Facile and scalable expression and purification of Transcription factor IIH (TFIIH) Core complex

Reyes Sanles-Falagan, Biljana Petrovic-Stojanovska and Malcolm F White*

Biomedical Sciences Research Complex, School of Biology, University of St Andrews, St Andrews KY16 9ST, UK.

* for correspondence: email mfw2@st-andrews.ac.uk

HIGHLIGHTS

- Auto-cleavable 2A-like sequences effectively simplify the cloning of multi-subunit complexes.
- A versatile 6xHis-V5 epitope-TEV cleavable tag facilitates streamlined purification.
- 7-subunit, ~350 kDa sub-complex cloned using the MutiBac™ system.
- Pure sub-complex obtained in mg quantities using High Five™ insect cells as an expression system.

ABSTRACT

Transcription factor IIH (TFIIH) plays essential roles in both the initiation of RNA Polymerase II-mediated transcription and the Nucleotide Excision Repair (NER) pathway in eukaryotes. In NER, the 7-subunit TFIIH Core sub-complex is responsible for the opening and extension of the DNA bubble created at the lesion site, utilizing the molecular motors XPB and XPD. Mutations in Core subunits are associated with a series of severe autosomal recessive disorders characterised by symptoms such as mild-to-extreme photosensitivity, premature ageing, physical and neurological anomalies, and in some cases an increased susceptibility to cancer. Although TFIIH Core has been successfully obtained in the past, the process has always remained challenging and laborious, involving many steps that severely hindered the amount of pure, active complex obtained. This has limited biochemical and functional studies of the NER process. Here we describe improved and simplified processes for the cloning,

expression and purification of the 7-subunit TFIIH Core sub-complex. The combined use of auto-cleavable 2A-like sequences derived from the Foot-and-Mouth Disease Virus (FMDV) and the MultiBac™ cloning system, a powerful baculoviral expression vector specifically conceived for the obtaining of multi-subunit eukaryotic complexes, allowed us to obtain a single, 7-gene plasmid in a short time using regular restriction cloning strategies. Additionally, expression of the construct in High Five™ insect cells paired with a simple 5-step purification protocol allowed the extraction of a pure, active TFIIH Core sub-complex in milligram quantities.

KEYWORDS

TFIIH Core, 2A-Sequences, MultiBac™, *Sf*9, High Five™, immobilised metal ion affinity chromatography, heparin chromatography, TEV cleaving, size exclusion chromatography.

INTRODUCTION

The integrity of DNA is constantly threatened by internal and external factors, with the genome of a human cell typically suffering more than 10^5 lesions in a single day (Tomkinson *et al.*, 2006). In eukaryotes, the nucleotide excision repair (NER) pathway protects cells from mutagenic and carcinogenic processes by removing an extensive range of bulky adducts and lesions that introduce a significant distortion in the double helix. This pathway is characterized by its three defining steps of damage recognition, unwinding of the DNA duplex around the distortion and removal of the lesion (Compe & Egly, 2012). The NER process involves the coordinated action of more than 20 proteins in a complex, dynamic process that is still not fully understood at a molecular level.

TFIIH is a 10-subunit complex of about 500 kDa that plays important roles not only in NER, but also in transcription (Flores *et al.*, 1992) (Feaver *et al.*, 1993). The first published structures of TFIIH revealed a complex with two characteristic regions: a Core sub-complex that included subunits XPB, p62, p52, p44, p34 and p8, and a cyclin-dependent kinase (CDK)-activating kinase (CAK) sub-complex comprising subunits cyclin-dependent kinase 7 (cdk7), cyclin H

and ménage-à-trois 1 (MAT1), with subunit XPD acting as a structural bridge between both sub-complexes (Schultz *et al.*, 2000) (Abdulrahman *et al.*, 2013) (Giglia-Mari *et al.*, 2004). The recent high resolution electron cryomicroscopy structures of yeast and human TFIIH confirmed the unmistakable position of every Core subunit within the horseshoe-shaped arranged sub-complex, with the CAK sub-complex protruding from it (Greber *et al.*, 2017) (Kokic *et al.*, 2019). The roles carried out by TFIIH in NER and transcription require different structural arrangements of the complex, with the intact 10-subunit complex participating in the latter, while only the Core sub-complex is needed to complete the repair reaction (Svejstrup *et al.*, 1995) (Tirode *et al.*, 1999). This Core sub-complex presents two enzymatic subunits: the ATP-dependent molecular motors XPB and XPD, which are both essential for NER (Coin *et al.*, 1998) (Oksenych *et al.*, 2009), as their coordinated actions will open and extend the DNA bubble to allow the completion of the repair process.

Subunits XPB, XPD and p8 can be affected by different mutations which are the cause behind a number of severe and sometimes even lethal genetic syndromes that show an exceptional diversity in their genetic make-up and clinical manifestations. Patients affected by xeroderma pigmentosum (XP) display a characteristic photosensitivity and a higher predisposition to suffer from cancer as a result of an impaired NER pathway (Peng *et al.*, 2011). By contrast, Cockayne syndrome (CS) and trichothiodystrophy (TTD) patients exhibit a remarkable heterogeneity of physical and mental defects that severely hinder their development (Lehmann, 2003), but do not present a higher frequency of cancer, consistent with defects in transcription rather than impairment of the repair process (Dubaele *et al.*, 2003) (Bootsma & Hoeijmakers, 1993).

For a long time, the purification of eukaryotic multi-protein complexes has remained a challenge due to the limitations of the available cloning and expression tools. The biochemical and structural analyses of these complexes has therefore been restricted due to the high quality and quantity of the sample required for these studies, rarely achievable with the technologies available until recently (Bieniossek *et al.*, 2012). The routine use of baculoviral expression vectors (BEVs) for overexpression of recombinant proteins in laboratories has helped overcome this problem, especially with the development of production tools specifically conceived to obtain eukaryotic multi-protein complexes, such as the MultiBac™ system (Berger *et al.*, 2004). The use of BEVs for expression of heterologous proteins offers a number of advantages over the more traditional bacterial expression systems, whose protein production machinery frequently cannot synthesise many eukaryotic complexes presenting

subunits of large size. This doesn't pose a problem for insect cells, which in addition offer post-translational modifications and a folding apparatus similar to mammalian cells, and while expression of heterologous protein requires shorter times using the *E. coli* system, the sample obtained from insect cells frequently offers a considerably higher yield and improved quality (Bieniossek *et al.*, 2012) (Assenberg *et al.*, 2013).

The first reported purification of TFIIH, from HeLa cell extracts, included three fractionation procedures, four chromatography columns and four dialysis steps (Flores *et al.*, 1992). The purification of active TFIIH stills remains a challenging and arduous process. Our lab has simplified substantially both the cloning and expression processes for the TFIIH Core sub-complex by using the MultiBac™ expression system, further aided by the introduction of short, multi-purpose purification tags designed by our group into specific Core subunits (Rouillon *et al.*, 2019) and the combination of pairs of genes into a single unit by linking their sequences through a 2A-like auto-cleavable peptide (Luke *et al.*, 2009). The cloned 7-gene construct was transposed into a baculoviral genome engineered for improved protein expression (Bieniossek *et al.*, 2012) (Berger *et al.*, 2004), and this bacmid was purified and transfected into a monolayer of *Sf*9 insect cells. The recombinant viral stock obtained was finally used to infect a High Five™ suspension culture (Bieniossek *et al.*, 2012) that ultimately yielded ~ 1mg of active, purified TFIIH Core sub-complex for every 1 L of culture following a five-step purification process that can be completed in 2 days. This procedure opens the door to much-needed functional and biochemical studies of the NER pathway *in vitro*.

MATERIALS AND METHODS

TFIIH Core engineered genes and self-cleaving 2A sequences

Synthetic genes (not codon-optimized) for all seven *Homo sapiens* TFIIH Core subunits were ordered from Integrated DNA Technologies (IDT). A six-residue polyhistidine tag was designed to be attached to different TFIIH Core subunits for purification purposes. The tag also included a TEV cleaving site for removal of the tag, and a V5 epitope for protein detection in a western blot (Rouillon *et al.,* 2019) (figure 1).

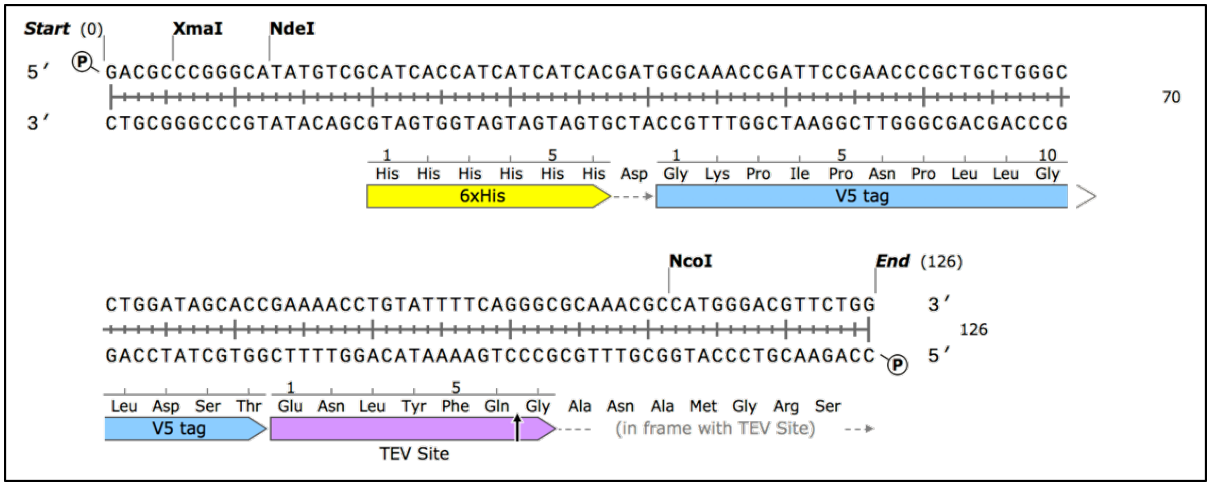

Figure 1: 6xHis-V5-TEV purification tag.

We designed this 6xHis-V5-TEV multi-purpose tag with several functions in mind: (1) purification of our proteins in an IMAC by means of the 6-residue His tag, (2) Western blot detection and potential immunoprecipitation of the complex through the V5 epitope, and (3) removal of the tag from the purified protein if necessary thorough cleaving at the TEV site.

The genes for Core subunits p44/p34 and p52/p8 were designed as a single unit. The p44/p34 genes were separated by the 2A-like sequence T2A (EGRGSLLTCGDVEENPGP) and the p52/p8 genes by the sequence F2A (GKPIPNPLLGLDSTQTLNFDLLKLAGDVESNPGP) (figure 2A).The 2A auto-cleavable sequence from which these F2A and T2A peptides were derived was originally observed in the genome of the Foot-and-Mouth Disease Virus (FMDV). During translation, the ribosome becomes stalled upon encountering this sequence, and the nascent protein is released by self-cleaving at the NPG'P site before resuming translation by translocation of the ribosome to the next in-frame starting point. The cleaved site remains attached to the C-terminal end of the newly translated protein without generally affecting its function or expression (Luke *et al.*, 2009).

Cloning using the MultiBac™ multiplication module

The 7-subunit TFIIH Core sub-complex was cloned using the MultiBac™ baculovirus expression system (Geneva Biotech). As we cloned all of our genes using the homing endonuclease / *Bst*XI multiplication module, only the acceptor plasmids pACEBac1 (2904bp) and pACEBac2 (2761bp) were used (Bieniossek *et al.*, 2013).

In a first step each TFIIH Core gene was cloned into either pACEBac1 or pACEBac2 using restriction enzymes *Bam*HI/*Sal*I and *Nco*I/*Kpn*I (Fermentas), respectively, following the

manufacturer's instructions. Afterwards all seven genes were cloned, one gene at a time, into a single plasmid using the multiplication module I-*Ceu*I/*Bst*XI (New England Biolabs) as indicated by MultiBac™'s manufacturer, until a single recombinant plasmid containing XPD, XPB, p62, p52/F2A/p8 and p44/T2A/p34 was achieved (Bieniossek *et al.*, 2009). All constructs were expressed in DH5α cells.

Intermediate constructs were verified by restriction with the appropriate combinations of different enzymes (*Nco*I, *Kpn*I, *Bam*HI, *Hind*III, etc.) (Fermentas) following the manufacturer's instructions. The final TFIIH Core construct was further analysed by PCR using the MyTaq™ DNA Polymerase (Bioline) and the primers listed in table 1. The presence of all seven genes was ultimately verified by sequencing (GATC Biotech).

Table 1: TFIIH Core sub-complex PCR primers

Oligonucleotide	Sequence 5' to 3'
XPD Fwd	TACATGCGGGAGCTCAAACGC
XPD Rev	TCTGCTCTATCCTCTTCAGC
XPB Fwd	GGCCATATCTTCTTGGAAGC
XPB Rev	GCCAGGACTTTCTGTAAGAGC
p62 Fwd	CCTCATCTGAAGTTTTGC
p62 Rev	CGTTTTCTTCATCAGACGCCG
p52/p8 Fwd	GGATCTGGCACACACAGCTGC
p52/p8 Rev	CTTGAATGATGAACTTCTTCC
p44/p34 Fwd	TTAGAGTATCTGTTATTGG
p44/p34 Rev	GACCCTTTATGTCACTTTTGG

Tn7 Transposition and bacmid purification

The recombinant plasmid carrying the TFIIH Core genes was transposed into chemically competent DH10 MultiBac™ cells (provided with the MultiBac™ kit) containing an engineered baculoviral genome derived from the *Autographa californica* multiple nucleopolyhedrovirus

*(Ac*MNPV) (Bieniossek *et al*., 2012) and a helper plasmid containing a Tn7 transposase, as per manufacturer's instructions.

A minimum of two recombinant white colonies per construct were picked and re-streaked to confirm their white phenotype. At the same time, these colonies were used to start 3 ml Luria-Bertani (LB) broth cultures containing kanamycin 50 μg/ml and gentamycin 7 μg/ml. The cultures were incubated at 37 °C, 200 rpm overnight, then centrifuged at 15000 rpm for 1 min at room temperature (centrifuge 5424, Eppendorf). Pellet was resuspended in buffer P1, then buffer P2 and finally buffer N3 from a QIAprep Spin Miniprep Kit (Qiagen). The suspension was centrifuged at 15000 rpm for 10 min at room temperature, and supernatant was transferred to a fresh tube. The baculoviral DNA present in the supernatant was precipitated with isopropanol and washed twice with ethanol 70%. All work following the second washing step was performed in a sterile S@feflow 1.2 hood (Bioair Instruments). Ethanol was removed from the tube, and sterile bacmids were finally resuspended in 30 μl deionized water and kept at 4 °C until transfected.

The transposition protocol was adapted for each construct according to its size: (1) incubation on ice varied from 30 min for smaller constructs to 2 h for the 7-gene Core construct; (2) heat shock went from 60 to 90 s for the bigger constructs; (3) chilling on ice was extended from 5 min up to 20 min; (4) the recovery step was extended from 6 to 24 h at 37 °C with shaking as the construct grew bigger. Each culture was then plated into LB agar containing kanamycin 50 μg/ml, gentamycin 7 μg/ml, tetracycline 10 μg/ml, ampicillin 100 μg/ml, isopropyl β-D-1-thiogalactopyranoside (IPTG) 40 μg/ml and X- gal 20 μg/ml and incubated at 37 °C for 2-3 days.

Insect cell transfection

Recombinant bacmids were transfected into a monolayer of *Spodoptera frugiperda* (*Sf*9) cells (Thermo Scientific) using X- tremeGENE HP DNA Transfection Reagent (Roche). $1x10^6$ cells were seeded to each well in a 6-well plate (Greiner Bio-One), and volume was topped up to 3 ml with fresh *Sf*900™ III SFM medium (Thermo Scientific). The plate was incubated at 27 °C for a minimum of 2 h to allow cells to attach to the surface of the well. Meanwhile, each bacmid to be transfected was resuspended in 200 μl *Sf*900™ III SFM medium. A mixture of 100 μl *Sf*900™ III SFM medium and 10 μl transfection reagent was added to each bacmid, and tubes were incubated at room temperature for 30 min. Afterwards, the bacmid/transfection reagent

mixture was added to the *Sf*9 cell monolayer (2 wells for every bacmid to be transfected), and plates were incubated at 27 °C for 60 h before recovering a P0 viral stock.

TFIIH Extraction and expression tests

Cell pellet from a 50 ml *Sf*9 suspension culture infected with 1 ml P1 virus and incubated at 27 °C for 72 h was harvested by centrifugation and resuspended in 5 ml lysis buffer phosphate-buffered saline (PBS) 163 mM NaCl plus a protease inhibitor cocktail (Roche) per gram of pellet. The resuspended cells were subsequently lysed with a douce homogenizer (Fisher Scientific) (20 strokes with a tight pestle, performed on ice) and centrifuged at 13000 rpm for 30 min at 4 °C (centrifuge 5415, Eppendorf). The cleared supernatant was then transferred to a fresh tube and analysed in a BioSprint station (Qiagen): 500 μl supernatant and 10 μl magnetic Ni beads (Promega) were loaded into the first column of the BioSprint sample tray. After binding, the beads were washed twice with 500 μl buffer PBS 163 mM NaCl 30 mM imidazole, and bound proteins were eluted in 50 μl buffer PBS 163 mM NaCl 500 mM imidazole. The eluted fraction was further analysed in an SDS-PAGE gel to determine TFIIH Core expression levels.

Insect cell infection

Virus amplification was carried out in *Sf*9 suspension cultures. Infections were carried out in High Five™ suspension cultures with a starting cell density of $8x10^5$ cells/ml in a total volume of 250 ml Express Five® Serum-Free Media (ThermoFisher Scientific). Each 250 ml suspension was infected with 3 ml of a P1 – P2 viral stock and incubated for 48 h in a sterile shaker flask (Corning) at 27 °C with 110 rpm orbital shaking. Afterwards, cell pellet and medium containing viral stock were collected by centrifugation (CF20 centrifuge, Awel) at 1000 rpm for 5 min and kept at -80 °C until tested. Viral stocks were kept at 4 °C for a maximum of 4 months.

Immobilised metal ion affinity chromatography - Talon® Superflow™ cobalt resin

A Talon® Superflow™ cobalt resin (GE Healthcare), which was prepared for purification following the manufacturer's indications, was used to maximize the contact between the sub-complex and the metal ion affinity matrix. Cleared supernatant extracted as previously described was added to 5 ml resin and incubated at 4 °C for 1 h with rotation. Resin and supernatant were then transferred into an empty 10 ml gravity-flow column previously

prepared as indicated by the manufacturer (Thermo Scientific). Resin was allowed to settle for 15-30 min, and after that time the supernatant was eluted. This step was repeated until all of the resin had been packed and all the supernatant had been eluted. A washing step was performed twice using ten times resin bed volume of buffer A (PBS pH 7.5 163 mM NaCl 30 mM imidazole 10% glycerol); the buffer was incubated at room temperature for 10 min in the column, then slowly eluted. The TFIIH Core sub-complex was eluted with 5-10 times resin bed volume of buffer B (PBS pH 7.5 163 mM NaCl 300 mM imidazole 10% glycerol): first 1 ml buffer was eluted and collected. The rest of the volume was incubated with the resin at room temperature for 10 min, then eluted and collected in 1 ml fractions. The different fractions were analysed by SDS-PAGE.

Heparin chromatography

Fractions of interest collected after IMAC were first concentrated, then diluted with buffer PBS pH 7.5 to reduce the amount of NaCl in the sample down to approximately 200 mM before loading it into a HiTrap™ Heparin HP column (GE Healthcare). The column was washed with buffer A (PBS pH 7.5 5% glycerol), before eluting bound proteins in a gradient of salt created by combining buffer A and buffer B (PBS pH 7.5 863 mM NaCl 5% glycerol) in different proportions. The purification was carried out using a BioLogic DuoFlow™ system (Bio-Rad).

TEV Cleavage

Fractions containing the TFIIH Core sub-complex collected after the heparin chromatography were pooled together and an excess (500 μl) of 1 mg / ml TEV protease was added to the sample. Cleaving was allowed overnight at room temperature, with gentle rolling.

Immobilised metal ion affinity chromatography - HisTrap™ FF column

The TEV-cleaved sample was subsequently loaded into a 1 ml HisTrap™ Fast Flow (FF) column (GE Healthcare) and purified using a BioLogic DuoFlow™ system (Bio-Rad). After washing the column with buffer A (PBS pH 7.5 163 mM NaCl 30 mM imidazole), bound proteins were eluted in an imidazole gradient created by buffer B (PBS pH 7.5 163 mM NaCl 300 mM imidazole). Purified fractions containing the TFIIH Core sub-complex were pooled together and concentrated using an Amicon® concentrator with a 30K cutoff (Millipore).

Size exclusion chromatography – Superose® 6 column

The concentrated sample obtained in the previous IMAC step was loaded into a size exclusion Superose® 6 HR 10/30 column (GE Healthcare) and eluted with PBS pH 7.5 63 mM NaCl. The purification was carried out using an Äkta system (GE Healthcare). Fractions containing the TFIIH Core sub-complex were pooled together, concentrated and stored at – 80 °C.

DNA substrate preparation

Oligonucleotides were purchased from IDT and resuspended in 50 mM Tris pH 7.5 to a final concentration of 500 μM. The sequences used were as follows:

5'-AGCTACCATGCCTGCACGAATTAAGCAATTCGTAATCATGGTCATAGCT/Dabcyl/-3'
and 5'-Cy3/AGCTATGACCATGATTACGAATTGCTTGGAATCCTGACGAACTGTAG-3'

A double stranded substrate was generated by annealing in a 50 μl reaction containing buffer 50 mM Tris pH 7.5 100 mM NaCl and the two single strands to a final concentration of 20 μM. The mixture was heated in a water bath at 90 °C for 5 min and left to slowly cool down overnight. A final concentration of 3% ficoll was added to the reaction now containing the double-stranded substrate, which was subsequently purified in a 10% polyacrylamide native gel, run in Tris – Borate – Ethylenediaminetetraacetic acid (EDTA) (TBE) buffer 10 mM NaCl at 4 °C, 120 V for 3-4 h, protected from light. The band of interest was cut, placed in an Eppendorf tube and slightly crushed, then covered with buffer 50 mM Tris buffer pH 7.5 50 mM NaCl. The tube was incubated at 4 °C overnight with rocking to promote diffusion of the substrate from the acrylamide towards the buffer. Afterwards, the buffer was separated from the acrylamide and the double-stranded substrate was precipitated with 10% volume sodium acetate 3 M pH 5.2 and 250% volume chilled EtOH 100% and incubated at -20 °C overnight. Finally, the reaction was centrifuged at 15000 rpm at 4 °C for 30 min (centrifuge 5424, Eppendorf), the supernatant was immediately removed, and the pelleted DNA substrate was resuspended in an appropriate volume of 50 mM Tris pH 7.5 50 mM NaCl buffer.

Fluorescence-based helicase assays

The unwinding activity of TFIIH Core was evaluated in a fluorescence helicase assay using a splayed duplex substrate carrying both a Cy3 fluorophore and a quencher strategically placed to suppress the dye. The quencher will be removed upon unwinding by the sub-complex, thus permitting the detection of the fluorescence emitted by Cy3 (Kuper *et al.*, 2012). The helicase activity assay was carried out in buffer 20 mM HEPES pH 7.0 50 mM NaCl containing 0.1

mg/ml BSA, 1 mM $MgCl_2$, 25 nM splayed duplex substrate and 200 nM TFIIH Core sub-complex, performed at 25 °C. The reaction mixture was then incubated at room temperature for 10 min before being added to a cuvette containing 1 mM ATP, previously incubated at room temperature for 10 min, too. Activity was measured for 15 min in a Varian Cary Eclipse fluorescence spectrophotometer using the Kinetics application included in the software package provided with the instrument. The Cy3 dye was excited at a wavelength of 547 nm, and emitted fluorescence was collected at a wavelength of 565 nm. The obtained fluorescence emission intensity values were plotted against TFIIH Core concentration.

RESULTS & DISCUSSION

Initial efforts at TFIIH Core cloning and expression

The first consideration we took into account when we first approached the cloning of the TFIIH Core sub-complex was the number and position of the purification tags. We designed a multi-purpose tag consisting of 6xHis-V5-TEV that would allow us to purify our proteins in an IMAC (6xHis), to perform Western blot detection and immunoprecipitation of the complex, if required, through the V5 epitope, and to remove the tag from the purified protein by cleaving at the TEV site.

For our first attempt we opted for placing this tag at the N-terminal end of subunit p52, due to its strategic position at the base of the ring, and another tag at the N-terminal end of the XPD helicase which, as the bridge between the Core and CAK sub-complexes (Schultz *et al.*, 2000) (Greber *et al.*, 2017), is the most flexible subunit within TFIIH. To facilitate the cloning of the sub-complex the genes for subunits p52 and p8, and p44 and p34, were designed as a single unit, with each pair of genes separated by a 2A-like auto-cleavable peptide (figure 2 (A)) (Luke *et al.*, 2009), meaning that only five cloning reactions were needed to obtain the 7-gene construct. Unfortunately, this construct did not offer the expected results, showing low expression levels and a worrying loss of subunit XPD, whose coordinated action with subunit XPB is responsible for the opening and extension of the repair bubble to allow the following repair factors access to the damaged site. A Western blot analysis performed with a V5 primary antibody and an IRDye® 800CW goat anti-mouse IgG secondary antibody (Li-COR) showed an acceptable cleavage efficiency for the F2A auto-cleavable peptide, with 87% of the *p52/F2A/p8* gene actively translated into independent subunits p52 and p8 (figure 2B).

Our experiments showed no evidence of uncleaved construct p44/T2A/p34.

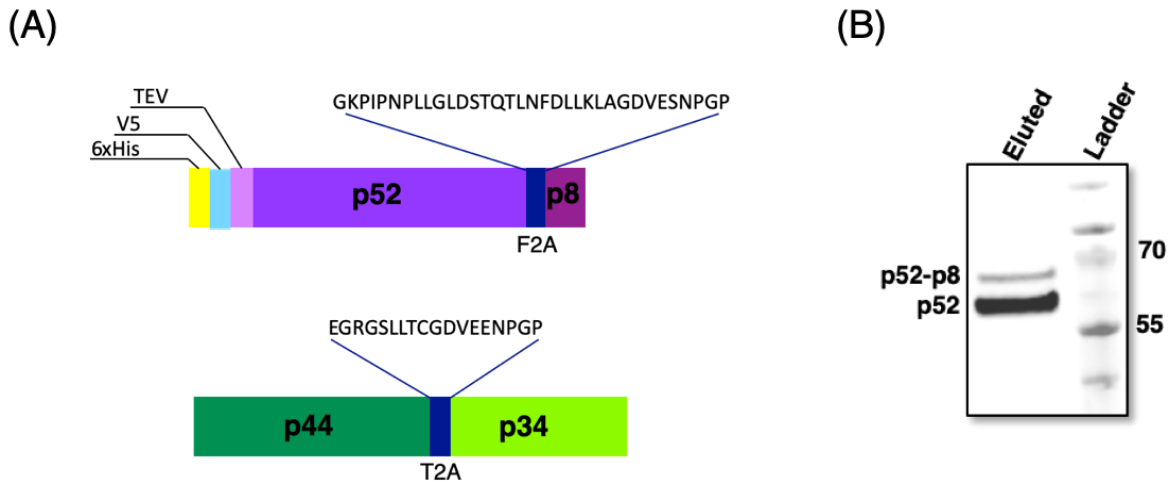

Figure 2: Cleaving efficiency of the F2A auto-cleavable sequence is close to 90%.

Cartoon representation showing TFIIH Core constructs p52/V5/F2A/p8 and p44/T2A/p34 (A). Each pair of genes was cloned as a single unit thanks to the F2A and T2A auto-cleavable peptides joining the two selected genes. These 2A-like sequences later allowed the translation of all four genes as individual proteins. (B) Western blot performed with an anti-V5 primary antibody showing a BioSprint purification test for an infection with a P0 recombinant virus containing construct TFIIH Core. Although a small amount of the p52/F2A/p8 construct remains (13%), cleaving efficiency of the F2A sequence is 87%. Ladder sizes are indicated in kDa.

Scaling-up and optimisation of the protocol for the purification of this TFIIH Core construct finally resulted in a procedure that included an IMAC step, followed by size-exclusion chromatography and finally cation exchange chromatography, but the impossibility of retaining sufficient amounts of subunit XPD in our sample led us to reconsider our strategy regarding the position of the purification tags, and several other constructs were produced, always maintaining a tag at subunit p52 due to its central position at the base of the ring, while varying the position of the second tag. Many of the intermediate constructs required for the cloning of these new TFIIH Core sub-complexes had already been obtained during the cloning process of our original 7-gene plasmid. The characteristics of the MultiBac™ system allow the use of these intermediate multi-gene plasmids (Bieniossek *et al.*, 2012), which meant that each new 7-gene construct was obtained in only two or three cloning steps. This advantage considerably reduced the time scale necessary to obtain a new full TFIIH Core sub-complex.

Steps towards the purification of an intact TFIIH Complex

A common characteristic to our first TFIIH Core constructs was their inability to retain subunit XPD as part of the sub-complex. A Western blot performed with a human/rat/mouse XPD (R&D Systems) primary antibody and an IRDye® 800CW donkey anti-goat IgG (Li-COR) secondary antibody proved that XPD was present in our cell extract, while also confirming that major losses occurred at every step, finally recovering a mere 0.1% XPD by the end of the purification process (figure 3). These results lead us to think that XPD was more susceptible to dissociation from the sub-complex after extraction, and this was further aggravated by the fact that binding to the three different chromatography columns employed in the purification process was not as strong for the helicase as it was for the rest of the TFIIH Core subunits.

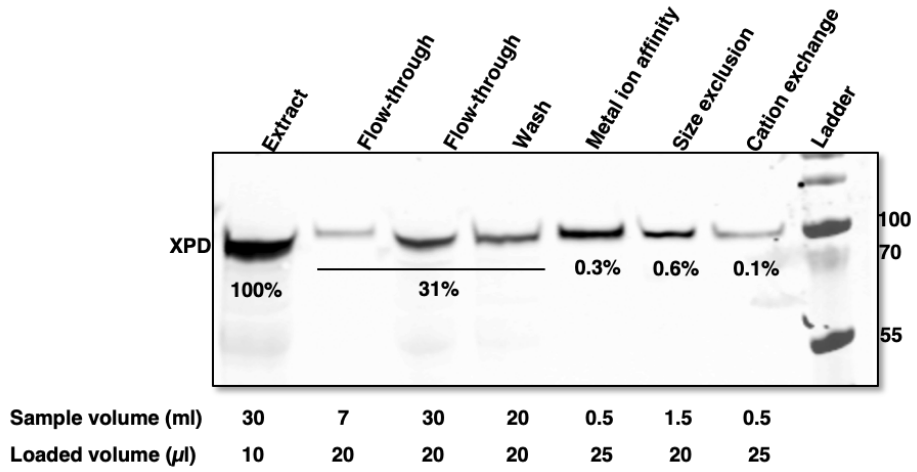

Figure 3: Subunit XPD is lost in the purification of TFIIH Core.

Western blot performed with an anti-XPD primary antibody showing XPD after every purification stage included in our initial protocol. A considerable amount (31%) of the helicase was lost due to inefficient binding at the initial stages (flow-through, washing), and the losses continued through the three different chromatographic steps, to finally retrieve only 0.1% of the initial XPD available in the extract. Ladder sizes are indicated in kDa.

Several different strategies were approached to try to solve this problem, including: 1) extraction and purification of the sub-complex under anaerobic conditions that would prevent the iron-sulfur (4FeS) cluster present in XPD (Rudolf *et al.*, 2006) (Liu *et al.*, 2008) *(Fan et al.*, 2008) (Wolski *et al.*, 2008) from being oxidized; 2) incorporation of a series of detergents to our extraction buffer (Pullara *et al.*, 2013); 3) expression and purification of XPD as an individual protein to later perform a co-infection with a virus containing a 6-subunit TFIIH Core,

thus obtaining the complete 7-subunit sub-complex; and 4) cloning, expression and purification of a 7-subunit TFIIH Core with the p52 subunit carrying our 6xHis-V5 epitope-TEV tag and XPD carrying a Twin-Strep-tag®, with a reported affinity in the pM range for the Strep-Tactin®XT Superflow® purification resin. However, none of these approaches offered an acceptable improvement of XPD yield in our purifications, and although the construct carrying the Twin-Strep-tag® finally allowed us to obtain helicase XPD in similar amounts as TFIIH Core subunits XPB, p62, p52, p44, p34 and p8 (confirmed by MS, data not shown), the tag on XPD tended to separate the XPD protein from the rest of the Core complex.

Cloning and expression of the final TFIIH Core construct

A 7-gene TFIIH Core construct carrying our 6xHis-V5-TEV tag at the N-terminal end of subunits p52 and p62 (chosen for its central position within the ring-like structure) finally offered encouraging expression levels for all seven subunits, helicase XPD included (figure 4).

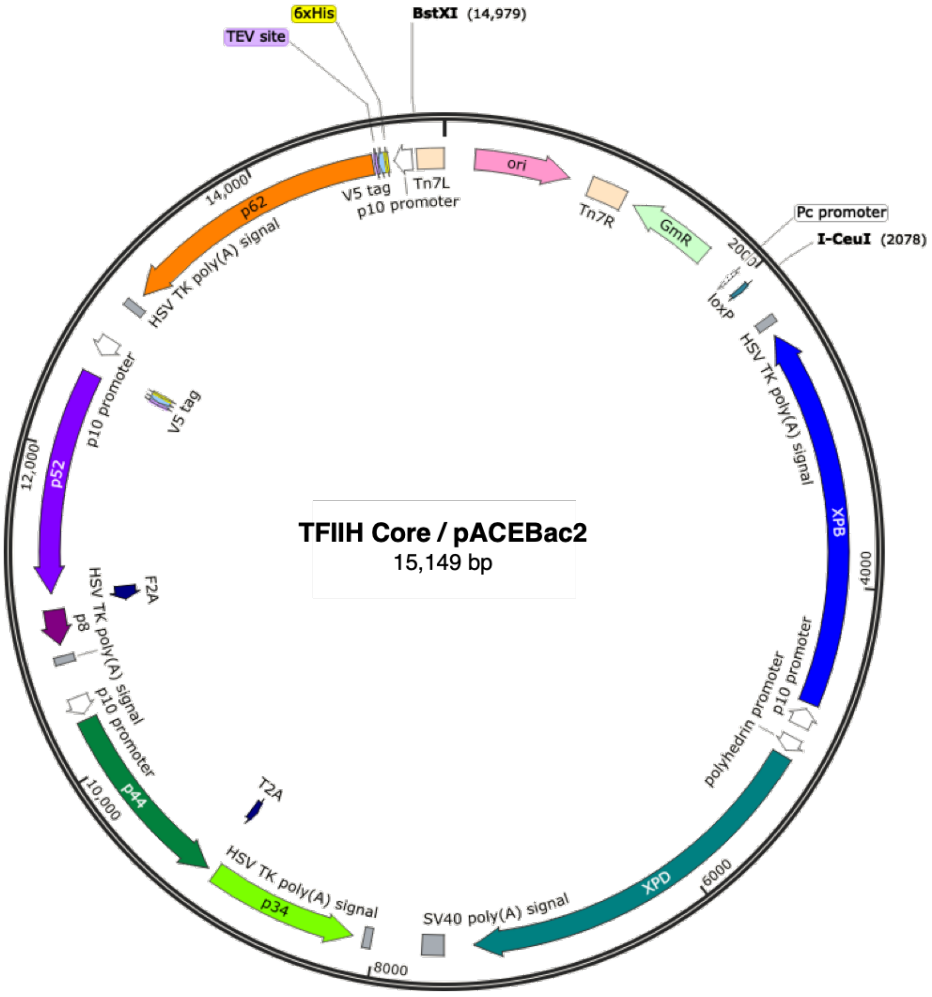

Figure 4: Cartoon representation of the TFIIH Core / pACEBac2 construct, with subunits p52 and p62 carrying a 6xHis-V5-TEV purification tag at their N-terminal ends.

The 15,149bp plasmid contains genes *XPB*, *p62*, *p52/F2A/p8* and *p44/T2A/p34,* each one of them under the control of a p10 promoter, and gene *XPD* under the control of a polyhedrin promoter. The plasmid also contains a *I-Ceu*I/*Bst*XI multiplication module, the Tn7L/Tn7R transposition elements for later transposition of the 7-gene cassette into the baculoviral genome, and a gentamicin (GmR) antibiotic resistance marker.

The seven TFIIH Core genes had originally been cloned into individual pACEBac2 or pACEBac1 plasmids, to subsequently be cloned one by one into growing multi-gene constructs using the I-*Ceu*I/*Bst*XI multiplication module as previously described (figure 5, step 1). These intermediate constructs were verified by restriction with different combinations of several enzymes, comparing the band pattern offered by the restricted plasmid in a 1% agarose gel with the one predicted *in silico* by the NEBcutter® tool (New England Biolabs). The final 7-gene construct was further analysed by PCR and ultimately, all TFIIH Core genes were sequenced (GATC Biotech) using the primers listed in table 1.

The recombinant plasmid containing all seven TFIIH Core genes was next transposed into the baculoviral genome as previously described (figure 5, step 2). The transposition procedure became less and less efficient as the transposed constructs grew bigger, and our protocol had to be adapted accordingly, increasing both the incubation and recovery times and using freshly made chemically competent cells for each reaction. Even so, the number of white colonies obtained per reaction was frequently still lower than five.

The successfully transposed recombinant plasmid, or bacmid, was purified and transfected into a monolayer of *Sf*9 cells (figure 5, step 3). The P0 virus recovered from this transfection was used to infect a 50 ml suspension culture, which was harvested and tested for TFIIH Core expression after 60 hours of incubation. Large-scale infections for over-expression of the sub-complex were initially performed in *Sf*9 cells but our protocol was later adapted to perform the infections in High Five ™ cells, which offered a considerably increased yield (figure 5, steps 4 and 5).

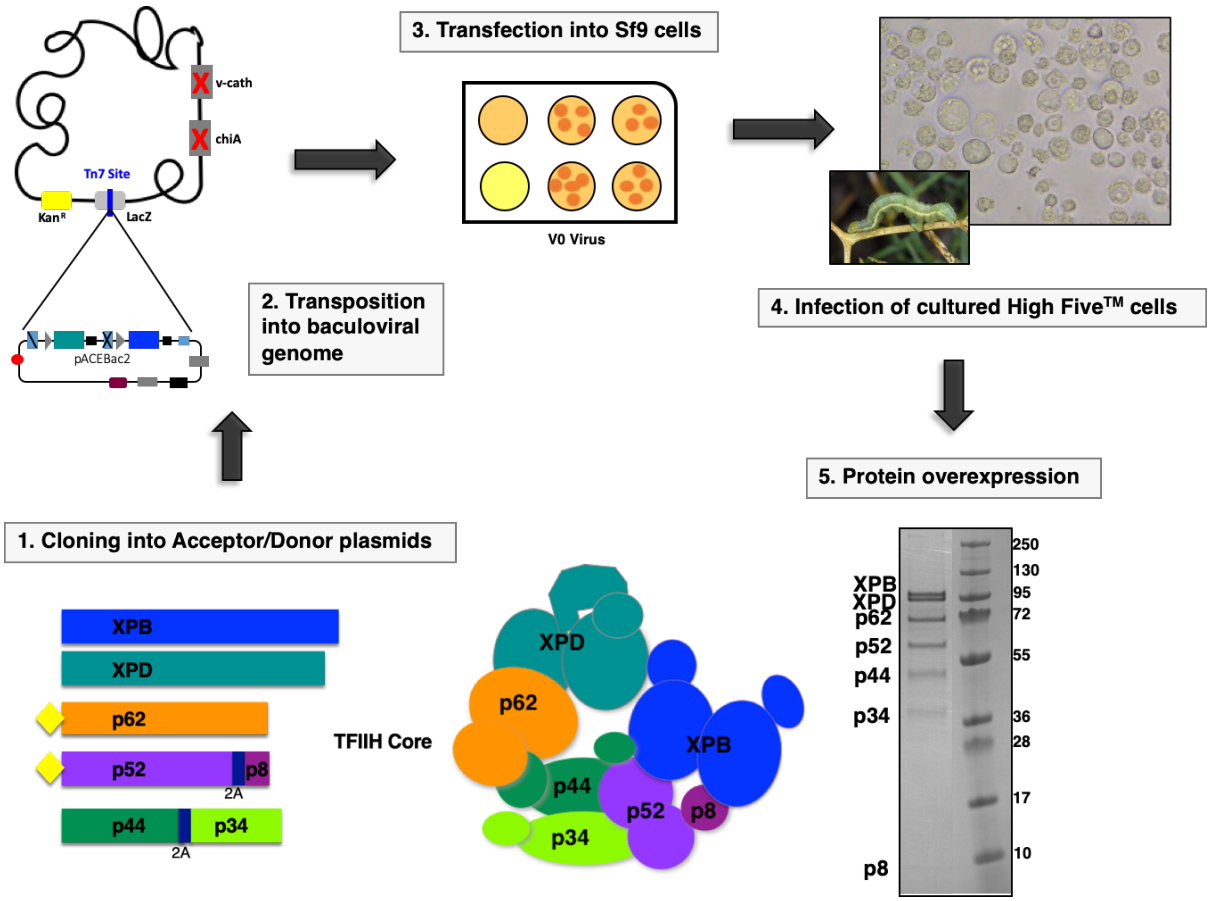

Figure 5: Overview of the cloning and expression process for the TFIIH Core sub-complex.

Summary of all the steps required for the cloning and expression of the TFIIH complex using the MultiBac™ system. The 6xHis-V5-TEV purification tag is represented by a yellow diamond. In step 5, the size of the protein markers are indicated in kDa.

Purification of TFIIH Core

Initial expression levels for every TFIIH Core sub-complex construct were verified in a 50 ml BioSprint purification test (see Materials and Methods). This quick and simple purification method allowed us to evaluate the suitability of each one of our constructs, discarding those that did not offer the expected results, before moving forward with the large-scale purification process, thus saving us a considerable time and effort. This method also permitted us to compare the expression efficiency of *Sf*9 cells vs High Five™ cells (figure 6), which led us to change our large-scale expression system to the latter. This change doubled the yield of our purifications from ~ 1 mg TFIIH Core sub-complex per 2 L culture to ~ 1 mg per every litre of culture.

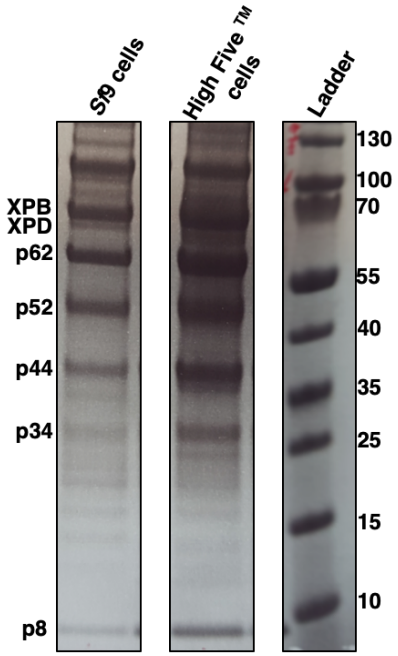

Figure 6: TFIIH Core extraction from *Sf*9 cells vs High Five™ cells.

Coomassie-stained gels showing two BioSprint purification tests (50 ml, infected with a P1 virus) for the TFIIH Core sub-complex extracted from *Sf*9 cells and High Five™ cells. Expression levels from High Five™ cells are significantly higher than those from *Sf*9 cells and performing our large-scale infections in High Five™ instead of *Sf*9 cells doubled our TFIIH Core yield. Ladder sizes are indicated in kDa.

The scaling up of our infections in High Five™ cells for the sub-complex carrying two 6xHis-V5-TEV tags at the N-terminal end of subunits p52 and p62 and the troubleshooting of our initial purification protocol in aerobic conditions further helped to solve the issues affecting XPD. Substitution of the nickel column originally employed for a Talon® Superflow™ cobalt resin in the first IMAC step resulted in a higher recovery of the sub-complex, as purification with a loose resin instead of the pre-packed matrix present in the HisTrap™ FF column we had been using so far allowed us to increase the binding time.

Afterwards, the sample was further cleaned through a HiTrap™ Heparin HP column, and fractions of interest obtained from this step were pooled together for overnight TEV-cleaving of the two 6xHis-V5-TEV tags. The TEV protease was removed from the sample through a HisTrap™ FF column, and the TEV-free fractions containing the Core sub-complex were finally loaded into a Superose™ 6 HR 10/30 column. Fractions of interest were analysed in a SDS-PAGE gel after every purification step, showing a pure complex containing stoichiometric amounts of subunits XPD, XPB, p62, p52, p44, p34 and p8 (figure 7). The scaling up and

optimisation of the expression and purification procedures finally allowed us to obtain a yield to 1 mg of pure sub-complex for every 1 L culture. We consistently observe in our purification gels a band of approximately 180 kDa which was identified by mass spectrometry analysis as the FACT complex subunit SPT16, a chaperone originating from our insect cell expression system. This protein is not expected to interfere in biochemical assays of the TFIIH complex.

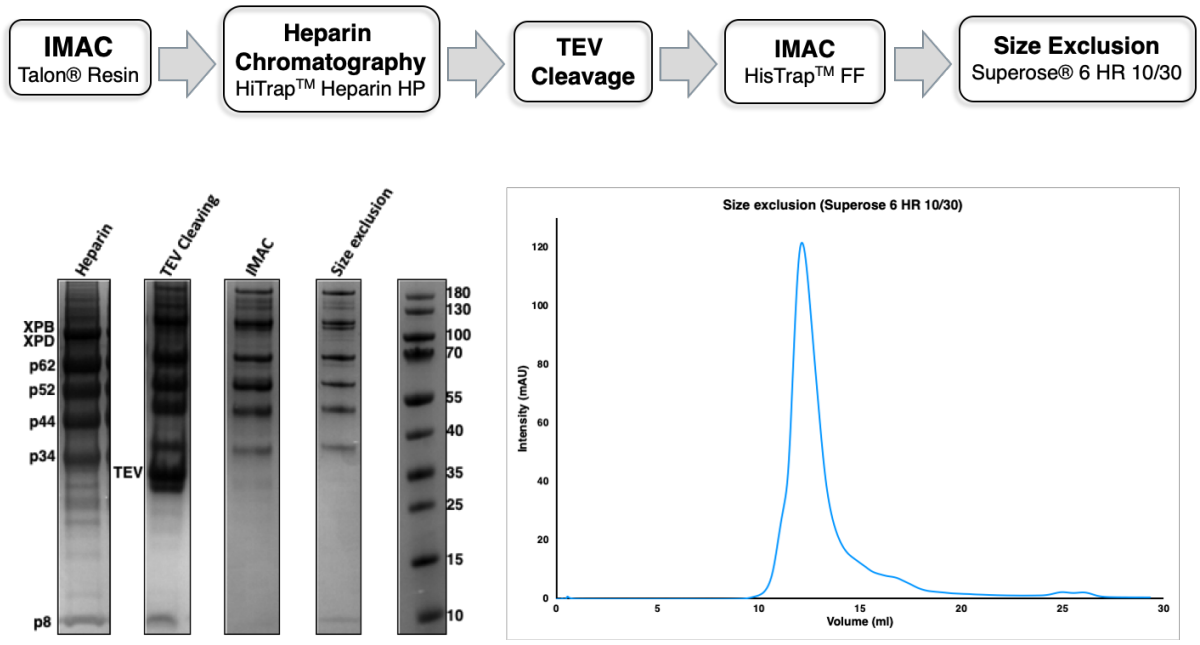

Figure 7: Purification of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.

Coomassie-stained gels showing peak fractions for the different chromatographic steps followed in the optimised purification of the TFIIH Core sub-complex with subunits p52 and p62 tagged at their N-terminal end with 6xHis-V5-TEV: heparin chromatography (HiTrapTM Heparin HP column), followed by overnight TEV-cleaving of the sample, then ion affinity (HisTrap™ FF column), and finally a size exclusion chromatography (Superose® 6 HR 10/30 column). Ladder sizes are indicated in kDa.

The purified TFIIH Core sub-complex shows 5' unwinding activity in a fluorescence-based helicase assay

The TFIIH Core sub-complex possesses two helicases of opposite polarities: XPB (3' to 5') and XPD (5' to 3'). However, XPB has recently been redefined as a translocase rather than a helicase in the canonical sense of the classification (Fishburn *et al.*, 2015) (He *et al.,* 2016) (Schilbach *et al.*, 2017). We demonstrated the 5' to 3' helicase activity of the sub-complex in a fluorescence-based unwinding assay with a splayed duplex substrate that includes a Cy3 dye, whose fluorescence is supressed by a quencher in the duplex (figure 8) (Kuper *et al.*,

2012). Upon binding of the TFIIH Core sub-complex, the XPD helicase will unwind the substrate in an ATP-dependent reaction, effectively removing the quencher and allowing the detection of the fluorescence emitted by the Cy3 dye. Unwinding of the substrate was increased by nearly 50% after the final size exclusion chromatography compared to unwinding of the sub-complex after the previous IMAC step, presumably due to the increased purity of the sample and a more accurate estimation of the TFIIH Core concentration. A TFIIH Core sub-complex with XPD carrying a K48R mutation, which nearly abolishes the unwinding capacity of the sub-complex, shows only background activity, potentially caused by the opening of the substrate due to XPB translocation.

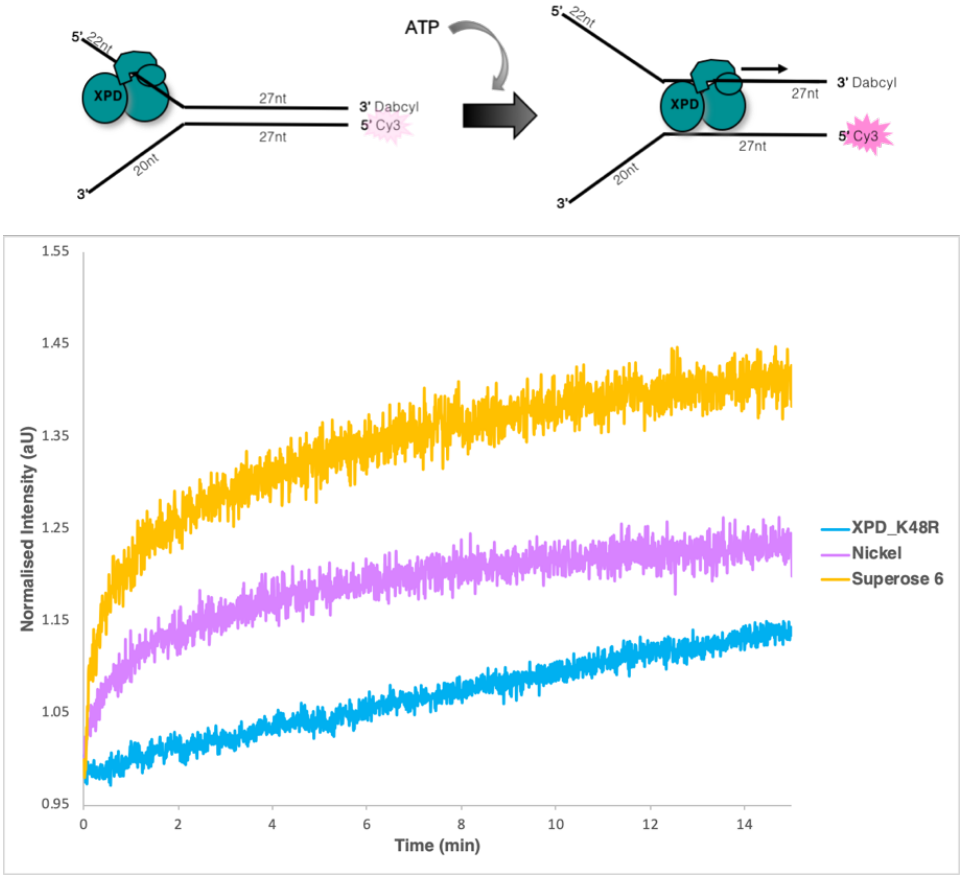

Figure 8: Unwinding of a splayed duplex substrate proves TFIIH Core 5' helicase activity. Fluorescence helicase assay showing unwinding of a splayed duplex substrate SD (25 nM) by the TFIIH Core sub-complex (250 nM). A more thorough elimination of contaminants and more exact determination of the TFIIH Core concentration probably caused the considerable increase of the unwinding capacity of the sample after the Superose 6 column (orange line), compared to the level of activity observed for the sample after the nickel column (purple line). The 5' to 3' unwinding capacity of our TFIIH Core sub-complex was further verified in a helicase assay performed with a TFIIH Core sub-

complex with XPD carrying a K48R mutation, which only showed a residual activity potentially attributable to XPB (blue line).

CONCLUSIONS

We have cloned and expressed the ~ 350 kDa, 7-subunit TFIIH Core sub-complex using the MultiBac™ system, a BEV specifically designed for the production of eukaryotic multi-protein complexes that allows the expression of every subunit in equimolar amounts (Berger *et al.*, 2004), as opposed to infection with several different viral stocks as it has been traditionally done. The cloning of the sub-complex was simplified by the introduction of 2A-like sequences (Luke *et al.*, 2009), designing the synthetic genes for subunits p52/p8 and p44/p34 as a single unit, with each pair of genes separated by these auto-cleavable peptides – hence reducing the number of cloning steps necessary to obtain the multi-gene construct. Cleaving efficiency was shown to be close to 90%. This approach is easily adaptable and might be highly beneficial in the cloning of many different multi-gene constructs.

Purification of the sub-complex was achieved after establishing a simple and efficient 5-step IMAC – heparin – TEV cleaving – IMAC - size exclusion chromatography as a standardized protocol that allows the extraction of up to 1 mg of pure sub-complex from a 2 L culture. The purification process was aided by the introduction of a 6xHis-V5-TEV tag at the N-terminal end of subunits p52 and p62. This small, versatile tag designed by our lab has a triple purpose: purification of the desired peptide through its His tail, Western blot detection by means of its V5 epitope and removal of the tag by cleaving at its TEV site. Taken as a whole, the method described here permits the obtaining of a high-yield, pure and active multi-subunit complex, suitable for both biochemical and structural studies.

Funding

This work was supported by the BBSRC [grant number BB/J01446X/1] and [grant number BB/R015570/1]

BIBLIOGRAPHY

Abdulrahman, W., Iltis, I., Radu, L., Braun, C., Maglott-Roth, A., Giraudon, C., … Poterszman, A. (2013). ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(8), E633-42. https://doi.org/10.1073/pnas.1213981110

Assenberg, R., Wan, P. T., Geisse, S., & Mayr, L. M. (2013). Advances in recombinant protein expression for use in pharmaceutical research. *Current Opinion in Structural Biology*, *23*(3), 393–402. https://doi.org/10.1016/j.sbi.2013.03.008

Berger, I., Fitzgerald, D. J., & Richmond, T. J. (2004). Baculovirus expression system for heterologous multiprotein complexes. *Nature Biotechnology*, *22*(12), 1583–1587. https://doi.org/10.1038/nbt1036

Bieniossek, C., Imasaki, T., Takagi, Y., & Berger, I. (2012). MultiBac: Expanding the research toolbox for multiprotein complexes. *Trends in Biochemical Sciences*, *37*(2), 49–57. https://doi.org/10.1016/j.tibs.2011.10.005

Bieniossek, C., Nie, Y., Frey, D., Olieric, N., Schaffitzel, C., Collinson, I., … Berger, I. (2009). Automated unrestricted multigene recombineering for multiprotein complex production. *Nature Methods*, *6*(6), 447–450. https://doi.org/10.1038/nmeth.1326

Bootsma, D., & Hoeijmakers, J. H. J. (1993). Engagement with transcription. *Nature Letters*, *363*.

Coin, F., Marinoni, J. C., Rodolfo, C., Fribourg, S., Pedrini, A. M., & Egly, J. M. (1998). Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nature Genetics*, *20*(2), 184–188. https://doi.org/10.1038/2491

Compe, E., & Egly, J.-M. (2012). TFIIH: when transcription met DNA repair. *Nature Reviews Molecular Cell Biology*, *13*(7), 476–476. https://doi.org/10.1038/nrm3377

Dubaele, S., De Santis, L. P., Bienstock, R. J., Keriel, A., Stefanini, M., Van Houten, B., & Egly, J. M. (2003). Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Molecular Cell*, *11*(6), 1635–1646. https://doi.org/10.1016/S1097-2765(03)00182-5

Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., … Tainer, J. A. (2008). XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell*, *133*(5), 789–800.

https://doi.org/10.1016/j.cell.2008.04.030

Feaver, W. J., Svejstrup, J. Q., Bardwell, L., Bardwell, A. J., Buratowski, S., Gulyas, K. D., … Kornberg, R. D. (1993). Dual roles of a multiprotein complex from *S. cerevisiae* in transcription and DNA repair. *Cell*, *75*(7), 1379–1387. https://doi.org/10.1016/0092-8674(93)90624-Y

Fishburn, J., Tomko, E., Galburt, E., & Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(13), 3961–3966. https://doi.org/10.1073/pnas.1417709112

Flores, O., Lu, H., & Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(23), 9158–9162. https://doi.org/10.1073/pnas.87.23.9158

Giglia-Mari, G., Coin, F., Ranish, J. A., Hoogstraten, D., Theil, A., Wijgers, N., … Vermeulen, W. (2004). A new, tenth subunit TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nature Genetics*, *36*(7), 714–719. https://doi.org/10.1038/ng1387

Greber, B. J., Nguyen, T. H. D., Fang, J., Afonine, P. V., Adams, P. D., & Nogales, E. (2017). The cryo-electron microscopy structure of human transcription factor IIH. *Nature*. https://doi.org/10.1038/nature23903

Hanawalt, P. C., & Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Reviews. Molecular Cell Biology*, *9*(12), 958–970. https://doi.org/10.1038/nrm2549

He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., & Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. *Nature*, *533*(7603), 359–365. https://doi.org/10.1038/nature17970

Kokic, G., Chernev, A., Tegunov, D., Dienemann, C., Urlaub, H., & Cramer, P. (2019). Structural basis of TFIIH activation for nucleotide excision repair. *Nature Communications*, *10*(1), 1–9. https://doi.org/10.1038/s41467-019-10745-5

Kuper, J., Wolski, S. C., Michels, G., & Kisker, C. (2012). Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. *EMBO Journal*, *31*(2), 494–502. https://doi.org/10.1038/emboj.2011.374

Lehmann, A. R. (2003). DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *Biochimie*, *85*(11), 1101–1111.

https://doi.org/10.1016/j.biochi.2003.09.010

Liu, H., Rudolf, J., Johnson, K. A., McMahon, S. A., Oke, M., Carter, L., … White, M. F. (2008). Structure of the DNA repair helicase XPD. *Cell*, *133*(5), 801–812. https://doi.org/10.1016/j.cell.2008.04.029

Luke, G. A., Escuin, H., Felipe, P. De, & Ryan, M. D. (2009). 2A to the fore – research, technology and applications. *Biotechnology and Genetic Engineering Reviews*, *26*(1), 223–260. https://doi.org/10.5661/bger-26-223

Oksenych, V., De Jesus, B. B., Zhovmer, A., Egly, J. M., & Coin, F. (2009). Molecular insights into the recruitment of TFIIH to sites of DNA damage. *EMBO Journal*, *28*(19), 2971–2980. https://doi.org/10.1038/emboj.2009.230

Peng, Y., Wang, H., Santana-santos, L., & Kisker, C. (2011). *Chemical Carcinogenesis*. (T. M. Penning, Ed.). Springer New York. https://doi.org/10.1007/978-1-61737-995-6

Pullara, F., Guerrero-Santoro, J., Calero, M., Zhang, Q., Peng, Y., Spåhr, H., … Levine, A. S. (2013). A general path for large-scale solubilization of cellular proteins: From membrane receptors to multiprotein complexes. *Protein Expression and Purification*, *87*(2), 111–119. https://doi.org/10.1016/j.pep.2012.10.007

Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S., & White, M. F. (2019). Investigation of the cyclic oligoadenylate signaling pathway of type III CRISPR systems. *Methods in Enzymology*, *616*, 191–218. https://doi.org/10.1016/bs.mie.2018.10.020

Rudolf, J., Makrantoni, V., Ingledew, W. J., Stark, M. J. R., & White, M. F. (2006). The DNA repair helicases XPD and FancJ have essential ion-sulfur domains. *Molecular Cell*, *23*(6), 801–808. https://doi.org/10.1016/j.molcel.2006.07.019

Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., & Cramer, P. (2017). Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature*, *551*(7679), 204–209. https://doi.org/10.1038/nature24282

Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., & Egly, J. M. (2000). Molecular structure of human TFIIH. *Cell*, *102*(5), 599–607. https://doi.org/S0092-8674(00)00082-9 [pii]

Svejstrup, J. Q., Wang, Z., Feave, W. J., Wu, X., Bushnell, D. A., Donahue, T. F., … Kornberg, R. D. (1995). Different forms of TFIIH for transcription and DNA repair: Holo-TFIIH and a nucleotide excision repairosome. *Cell*, *80*(1), 21–28. https://doi.org/10.1016/0092-8674(95)90447-6

Tirode, F., Busso, D., Coin, F., & Egly, J.-M. (1999). Reconstitution of the Transcription Factor TFIIH. *Molecular Cell*, *3*, 87–95. https://doi.org/10.1016/S1097-2765(00)80177-X

Tomkinson, A. E., Vijayakumar, S., Pascal, J. M., & Ellenberger, T. (2006). DNA ligases: Structure, reaction mechanism, and function. *Chemical Reviews*, *106*(2), 687–699. https://doi.org/10.1021/cr040498d

Wolski, S. C., Kuper, J., Hänzelmann, P., Truglio, J. J., Croteau, D. L., Van Houten, B., & Kisker, C. (2008). Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biology*, *6*(6), 1332–1342. https://doi.org/10.1371/journal.pbio.0060149

SUPPLEMENTARY MATERIAL

TFIIH Core / pACEBAc2 plasmid sequence (15,149 bp)

accggttgacttgggtcaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctagg
tgaagatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaa
aggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccgg
atcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagtt
aggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgata
agtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacac
agcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaag
ggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacg
cctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatgga
aaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattgacttg
ggtcgctcttcctgtggatgcgcagatgccctgcgtaagcgggtgtgggcggacaataaagtcttaaactgaacaaaatagatct
aaactatgacaataaagtcttaaactagacagaatagttgtaaactgaaatcagtccagttatgctgtgaaaaagcatactggact
tttgttatggctaaagcaaactcttcattttctgaagtgcaaattgcccgtcgtattaaagaggggcgtggccaagggcatgtaaag
actatattcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcg
gtacttgggtcgatatcaaagtgcatcacttcttcccgtatgcccaactttgtatagagagccactgcgggatcgtcaccgtaatctg
cttgcacgtagatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtggcaatgccctgcctccg
gtgctcgccggagactgcgagatcatagatatagatctcactacgcggctgctcaaacttgggcagaacgtaagccgcgagag
cgccaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccgaggtaatcggagtccg
gctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcatggatttgacttggtcag
ggccgagcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagggcgactgccctgctgcgtaacatcgttgct

gctgcgtaacatcgttgctgctccataacatcaaacatcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcataga

ctgtacaaaaaaacagtcataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggttctggacc

agttgcgtgagcgcatacgctacttgcattacagtttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgtttccac

ggtgtgcgtcacccggcaaccttgggcagcagcgaagtcgccataacttcgtatagcatacattatacgaagttatctgtaactat

aacggtcctaaggtagcgagtttaaacgtacccgtagtggctatggcagggcttgccgccccgacgttggctgcgagccctggg

ccttcacccgaacttgggggttggggtggggaaaaggaagaaacgcgggcgtattggtcccaatggggtctcggtggggtatc

gacagagtgccagccctgggaccgaaccccgcgtttatgaacaaacgacccaacacccgtgcgttttattctgtcttttattgccg

tcatagcgcgggttccttccggtattgtctccttccgtgtttcagttagcctcccccatctcccggtacCTCATTTCCTAAAGC

GCTTGAAGAGCGGGTGTACATGTTTGCTGGGCGCCTTGCTCCGCGATGAGTGGTACTC

CATGTACACAGTGTCGTCGGCCCCAGACATAGAACTCATGGTGCCAAAGCGCCGAGAT

GCCTGGCTGGATCTGGAGCCAAATTCCCCAGCCACCACCTCCTCCTCGGCATCCAGGT

CAGTTGCTGCCAGGACTTTCTGTAAGAGCTGCTGTTGCTCTTCTTTTGTCGAAAACGCC

AAGTCTTCCTCCTCCATGCCAGCGAGTTTCGTGATCACCTTGAAGCTATAACCTTGATC

TACCAAGAATCTCTGCCGCTTGGTTGAGTAAGCCATTTCCTGTGTGTCCTGGGATACCA

GTGAGTAGAAAAAGGCATTGTACTCTTCTGCAACCATCCCTTTTTTAGCTCGAAGCACC

CGCCCTAGCCTTTGGGCTTCCTGACGCCTGGAGCCACCGTGGGATGAGATCTGAATGA

GGACATTTGCTTCCGGCAGATCAAACGAAGTGTCACCTACCTTGGATATGAAGATGGTG

TTAATTTTGGGGTTGTGCTTGAAATTCTGGAGAATTTGCATCCTTTCCCCCTGAGACGTA

GGTCCGTAGATATAGGGTTTGTTCAGTCGAATGGCATATTCCTTTAGGGCAAACACATT

GTCAGCAAAGACAATAATCTTGTCATTCCTCCTTTCATGAAACTTGATCAGAAACTGGCA

AGCTCTAAATTTGTTGGGGTTCATGGTGTACAGCAAGATTCGTTTCTTGGTTTTGATTGC

CACATATTCCCGGTAAAATTCAGGAGACATAGGGCACCAGACCTCAGCACACTGGACT

TTGGCGATGTAGCCATTATTCTGCAGCTCCATCCAGTTAGCTTCGTAGAGCTTAGGCCC

AATCAGAAAATTTAAATCCACAATTTTGTCATCTTCGCGGACGAGGGTCGCAGTCAAAC

CCAGCTTACAGTGGGCCTGCACGATGGTGAGCACCCTTCGGAACATCTTGGCTGGTAT

GGTGTGCACTTCATCCAGGATCATGAGGCCCCACTCCTGGGTCTTGAGCCACTCCATG

ACTCGCTCGGCCTCCCAGGACCTTTTGGTGGTGTGGCCCAGCATGGAGTAGGTGCTAA

TGGCAACGGAGCAGCCGATGGGCTTGTCCTTGGCATCGGAGGTGAACCGGCAGATCT

GGCTGTCGTCAATGGTGGACCACATCTTGAACTGGGCTTTCCACTGCTCCACAGAAAC

AGCTGAGTTGCCCAGCACCAGACAGCGTTTTCTGACAGTGCATGCAGCAGTCACACCA

ACCAGGGACTTTCCAGCACCGCAGGGAAGAACAATGACCCCCGAACGTGCACGCCCG

TTTCCAAACATCTTTCGCAAGCTCTTCTCCTGATAGGGTCTGAGGACAGCTGTGGGCTT

TAGGTCAATGTTGATATCAGGGTTGACAGAATCATTCCGGAAGTCATATTCTGCCAACA

GAGGGTACTCCAGGTGGATGCAACGTTTCTGGAGTTCCTCAATCATTTCCTGCTTGACT

TCAAAAGACACTGTCTGTGTCTCTTCTTCTTCTTCTTCATCCTTGTCCATTTGCTCATAGA

AGTCAAACAGGTCCATTGGGATGTCAGATTTACCCTGTGGATCTGTCACTCGGGAAGT
GGAGGGCCCACCACTGCTTTCAGCAGTCTTAGAAATGGCAGATTTGCTTGTGAAAGTCT
CTGTGATGAGCTCAGTGGCCTCCCCTTCAGAGTTTCTTAAGCGGCATTCTCGGATCACG
GGGTCCTGGAGAAGATGCTGGATTACATCAGGGTGGCAACTTTCAACGAAGTATCTGT
TGTGCTTCAAGACCAGCTTGACTTTTCCATAGCTGACAGTACACAACTTAATAAACTGCA
TAATTCCATCAGGGACTCCAGTCTTGCTGAGCTTCCTGAGGTACTCGGTGATGTCACTG
GTTTGCAGCCCAACGCTGACAGCTGCATACAAGGAGTAGGCAGTTAGTTTGTACTCAT
GCACATGGGTTGGTCGGCACACTGGCTCTGCAATAGCCACCAAGAAGTCTTGGGCATA
TTTGTAAACTGGAGAGAAGGCTTCCAAGAAGATATGGCCATCGGGAGCCACCCAGAGG
GGCCTGGAGGTGTGGTCGTCCTTCAGCGGCATTTGCAGCCTGTAGTCCTTGGCTCCAT
ATTCATCCACTTTGGTGCCTGACTCATCCACCTGCTTCCCCGCCGCCGAGGGAACCGC
TTCCTGAGGGTCGTTCCCCGGGGCGTCCTCTTCATCATCCTCTTCATCCTCATAGTGCC
GCTTCCTGGATTTCTTCTTGTCGCGGTCCGCTCGGTCTCTTTTGCCcatggctcgagatcccgg
gtgatcaagtcttcgtcgagtgattgtaaataaaatgtaatttacagtatagtattttaattaatatacaaatgatttgataataattcttat
ttaactataatatattgtgttgggttgaattaaaggtccgtatactagggtatacccatctaaggtagcgagtttaaacactagtatcg
attcgcgacctactccggaatattaatagatcatggagataattaaaatgataaccatctcgcaaataaataagtattttactgttttc
gtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtcccaccatcgggcgcgGATCCCCATGGA
GCTCAACGTGGACGGGCTCCTGGTCTACTTCCCGTACGACTACATCTACCCCGAGCAG
TTCTCCTACATGCGGGAGCTCAAACGCACGCTGGACGCCAAGGGTCATGGAGTCCTG
GAGATGCCCTCAGGCACCGGGAAGACAGTATCCCTGTTGGCCCTGATCATGGCATACC
AGAGAGCATATCCGCTGGAGGTGACCAAACTCATCTACTGCTCAAGAACTGTGCCAGA
GATTGAGAAGGTGATTGAAGAGCTTCGAAAGTTGCTCAACTTCTATGAGAAGCAGGAG
GGCGAGAAGCTGCCGTTTCTGGGACTGGCTCTGAGCTCCCGCAAAAACTTGTGTATTC
ACCCTGAGGTGACACCCCTGCGCTTTGGGAAGGACGTCGATGGGAAATGCCACAGCC
TCACAGCCTCCTATGTGCGGGCGCAGTACCAGCATGACACCAGCCTGCCCCACTGCC
GATTCTATGAGGAATTTGATGCCCACGGGCGTGAGGTGCCCCTCCCCGCTGGCATCTA
CAACCTGGATGACCTGAAGGCCCTGGGGCGGCGCCAGGGCTGGTGCCCATACTTCCT
TGCTCGATACTCAATCCTGCATGCTAATGTGGTGGTTTATAGCTACCACTACCTCCTGG
ACCCCAAGATTGCAGACCTGGTGTCTAAGGAACTGGCCCGCAAGGCCGTCGTGGTCTT
CGACGAGGCCCACAACATTGACAACGTCTGCATCGACTCCATGAGCGTCAACCTCACC
CGCCGGACCCTTGACCGGTGCCAGGGCAACCTGGAGACCCTGCAGAAGACGGTGCTC
AGGATCAAAGAGACAGACGAGCAGCGCCTGCGGGACGAGTACCGGCGTCTGGTGGAG
GGGCTGCGGGAGGCCAGCGCCGCCCGGGAGACGGACGCCCACCTGGCCAACCCCGT
GCTGCCCGACGAAGTGCTGCAGGAGGCAGTGCCTGGCTCCATCCGCACGGCCGAGCA
TTTCCTGGGCTTCCTGAGGCGGCTGCTGGAGTACGTGAAGTGGCGGCTGCGTGTGCA

768 GCATGTGGTGCAGGAGAGCCCGCCCGCCTTCCTGAGCGGCCTGGCCCAGCGCGTGT
769 GCATCCAGCGCAAGCCCCTCAGATTCTGTGCTGAACGCCTCCGGTCCCTGCTGCATAC
770 TCTGGAGATCACCGACCTTGCTGACTTCTCCCCGCTCACCCTCCTTGCTAACTTTGCCA
771 CCCTTGTCAGCACCTACGCCAAAGGCTTCACCATCATCATCGAGCCCTTTGACGACAG
772 AACCCCGACCATTGCCAACCCCATCCTGCACTTCAGCTGCATGGACGCCTCGCTGGCC
773 ATCAAACCCGTATTTGAGCGTTTCCAGTCTGTCATCATCACATCTGGGACACTGTCCCC
774 GCTGGACATCTACCCTAAGATCCTGGACTTCCACCCCGTCACAATGGCAACCTTCACAA
775 TGACGCTGGCACGGGTCTGCCTCTGCCCTATGATCATCGGCCGTGGCAATGACCAGGT
776 GGCCATCAGCTCCAAATTTGAGACCCGGGAGGATATTGCTGTGATCCGGAACTATGGG
777 AACCTCCTGCTGGAGATGTCCGCTGTGGTCCCTGATGGCATCGTGGCCTTCTTCACCA
778 GCTACCAGTACATGGAGAGCACCGTGGCCTCCTGGTATGAGCAGGCGATCCTTGAGAA
779 CATCCAGAGGAACAAGCTGCTCTTTATTGAGACCCAGGATGGTGCCGAAACCAGTGTC
780 GCCCTGGAGAAGTACCAGGAGGCCTGCGAGAATGGCCGCGGGGCCATCCTGCTGTCA
781 GTGGCCCGGGGCAAAGTGTCCGAGGGAATCGACTTTGTGCACCACTACGGGCGGGCC
782 GTCATCATGTTTGGCGTCCCCTACGTCTACACACAGAGCCGCATTCTCAAGGCGCGGC
783 TGGAATACCTGCGGGACCAGTTCCAGATTCGTGAGAATGACTTTCTTACCTTCGATGCC
784 ATGCGCCACGCGGCCCAGTGTGTGGGTCGGGCCATCAGGGGCAAGACGGACTACGG
785 CCTCATGGTCTTTGCCGACAAGCGGTTTGCCCGTGGGGACAAGCGGGGGAAGCTGCC
786 CCGCTGGATTCAGGAGCACCTCACAGATGCCAACCTCAACCTGACCGTGGACGAGGG
787 TGTCCAGGTGGCCAAGTACTTCCTGCGGCAGATGGCACAGCCCTTCCACCGGGAGGA
788 TCAGCTGGGCCTGTCCCTGCTCAGCCTGGAGCAGCTAGAATCAGAGGAGACGCTGAA
789 GAGGATAGAGCAGATTGCTCAGCAGCTCTGAGGTACCGtcgacgagctcacttgtcgcggccgctttc
790 gaatctagagcctgcagtctcgacaagcttgtcgagaagtactagaggatcataatcagccataccacatttgtagaggttttactt
791 gctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataat
792 ggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaat
793 gtatcttatcatgtctggatctgatcactgcttgagcctagaagatccggctgctaacaaagcccgaaaggaagctgagttggctg
794 ctgccaccgctgagcaataactatcataaccctagggtatacccatctaaggtagcgagtttaaacgtacccgtagtggctatgg
795 cagggcttgccgccccgacgttggctgcgagccctgggccttcacccgaacttgggggttggggtggggaaaaggaagaaac
796 gcgggcgtattggtcccaatggggtctcggtggggtatcgacagagtgccagccctgggaccgaaccccgcgtttatgaacaa
797 acgacccaacacccgtgcgttttattctgtcttttattgccgtcatagcgcgggttccttccggtattgtctccttccgtgtttcagttagcc
798 tcccccatctcccggtacCTCAGGCAGACACTTTCAGTTTCTTTTTCTTGGCTTTCAGCACTGGA
799 GGCAGAGAAATTTTAAAGGCTGTCTCGCACGTAGTACAAATGGGGCTGAAATTGCAGA
800 ATATTGACAAACACACAGAACAGACATAACCAATTTCAATGAGATTTCGATGACAGAAG
801 CAAGCAGCCCTGTAGTCAACATGAACCGGGGGGTGGGAGGATTAACTGAGATCTCTGAT
802 CTTGATCGGGAAGAAACACCCACAGCAAATACTGCAGAAGAGAAGGCATCTGAGGCAC

803 CTTCAGGTACAGTCCTCCCGTGATGTCACAAGCCTGTTGGAGGAGCCCTGAGTCGGAG
804 TCTAAAACACAGGCATCAATCAAAATATTCTGTTTCTGTGCTGCAAAGATGACATTCATG
805 AAGTTCATATACTGCAACGCACTGTCTTCTGCAGCCTTAATCACCAATATCCTTGATTTC
806 ATTTCCTGATTGTCTTTAACTTCCTTGTTCATTCTATGAATGTAGCAAAGGGCTTTGGCC
807 AGGGAGCCTGCCAGCAAAGTTTCTGTATGTTGACCCTTTATGTCACTTTTGGTCATTAG
808 ATCTTTAATCTCTTCAACAATAACTTCATTTGCTGAGGTTAAAAGTTCGTATTTTCCATCT
809 TTACTCCCAGAGGGATTAAATTCAGGAGGGTTGCCAGGGTCTCCGAAGAAGTCTCCAA
810 GTCTGCCATTCTTTCCAGGATATAAGAATCGGCTTTCTTGAATGTGACTTGCTATCACAG
811 CAAGTTTGTTGGAACGATTCATGAATAAATGCGAATTTCCCAGCACCATCACGGCATCT
812 ATGCATTTGGATAAAGTGAACTGAGATTCCTTTAATGCTTGCTTTCCCCACCAAATAGGG
813 TTGGCATCAACTACAATAACCAGAAGATTCAATTCATCTTCGTCTGAAACCATAGGGCC
814 GGGGTTCTCCTCCACGTCGCCGCAGGTCAGCAGGCTGCCCCGGCCCTCAACACCTGA
815 AGGAGCTGGAATCTTATGAATACAGCCAGGGCAACAGTGTAGAGAATCATGAACAAAAA
816 CATCACAGTCCACACAGAAAACATTTTGGCACACAGCACAAACATAAACATGTTGGTCT
817 TTCAATTCCCCCTGACATCCATAACAAAATCTTTCTCCATTATATTCTTCTAGGGGAATTT
818 CTTGAAAAGCATCCAAAGGAAACAAATGATGGTAAGACCGTGCCAAGTGGGGAGCAGA
819 CACCAAAGTAAGACCACAGATTTTACATTCAACAGGTAGCTCACAGTACTTTGCCCGAC
820 ACTGTGGGCAGAAATAGCCTCCTAATGTAAGCCCTGGCTCAGTATTGCCATCCAAATGC
821 GCCATGCTGAAAGAGGGTTTTGCATCCTGGTCAGATAAAGAAGCAATGGTGTGCTGAG
822 GAAATCCCATACGAATAAGTGAGCATTCAGAACTTGAGCTAGCAGGAGGAGGACTAAC
823 ATGATGTGTGAGCAACTCTTTGTAATGGCTTTCATCTAAAATAACATGGTACGTGCCACC
824 AGTTTCACGAGCAAGTACAGTGCAAACGCGAACTTCTGCAGACAATCCAATAACAGATA
825 CTCTAATTTTAGCTGCCTTTAGGGTCTTGATTAGATCATAAATATTAGATGGATCGCAAG
826 TTGTAAGGCTGCTAAAGATGATTAGTACTTCTCGACTTGTATGTCCAGGCATGTGTTTTA
827 GAGTCTGCATAGCTATGCTTAGGGAATTATAAAGAGATGGCTCTCCGTGGCAGGTCATA
828 TCCACAGCTTTCTTCAAAGACGTTATATGTTTTCTTGGGTTTCCTGAAAGTTCAGTCAAT
829 TTTTCAGCTCTTTTACTCTTAGTTACAATTATTCCAATCTGACTAATAGGATTTTGATCAA
830 AATATTCCTCTACAAAGTATTCCAACAACTTTAAAGTACACGTCAGTCTATTAGGCTTTAA
831 ATCTTGGTCTTCCATTGTTCTTGATCCATCTACTACCACATAAAGGTGGCGCATCATTCC
832 AAGTCGAACTTGTCCATGATGCTCAAATACTCTTTTTCTCTTTGCCTTGAATAGAATGTC
833 TTCTATTGTAGCTTTAAGTGATCCAGATTCATCTTCTTTAAGAATCTCCCATGTTCTTTCA
834 TAGCCTCCTTCCCATCGCTTAGTTCTTTCAGGTTCTTCATCcatggctcgagatcccgggtgatcaag
835 tcttcgtcgagtgattgtaaataaaatgtaatttacagtatagtattttaattaatatacaaatgatttgataataattcttatttaactataa
836 tatattgtgttgggttgaattaaaggtccgtatactagggtatacccatctaaggtagcgagtttaaacgtacccgtagtggctatggc
837 agggcttgccgccccgacgttggctgcgagccctgggccttcacccgaacttgggggttggggtggggaaaaggaagaaacg

838 cgggcgtattggtcccaatggggtctcggtggggtatcgacagagtgccagccctgggaccgaaccccgcgtttatgaacaaac
839 gacccaacacccgtgcgttttattctgtcttttattgccgtcatagcgcgggttccttccggtattgtctccttccgtgtttcagttagcctc
840 ccccatctcccggtacCTCATTTCTGGGTAAGGGAAAAAGCATTTTGGTCCATTAATTCACCCA
841 CTCGCTCCTGGAGGACATTAACCAATTCTGCTATTACAAAGACGTGAGTGTCATCAATG
842 TCTTGAATGATGAACTTCTTCCCCAGGGCATTTGACTCATCCAAGTACAGCAGAAACTG
843 CTTCATGGCAGGATCACATTCTATAAGCACTCCTTTCAAGACGTTGACCATAGGGCCGG
844 GGTTGCTCTCCACGTCGCCGGCCAGCTTCAGCAGGTCGAAGTTCAGGGTCTGGGTGC
845 TGTCCAGGCCCAGCAGGGGGTTGGGGATGGGCTTGCCGGAGCTATGTTTCTGCCGCT
846 TCCAAAAGCGCTTGACGTCGCTGTGCCCGGCCGGGGTCACCACCATGAGCCGCTTGG
847 CCGAGTTCTCGAACACGAGCACGCCCAGCTCCCGCGCGTGGGCCAGCAGCAGCTCAA
848 AGTCCACTTGCGACAGGAACTGGTTATACAGGACACCCTCAGTGAACCGGAGTCTGTC
849 CCTTTCCAGCTCCCAGAGCCGGATCTGGTCGGTGATGGTGGGGGGCAGCACAGGTGT
850 CTGTTTGAGCATCACTGGGTGGGCTCTTGTCCTTAGGAAATGGATTATCTGCTGGGCTG
851 TGATGCCACTGGCGATTGCCTGCTGCACACTCTCCCGGGTCACCTGCGCCACCACCAT
852 GTTAGGGAACCGATAGAGCATCTCAGAGAAGAGGGCAATGAGGGCAATCTGCAGCTCC
853 GACTCCGTGTAGGCATACAGTCGGTAATTGGTTTCCACGACAATGAAACCTGGCTGAT
854 GCACAGTGCCCCCAGCTCCAGAGACACCTGATGAGAGATTGATGGCCAGGCGTGTAG
855 GGTAGTAACGCCGAGATTTCCTCTTCCTCTGGAAAACAAGCCCAAACTCACGCAGATGT
856 TGCAGGAAGTTCAACAGAGAATCACTCATACCTTCCACAGAGTAATCCTTGCCCAGAGT
857 AGAGAAGCTGAGCTGGAAGAGGAAGGAGAGAATCTCTACCAGGTCCATGCCCCGGCT
858 CTGGGCTGTCTGCAAATACTGCAACATAAAGTACCAGAGCTGAGCCGGGGTGTCCAGC
859 AACAGGAACTGGAAGCCAGCGGAAGTAATGCAGGGCGGCTCTCCAGGTTCAGTACTCT
860 TCATGAGCCCAGCCTGGCTGAGGAGCTGAGCCAAGTCCTGGCTGACAGCTGCACTGG
861 GGGAGCCCACCATGAAGTGCAAGACCACCTCCCATCGCTCCTCGGCGTACTTGTCAAG
862 GGAGGGAACGTCCCGGGCATGCTTGTCTGGTCCCAGCTGACTTGTGTCATCAGACCAG
863 GCCTTCCCCCCACCCAGAAGGGCAATGCGGAGGTTCTGGCGGAAAATGGGGTTGAGG
864 ATGAGGCCCTGGAGCCCGCCTGGGAGCAGCTGTGTGTGCCAGATCCGGAGGCCGCTC
865 AGCAGCCCTGTACTTTCCTCCTGAGCCTTGCTGAATTCCTTCTTTACCCACAGAGCTAC
866 AGCAGCCTGTGGCAAAGGCTGCTCCAGAAAGAGCATCCGCATCACCCAGTTCTTAGCC
867 AAGGATGGGAGCTCCCTGAAGACAGCCAGACATGTAGCAGGGTGCCCATACAATCGGT
868 CCAATACCCCAGGGCTCAGGCCCCCTAAGAATTCCTGCAGATTCCTGCATTGTAGGTG
869 TACTCGGTTCAGTCCCCTTGAAGGGGTGCTCTCCATGGCGTTTGCGCCCTGAAAATAC
870 AGGTTTTCGGTGCTATCCAGGCCCAGCAGCGGGTTCGGAATCGGTTTGCCATCGTGAT
871 GATGATGGTGATGCGACATATGCccgggtgatcaagtcttcgtcgagtgattgtaaataaaatgtaatttacagtat
872 agtattttaattaatatacaaatgatttgataataattcttatttaactataatatattgtgttgggttgaattaaaggtccgtatactagggt

873 atacccatctaaggtagcgagtttaaacgtacccgtagtggctatggcagggcttgccgccccgacgttggctgcgagccctggg
874 ccttcacccgaacttgggggttggggtggggaaaaggaagaaacgcgggcgtattggtcccaatggggtctcggtggggtatc
875 gacagagtgccagccctgggaccgaaccccgcgtttatgaacaaacgacccaacacccgtgcgttttattctgtctttttattgccg
876 tcatagcgcgggttccttccggtattgtctccttccgtgtttcagttagcctcccccatctcccggtacCTCACGTTTTCTTCA
877 TCAGACGCCGTGACTGCCATGTGTGGAGCTTGTTGTAGGCTGTCTGGAGCATCTCTTC
878 TATGTGACTTACCAAATTTGTGCTTAAATACTGTCTCCGAATCTTTTCTTGGAATGGACA
879 GAGCTTCGTAACTTGGAATCGTTCCAAATTACTTTTCATTTTCACTACCTTTTCTTCTAGG
880 AATGGCGTATTAACAGGAAAGCAGGACCAGAAATGTCGTAGAAGTTCTCCAACAGCTAC
881 ATATAAGTGTTTCAATTCAGATTGAATATCATTTGGCACCATCTGGTTTATGGCTTGCTG
882 TGTTCCTCCCTGCATAAGTGCCCCTCCAGGTGACAGTGCTGTGATGGTACTACTGGCA
883 GCACTACTTGAGAGAACCTGAGTTAACTTGGGTGTATAAGCTTCCATTTCTTGTCTAATA
884 CTTTGAAAAGAATTAATAATGTCCTGACTTGTTGCATACTGTAGTGACTGGATTGGAGTT
885 GGACCATGATAATACCTATCTGACTTCTTGAGGTTTAGTGCAATCGTTTTTACAGAATTA
886 TTTTTCCCCAAGTCTTCATATTCAATGGACTCTTGTAATTTCGCCCTTTTGACTGCTGGC
887 TGAAAGCAGTCTGCATCTCCGGAATTTCCATCCATGTTGCTAGGCTCACTAGTTTGTTC
888 ATTTTGTGCTTCTTGTTTTCTGAGTCCAGCTGCCAGGACCATAGCACTGTGATGGTTAAA
889 TCTCTTGATGATGGCAGCATTACTATTCTCTTTTATGGATTTAGAATTGGAAGCAGATGG
890 CACAGAGGAAATGCCATAGCCCTCATCTAATGGTTTATCTTCCAAAGCTGTTAAATCTAG
891 TAGTGGGTTTTTCACTCCTAATGAAACCATTGTTTTTAGGCCTTTTTCATCTATTTTGGCA
892 CATTCTGCAAAGAGATCCTTTGACCCTGTATTCAGCCGATCCCTGTGAAAATAATGGGA
893 CTGGAAAAAACGTGTCCAGAATTCCTTCTCTGTCATGTTGTGGGGAACATTTTCTGCATA
894 TTTCATTTTTACTGCTGGATAGGTCCTAAATATGGACTCAATGATATCAGAAGTTAAATTA
895 TATCTTAGACCGTTACAGCCATCAGTTTGAGGCCGGACATCAGCCAGAAATGCAGCAG
896 AAATGCCAACATCCTGCTTATGATTGGATGTGGAAGAACTATCTGTTGCATTCACATTTA
897 AACGATTGGCCCAGAATTCCTCAGCACTGATCACTTGACTCACAACAAGGTCTTTATAA
898 AGCTGAAACAAAACAGGATCTTCTTGCAGCATTCTGTTCTTCTCTTCCAGTTCTTTATTT
899 GCTTTCCTCTTGAATTTGGGCAGCAGCTGCTGAAGAAGGTCTTTTACTGCATCTCGCTC
900 TTTCACTGCTGTGCTTTCATTGGAAAAATGGAAGTTAGTTGTGTCCCCTGCATGTAGGA
901 CCAGCTGAAGCTGAATTTTAGCTTTTCCTTCTGGACTAATTTTCTGGCATTTAATATCTG
902 CATACATATGGCTGATTGTAAATCTATCTTTGCCTTCAGGTGCCCAAGCAATTCTTTCTG
903 CCATGAGGTACAGAGCTCCATCCTGCTTCTTTTGACGCACTTTCTTTACAATCAGCAAAA
904 CTTCTTCAGATGAGGTTGCCATGGCGTTTGCGCCCTGAAAATACAGGTTTTCGGTGCTA
905 TCCAGGCCCAGCAGCGGGTTCGGAATCGGTTTGCCATCGTGATGATGATGGTGATGCG
906 ACATATGCccgggtgatcaagtcttcgtcgagtgattgtaaataaaatgtaatttacagtatagtattttaattaatatacaaatg
907 atttgataataattcttatttaactataatatattgtgttgggttgaattaaaggtccgtatactagggtatacccatctaattggaaccag

ataagtgaaatctagttccaaactattttgtcatttttaattttcgtattagcttacgacgctacacccagttcccatctattttgtcactcttc

cctaaataatccttaaaaactccatttccacccctcccagttcccaactattttgtccgcccaca