The Basics of Transcription - I

Methods used to study transcription: in vitro assays

Transcription proceeds in several steps

DNA-dependent RNA polymerases

General Transcription Factors: GTFs

RG Clerc  February 26, 2014
In Vivo Transcription Assays

**Promoter of interest**
- Reporter gene (e.g., luciferase or CAT)
  - For promoter analysis

**Constitutive promoter** (e.g., SV40, HSV-TK)
- Reporter gene
  - For distant control region analysis

**Transfect cultured cells** with reporter plasmid (Include normalization plasmid, if desired [see text]).

**Incubation for 24–72 hr.**
- Transcription from episomal plasmid and translation of protein.

**Harvest cells. Prepare protein extract or mRNA.**

**Measure enzymatic activity of reporter gene product.**

**Measure reporter mRNA levels.**

**FIGURE 5.1.** Transient transfection assay.
**In Vitro Transcription Assays**

- Purified DNA templates (plasmids) are transcribed in the test tube with nuclear extracts, highly purified or recombinant transcription factors.

- Criteria: **accuracy** and **quantity** of transcripts.

- *The ultimate goal is the reconstitution of regulated gene expression from recombinant (i.e. completely defined) components.*
In Vitro Transcription Assays

- All methods require some knowledge about the DNA fragment containing the promoter of interest
  - where is the natural start site (+1)?
• Many transcripts will be accurately initiated on the promoter of interest, but there will be a degree of non-specific background transcription
  – spurious promoters; gaps/nicks in template DNA

• Different methods are available for quantitating *in vitro* transcripts:
  – primer extension assay
  – “run-off” assay
  – „G-less“ cassettes
  – S1 nuclease protection assay
**Methods:** primer extension, "run off" transcripts and G-less cassettes

<table>
<thead>
<tr>
<th>Primer extension reaction</th>
<th>Runoff transcription</th>
<th>G-less transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Promoter</strong></td>
<td><strong>Promoter</strong></td>
<td><strong>Promoter</strong></td>
</tr>
<tr>
<td>Transcription with ATP, CTP, GTP, UTP</td>
<td>Linearize</td>
<td>Transcription with ATP, CTP, $^{32}$P]UTP (add O-Me-GTP in crude extracts)</td>
</tr>
<tr>
<td>5$^{\prime}$--$\cdots$--3$^{\prime}$ mRNA</td>
<td>5$^{\prime}$--u$^{\star}$--u$^{\star}$--u$^{\star}$--3$^{\prime}$ mRNA</td>
<td>5$^{\prime}$--u$^{\star}$--u$^{\star}$--u$^{\star}$--3$^{\prime}$ mRNA</td>
</tr>
<tr>
<td>5$^{\prime}$--$\cdots$--3$^{\prime}$ 3$^{\prime}$--$\cdots$--5$^{\prime}$</td>
<td>3$^{\prime}$--$\cdots$--5$^{\prime}$</td>
<td>3$^{\prime}$--$\cdots$--5$^{\prime}$</td>
</tr>
<tr>
<td><strong>G+C+T</strong> Extension product</td>
<td><strong>G+C+T</strong> mRNA</td>
<td><strong>G+C+T</strong> mRNA</td>
</tr>
<tr>
<td><strong>Primer</strong></td>
<td>Analyze radio-labeled RNA on sequencing gel.</td>
<td>Analyze radio-labeled RNA on sequencing gel.</td>
</tr>
</tbody>
</table>

**FIGURE 14.1.** Primer extension, runoff, and G-less transcription.
Methods: S1 Nuclease Protection Assay

• Single strand-specific nuclease enzymes will not digest double-stranded nucleic acid substrates
  – this includes RNA:RNA or RNA:DNA hybrids
  – S1 endonuclease

• mRNA is hybridised to an uniformly *radiolabelled* single-stranded DNA fragment containing the start site; the probe can also be end-labelled

• mRNA hybridised to the DNA-fragment will protect it right up to the start sequence from degradation
Methods: S1 Nuclease Protection Assay

1. *In vitro* Transcription

2. Prepare radiolabeled probe

3. Hybridize radiolabeled probe with RNA

4. Digest with nuclease S1
The Players in Transcription Regulation

- DNA-binding transcription factors (upstream factors)
- Chromatin regulators
- Coactivators and corepressors: Mediator, etc..
- Basal Machinery: RNA PolII, GTFs
Basal vs Activated Transcription: an *in vitro* concept

**Basal machinery**

- RNAPII
- TFIIA
- TFIIB
- TFIID (TBP) ca. 14
- TFIIE
- TFIIF
- TFIH

**upstream (regulatory, enhancer)**

- Transcription activators
  - NF-kB
  - Fos
  - Pu.1
  - Oct-1
  - CBFA-1
  - etc...

**core**

- TATA
- INR

**Cofactors, Mediator**

- RNAPII 12
- TFIIA 3
- TFIIB 1
- TFIID (TBP) ca. 14
- TFIIE 2
- TFIIF 2
- TFIH 10
Transcription Cycle: Multiple Steps, Multiple Opportunities for Regulation

INITIATION

1. Polymerase binds to promoter sequence in duplex DNA. "Closed complex"

2. Polymerase melts duplex DNA near transcription start site, forming a transcription bubble. "Open complex"

3. Polymerase catalyzes phosphodiester linkage of two initial rNTPs.

ELONGATION

4. Polymerase advances 3' → 5' down template strand, melting duplex DNA and adding rNTPs to growing RNA.

TERMINATION

5. At transcription stop site, polymerase releases completed RNA and dissociates from DNA.
Transcription Cycle: Status of Promoter-Proximal Pol II: Paused, Stalled, Poised, etc?

**Paused:** elongation complexes halted temporarily

**Stalled:** elongation complexes that have stopped RNA synthesis

**Poised:** RNA polymerase detected near a gene promoter (ChIP assay terminology)

**Backtracked:** halted elongation complexes that move backwards

** Arrested:** complexes cannot restart transcription without additional factors
Transcription Initiation Steps

- Promoter (Start Site) Recognition
- Promoter Binding
- Promoter Melting
- Transcript Initiation
- Promoter Escape/Clearance
- Transcript Elongation

Increasing Commitment
Sequential assembly of Initiation Complex

Stepwise recruitment of general transcription factors or ...

PIC: preinitiation complex (DNA, RNAPII, TBP, TFIIA, TFIIB, TFIIF, TFIIE, TFIIH)

Figure II: Schematic Model of the Molecular Structure of the Complexes

Buratowski S and Sharp PA. Cell 56:549
Recruitment of preassembled RNAPII holoenzyme and general transcription factors "two component pathway"
alpha-Amanitin, a toxin from the mushroom Amanita phalloides, is a potent inhibitor of DNA-dependent RNA polymerase II
Eukaryotic RNA polymerases: Contacts between Alpha-amanitin and RNA Pol II as of Today

alpha-Amanitin chemical formula and stick model shown in orange, RNA Pol II residues and as ribbon model in grey

RNA Polymerase I Initiation
RNA Polymerase I : Purification and Crystalization

RNA Polymerase I

RNA Polymerase II

- All eukaryotic RNAPIIIs characterized so far contain 12 different subunits (involved in transcription initiation, elongation and termination)
  - RPB1 to RPB12

- RNAPIIIs are highly conserved across eukaryotic species
  - 53% overall identity between yeast and human RNAPII

- The two largest subunits, RPB1 and RPB2, contain the catalytic center

- Four subunits (RPB3, RPB10, RPB11 and RPB12) serve as an assembly scaffold for RPB1/RPB2

- Post-transcriptional mechanisms also depend on RNAPII
  - capping
  - splicing
  - 3’ end formation
Structural Conservation: 53% overall Identity between Yeast and Human all across distributed Residues *Identical* Between Yeast and Human RNAPII

Residues *Identical* Between Yeast and *E. coli*
Structural Conservation: Homology

Sequence Homology Between Yeast RNAPII and Bacterial RNAPs

Structural Homology Between Yeast RNAPII and Bacterial RNAPs
Differences between Pol I and Pol II
The Carboxy Terminal Domain: CTD

repeating heptapeptide sequences: 52 repeats human, 26 in yeast

Tyr-Ser-Pro-Thr-Ser-Pro-Ser

target for phosphorylation
RNA Pol II O = phosphorylated
RNA Pol II A = unphosphorylated form

RNA Pol II is phosphorylated at Ser5 to initiate elongation
RNA Pol II is phosphorylated at Ser2 after bp +50 during elongation
Phosphorylation is removed for termination
Proportions of CTD Relative to RNAPII

YSPTSPS\textsubscript{n}

Yeast: n=26
Mouse: n=52

Science 292, 1863-1876
RNA Polymerase III: Rapid Repression by Maf1 ensures cell survival during stress

Pol III specific subunits on Pol II structure (C53, C34, C82)

ribbon diagram of Maf1 crystal

Prokaryotic vs eukaryotic RNA Polymerases

• In the presence of $\sigma$ factors, bacterial RNA polymerases (RNAPs) can recognize promoters without the help of any other transcription factors.

• Although eukaryotic RNAPs are structurally much more complex, they do not have any obvious sequence-specific binding activities by themselves. Promoter recognition requires large number of additional factors (GTFs).
(a) Bacterial RNA polymerase

(b) Yeast RNA polymerase II
### Table 1  RNA polymerase subunits

<table>
<thead>
<tr>
<th>RNA polymerase</th>
<th>Pol I</th>
<th>Pol II</th>
<th>Pol III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten-subunit core</td>
<td>A190</td>
<td>Rpb1</td>
<td>C160</td>
</tr>
<tr>
<td></td>
<td>A135</td>
<td>Rpb2</td>
<td>C128</td>
</tr>
<tr>
<td></td>
<td>AC40</td>
<td>Rpb3</td>
<td>AC40</td>
</tr>
<tr>
<td></td>
<td>AC19</td>
<td>Rpb11</td>
<td>AC19</td>
</tr>
<tr>
<td></td>
<td>A12.2</td>
<td>Rpb9</td>
<td>C11</td>
</tr>
<tr>
<td></td>
<td>Rpb5 (ABC27)</td>
<td>Rpb5</td>
<td>Rpb5</td>
</tr>
<tr>
<td></td>
<td>Rpb6 (ABC23)</td>
<td>Rpb6</td>
<td>Rpb6</td>
</tr>
<tr>
<td></td>
<td>Rpb8 (ABC14.5)</td>
<td>Rpb8</td>
<td>Rpb8</td>
</tr>
<tr>
<td></td>
<td>Rpb10 (ABC10α)</td>
<td>Rpb10</td>
<td>Rpb10</td>
</tr>
<tr>
<td></td>
<td>Rpb12 (ABC10β)</td>
<td>Rpb12</td>
<td>Rpb12</td>
</tr>
<tr>
<td>Rpb4/7 subcomplex</td>
<td>A14</td>
<td>Rpb4</td>
<td>C17</td>
</tr>
<tr>
<td></td>
<td>A43</td>
<td>Rpb7</td>
<td>C25</td>
</tr>
<tr>
<td>TFIIF-like subcomplex&lt;sup&gt;a&lt;/sup&gt;</td>
<td>A49</td>
<td>(Tfg1/Rap74)</td>
<td>C37</td>
</tr>
<tr>
<td></td>
<td>A34.5</td>
<td>(Tfg2/Rap30)</td>
<td>C33</td>
</tr>
<tr>
<td>Pol III-specific subcomplex</td>
<td>-</td>
<td>-</td>
<td>C82</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>C34</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>C31</td>
</tr>
<tr>
<td>Number of subunits</td>
<td>14</td>
<td>12</td>
<td>17</td>
</tr>
</tbody>
</table>

<sup>a</sup>The two subunits in Pol I and Pol III are predicted to form heterodimers that resemble part of the Pol II initiation/elongation factor TFIIF, which is composed of subunits Tfg1, Tfg2, and Tfg3 in <i>Saccharomyces cerevisiae</i>, and of subunits Rap74 and Rap30 in human.

---

**Figure 20.2** Eukaryotic RNA polymerase II has >10 subunits.

<table>
<thead>
<tr>
<th>Size</th>
<th>Stoichiometry</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>1</td>
<td>Related to bacterial subunit binds DNA, has CTD = (YSPTSPS)&lt;sub&gt;n&lt;/sub&gt;, yeast n = 26; mouse n = 52</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>Related to bacterial subunit binds nucleotides</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>Related to bacterial subunit</td>
</tr>
<tr>
<td>25</td>
<td>&lt;1</td>
<td>Common to all 3 polymerases</td>
</tr>
</tbody>
</table>

---

E. coli RNA polymerase

# Table 10-1. Sigma Factors of *E. coli*

<table>
<thead>
<tr>
<th>Sigma Factor</th>
<th>Promoters Recognized</th>
<th>Promoter Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^{70}$</td>
<td>Most genes</td>
<td>TTGACAT</td>
</tr>
<tr>
<td>$\sigma^{32}$</td>
<td>Genes induced by heat shock</td>
<td>TCTCNCCCTTGAA</td>
</tr>
<tr>
<td>$\sigma^{28}$</td>
<td>Genes for motility and chemotaxis</td>
<td>CTAAA</td>
</tr>
<tr>
<td>$\sigma^{38}$</td>
<td>Genes for stationary phase and stress response</td>
<td>?</td>
</tr>
<tr>
<td>$\sigma^{54}$</td>
<td>Genes for nitrogen metabolism and other functions</td>
<td>-24 Region</td>
</tr>
</tbody>
</table>
$\sigma^{54}$ is dedicated for long-range activation
Table 1. RNA polymerase II transcription machines. Mass data are for yeast proteins. Details about function are in the text. TAF, TBP-associated factor.

<table>
<thead>
<tr>
<th>Component</th>
<th>Subunits</th>
<th>Mass (kD)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol II</td>
<td>12</td>
<td>520</td>
<td>RNA synthesis</td>
</tr>
<tr>
<td>TFIIB</td>
<td>1</td>
<td>38</td>
<td>Start site determination</td>
</tr>
<tr>
<td>TFIID (TBP)</td>
<td>1</td>
<td>27</td>
<td>Bending TATA box DNA around TFIIB and pol II</td>
</tr>
<tr>
<td>(TAFs)</td>
<td>14</td>
<td>749</td>
<td>Promoter recognition</td>
</tr>
<tr>
<td>TFIIE</td>
<td>2</td>
<td>92</td>
<td>Coupling pol II–promoter interaction to recruitment of TFIIH</td>
</tr>
<tr>
<td>TFIIF</td>
<td>3</td>
<td>156</td>
<td>Interaction with nontemplate DNA strand</td>
</tr>
<tr>
<td>TFIIFH</td>
<td>9</td>
<td>525</td>
<td>Promoter opening, pol II phosphorylation</td>
</tr>
<tr>
<td>Mediator</td>
<td>20</td>
<td>1003</td>
<td>Regulatory signal transduction</td>
</tr>
</tbody>
</table>
The core Pol II promoter in more detail

BRE  TATA box  Initiator  DPE
G/CWG/CWG/CGGCGC  TATAAA  PyPyANT/APyPy  GA/TCGTG
-32  -26  +1  +31

TFIIB  TFIID  TFIID  TFIID
TFIIA  TBP  TAFII250  TAFII60
       TAFII150  TAFII40
TRANSCRIPTION INITIATION, PROMOTER CLEARANCE AND ELONGATION
Purification of General Transcription Factors (GTFs)

Nuclear Extract

- **P-11**
  - 0.1 M
    - DE-52
      - 0.1 M
        - DE-52
          - TFIIA
      - 0.3 M
  - 0.6 M
    - DE-52
      - 0.15 M
        - DE-52
          - USA
      - 0.3 M
        - DE-52
          - TFIIB/TFIIE/TFIIF/TFIIH/Pol
  - 0.85 M
    - DE-52
      - 0.1 M
        - USA
      - 0.25 M
        - DE-52
          - TFIID
      - 0.5 M
<table>
<thead>
<tr>
<th>Factor</th>
<th>Number of Subunits</th>
<th>Mw. (kD)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFIID - TBP</td>
<td>1</td>
<td>38</td>
<td>Recognize core promoter; Recruit TFIIB</td>
</tr>
<tr>
<td>TFIID - TAFs</td>
<td>12</td>
<td>15 - 250</td>
<td>Assist transcription activation; Assist promoter recognition</td>
</tr>
<tr>
<td>TFIIA</td>
<td>3</td>
<td>12, 19, 35</td>
<td>Stabilize TFIID and promoter binding</td>
</tr>
<tr>
<td>TFIIIB</td>
<td>1</td>
<td>35</td>
<td>Recruit RNA Pol II and TFIIIF</td>
</tr>
<tr>
<td>TFIIIF</td>
<td>2</td>
<td>30, 74</td>
<td>Assist RNA Pol II to reach promoter</td>
</tr>
<tr>
<td>TFIIIE</td>
<td>2</td>
<td>34, 57</td>
<td>Recruit TFIIH; Modulate TFIIH helicase, ATPase and kinase activities</td>
</tr>
<tr>
<td>TFIIIH</td>
<td>9</td>
<td>89, 80, 62, 52, 44</td>
<td>Promoter melting using helicase activity; DNA repair</td>
</tr>
</tbody>
</table>
TFIID

• TFIID is a multiprotein complex
  – (~10-14 different subunits)

• The subunit responsible for recognizing the TATA box is called ‘TATA-Binding Protein’ (‘TBP’)

• TBP can specifically bind to TATA-boxes on its own
Evolutionary Conservation of TBP C-Terminal Domain

- **Human**: 100%
- **Mouse**: 88%
- **Drosophila**: 83%
- **Arabidopsis**: 82%
- **Potato**: 80%
- **S. pombe**: 81%
- **S. cerevisiae**: 84%
- **Acanthamoeba**: 75%
- **Dictyostelium**: 38%
- **Plasmodium**
The TATA-Box is Specifically Recognized by TFIID
The TATA-Binding Protein: a Molecular Saddle
TATA-Binding Protein: “molecular saddle”
TFIIA and TFIIIB Stabilize TFIIID Binding
TBP Binding is Stabilized by TFIIA and TFIIB
EM Structure of TFIID-IIA-IIB Complex

TFIIH

• TFIIH is a multiprotein complex required for both transcription and DNA repair
• Human TFIIH is composed of nine polypeptides with a total molecular weight of 460 kDa
• TFIIH is responsible for three critical functions in transcription
  – phosphorylation of the COOH-terminal domain (CTD) of the RPB1 subunit of RNA polymerase II (RNAPII)
  – promoter melting (DNA helicase activity for promoter opening)
  – promoter clearance (DNA dependent ATPase activity for transcription initiation)

• Two biochemically separable complexes:
  – core TFIIH
  – Cdk Activating Kinase (CAK) complex
TFIIH: Core and CAK complexes

- Core TFIIH contains two ATP-dependent DNA helicases with opposite polarity, p89XPB and p80XPD
  - plus p62, p52, p44, p40/CDK7, p38/CyclinH, p34, p32/MAT1 and p8 polypeptides

- Mutations in human XPB and XPD genes can give rise to three genetic disorders: xeroderma pigmentosum, Cockayne’s syndrome, and trichothiodystrophy

- Phenotypes of these disorders can be explained by deficiencies in both transcription and DNA repair

- The CAK subcomplex is composed of the cdk7 kinase, the cyclin H, and MAT1, which stabilizes the interaction between the two other subunits
XP Individuals homozgyous for the mutation are unable to reverse T^T dimerization induced by sunlight. XP is characterized by extremely dry skin (xeroderma) and numerous malignant pinpoint tumors that are induced by brief exposure to sunlight.
TFIIE

• The general transcription factor TFIIE plays important roles at two distinct but sequential steps
  – open complex formation
  – transition from initiation to elongation.

• The large subunit of human TFIIE binds to and facilitates the enzymatic functions of TFIIH, but TFIIE also functions independently from TFIIH

• Human TFIIE consists of two subunits, TFIIE\(\alpha\) (57 kDa) and TFIIE\(\beta\) (34 kDa)

• The subunits combine into a \(\alpha_2\beta_2\) heterotetramer with a combined molecular mass of 180 kDa

• TFIIE plays essential roles in the regulation and recruitment of TFIIH activities, involved in promoter clearance
  – CTD kinase, ATPase and DNA helicase activities are positively regulated
TFIIF

- TFIIF binds to RNAPII in solution and facilitates delivery of the polymerase to the TFIID-TFIIB-DNA complex on the promoter

- Human TFIIF contains two separate subunits
  - RAP74 (517 AA) interacts with RAP30, TAF1, TFIIB and RNAPII
  - RAP30 (249 AA) DBD from the winged HTH family

- TFIIF is required for entry of TFIIE and TFIIH into the pre-initiation complex
Recognize core promoter

Targets Pol II to promoter

Modulates helicase

CTD protein kinase

CTD of large subunit of Pol II