Published in final edited form as:

Methods. 2019 April 15; 159-160: 82–89. doi:10.1016/j.ymeth.2019.03.016.

In vitro reconstitution of yeast RNA polymerase II transcription initiation with high efficiency

Rina Fujiwara¹ and Kenji Murakami^{1,*}

¹Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

Abstract

Transcription initiation can be reconstituted from highly purified general transcription factors (GTFs), RNA polymerase II (pol II), and promoter DNA. However, earlier biochemical reconstitution systems had a serious technical limitation, namely very poor initiation efficiency. Due to the poor efficiency of the reaction and trace amounts of proteins involved in the pre-initiation complex (PIC) assembly, detection of transcription and PIC formation was only possible by the synthesis of a radiolabeled transcript and by immunoblotting for PIC components on templates. Here we describe a transcription system that is capable of initiating transcription with >90% efficiency using homogeneous, active yeast components including TFIIA, TFIIB, TBP, TFIIE, TFIIIF, TFIIH, Sub1, and pol II. The abundant specifically assembled PICs on promoter DNA can be separated from free general transcription factors (GTFs) and pol II by density gradient sedimentation, irrespective of the length of promoter DNA. The system is robust, and can be modified to accommodate many other transcription factors, and the resulting complexes can be analyzed by SDS-PAGE followed by Coomassie Blue staining. This technical advance now paves the way to conduct definitive biochemical and structural studies of the complete process of pol II initiation from the PIC, through promoter escape, and finally to productive elongation.

Keywords

transcription; initiation; RNA polymerase II; in vitro reconstitution; pre-initiation complex; yeast

1. Introduction

It has been well established that RNA polymerase II (pol II) and the general transcription factors (GTFs) are recruited to promoter DNA to form the pre-initiation complex (PIC) that is responsible for transcription initiation (1, 2). The requirement for GTFs (TFIIB, TBP (a subunit of TFIID), TFIIE, TFIIIF, and TFIIH) was determined in the early 1990s using a rat liver system, a yeast system, and subsequently other systems such as human (1, 3, 4). The five GTFs are conserved across all eukaryotes and are indispensable for initiation reaction (therefore defined as general transcription factors), whereas other transcription factors are dispensable at least in terms of PIC formation and transcription initiation using an *in vitro* reconstituted system with a linear promoter template.

*Correspondence: kenjim@pennmedicine.upenn.edu.

Although the formation of a PIC only requires pol II, the five GTFs, and promoter DNA, the resulting complex does not transcribe with high efficiency. We sought to reconstitute highly efficient transcription from purified yeast components and a linear promoter template. While transcription in yeast has been perceived as having some unique attributes such as transcription start site (TSS) scanning (5), the high level of conservation of its transcription machinery with other eukaryotes and the prior history of insights from *S. cerevisiae* transcription mechanisms makes the yeast system a powerful tool for deciphering details of eukaryotic transcription. As a first step, we developed methods to isolate pure, homogeneous GTFs without any nucleic acid contaminants (6, 7). With the highly purified GTFs in suitable form, we sought a way to assemble homogeneous active PIC, resulting in transcription efficiency per PIC of ~20% at best (8). Finally, by addition of ~4–6-fold molar excess pol II and TFIIF, the transcription system achieved >90% initiation efficiency (per template), and differs from traditional systems (including traditional yeast systems) in details (promoter DNA, protein purifications, PIC assembly). Here we describe methods for isolation of these yeast factors, reconstitution of highly efficient transcription, and isolation of the reconstituted PIC, and discuss applications to other systems, e.g., transition from initiation to elongation, transcription regulation including Mediator.

2. Oligonucleotides for promoter DNA

Two important aspects to consider when designing a promoter template with highly efficient transcription are 1) the length of promoter DNA fragment and 2) the use of strong promoter sequence. Promoter DNA is indispensable for transcription initiation, but naked non-specific DNA itself is a strong inhibitor of transcription initiation. Thus the length of promoter DNA can be optimized experimentally. In general, transcription activity is optimal with a promoter DNA of ~150 to 250 bp in length, starting ~20–30 bp upstream from the TATA box and extending at least ~50 bp downstream from TSS (Fig. 1), whereas transcription activity strongly decreases with DNA exceeding ~250 bp in length. Promoter templates can be prepared by large scale PCR or restriction digestion of a concatemeric form of plasmid DNA followed by purification using gel filtration (8). It is noteworthy that there is much less transcription activity with promoter DNA obtained by annealing equimolar amounts of complementary oligonucleotides, although extensively used for structural studies, likely due to contamination of single-stranded oligonucleotides.

The level of transcription *in vitro* is largely dependent on promoter sequences. Highly efficient *in vitro* transcription can be achieved by using a strong promoter template; for example, U2 snRNA promoter (*SNR20*) showed higher initiation efficiency than other typical strong promoters (e.g., *HIS4*) (8), offering a significant advantage in single-molecule and definitive biochemical studies (9, 10). Although sequences of the eight residues of the TATA-box (11) and ~10 residues at TSSs (12) are the most important factors that determine transcription efficiency, any mutations on the promoter template may have a positive or negative effect on transcription.

3. Protein purifications

Milligram quantities of all GTFs can be obtained in pure, homogeneous form (Table 1, Fig. 2). Two GTFs – TFIIB and TBP – are available from overexpression in bacteria in yields of 2–10 mg/Liter. The remaining GTFs – TFIIE, TFIIF, TFIIH, TFIK – are available from a protease-deficient strain of the yeast *S. cerevisiae* in yields of 1–3 mg from a 100L culture. In addition, we add recombinant forms of TFIIA and Sub1 to the reconstituted transcription system as they form a stable complex with the PIC (13) and enhance transcription by ~1.1 fold and ~1.4 fold, respectively (Fujiwara et. al, in preparation). Recombinant forms of TFIIE and TFIIF can be obtained by overexpression in bacteria, but recombinant TFIIE has ~2-fold less transcription activity than TFIIE isolated from *S. cerevisiae* (14). All factors are purified with > 90% purity without any detectable nucleic acid contamination. To avoid nucleic acid contaminants, which significantly reduce transcription activity, each GTF is extensively washed with high salt during affinity chromatography followed by an ion exchange column, a heparin column, or a gel filtration column. Each protein can be concentrated up to at least ~3 μ M (for the purpose of transcription assays), aliquoted, frozen in liquid nitrogen, and stored at –80°C without loss of transcription activity. All the *S. cerevisiae* strains and plasmids for expression and purification are available upon request (Table 1).

Purification protocols for bacterially expressed factors (TBP, TFIIB, TFIIA, Sub1) were previously published (7, 9, 15) and will not be detailed here. We emphasize that it is important to make sure purified proteins are free of nucleic acid contaminants; typically, we aim to have A260/A280 lower than 0.6. For example, after affinity purification, TFIIB is slightly contaminated with nucleic acids, albeit >90% pure on SDS-PAGE, therefore requiring further purification by a heparin column.

Unlike bacterial cells, yeasts are difficult to lyse with enzymes, detergents, or sonication. Accordingly, cell lysis is processed by mechanical disruption by bead beating, which tends to generate heat and activate proteolysis. Extra care should be taken since GTFs are susceptible to proteolysis and not abundant in cells; for example, only small amounts of GTFs can be obtained without the use of a protease-deficient strain (*pep4*, *prb1*, *prc1*) (16) or without the use of bead beating in continuous flow. After cell lysis, all GTFs are purified by one-step affinity purification through TAP tag on Tfg2 of TFIIF, Tfa2 of TFIIE, or Tfb4 (Tfb3) of TFIIH (16). For this purpose, IgG Sepharose Fast Flow is prepared by coupling IgG, which can be purified from rabbit serum using a Protein A column (GenScript), to CNBR activated Sepharose Fast Flow by a standard method described in the manufacture's instructions. Also, to cleave the TAP tag and elute GTFs from the IgG column, His6-TEV is overexpressed in BL21-CodonPlus (DE3)-RIL strain, purified by Ni-NTA agarose, and frozen in ~1mg aliquots in a volume of 500-800 μ L buffer containing Tris (pH 7.5), 300 mM NaCl, 500 mM imidazole, 30% glycerol, and 5 mM DTT. Further below, we describe methods for cell growth and purification for isolation of milligram quantities of GTFs as guidelines, which are fairly robust and consistent. However, we emphasize that it is necessary to optimize PEI concentration (17) within range of 0-0.25% for the reason described below. For optimization, we typically proceed with small scale purification (~40g of cells from ~4 L YPAD) by using cryogenic grinding (Cryomill, RETSCH) followed by

brief sonication (amplitude 50, a total of 10 min by Q700 sonicator, Qsonica), precipitation with 0, 0.1, 0.2, and 0.25% PEI, and affinity purification using 100–500 μ L of IgG resin in a small open column (Bio-Rad). The affinity tag is cleaved by TEV, and proteins can be analyzed by SDS-PAGE followed by detection by Coomassie Blue.

3.1 Isolation of TFIIH and TFIK

TFIIH tends to dissociate into subcomplexes upon isolation. Deleting *TFB6* prevents the dissociation of Ssl2 (6, 7), enabling isolation of more complete TFIH in a yield much greater than from wild type yeast.

1. Yeast cells are grown in 25 mL synthetic media (SC-Trp) supplemented with 150 μ g/mL ampicillin at 30°C for 2 days (until the 25 mL culture is saturated), and then transferred to 1 L YPAD with 150 μ g/mL ampicillin, and grown for 24 hours. If ampicillin is omitted, more caution should be taken to prevent bacterial and fungal contamination.
2. The 1 L preculture is used to inoculate 100 L of YPAD (1.5 kg peptone, 3 kg yeast extract, 16g adenine, and 3 kg glucose) in a bioreactor at 30 °C with 200 rpm, 10 psi pressure, and 25 L/min airflow until late-log phase. The approximate number of cells are determined during cell growth by the optical density (OD) at 600 nm. Of note, cultures should be diluted, so that the reading should be <1.0. We plot OD₆₀₀ of the culture every ~2-3 hours to generate a growth curve and experimentally determine late-log phase. When the grow rate significantly slows down (usually at OD₆₀₀ 10-15, which may differ depending on yeast strains), cells are harvested with a continuous flow centrifuge, typically yielding 1.0–1.5 kg cell paste.
3. The cell paste is washed with water, and resuspended with an appropriate volume of a lysis buffer, such that a final volume of 2.0–2.4 L contains 50 mM HEPES (pH7.6 adjusted at 4°C), 1 mM EDTA, 400 mM potassium acetate, 5% (w/v) glycerol, 5 mM beta-mercaptoethanol, and protease inhibitors (1 mM benzamidine, 100 μ M leupeptin, 10 μ M pepstatin A, and 1 mM PMSF). Resuspended yeast cells can be frozen and stored at –80°C until purification.
4. The cell resuspension is disrupted by bead beating in continuous flow (DYNO-MILL, Glen Mills Inc.) at a flow rate of 60–70 mL/min. The disruption efficiency should be monitored by counting cells before and after bead beating using a hemocytometer. A typical disruption efficiency is ~80–90%.
5. The cell extract is stirred with 100 mM ammonium sulfate and 0.2 % polyethyleneimine (PEI) (add 10% PEI (P3143, Sigma) pH 7.6 slowly) for 30 min, and centrifuged at 13,800 $\times$ g for 75 min. The concentration of PEI needs to be optimized within a range of 0–0.2%. It should be noted that the amount of PEI to be added depends on the amount of cell extracts in contrast to ammonium sulfate precipitation because PEI binds nucleic acids and acidic proteins very tightly and essentially titrates them (18). For example, 0.2% PEI precipitates acidic components while leaving TFIH in the supernatant in our routine

procedures, whereas only 0.1% PEI would be needed to achieve the same PEI precipitation when the cell lysate is diluted by 2-fold.

6. The supernatant is passed through layers of cheesecloth (Miracloth, EMD Millipore), and loaded onto a 40–50 ml IgG Sepharose Fast Flow settled in a gravity flow column (5.0 × 10 cm, Bio-Rad) preequilibrated in buffer 300 (20 mM HEPES (pH7.6), 300 mM potassium acetate, 1 mM EDTA, 5% (v/v) glycerol, 5 mM DTT, and protease inhibitors). The column is washed with ~3 L of buffer 300 and ~1 L of buffer 300 without PMSF, and is incubated with 1 mg TEV protease for 15 hours at 4°C, before eluting TFIIH with ~6 column volumes of buffer 300. The eluate is collected in 6–10 fractions (~50 mL per fraction), and analyzed by SDS-PAGE. Of note, washing with buffer 300 during the affinity purification yields a mixture of 10-subunit holoTFIIH and 7-subunit core TFIIH (lacking TFIID), whereas washing with buffer 300 containing 400 mM ammonium sulfate yields pure core TFIIH (10).
7. The IgG eluate is applied to a 5 ml HiTrap Q (GE Healthcare) or a 1 mL UnoQ (Bio-Rad) column connected to FPLC (AKTA pure), and TFIIH is eluted in a linear gradient from 300 mM to 1200 mM potassium acetate in 20 mM HEPES (pH7.6), 5% (v/v) glycerol, and 5 mM DTT and is collected in 500 µL fractions. The recommended linear gradient volume is ~100 mL and ~30 mL for 5 mL HiTrap Q and 1 mL UnoQ, respectively. The former gives one single peak containing the holoTFIIH and core TFIIH at ~800 mM potassium acetate, while the latter gives two separated peaks containing holo and core TFIIH at ~525 mM and ~575 mM potassium acetate, respectively. Absorbance at 260 nm (for DNA contamination) and 280 nm (protein) is monitored during chromatography. Proteins in peak fractions are analyzed by SDS-PAGE, pooled, and concentrated using a 100k MWCO concentrator (Sartorius).

TFIID (TAP tag on Tfb3) is isolated in the same manner as described above except that 0.05–0.1% Zwitterionic detergent 3-(decyldimethyl-ammonio) propanesulfonate (Sigma) is added in all buffers and that the IgG eluate is purified by a gel filtration column Superdex 200 instead of HiTrap Q at step 7.

3.2 Isolation of TFIIE

1. Cells are grown in the same manner as for TFIIH and resuspended in lysis buffer (2.0–2.4 L final volume for 1.0–1.2 kg cell paste) such that resuspended cell solution contains 50 mM HEPES (pH7.6), 1 mM EDTA, 250 mM ammonium sulfate, 5% (w/v) glycerol, 5 mM beta-mercaptoethanol, and protein inhibitors (1 mM benzamidine, 100 µM leupeptin, 10 µM pepstatin A, and 1 mM PMSF).
2. After cells are lysed by bead beating, 0.1 % PEI is added drop by drop to the cell extract, and the extract is stirred for 30 min and centrifuged at 13,800 ×g for 75 min.
3. The PEI supernatant is filtered using layers of cheese cloth (Miracloth, EMD Millipore) and loaded onto a 40–50 ml IgG resin in a gravity flow column (5.0 × 10 cm), and washed with ~2 L of buffer 300 plus 400 mM ammonium sulfate

and then ~1 L of buffer 300 and ~1 L of buffer 300 without PMSF. Protein is incubated with 1 mg TEV protease for 15 hours at 4°C, and is eluted with ~6 column volumes of buffer 300.

4. The IgG eluate is concentrated using a 30k MWCO concentrator to 1-2 mL. The concentrated sample should be resuspended well by pipetting in the concentrator prior to transfer to a 2 mL microcentrifuge tube. The sample is centrifuged at $20,627 \times g$ for 15 min to remove any aggregated protein. The supernatant is loaded onto a gel filtration column Superdex 200 16/600 equilibrated with buffer containing 300 mM potassium acetate, 20 mM HEPES, 5% (v/v) glycerol, 2 mM magnesium acetate, and 5 mM DTT. Fractions that contain TFIIE are identified by SDS-PAGE, pooled, and concentrated using a 30k MWCO concentrator (Sartorius).

3.3 Isolation of TFIIF and pol II

TFIIF and pol II can be isolated in a similar manner as for TFIIH, TFIIF, and TFIIE, but are subjected to more rigorous isolation by ammonium sulfate precipitation to remove strongly bound contaminants.

1. Yeast TFIIF is TAP-tagged on the C-terminus of Tfg2. Cells are grown in the same manner as for TFIIH. Cells are resuspended in lysis buffer such that the final resuspension (2.0–2.4 L for 1.0–1.2 kg cell paste) contains TEZ (200) (50 mM Tris (pH 7.5), 1 mM EDTA, 1 uM $ZnCl_2$, 0.15% 3-(decyldimethylammonio) propanesulfonate (Sigma), 3 mM DTT, and protease inhibitors (1 mM benzimidazole, 100 μ M leupeptin, 10 μ M pepstatin A, and 1 mM PMSF) with the mM concentration of ammonium sulfate in parentheses).
2. After cells are lysed by bead beating, the cell extract is stirred with 0.25% PEI for 30 min and centrifuged at $13,800 \times g$ for 90 min. The concentration of PEI needs to be optimized within a range of 0–0.25%.
3. The supernatant is passed through layers of cheesecloth. Ammonium sulfate (final conc. 55%) is added drop by drop to the supernatant to precipitate proteins and the solution is stirred for 30 min before centrifugation at $13,800 \times g$ for 90 min.
4. The pellet resulting from ammonium sulfate precipitation is resuspended with ~1 L of TEZ (25), centrifuged at $13,800 \times g$ for 90 min, and loaded onto an 40-50 mL IgG resin in a gravity flow column (5.0×10 cm). TFIIF is washed with 1 L each of TEZ (250), TEZ (500), and TEZ (25), then 1 L of TEZ (200) without PMSF, followed by incubation with 1 mg TEV protease for 15 h at 4 °C.
5. The protein is eluted with ~6 column volumes of TEZ (200) without PMSF, applied to a 1 or 5 ml heparin column, and eluted in a linear gradient of ammonium sulfate from 75 mM to 800 mM. The recommended volumes of the linear gradient are ~30 mL for a 1 mL heparin column and ~100 mL for a 5 mL heparin column. TFIIF elutes from a heparin column at ~420 mM ammonium sulfate. Fractions containing TFIIF are identified by SDS-PAGE, pooled, and

concentrated using a 100k MWCO concentrator (Sartorius). While concentrating, the buffer is exchanged to storing buffer containing TEZ(400) without protease inhibitors and supplemented with 10% glycerol.

Isolation of pol II is essentially the same as that for TFIIF except that Zwitterionic detergent 3-(decyldimethyl-ammonio) propanesulfonate (Sigma) is omitted in all buffers. Also we add ~2 mg of recombinant Rpb4/7 per ~1 kg of pol II yeast cells (~100 L culture) at the time of addition of TEV protease to the IgG column, as Rpb4/7 tends to dissociate during *S. cerevisiae* pol II isolation. ~2 mg of recombinant Rpb4/7 is sufficient enough to complement pol II that lacked endogenous Rpb4/7, and excess recombinant Rpb4/7 can be removed by following purification using a heparin column at step 5.

The resulting pol II contains the full complement of 12 subunits and unphosphorylated CTD of Rpb1. Rpb1 CTD becomes phosphorylated upon initiation, although CTD phosphorylation is not required for transcription initiation.

Expression and purification of His6-Rpb4/7 are described below.

1. Rpb4/7 plasmid is transformed into *Escherichia coli* Rosetta 2(DE3) competent cells. Single colony is used to inoculate 25 mL LB supplemented with 150 µg/mL ampicillin and the preculture is grown for overnight.
2. 10 mL of the preculture is used to inoculate 2 L of Terrific Broth (TB) media supplemented with 150 µg/mL ampicillin. Expression is induced by addition of 0.1 mM IPTG at OD₆₀₀ ~0.6, and then the culture is grown for 3 hours at 30°C.
3. Cells are harvested, and then washed once with buffer containing 20 mM Tris (pH 7.5) and 300 mM sodium chloride.
4. Cells are resuspended in lysis buffer (20 mM Tris pH 7.5 and 300 mM sodium chloride, 30 mM imidazole, 0.25% Triton, 2mM DTT, and protease inhibitors) and lysed by sonication.
5. The cell debris is removed by centrifugation at 41,657 ×g for 1 hour, and the supernatant is filtered using a syringe filter with a 0.45 µm pore size and loaded onto a Ni column. Rpb4/7 is eluted in a linear gradient of buffer A (20 mM Tris pH 7.5, 500 mM sodium chloride, 10 mM imidazole) to buffer B (20 mM Tris pH 7.5, 500 mM sodium chloride, 500 mM imidazole).
6. The fractions containing Rpb4/7 are identified by SDS-PAGE, pooled, dialyzed to dialysis buffer (20 mM Tris pH 7.5, 100 mM sodium chloride, 2 mM DTT), and loaded on a HiTrapQ column. Rpb4/7 is eluted by a linear gradient of 100 mM to 1M sodium chloride in 20 mM Tris pH 7.5 and 2 mM DTT. Rpb4/7 elutes at ~415 mM sodium chloride. Fractions containing Rpb4/7 are supplemented with 20% glycerol and 70 mM ammonium sulfate, and stored at -80°C.

4. Reconstitution of PIC

Here we describe two methods to reconstitute the PIC (8): one is for large scale reconstitution (up to ~300 pmol (~450 µg) of PIC) and the other for small scale

reconstitution (less than ~80 pmol (~120 µg) of PIC). Large scale reconstitution is done by stepwise salt dialysis, followed by isolation by gradient sedimentation, and used for biophysical and structural analyses. Small scale reconstitution is done by mixing all factors at high ionic strength and diluting into moderate ionic strength in a 1.5 mL microcentrifuge tube, and is used for biochemical studies as well as biophysical and structural analyses. Small-scale and large-scale reconstitutions described here circumvent formation of nonspecific complexes by starting the assembly at high ionic strength, in which all GTFs and pol II are soluble by themselves, followed by dilution into lower ionic strength solution (~150 mM potassium acetate), upon which the PIC assembly occurs.

4.1 Small-scale reconstitution

Small scale reconstitution without isolation of the resulting complex is suitable for transcription assays and DNA footprinting assays using radiolabeled RNA transcripts and DNA template (8). All factors and promoter DNA are prepared in a high ionic strength solution (300 mM), in which all GTFs and pol II are soluble by themselves. We note at least 300 mM ionic strength is required for solubility of TFIID. To assemble the PIC, GTFs, pol II, and DNA are combined and the mixture is diluted into lower ionic strength solution (~150 mM potassium acetate). Transcription is initiated by adding an appropriate volume (typically 2×) of solution containing NTPs (final conc. 800 mM ATP, 800 mM CTP, and 500 mM UTP), and the resulting transcription reaction contains ~80–100 mM ionic strength and ~40–80 nM PIC. Transcription initiation efficiency is reduced when ionic strength is higher than ~120 mM or when PIC concentration is lower than 4 nM or higher than 100 nM.

The following steps are used for small scale reconstitution and transcription reaction. A typical assembly (~1.3 pmol of PIC in 10 µL reaction volume) is performed by mixing ~1.3 pmol of promoter DNA with a slight molar excess of the general transcription factors (1.5–3.0 pmol) except TFIIF, and more excess pol II and TFIIF (5–8 pmol) (Fujiwara et al. in preparation). The amounts of each factor to be added are given only as guidelines, and should be experimentally determined considering batch-to-batch difference of protein activity.

1. All reconstitution experiments are performed on ice. All factors and DNA are diluted with buffer 300 (50 mM HEPES (pH 7.6), 300 mM potassium acetate, 5 mM DTT, and 5% (v/v) glycerol) to ~1–20 µM each. Our typical assembly (1.3 pmol of PIC) contains 1.3 pmol of DNA template, 3.0 pmol of TFIIB, 2.0 pmol of TFIIA, 1.5 pmol of TBP, 3.0 pmol of TFIIE, 1.5 pmol of TFIID, 1.8 pmol of TFIIF, 1.5 pmol of pol II, 1.5 pmol of TFIIF. Purified TFIID described above often contains substoichiometric amounts of TFIIF and thus supplemented with TFIIF that is separately isolated from Tfb3-TAP tagged yeast strain (Table 3). Excess TFIIF has no effect on transcription in the in vitro system. Additionally, there is a significant (~2–3-fold) increase in transcription with increasing amounts of pol II and TFIIF (up to ~5–8 pmol) over 1.5 pmol of each factor (Fujiwara et al., in preparation). Also there is some (~1.4-fold) increase with addition of Sub1 (3.0 pmol), a yeast homologue of PC4 (19). To test the effects of additional transcription factors on initiation, amounts of salt brought into the

reaction must be taken into account, given transcription initiation *in vitro* is highly sensitive to ionic strength.

2. All factors are mixed in 5 μ L of buffer 300, and then the mixture is diluted with 5 μ L of buffer 10 (20 mM HEPES [pH 7.6], 10 mM potassium acetate, 5 mM magnesium sulfate, 5 mM DTT) and incubated on ice for 24 hours.
3. For transcription, 10 μ L of the PIC is preincubated at 30°C for 20 min, and is mixed with 10 μ L of 2 $\times$ NTPs containing 20 mM HEPES [pH 7.6], 10 mM potassium acetate, 5 mM magnesium sulfate, 5 mM DTT, 1.6 mM ATP, 1.6 mM GTP, 1.6 mM CTP, 1 mM UTP, 1 unit of RNaseOUT, 99 nM [α -³²P] UTP (3 μ Ci). Reaction is ~80% completed in 10 min and completed in 20 min (8), and is stopped by adding 190 μ L of stop buffer (10 mM Tris, 300 mM sodium acetate (pH 5.5), 5 mM EDTA, 0.7% SDS, 0.1 mg/mL glycogen, 0.013 mg/ml of proteinase K).
4. Transcripts are precipitated by the addition of 3 volumes of ethanol, incubated overnight (to precipitate all transcripts including short RNAs) at -20°C and centrifuged at 20,627 $\times$ g for 60 min at 4°C. The pellet is dried and resuspended with ~10 μ L of 1 $\times$ RNA loading dye containing 0.025% (w/v), Bromophenol, 0.025% (v/v) SDS, and 5 mM EDTA (pH 8.0). Transcripts are analyzed by a denaturing 4–12% acrylamide gel containing 7 M Urea and 1 $\times$ TBE using S2 Sequencing Gel Electrophoresis System (Gel Company). Undried gels are exposed to PhosphorImager screens at -20°C for visualization of radiolabeled transcripts.
5. To calculate the transcription initiation efficiency, a dilution series of the stock radiolabeled NTP is blotted on a filter paper and exposed to a phosphorimager screen with the assay gel. Band intensities of transcripts initiated from TSSs and standard blots are measured using Image J. A linear regression is fitted through the standard points and its equation is used to calculate moles of ³²P NTP in the transcripts. Then transcription efficiency is determined as follows:

$$\text{transcription efficiency} = \frac{\text{moles of } ^{32}\text{P NTP} \div \text{number of Ns in a transcript} \times ([\text{cold NTP}])}{([\text{hot NTP}]) \div \text{moles of preinitiation complex}} \times 100.$$

It is important to note that intensities of transcripts and standard blots should not be saturated when measuring in order to accurately determine the transcription efficiency.

4.2 Isolation of PIC by gradient sedimentation

Up to ~80 pmol (~120 μ g) of PICs assembled on the small-scale method can be isolated by glycerol gradient sedimentation using a SW60 rotor (Beckman) (Fig. 3). For example, for structural and biophysical analyses, we load as much as ~80 pmol (~120 μ g) of PICs (corresponding to ~600 μ L) onto a gradient, whereas, to determine the stoichiometry and the assembly of the resulting complex by SDS-PAGE with Coomassie, we typically load ~26 pmol (40 μ g) of the PICs onto a 10–40% glycerol gradient (3.5 - 4.0 mL), followed by

fractionation and TCA precipitation. We note that although addition of excess pol II and TFIIF relative to the PIC is essential in achieving high initiation efficiency, addition of equimolar amounts of pol II and TFIIF should be sufficient to assemble the PIC.

Here we describe a protocol for isolation of ~26 pmol (40 µg) of PICs by gradient sedimentation.

1. A glycerol gradient is prepared using a Gradient Maker (BioComp) by a standard method described in the manufacture's instructions (Fig. 3A) using two buffers: one contains lower concentration of glycerol (20 mM HEPES (pH7.6), 80 mM KOAc, 5 mM DTT, 2 mM MgOAc, 10%(v/v) glycerol), and the other is the same except it contains a high concentration of glycerol (40%(v/v)).
2. Centrifugation conditions are chosen so that the complex migrates into ~2/3 of the gradient. The PIC is sedimented at 30,000 rpm (92,434 ×g) in a Beckman SW60 rotor for 14 h.
3. Gradients are fractionated from bottom to top by using a peristaltic pump or, if available, from top to bottom using a Piston Gradient Fractionator (BioComp Instruments, Inc.) (Fig. 3A). Fractionation using the latter provides better separation of the PIC from free unassembled proteins than the former. We typically collect 20–40 fractions (~100–200 µL per fraction) depending on the resolution that needs to be achieved.
4. TCA precipitation (steps 4–6) is required when small amounts (e.g., ~26 pmol (40 µg)) of PICs are subjected to a gradient. Add 1/5 volume (40 µL) of 100% TCA to each fraction (200 µL), vortex, incubate on ice for 1 hour (do not incubate overnight), and spin at 20,627 g for 15 min at 4 °C.
5. Wash the pellet with 500 µL of solution containing 10% TCA, 0.1M Tris-base (without adjusting pH), and spin at 20,627 ×g for 15 min at 4 °C.
6. Add 500 µL of cold acetone to the pellet, invert the tube several times, and spin at 20,627 ×g for 15 min at 4 °C. Decant the supernatant and dry the tube at room temperature.
7. Add 20 µL NuPAGE 1× LDS sample buffer (4× diluted with water and supplemented with ~5 mM DTT) to the pellet, vortex for 5 min, heat at 90 °C for 5 min, and load 20 µL onto a SDS-PAGE gel.

4.3 Large-scale reconstitution

When large amounts of PICs need to be isolated for biophysical and structural studies, the small-scale method is not preferred, as simple mixing of highly concentrated factors results in precipitation (8). Thus we developed a procedure for large scale reconstitution, which entails combining all GTFs except TFIIF with promoter DNA at high ionic strength, dialysis into moderate ionic strength, then the addition of pol II and TFIIF, followed by dialysis into lower ionic strength. The reconstituted PICs are isolated by gradient sedimentation, as described above. This large-scale reconstitution is designed to represent a process for

transcription re-initiation, in which the GTF-DNA complex forms a structural scaffold prior to recruitment of pol II-TFIIF (20).

As an example, here we describe procedures for reconstitution and isolation of ~ 150 pmol of PICs:

1. All reconstitution experiments are performed at 4 °C. Promoter DNA (0.15 nmol) is combined with 0.75 nmol of TFIIB, 0.75 nmol of TFIIA, 0.5 nmol of TBP, 0.6 nmol of TFIIE, and 0.24 nmol of TFIIH, 0.2 nmol of TFIIF, and 0.2 nmol of pol II in buffer 500 (50 mM HEPES (pH 7.6), 5 mM DTT, 5% glycerol, and 500 mM potassium acetate). The total volume of the mixture is typically less than 150 µL.
2. The protein mixture is dialyzed using Slide-A-Lyzer MINI Dialysis Devices (Thermo), 20K MWCO in steps of 50-100 mL each of buffer 300, buffer 220, and buffer 150 for at least 4 h at each step to form the GTF-DNA complex and then combined with 0.25 nmol of pol II and 0.25 nmol of TFIIF, allowing PIC assembly. The dialysis into buffer 300 and buffer 220 can be omitted, whereas the dialysis into buffer 150 is important as the GTF-DNA complex is unstable at ionic strength higher than ~150 mM, and is insoluble at ionic strength lower than ~150 mM.
3. The mixture is further dialyzed into 50-100 mL buffer 80 before loading on a ~3.5 mL 10–40% (v/v) glycerol gradient containing 80 mM ionic strength. The gradient is fractionated as described above, frozen in liquid nitrogen, and stored at –80°C.

5. Discussions

Earlier biochemical reconstitution systems presented a longstanding serious technical limitation, namely very poor initiation efficiency. The inefficiency is partly due to inefficient assembly of the pre-initiation complex (PIC) on a promoter template. We overcame this technical limitation by isolation of pure homogeneous GTFs and stepwise reconstitution of the PIC (8). When GTFs, pol II, and DNA are simply mixed at low ionic strength, there are many nonspecific GTFs-DNA complexes, which fail to convert to specific PICs over a reasonable time frame. Both small-scale and large-scale reconstitutions described here circumvent formation of these nonspecific complexes by starting the assembly at high ionic strength. In addition, specifically assembled PICs on promoter DNA by using either the small-scale or the large-scale method can be separated from excess GTFs and pol II that nonspecifically bind to DNA by a glycerol gradient sedimentation, irrespective of the length of promoter DNA. Indeed, the PIC can be assembled on a 2.5 kbp *HIS4* promoter DNA fragment and isolated in complete homogeneous form (9).

However, we should emphasize that transcription activity of such highly-purified homogeneous PIC is ~20% at best with *SNR20* promoter (8) and is even poorer with DNA exceeding ~250 bp in length. Evidently, even such highly purified homogeneous PIC has technical limitations in initiation; PIC may dissociate upon addition of nucleotides without RNA synthesis and fail to re-assemble especially in the presence of naked non-specific

DNA. Lastly it should be noted that we have recently achieved highly-efficient (>90%) transcription by addition of ~4–6 fold molar excess pol II and TFIIF to transcription reaction (Fujiwara et al., in preparation). This may suggest that, upon addition of nucleotides, ~80% PICs become inactivated without synthesizing transcripts, but can be rescued by recruiting new pol II-TFIIF molecules, while recycling GTFs and DNA.

In addition to the highly-efficient transcription, we have established a robust method to isolate an abundant homogeneous PIC, which opens the way to definitive biochemical and structural studies. Of note, after isolation by gradient sedimentation, we observe some partial PICs lacking TFIIF (fractions 9–12 in Fig. 3B), which more slowly migrate into a gradient than complete PICs (fractions 3–7 in Fig. 3B). The partial PICs are observed even when more excess TFIIF is added to the reaction, indicating that the loss of TFIIF is due to the rigors of isolation by gradient sedimentation. The loss of TFIIF was not observed when the PIC was assembled with Mediator (21), and, in this regard, the PIC-Mediator may represent a more fully transcription-competent initiation complex.

The stepwise reconstitution method yields very robust reproducible PIC assembly, and can be easily modified to accommodate many other transcription factors. For example, as mentioned above, Mediator for transcription regulation can be incorporated into a PIC and isolated as a complex in homogeneous form in a slightly modified procedure (21). In addition, our highly efficient system now achieves a complete reconstitution of pol II transcription from initiation by the PIC through mRNA capping and promoter escape, to finally reach productive elongation. (Fujiwara et al., in preparation). It will be of great interest to pursue how each process is influenced by other factors, such as activators, coactivators, and nucleosomes.

ACKNOWLEDGEMENTS

The system is based on studies in the Kornberg lab at Stanford University and further improvement at University of Pennsylvania. The plasmid His6-SUMO-TFIIF is a gift from the Liu lab at UT Southwestern. We thank all past and current lab members who shared techniques and knowledge. A part of this research was supported by NIH Grant GM123233.

REFERENCES

1. Conaway RC & Conaway JW (1993) General initiation factors for RNA polymerase II. Annual review of biochemistry 62:161–190.
2. Orphanides G, Lagrange T, & Reinberg D (1996) The general transcription factors of RNA polymerase II. Genes & development 10(21):2657–2683. [PubMed: 8946909]
3. Sayre MH, Tschochner H, & Kornberg RD (1992) Purification and properties of *Saccharomyces cerevisiae* RNA polymerase II general initiation factor α . The Journal of biological chemistry 267(32):23383–23387. [PubMed: 1429681]
4. Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci 21(9):327–335. [PubMed: 8870495]
5. Giardina C & Lis JT (1993) DNA melting on yeast RNA polymerase II promoters. Science 261(5122):759–762. [PubMed: 8342041]
6. Gibbons BJ, et al. (2012) Subunit architecture of general transcription factor TFIIF. Proceedings of the National Academy of Sciences of the United States of America 109(6):1949–1954. [PubMed: 22308316]

7. Murakami K, et al. (2012) Tfb6, a previously unidentified subunit of the general transcription factor TFIID, facilitates dissociation of Ssl2 helicase after transcription initiation. *Proceedings of the National Academy of Sciences of the United States of America* 109(13):4816–4821. [PubMed: 22411836]
8. Murakami K, et al. (2013) Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. *The Journal of biological chemistry* 288(9):6325–6332. [PubMed: 23303183]
9. Fazal FM, Meng CA, Murakami K, Kornberg RD, & Block SM (2015) Real-time observation of the initiation of RNA polymerase II transcription. *Nature* 525(7568):274–277. [PubMed: 26331540]
10. Murakami K, et al. (2015) Uncoupling Promoter Opening from Start-Site Scanning. *Molecular cell* 59(1):133–138. [PubMed: 26073544]
11. Nikolov DB, et al. (1996) Crystal structure of a human TATA box-binding protein/TATA element complex. *Proceedings of the National Academy of Sciences of the United States of America* 93(10):4862–4867. [PubMed: 8643494]
12. Lubliner S, et al. (2015) Core promoter sequence in yeast is a major determinant of expression level. *Genome research* 25(7):1008–1017. [PubMed: 25969468]
13. Murakami K, et al. (2015) Structure of an RNA polymerase II preinitiation complex. *Proceedings of the National Academy of Sciences of the United States of America* 112(44):13543–13548. [PubMed: 26483468]
14. Fishburn J & Hahn S (2012) Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. *Molecular and cellular biology* 32(1):12–25. [PubMed: 22025674]
15. Ranish JA, Lane WS, & Hahn S (1992) Isolation of two genes that encode subunits of the yeast transcription factor IIA. *Science* 255(5048):1127–1129. [PubMed: 1546313]
16. Borggreffe T, Davis R, Bareket-Samish A, & Kornberg RD (2001) Quantitation of the RNA polymerase II transcription machinery in yeast. *The Journal of biological chemistry* 276(50):47150–47153. [PubMed: 11591727]
17. Yang C, Khaperskyy DA, Hou M, & Ponticelli AS (2010) Improved methods for expression and purification of *Saccharomyces cerevisiae* TFIIF and TFIID; identification of a functional *Escherichia coli* promoter and internal translation initiation within the N-terminal coding region of the TFIIF TFG1 subunit. *Protein Expr Purif* 70(2):172–178. [PubMed: 19818408]
18. Burgess RR (2009) Protein precipitation techniques. *Methods in enzymology* 463:331–342. [PubMed: 19892180]
19. Henry NL, Bushnell DA, & Kornberg RD (1996) A yeast transcriptional stimulatory protein similar to human PC4. *The Journal of biological chemistry* 271(36):21842–21847. [PubMed: 8702984]
20. Yudkovsky N, Ranish JA, & Hahn S (2000) A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408(6809):225–229. [PubMed: 11089979]
21. Robinson PJ, et al. (2016) Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* 166(6):1411–1422 e1416. [PubMed: 27610567]
22. Bratkowski M, et al. (2018) Structural dissection of an interaction between transcription initiation and termination factors implicated in promoter-terminator cross-talk. *The Journal of biological chemistry* 293(5):1651–1665. [PubMed: 29158257]

Figure 1.
The non-template strand sequences of promoters *SNR20* (-122/+53) and *HIS4* (-96/+74) are shown (8). The TATA box is in red. The in vitro transcription start sites are indicated by black arrows. Minor upstream TSSs (10) are indicated by red arrows.

Figure 2.
Analysis of purified transcription factors by SDS-PAGE.

Figure 3.

Isolation of the 31-subunit PIC. A. The gradient sedimentation workflow. Glycerol gradient is prepared using a Gradient Maker (step1). Up to 600 μ L of sample is loaded onto a 4 mL gradient for SW60 (Beckman) (step 2). Sample is sedimented (step 3) and fractionated by a peristaltic pump (step 4-a) or a piston fractionator (step 4-b). B. A representative fractionation of the PIC. The 31-subunit PIC was assembled with 82-bp *HIS4* promoter fragment, sedimented on a 4 mL glycerol gradient, and fractionated (100 μ L each) by a piston fractionator, and every other fraction was analyzed by SDS-PAGE, as previously published (8).

Table 1

Transcription factors required for highly-efficient transcription initiation.

Component	Number Subunits	Mass (kDa)	Functions	Isolation methods	yield
TFIIA	2	46	Stabilizing TBP on promoter DNA	overexpression in E.coli	~2 mg/L
TFIIB	1	38	Start site determination	overexpression in E.coli	~2 mg/L
TBP	1	1	Binding TATA box DNA	overexpression in E.coli	~10 mg/L
TFIIE	2	92	DNA binding, Recruitment of TFIIH	S. cerevisiae	1.5–3 mg/L
TFIIF	3	156	DNA binding	S. cerevisiae	1.5–3 mg/100 L
TFIIH	10 (+Tfb6)	525 (+39)	Translocase activity	S. cerevisiae	1–3 mg/100 L
TFIIK	3	120	CTD phosphorylation	S. cerevisiae	1.5–3 mg/100 L
Sub1	1	1	ssDNA binding	overexpression in E.coli	~1 mg/L
pol II	12	520	RNA synthesis	S. cerevisiae	>5 mg/100 L

Table 2

Plasmids to overexpress GTFs for PIC assembly

Protein name	Expression vector	References
TBP	pRSF-TBP	Murakami et al., 2012 (7)
Toa1 (TFIIA)	pRSF-TOA1	Murakami et al., 2012 (7)
Toa2 (TFIIA)	pRSF-TOA2	Murakami et al., 2012 (7)
His6-SUMO-TFIIB	pET28-SUMO-SUA7	Bratkowski et al., 2017 (22)
His6-Sub1	pCold II-SUB1-His6	Fazal et al., 2015 (9)

Table 3Strains of *S. cerevisiae* used for PIC assembly

Protein name	Yeast strain	Genotype	References
	CB010	<i>MATa pep4::HIS3 prb1::LEU2 prc::HISG can1 ade2 trp1 ura3 his3 leu2-3,112</i>	Borggrete et al., 2001 (16)
TFIIE	Tfa2-TAP	CB010 except <i>TFA2::TAP::KI.TRP1</i>	Borggrete et al., 2001 (16)
TFIIF	Tfg2-TAP	CB010 except <i>TFG2::TAP::KI.TRP1</i>	Borggrete et al., 2001 (16)
TFIIH	Tfb4-TAP delta-tfb6	CB010 except <i>TFB4::TAP::KI.TRP1 tfb6::kanMX6</i>	Murakami et al., 2012 (7)
TFIIK	Tfb3-TAP delta-tfb6	CB010 except <i>TFB3::TAP::KI.TRP1 tfb6::kanMX6</i>	Murakami et al., 2012 (7)
pol II	Rpb3-TAP	CB010 except <i>RPB3::TAP::KI.TRP1</i>	Borggrete et al., 2001 (16)