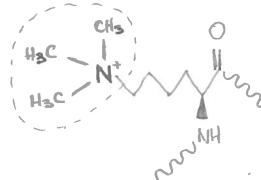
uncharged

acetylation changes charge  
(from positively charged  
to uncharged)

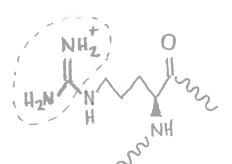
methylation of lysine



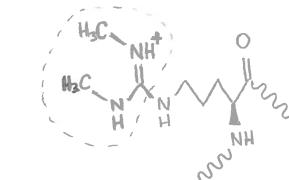
methylated



methylation of arginine



methylated



methylation "locks-in" the  
charged state (it is not  
possible to acetylate  
methylated amino acid)

negatively charged

## EFFECTS OF HISTONE MODIFICATIONS

### DIRECT

- acetylation [more accessible DNA (acetylated, i.e. uncharged, histone tails do not interact with DNA) as good as positively charged tails)

interferes w/ 30nm fiber formation (there are indications of histone tails interacting w/ other histones in order to form 30nm fiber and these interactions are disrupted by acetylation) this is not fully proven theory

- favors accessible DNA (and therefore transcription)

- methylation - does NOT strengthen interactions disrupted by acetylation methylation does NOT have symmetrically opposite effect to acetylation !

INDIRECT - recruiting other proteins to the site of modification (mediated by small protein domains that bind to specific modifications at specific sites)

applies to all of these

chromodomains - recognize sites of acetylated lysine (specific to histone and site, e.g. H3K27 recognize acetylation on lysine 27 of H3)

chromodomains - recognize methylated lysine/arginine

TUDOR  
PHD finger (plant homeo domain)  
BATH (bromo adjacent homology)

} recognize methylation sites

quite possibly many more

"READERS"

- single one for acetylation  
- multiple for methylation  
probably because methylation does not have any effect just by itself

- histone modification reader domains are typically part of either large protein or protein complex

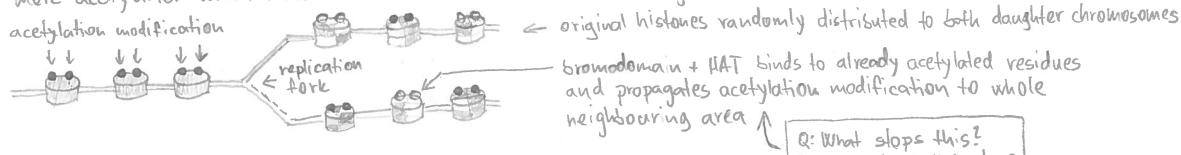
- examples

- H3Kme: usually associated w/ turning off transcription

- chromodomain for H3Kme + HDAC (histone deacetylase) binds the methylated site and the associated deacetylase removes acetylation modifications, restricting access to DNA and transcription promoters

- propagation of histone modifications during/after replication

- bromodomain + HAT (histone acetyltransferase) bind the acetylated site and the associated acetyltransferase adds more acetylation modification to the histone (and also adjacent histones as well)



Q: What stops this?  
A: Answer in next lecture

- activators bind to DNA associated w/ histone (possibly 3Dm fibre), recruit HAT, which then acetylates surrounding region, allowing access to DNA for e.g. transcription (by disrupting 3Dm fibre & histone-DNA interactions)

- usually, DNA needs to be accessible for binding, activators binding like this are more exception than rule

- TFIID (fundamental factor that localizes RNA Pol to DNA - like e.g. G factor) has bromodomain

- recruits RNA Pol to sites of acetylated (i.e. accessible) DNA

many more

- Histone Code Hypothesis

- histone modifications lead to specific consequences

- full coding capacity does NOT seem to be used by far

- e.g. one study in drosophila genome described 18 different modifications commonly present

$2^{18} = \approx 250,000$  possible combinations on each nucleosome } "class invariants" (as in object oriented programming)  
only 9 distinct combinations seem to be present

- there are clear examples of recognition of COMBINATIONS of modifications (by individual proteins)

- there are probably different types of modifications: dominant, combinatorial, additive, ...

- but also simple on/off switches (e.g. H3K4me → activated chromatin, H3K9me → repressed chromatin)

almost always

- different meanings at different promoters

## NUCLEOSOME POSITIONING

- where nucleosomes are located on DNA (they are NOT positioned randomly - see Nucleosomes and DNA Accessibility, pg 5.1)

### DNA SEQUENCE-DEPENDENT POSITIONING

- most effective at predicting nucleosome-free regions

- based solely on DNA sequence - nucleosomes occur at regions of alternating AT-rich and GC-rich sequences ⇒ regions not fulfilling this criteria will, with high probability, not contain nucleosomes

either only AT-rich or only GC-rich

each ~5bp

promoters are generally more AT-rich than surroundings  
- easier to separate DNA strands  
- favors nucleosome-free regions !!!

### PROTEIN DEPENDENT POSITIONING

- most DNA-binding proteins can NOT interact w/ nucleosome-bound DNA (this is true also in opposite direction: presence of these proteins interferes w/ nucleosome formation)



, but



## ASSAYS FOR NUCLEOSOME REGULATORS

### HISTONE MODIFYING ENZYMES (in vivo)

- $\alpha\beta$ s (antibodies) that recognize specific modifications (it is not easy to design them, because they need to recognize a) specific histone subunit, b) specific amino-acid at its specific position, c) tiny modification (acetylation; mono-, di-, tri-methylations; ...)
- ChIP / ChIP Seq to map sites of modification
- typically, these assays are done +/- inducer or +/- transcription factor (need to compare two states in order to see difference of a modification at particular promoter site, not just genome-wide)
- process: crosslink protein to DNA using formaldehyde

- shear DNA
- immunoprecipitate target protein and associated DNA
- reverse crosslink & purify DNA
- prepare DNA for sequencing
- sequence DNA libraries & map reads back to the genome

when looking for information if two modifications are on the same DNA molecule:  
1. immunoprecipitate the first modification  
2. on resulting DNA-protein complexes, immunoprecipitate the second modification  
3. repeat 2) for more modifications if any  
4. sequence & map DNA

for more

### MONITOR NUCLEOSOME POSITIONING (in vivo)

- MNase Seq (may be good for in vitro as well, but it is expensive)
- +/- inducer/activator

to see if an enzyme modifies a histone at all, it is much easier to treat just histone octamer w/ that enzyme and do mass spec

### NUCLEOSOME REMODELLING (in vitro)

- process: assemble nucleosome (single one) on DNA ( $\sim 280$  bp long)

- purified histone subunits (e.g. H<sub>2</sub>A, H<sub>2</sub>B, H<sub>3</sub>, H<sub>4</sub>)

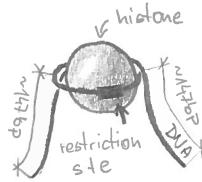
- DNA

- positioning signal (AT-rich, GC-rich alteration each  $\sim 5$  bp)

- restriction site (such that if the region of DNA w/ the restriction site will interact w/ histone  $\rightarrow$  can not be cut)

- high salt buffer (proteins interact poorly in high salt buffer)

- slowly reduce salt (slowly improving protein-DNA interactions will result a nucleosome forming on DNA)



can add more restriction sites for different restriction enzymes in order to determine where has the nucleosome been moved to

- treat w/ putative remodelling complex (or not - for negative control)

compare ability of restriction enzyme to cleave DNA

- will be able to cleave much better if nucleosome has been moved, exposing restriction site

- acetylation of histone may interfere, because it loosens histone-DNA interactions, somewhat exposing restriction site

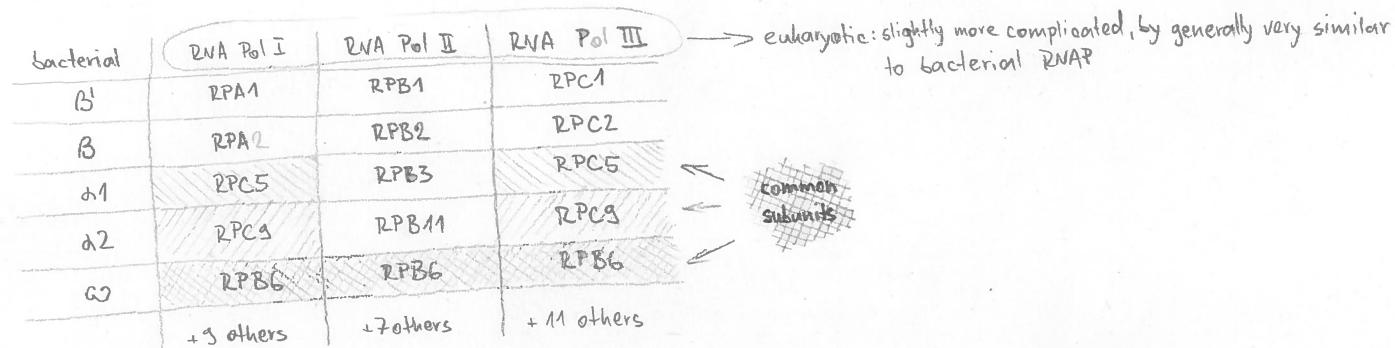
↑ acetylating agent is NOT nucleosome remodelling agent

- removing histone octamer will have same effect as moving it away from restriction site

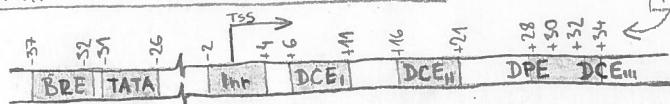
# EUKARYOTIC RNA POLYMERASES

	% of total RNA	# of genes transcribed	product	$\alpha$ -amanitin sensitivity
RNA Pol I	~80%	1	rRNA	insensitive
RNA Pol II	1-5%	~20,000	mRNA	very sensitive
RNA Pol III	~15%	100	tRNAs & small RNAs	weakly sensitive

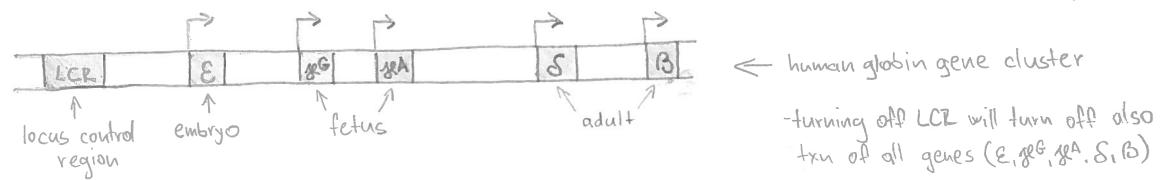
Toxin produced by amanitoid mushrooms:  
 - Amanita verna  
 - Amanita bisporigera } "destroying angel"  
 - Amanita ocreata  
 - Amanita phalloides: "death cap" (mucho-trávica zelená)  
 - Amanita verna: "fool's mushroom"  
 (oral LD<sub>50</sub> ~ 0.1 mg/kg)



## RNA POL II PROMOTER STRUCTURE



## Locus control regions / super-enhancers ("enhancers on steroids")

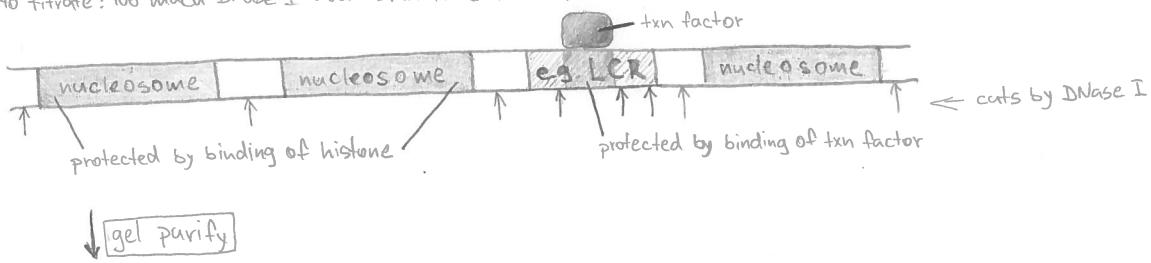


- act on a set of proximal genes
- act to open chromatin of a large region - by recruiting histone modifying enzymes/remodelling complexes
- tend to be larger than conventional enhancers  
[located farther from target gene(s)]

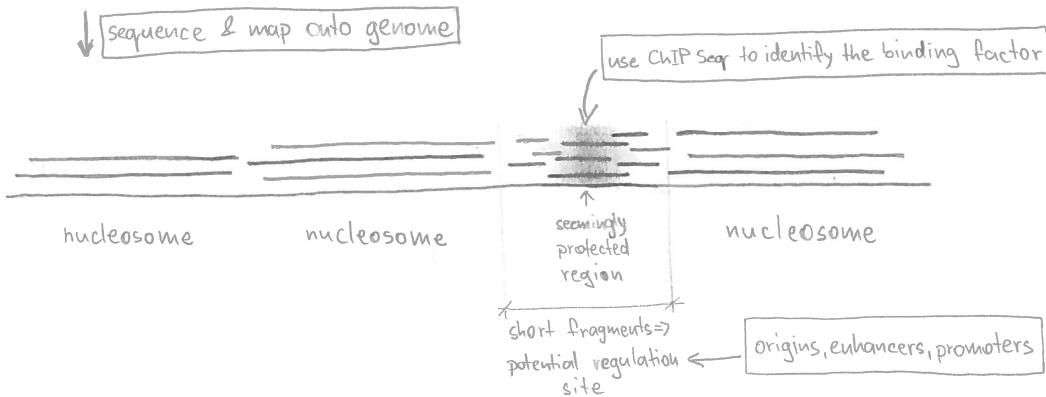
## ASSAYS

### DNase I Seq

- using the same enzyme as DNase I footprinting
- DNase I vs. MNase: both are endonucleases, but MNase is also exonuclease (i.e. will chew fragments until they have reached their minimal protected size)
- sites involved in regulation are typically hypersensitive to DNase I (but the part bound by various factors is still protected)
- need to titrate: too much DNase I would start to cut also into nucleosome-bound DNA !



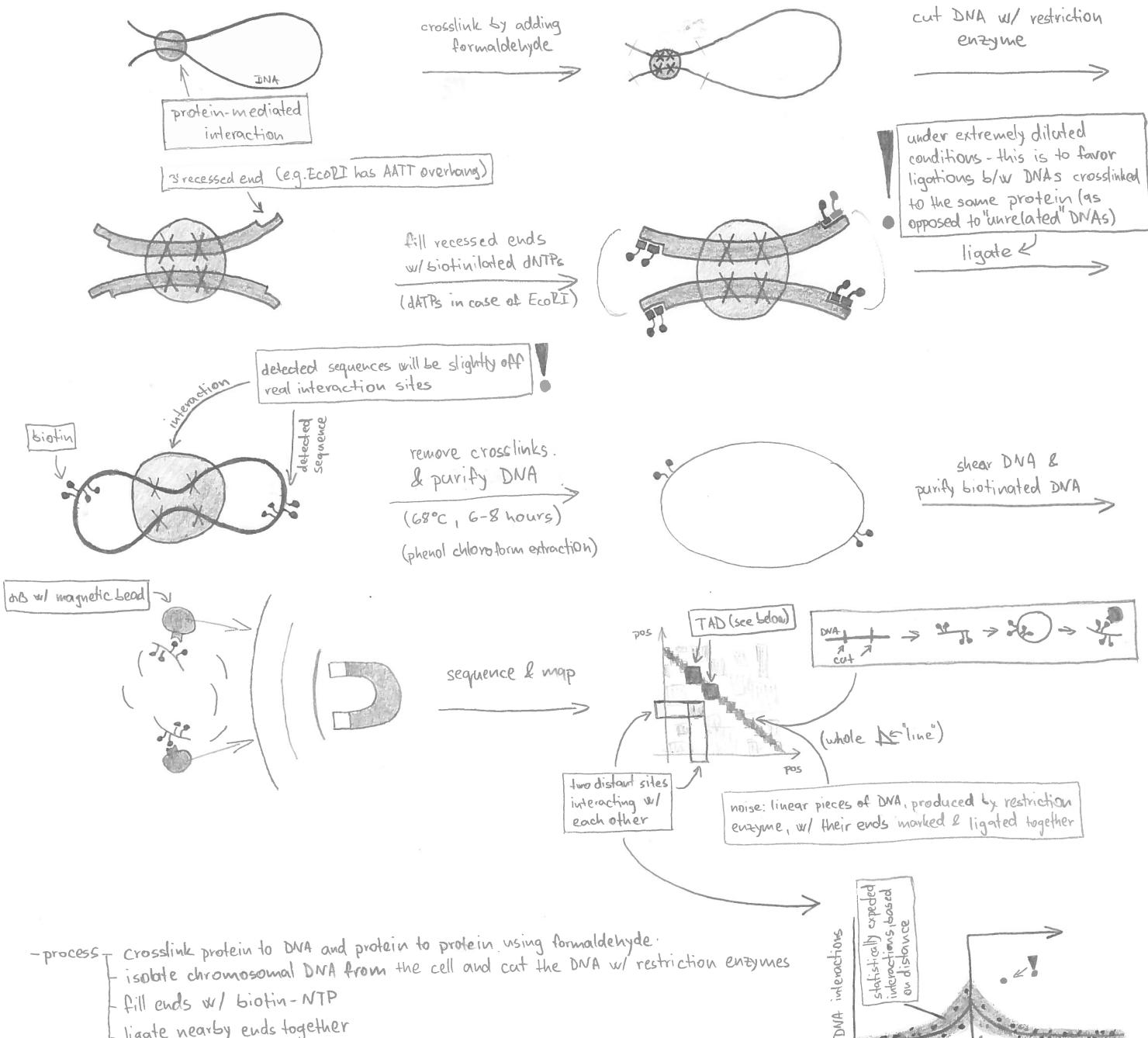
- long fragments (145-150bp) from nucleosomes
- short fragments from nucleosome-free regions (regions involved in regulation)



- used to comprehensively identify regulatory regions throughout the genome

## CHROMOSOME CONFORMATION CAPTURE (3C)

- Hi-C: 3C + deep sequencing



- process
    - crosslink protein to DNA and protein to protein using formaldehyde
    - isolate chromosomal DNA from the cell and cut the DNA w/ restriction enzymes
    - fill ends w/ biotin - NTP
    - ligate nearby ends together
    - reverse crosslink and purify DNA from protein
    - shear the DNA into lengths appropriate for sequencing
    - pull down DNA w/ biotin using streptavidin - coated beads
    - deep sequence the DNA fragments
    - search for junctions b/w distant sequences

-observations - very little interaction b/w different chromosomes

within chromosomes, there are compartments [ A: highly transcribed, open chromatid, gene-rich  
 B: poorly transcribed, closed chromatid, gene-poor ] ~5M6 on avg in size

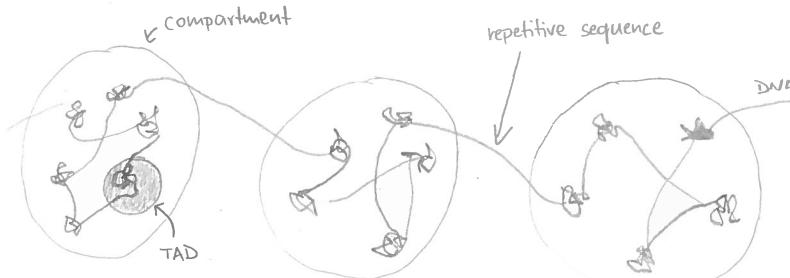
- boundaries b/w compartments change depending on the cell type  
topologically associated domains (TADs)

- 0.1-1 mb in size

- do not change in different cell types

- very high interact

- very high inter-action rates



## ChIA-PET (Chromatin Interaction Analysis using Paired-End Tag sequencing)

- answering the question: "Is the given protein involved in looping?"

- ChIP + 3C: protocol very similar to 3C/Hi-C

  | does NOT use biotinylated dNTPs (?)

  | uses  $\alpha\beta$  against the protein itself to purify DNA of interest

- observations - there is a lot of both [ interactions within single chromosome  
[ interactions b/w chromosomes ] ]

CTCF protein

- Hi-C can search for all distant DNA interactions throughout the genome, ChIA-PET targets interactions mediated by individual proteins

## MECHANISM OF POL II TXN INITIATION

- RNA Pol II GTFs - act similar to sigma factor in prokaryotes

- TFIID\*
- TFII B\*
- TFII C: binds to nicked DNA to block it from RNA Pol II binding
- TFII D+
- TFII E\*
- TFII F\*
- TFII G: part of TFII H

- steps of initiation

TFIID interacts w/DNA

- it contains TBP (TATA Binding Protein), which mediates the interaction in presence of TATA-box
- TBP binds to the MINOR groove and bends DNA dramatically (it is believed that bending is in order to strain and distort DNA prior to unwinding)
- sequence specific (as opposed to RNA Pol I & III) interaction of TBP

- TFIIA

- the only factor that is not absolutely necessary for initiation
- binds to TFIID and stabilizes its interaction w/ DNA

- TFIIB

- binds to TFIID
- BRE element, if present

- RNA Pol II + TFIIF

- RNA Pol II binds to TFIIB

- TFIIE

- required to recruit TFIIH

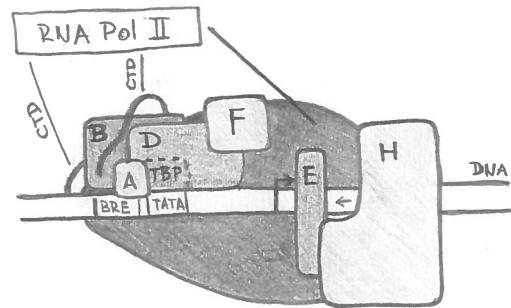
- TFIIH

- ds translocase (ATPase): pushes dsDNA towards RNA Pol II
- by forcing DNA helix to occupy less space, DNA helix "pops open" at TSS



\*: single protein  
+: large protein complex

not sequence-specific interaction of TBP for RNA Pol I & III.  
(it is believed that it is used for bending DNA prior to unwinding)



- kinase (cyclin C1/Cdk9)
- phosphorylates C-terminal domain (CTD) of RNA Pol II

! ATP hydrolysis is required for DNA unwinding (unlike prokaryotic txn) !

- RNA Pol II CTD (C-terminal domain):  $(\text{Tyr}_1 \text{Ser}_2 \text{Pro}_3 \text{Thr}_4 \text{Ser}_5 \text{Pro}_6 \text{Ser}_7)_x \sim 25-52$

- the largest subunit of RNA Pol II, at its C-terminal region, contains many repeats of (N) YSPTSPS (C)

- yeast: 26 repeats

- humans: 52 repeats

- extended unstructured region ("tail" of RNA Pol II)

- unphosphorylated CTD binds big multi-protein complex, the Mediator

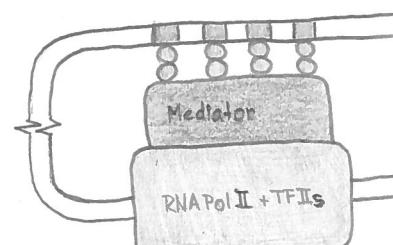
- Mediator interacts w/ txn factors

- phosphorylation of CTD releases the Mediator

- stimulates promoter clearance ("lizard's tail")

- phosphorylation (by TFIIH) occurs on Ser 5 of repeat

- post-initiation: see Q.1



## POST INITIATION EVENTS

### - promoter proximal pausing

- 2 factors bind RNA Pol II after initiation of txm
  - NELF (Negative Elongation Factor)
  - DSIF (DRB Sensitivity-Inducing Factor) - nucleotide analog
  - they inhibit early processive txm (arrest at ~ 25-60 downstream)

### - PTEFb is required for RNA Pol II to escape

- modifies Ser 2 in the CTD (see 9.1)
- modifies NELF and DSIF → causes release of NELF

modifies = phosphorylates

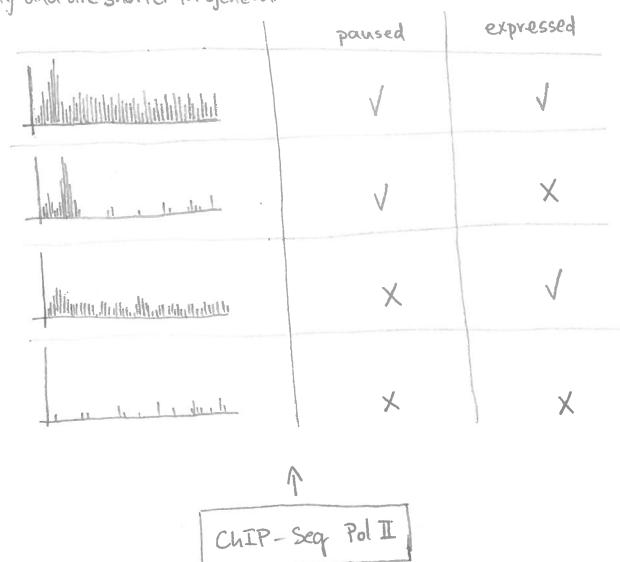
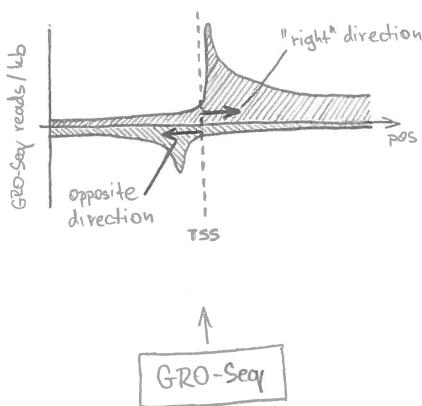
- in some promoters, all of this happens rapidly, but in others these pauses are very pronounced

## BIDIRECTIONAL TXN

- TATA box
  - absent: lots of bidirectional txm
  - strong match: mostly unidirectional txm

- purpose is unclear

- transcripts from opposite direction tend to be degraded rapidly and are shorter in general



## CAPPING AND POLYADENYLATION OF TRANSCRIPTS

- CTD phosphorylation controls post-initiation events

- Ser5-P<sub>i</sub> - mediated by TFIH

- causes release from Mediator

- recruits & activates 5' end modifying enzymes (they add 5' cap)

- 5' Cap - stabilises 5' end of RNA (it would have been rapidly digested by 5'→3' exonucleases)

- recognition element for mRNA translation (in most mRNAs)

- capping increases by 10<sup>5</sup>-fold when capping enzymes bind CTD

these are designed to degrade RNA that is somehow defective

CTD emerges immediately next to the RNA-exit channel

- Ser2-P<sub>i</sub> - mediated by PTEFb (see 8.2)

leads to recruitment of RNA splicing and poly(A) addition machinery

phosphorylation

- 3' end processing (metazoan model)

- polyA signal sequence in nascent mRNA: AUAAA

- CPSF (Cleavage and Polyadenylation Specificity Factor)

- CstF (Cleavage Stimulation Factor)

not conserved

- CPSF & CstF recognize polyA signal sequence + some other sequence (not understood, probably has to do w/ structure of RNA)

- CPSF cuts mRNA (CstF is bound to the part of mRNA being cut off → cut off mRNA & CstF "floats away")

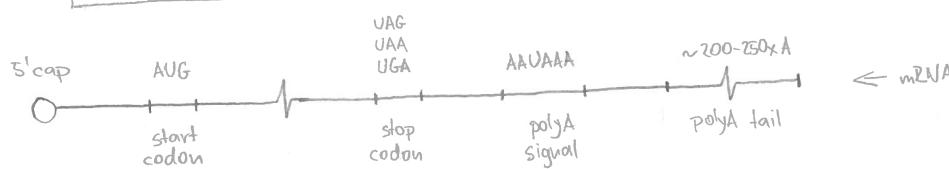
- CPSF recruits polyA polymerase that adds untemplated AMP to the mRNA

- polyadenylation by polyA polymerase creates binding site for PolyA Binding Protein that coats polyA tail synthesized by PolyA Pol

- PolyA Binding Proteins cause PolyA Pol to dissociate from mRNA

↑ cooperative activation ← the more PolyA Binding Proteins are present, the higher chance to dislodge PolyA Pol

regulates length of polyA tail (~200-250 nt)



### RNA POL II TXN TERMINATION

- RATI - processive RNA 5'→3' exonuclease

- can NOT act on capped RNAs

- recruited in a Ser2 CTD phosphorylation-dependent manner

- removes & degrades the remaining RNA (pulls it from RNA Pol II) - similar to Rho in bacteria: see 4.1)

- DNA strands in txn bubble reanneal to form dsDNA

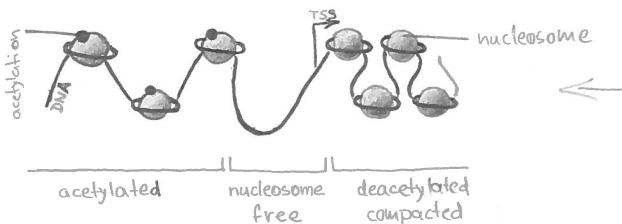
- collapse of txn bubble causes release of RNA Pol II

when the two DNA strands reanneal, txn factors needed to separate them (most notably TFIH) are not available

- phosphorylation of CTD gets rapidly removed (in order to bind to Mediator for start of new round of txm)

remaining Ser2-P<sub>i</sub> needs to be removed before starting new round of txm, so that e.g. RATI does not get prematurely recruited (need to cap RNA first)

### CHROMATIN STRUCTURE AND REGULATION OF TXN



- upstream region: acetylated, promoter sequences accessible to proteins  
- coding region: deacetylated, compacted, NOT accessible  
- prevents txm from starting in the middle of coding region

transcribed DNA (especially highly transcribed) is left in highly compacted deacetylated state (but methylated on H3K36)

- CTD phosphorylation of Ser2/Ser5 recruits a Histone Methyl Transferase SET2 (methylates H3K36)  
- H3K36-Me recruits RPD3s (multi-protein complex) - Histone Deacetylase (removes H3 and H4 acetylation)

- How does RNA Pol II txb through chromatin

- FACT factor (Facilitates Chromatin Txm): SSRP1 + SPT16

- removes one H2A/H2B heterodimer from nucleosome, leaving behind Histone hexamer (Hexosome)

- Hexosome is easier for RNA Pol II to txb through

- SPTG joins FACT to put H2A/H2B heterodimer back (not necessarily the original one)

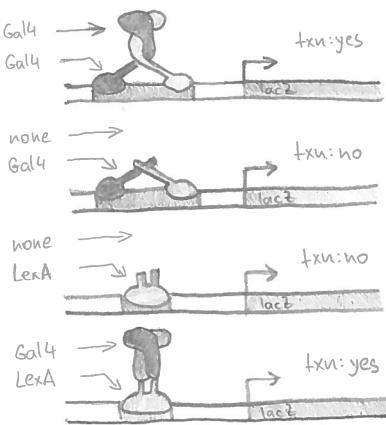
1x H2A/H2B + 2x H3/H4

it is possible to loose/change H2A/H2B modifications

! most modifications are on H3/H4, not H2A/H2B !

## - Regulation of txm

- RNA Pol II txm is mostly POSITIVELY regulated (there already is global negative effect of nucleosomes that needs to be counteracted)
- potential rate-limiting steps
  - inhibition by nucleosomes/chromatin preventing DNA accessibility
    - 30 nm fiber, or
    - nucleosomes occupying the binding sites of DNA-binding txm factors or GTFs
  - recruitment of RNA Pol II / GTFs ('closed complex formation')
  - activity of TFIIH ('open complex formation')
  - promoter clearance: TFIIH kinase function
  - promoter proximal pausing of RNA Pol II: P-TEFb factor (see 8.2)
- much regulation of promoter activity is mediated by DNA-binding txm factors (sequence specific)



General Txm Factors: TFIIIX

- Mark Ptashne experiment (in yeast)

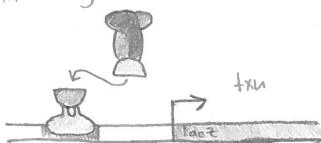
- LexA: bacterial factor inhibiting txm

- proteins are not atomic

- mere binding of a protein does not necessarily provide the effect  
- presence of a part of protein, that conveys the function, at the correct place in correct time is what really matters

ground-breaking work at the time (~25 years ago)

- 2-hybrid assay (extension of Mark Ptashne's experiment)



# SEQUENCE-SPECIFIC DNA-BINDING TXN FACTORS

## - DNA-binding motifs

### - helix-turn-helix (see also 4.2)

- two  $\alpha$ -helices, one of them is inserted into major groove and recognize specific sequence

### - homeodomain

- similar to helix-turn-helix, just not quite as compact

- 3  $\alpha$ -helices (3rd one is inserted into major groove)

- non-specific interaction on the 1st  $\alpha$ -helix (Arg)

### - zinc finger

- recognition  $\alpha$ -helix inserted into major groove

- $\alpha$ -helix held in place by bound zinc molecule that interacts w/ [ 2x Cys, 2x His, or  
4x Cys, 4x His ]

### - bZIP (basic leucine zipper)

- coiled coil: two long  $\alpha$ -helices intertwined together to form helix

- ends of both  $\alpha$ -helices are inserted into major groove, recognizing specific sequence

### helix-loop-helix

- two pairs of  $\alpha$ -helices separated by loop

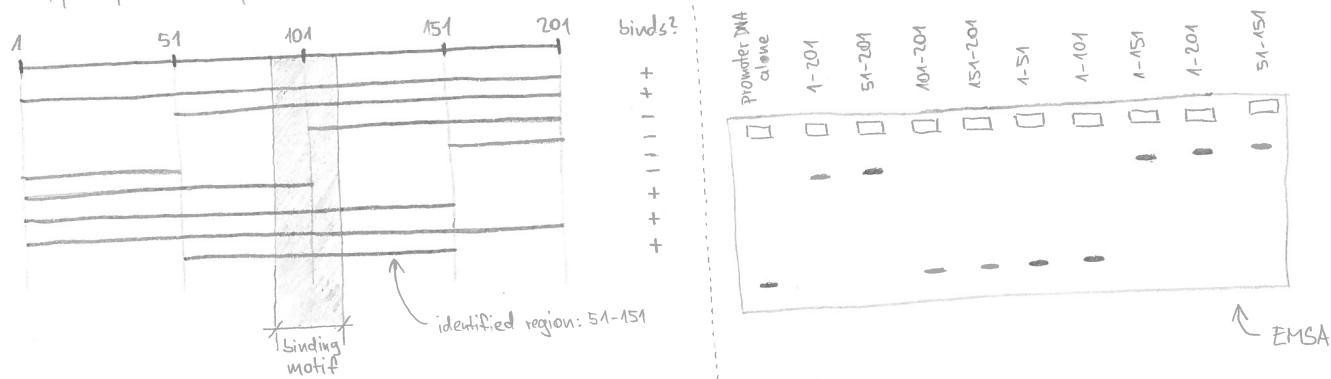
- first  $\alpha$ -helix is inserted into major groove, recognizing specific sequence

- second  $\alpha$ -helix is interacting w/ its counterpart from the other pair, forming dimer

## - mapping DNA-binding motifs

- design truncation constructs of the DNA coding for DNA-binding protein

- expose proteins expressed from these constructs to an appropriate promoter DNA w/ an EMSA assay



## SELEX: SYSTEMATIC EVOLUTION OF LIGANDS BY EXPONENTIAL ENRICHMENT (in vitro)

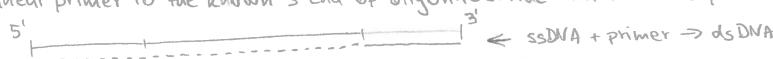
- identifies recognition motif if DNA-binding domain is available

- process: generate a library of oligonucleotides containing random sequences flanked by known ends



number of possible random sequences grows exponentially w/ length of random sequence: NOT all possible sequences will be present (technically impossible - too big volume)

- anneal primer to the known 3' end of oligonucleotide and extend the primer to create dsDNA



- this will lead to better match b/w the two strands (almost 100% matching) than creating anti-parallel pairs of oligonucleotides and annealing them to form dsDNA

- incubate w/ DNA-binding protein of interest

- select for DNA bound to proteins (EMSA or IP)

- purify protein-bound DNA and PCR amplify

- using error-prone DNA Pol is beneficial here, because it adds slight variations of bound DNA sequences.

that bind better than original sequence: climbing up the hill  $\rightarrow$  these will be selected for in next round

bind worse than original sequence: NOT climbing up the hill  $\rightarrow$  these will be discarded by selection step

- repeat incubate/select/purify steps several times ("genetic algorithm")

- sequence oligonucleotides

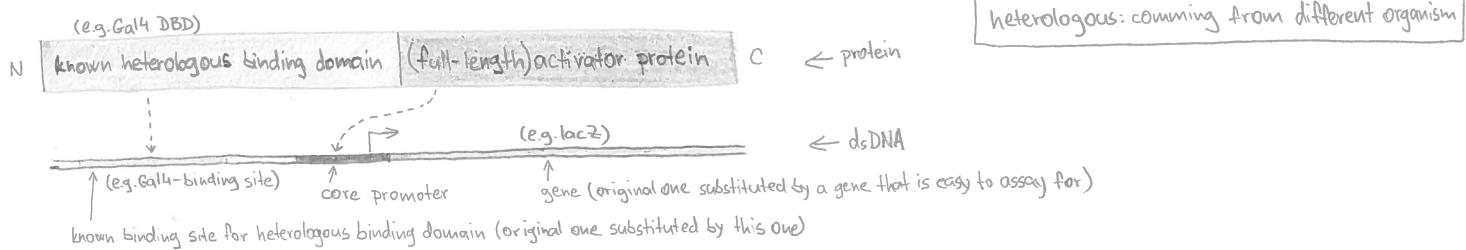
- create a LOGO to visualise the consensus sequence

error-free DNA Pol should be used in the last round  
(there is no benefit in random variation w/o selection)

## - targets for activation domains

- DNA Pol II / Mediator : facilitate closed-complex formation
  - >20 proteins in Mediator (depending on species)
- GTFs (General Trn Factors)
  - typically, this will happen at promoters that are weakened (e.g. TFIIB recruiting trn factor  $\Rightarrow$  promoter would lack a BIE)
- P-TEFb : facilitates release of a paused RNA Pol
- chromatin modifying factors

## - mapping trn activation domains



- process is analogous to mapping DNA-binding motifs

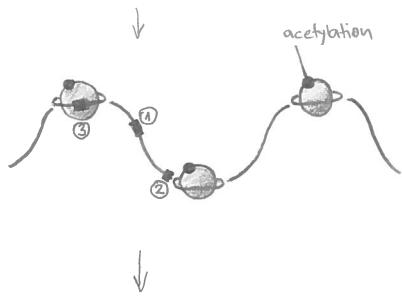
- design truncation constructs of the DNA coding for known heterologous binding domain and some part of activator protein
- expose proteins expressed from these constructs to DNA (as described above) and assay for activity of the protein coded for by the gene

- activation domains do not fall into categories like DNA-binding motifs do, because they interact with variety of other proteins (as opposed to interaction w/ DNA only in DNA-binding domains)

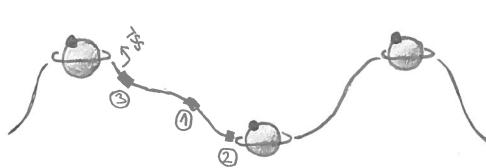
## - typical series of activation events



binding sites  
 ①: linker region - trn factor can bind and recruit histone acetyltransferase (HAT), that will acetylate nearby histones  
 ②: boundary b/w linker region and nucleosomal DNA  
 ③: nucleosomal DNA



- acetylation
- after opening of chromatin due to acetylation, ② can be bound by trn factor that recruits nucleosome remodelling complex
  - nucleosome blocking binding site ③ will be moved to expose this binding site
    - this can be done w/o spending any extra energy, because it is thermodynamically favourable (histone energy state is the same regardless of its position and trn factor binding DNA is in lower energy state)



after moving the nucleosome, trn factor can bind site ③ and initiate trn (e.g. recruit RNA Pol II / Mediator, recruit GTF, interact w/ P-TEFb, etc.)

## - examples of repressive trn regulation (relatively rare)

### - competition

- binding of repressor interferes w/ DNA-binding event that is necessary or at least important for high levels of trn
- e.g. sharing (part of) DNA binding site of repressor and GTF (or some other trn factor)
  - repressor interfering w/ function of GTF, not allowing it to perform its function, when repressor is bound to nearby DNA
  - recruiting chromatin modifier that represses trn (e.g. histone deacetylase)
  - blocking activation domain (by binding directly to it instead of DNA) under certain conditions

e.g. Gal80 binds Gal4 trn factor in presence of glucose, but not in presence of galactose

e.g. Gal trn is suppressed by Mig1 that recruits Tup1, which recruits deacetylase - all in presence of glucose, but not in presence of raffinose

## - Gene silencing

"bad neighbourhood"

- silenced region of genome turns off any txm unit that may be present, i.e. NOT gene-specific (all genes present will be silenced)
- works by chromatin modulation
- silenced regions can spread over many kb
- heterochromatin (most of genes are suppressed by heterochromatin, but there are some that will be expressed)
  - gene poor
  - telomere proximal and centromere proximal
  - repetitive sequences
- control / inhibit transposition ← [this is our idea for reason of silencing]
- polycomb group of proteins create silenced regions along the chromosome
- example: telomeric/mating type silencing in yeast
  - requires three SIR (Silent Information Regulator) proteins (SIR 2,3,4)
  - SIR 3,4 proteins are recruited by seq-specific DNA-binding protein
    - [at telomere: RAPI
    - [at silent mating type loci: ORC, RAPI, ABFI (any two of them are sufficient)]
  - SIR 3,4 bind SIR2 (SIR2 is HDAC - Histone DeAcetylase)
  - SIR2 deacetylates the nearby nucleosomes
  - SIR 3,4 bind also deacetylated nucleosomes
    - creates inaccessible chromatin structure (SIR 3,4 are relatively big proteins and they "coat" chromatin)
    - SIR 3,4 continue to recruit SIR2, which continues to deacetylate nucleosomes → self propagating structure
    - what stops the propagation
      - HATs (Histone Acetyl Transferases) compete w/ SIR2
        - SAS2 (Something About Silencing) ⇒ HAT that counteracts SIR2 HDAC activity
      - H3K9 methylation performed by DOTI (HMT - Histone Methyl Transferase)
        - inhibits SIR2 nucleosome binding
  - strong txm
    - tRNAs flank the mating type loci that seem to create boundaries for gene silencing
      - only ~80nt long
    - tRNAs are transcribed by RNA Pol III
    - sequence-dependent mechanism (sequence of tRNA gene) → more distinct transition from silenced to active chromatin across a population of cells

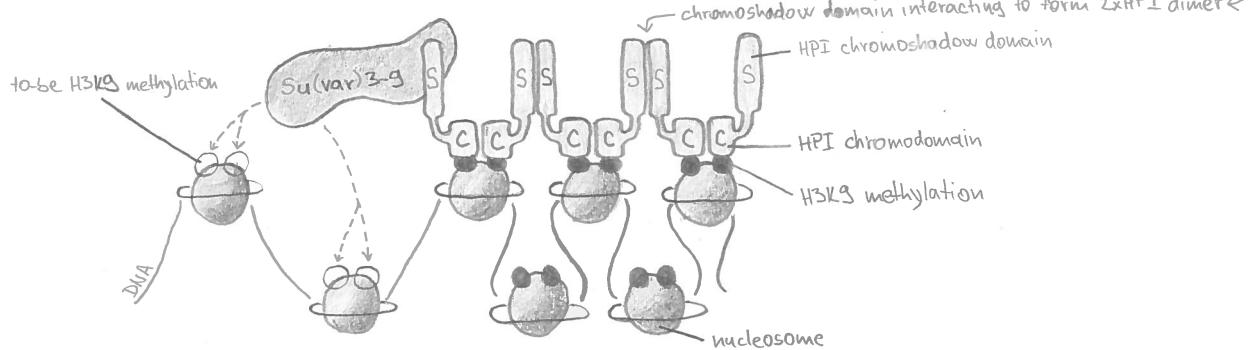
## - example: heterochromatin protein I (HPI)

- heterochromatin often has H3K9 methylation modification

- HPI has chromodomain that binds to H3K9 methylated sites

chromoshadow domain that dimerizes HPI protein (chromoshadow domains of two HPI proteins interact, forming 2x HPI dimer)
 

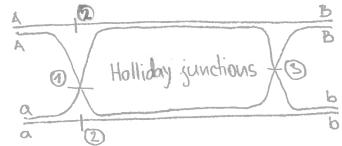
- this brings together nucleosomes HPIs bind to, facilitating formation of 30nm fiber
- recruits (Su(var)3-9) that propagates H3K9 methylation
- Suppressor of variegation



## GENETIC RECOMBINATION

### - Homologous recombination

- b/w any two similar regions of DNA (e.g. homologous chromosomes)
- see 7.28.1x: Double-stranded break repair (ds-breaks, restart of collapsed replication forks)



① + ③ → patch → (A/A, B/B), (a/a, b/b)  
 ② + ③ → crossover → (a/a, B/B), (A/A, b/b)

...

- meiosis (see 7.00x, pg. 11)

- homologous recombination is programmed part of meiotic process

- frequency of homologous recombination b/w two genes depends on distance b/w those two genes (see 7.00x, pg. 13)
  - genetic linkage maps (constructed WITHOUT sequencing DNA): relatively very accurate
  - physical DNA maps (constructed WITH sequencing DNA)

### - Conservative site-specific recombination

DNA transposition ← primary focus of this lecture

## TRANSPOSSABLE ELEMENTS

- move by very different mechanism than homologous recombination

- can insert DNA into multiple places in chromosome

- can disrupt genes  
 alter gene expression

linkage maps

- ubiquitous in nature, highly active in microbes, plants, animals (including humans), ...

- multiple mechanisms - can excise from their chromosomal location (causing breaks in DNA)

↓ DNA rearrangements

can induce DNA damage response

↓  
 perceived by cell as a problem

- used in experimental genetics as tools for - walking mutations  
 mapping

to integrate their genome into the genome of the host cell being infected by the retrovirus

- one of the pathways for DNA transposition is essentially identical to pathway used by retroviruses
- progenitor of the type of recombination that helps our immune systems to have such diversity

- V(D)J recombination (evolved from transposable element)

- both pathogens and immune system depends on this class of transposition to gain their diversity

e.g. anti-biotics resistance

- cause of some genetic diseases (especially pronounced in males & X-linked diseases)

- examples - hemophilias (blood clotting factors VIII, IX invaded and disrupted by transposable elements)

muscular dystrophies, cardiomyopathies (transposition events at large dystrophin locus leading to disease)  
 cancers (transposition in somatic cells)

- can have also positive effect: documented case of a woman with a genetic disorder that have been cured by a random transposition event

- Barbara McClintock 8 maize genetics: controlling elements and DNA breaks are not always in the same place



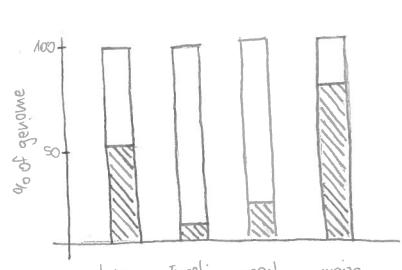
- examples - corn kernels coloring  
 coloring of bacterial colonies

transposition occurred early in the colony development

plants (e.g. flower coloring)



high rate of transposition



█ = repeats: transposon-like

not necessarily active transposons, but evolved from transposons

higher eukaryotes (e.g. humans): these are degenerate damaged no longer functioning remnants of what once have been transposons

it is NOT healthy for an organism to have lots of transposition going on

# DNA TRANSPOSITION

## - consequences

- insert & disrupt gene/regulation
- ds-breaks  $\Rightarrow$  DNA damage response
- horizontal gene transfer
- tools in experimental genetics
- evolution

## - classes of transposable elements

### - classical DNA elements

#### - replicative

- examples: Ac/Ds (observed by B. McClintock in corn), P-elements (in *drosophila*), Tn family (anti-biotic resistance)

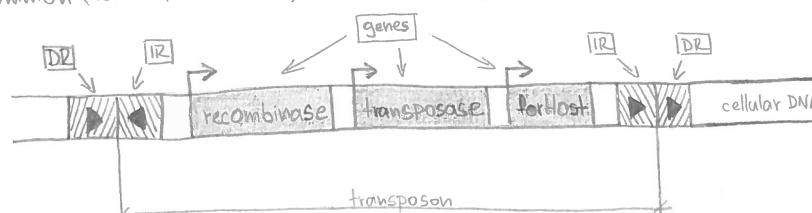
### - LTR (Long Terminal Repeat) retrotransposons

#### - viral-like elements

### - non LTR retrotransposons (polyA retrotransposons)

- some of the most active elements in higher eukaryotes

## - common (for all three classes) core of the biochemical mechanism



the order of genes does NOT matter

- recombinase: critical - carries out basic recombination reaction

- transposase: usually present, sometimes replaced by couple of genes that contribute to basic recombination reaction

- forHost: gene(s) used by host organism

- repeats - IR: inverted repeats  
DR: direct repeats

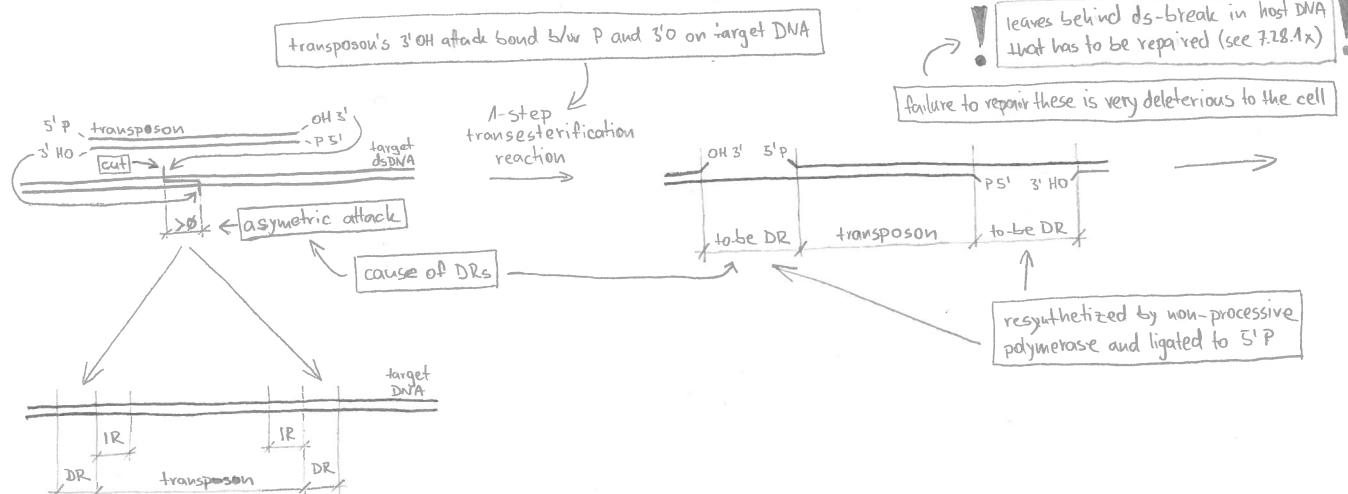
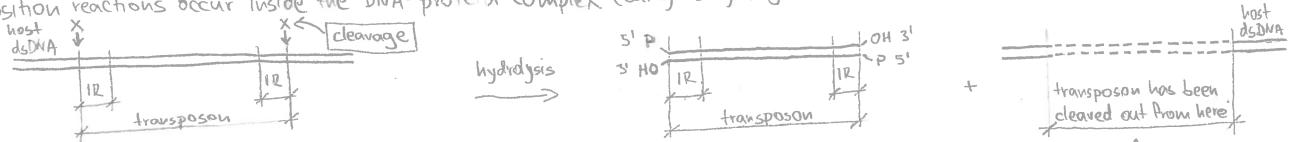
# TRANPOSITION MECHANISMS

## Cut and Paste

- inverted repeat sequences encode binding sites for transposase (specialized recombinase)

- transposase binds the two ends of the transposone and brings them together in very tight DNA-protein complex

- transposition reactions occur inside the DNA-protein complex (catalyses hydrolysis and 1-step transesterification reaction - see below)

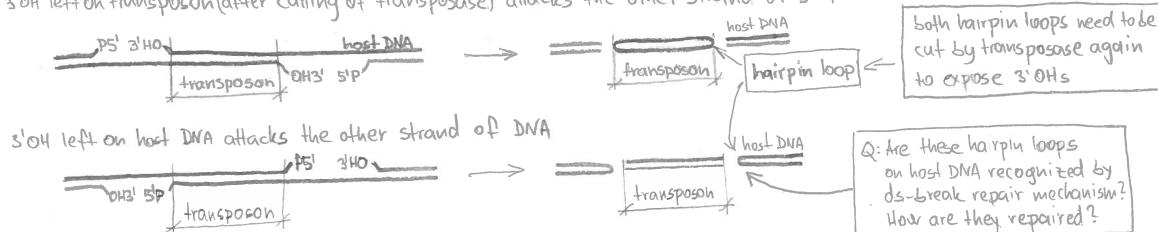


## transposon excision mechanisms

sequential cutting: two DNA strands are cut by two metabolising enzymes

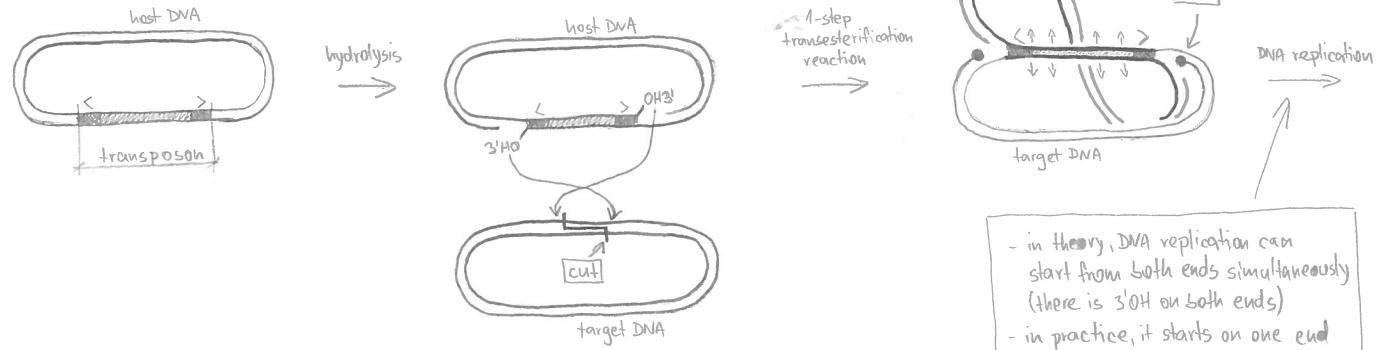
- transposon encodes for (part of recombinase and "forHost" genes)  $tnsA$ : transposase (leaves 3'OH on transposon after cutting)  
 $tnsB$ : cleaves the other strand (leaves 5'P on transposon after cutting)

transesterification 3'OH left on transposon (after cutting of transposase) attacks the other strand of DNA

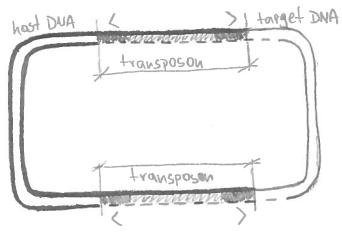
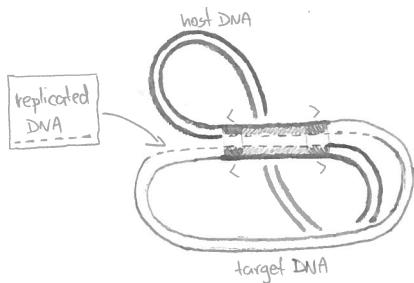


## Relicative

- similar to cut and paste, uses different substrate



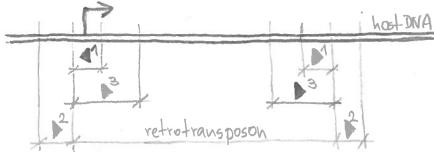
- in theory, DNA replication can start from both ends simultaneously (there is 3'OH on both ends)
- in practice, it starts on one end and repairs whole gap



- possible transposons:

this mechanism allows transposons to get more and more complex

## Retrotransposition



- 1: inverted repeats } present in all transposons (see 10.2: IR, DR)
- 2: direct repeats }
- 3: direct repeats (include 1, 2)

therefore never become retroviruses

- genomes of many organisms contain these structures that look a lot like retroviruses, but never escape the organism
- mechanism



txn event near the end of the element that makes almost full-length RNA copy of the element genetic material

reverse transcription



Synthesis of new retrotransposon from the RNA (LTRs allow for generation of full-length cDNA from partial-length RNA)

attack new target site (retrotransposon)

viral packaging (retrovirus)

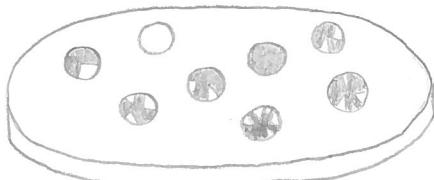
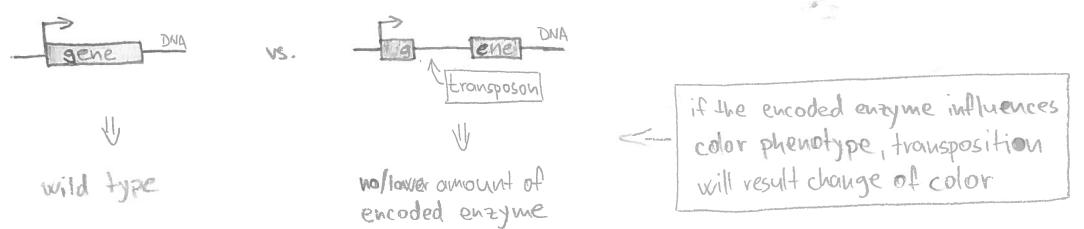
integrase (transposase) cleaves off a couple of nucleotides from each 3' end to reveal new 3' OH ends

asymmetric attack of target DNA: see 10.2: cut and paste mechanism

P 5' overhangs are replaced by newly synthesised DNA (repair of gaps after attacking target DNA)

## SECTORING ASSAYS

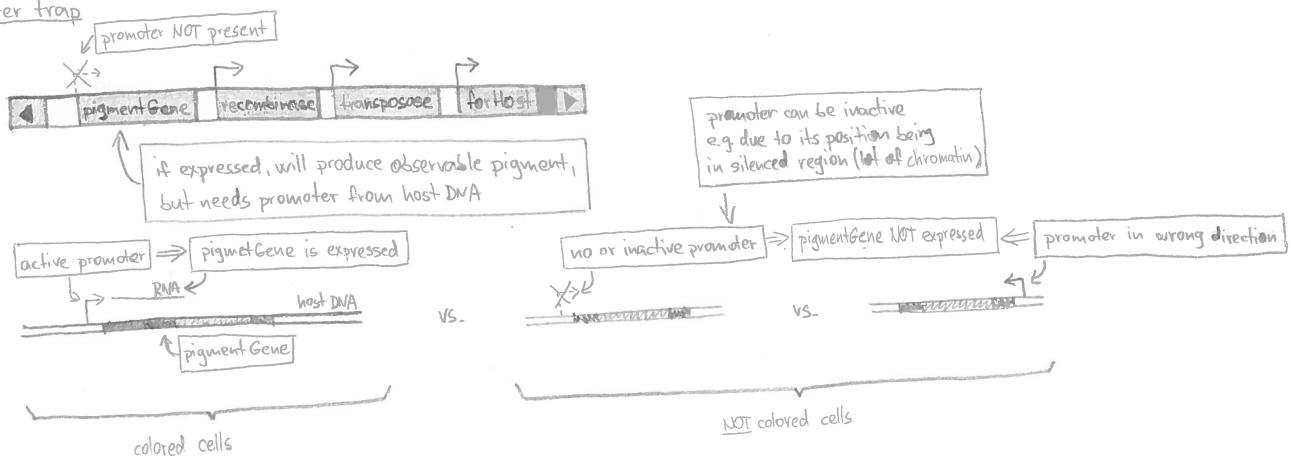
- based on change of color (observable phenotype) after transposition (sectoring)



frequency of transposition event: vs.

timing of transposition event: vs.

- Promoter trap



- can assay frequency, but underestimates total number of transposition events

- very sensitive (change of color, e.g. from white to red, is very apparent):

even b/w different species

## CONJUGATION ASSAY (a.k.a. Mating-out assay)

- in bacteria: takes advantage of bacteria carrying special type of plasmid that can move b/w individual bacterial cells  $\leftarrow$  conjugation

- donor cell: chromosome (does not contain any transposons, no markers)

- donor plasmid: contains transposon + markers



T: tetracycline resistance gene  
K: kanamycin resistance gene } markers

conjugal plasmid (F-plasmid): conjugal plasmid containing genes required to move DNA b/w two types of cells + marker



A: ampicillin resistance gene } marker

- recipient cell: initially does not have any of the mobile elements (D,F), contains only chromosome

- process: mix donor and recipient cells, collect recipient cells and determine which mobile elements have been acquired by growing them on selective media:

DONOR	RECIPIENT	RESISTANCE			NOTE
		A	T	K	
		+	-	-	only F-plasmid moved to the recipient cell (no transposition)
		+	+	-	transposon copy-pasted onto F-plasmid and moved to the recipient cell
X		+	+	+	cointegrate moved to the recipient cell

example: two different transposons, markers exactly the same

TRANSPOSON	-	A	A,T	A,K	A,T,K
I	1000	303	102	1	2
II	1000	300	60	62	61

number of colonies

A-plate: base rate of conjugation !!!!!!!

I:  $\sim 300$  on AT-plate, but essentially none on A,K and A,T,K

- frequency:  $100(\text{AT})/303(\text{A}) \approx 30\%$

- mechanism: K-plates:  $\Rightarrow$  copy-paste

II:  $\sim 300$  on AT-plate,  $\sim 60$  on all other plates

- frequency:  $60(\text{A,T/A,K/A,T,K})/300(\text{A}) \approx 20\%$

- mechanism: AT  $\approx$  AK  $\Rightarrow$  replicative transposition

## SLEEPING BEAUTY

- eukaryotic cells

	-	+	+
forHost	-	+	+
Transposase	-	-	+
forHost gene expression	-	-	++++

- performed by growing cells on restrictive media (e.g. forHost = antibiotics resistance gene & using media containing that antibiotics), i.e. observing phenotype coded for by the forHost gene

- used to infer origins of transposon-like repeats in (higher) eukaryotes

the original transposon the sequence originated from

## GENOMIC SOUTHERN BLOT

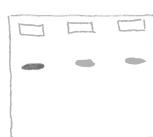
- allows to probe whole chromosome for specific sequences

{ parent cell: one transposon at position A  
daughter cell: one transposon at position A (same as parent), another copy of that transposon at position B (due to replicative transposition)

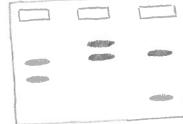
- assay identifies movement of sequences through genome

- process: isolate DNA from cells of interest

- cut isolated DNA into medium-sized fragments (must not cut sequence of interest)
  - DNA in individual samples has to be always cut at the same places (use e.g. sequence-specific endonuclease)
- separate mid-sized DNA fragments on gel
- transfer from gel to nitrocellulose membrane
- use labelled DNA probes to identify sequences of interest
- labelled DNA fragment will anneal to complementary sequences on DNA attached to nitrocellulose



sequence of interest  
NOT moving (it is  
at the same place  
in all samples)



replicative  
transposition



copy-paste  
transposition

, blot involves the transfer of — from a gel to a membrane followed by probing with a —.

- western, proteins, specific antibody
- southern, DNA, labelled oligonucleotide
- northern, RNA, labelled oligonucleotide

## INSIDE-OUT PCR

- identifies which gene the transposon resides in (which gene has been disrupted by transposition)

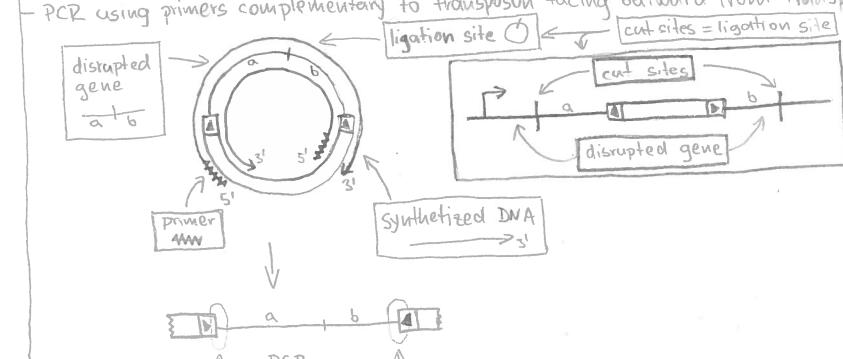
- process: isolate DNA from cells of interest

- cut isolated DNA into fragments (must not cut transposon of interest)

- ligate fragments to form plasmids

- low DNA concentration conditions (do not want to ligate two individual fragments to one another)

- PCR using primers complementary to transposon facing outward from transposon



sequence and analyse junctions b/w transposon sequence (known) and gene (reference genome must be available)

## SUMMARY: ASSAYS FOR TRANSPOSITION

	rate of transposition	mechanism of transposition	sequence of target genetic locus
promoter trap (sectoring assay)	✓	✗	✗
sleeping beauty	✓	✗	✗
mating out (conjugation assay)	✓	✓	✗
genomic Southern blot	✓	✓	✗
inside-out PCR	✗	✗	✓