Study of the co-translational assembly mechanism of transcription complexes in mammalian cells

Pooja Mukherjee

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE DE STRASBOURG

IGBMC - CNRS UMR 7104 - Inserm U 1258

THÈSE

présentée par :

Pooja MUKHERJEE

soutenue le : **20 Septembre 2019**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

Study of the co-translational assembly mechanism of transcription complexes in mammalian cells

Etude des mécanismes d'assemblage co-traductionnel des complexes protéiques impliqués dans la transcription dans les cellules de mammifère

THÈSE dirigée par :

Dr. TORA László — Directeur de Recherche, IGBMC, Université de Strasbourg

RAPPORTEURS :

Dr. COLLART Martine — Professeur, University of Geneva

Dr. MARSH Joseph — Reader, University of Edinburgh

AUTRES MEMBRES DU JURY :

Dr. SEXTON Thomas — Chargé de Recherche, IGBMC, Université de Strasbourg

Acknowledgements

Firstly, I would like to thank my supervisor and mentor Laszlo for giving me the opportunity to work in his lab. I am really very very grateful to you. It was a very difficult phase for me 4 years back in 2015 when I decided to leave my country and started applying to positions abroad. You have no idea how your reply to my application on the 14th of January 2015 morning (yes, I still remember the date and time!) changed my life in a good way, really. And then comes the very interesting project on co-translational assembly, I am so thankful to you for making me a part of the project. I feel I was at the right time at the right place doing the right project! Your enthusiasm and motivation were so important for me throughout my PhD; I highly enjoyed the freedom of planning and carrying out experiments in your lab and this also helped me develop my confidence as a researcher. Your suggestions and criticisms are always so helpful. I had a very very enriching experience in your lab and I wish I always get a boss like you. Thanks a ton.

I would also like to thank Ivanka. I believe life in the new lab would have been difficult without you. Right from day 1, you were very kind and helpful. I thank you for all the useful scientific discussions that I had with you. I believe my PhD would not have been so successful without you.

I would like to thank the jury members of my thesis, Dr. Collart, Dr. Marsh and Dr. Sexton. Thank you very much for being a part of my thesis evaluation and for your suggestions and advices.

My heartfelt thanks to all the members of the Tora lab, both present and former. You guys are amazing! Our lab has such a wonderful working environment and its only because of the lovely lab members. Thanks, Tiago, for helping me during my initial days in Strasbourg. Thanks Federica, Nicholaos for the wonderful time we had together. Thanks very much Didier for the occasional scientific and non-scientific discussions that we had. Thanks, Stéphane, for your useful scientific suggestions and also insightful career advices. Thanks a lot Kenny, Vincent and Matthieu for being a great labmate. Thanks very much Veronique for all the amazing moments we spent and for your help with French translations. Thanks very much Farrah for being such a kind friend. Thanks so much Fang for your kind words in stressful situations and also for the wonderful dishes that you cook. Thanks very much Sascha for your help and answering all my queries about imaging. Thanks a lot Changwei for all the wonderful scientific and career discussions we had and also for the amazing time we spent outside lab. Thanks a lot Paul for being such an amazing team member. Thanks a lot my fashion girls, Eli and Gizem. Life in lab was more interesting with you guys. Thanks for all the crazy fun we had together. I am grateful to you guys for your suggestions and help with ALL my problems, scientific and non-scientific. A special thanks to Eli for keeping the lab reagents, plasmids, antibodies so so organized and for always helping with perfectly optimised protocols and enormous wet lab experience. Really amazing!

Thanks a lot to all my other friends in IGBMC, specially Ben and Pietro. Even though I am not a very social person, but still you guys never forget to talk to me.

Thanks to ARC for supporting me with the 4th year fellowship.
Special thanks goes to Alka, Akinchan and Sweta. You guys made my initial days in Strasbourg very comfortable, when I was dealing with culture shock with a completely new independent life away from home comforts. Thanks a lot!
Thanks to Sanjay (Chahar), Dhanvantri, Atish, Sankari, Banku for the all fun discussions and occasional great food. Thank you, Sanjay (Dey), Nisha and Abbas, for your insightful career advices.
A very special thanks to Ujjwal. I cannot express in words how grateful I am to you for multiple reasons. Its only because of you that I can cook tasty food today! I am forever indebted to you. Life in Strasbourg would have been boring without you. Thanks for helping me in every stage of my PhD life and my postdoc applications with your amazing problem-solving solutions each time. Discussions with you improve my GK every time!! Thanks a ton!
Finally, I would like to thank my family and friends in India. Thanks, Shalini and Diptasree, my school friends, for bearing with me even though I disappear occasionally. You guys never fail to ask how I am from time to time!
Thanks Upamanyu for helping me with my PhD applications when I was struggling to find a position four years back. Thanks Naibedya, my ex-labmate, for helping me sail through the rough times in 2015.
A very special thanks to Arnab. You are the only one person who is always more confident about me than even myself! I would have left science long back if I had not known you. Your motivation and encouragement helped me pass through the darkest days of my life. And as I always say, it is because of you that I am in Strasbourg today. My heartfelt thanks to you.
Thanks to Pipi, Boromashi, Mejdi, Pibaba, Alokda, Sumitda, Bapindada, Mithundada, Noddy, Gublu, Tumpadi, Rajan, Debasishmama, Bob. Love you guys!
Loads of thanks to my family, Gigi, Dada, Bhaiya, Tutu. You guys never failed to celebrate even the smallest of my victories. I am so lucky to have such a great family.
Thanks a ton, my parents. You taught me how to dream big and achieve it. In every life situation, I had and still have you behind me to hold me if I fall. Failures in life are never a big deal because you taught me how to see positivity in every situation. Have no words to express my gratitude. I am lucky indeed.

Pooja Mukherjee

Table of Contents

List of Figures

List of Tables

Abbreviations

ADA2 : Adenosine deaminase 2

ADA3 : Alteration/Deficiency in Activation 3

ADP-ribosylation : Adenosine diphosphate-ribosylation

ARE : AU-rich element

Arp4 : Actin-related protein 4

Ascl1 : Achaete-Scute Family BHLH Transcription Factor 1

ATAC : Ada2a-containing complex

ATRX : α-thalassaemia/mental retardation syndrome X-linked

ATXN7L3 : Ataxin 7 like 3

BAF : BRG1-Associated Factor

BLOC-1 : biogenesis of lysosome-related organelles complex 1

BORC : BLOC-one-related complex

Bre : Brain and Reproductive organ-expressed protein

BRF1 : B-related factor 1

BTAF1 : B-TFIID TATA-Box Binding Protein Associated Factor 1

CAF-1 : Chromatin Assembly Factor-1

CARM1 : Coactivator Associated Arginine Methyltransferase 1

CBP : CREB-binding protein

Ccr4 : C-C Motif Chemokine Receptor 4

CCT : Chaperonin Containing TCP1 or TriC-TCP-1 Ring Complex

CHD : chromodomain helicase DNA-binding

COMPASS : complex proteins associated with Set1p

CPSF : cleavage and polyadenylation specificity factor

cryo-EM : cryo-electron microscopy

CSB : Cockayne syndrome group B

CSTF : cleavage stimulation factor

CTD : C-terminal domain

DOT1L : Disruptor of telomeric silencing 1-like

DP : Aspartic acid-Proline

DSIF : DRB sensitivity inducing factor

DSS1 : deleted in split hand/split foot

DUB module : Deubiquitination module

Eaf3 : Esa1-associated factor 3

ECDs : extracellular domains

EGFR : epidermal growth factor receptor

EJC : exon junction complex

ENY2 : Enhancer of Yellow 2

ER : Endoplasmic reticulum

FACT : facilitates chromatin transcription

FoxA : Forkhead box A

GABAA : γ-aminobutyric acid

GANP : Germinal-center Associated Nuclear Protein

GAPDH : Glyceraldehyde 3-phosphate dehydrogenase

Gata : (A/T)GATA(A/G)

GCN5 : General Control Nondrepressible 5

GDP : guanosine diphosphate

GFP : green fluorescent protein

GRE : GU-rich element

GTFs : General Transcription Factors

GTP : guanosine triphosphate

HAT : Histone acetyltransferase

HEAT : Huntingtin-elongation factor 3 (EF3)- protein phosphatase 2A (PP2A)- yeast kinase TOR1

hERG : human ether-à-go-go-related Gene

HFD : histone fold domain

Hsc70 : heat shock cognate70

Hsp70 : heat shock protein 70kD

Hsp70L1 : Hsp70-like protein 1

HSPB1 : heat shock protein beta-1

ICDs : intracellular domain

IF : Immunofluorescence

IF : Intermediate filament

IF-smFISH : Immunofluorescence single molecule fluorescent in situ hybridization

Ig : immunoglobulin

INO80 : inositol requiring 80

IP : Immunoprecipitation

ISWI : imitation switch

J protein : J domain protein

KDM5 : lysine (K) demethylase 5

Klf4 : Kruppel Like Factor 4

KO : Knockout

Lid : Little imaginal discs

MCP : MS2-coat protein

MERFISH : multiplexed error robust FISH

mESCs : mouse embryonic stem cells

miRNA : microRNA

MLL : mixed lineage leukemia

MPP11 : M phase phosphoprotein 11

mRAC : mammalian RAC

mRNA : messenger RNA

mRNP : mRNA-ribonucleoprotein

MS2-TRAP : MS2-tagged RNA affinity purification

MYB : myeloblastosis

NAC : nascent-chain-associated complex

NC2beta : Negative co-factor 2

NCA : Not1 containing assemblysomes

NEFs : nucleotide exchange factors

NELF : Negative elongation factor

NGD : no-go decay

NLS : Nuclear Localisation Signal

NMD : nonsense-mediated decay

Not : negative on TATA

NPCs : Nuclear pore complexes

NSD : non-stop decay

NTD : N-terminal domain

NTP : Nucleoside triphosphate

NuA4 : Nucleosome acetyltransferase of H4

NuRD : nucleosome remodelling and deacetylase

NURF : Nucleosome Remodeling Factor

Oct3/4 : octamer-binding transcription factor 3/4

ORF : Open reading frame

P bodies : Processing bodies

PAF1C : Polymerase associated factor-1 complex

PAQosome : particle for arrangement of quaternary structure

PAS : polyadenylation signal

Pax7 : Paired box 7

PCID2 : PCI domain containing 2

PCNA : Proliferating Cell Nuclear Antigen

Pfd : Prefoldin

PHD : plant homeodomain

PIC : pre-initiation complex

PPIs : protein-protein interactions

PRMT1 : Protein arginine N-methyltransferase 1

PS1 : Presenilin-1

PSII : Photosystem II

PTC : premature stop codon

P-TEFb : positive transcription elongation factor b

R2TP : Rvb1–Rvb2–Tah1–Pih1

RAC : ribosome-associated complex

RAP30/74 : RNA polymerase II-associating protein 30/74

RAPP : Ribosome-associated protein quality control

RIP : RNA Immunoprecipitation

RNA FISH : RNA Fluorescent in situ hybridization

RNA-BP : RNA-binding protein

RPBs : ribosome-bound protein biogenesis factors

Rpd3 : Reduced potassium dependency-3

Rpt : Proteasome Regulatory particle base subunit

rRNA : ribosomal RNA

SAGA : Spt-Ada-Gcn5 acetyltransferase

SeRP : selective ribosome profiling

SET1 : Su(var) 3-9, Enhancer of zeste (E(z)), Trx 1

SETD2 : SET Domain Containing 2

SL1 : Selective factor 1

SMAT : small TAF complex

smFISH : single molecule fluoresencent in situ hybridization

SMG1 : Serine-threonine protein kinase

smiFISH : single molecule inexpensive fluorescent in situ hybridization

snRNA : small nuclear RNA

Sox2 : SRY-Box 2

SPT : Suppressor of Ty

SRα : signal recognition particle receptor α

SUMOylation : Small Ubiquitin-like Modifier (SUMO) -ylation

SURF : SMG1-UPF1-eRF

SWI/SNF : switch/ sucrose non-fermentable

SWR1 : SWI2/SNF2-Related 1

TADs : topologically associating domains

TAF : TBP associated factor

TAFH : TAF4 homology domain

TAND : TAF1 N-terminal domain

TBP : TATA-box binding protein

TCP-1 : tailless complex polypeptide-1

TCR : T-cell receptor

TF : Trigger Factor

TFII : Transcription factor II

TIGER : TIS Granules-ER

Tip60 : Tat interactive protein 60-kDa

TIS : TPA-induced

TLF : TBP-like factor

TLP : TBP-like protein

TP53 : Tumor protein p53

TRRAP : Transformation-transactivation domain-associated protein

TREX-2 : TRanscription and EXport complex 2

TRF : TBP-related factor

tRNA : transfer RNA

TRP : TBP related protein

UPF : Up-frameshift

UPF1 : Regulator of nonsense transcripts 1

USP22 : Ubiquitin Specific Peptidase 22

UTR : Untranslated region

WD-40 : Trp-Asp (W-D)-40

XRN2 : 5′-3′ exoribonuclease 2

YEATS2 : YEATS Domain Containing 2

ZZZ3 *:* ZZ-type zinc finger-containing protein 3

Abstract

Biological processes in the cell are mainly are carried out by multisubunit protein complexes and a significant amount of energy is required by the cells to build these huge complexes. Unlike bacteria, genes encoding proteins are dispersed in the genome of eukaryotes and this makes the assembly of protein complexes more complicated. For many years, it was thought that protein complexes are formed by random collisions of their subunits diffusing freely in the cell cytoplasm. However, random collisions could also lead to non-specific interactions and aggregations in the extremely crowded cellular environment. In this respect, co-translational assembly of nuclear and cytoplasmic protein complexes has been put forward in bacteria and yeast that involves the association of nascent ribosome-associated proteins subunits with each other, thereby preventing unwanted interactions. It was shown earlier that operon organisation in bacteria facilitates co-translational assembly of their protein complexes due to proximity of the encoding genes. Additionally, several well-characterised protein complexes in yeast were also reported to assemble co-translationally. In this study, we show that the mammalian multisubunit transcription complexes assemble co-translationally by using several alternate approaches like RNA immunoprecipitation followed by genome-wide detection of mRNAs by microarray analysis, single molecule RNA FISH, immunofluoresence, knock-out mouse embryonic stem cells and domain swapping approaches. We also demonstrate that the dimerization domains and their positions in the interacting subunits determine the co-translational assembly pathway (simultaneous or sequential). Furthermore, cytoplasmic IF-smFISH and two-colour smFISH experiments indicate that the described co-translational assembly is clearly occurring in the cytoplasm of human cells. Identical results in yeast, mouse and human cells suggest that co-translational assembly is a general mechanism of protein complex assembly in eukaryotes.

Thesis summary in English

Introduction

Majority of the biological processes are carried out by multisubunit protein complexes in mammalian cells. Transcription is the process by which the information in the DNA is copied into RNA molecules. Eukaryotic RNA Polymerase II transcription is driven by six general transcription factors namely TFIID, TFIIA, TFIIB, TFIIF, TFIIE, TFIIH and the multisubunit mediator complex. In mammalian cells TFIID nucleates the assembly of other transcription factors in most of the expressed protein coding gene promoters and loss of TFIID was shown to be embryonic lethal. TFIID is composed of TBP (TATA-box binding protein) and 13 TAFs (TBP associated factors). Majority of the TAFs contain a common structural motif called the histone fold domain (HFD), which facilitates pairwise interaction between specific TAFs, specifically, TAF10-TAF8, TAF3-TAF10, TAF6-TAF9, TAF4-TAF12, TAF11-TAF13 heterodimers. The histone fold is not conserved at the level of sequence but is conserved structurally; it is composed of three α-helices connected by two loops, which allow heterodimeric interactions between specific TAFs. Moreover, the histone fold containing proteins are not soluble when they are expressed individually. In spite of extensive studies on the structure and function of the multisubunit transcription complexes, very few studies have focused on the assembly mechanism and pathways of the multisubunit protein complexes. It has been reported previously that the assembly of TFIID takes place in a stepwise manner before it embarks on its function in the nucleus. Several submodules of TFIID have been shown to assemble in the cytoplasm before it forms the holo-TFIID complex. Stable heterotrimers of TAF5-TAF6-TAF9, TAF2-TAF8-TAF10 and TAF7-TAF11-TAF13 have been shown to form in the cytoplasm. In this context, the main aim of my project was to study the mechanism of assembly of TFIID submodules in the cytoplasm. A protein dimer can assemble in two possible ways: “posttranslational assembly”, where individual subunits are fully translated and released at random places in the cytoplasm eventually finding their interacting partners, or “co-translational assembly”, where protein-protein interactions form in the cytoplasm during the translation of the interacting protein partners. Co-translational assembly could be beneficial for the cell to prevent non-specific interactions in the very crowded cytoplasmic environment of the cell. One of the earliest evidences of co-

translational protein assembly was reported in 2009 where the authors studied 31 proteins lacking RNA-binding domains from *Schizosaccharomyces pombe and among them* ~38% co-purified with mRNAs that encode interacting proteins. This observation was further supported by a very recent report which showed co-translational assembly in *Saccharomyces cerevisiae* by selective ribosome profiling. Taken together, various evidences point towards the fact that co-translational assembly could be a widespread mechanism for assembly of a large number of complexes in yeast. Here, we study the co-translational assembly mechanism of the general transcription factor TFIID in detail and further extend our observations to other multisubunit complexes in mammalian cells.

Aims:

1. Do TFIID histone-fold domain pairs assemble co-translationally?
2. What drives the co-translational assembly of TFIID HFD pairs?
3. Do non-HFD pairs also assemble co-translationally?
4. Do other multisubunit complexes also assemble co-translationally?
5. Study the co-localisation of co-translationally assembling proteins and RNAs?
6. Study the role of chaperones in co-translational assembly?

Results:

To test whether HFD-containing TAFs assemble co-translationally, we used a monoclonal antibody against the N-terminus of the HFD containing TAF10 to immunoprecipitate (IP) endogenous TAF10 from human HeLa cell cytosolic polysome extracts. Immunoprecipitation of a protein followed by the study of its associated mRNAs is called RNA Immunoprecipitation (RIP) assay. Protein–protein interactions between nascent proteins still associated with translating ribosomes would be revealed by enrichment of mRNAs coding for the interacting partners in the IPs. Global microarray analysis of mRNAs precipitated by the anti-TAF10 RNA IPs (RIPs) revealed enrichment of *TAF8* mRNA, suggesting that the well-characterised TAF8-10 HFD dimer forms co-translationally. The microarray results were further supported by anti-TAF10 RIP coupled to RT-qPCR in both HeLa and mouse embryonic stem cells (mESCs). Puromycin treatment resulted in a significant loss of the enrichment of associated mRNAs thereby ruling out the possibility of direct interaction between *TAF8* mRNA and TAF10 protein. Next, we were interested in studying the underlying

mechanism of TAF10-TAF8 co-translational assembly. We generated expression vectors expressing TAF10 and TAF8 with either N-terminal or C-terminal tags. These constructs enabled us to pulldown either nascent (with N-terminal tag) or full-length protein (with C-terminal tag) and thereby determine the order of co-translational assembly of the proteins. In accordance with our endogenous immunoprecipitation (IP) results, nascent TAF10 RIP revealed the presence of its own mRNA, along with *TAF8* mRNA. Interestingly, however, *TAF10* mRNA was not co-immunoprecipitated with nascent TAF8 protein. This observation was further supported by pulldown experiments with C-terminal tagged TAF10 and TAF8 constructs. Fully translated mature TAF10 protein revealed the presence of only *TAF8* mRNA, but not its own mRNA (as expected) but mature TAF8 protein did not yield *TAF10* or *TAF8* mRNA. In all cases, TAF8 protein was co-immunoprecipitated with TAF10 protein thereby ruling out the possibility of unsuccessful protein IP experiment. Together these results indicate that TAF10 protein is assembling co-translationally with TAF8 nascent protein unidirectionally by sequential assembly model. We hypothesised that this sequential binding is specific to the localisation of the dimerization domains of TAF10 and TAF8, HFD of TAF10 being towards C-terminus and HFD of TAF8 being towards N-terminus. To test this, we engineered a mutation in the HFD of TAF8 disrupting the interaction between TAF10 and TAF8. This mutation resulted in a nearly complete loss of the co-precipitated *TAF8* mRNA, as compared with the wild-type controls, indicating that the dimerization of TAF8 and TAF10 through their HFDs is crucial for co-translational assembly. In addition to that, IF-smiFISH co-localization experiments in fixed HeLa cells showed significant co-localisation between TAF10 protein and *TAF8* mRNA in the cytoplasm by confocal microscopy. Importantly, the co-localisation between mutant *TAF8* mRNA and TAF10 protein was lost. In addition, TAF8 protein detection by IF and *TAF10* mRNA by smiFISH, showed no significant co-localisation, thereby further lending support to the sequential model of TAF10-TAF8 co-translational assembly. Next, we wanted to study if co-translational assembly is essential for the cell. To answer this question, we applied an indirect approach. We hypothesized that if nascent chains of a subunit cannot co-translationally interact with its partner, it may become prone to misfolding and degradation by the proteasome, but the fully translated partner should stay stable. By using these mouse ESCs we observed that the deletion of Taf10 not only ablated *Taf10* mRNA and TAF10 protein levels, but significantly reduced both *Taf8* mRNA and TAF8 protein expression. These results

were also confirmed in Taf10 KO mouse embryos. In contrast, the deletion of Taf8, decreased only its own mRNA and protein levels, without affecting the *Taf10* mRNA expression and TAF10 protein levels. In both KO mouse ESCs, other tested TFIID subunits remained unchanged, thereby ruling out the possibility of a primary transcriptional response. Additionally, we also observed that TAF10 re-expression rescued TAF8 from degradation. Thus, the nascent TAF8 HFD, in the absence of its interaction partner TAF10, may serve as a signal for both protein and mRNA degradation, while TAF10 is stable in the absence of TAF8.

Having shown the sequential co-translational assembly of TAF10 and TAF8, we next studied the assembly of a protein pair whose dimerization domains are localised towards the N-terminus. To test this, we studied the co-translational assembly of TAF6-TAF9 HFD pair as they interact through their N-terminal HFDs. Our nascent RIPs revealed that both TAF6 and TAF9 co-IP their partners' mRNA suggesting that they assemble in a bidirectional way, presumably as the neosynthesised interaction domains of both proteins are exposed early during their synthesis on the ribosomes. Further evidence from two colour smiFISH co-localization experiments showed a significantly higher co-localisation of the *TAF6* and *TAF9* mRNAs in the cytoplasm than several unrelated negative control mRNAs.

In order to rule out the possibility that this form of assembly is specific to only HFD protein pairs, we wanted to study the co-translational assembly of a non-HFD interacting protein pair. In TFIID, the evolutionary conserved core domain of TBP interacts with TAF1 via N-terminal TAND region of TAF1. Genome-wide microarray analysis of TBP-associated RNAs from HeLa cell polysome extracts revealed an enrichment of its own mRNA as well as 19 coding and non-coding RNAs. Among these, we found mRNAs coding for known TBP-interacting proteins: BRF1 coding for a factor important for Pol III transcription, BTAF1 coding for a B-TFIID subunit, as well as TAF1. RIP-qPCR analysis in human HeLa cells and mouse ESCs confirmed the microarray data and revealed a strong enrichment of the *TAF1* mRNA. To further investigate the specificity of TBP-TAF1 interaction, we generated a ΔTAF1 expression vector, in which sequences coding for the first 168 residues containing the TAND region were deleted. Anti-TBP RIPs from cells expressing ΔTAF1 resulted in complete loss of *TAF1* mRNA enrichment and a reduction of the co-immunoprecipitated protein. These results are consistent with a requirement of the N-terminal TAF1 domain to

recruit TBP to the nascent TAF1 polypeptide. As the protein interface is formed by the C-terminal portion of TBP and the very N-terminus of TAF1, we predicted that similar to TAF8-TAF10 assembly, a sequential assembly is also involved in the TBP-TAF1interaction. Indeed, nascent anti-TAF1 RIP from an engineered GFP-TAF1 HeLa cell line resulted in the enrichment of *TAF1* mRNA, but not that of *TBP*, thus supporting the sequential co-translational assembly model of TBP-TAF1 by the sequential pathway.

To extend our findings beyond TFIID, we examined co-translational assembly of the ENY2 subunit with its respective partners. ENY2 is subunit of the TREX-2 mRNA export complex and the DUB module of the SAGA transcription coactivator. In TREX-2, two ENY2 proteins wrap around the central portion of the large GANP helical scaffold. Similarly, human ENY2 wraps around the N-terminal helix of human ATXN7L3 in the highly intertwined SAGA DUB module. To test whether the co-translational model is generally applicable to multisubunit complexes, we analysed ENY2- associated mRNAs from HeLa cells stably expressing ENY2 with an N-terminal GFP-tag. Interestingly, we found that an anti-GFP-ENY2 RIP co-immunoprecipitates predominantly endogenous GANP mRNA and protein (the partner of ENY2 in TREX-2), and also endogenous ATXN7L3 mRNA and protein (the binding partner of ENY2 in the SAGA DUB module).

Conclusions:

Together, our study demonstrates that co-translational assembly is involved in the assembly of mammalian transcription complexes of diverse architecture and function. We also show that the localisation of the interacting domains of a protein dimer determines the co-translational assembly pathway, sequential or simultaneous. We also demonstrate that the nascent protein partner and its mRNA prone are to degradation in the absence of the fully translated co-translationally assembling partner.

Publication:

Kamenova I*, Mukherjee P*, Conic S, Mueller F, El- Saafin F, Bardot P, Marie Garnier J, Dembele D, Capponi S, HT Timmers M, Vincent S, Tora L (2019) Co-translational assembly of mammalian nuclear multisubunit complexes. Nature Communications

*equal first authors, names appear in alphabetical order

Communications:

1. **Flash talk and poster presentation**: Co-translation drives the assembly of mammalian nuclear multisubunit complexes
   Pooja Mukherjee, Ivanka Kamenova, Laszlo Tora
   SFB 1036 conference "Networks of Cellular Surveillance Mechanisms" October 22-24, 2018, Heidelberg, Germany.

2. **Poster presentation:** Co-translational assembly of mammalian nuclear multisubunit complexes
   Pooja Mukherjee, Ivanka Kamenova, Laszlo Tora
   TriRhena Transcription and Chromatin Club, March 26, 2019, Strasbourg, France.

3. **Short talk:** Co-translational assembly of mammalian nuclear multisubunit complexes
   Pooja Mukherjee, Ivanka Kamenova, Laszlo Tora
   Cell Symposia: Regulatory RNAs, May 12-14, Berlin, Germany.

Thesis summary in French

Introduction

La majorité des processus biologiques sont réalisés par des complexes protéiques contenant plusieurs sous unités. La transcription est le processus par lequel l'information contenue dans l'ADN est copiée dans des molécules d'ARN. La transcription par l'ARN polymérase II chez les eucaryotes est régie par six facteurs généraux de transcription, à savoir TFIID, TFIIA, TFIIB, TFIIF, TFIIE, TFIIH et le complexe multi protéique médiateur. Dans les cellules de mammifère, TFIID génère l'assemblage d'autres facteurs de transcription sur les promoteurs des gènes codants pour des protéines et la perte de TFIID est létale à l'état embryonnaire. TFIID est composé de la protéine TBP (*TATA-binding protein*) et de 13 TAF (*TBP associated factors*). La majorité des TAF partagent un domaine structurel commun appelé *Histone Fold Domain* (HFD), qui permet l'hétérodimérisation entre certains TAF, en particulier TAF10-TAF8, TAF3-TAF10, TAF4-TAF12 et TAF11-TAF13. Le domaine HFD n'est pas conservé au niveau de la séquence mais est conservé d'un point de vue structurelle, il est composé de trois hélices α reliées par deux boucles, qui permettent des interactions entre les TAF spécifiques. De plus, les protéines contenant un HFD ne sont pas solubles lorsqu'elles sont exprimées individuellement. Malgré des études approfondies sur la structure et la fonction de complexes transcriptionnels multi-protéiques, peu d'études se sont intéressées aux mécanismes d'assemblage de ces complexes. Il a déjà été montré que l'assemblage de TFIID se fait par étapes avant qu'il ne soit fonctionnel dans le noyau. Plusieurs sous-modules de TFIID ont été mis en évidence dans le cytoplasme avant la formation du complexe holo-TFIID. La présence d'hétérotrimères stables TAF5-TAF6-TAF9, TAF2-TAF8-TAF10 et TAF7-TAF11-TAF13 a été démontrée dans le cytoplasme. Dans ce contexte, le principal objectif de mon projet était d'étudier le mécanisme d'assemblage des sous-modules TFIID dans le cytoplasme. Un dimère protéique peut s'assembler de deux manières : un assemblage post-traductionnel où des sous-unités individuelles sont entièrement traduites et distribuées de manière aléatoire dans le cytoplasme, ou un assemblage co-traductionnel, où des interactions protéine-protéine se produisent au cours de la traduction des partenaires protéiques. L'assemblage co-traductionnel pourrait être

bénéfique pour la cellule en permettant d'éviter des interactions non spécifiques dans l'environnement cytoplasmique très encombré de la cellule. Une des premières observations d'assemblage co-traductionnel a été obtenue en 2009 où les auteurs ont étudié 31 protéines ne contenant pas de domaine de liaison à l'ARN de *Schizosaccharomyces pombe*. Parmis ces protéines, ~ 38% co-purifient avec des ARNm codant pour des protéines partenaires. Cette observation a été corroborée par une publication très récente qui montre un assemblage co-traductionnel chez S*accharomyces cerevisiae* par profilage sélectif des ribosomes. Pris ensemble, ces diverses évidences montrent que l'assemblage co-traductionnel pourrait être un mécanisme répandu pour l'assemblage d'un grand nombre de complexes multi-protéiques dans la levure. Ici, nous avons étudié en détail le mécanisme d'assemblage co-traductionnel du facteur général de transcription TFIID et avons étendu nos observations à d'autres complexes multi-protéiques dans des cellules de mammifère.

Objectifs:

1. Est-ce que les hétérodimères TAFs contenant des HFD s'assemblent de manière co-traductionnelle ?
2. Qu'est-ce qui dirige l'assemblage co-traductionnelle de hétérodimères TAFs à HFD ?
3. Est-ce que les hétérodimères TAFs ne contenant pas d'HFD s'assemblent également de manière co-traductionnelle ?
4. Est-ce que d'autres complexes multi-protéiques s'assemblent également de manière co-traductionnelle ?
5. Étude de la co-localisation des ARN et des protéines assemblées de manière co-traductionnelle.
6. Étude du rôle des chaperons dans l'assemblage co-traductionnel.

Résultats:

Afin de tester si les TAF contenant un HFD s'assemblent de manière co-traductionnelle, nous avons utilisé un anticorps monoclonal dirigé contre l'extrémité N-terminale de TAF10 contenant un HFD, pour immunoprécipiter la protéine TAF10 endogène à partir d'extraits de polysomes cytosoliques de cellules HeLa. L'immunoprécipitation d'une protéine associée à l'étude des ARNm associés est

appelée *RNA – ImmunoPrecipitation* (RIP). Les interactions protéine-protéine entre les protéines naissantes encore associées aux ribosomes sont révélées par l'enrichissement de l'ARNm codant pour les partenaires en interaction dans les IP. L'analyse globale de puces à ADN des ARNms précipités par les IP anti-TAF10 (RIP) a révélé un enrichissement de l'ARNm codant pour TAF8, suggérant que le dimère HFD TAF8-TAF10 bien caractérisé, se forme de manière co-traductionnelle. Les résultats des puces ont également été corroborés par le RIP anti-TAF10 couplé à la RT-qPCR dans les cellules HeLa et les cellules souches embryonnaires de souris (ESCs). Le traitement à la puromycine entraîne une perte significative de l'enrichissement des ARNm associés, excluant ainsi la possibilité d'une interaction directe entre l'ARNm codant pour TAF8 et la protéine TAF10. Ensuite, nous nous sommes intéressés à l'étude du mécanisme sous-jacent de l'assemblage co-traductionel de TAF10-TAF8. Nous avons généré des vecteurs d'expression exprimant TAF10 et TAF8 avec des étiquettes N-terminales ou C-terminales. Ces constructions nous ont permis d'immunoprécipiter la protéine naissante (étiquette N-terminale) et la protéine complète (étiquette C-terminale) et de déterminer ainsi l'ordre d'assemblage co-traductionnel des protéines. Conformément à nos résultats d'immunoprécipitation endogène (IP), le RIP de TAF10 naissant a révélé la présence de son propre ARNm, ainsi que l'ARNm de *TAF8*. Cependant, l'ARNm de *TAF10* n'a pas été co-immunoprécipité avec la protéine naissante de TAF8. Cette observation est corroborée également par des expériences d'immunoprécipitation avec des constructions TAF10 et TAF8 marqué à leurs extrémités C-terminale. La protéine mature complètement traduite de TAF10 a révélé la présence de seulement l'ARNm de *TAF8*, mais pas de son propre ARNm (comme attendu), mais la protéine mature de TAF8 ne donnait pas l'ARNm de *TAF10* ou *TAF8*. Dans tous les cas, la protéine TAF8 a été co-immunoprécipitée avec la protéine TAF10, ce qui exclut la possibilité d'une expérience IP infructueuse. Dans l'ensemble, ces résultats indiquent que la protéine TAF10 s'assemble de manière co-traductionnelle avec la protéine naissante de TAF8 de manière unidirectionnelle par un modèle d'assemblage séquentiel. Nous avons émis l'hypothèse que cette liaison séquentielle est spécifique à la localisation des domaines de dimérisation de TAF10 et TAF8, le HFD de TAF10 étant vers l'extrémité C-terminale et le HFD de TAF8 étant vers l'extrémité N-terminale. Pour tester cela, nous avons réalisé une mutation dans le HFD de TAF8 perturbant l'interaction entre TAF10 et TAF8. Cette mutation a entraîné une perte presque

complète de l'ARNm de TAF8 co-précipitée par rapport à la condition contrôle, ce qui indique que la dimérisation de TAF8 et TAF10 via leurs HFD est cruciale pour l'assemblage co-traductionnel. En outre, des expériences de co-localisation IF-smiFISH dans des cellules HeLa fixées ont montré une co-localisation significative entre la protéine TAF10 et l'ARNm de TAF8 dans le cytoplasme par microscopie confocale. Notablement, la co-localisation entre l'ARNm de TAF8 mutée et la protéine TAF10 a été perdue. En outre, la détection de la protéine TAF8 par IF et l'ARNm de TAF10 par smiFISH n'a montré aucune co-localisation significative, ce qui a permis de soutenir davantage le modèle séquentiel d'assemblage co-traductionnel de TAF10-TAF8. Ensuite, nous voulions étudier si l'assemblage co-traductionnel est essentiel pour la cellule. Pour répondre à cette question, nous avons utilisé une approche indirecte. Nous avons supposé que si les chaînes naissantes d'une sous-unité ne peuvent pas interagir de manière co-traductionnelle avec son partenaire, elles peuvent devenir sujettes au mauvais repliement et à la dégradation par le protéasome, mais le partenaire entièrement traduit devrait rester stable. En utilisant ces cellules souche de souris, nous avons observé que la suppression de Taf10 non seulement supprimait les niveaux d'ARNm de *Taf10* et de la protéine de TAF10, mais réduisait de manière significative l'expression de l'ARNm de *Taf8* et la protéine de TAF8. Ces résultats ont également été confirmés chez des embryons de souris KO pour Taf10. En revanche, la suppression de Taf8 n'a entraîné que la diminution des niveaux de ses propres ARNm et protéines, sans affecter l'expression de l'ARNm de Taf10 ni les taux de protéines TAF10. Dans les deux lignes KO des cellules souches de souris, les autres sous-unités TFIID testées sont demeurées inchangées, excluant ainsi la possibilité d'une réponse transcriptionnelle primaire. De plus, nous avons également observé que la ré-expression de TAF10 a empêché la dégradation de TAF8. Ainsi, le HFD naissant de TAF8, en l'absence de son partenaire d'interaction TAF10, peut servir de signal pour la dégradation de ses protéines et ses ARNm, tandis que TAF10 est stable dans l'absence de TAF8. Après avoir montré l'assemblage séquentiel co-traductionnel de TAF10 et TAF8, nous avons ensuite étudié l'assemblage d'une paire de protéines dont les domaines de dimérisation sont localisés vers l'extrémité N-terminale. Pour tester cela, nous avons étudié l'assemblage co-traductionnel de la paire de HFD de TAF6-TAF9 lors de leur interaction par leurs HFD N-terminaux. Nos RIP naissants ont révélé que TAF6 et TAF9 co-immunoprécipitent les ARNm de leurs partenaires suggèrent qu'ils s'assemblent de manière bidirectionnelle, vraisemblablement au

moment où les domaines d'interaction néosynthétisés des deux protéines sont exposés tôt pendant leur synthèse par les ribosomes. D'autres preuves provenant d'expériences de co-localisation smiFISH à deux couleurs ont montré une co-localisation significativement plus élevée des ARNm de TAF6 et TAF9 dans le cytoplasme que plusieurs ARNm de contrôle négatif non corrélatifs. Dans le but d'exclure la possibilité que cette forme d'assemblage protéique est spécifique aux sous-unités de TFIID contenant des *HFDs*, nous voulions étudier l'assemblage co-traductionnel de pair de protéine ne contenant pas d'*HFDs*. Au sein de TFIID, l'interaction entre TBP et TAF1 se fait d'une part par un domaine central évolutivement conservé et le côté N-terminal contenant une région TAND respectivement. Une analyse sur micro-puce à l'échelle du génome global des ARNm associés à TBP, obtenus à partir d'extrait de polysome de cellules HeLa a révélé que la protéine TBP était majoritairement associée à ses propres ARNm ainsi qu'à 19 autres ARNs codant et non codant. Parmi eux se trouvent les ARNm codant pour des protéines connues pour interagir avec TBP comme BRF1 qui est important pour la transcription médiée par l'ARN polymérase III, BTAF1 présent dans le complexe B-TFIID ainsi que TAF1. Des analyses par *RIP-qPCR* à partir d'extrait de cellules HeLa et de cellules *ES* murines ont confirmé les résultats obtenus sur micro-puces, dont un fort enrichissement de l'ARNm TAF1. Afin d'étudier la spécificité de l'assemblage de TAF1-TBP, un vecteur exprimant une version mutée de TAF1 (noté ΔTAF1) a été généré dans laquelle 168 acide-aminés présent au sein de la région TAND ont été supprimés. Une expérience de *RIPs* contre TBP à partir des cellules exprimant ΔTAF1 a montré une complète absence d'enrichissement en ARNm TAF1 et une réduction de la quantité de protéine TAF1 immunoprécipité. Ces résultats sont en accord avec le fait que le domaine en position N-terminal de TAF1 est requis pour recruter TBP au niveau de TAF1 en train d'être synthétisé. Comme l'interaction se fait par l'extrémité C-terminal de TBP et par l'extrémité N-terminal de TAF1, nous prédisons que similairement à l'assemblage de TAF8-TAF10, un assemblage séquentiel a aussi lieu pour l'interaction TBP-TAF1. En effet, une expérience de *RIPs* contre TAF1 naissant à partir d'une lignée cellulaire HeLa exprimant GFP-TAF1 a montré un enrichissement pour l'ARNm *TAF1*, mais pas de l'ARNm *TBP*, ce qui supporte cette hypothèse.

Pour élargir ces découvertes au-delà de TFIID, nous avons examiné l'assemblage co-traductionnel de ENY2 avec ses partenaires respectifs. ENY2 est une sous-unité du

complexe d'export des ARNm nommé TREX-2 ainsi que du module DUB du co-activateur de la transcription SAGA. Dans le complexe TREX-2, deux protéines ENY2 s'enroulent au niveau du squelette hélicoïdal de la protéine GANP. De façon similaire, dans le module DUB, ENY2 humain s'enroule autour de l'hélice en position N-terminal d'ATXN7L3. Pour tester si le modèle d'assemblage co-traductionnel peut-être généralement appliqué à l'assemblage des complexes protéiques, nous avons analysé les ARNm associés à ENY2 au sein de cellules HeLa exprimant de façon stable ENY2 fusionné en N-terminal à un tag GFP. Nous avons trouvé que GFP-ENY2 co-immunoprécipite principalement les ARNm de la protéine GANP endogène ainsi que GANP lui-même (le partenaire d'ENY2 dans TREX-2), et aussi les protéines et les ARNm de ATXN7L3 endogène (le partenaire d'ENY2 dans le module DUB).

Conclusions :

Notre étude démontre que l'assemblage co-traductionnel est impliqué dans l'assemblage de complexes de transcription de mammifères d'architecture et de fonctions diverses. Nous montrons également que la localisation des domaines en interaction d'un dimère de protéine détermine la voie d'assemblage co-traductionnel, séquentielle ou simultanée. Nous démontrons également que le partenaire protéique naissant et son ARNm sont susceptibles de se dégrader en l'absence du partenaire d'assemblage complètement traduit.

INTRODUCTION

Introduction

The cell uses a vast array of protein complexes to carry out important biological functions. Such protein complexes are made up of either multiple copies of same subunits (homomeric protein complexes) or different subunits (heteromeric protein complexes). The latter group includes various transcription regulatory and chromatin remodelling complexes. In order to achieve correct assembly of protein complexes, the genes encoding the subunits (dispersed in the genome) must be transcribed in the nucleus, transported to the cytoplasm, translated by the ribosomes and finally find their interacting partners. All these processes must be correctly orchestrated to achieve efficient assembly of protein complexes. Following translation in the cytoplasm the protein subunits can find their interacting partners either post-translationally by random collision in the cytoplasm. This can lead to non-specific interactions and the cell might end up spending more energy to assemble protein complexes by this method. Co-translational association of protein complexes can result in efficient formation of protein complexes preventing unwanted interactions. In fact, this model of protein assembly has been put forward in bacteria and yeast.

The major step linking our genome and its final product, protein that carries out functions in the cell, is transcription (synthesis of RNA from DNA). One of the key regulatory steps is transcription initiation. The protein complexes involved in transcription initiation include general transcription factors (GTFs), activators, co-activators, etc. As mentioned above, majority of these complexes are multisubunit. Though their functional aspects *in vivo* have been studied extensively, their assembly is not well-studied so far. Certain reports suggest the existence of submodules of transcription complexes including RNA Polymerases and TFIID but there has been no concrete report so far on their order of assembly. Another intriguing aspect that is still unanswered is how the cell allocates the common subunits to different complexes as many of these transcription complexes consist of common subunits.

The first two sections of my introduction will focus on the various steps of transcription and the key protein complexes involved in it. The third section mainly deals with different types of protein complexes in the cell. The fourth, fifth and seventh sections describe the various aspects of co-translational protein assembly. Finally, the sixth section describes the various mRNA and protein quality control mechanisms.

1. Eukaryotic gene expression

1.1 Chromatin organisation

Living organisms store genetic information in the form of the nucleic acids DNA and RNA. The message in our genome is transferred to RNA by a process called transcription. These RNAs in turn act as messengers of protein synthesis or translation and it is the protein that carries out various functions in the cell. This sequential flow of information constitutes the central dogma of molecular biology (Crick 1970). Although this is the basis of molecular biology, certain discoveries also describe the backward flow of genetic information from RNA to DNA (Iwre 1970; Baltimore 1970). Our genetic material DNA is organised hierarchically in the form of chromatin in the cell (***Figure 1***). The first level of organisation is the wrapping of 145-147 bp DNA around a scaffold of core histone octamer (two copies each of histone H2A, H2B, H3, H4) forming a dynamic protein-DNA complex called the nucleosome core particle (NCP) (Arents et al. 1991; Fa 1973; Richmond et al. 2003; Woodcock et al. 1976; Richmond et al. 1997; Arents 1993). H2A-H2B and H3-H4 form heterodimeric protein pairs through a specific domain called the histone fold domain (HFD) and their long N-terminal unstructured regions exist in the form histone tail extensions (Richmond et al. 1997; Arents et al. 1991). The DNA is further packaged into chromatosome core particle which is a recurring structural unit consisting of linker histone H1 bound to the nucleosome core particle (Simpson 1978). The linker histone further aids in the formation of a chromatin fiber of diameter 30 nm (Finch and Klug 1976). In addition, chromatin can also form loops and topologically associating domains (TADs) (Sexton et al. 2012; Hou et al. 2012; Dixon et al. 2012). Several TADs can associate and form a chromatin superdomain (Pueschel et al. 2016). The chromatin structure finally gets condensed to a chromosome. (***Figure 1***) (R. Kornberg 1974). The structure of chromatin is essential for the regulation of several DNA-related metabolic processes including transcription, recombination, DNA repair, replication, kinetochore, centromere formation, etc (Fyodorov et al. 2018). Over ten decades ago, it was established that some fractions of the chromosomes were stained very intensely with nuclear dyes, while other areas were only weakly stained (Passarge 1979). This finding formed the basis for defining “euchromatin” and “heterochromatin” respectively. Euchromatin is gene-rich and has open chromatin conformation whereas

heterochromatin is condensed and contains mainly repetitive elements (Huisinga et al. 2006; Grewal and Jia 2007).

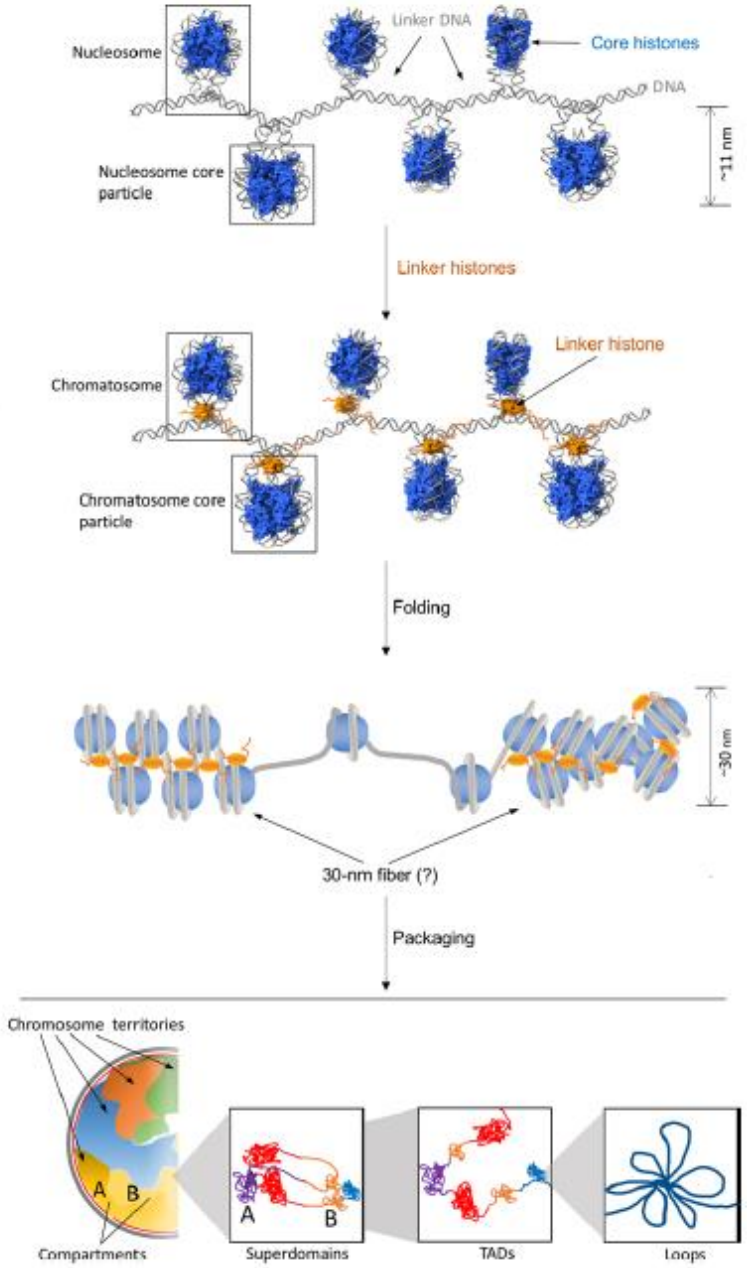


Figure 1: Multiple levels of chromatin folding. DNA is wrapped around the nucleosome and undergoes compaction to accommodate the DNA in the limited volume of the nucleus. Image is from (Zhou and Bai 2019).

1.2 Nucleosome remodelling

Chromatin structure imposes significant obstacles upon the transcription machinery. The multiple interactions between histone and DNA makes the nucleosome one of the most stable protein-DNA complexes under physiological conditions (Richmond et al. 1997). At the same time, chromatin is dynamic as well and is tightly regulated by various protein complexes involved in histone modification, chromatin remodelling, histone variant incorporation, and histone eviction. Both **histone-modifying enzymes** and **chromatin remodelling complexes** bind and make DNA accessible to transcription machinery.

1.2.1 Histone-modifying enzymes

Histone-modifying enzymes target both histone tails and globular domains for posttranslational modifications. Some of the main modifications and their functions are shown in ***Figure 2*** and ***Figure 3***. Modifications associated with active transcription include acetylation of H3 and H4, di- or trimethylation of H3K4 and are commonly referred to as euchromatin modifications. Modifications can also occur at specific amino acid residues, for example, methylation of arginine residues, phosphorylation of serine and threonine, lysine residues can undergo a wide range of modifications including methylation, acetylation, ubiquitination, ADP-ribosylation and SUMOylation. Other modifications like propionylation and butyrylation have also been described very recently (Kebede et al. 2015). A detailed description of all histone modifications and their functions are reviewed in (Lawrence et al .2016).

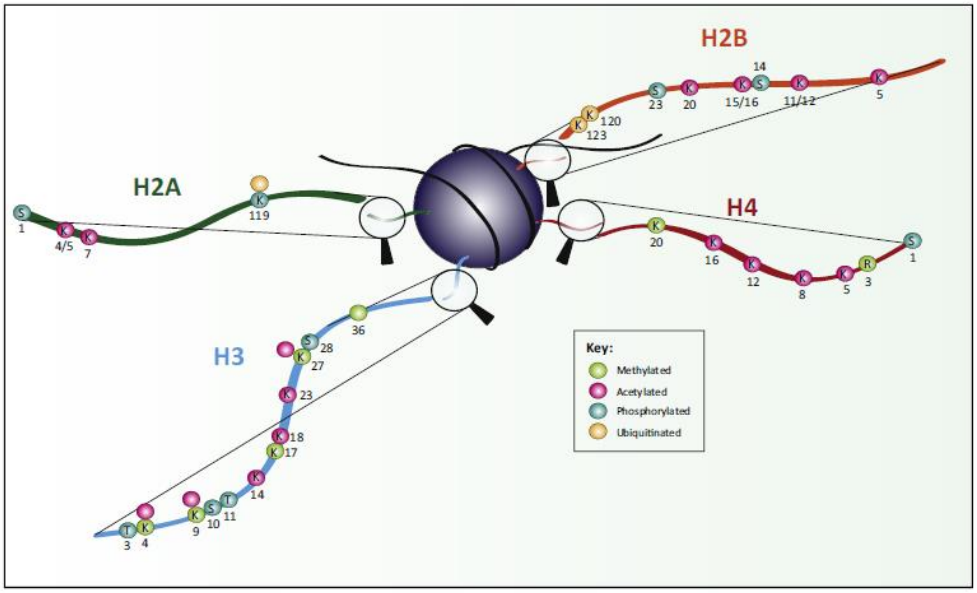


Figure 2: Schematic representation of post-translational modifications of histone tails. The amino acid modified is shown (K = lysine, R = arginine, S = serine, T = threonine) and the position of each modification is shown in black. Colours depict the nature of modification of each residue (green = methylated, pink = acetylated, turquoise = phosphorylated, beige = ubiquitinated). Image is from (Lawrence et al. 2016).

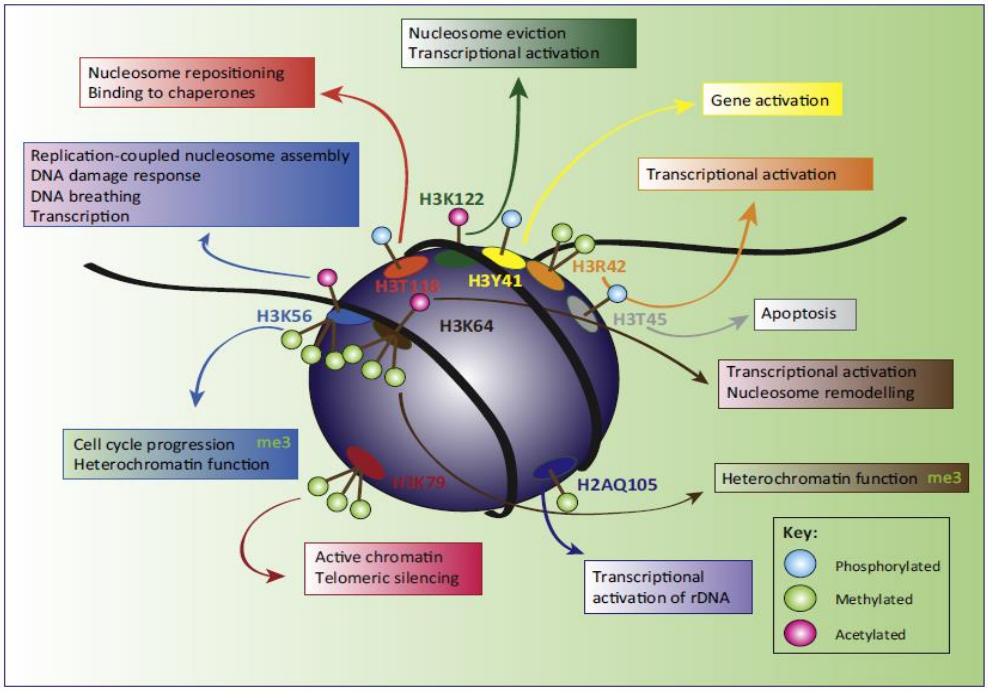


Figure 3: Schematic representation of key post-translational modifications within globular domains of histones and their functions. Methyl marks are shown in light green, acetyl marks in pink, and phosphorylated residues in light blue. Image is from (Lawrence et al. 2016)**.**

1.2.2 Chromatin remodelling complexes

Chromatin remodelling complexes (also called chromatin remodellers, (Clapier et al. 2017) are protein complexes that alter nucleosome structure in an ATP-dependent manner. Various subfamilies of chromatin remodelling complexes carry out diverse functions including transient unwrapping of DNA-histone octamers, forming DNA loops, and nucleosome sliding. Based on the subunit composition and their catalytic ATPases, chromatin remodellers can be classified into four subfamilies: 1) imitation switch (**ISWI**), 2) chromodomain helicase DNA-binding (**CHD**), 3) switch/ sucrose non-fermentable (**SWI/SNF**) and 4) inositol requiring 80 (**INO80**). Higher eukaryotes contain multiple remodeller subtypes within each family that are specific to certain cell types or developmental stages (Lessard and Crabtree 2010; Bao and Shen 2007). In addition to this, other remodellers, which do not belong to the previously mentioned subfamilies (for example, α-thalassaemia/mental retardation syndrome X-linked (ATRX (Law et al. 2010; Lewis et al. 2010) and Cockayne syndrome group B (CSB) (Citterio et al. 2000)) also exist but are less well-characterized mechanistically.

ISWI family of remodellers include ISWI complexes in yeast and NURF (Nucleosome Remodeling Factor) complex in humans. They are made up of 2 to 4 subunits and their functions include organising nucleosome spacing to facilitate chromatin assembly and therefore transcription repression. However, certain complexes in this family can also assist in RNA Polymerase II (RNA Pol II) activation, thereby suggesting that the functional diversity is mainly imparted by the presence of distinct subunits in each complex (Becker and Workman 2013). The **CHD** family of remodellers consists of complexes that can promote transcription by sliding or ejecting nucleosomes or they may also have transcription repressive roles. For example, the vertebrate NuRD (nucleosome remodelling and deacetylase) complex plays a role in chromatin compaction by its histone deacetylase activity (Fei et al. 2015). **SWI/SNF** family of chromatin remodellers include SWI/SNF complex in yeast and BAF (BRG1-Associated Factor complex) complex in human, are composed of 8 to 14 subunits. This family of remodellers has a role in sliding and ejection of nucleosomes at many loci (Narlikar et al. 2013). **INO80** group of remodellers include SWR1-related (SWI2/SNF2-Related 1) complexes, (which was first purified from yeast) and human Tip60/TRAPP (Tat interactive protein 60-kDa/ transformation-transactivation domain-associated protein) complex. They are made up of more than 10 subunits and play

roles in diverse functions including transcriptional activation and DNA repair. Although belonging to INO80 group, SWR1 has the unique ability to replace H2A of canonical H2A-H2B dimer with histone variant H2AZ to form H2A.Z-H2B dimer. Due to functional differences among the various families of remodelling complexes, earlier models suggested the use of distinct enzymatic mechanisms to achieve their diverse functions. In contrast, a more unified model was proposed recently which puts forward the fact that all histone remodellers use a common ATP-dependent DNA translocation mechanism to move DNA along the histone surface (Clapier et al. 2017). These complex series of events finally result in the accessibility of nucleosome wrapped DNA to transcription factors. Following this, transcription is carried out on “naked” DNA by the transcription machinery. The basic mechanism of transcription is conserved between prokaryotes and eukaryotes even though the transcription machinery is more complex in eukaryotes (Hahn 2004).

1.3 RNA Polymerases

Dedicated transcription machinery exists in the cell that carries out transcription. The enzyme responsible for transcription is RNA Polymerase. The enzymatic activity of RNA Polymerase was first discovered decades ago in rat liver nuclei (Weiss and Gladstone 1959). It was later discovered in *E.coli* as well (Hurwitz et al. 1961; A. Kornberg 1961) . Five different RNA polymerases (named I to V) have been discovered so far in eukaryotes whereas only one RNA Polymerase has been identified so far in prokaryotes and Archaea. The archaeal RNA Polymerase and eukaryotic RNA Polymerase II are structurally and mechanistically closely related (Korkhin et al. 2009). In eukaryotes, different polymerases transcribe different classes of cellular RNAs (Kedinger et al. 1970; Weinmann et al. 1974; Warfel 1970; Zylber and Penman 1971). Pol I synthesizes 18S and 28S ribosomal RNAs (rRNA), Pol II synthesizes messenger RNAs (mRNAs), small nuclear RNAs (snRNAs) and microRNAs (miRNAs), Pol III synthesizes transfer RNAs (tRNAs) and 5S rRNAs and certain viral RNAs (Zylber and Penman 1971; Weil and Blatti 1976; Roeder and Rutter 1970; Reinberg et al. 2004). The recently identified RNA Polymerase IV and V are required for the production of siRNAs (small interfering RNAs) in plants, mediating RNA-directed DNA methylation, transcriptional silencing and heterochromatin formation (Wierzbicki et al. 2009; Onodera et al. 2005; Kanno et al. 2005). Nevertheless, despite differences in functionality, all RNA Polymerases contain a

structurally and functionally conserved core (Vannini and Cramer 2012) and they require a set of general transcription factors, specific to each of them, to carry out their functions (Roeder 1996; Orphanides et al. 1996). Since RNA Pol I, III, IV and V are mainly involved in the transcription of non-coding RNAs, we will be mainly focusing on the mechanism of Pol II transcription and the various factors involved in it.

1.4 RNA Polymerase II transcription

The starting point for the study of the intricate details of RNA Pol II transcription was the isolation of a transcriptionally active form of Pol II from yeast (Kornberg 1987). Pol II is aided by many different transcription factors that allow the Pol II to gain access to the DNA and transcribe a gene efficiently. The DNA sequences necessary for transcription consist of the following sequence elements: the core promoter which drives the assembly of the pre-initiation complex (PIC), promoter-proximal regions and distant enhancer sequences that recruit transcription factors activating (activator) or repressing (repressor) transcription (Venter et al. 2001).

The process of transcription can be carried out in eight distinct steps which begins with the remodelling of chromatin to allow the access of Pol II and GTFs to the DNA (**Step 1**). The pre-initiation complex (PIC) then assembles on the core promoter (**Step 2**) followed by unwinding of DNA, transcription bubble formation and transcription initiation (**Step 3**). After clearing the promoter, Pol II proceeds to the promoter-proximal pause region (**Step 4**) where it is hyperphosphorylated and proceeds to elongation (**Step 5**). In case Pol II is unable to carry out elongation, it undergoes termination. Otherwise, it productively elongates the entire gene (**Step 6).** Following elongation, Pol II under goes termination (**Step 7**) and reinitiates a new round of transcription (**Step 8**).

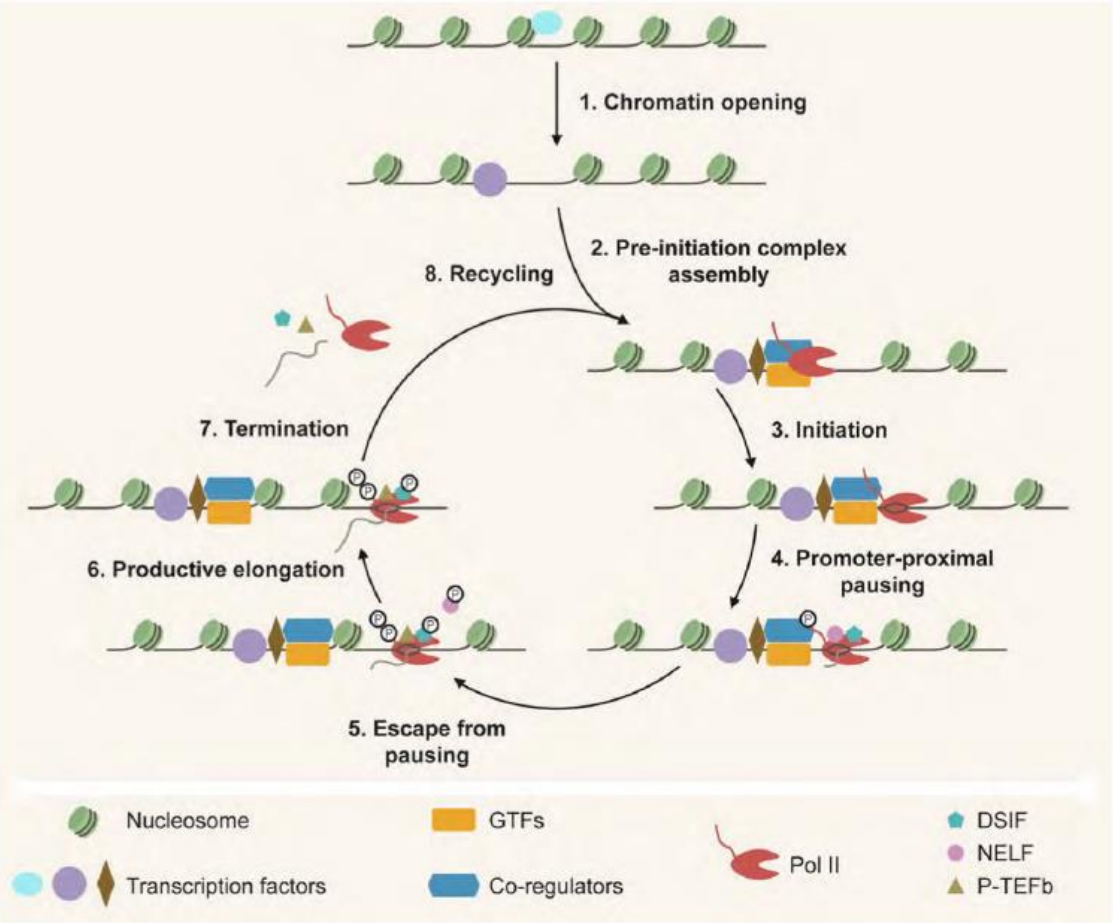


Figure 4: **Schematic representation of the different steps of RNA Pol II transcription** (based on (Fuda, Ardehali, and Lis 2009)).

1.4.1 Chromatin opening

1.4.1.1 Activators

Transcription is generally blocked due to chromatin structure formed by nucleosome bound DNA. Activators recruit co-activators which are nucleosome remodellers (described in section 1.2.1) and nucleosome-modifying enzymes (described in section 1.2.2) that allow the transcription machinery to access the promoter. Activators bind to distal cis-regulatory elements called enhancers (Hu and Tee 2017; Jin et al. 2013). Enhancers can exist in three different states based on their histone modification: typically, the active state shows methylation of lysine 3 and acetylation of lysine 27 of histone H3 (H3K3me3 and H3K27ac), the silent or repressive state shows histone H3 lysine 27 trimethylation (H3K27me3) (Tee and Reinberg 2014; Ernst and Kellis 2010). The third state called the 'poised' state harbour both repressive (H3K27me3) and active (H3K27me1) histone marks. These enhancers are associated with

developmental genes which are lowly expressed in embryonic stem cells (ESCs) but expression increases with the onset of differentiation (Rada-Iglesias et al. 2012; Bernstein et al. 2006). An activator is made up of a DNA binding domain that binds the DNA and an effector domain that carries out the activator function. Activators bind to DNA either by “DNA sequence” specificity (Rohs et al. 2010) or by “DNA shape” specificity (Stella et al. 2010) or by an interplay of both the mechanisms. Transcription factors (TFs) (both activators and repressors) can be classified into three groups: pioneers, settlers and migrants (Slattery et al. 2014). ‘Pioneer TFs’ can bind to the closed chromatin and make it accessible for gene activation (Cirillo et al. 2002). Pioneer factors belong to diverse structural classes like FoxA, Gata, Oct3/4, Sox2, Klf4, Pax7, Ascl1, p53 but they have common features that include binding to closed chromatin and leading to its opening (Iwafuchi-Doi and Zaret 2016). ‘Settler TFs’ can only bind to all of their specific DNA target sites in accessible chromatin regions but cannot bind to inaccessible regions. ‘Migrant TFs’ can only bind to a subset of their DNA target sites in accessible chromatin regions. Additional interactions with other cofactors are necessary to efficiently bind to their target sites (Sherwood et al. 2014; Slattery et al. 2014).

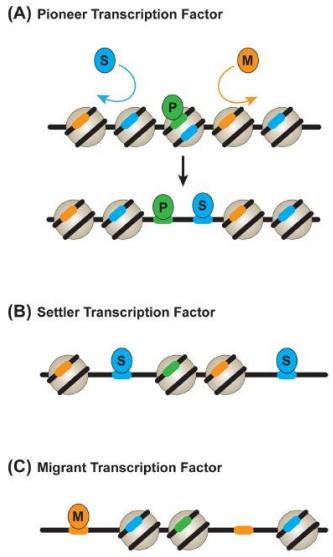


Figure 5: Transcription factor binding models. A) Pioneer factor (in green) can bind to inaccessible, nucleosome bound DNA sites and create an open chromatin environment necessary for the binding of non-pioneer transcription factors. B) Settler transcription factor (in blue) can bind to all their specific DNA target sites in open chromatin. C) Migrant transcription factor (in orange) can bind to a subset of their DNA target sites. Image is from (Slattery et al. 2014)**.**

1.4.1.2 Coactivators

Recruitment of coactivators is triggered once activators bind to the promoter. Coactivators include chromatin remodelling complexes, histone-modifying enzymes, and mediators. They act by facilitating stronger binding of activators to DNA as well as facilitating the binding of general transcription factors (GTFs) to DNA (Thomas and Chiang 2006). ***Table 1*** (Krasnov et al. 2016) is a list of different coactivators and their function in different steps of gene expression.

Table 1: **Different coactivator families and their functions** (Adapted from (Krasnov et al. 2016))

Family	**Function**
Mediator	Nucleosome removal, GTF and Pol II recruitment and stabilization.
Chromatin remodelling complexes (SWI/SNF, ISWI, CHD, Ino80)	Nucleosome removal, Activator and GTF recruitment, transcription initiation and productive elongation.
Methyltransferase (Set1/MLL, CARM1 PRMT1, Set2)	Histone modification leading to nucleosome removal and productive elongation.
Demethylase (KDM5/ Lid)	Removal of H3K4me3 modification and initiation of RNA synthesis.
Acetyltransferase (SAGA, CBP/p300)	Activator and GTFs recruitment.
Ubiquitin ligase (Bre)	Productive elongation.

Deubiquitinase (SAGA DUB module)	Repressor removal and productive elongation.

1.4.2 Transcription preinitiation

1.4.2.1 Preinitiation complex

The necessity of accessory factors in RNA Pol II transcription was first observed by invitro transcription experiments, where crude subcellular fractions were supplemented to accurately transcribe adenovirus DNA template (Weil et al. 1979). Purification of several factors from this subcellular fraction led to the discovery of the general transcription factors (GTFs) and they were named TFIID, TFIIA, TFIIB, TFIIF, TFIIE, TFIIF. Apart from these, activators and coactivators are also necessary for transcription as mentioned above in section 1.4.1 (Thomas and Chiang 2006).

The cooperative assembly of these basal/general transcription factors (GTFs) with RNA Polymerase II on the core promoter form the preinitiation complex (PIC) (Buratowski 1994; Orphanides et al. 1996; Roeder 1996). GTFs include TATA-binding protein (TBP), TFIIA, TFIIB, TFIIE, TFIIF, and TFIIH. Among them, TFIIA is dispensable and is only essential under certain conditions (Ozer et al. 1994; Yokomori et al. 1994; Sun et al. 1994). TFIIE, TFIIF and TFIIH are necessary for transcription from negatively supercoiled templates (Goodrich and Tjian 1994; Timmers 1994; Parvin et al. 1994; Parvin and Sharp 1993). Subunit composition and function of each GTF is written in ***Table 2***.

Table 2: General transcription factors (GTFs) involved in Pol II transcription. (Adapted from (Thomas and Chiang 2006))

Factor	**Composition**	**Function**
TFIID	TBP, TAFs (TAF1-13)	Core-promoter binding factor; Coactivator; Many TAFs were suggested to bind activators

TFIIA	p35 (α), p19 (β), p12 (γ)	Antirepressor; stabilizes TBP-TATA complex; coactivator
TFIIB	p33	Transcription start site selection; stabilises TBP-TATA complex; Pol II/ TFIIF recruitment
TFIIF	RAP30 and RAP74	Binds Pol II and facilitates its recruitment to promoter; recruits TFIIE and TFIIH; functions with TFIIB and Pol II in start site selection; facilitates Pol II promoter escape; enhances efficiency of Pol II elongation
TFIIE	p56 (α) and p34 (β)	Recruits TFIIH; facilitates formation of initiation-competent Pol II; involved in promoter clearance
TFIIH	P89/XPB, p80/XPD, p62, p52, p44, p40/CDK7, p38/Cyclin H, p34, p32/MAT1, p8/TFB5	ATPase activity for transcription initiation and promoter clearance; helicase activity for promoter opening; transcription-coupled nucleotide excision repair; kinase activity for phosphorylating Pol II CTD; E3 ubiquitin ligase activity

1.4.2.2 *Preinitiation complex assembly*

General transcription factors TFIID, TFIIA and TFIIB help to position the RNA Pol II to the core promoter, thereby forming a “closed form” of the preinitiation complex (PIC). This form of PIC is in its inactive state. The ATP dependent helicase activity of TFIIH

then aids in the formation of an “open complex” by melting 11-15 bp of DNA and positioning the single-strand template DNA to initiate RNA synthesis (Wang et al 1992; Kim et al. 2000). This is followed by the phosphorylation of the carboxy-terminal domain (CTD) of RNA Pol II by TFIIH while transcribing the first 30 bp of template DNA. This leads to the loss of contact with other GTFs and RNA Pol II proceeds into the elongation stage and starts transcribing in a highly processive manner. The part of the complex that is still attached to the promoter acts as a complex for re-initiation of transcription following the addition of TFIIB and TFIIF/RNA Pol II (Hahn 2004; Yudkovsky et al. 2000). The phosphorylated CTD starts to recruit factors necessary for productive elongation and mRNA processing (Buratowski 2003; Li et al. 2007). Release of the newly synthesized RNA marks the termination of transcription (Saunders et al. 2006).

Apart from the model described above, another model named as the RNA Pol II holoenzyme model has been put forward. Here, TFIID bound to the core promoter and stabilised by TFIIA, recruits preassembled RNA Pol II holoenzyme complex (RNA Pol II along with other GTFs) (Ossipow et al. 1995; Thomas and Chiang 2006). Although *in vitro* studies provide evidence for both the models, there has been no conclusive evidence of the *in vivo* prevalence of either models so far (Thomas and Chiang 2006).

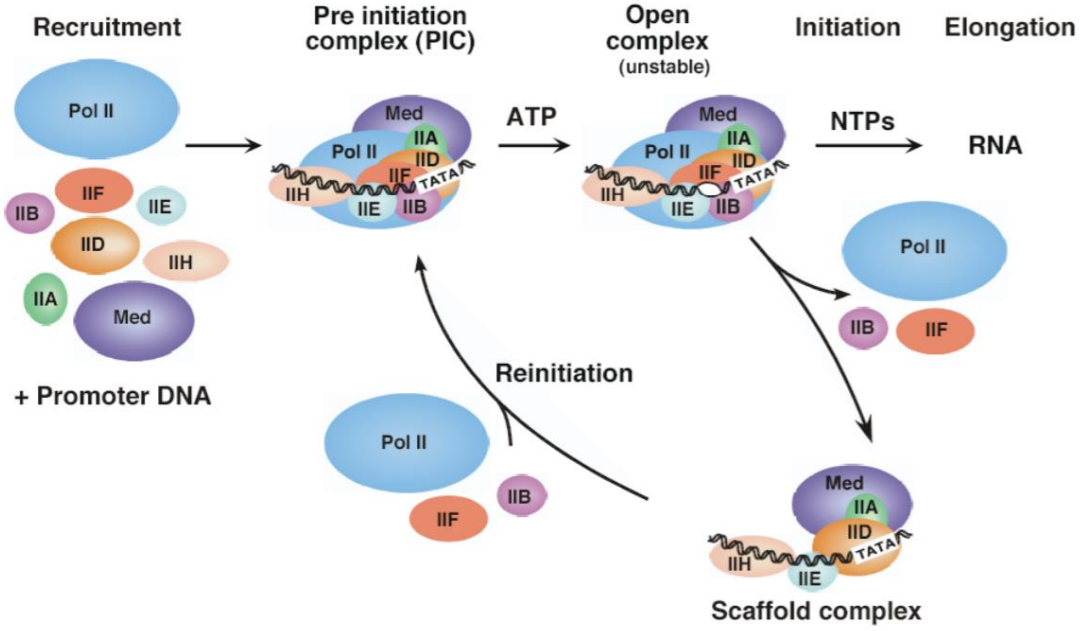


Figure 6: RNA Pol II transcription PIC assembly and transcription reinitiation. PIC is formed on the promoter by the recruitment of GTFs TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH along with Mediator (Med) and Pol II. Following the start of elongation, Pol II is released

from the PIC and continues elongation. A scaffold complex made up of TFIIA, TFIID, TFIIE, TFIIH and Mediator remains bound to the core promoter and helps in the reinitiation of second round of transcription. Image is from (Hahn 2004).

1.4.3 Transcription initiation

The ATP-dependent helicase activity of bound TFIIH within the promoter bound PIC is essential for unwinding the promoter and forming a 'transcription bubble'. This leads to the formation of an open PIC. TFIIB, TFIIE and TFIIF helps in the formation and stabilisation of the open promoter complex. Following this, Pol II starts synthesising RNA in the presence of NTPs. It dissociates from the promoter bound GTFs after being phosphorylated at Ser5 and Ser7 by TFIIH and continues transcription (Sainsbury et al. 2015).

1.4.3.1 Promoter proximal pausing and escape

Following promoter escape, many genes in metazoans undergo Pol II pausing, which typically happens after Pol II synthesises 30-50 nucleotides (Haberle and Stark 2018; Adelman and Lis 2012). Negative elongation factor (NELF) and DRB sensitivity inducing factor (DSIF) have been reported to interact with nascent RNA and mediate Pol II pausing independent of promoter sequence and chromatin structure (Yokoyama et al. 2017; Bernecky et al. 2017; Vos et al. 2018). Several hypotheses have been put forward to understand the functional aspect of promoter proximal pausing and include regulatory signals, checkpoint for transcription elongation-RNA processing coupling, establishing permissive chromatin and rapid gene activation (Adelman and Lis 2012). Escape from this pause is facilitated by factors like positive transcription elongation factor b (P-TEFb), which phosphorylates NELF, DSIF and Ser2 of Pol II CTD (Kwak 2013). This phosphorylation leads to the release of NELF, conversion of DSIF to positive transcription elongation factor and release of Pol II pause (Jonkers and Lis 2015).

1.4.4 Transcription elongation

Productive transcription elongation is tightly linked to chromatin structure through co-transcriptional covalent modification of histones and DNA and nucleosome turnover. A fine balance between nucleosome disassembly and reassembly exists to facilitate the forward movement of RNA Pol II as well as preventing cryptic transcription from intergenic regions (Talbert and Henikoff 2017; Lai and Pugh 2017; Venkatesh and

Workman 2015). For instance, FACT (facilitates chromatin transcription) along with other chromatin remodellers such as CHD1 and ISWI plays role in the dynamic regulation of nucleosomes during transcription (Petesch and Lis 2012; Teves et al. 2014). Polymerase associated factor-1 complex (PAF1C) travels with the elongating Pol II and facilitates co-transcriptional processing by acting as a scaffold to recruit various nucleosome remodellers CHD1 (Simic et al. 2003), histone chaperones SPT6 (Kaplan et al. 2005) and FACT (Pavri et al. 2006; Pruneski et al. 2011) and histone modifiers (complex proteins associated with SET1 (COMPASS) (Dean et al. 2004), E3 ubiquitin- protein ligase BRE1 (Wood et al. 2003) and histone- lysine *N*-methyltransferase, H3K79 specific (also known as DOT1L) (Wood et al. 2003; Dean et al. 2004).

A variety of posttranslational modifications are also associated with transcription elongation. For instance, H3K36me3 is deposited in gene bodies by the histone methyltransferase SETD2 which was shown to be associated with the phosphorylated CTD of Pol II. H3K36me3 affects RNA splicing and prevents cryptic transcription (McDaniel and Strahl 2017; Venkatesh and Workman 2015). Other modifications like H3K79me2 and H3K79me3 have also been discovered but the significance of these modifications is not well understood (Zhu et al. 2005; Nakanishi et al. 2009; Lee et al. 2007; Wang et al. 2013; Dean et al. 2004). Pol II elongation rate is not uniform within different regions of a gene and among different genes. In mammals, productive elongation ranges from around 0.5kb/min within the first few kilobases to 2-5kb/min after around 15 kilobases (Jonkers and Lis 2015). In addition, Pol II can be slowed by various co-transcriptional RNA processing events like splicing (Jonkers and Lis 2015).

1.4.5 Transcription termination

Transcription termination is not only essential for defining the boundaries of genetic information but also influences the fate and half-life of the newly synthesised mRNA since many RNA processing factors associate co-transcriptionally with the terminating RNA. Nevertheless the transition of elongating RNA Pol II to transcription termination at the 3' end of genes is coupled to RNA cleavage and polyadenylation carried out by various protein complexes like cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CSTF) (Kuehner et al. 2011; Porrua and Libri 2015). CSTF binds to the phosphorylated Ser-2 residue of Pol II CTD (Kuehner et al. 2011; Porrua and Libri 2015) and helps CPSF to recognise the polyadenylation signal

‘AAUAAA’ and cleave the nascent RNA downstream of it (Porrua and Libri 2015). Two different models for Pol II expulsion from DNA template following transcription termination have been put forward, namely, the torpedo model and the allosteric model. According to the torpedo model, cleavage of the nascent transcript is coupled with the access of 5′-3′ exoribonuclease 2 (XRN2) to the transcript, which then chases down and promotes the eviction of Pol II. The allosteric model on the other hand, proposes the release of Pol II from chromatin following poly (A) signal dependent conformational change of Pol II (Porrua and Libri 2015). Transcription is terminated following the release of Pol II from the chromatin and free Pol II can be recycled back to the promoter for a new round of transcription. It has been reported that a subset of GTFs remain associated to the promoter, acting as a platform for the assembly of a new PIC (Orphanides and Reinberg 2016; Sandaltzopoulos and Becker 1998).

2.Transcription complexes and subunit sharing

2.1 Subunit sharing

A very interesting aspect of multisubunit complexes is that a protein subunit is often shared between different complexes, each of which carries out distinct function within the cell. For example, several Rpb subunits (Rpb5, 6, 8, 10, 12) are shared between the three RNA Polymerases (Yudkovsky et al. 2000). Apart from functionally related RNA Polymerases, several transcription complexes carrying out distinct functions in the process also share subunits amongst themselves. For instance, NuA4 histone acetyltransferase complex shares Eaf3 subunit with Rpd3 histone deacetylase complex, and Act1 and Arp4 subunits with chromatin remodelling complexes SWR1 and IN080 (Van Attikum and Gasser 2005; Smith and Shilatifard 2010). TBP is shared between SL1, TFIID and TFIIIB complexes (Burley 1996). Moreover, TFIID shares several of its TAF subunits, TAF9, TAF10 and TAF12 with the coactivator SAGA (Spt–Ada–Gcn5 acetyltransferase) complex. TAF5 and TAF6 are also present in the yeast SAGA complex but the human SAGA consists of the paralogues TAF5L and TAF6L (Spedale et al. 2012; Helmlinger and Tora 2017). SAGA in turn shares three of its subunits GCN5, ADA3, SGF29 with another HAT complex called ATAC (Ada2a-containing complex) and its ENY2 subunit with TREX-2 (transcription and export complex 2) complex (Helmlinger and Tora 2017). Apart from the transcription-related complexes mentioned above, lysosome associated complexes BORC (BLOC-one-

related complex) and BLOC-1 (biogenesis of lysosome-related organelles complex 1) also share three subunits between them (Pu et al. 2015; Langemeyer and Ungermann 2015). This indicates that subunit sharing is quite common among large multisubunit complexes. Understanding the mechanism underlying the allocation of common resources to different complexes would be an interesting direction of inquiry. Due to the sharing, it is difficult to study the role of a common subunit separately in different complexes. Nakabayashi et al. 2014 developed a technique to study the function of a common subunit by genetically fusing the common subunit to a subunit that is specific to a complex and then point mutating it. In that way, the resulting phenotype would be specifically due to the function of the common subunit in that complex only.

Transcription complexes relevant to our study, TFIID, SAGA and TREX-2 will be discussed below in **section 2.2, 2.3 and 2.4**.

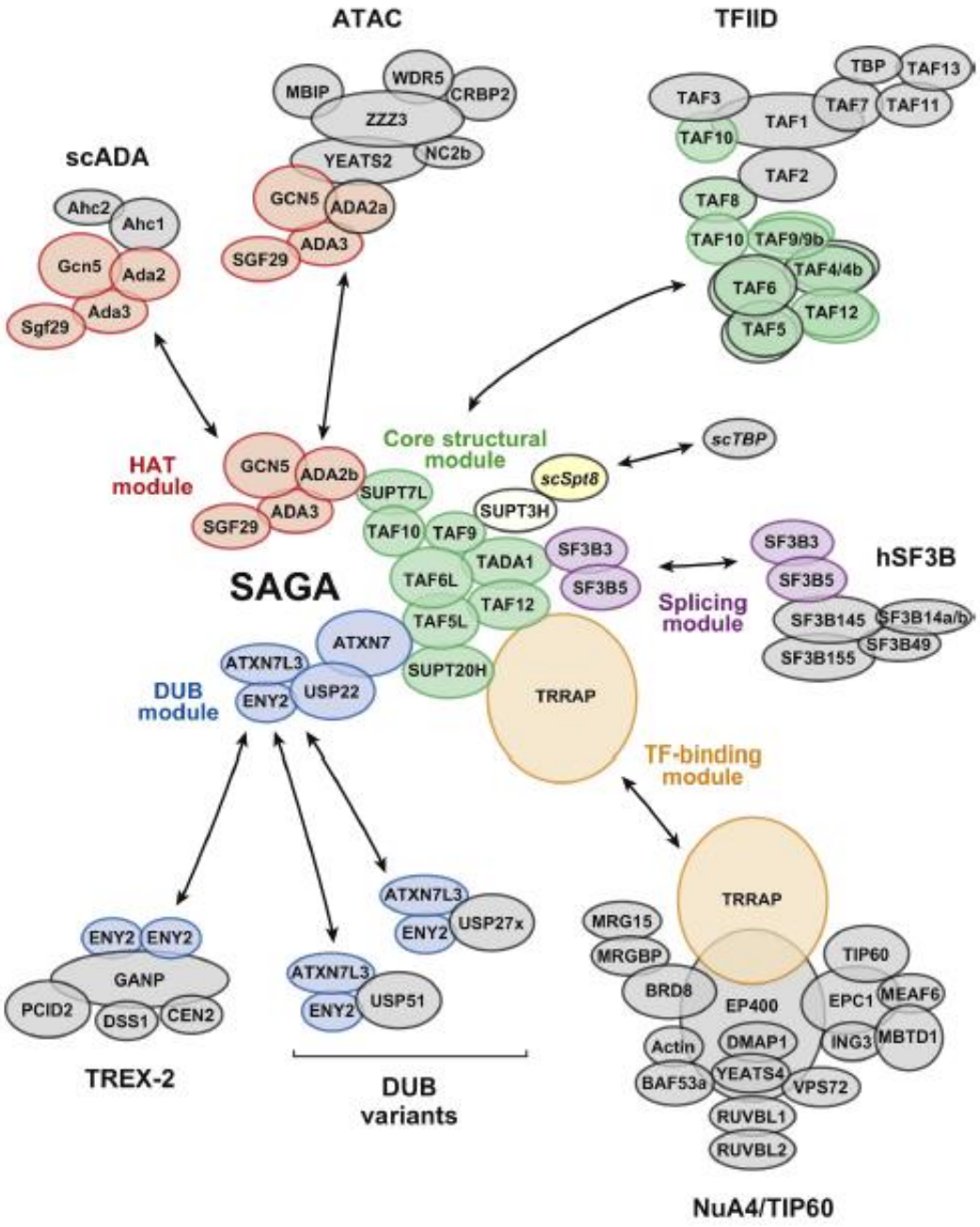


Figure 7: Subunit sharing between different transcription complexes. Subunit composition of different transcription regulatory complexes and the subunits shared among them are shown. Image is from (Helmlinger and Tora 2017).

2.2 General transcription factor TFIID

2.2.1 TFIID structure

TFIID is the leading GTF that recognises the core promoter, thereby acting as a scaffold for the assembly of other general transcription factors to form the pre-initiation complex (PIC). Structurally, TFIID has a trilobed structure (Patel et al. 2018) consisting of TATA box binding protein (TBP) and 13 evolutionarily conserved TBP- associated factors or TAFs (TAF1 to TAF13). Six out of the 13 TAFs are present in two copies in TFIID structure. (***Fig 8***) The TAFs were identified across many different species including yeast, Drosophila, C. elegans and human and were named according to their predicted molecular weight in the species identified. A unified nomenclature was later proposed for TAFs and they were named in increasing order of their molecular weights from TAF13 to TAF1 respectively (Tora 2002) TAF4, TAF5, TAF6, TAF9, TAF10, TAF12 are present in two copies which impart a two-fold symmetry to the TFIID structure (Sanders et al. 2002; Leurent et al. 2002; Hoffmann and Roeder 2002).

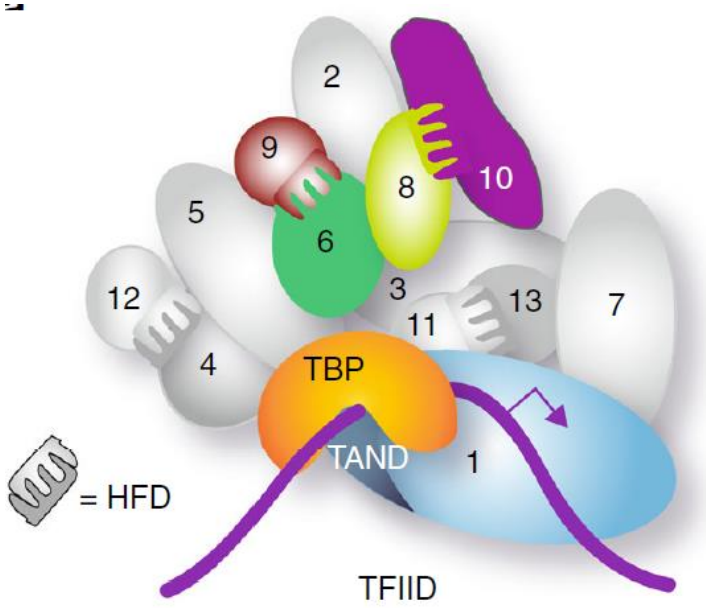


Figure 8: Schematic representation of general transcription factor TFIID. TFIID consists of TBP and 13 TAFs (TAF1-13) as shown here. 8 out of 13 TAFs dimerise with their partners via a specific domain called the histone fold domain as shown above.

2.2.2 Domain organization of TAFs

Majority of the TAFs dimerise via a specific domain called the histone fold domain (HFD). HFD is the key TAF-TAF interaction within TFIID (Gangloff et al. 2001) and is structurally conserved across different TAFs. The HFD is composed of three α helices connected by two loops, which allow heterodimeric interactions between specific TAFs. Structures of TAF HFD heterodimers TAF6/TAF9 from *Drosophila*, and TAF11/TAF13, TAF4/TAF12, TAF8/TAF10 from human were determined and reported to be structurally similar to histone heterodimers H3-H4 and H2A-H2B (Xie et al. 1996; Wertent et al. 2002; Trowitzsch et al. 2015; Birck et al. 1998). Moreover, the histone fold containing proteins are not soluble when they are expressed individually (Fribourg et al. 2001). Apart from histone fold domains, TAFs also consist of other domains as shown in the schematic below ***(Figure 9)***. The N-terminal TAND domain of TAF1 is essential for its interaction with TBP (Anandapadamanaban et al. 2013; Mal et al. 2004; Liu et al. 1998). The C-terminal double bromo domain of TAF1 is essential for interaction with acetylated histones thereby facilitating the binding of TFIID to nucleosome bound DNA (Jacobson et al. 2000; Bhattacharya et al. 2014). TAF2 is reported to interact with the initiator sequence of the core promoter (Kaufmann et al. 1996; Verrijzer et al. 1995). Two arginine and lysine rich loops of the aminopeptidase domain of TAF2 are essential for its association with promoter DNA (Kolesnikova et al. 2018). It was reported earlier that TAF3 binds to TAF10 through its HFD and facilitates the binding of NLS lacking TAF10 to importinβ *in vitro* and thus transport to the nucleus (Soutoglou et al. 2005). But structural information of TAF10/TAF3 HFD is still lacking. The plant homeodomain (PHD) finger of TAF3 binds to the H3K4me3 mark of active promoters (van Ingen et al. 2008). TAF4 and TAF12 interact with each other through their HFD. Very recently it has been shown that TAF4-TAF12 heterodimer regulates the transactivation of MYB and plays a role in leukemogenesis. Disrupting this interaction by ectopically overexpressing TAF4 histone-fold fragment and perturbing TAF4-TAF12 heterodimer formation prevents acute myeloid leukemia in mouse models (Xu et al. 2018). Apart from HFD, TAF4 homology domain (TAFH) of TAF4 has a role in gene activation through activator binding (Wright et al. 2006). Recent crystallization experiments have shown the importance of TAF5 domains. TAF6 and TAF9 interact with each other through their histone fold domains and are sandwiched between the NTD and WD-40 repeat domains of TAF5 (Antonova et al.

2018). In addition to HFD, TAF6 also contains five conserved C-terminal HEAT repeat domains which are important for TAF6-TAF9 binding (Scheer et al. 2012). TAF8 and TAF10 interact with each other through their HFD and this interaction is enhanced by the presence of proline rich domain of TAF8 (Demény et al. 2007).

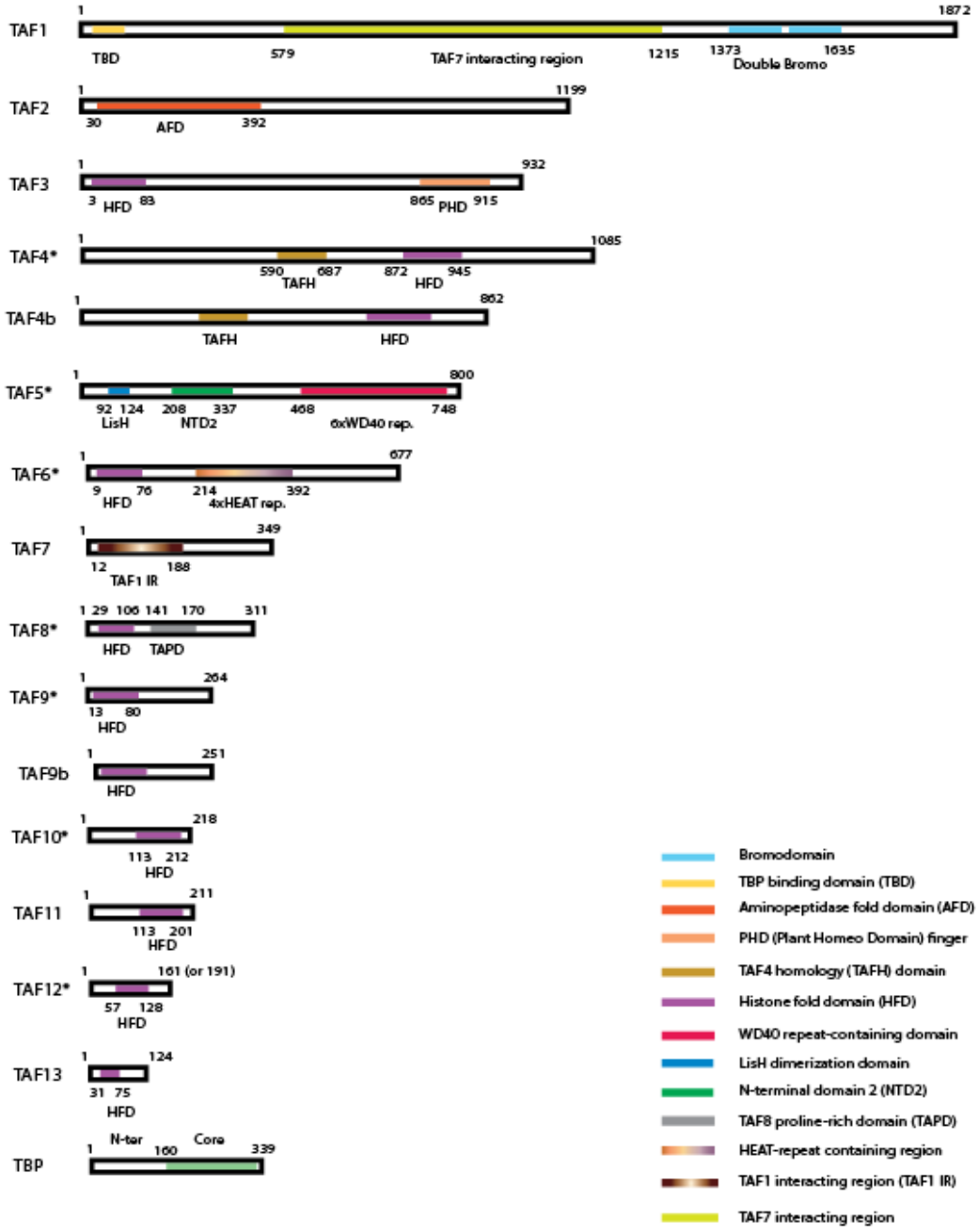


Figure 9: Domain organization of TAFs. Different domains present in all TAFs and their amino acid position. Image modified from (Müller and Tora 2004).

2.2.3 TAF paralogues and their functions

Several TFIID paralogues were reported in different organisms which carried out diverse functions and also gave rise to different TFIID variants (Müller and Tora 2004). For example, various TBP paralogues have been identified over the years. Insect-specific TRF1 (also known as TRF), metazoan-specific TRF2 (also known as TBPL1, TLP, TRP and TLF) and vertebrate-specific TBP2 (also known as TBPL2 and TRF3) were discovered and suggested to function differently in different organisms. TRF1 mainly functions in RNA Pol III transcription, although it has been shown to regulate Pol II transcription *in vitro* (Hansen et al. 1997; Verma et al. 2013; Takada et al. 2000; Holmes and Tjian 2000). TRF2 has been reported to function mainly in embryogenesis in most metazoans but only in spermatogenesis in mammals (Teichmann et al. 1999; Ohbayashi et al. 1999; Martianov et al. 2002). Previous work from our lab suggests the role of TBP2 in vertebrate ovary and oogenesis (Gazdag et al. 2007). Apart from TBP, some of the TAFs also have paralogues involved in various functions. A list of paralogues and their similarity to TAFs is shown in ***Table 4***. TAF1L and TAF7L are paralogues of TAF1 and TAF7 respectively that are involved in male spermatogenesis (Pointud 2003; Wang 2002). TAF5L and TAF6L are paralogues which are part of the SAGA complex coactivator complex (described further in section 2.3). TAF4b was initially isolated from B lymphocytes (Dikstein et al. 1996) and was later reported to be present in mouse testis and ovary (Freiman et al 2001; Falender et al. 2005). TAF9b, a paralogue of TAF9 is a part of both TFIID and SAGA complex (Frontini et al. 2005). TAF9b has been reported to be essential in controlling neuronal differentiation where it is mainly a part of the SAGA complex (Herrera et al. 2014).

Table 3: TAF paralogues and their sequence similarity

Paralogue pair	**% Identity (Peptide sequence)**
TAF4 and TAF4b	59%
TAF9 and TAF9b	81%
TAF7 and TAF7L	56%

TAF1 and TAF1L	93%
TAF6 and TAF6L	26%

2.2.4 TFIID assembly

Few reports suggest that TFIID assembly is a stepwise process and not a random assembly of all subunits before it is transported to the nucleus (Gupta et al 2017; Antonova et al. 2018; Trowitzsch et al. 2015). One of the very first reports of TFIID submodule was of a small TAF complex called SMAT consisting of TAF10, TAF8 and SPT7L, though the function of the complex is still unclear. TAF10 is a subunit of both general transcription factor TFIID and co-activator SAGA complex. The authors speculated that the SMAT complex is essential for maintaining TAF10 distribution equilibrium between TFIID and SAGA complex (Demény et al. 2007). Compositional variation in TFIID complexes, sometimes consisting of a subset of TAFs have also been reported and they also carry out unique roles in transcription regulation (Maston et al. 2012; Müller and Tora 2004). Five (TAF4, TAF5, TAF6, TAF9, TAF12) out of the fourteen TFIID subunits form the symmetric core-TFIID complex in two copies each (***Figure 10***). This core-TFIID was first identified in Drosophila nuclei (Wright et al. 2006) and later its architecture was determined by cryo-EM (Bieniossek et al. 2013). A heterotrimeric TAF2-TAF8-TAF10 complex exists in the cytoplasm which binds to the symmetric core complex to form the 8TAF complex. This binding of the heterotrimer to the core-TFIID breaks its symmetry and facilitates the binding of the remaining TAFs (Trowitzsch et al. 2015). A similar TAF11-TAF13-TBP trimer could also be assembled in vitro but could not be detected in the cytoplasm. It has been hypothesized that TAF11-TAF13 HFD dimer plays role in the dynamics of TBP association within TFIID (Gupta et al. 2017). Additionally, a second heterotrimer, TAF5-TAF6-TAF9 was also observed in the cytoplasm of cells very recently (Antonova et al. 2018). However, its role with respect to stepwise TFIID assembly has not been studied so far.

Another interesting aspect about the involvement of chaperones in TFIID assembly was shown very recently (Antonova et al. 2018). The authors argue that CCT (Chaperonin Containing TCP1 or TriC-TCP-1 Ring Complex) complex is involved in

the folding of WD40-repeat domain containing TAF5 and its subsequent handover to TAF6-TAF9 heterodimer, thereby facilitating orderly assembly of TFIID. Chaperones are described more in detail in ***Section 6***.

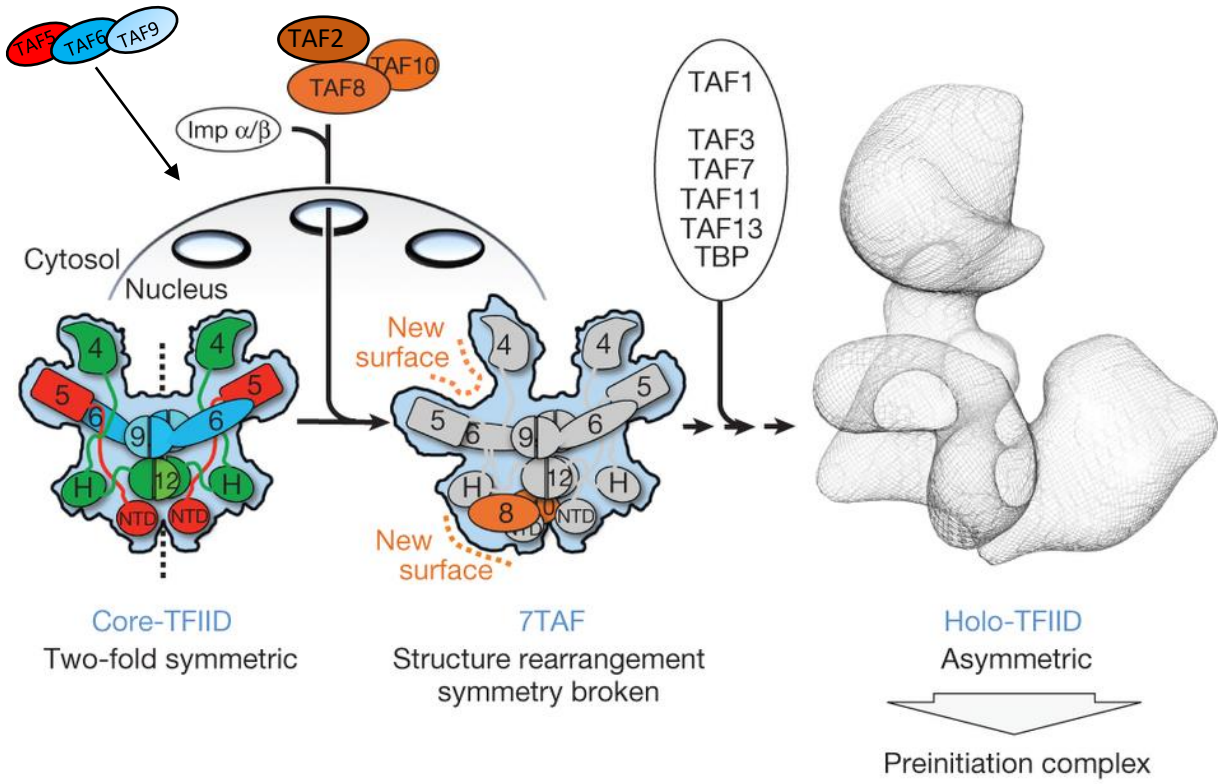


Figure 10: Stepwise assembly of TFIID complex. Schematic representation of the assembly of TFIID complex. Heterotrimeric submodule of TAF2-TAF8-TAF10 exists in the cytoplasm which gets imported to the nucleus by importin α/β and binds to the core TFIID complex, causes structural transitions and enables the binding of remaining TAFs. TAF5-TAF6-TAF9 also exists in the cytoplasm, however it is not clear yet if it exists as a heterotrimer of one copy each of TAF5-TAF6-TAF9 or in two copies and how it is imported to the nucleus. Image modified from (Trowitzsch et al. 2015).

2.3 Coactivator SAGA complex

The Spt–Ada–Gcn5 acetyltransferase (SAGA) complex is an evolutionarily conserved multisubunit coactivator complex in eukaryotes (Spedale et al. 2012). It is composed of 18-22 subunits organised in several functional modules namely activator-binding module, histone acetyltransferase (HAT) module, histone deubiquitinase (DUB) module, core structural module and a metazoan specific splicing module (Helmlinger and Tora 2017; Koutelou et al. 2010) ***(Figure 11).*** SAGA plays multiple roles in transcription regulation, as indicated by the module names. Activators are suggested to recruit SAGA to the promoter through its largest subunit, TRRAP (Tra1 in yeast). TRRAP interacts with different transcription factors and may serve as the target of major promoter bound activators thereby enabling the recruitment of SAGA to promoters (Weake and Workman 2012; Bhaumik et al. 2004; McMahon et al. 1998; Helmlinger et al. 2011). Other alternate mechanisms for SAGA recruitment include interactions with chromatin marks, TAF12 interactions with activator or interactions with TBP in yeast (Weake and Workman 2012). Following recruitment the histone acetyltransferase module of SAGA induces acetylation of lysine (K) 9 of histone H3 (H3K9ac) and K14 of H3 (H3K14ac). This activity is catalyzed by the GCN5 enzyme (Grant et al. 1997). The deubiquitination module (DUBm) of SAGA removes ubiquitin modification from both H2Bub and H2Aub in gene bodies (Zhang et al. 2008; Zhao et al. 2008) as well as nonhistone substrates (Weake and Workman 2012). It has been shown recently that SAGA is required for the transcription of all active genes in both yeast and human cells (Baptista et al. 2017; Bonnet et al. 2014). As shown in ***Figure 7*** SAGA shares many of its subunits with different regulatory complexes thereby adding more complexity to its assembly mechanism.

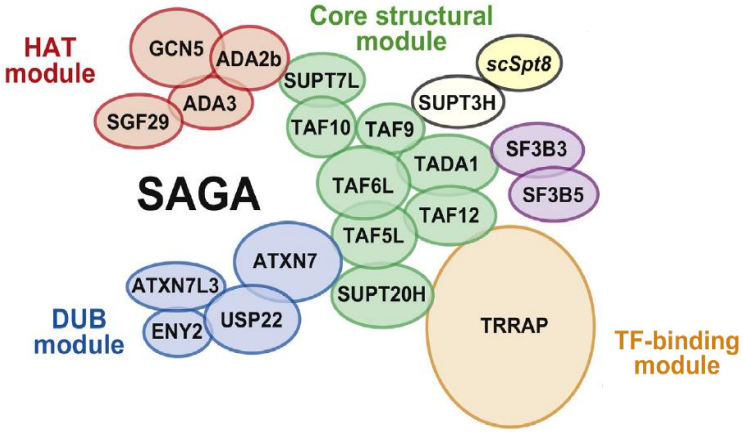


Figure 11: Schematic representation of SAGA complex. SAGA complex is made up of ~20 subunits, arranged in several modules each carrying out different functions. Image adapted from (Helmlinger and Tora 2017).

2.4 TRanscription and EXport complex 2 (TREX-2)

The presence of different subcellular compartments is one of the main features that distinguishes eukaryotes from prokaryotes. So, eukaryotic cells have to ensure appropriate targeting of various functional macromolecules into target organelles for efficient functioning of cells. While mRNAs are transcribed and processed in the nucleus, protein translation by ribosomes takes place in the cytoplasm. So, mRNAs must cross the nuclear membrane barrier and get transported to the cytoplasm by active and selective mechanisms. Nevertheless, this functional orchestration between nuclear export of mRNAs and different steps of gene expression is essential for the maintenance and fidelity of gene expression. Nuclear pore complexes (NPCs) act as passages for the trafficking of proteins and RNAs between nucleus and cytoplasm. The evolutionarily conserved TREX-2 multiprotein complex binds to the NPC basket structure (Fischer et al. 2004) and facilitates a variety of functions including transcription and mRNA export (Umlauf et al. 2013; Gallardo et al. 2003), as well as for genomic integrity (Evangelista et al. 2018; Kotani et al. 2009a). It has been reported that TREX-2 interacts directly with Mediator and it is the Mediator which established a connection between TREX-2 and Pol II (Schneider et al. 2015). Human TREX-2 has also been shown to stably associate with NPC in contrast to earlier reports which

pointed towards a rather transient interaction (Umlauf et al. 2013). Human TREX-2 is composed of five subunits, GANP, ENY2, PCID2, Centrin 2/3, and DSS1 (which in yeast are Sac3, Sus1, Thp1, Cdc31, and Sem1, respectively) as shown in ***Figure 12***. TREX-2 shares its ENY2 subunit (Sus1 in yeast) with SAGA complex (as described above). In TREX-2, ySus1/hENY2 binds to the scaffold protein ySac3/hGANP in two copies (Jani et al., 2012) and in SAGA, hENY2 binds to ATXN7L3. hENY2, together with ATXN7 and USP22 forms the deubiquitination or DUB module (Zhang et al. 2008; Lang et al. 2011).

Therefore, it would be interesting to understand how the cell distributes the common subunits to different complexes and the factors guiding it.

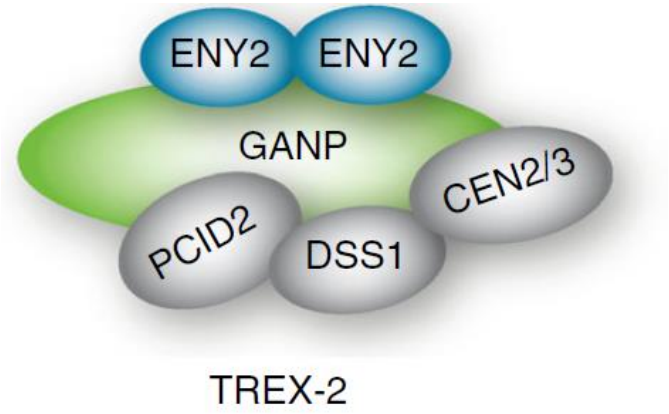


Figure 12: Schematic representation of human TREX-2 complex. Two ENY2 subunits, one subunit each of PCID2, DSS1 and CEN2/3 wrap around the central scaffold of the huge GANP subunit to form the TREX-2 complex.

3. Protein complexes

The structure of a protein is defined by four levels of complexity: primary, secondary, tertiary and quaternary. The primary structure of a protein is the sequence of amino acids synthesised by the ribosomes. The interaction between adjacent amino acids leads to a regular arrangement of the polypeptide chain into secondary structures like alpha helices, beta sheets, turns or loops. The next level of protein structure, the tertiary structure is a three-dimensional conformation of a polypeptide caused by numerous interactions between amino acid side chains and secondary structure elements. Monomeric proteins only have a single folded polypeptide chain. Quaternary structure exists for oligomeric proteins which are made up of multiple folded subunits. Majority of proteins in a cell assemble into complexes to carry out their function. It is therefore important as well as necessary to understand the physicochemical properties of protein interactions. Homomeric proteins are formed by the assembly of identical protein subunits, whereas heteromeric proteins are formed by the association of multiple distinct protein subunits.

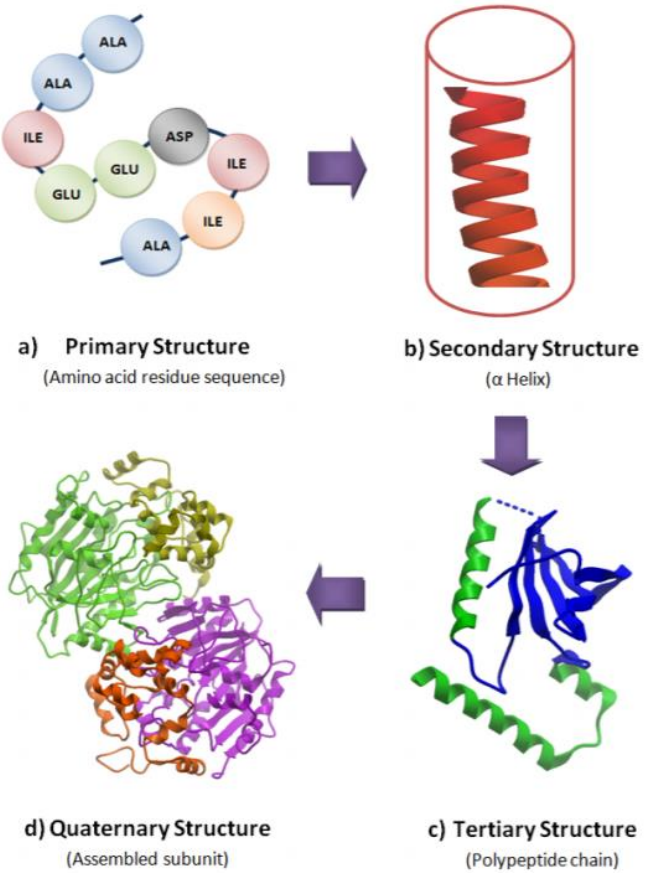


Figure 13: Different levels of protein organization. The peptide sequence of a protein undergoes various levels of organisation to finally form the globular or tertiary structure which in turn binds to its partner proteins to form the quaternary structure. Image adapted from (Aziz 2011).

3.1 Homomeric complexes

Many homomeric protein complex structures have been solved by X-ray crystallography. In fact, 87% of crystal structures available are of monomers and among them, 54% form homomers (Marsh and Teichmann 2014b). Despite the bias towards the number of available crystal structures for homomers, the fact that the formation of homomeric protein complexes is common in the cells cannot be ruled out. Homomeric complexes carry out various important functions in the cell. For example, the homo tetramer phosphoglycerate mutase plays important role in glycolysis (Winn et al. 1981). More examples are described next.

Depending on the number of subunits, homomers can be classified into several categories, namely homodimer, homotrimer, homotetramer and so on. Homodimerization is an important event in cellular signalling pathways. One of the classic examples in this case is the homodimerization of epidermal growth factor receptor (EGFR) family protein ErBb2. The extracellular domains (ECDs) of ERBb2 contribute to the direct binding of ERBb2 monomers. This dimerization activates the phosphorylation of the intracellular domains (ICDs) of ERBb2 which further leads to downstream signal transduction (Hu et al. 2015; Penuel et al. 2002; Chantry 1995). Nuclear hormone receptors like retinoid, thyroid, vitamin D receptors form homodimers under specific conditions, even though they mainly act as heterodimers (Mangelsdorf and Evans 1995). Depending on the type of ligand interacting, thyroid hormone receptor (TR) could exist in either homodimeric or heterodimeric forms (Lehmann et al. 1993). Individual proteins of a huge protein complex could also form homodimers amongst themselves before combining with the holo-protein. For example, Gamma-secretase complex is comprised of four major components: PS1, nicastrin, Aph-1, and Pen-2. Among them, PS1, which is the catalytic component of the complex, forms homodimer during normal functioning in the cell (Herl et al. 2006). Homotrimers also perform important functions in the cell. Proliferating Cell Nuclear Antigen (PCNA) exists as a double homotrimer complex in the cell, which enables it to bind to both

DNA Polymerase and Chromatin Assembly Factor-1 (CAF-1) simultaneously thereby allowing DNA Replication coupled to chromatin remodelling. The shape attributed by the trimeric ring enables PCNA to act as both sliding clamp and docking station for a number of proteins (Naryzhny et al. 2005). One of the primary gatekeepers of inhibitory neurotransmission of the central nervous system, GABAA receptors form a pentameric arrangement of five identical subunits (Claxton and Gouaux 2018). Thus, the functional role of homomers in the cell is mainly dependent on the shape attributed by the arrangements of the subunits in the complex. Symmetry is a common feature in the structure of most homomeric complexes. All homomers can be classified into different groups depending on their structure: Twofold dimeric complexes, Cyclic complexes, Dihedral complexes, Cubic complexes, Helical complexes and Asymmetric homomers (Bergendahl and Marsh 2017; Marsh and Teichmann 2014b). Each of these homomer groups is involved in specific functional aspects of a cell (Bergendahl and Marsh 2017). ***(Figure 14)*** For example, two-fold dimeric complexes are mainly associated with “biosynthetic processes” in the cell and also “DNA-templated transcription”. Cyclic homomers are mainly associated with functions related to the cell membrane since it is important to form a well-defined channel structure through the two-dimensional cell membrane. Most of the dihedral complexes are associated with metabolic processes. Dihedral symmetry could be an easy structure to form with four or more subunits, thereby bringing multiple enzymes together in a single complex to provide a higher number of active sites for catalysis. Cubic homomers form large hollow shells which are convenient for storage, thereby making it the ideal candidate for proteins involved in homeostasis and metal ion binding. Helical complexes on the other hand are well-suited for proteins involved in the formation of long fibres like microtubules and actin filaments. Asymmetric complexes although rare (Swapna et al. 2012) are mainly involved in signal transduction processes (Lee and Dominguez 2010; Birck et al. 2003). The fact that monomers are also essential for the cell cannot be ruled out. Most of the biological macromolecule modifying proteins in the cell are monomers. This is in accordance with the idea that monomers are more suitable to accommodate large macromolecular substrates.

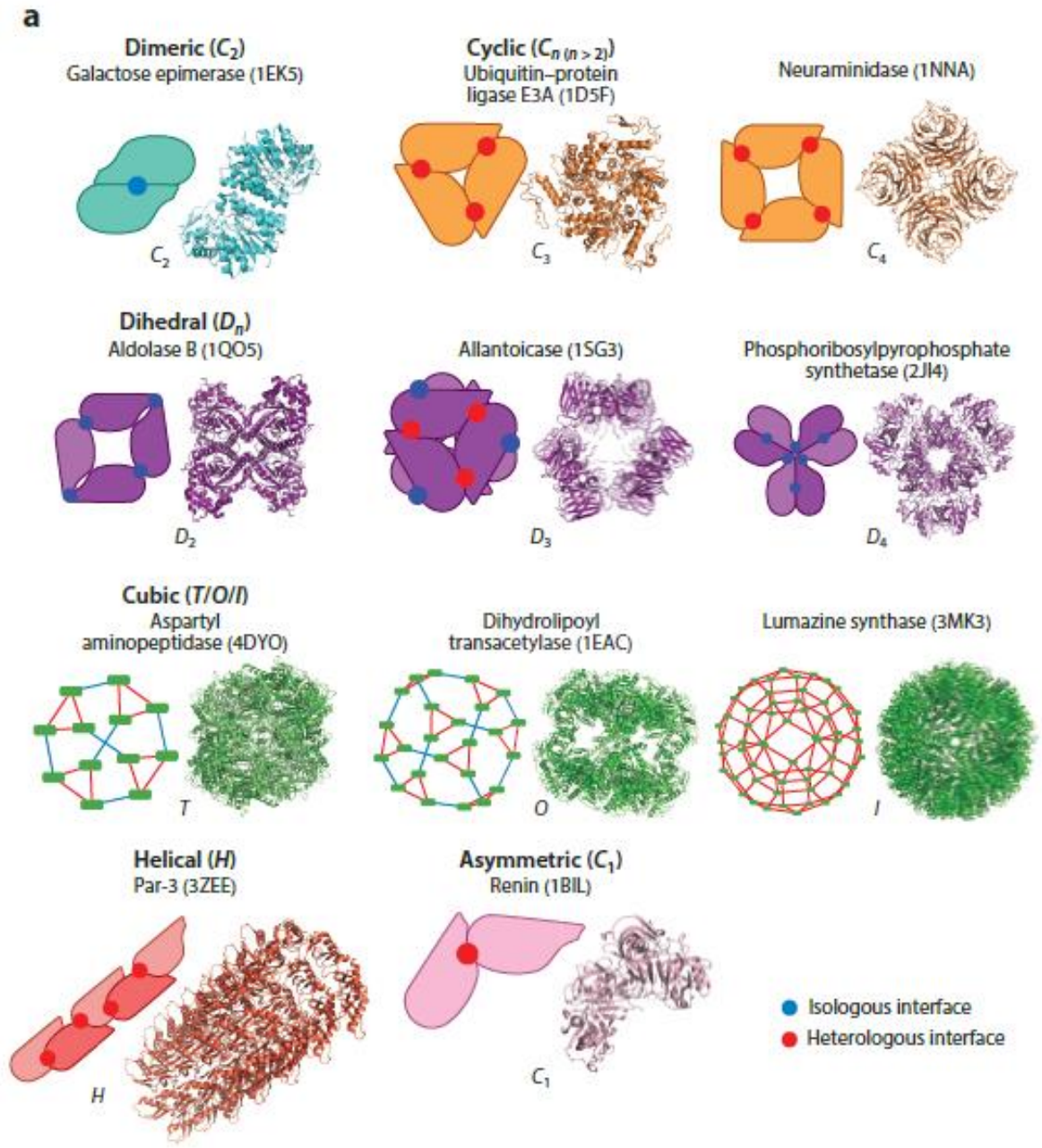


Figure 14: Classification of homomeric protein complexes based on shape. Homomeric protein complexes can be grouped into several categories depending on the shape attributed by the arrangement of their subunits. Image is from (Marsh and Teichmann 2014b).

Homomer formation attributes unique properties to the proteins which monomers do not have and is essential for diverse functions of a cell (Jones and Thornton 1995). However, aberrant homomer formation leads to diseased condition in the cell. For example, collagen I, the major extracellular fibrillar collagen exists in both heteromeric

and homomeric forms in the tissues. A naturally occurring missense mutation shifts the equilibria to homomer formation and this increased homomer formation of collagen I leads to various diseased conditions (Sharma et al. 2017).

3.1.1 TBP-associated factor (TAF) homodimerization

In the context of this thesis, it is important to note that the multisubunit heteromeric general transcription factor TFIID also consists of homodimers of specific subunits (TAF4, TAF12, TAF5, TAF6, TAF9). These five subunits in two copies form a two-fold symmetrical functional scaffold called the core TFIID complex (***Figure 15***) which was first revealed in Drosophila nuclei and is important for the assembly and integrity of holo-TFIID complex (Jones and Thornton 1995). In contrast, other TAFs bind to the core-TFIID as single copies and break the symmetry of TFIID-core complex (Bieniossek et al. 2013; Cler et al. 2009).

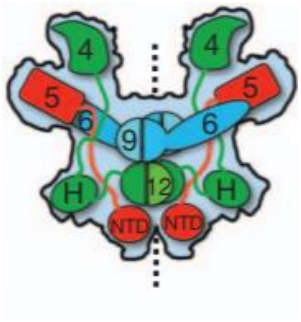


Figure 15: Schematic representation of core-TFIID complex. Two copies of each of TAF4, TAF5, TAF6, TAF9, TAF12 subunit form a symmetrical scaffold called the core-TFIID to which other subunits of TFIID bind to form the holo-TFIID complex. Image adapted from (Bieniossek et al. 2013).

3.2 Heteromeric complexes

Despite the presence of a vast majority of heteromeric protein complexes in cells, fewer structures for heteromers have been solved than homomeric protein complexes (Perica et al. 2012; Marsh and Teichmann 2014a). Based on the structure attributed by the arrangement of subunits in heteromeric protein complexes, they can be classified into the following types (Levy et al. 2006; Marsh and Teichmann 2014b): 1) Paralogous heteromers, 2) Symmetric heteromers, 3) Mixed-symmetry heteromers

and 4) Asymmetric heteromers. (***Figure 16***) Paralogous heteromers are formed by subunits which are paralogues of each other. They may or may not form a symmetrical structure. This group includes the human Rad9-Hus1-Rad1 heterotrimer which forms a cyclic trimer, the archaeal chaperonin thermosome consisting of only two paralogous subunits but forming a symmetric structure, etc. (***Figure 16a***) Non-paralogous subunits can form symmetric structures as well and these complexes form the symmetric heteromer family. Tryptophan synthase and formate dehydrogenase have two and three different subunits respectively, arranged in a symmetrical structure. Some complexes form a symmetric structure with both paralogous and non-paralogous subunits, like the complex formed between the proteasome and the proteasome-activating nucleotidase (PAN) assembly. ***(Figure 16b)*** Some heteromers possess different types of symmetry together in the same structure which are made up of different numbers of each type of subunit. ***(Figure 16c)*** Another type includes asymmetric heteromers that have no symmetry at all such as transcription complexes like TFIID, SAGA, RNA Polymerase II and so on. Although the past decades have witnessed crystal structure of fewer heteromeric complexes than homomers, advent of newer techniques like cryo-EM opens new avenues for solving heteromeric protein complexes structures (Marsh and Teichmann 2014b).

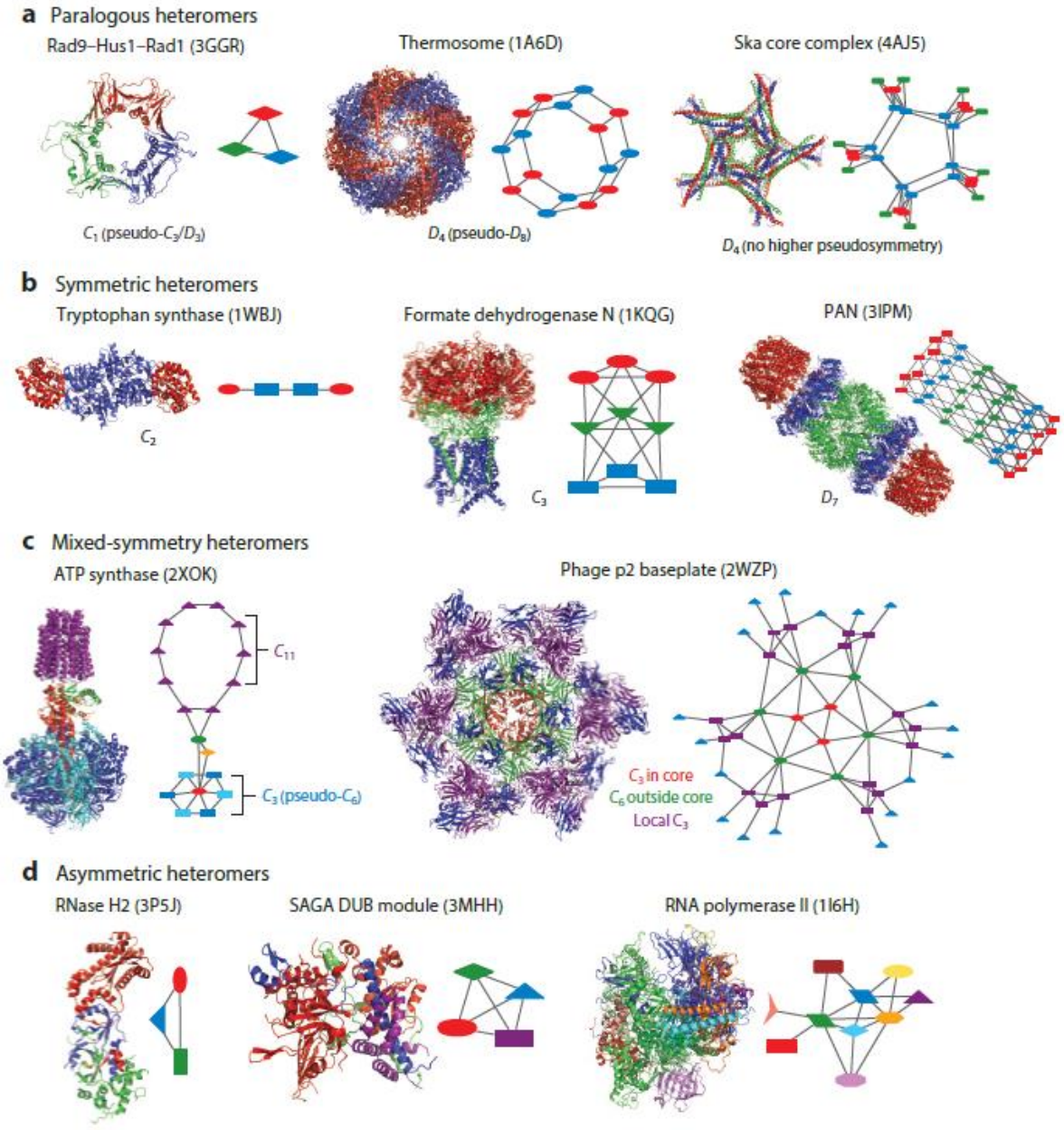


Figure 16: Classification of heteromeric protein complexes based on shape. Heteromeric protein complexes can be grouped into several categories depending on the shape attributed by the arrangement of their subunits. Image is from (Marsh and Teichmann 2014b).

3.3 Other types of protein complex classification

3.3.1 Obligate and non-obligate protein complexes

A protein that can fold and exist independently *in vivo* is called non-obligate protein. These proteins are also functionally independent. This group includes intracellular signalling complexes (like RhoA-RhoGAP), antibody-antigen, receptor-ligand and enzyme-inhibitor complexes. The components of these complexes are independently stable and are often not co-localising in the cell. On the other hand, proteins that cannot fold and form stable structures independently and exists as part of a complex (where each subunit is stabilised by its partner) are obligate proteins. Such complexes are also functionally obligate. For example, the Arc repressor dimer acts as obligate protein complex and is essential for DNA binding (Nooren and Thornton 2003).

3.3.2 Transient and stable protein complexes

Based on the lifetime of interactions between proteins, they can form transient or stable protein complexes. Stable complexes are formed by irreversible interactions between protein subunits. Transient protein complexes are reversible interactions that are formed and broken continuously. An example of both type of interactions is the heteromeric G protein. It dissociates into Gα and Gβγ subunits upon guanosine triphosphate (GTP) binding but forms a stable trimer with guanosine diphosphate GDP. Generally, obligate interactions (described in section 3.3.1) are stable and non-obligate interactions may be transient or stable structurally and functionally (Nooren and Thornton 2003; Wall et al. 1995).

3.4 Assembly of protein complexes

Protein assembly pathways are mainly studied by mass spectrometry approaches which enables the detection of assembly intermediates of protein complexes (Hall et al. 2013; Hernández and Robinson 2007; Ruotolo et al. 2008). With this technique, nearly all dihedral and tetrahedral homomeric protein complexes were shown to have cyclic intermediates. No further subcomplex of cyclic intermediates were observed suggesting that most homomers assemble via a specific assembly pathway (Levy et al. 2008). Heteromeric protein complexes like membrane bound protein complexes and ribonucleoprotein complexes (ribosome and spliceosome) also assemble via intermediates. These assembly intermediates are built independently and then

assembled into a huge complex (Ackerman and Tzagoloff 2005; Pandit and Skolnick 2008; Henras et al. 2008). Additionally, assembly intermediates of RNA Polymerases have also been described along with several factors mediating assembly (Van Nguyen et al. 1996; Forget et al. 2010; Boulon et al. 2010). Based on these observations, a model has been put forward which shows the formation of RNA Polymerase II from at least two major assembly intermediates and the intermediates are stabilised by assembly factors (Wild and Cramer 2012). Although the already mentioned studies suggest assembly of complexes in stepwise fashion, there has been no concrete evidence on the order of assembly of subunits so far. Nevertheless, an obvious way of subunit assembly is post-translational, following the release of fully translated protein from ribosomes. The crowded cytoplasm could be a challenging environment for the cell to find its correct interacting partner. An alternative method of assembly that the cell could employ is co-translational assembly where a protein assembles with its interacting partner protein while it is still associated with the translating ribosome (***Figure 17***). A number of studies have recently shown the occurrence of this type of assembly for multisubunit complexes which are described in detail in ***section 4***.

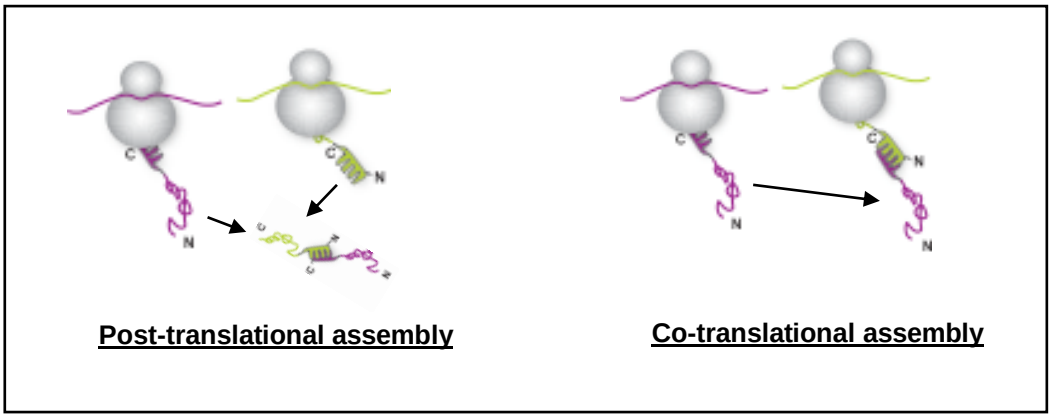


Figure 17: Models of protein complex assembly. Two models have been put forward for pairwise protein subunit assembly, post-translational assembly involves the interaction of two protein subunits following their release from the ribosomes and co-translational assembly involves the association of proteins while still associated with the ribosome.

4. Co-translational assembly of protein complexes

Proteins carrying out biological processes mostly act as macromolecular complexes. These complexes could be either homomeric like the homotetrameric GAPDH or heteromeric made up of many different subunits like the 26S proteasome (Voges, Zwickl, and Baumeister 1999) as described in ***section 3***. Regardless of the nature of subunit composition, it is very important to understand their mechanism of assembly. A simple model of assembly is post-translational assembly where fully translated protein subunits find their interacting partners by random collision in the cytoplasm. Another possibility is co-translational assembly where protein partners start interacting with each other while still attached to the translating ribosome **(*Figure 17*).** This form of assembly would prevent unimportant aggregation in the crowded environment of the cell. There have been several reports on co-translational assembly of homomeric and heteromeric protein complexes as discussed below.

4.1 Co-translational assembly of homomeric protein complexes

4.1.1 Bacterial beta-galactosidase enzyme

One of the first reports on co-translational assembly was in 1964 where the authors observed beta-galactosidase enzyme activity on the polyribosomes of the bacterium *Escherichia coli*. The authors postulated two possible models of polysome-associated assembly for β-galactosidase: 1) Each fully synthesised protein monomeric units are released from the ribosome followed by its association with the translating polypeptide chain on the polysomes. 2) Another alternate mechanism is the transfer of monomeric chains from the ribosome synthesising it to the next ribosome. So, a protein monomer is transferred to its adjacent translating monomer to form a dimer which is then transferred to its adjacent monomer forming a trimer and so on. The ribosome holds the intermediate multimers through the protein chain being synthesised (Kiho and Rich 1964).

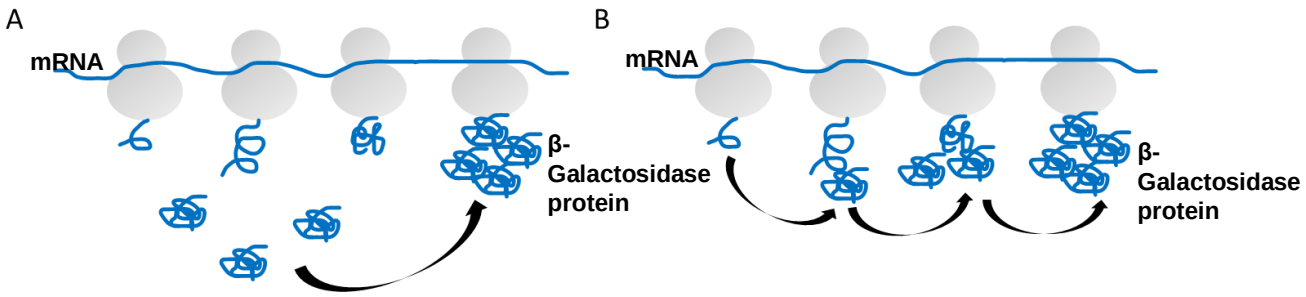


Figure 18: Models depicting β-galactosidase co-translational assembly. Model A shows the release of fully translated monomers and their co-translational association to form homotetramer, Model B shows the sequential transfer of monomers on the polysome to form the homotetramer co-translationally.

4.1.2 Myosin heavy chain

More than a decade after the first observation of bacterial co-translational assembly was made, more reports came out on the co-translational assembly of proteins in the cytoskeleton. Cytoskeleton is the backbone of a cell which gives shape to the cell. It is made up of microtubules, actin filaments and intermediate filaments. Myosin is a molecular motor protein that binds to actin and regulates muscle contraction and intracellular motility (Isaacs and Fulton 1987; Redick et al 1995). It was reported that the co-translational association of myosin heavy chain into the cytoskeleton takes place during development. By using an antibody to immunoprecipitate S^{35} radiolabelled nascent chain, the authors showed that a significant fraction of heavy chain is co-translationally associated with the cytoskeleton (Isaacs and Fulton 1987).

4.1.3 Tenascin intermediate filament

Tenascin is an extracellular matrix protein which exists as a disulphide-bonded hexamer. Electron micrograph images show tenascin as a complex made up of two trimers attached to each other by a central knob (Erickson and Inglesias 1984). The neo-synthesized full-length tenascin exists as disulphide-bonded hexamers as soon as they are synthesised. The authors could not detect multimeric intermediates in the cell lysate. Several other extracellular matrix proteins with both N- and C-terminal assembly domains are formed post-translationally unlike tenascin (Yu et al 1983; Choi

and Hynes 1979; Counts et al 1978; Vuorio et al 1990). It is however not known clearly how the translating ribosomes are arranged close to each other to form this huge organised hexameric complex (Redick et al 1995).

4.1.4 Reovirus cell attachment protein

A very interesting case of two forms of assembly, post and co-translational was reported in a single protein complex. The reovirus cell attachment protein σ1 undergoes N-terminal trimerization co-translationally and is ATP-independent but the C-terminal trimerization is post-translational and require ATP. Additionally, Hsp70 only associates with the already N-terminal trimerized but non-C terminal trimerized protein complex, which suggests that C-terminal post-translational trimerization requires both ATP and Hsp70 (Leone et al. 1996).

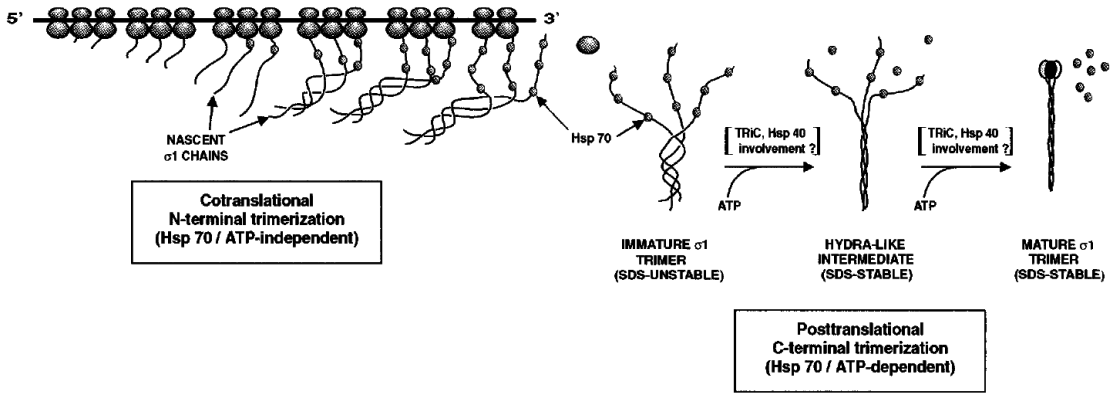


Figure 19: Model for the biogenesis of reovirus σ1 trimer. Assembly of three σ1 nascent chains occur co-translationally at the N-terminus without the involvement of Hsp70 and ATP. As the triplex moves down the polysome, it associates with more Hsp70 thereby preventing misfolding and aggregation. The C-termini finally assembles post-translationally in the presence of ATP to form the mature σ1 trimer. Image is from (Leone et al. 1996).

4.1.5 NFKB

NFKB1 gene encodes both p50 and the larger protein p105 (Kieran et al. 1990; Ghosh et al. 1990). Co-translational dimerization of the Rel homology domain of p50 coupled to proteasome capture and release is required for the efficient production of p50-p105 heterodimer. In the absence of this interaction, only full-length p105 monomer is formed. Thus, co-translational dimerization is necessary to maintain the intracellular balance of both p50 and p105 proteins and imbalance might lead to various diseased conditions in the cell (Lin 2000).

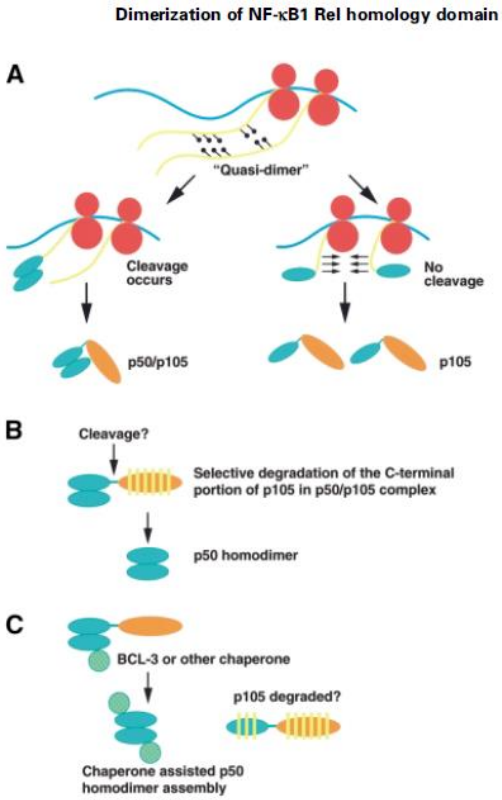


Figure 20: Models of p50 homodimer formation. Co-translational dimerization facilitates the formation of p50-p105 heterodimer (A) which is followed by the formation p50 homodimer by either selective cleavage (B) or by chaperone assisted p50 homodimer formation. Image is from (Lin 2000).

4.1.6 p53

p53 is essential for its tumor suppressor activity. It induces cell cycle arrest and apoptosis in response to genotoxic stress. TP53 gene mutations, which reduce tumor suppressor activity of p53, are observed in more than 50% of the human tumors. Apart from this, p53 also plays a role in various other functions of the cell including autophagy, stem cell self-renewal, and reprogramming of differentiated cells into stem cells, immune system, and metastasis (Kamadaa et al. 2007). p53 acts as a homotetramer, with two dimers binding to two consensus DNA half sites in the form of a clamp which gives more stability to the p53-DNA complex (Kamadaa et al. 2007) (Kevin G.McLure 1998). It has been reported that p53 forms a homodimer co-translationally on the polysome and it forms the tetramer post-translationally in solution. The authors carried out a series of mutant/ wild type co-expression and DNA-binding assays to prove the assembly mechanism (Nicholls et al. 2002).

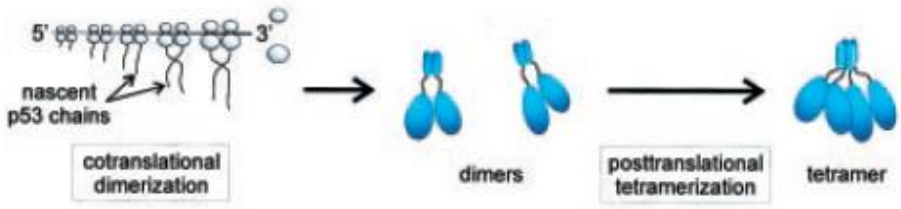


Figure 21: Model for p53 homodimer formation. p53 nascent chains dimerize co-translationally but the individual dimers form tetramers post-translationally. Image is from (Nicholls et al. 2002).

4.1.7 Peripherin

Peripherin is a neuronal intermediate filament (IF) protein expressed mainly in the peripheral nervous system. It has been reported that peripherin mRNA-ribonucleoprotein (mRNP) particles move along the microtubules. These particles are generally translationally inactive, but they become active when they stop their movement. The regulation of their motility is not known clearly and could involve inactivation of their associated motors. Once the peripherin proteins are synthesised, they are co-translationally assembled to non-filamentous IF precursor particles. Following synthesis of adequate amount of proteins, the mRNP IF precursor particles move away from the site of synthesis and assemble into IFs. The authors used a series of imaging techniques to show the dynamic cotranslation of peripherin (Chang et al. 2006).

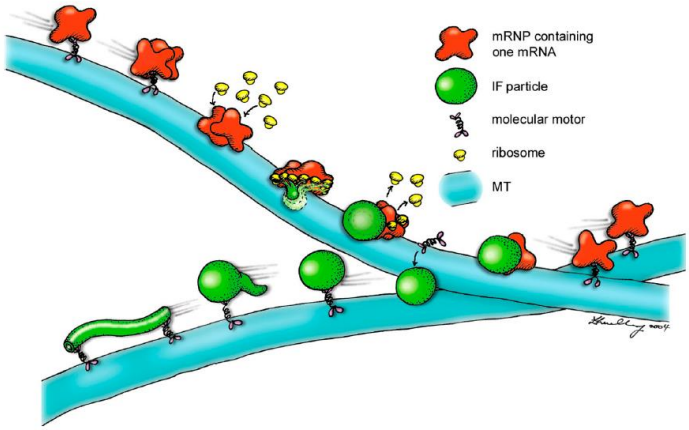


Figure 22: Model depicting dynamic co-translational assembly of intermediate filaments. mRNPs (red) containing multiple peripherin mRNAs move along the microtubule (blue) by molecular motors kinesin and dynein. The motile mRNPs are translationally silent. When these complexes stop moving, they engage in translation activity with ribosomes (in yellow). The mRNP particles contain multiple mRNAs and as the IF protein is synthesised from multiple mRNAs, the protein chains assemble co-translationally into higher order nonfilamentous particles (green). The newly synthesized IF particles could assemble into IF at their site of synthesis or recruit motors to begin their journey as IF precursors to different regions of the cytoplasm. Once they reach their targets, they assemble into short IF (squiggles) that link up in tandem to form longer filaments. Image is from (Chang et al. 2006).

Despite many reports on cotranslational assembly of homomeric proteins, the fact that the close proximity of two interacting nascent chains might lead to their premature assembly cannot be ruled out. In fact, a recent study across diverse proteomes showed the enrichment of the interaction domains of homomeric proteins towards their C-terminus, thereby enabling the folding of majority of the protein before initiating assembly (Natan et al. 2018).

4.2 Co-translational assembly of heteromeric protein complexes

4.2.1 Immunoglobulin

The very first report on co-translational intrachain disulphide bond formation in heterodimers was published in 1979 (Lawrence et al. 1979). An immunoglobulin molecule consists of two heavy and two light chains linked by both intra and interchain disulphide bonds. Each light chain has two globular domains, each of these domains are made up of 110 amino acids and an intrachain disulphide bond. Using chromatography, the authors showed the formation of intrachain disulphide bond of the first domain co-translationally while the formation of the bond for the second domain was post-translational (Lawrence et al. 1979).

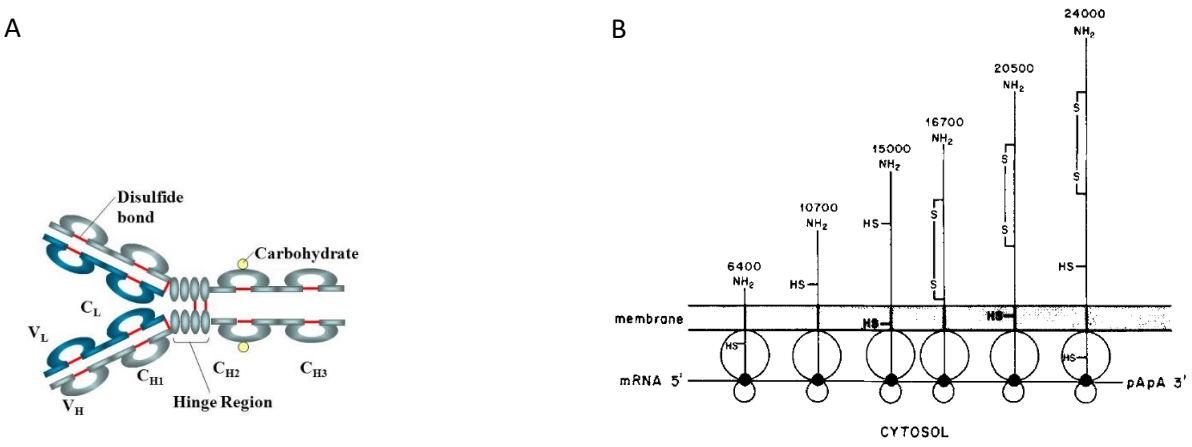


Figure 23: Model showing the structure of a typical immunoglobulin molecule (A) and co-translational formation of intrachain disulphide bond of an immunoglobulin molecule (B). An immunoglobulin molecule consists of a heavy and light chain connected to each other by interchain disulphide bond. Each heavy and light chain consists of four and two domains respectively and each domain contains an intrachain disulphide bond (A). The intrachain disulphide bond in the variable region of light chain is formed co-translationally (B). V_L and C_L are variable and constant region of light chain respectively; V_H and C_H are variable and constant region of heavy chain respectively. Image adapted from (Lawrence et al. 1979).

4.2.2 Signal recognition particle receptor

One of the earliest reports of co-translation assembly of proteins into membranes was that of signal recognition particle receptor (SRα). The authors discovered a translation pause site at nucleotide 507 of the mRNA encoding SRα which induces a brief ribosomal pause that enables the protein to insert itself into the endoplasmic reticulum (ER) membrane. Change in this nucleotide sequence without changing the protein sequence disrupts the co-translational association of SRα into ER membrane (Young and Andrews 1996).

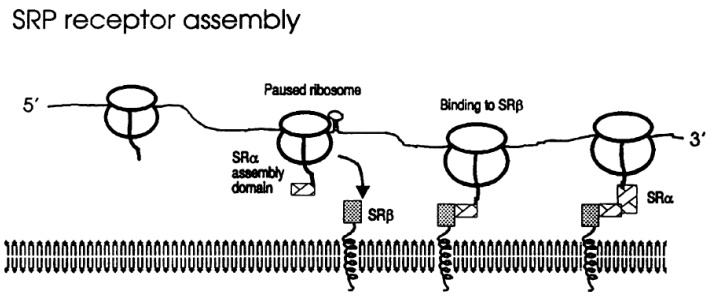


Figure 24: Model of SRα co-translational membrane assembly. Translation pause facilitates the folding of the N-terminal membrane-anchoring domain of SRα followed by targeting into the ER membrane and assembly on SRβ. Image is from (Young and Andrews 1996).

4.2.3 D1 protein of photosystem II

Photosystem II (PSII) is a large multiprotein complex located in the thylakoid membrane of plants, algae and cyanobacteria, and functions in oxygenic photosynthesis. The D1 protein of PSII has been shown to associate with the thylakoid membrane as well as assemble to the PSII complex co-translationally (Zhang et al. 1999). As discussed above for SRα protein, membrane proteins contain a stretch of rare codons downstream of the region encoding hydrophobic amino acid stretch (required to attach to the membrane) which might cause ribosome pausing (Young and Andrews 1996; Képès 1996). This ribosomal pause following the translation of the membrane anchoring domain enables D1 protein to attach to the membrane as well as to PSII complex co-translationally (Kim et al. 1991; Zhang et al. 1999).

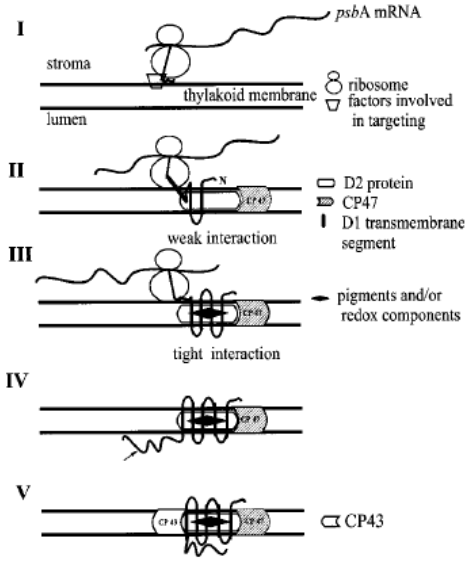


Figure 25: Model showing stepwise co-translational assembly of D1 protein into PSII complex and thylakoid membrane. D1 protein co-translationally inserts itself into thylakoid membrane (I, II) and also simultaneously associates with PSII complex (II-V). Image is from (Zhang et al. 1999).

4.2.4 Membrane ion channel

Co-translational assembly of membrane associated ion channels have also been reported. For instance, homo-oligomeric K^+ ion channels assemble co-translationally through their N-terminal interaction domain (Tu et al. 2000; Robinson and Deutsch 2005; Lu et al. 2001). In addition, heteromeric ion channel human ether-à-go-go-related Gene (hERG), consisting of hERG 1a and 1b subunits also assemble co-translationally (Liu et al. 2016; Phartiyal et al. 2007). The authors show that knocking down one of the subunits (hERG1a or hERG1b) reduces the level of both the target subunit as well as the partner subunit. In contrast, inhibition of hERG1b translation did not affect the association of hERG1a and 1b transcript. This suggests that the two mRNAs are associated with each other independent of their translation. Their association might be through RNA-binding protein which enables them to be in close proximity in the cell thereby promoting co-translational assembly (Liu et al. 2016).

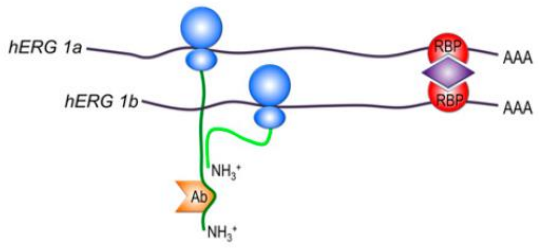


Figure 26: Model for co-translational association of hERG subunit mRNAs. hERG1a and hERG2a mRNA transcripts associate with each other through an RNA-binding protein (RBP) (not known yet) that enable the nascent proteins encoded by them to assemble co-translationally. Image is from (Liu et al. 2016).

4.2.5 IgE receptor (Fc€RI)

Fc€RI is a cell surface high affinity receptor expressed as a tetramer ($\alpha\beta\gamma_2$) on basophils and mast cells and as a trimer ($\alpha\gamma_2$) on antigen-presenting cells. It has been shown by *in vitro* experiments that the α subunit acts as a core and the other subunits of the complex assemble co-translationally. Expressing the subunits (α and γ) separately and then mixing the lysate does not lead to complex formation, which proves that the subunits bind to each other solely co-translationally (Fiebiger et al. 2005).

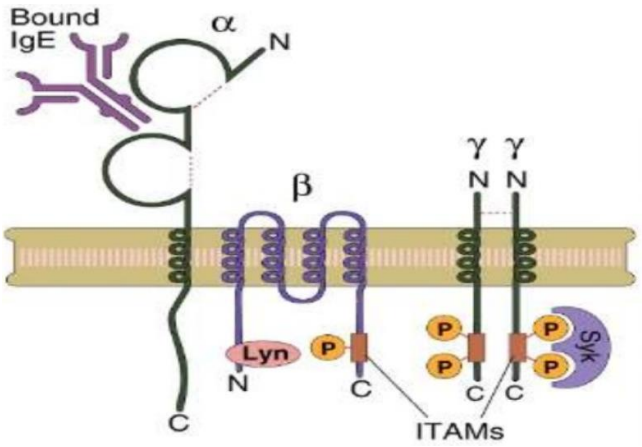


Figure 27: Schematic representation of Fc€RI receptor. A high affinity Fc€RI receptor consists of an α chain, a βchain and two γ chains. Lyn and Syk are two tyrosine kinases, P denotes phosphorylation and ITAM stands for immunoreceptor tyrosine-based activation motif. Image is from https://www.slideshare.net/MisHanif/immune-response-45935874.

4.2.6 Co-translation assembly of diverse heteromeric complexes in yeast

One of the earliest studies on co-translational assembly of non-membrane multisubunit complexes was reported in yeast (Halbach et al. 2009). While trying to study the RNAs associated with RNA-binding domain containing SET1C protein, the authors observed co-translational assembly of four out of the eight subunits of SET1C complex on *SET1C* mRNA. The deletion of RNA-binding domain of SET1C protein did not abolish these interactions. In turn translation inhibition affected the interactions, thereby ruling out the possibility of direct interaction of the proteins with *SET1C* mRNA. This was the first evidence of co-translational assembly of a multisubunit complex that functions (methylates histone) in the nucleus of eukaryotes (Halbach et al. 2009).

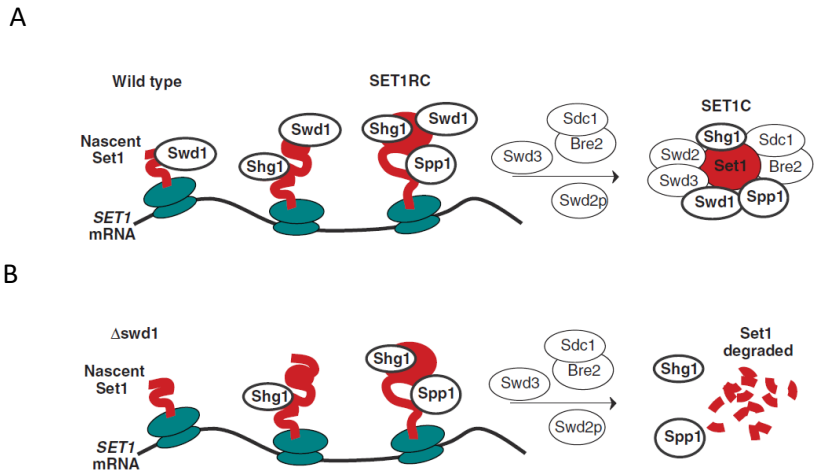


Figure 28: Model for co-translational assembly of the SET1C complex. Comparison of model for co-translational formation of the SET1RC (SET1 mRNA- associated complex) complex and maturation of SET1C in wildtype (A) and Δswd1 (B) strains. The absence of Swd1 protein did not abolish SET1RC formation but the protein SET1C was targeted for degradation. Nascent Set1 protein is indicated by a red line, *SET1* mRNA is indicated by a black line, mature Set1 and other SET1C subunits are indicated as ovals, ribosomes are depicted in turquoise and proteins found in SET1RC are indicated in bold letters. Image is from (Halbach et al. 2009).

Following this, there was another report that showed the widespread occurrence of co-translational assembly (Duncan and Mata 2011). In this paper, the authors studied a set of 31 *Schizosaccharomyces pombe* proteins, which were known subunits of different protein complexes, but did not contain RNA-binding domains. Interestingly, approximately 38% of the proteins co-purified mRNAs of their interacting protein partners from their respective protein complexes (Duncan and Mata 2011). Thus, co-translational association of two interacting protein partners was common in yeast. Note, however, that co-translational assembly of more than two mRNAs is also possible. Kassem et al. 2017 showed that three mRNAs of SAGA complex (SPT20, GCN5, ADA2) are tethered together by Ccr4-Not complex subunits and the glycolytic enzyme, glyceraldehyde-3-phosphate dehydrogenase (Tdh3), which aid their co-translational assembly (Kassem et al. 2017).

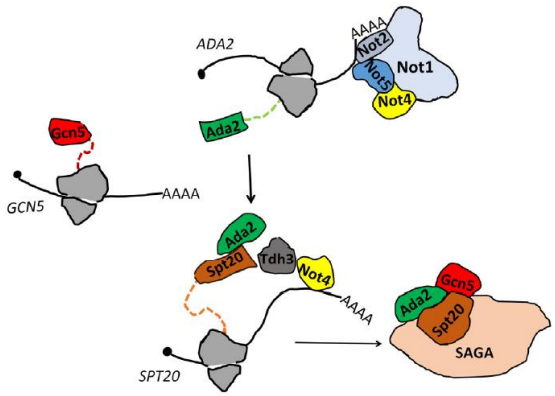


Figure 29: Co-translational assembly of SAGA complex in yeast. The *ADA2*, *SPT20* and *GCN5* mRNAs are tethered together at the site of translation in the presence of several Not subunits and Tdh3 as indicated. Image is from (Kassem et al. 2017).

The role of Ccr4-Not complex in co-translational assembly was furthered supported by another study from the group that showed the co-translational assembly of two subunits (Rpt1 and Rpt2) of the huge proteasome complex in a Ccr4-Not containing granule in the cytoplasm. In addition, their ribosome profiling data showed a brief ribosomal pause following the translation of the interaction domain of co-translationally assembling proteins, thereby enabling them to associate co-translationally (Panasenko et al. 2019).

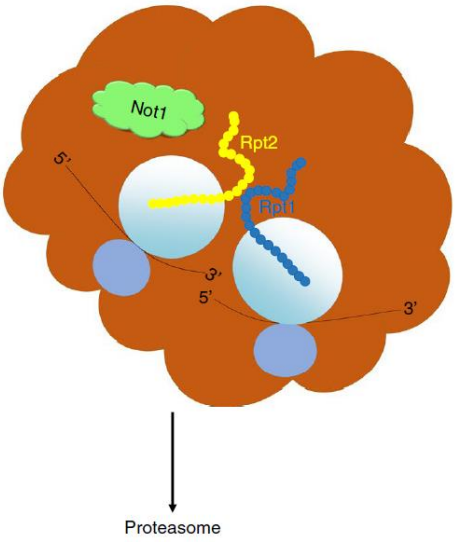


Figure 30: Model for co-translational assembly of Rpt1 and Rpt2 in Not1 containing assemblysomes. Rpt1 and Rpt2 encoding mRNAs assemble co-translationally through their N-terminal domains in Not1 containing membrane less organelles called assemblysomes which ultimately leads to the formation of proteasome. Rpt1 and Rpt2 nascent chains are shown in blue and yellow respectively. Image is from (Panasenko et al. 2019).

Finally, with the advent of newer high throughput techniques, Shiber et al studied the co-translational assembly of a group of well-characterised cytosolic protein complexes by selective ribosome profiling (described in Section 7 of Introduction). Nine out of the twelve complexes studied by them assembled co-translationally and the nascent proteins engaged in the assembly were prone to aggregation and degradation in the absence of their co-translationally assembled full-length partners. The authors also argued that the complexes that did not assemble co-translationally have dedicated chaperones or inhibitors (Tallec et al. 2007; Meurisse et al. 2014; Smardon et al. 2002) that might protect them from non-specific aggregations during assembly (Shiber et al. 2018).

4.2.7 Bacterial Lux Operon

Several decades after the initial report on bacterial co-translational assembly, another interesting study showed the importance of operon structure and polycistronic mRNAs in prokaryotes for spatial regulation of protein complex assembly. The authors argued that the gene arrangement in the form of operon is important for the proximity of interacting protein partners thereby facilitating their co-translational assembly. They studied the co-translational assembly of heterodimeric luciferase subunits LuxA and LuxB into complexes at the site of translation. Assembly of LuxA and LuxB initiates co-translationally on nascent LuxB protein and not vice versa, after the interaction domain of LuxB protein is synthesised by the ribosome. The ribosome associated chaperone trigger factor prevents premature co-translational interaction between them until the interaction domain of LuxB is fully exposed (Shieh et al. 2015). Therefore, the organization of genes into operon system in bacteria is essential for co-translational assembly of protein complexes.

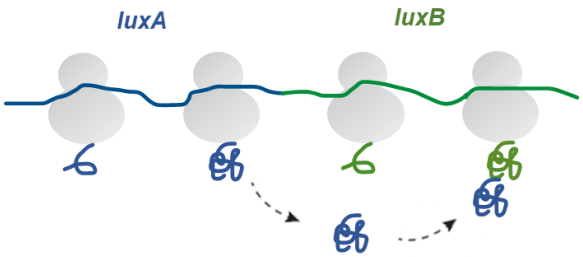


Figure 31: Co-translational assembly bacterial Lux operon subunits. Full-length LuxA protein (shown in blue) assembles co-translationally to nascent LuxB protein (shown in green) at its site of translation.

4.2.8 Co-translational assembly of protein complexes in mammalian cells

In spite of several studies on co-translational assembly of membrane proteins in mammals, there were no reports so far on this form of assembly in multisubunit cytoplasmic or nuclear complexes. We showed for the first time that nuclear transcription complexes assemble co-translationally in mammalian cells and the position of the interaction domain guide the directionality of the assembly (Discussed more in the Results section, (Kamenova, Mukherjee et al. 2019)). In parallel, the authors of the study which showed the co-translational assembly of two subunits of

proteasome complex by ribosome pausing in yeast also reported the same in mammalian cells (Panasenko et al. 2019).

4.3 Ribosomal pause associated with co-translational assembly

In some of the earliest reports of co-translational membrane protein assembly, ribosomal pausing has been reported following the translation of the membrane targeting domain to facilitate its targeting into the membrane (Képès 1996; Chartron, Hunt, and Frydman 2016; Shen and Shan 2010). Bacterial and mitochondrial proteins are inserted post-translationally into the membranes unlike yeast membrane proteins. Sequence analysis of a cluster of membrane proteins in yeast revealed the presence of a stretch of approximately 17 rare codons 56-75 codons downstream of the hydrophobic amino acid stretch encoding codons in the mRNA. This pause also known as the “+70 pause” aids in the folding and interaction of the membrane targeting domain with targeting factors (Képès 1996). It has been reported in the same year that the signal recognition particle receptor subunit α (SRα) undergoes a pause in the ribosome at nucleotide position 507 for co-translational membrane anchoring. The authors generated ribosome footprints of *invitro* translated SRα protein and studied the regions by hybridization and primer extension methods (Shen and Shan 2010; Chartron et al. 2016). Finally, with the emergence of new high throughput sequencing methods, it is now possible to study ribosome footprints *in vivo* at single nucleotide resolution with much higher precision. Using ribosome profiling (described above) it was shown recently that both Rpt1 and Rpt2 (subunits of the proteasome complex) assemble co-translationally following a ribosome pause at a specific codon pair DP (Panasenko et al. 2019). We (Kamenova, Mukherjee et al. 2019) and (Shiber et al. 2018) showed that the absence of co-translationally assembling full-length partner makes the nascent assembling partner prone to aggregation and degradation. Thus, summing up, ribosome pause might be essential to facilitate co-translation binding of partner proteins and essential factors and acts as a quality control checkpoint for translating mRNAs and nascent proteins. However, the pathway and the factors associated with co-translational assembly linked mRNA quality control is yet to be discovered. A general discussion on different mRNA quality control pathways is described in Section 6.1.

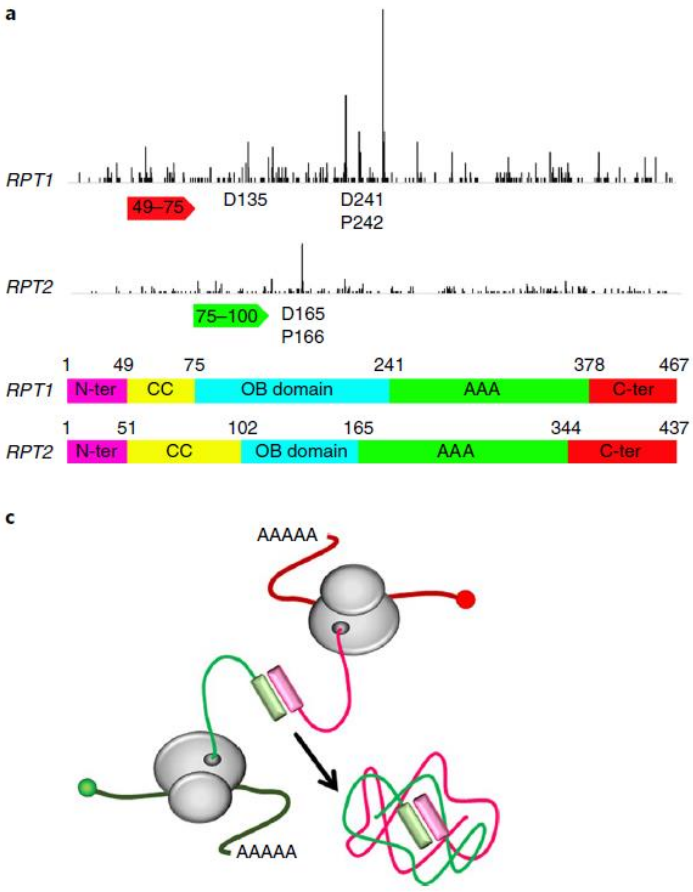


Figure 32: An example of ribosome pausing during co-translational assembly from (Panasenko et al. 2019). The interaction domains of RPT1 and RPT2 are shown in red and green respectively (a and c). Ribosome profiling data shows the presence of peaks following the synthesis of interaction domains of both RPT1 and RPT2, suggesting ribosome pause to facilitate co-translational assembly (a).

5. How are co-translationally assembling partners brought in close proximity to each other?

There could be several mechanisms by which proteins assembling co-translationally are brought close to each other. The possible mechanisms include: 1) chaperone-assisted co-translational assembly ***(Section 5-1)***, 2) cis-acting element driven co-translational assembly: close proximity of translating mRNAs and physical linkage by a bridging protein ***(Figure 33a, Section 5-2)***, or 3'UTR guided co-translational association of interacting proteins (***Figure 33c-d, Section 5-2***). 3) close proximity of mRNAs without physical linkage in RNA granules ***(Figure 33b, Section 5-3)***.

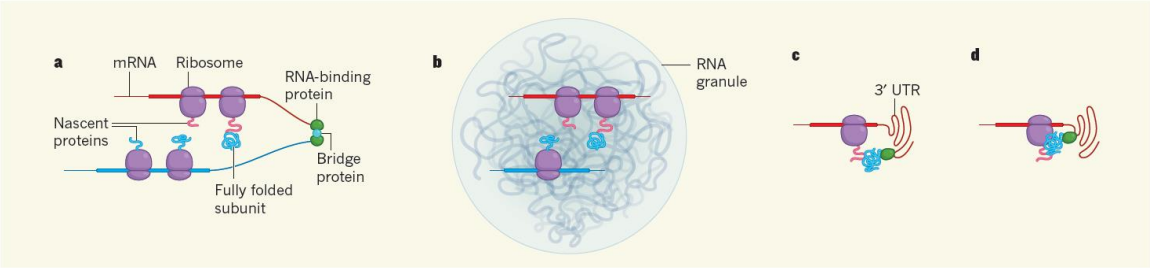


Figure 33: Possible mechanisms by which co-translationally assembling protein partners are brought near each other. a) An RNA-binding protein bridges two mRNAs either directly (not shown) or indirectly (shown) through another protein. Here, a fully folded subunit translated from one mRNA interacts with nascent protein on another mRNA. Thick regions of mRNAs represent translated region and thin regions are the untranslated regions. b) The mRNAs come into close proximity in RNA granules. c) 3'UTR of mRNAs might recruit the fully folded co-translationally assembling partner using an RNA-binding protein, bringing the subunit close to the mRNA's nascent protein. d) The recruited subunit could be temporarily deposited on the ribosome before being transferred to the nascent protein. Image is from (Mayr 2018).

5.1 Chaperones associated with co-translational assembly

Very little is known about the role of chaperones in co-translational assembly and their mechanism of action. Bacterial chaperone Trigger Factor (TF) is a ribosome associated chaperone that is involved in the temporal regulation of co-translational assembly interactions of Lux heterodimer. The initial contact of nascent LuxB with LuxA is delayed in the presence of TF thereby preventing non-specific interactions (Shieh et al. 2015). In parallel, the same group also put forward the role of ribosome associated Hsp70 chaperone Ssb in co-translational assembly of protein complexes

in yeast. The protein Ssb engages with partially synthesized interaction domains and dissociates prior to the onset of interaction with its co-translationally assembling partner (Shiber et al. 2018). Another report also shed some light on the protein glyceraldehyde-3-phosphate dehydrogenase Tdh3 being the chaperone playing role of the subunits of SAGA complex in yeast (Kassem et al. 2017). A detailed discussion on the different groups of chaperones involved in protein folding is provided in the following ***Section 6***.

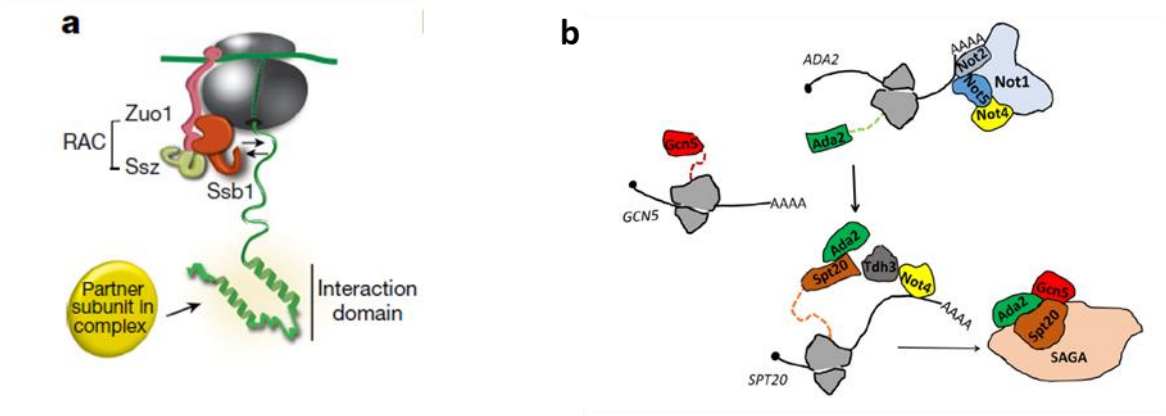


Figure 34: Chaperones assisting co-translational assembly in yeast. a) Schematic illustration of chaperones associated with ribosome bound nascent protein chain in yeast. The chaperones stay bound to the nascent chain until the interaction domain is ready to bind to the partner subunit. b) Tdh3 protein acts as a chaperone facilitating the co-translational association of Spt20, Gcn5 and Ada2 subunits of SAGA complex. Image is from (Shiber et al. 2018; Kassem et al. 2017).

5.2 Cis-acting RNA sequence elements

There has been no evidence so far on the involvement of cis-acting elements in co-translational assembly. Our results with ectopically expressed proteins (from cDNA constructs without 3'UTR) rules out the role of 3'UTR in co-translational assembly of the protein complexes studied by us. But its involvement in the assembly of other complexes cannot be ruled out. Nevertheless, a general discussion on the various mRNA cis elements is described.

Cis-acting RNA regulatory elements play an important role in the posttranscriptional control of gene expression. These elements are sequences located in various regions

of pre-mRNAs, like 5'UTRs, 3'UTRs, introns, coding regions, etc. Trans-acting factors binding to these elements modulate specific functions driven by the elements. Trans-acting factors include RNA-binding proteins (RNA-BPs) and microRNAs (miRNAs). This section describes the various types of cis-regulatory elements, their localization and function in the precursor or mature mRNA.

5.2.1 AU-rich element (ARE)

Repeating pentamer (AUUUA) with 1 or 2 A to U substitutions defines the structure of AU-rich elements. These elements are part of a group of genes whose expression requires fine regulation including cytokines, immune-regulatory genes and proto-oncogenes (Barreau et al. 2005; Chen and Shyu 1995). In fact, the first elements discovered in the 3'UTRs were the AU rich elements which facilitated the rapid decay of mRNAs (Barreau et al. 2005; Chen and Shyu 1995). Approximately 5% of the transcriptome contains 3'UTRs with AU-rich elements (Bakheet 2001). Transcripts containing these elements generally have short half-lives, but it can also stabilise the transcript under certain conditions depending on the trans-acting factor binding to it (Vlasova et al. 2005; Tebo et al. 2003). Hence, the trans-acting factor binding to it determines the outcome of the transcript. The basic structure of ARE elements is generally a repeating pentamer (AUUUA) with 1 or 2 A to U substitutions (Chen et al. 1995). AU-rich elements are clustered into five groups depending on their sequence content and position of A or U, as shown in the table below ***(Table 4).*** AU-rich elements are also found in the introns of pre-mRNAs (Vogel et al. 2016; Yoon et al. 2014; Mukherjee et al. 2011; Lebedeva et al. 2011). Majority of the trans-acting factors shuttle between nucleus and cytoplasm and so they can bind to the pre-mRNAs and carry out functions different from their function in the cytoplasm (Al-Ahmadi et al. 2009).

5.2.2 GU-rich element (GRE)

8% of transcripts of human transcriptome contain GU-rich elements and they are essential regulators of mRNA processing and stability (Vlasova and Bohjanen 2008; Halees et al. 2011). 3'UTRs contain conserved GU-rich elements in the form of consensus $U(GUUUG)_n$ sequences or GU repeats and they confer instability to mRNAs when tested *in vivo* (Rattenbacher et al. 2010; Vlasova et al. 2008). Similar to AU-rich elements, depending on the family of proteins binding to it, GU-rich elements

can induce stability or degradation of mRNAs. For example, the CELF family of proteins binds GU-rich elements of mRNAs to promote their degradation and the ELAVL family of protein function as RNA stabilizers (Irina Vlasova-St. Louis and Calandra Sagarsky 2018).

Table 4*:* Different clusters of AU-rich elements (ARE) and GU-rich elements (GRE). Adapted from (Irina Vlasova-St. Louis and Calandra Sagarsky 2018)

Cluster	ARE sequences	GRE sequences
I	AUUUAUUUAUUUAUUUAUUUA	GUUUGUUUGUUUGUUUGUUUG
II	AUUUAUUUAUUUAUUUA	GUUUGUUUGUUUGUUUG
III	WAUUUAUUUAUUUAW	GUKUGUUUGUKUG
IV	WWAUUUAUUUAWW	KKGUUUGUUUGKK
V	WWWWAUUUAWWWW	KKKU/GUKUG/UKKK

5.2.3 Polyadenylation sequences

One of the most important steps of pre-mRNA processing is poly(A) tail addition. All cellular mRNAs except replication dependent histone mRNAs are polyadenylated (Marzluff et al. 2008). It takes place in two tightly coupled reactions, cleavage and polyadenylation, carried out by many protein factors. These two reactions are controlled by *cis* elements located upstream and downstream of the actual pA site. In mammals, upstream elements include the polyadenylation signal (PAS) which is a hexamer of sequence AAUAAA/AUUAAA or close variants, U-rich elements and UGUA elements. Downstream elements include U-rich and GU-rich elements (Tian and Manley 2008). It has been reported that majority of mammalian mRNAs harbor a conserved AAUAAA or a close canonical variant, AUUAAA (Hoque et al. 2013; Xiao et al. 2016). A great diversity of PAS utilisation is observed in genome wide PAS analysis in mammalian cells (Wang et al. 2016; Tian et al. 2005). Alternate mRNA 3'UTR isoforms are generated under different cellular conditions by differential usage

of alternate PASs that leads to the production of mRNAs with same coding sequence but different 3'UTR lengths ***(Figure 35;*** Mayr, 2016; Mayr & Bartel, 2009; Weng, Li, Xie, & Shi, 2016). More than half of the human and mouse genes generate alternative mRNA isoforms that differ in their 3'UTRs but encode proteins with identical amino acid sequences. For example, the protein CD47 exists in two different isoforms CD47-Long UTR (CD47-LU) and CD47-Short UTR (CD47-SU). Due to the differences in the length of 3'UTR, CD47-LU and CD47-SU reside in different cellular compartments (Berkovits and Mayr 2015). The difference in the length of the 3'UTR is due to the presence of multiple poly(A) sites in the pre-mRNA as shown in ***Figure 35***.

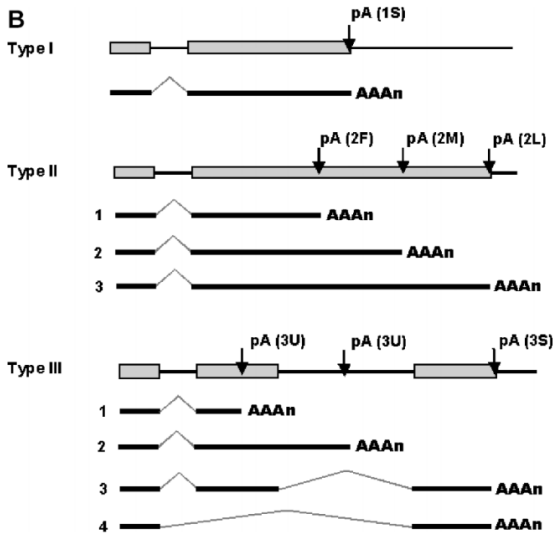


Figure 35: Schematic representation of the different types of isoforms generated due to the usage of different poly(A) sites in a particular mRNA. The arrows show the position of poly(A) site where the mRNA is cleaved. Three types of genes depending on the number of their poly(A) site are shown here. Type I gene has a single poly(A) site, type II gene has alternate poly(A) sites located in the 3' most exon and type III gene has alternate poly(A) sites located in different exons. Exons are represented as boxes, introns as lines and poly(A) site as pA. Image is from (Tian et al. 2005).

5.2.4 3'UTRs

Although the coding region of an mRNA is considered to be the most important since it translates to protein which ultimately carries out all the functions in the cells, untranslated regions at the 5' and 3' ends are also very important regulatory elements (Proudfoot & Brownlee, 1976***)***. The untranslated regions are more accessible to

regulatory factors since they are not coated by the ribosomes. It has been reported that the 3' UTR sequence can mediate important protein-protein interactions that ultimately define the function of a protein without altering its coding sequence (Berkovits and Mayr 2015). The 3'UTR length varies across different species depending on the genome size and hence complexity of various organisms. It ranges from approximately 60 nucleotides in bacteria and archaea, 150 nucleotides in yeast, 140 nucleotides in worms to 1200 nucleotides in human (Chen et al. 2012; Mayr 2016). It shows a high degree of sequence conservation across species (Xie et al. 2005; Siepel et al. 2005). Some of the most important functions of 3'UTR include 1) regulation of mRNA stability through AU rich elements and miRNAs, thereby affecting translation protein levels (Bartel 2009; Barreau et al. 2005; Chen and Shyu 1995), 2) regulation of mRNA localisation thereby enabling spatial regulation of translation (Niednery, Edelmanny, and Niessing 2014; Martin and Ephrussi 2009), 3) cleaving themselves and leading a separate life inside the cell as noncoding RNAs (Mercer et al. 2011; Kocabas et al. 2015; Chao and Vogel 2016), 4) driving protein-protein interactions (PPIs) that has widespread consequences starting from protein complex formation to protein localization (Chartron et al. 2016; Halbach et al. 2009; Duncan and Mata 2011; Berkovits and Mayr 2015). As mentioned in the previous section, alternative polyadenylation (APA) sites leads to mRNAs with short or long 3'UTRs. Around 51-79% of genes express alternate 3'UTRs (Derti et al. 2012; Mayr et al. 2013; Singh et al. 2018). It is expected that an RBP which binds a long 3'UTR isoform should also bind the short 3'UTR isoform as the sequence of the short UTR isoform is contained in the long UTR isoform. But this might not be the case necessarily as the long UTR can form secondary and tertiary structures thereby making the binding site of certain RNA binding proteins (RBPs) inaccessible unlike shorter UTRs (Pianka et al. 2007; Kristjánsdóttir et al. 2015). Additionally, RBPs also act in cooperation with each other and thus its functionality is also dependent on its neighbouring sequences and it might not act the same way in isolation (Wissink, Fogarty, and Grimson 2016; Jens and Rajewsky 2015; Campbell et al. 2012). Alternate 3'UTR ratios can vary across different tissues and cell types. Differentiated tissues tend to express longer 3'UTRs than embryonic tissues (Shi et al., 2011; Tian et al., 2005; Ulitsky et al., 2012***)***. Majority of the functions carried out by 3'UTRs depend on the effector proteins binding to it through RBPs. For example, the recruitment of deadenylase by RBPs for mRNA destabilization and recruitment of decapping enzymes by RBPs for repression of

translation (Zaessinger et al. 2006; Chen et al. 2014). 3'UTRs of membrane protein encoding mRNAs have been shown to recruit signal recognition particle so that it can co-translationally assemble with the nascent peptide containing the signal sequence exiting the ribosome (Chartron et al. 2016). Thus, mRNAs which are mostly considered as template for protein synthesis, can also function in regulation through their untranslated regions.

5.2.5 5'UTRs

The length of the 5'UTR did not change much during evolution unlike 3'UTRs (Lynch et al. 2005; Pesole et al. 2001; Ferreira et al. 2013)**.** The median length of 5'UTRs ranges from approximately 53 nucleotides in yeast to 218 nucleotides in humans (Hernández et al. 2010; Lingala and Ghany 2016). However, 5'UTR lengths may vary dramatically among individual genes in higher eukaryotes which points towards the possibility of fine regulation of a specific subset of mRNAs (Pot 2008; Mignone et al. 2002). The 5' cap and UTR act as the entry point for ribosomes for translation initiation. The ribosome scans the region for an AUG start codon to initiate translation (Kozak 1978). Some mRNAs species lack 5'UTR like mammalian mitochondria mRNAs whereas some mRNAs have highly structured 5'UTRs. Regulation of translation by 5' UTRs is mainly attributed by its structure. The secondary structures in the 5' UTRs block the entry of ribosomes (Pelletier and Sonenberg 1985; Kozak 1986; Hentze et al. 1987; Haimov et al. 2015). G quadruplexes reported to form in the 5'UTR are involved in translational repression by affecting the pre-initiation complex binding or by slowing down scanning (Bugaut and Balasubramanian 2012; Song et al. 2016; Halder et al. 2009; Bolduc et al. 2016; Beaudoin and Perreault 2010). Tertiary structure formed by interactions between secondary structures are also common. This is exemplified by pseudoknot structures consisting of two intercalated stem loop structures and is proposed to form in the 5'UTR of human interferon gamma (IFNG) mRNA (Cohen-Chalamish et al. 2009; Ben-Asouli et al. 2002). Internal ribosome entry site (IRES) in viral genomes is one of the best studied examples of RNA structure and function. Viruses have evolved this site to avoid cap-dependent translation initiation and recruit ribosome machinery directly to this site (Leppek et al. 2018). Around 10 to 15% of mammalian mRNAs are predicted to contain IRESs (Spriggs et al. 2008). Another example of 5' cap independent translation initiation was reported recently where the authors studied the function of multisubunit translation initiation factor eIF3

in directly regulating a subset of mRNAs by binding to the stem loop structures in their 5'UTRs (Lee et al. 2016).

5.3 RNA granules

Only one study (Panasenko et al. 2019) showed the presence of RNA granules in co-translational assembly so far (mentioned in section 4.2.6 and ***Figure 30*** and also described in this section). A general discussion on the evolution of RNA phase separation studies and RNA granule formation is described below.

A transformation of one physical state to another, like freezing of water into ice (liquid to solid) or water to vapour (liquid to gas), is called phase separation. A special form of phase separation that exists in living organisms is liquid-liquid phase separation. Here a homogenous solution of molecules inside the cell spontaneously separates into two co-existing liquid phases, a dense phase that is enriched for these molecules and a liquid phase that is depleted (Alberti 2017).This gives rise to membrane-less organelles inside cells consisting of macromolecules like RNAs and proteins. Examples of these type of organelles have been described previously like centrosome, processing or P bodies and nucleolus in the nucleus but it has never been studied so extensively until recently (Woodruff et al., 2017, Mitrea & Kriwacki, 2016; Uversky, 2017). Phase separation can be formed within one, two or multiple proteins and the factors that regulate their formation include protein concentration, solubility, affinity and valency of phase separating proteins (Li et al. 2012). Membrane-less organelles formed by phase separated proteins have been implicated in various cellular functions like stress response, gene expression regulation, control of signal transduction (Li et al., 2012; Voigts et al., 2016; Wheeler et al. 2016). These organelles have also been implicated in age-related neurodegenerative disorders (Murakami et al. 2015; Molliex et al. 2015; Conicella et al. 2016; Burke et al. 2015; Boeynaems et al. 2017). Cellular RNAs along with RNA binding proteins (RBPs) have also been shown to play role in phase separation very recently (Maharana et al. 2018; Langdon et al. 2018). This is evident in stress granules (SGs) consisting of cytoplasmic RNAs associated with multiple RBPs with low complexity regions (LCRs) (Langdon et al. 2018; Maharana et al. 2018; Molliex et al. 2015; Murray et al. 2017). Further insight into the mechanistic details of RNA-protein interaction driven phase separation was shown by two recent papers. Maharana et al., 2018 argued that the sequence specificity of RBP has no role

in phase separation. Instead the abundance of RNA and proteins regulate its formation; high RNA:protein ratio, the situation which is mostly observed in the nucleus, prevent phase separation and low RNA:protein ratio which is typically observed in the cytoplasm, facilitate phase separation (Maharana et al. 2018). In addition, Langdon et al., 2018 showed that RNA secondary structure has a role in forming specific type of phase separated granules, either by exposing sequences to bind complementary RNAs or by masking them (Langdon et al. 2018). Some of the mRNP granules reported are described below.

5.3.1 mRNP granules

Following transcription, cellular mRNAs are coated with proteins and they exist as messenger ribonucleoprotein (mRNP) complexes. The mRNP composition regulates various aspects of an mRNA lifecycle starting from its processing to its eventual degradation. Depending on its fate, several mRNPs can assemble with each other forming an mRNP granule (Zeitelhofer et al. 2008; Anderson and Kedersha 2009; Arkov and Ramos 2010; Buchan and Parker 2009; Eulalio et al. 2007; Franks and Lykke-Andersen 2008). Some of them are described below:

5.3.1.1 Stress granules and Processing bodies (P-bodies)

One of the best characterised cytoplasmic mRNP granules are stress granules and P bodies. Stress granules are made up of translation initiation stalled mRNAs which are formed when translation initiation is inhibited by drugs or stress response (Anderson and Kedersha 2009). P-bodies are cytoplasmic mRNP granules consisting of factors related to translation repression and mRNA decay. The formation of this type of granule is reversible and mRNAs can recycle back to translating polysomes from P-bodies (Parker et al. 2005). Moreover, P-bodies are not imperative for mRNA decay to occur (Huch and Nissan 2017; Decker et al. 2007; Eulalio et al. 2007) since mRNAs can be degraded outside P-bodies (Aizer et al. 2014). A recent model for P-body function has emerged which suggests that P-bodies act as storage sites for translationally repressed mRNAs and undergo liquid-liquid phase separation when the mRNA decay factors accumulate, thus decreasing RNA:protein ratio and promoting phase separation (Luo et al. 2018; Polymenidou 2018). Although stress granules and P-bodies share protein factors, come in contact with each other and are both induced by stress, they differ in molecular composition and function (Li et al. 2013). The

mRNAs within stress granules can also resume translation or can be targeted to lysosomes for autophagy (Protter and Parker 2016).

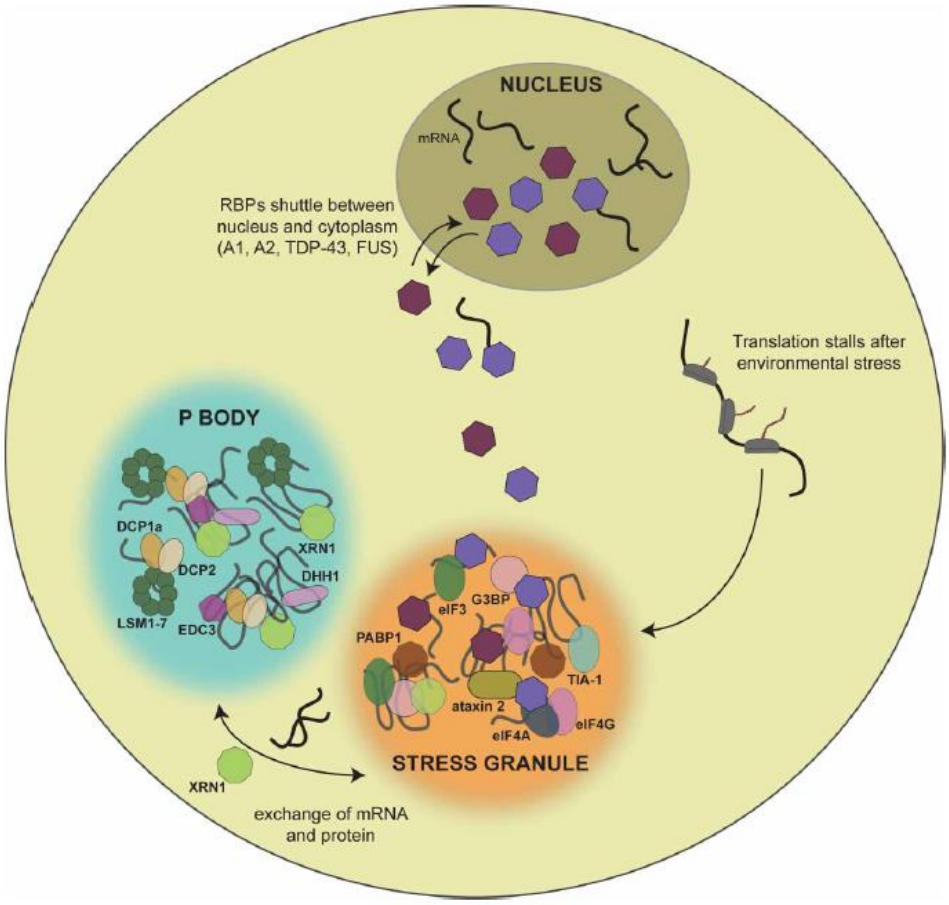


Figure 36*:* Schematic representation of cytoplasmic movement of mRNAs through P-body and stress granules. Stress granules are composed of translation initiation stalled polyadenylated mRNA transcripts, RNA-binding proteins, translation initiation factors and small ribosomal subunits. P-bodies consist of mRNA decapping and decay factor. Image is from (Harrison and Shorter 2017).

5.3.1.2 TIS granule

A more recently discovered membrane less organelle is the TIS granule formed by the broadly expressed RNA binding protein TIS11B. This granule is enriched in membrane protein encoding mRNAs with AU rich elements. It forms a subcellular compartment with endoplasmic reticulum (ER) called the TIGER domain (Ma and Mayr 2018). This compartment is essential for establishing specific protein-protein interactions between SET and membrane proteins, which are not possible outside. Although the concept of

biological phase transition has revolutionized our understanding of cellular compartmentalization, but this field is still in its infancy and a lot still needs to be uncovered.

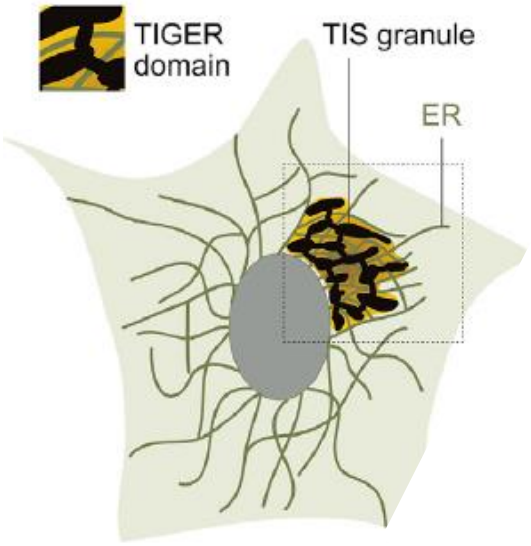


Figure 37*:* Schematic representation of TIGER domain cellular compartment. TIS granule (made up of TIS11B protein and membrane protein encoding mRNAs) and endoplasmic reticulum (ER) forms the cellular compartment TIGER domain. Image adapted from (Ma and Mayr 2018).

5.3.1.3 Not1 containing assemblysomes

Two subunits of the proteasome complex (Rpt1 and Rpt2) were shown to assemble co-translationally in heavy bodies in the cytoplasm by sucrose density gradient experiments (Panasenko et al., 2019). These bodies are different from polysomes as control experiments disrupting polysomes did not affect the assemblysomes. In addition to Rpt1 and Rpt2, the assemblysomes also contain the protein Not1, a subunit of the Ccr4-Not complex. Hence, the authors named the bodies as Not1 containing assemblysomes (NCA). Other than facilitating co-localization of mRNAs encoding co-translationally assembling partners, NCA might be involved in temporal regulation of co-translational association of protein partners thereby preventing translation initiation and ribosome collision until the partners are associated. Thus several translation factors might be present in the NCA and the composition of NCA might evolve and vary in different steps of assembly. NCA might also be responsible for the degradation

of a protein subunit if it is translated and released in the absence of its partner. Nevertheless, it would be interesting to study the composition and function of all the components of NCA in detail and also identify other RNA granules associated with co-translational assembly in the cytoplasm.

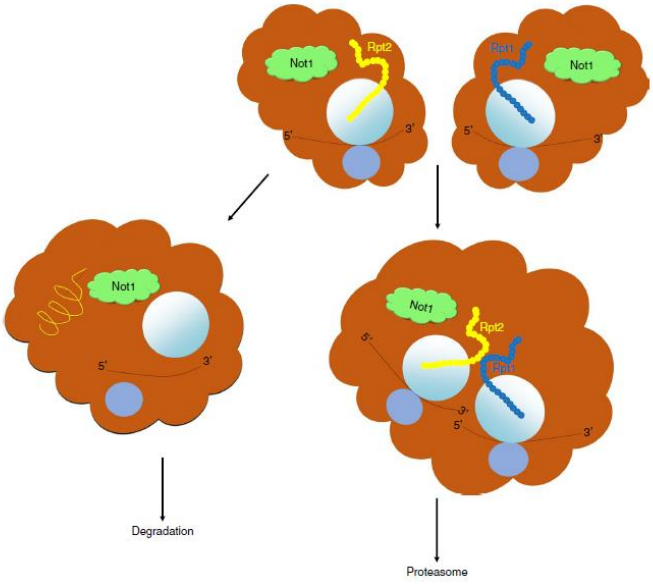


Figure 38: **Schematic representation of co-translational assembly of Rpt1 and Rpt2 in NCA.** Translation of *RPT1* and *RPT2* mRNAs are induced under proteotoxic stress and the ribonucleoprotein complexes (RNCs) assemble into Not1-containing particles (shown here by the orange could-like structure). Productive interaction between Rpt1 and Rpt2 in NCA leads to the formation of proteasomes. In the absence of nascent Rpt1 and Rpt2 assembly, they will be ultimately degraded. Image is from (Panasenko et al. 2019).

6. mRNA and protein surveillance in cells

As mentioned in Introduction Section 4.3, ribosomal pause might be essential to facilitate co-translation binding of partner proteins and essential factors and might act as a quality control checkpoint for translating mRNAs and nascent proteins. However, the pathway and the factors associated with co-translational assembly linked mRNA quality control is yet to be discovered. Another intriguing aspect that is still unknown is how protein folding is co-ordinated temporally with co-translational assembly. Nevertheless, a general discussion on different mRNA quality control pathways and different protein folding chaperones is provided in the sub-sections below.

6.1 mRNA Quality Control pathways

Most eukaryotic cellular mRNA transcripts undergo a number of quality control tests before being translated into functional proteins by ribosomes. Transcripts that fail to pass the quality checks are prevented from engaging in protein synthesis. The first level of mRNA quality control initiates in the nucleus and is generally coordinated with its synthesis. The second phase of surveillance is carried out by dedicated machineries in the cell cytoplasm. The different mRNA surveillance pathways in the cytoplasm include nonsense-mediated decay (NMD), no-go decay (NGD) and non-stop decay (NSD). They recognise and eliminate mRNAs with premature stop codon (PTC), truncated and translation-stalled mRNAs and mRNAs without natural stop codon respectively (Doma and Parker 2007; Popp and Maquat 2013; Welch and Jacobson 1999).

6.1.1 mRNA surveillance in the nucleus

Transcription and mRNA processing are tightly coupled in the cell nucleus. mRNA processing includes splicing and poly(A) tail addition to the 3' end and is facilitated by a series of RNA-binding proteins tightly bound to the mRNA forming the messenger ribonucleoproteins (mRNPs). Splicing factors are recruited during transcription by the general transcription factor TFIID and is transferred to the C-terminal domain of RNA Pol II (Maquat and Carmichael 2001). Splicing in turn is accurately coordinated with mRNA export. mRNA export factors are specifically recruited to the pre-mRNAs that have either engaged in or completed splicing (Maquat and Carmichael 2001). Finally, export-competent mRNPs are transferred to the cytoplasm through the nuclear pore. The faulty mRNPs are retained in the nucleus (Tutucci and Stutz 2011; Mital et al. 2005; Galy et al. 2004). In case these faulty mRNPs reach the cytoplasm, they are degraded by the cytoplasmic mRNA quality control mechanisms described below.

6.1.2 Nonsense- mediated decay (NMD)

Nonsense-mediated decay (NMD) mRNA surveillance pathway recognizes and eliminates mRNAs with premature stop codons (PTCs) in eukaryotes. A cluster of proteins called the exon junction complex (EJC) are deposited at the junction of two exons on the pre-mRNA during splicing (Gehring et al. 2009). This EJC complex is removed from the mRNA during the first round of translation by the ribosomes. But the presence of PTC more than 50-55 nucleotides upstream of the EJC causes the

ribosome to stall. The stalled ribosome fails to evict the downstream EJC and this triggers the NMD pathway (Popp and Maquat 2013). Note, however that the 50-55 nucleotide distance between PTC (i.e., stalled ribosome) and EJC is important; a distance shorter than that would enable the translating ribosome to advance enough distance to displace the EJC complex, thereby preventing the activation of the NMD pathway (**Figure 39**) (Dostie and Dreyfuss 2002; Alkalaeva et al. 2006).

An interesting question is how PTCs are at all formed in the mRNAs. There could be several possible reasons: a) transcription initiation upstream of the proper start site could lead to an mRNA with a nonsense codon upstream of or within translation reading frame, b) incorrect pre-mRNA splicing could result in an intron-derived premature nonsense codon, c) Programmed DNA rearrangements of T-cell receptor (TCR) and immunoglobulin (Ig) genes that generate diverse antigen receptors also leads to premature nonsense codons in approximately two out of three cases (Li and Wilkinson 1998). Additionally, natural NMD targets also exist in the cells, for example, selenocysteine codon UGA in selenoprotein mRNAs direct termination of translation in some situations (Wittmann et al. 2006; Sun et al. 2001; Moriarty et al. 1998).

6.1.2.1 Mechanism of NMD pathway

NMD takes place in three steps: 1) detection of PTC containing NMD transcripts, 2) tagging of the substrates, and finally 3) degradation of the substrates. NMD substrate is detected during the first round of translation. After detection, the PTC is tagged by formation of a complex with serine/threonine kinase SMG1, UPF1, and eukaryotic release factors eRF1-eRF3 called the SURF complex at the terminating ribosome. The factors associated and there are described in ***Figure 39***. (Hwang et al. 2010; Kashima et al. 2006).

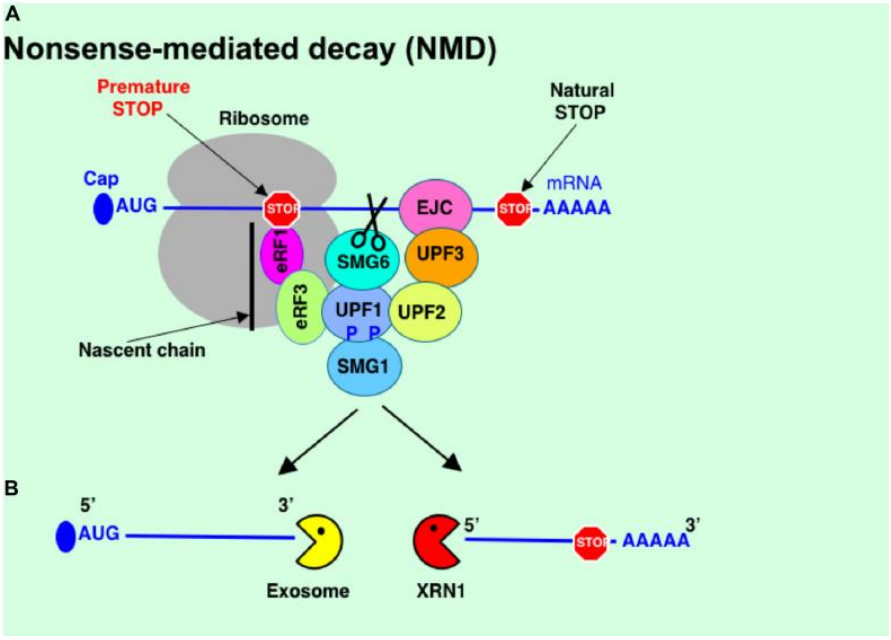


Figure 39: Mechanism of NMD pathway (EJC model). The UPF1-SMG1 binds to EJC via interaction with UPF2, which in turn binds to the EJC through interaction with UPF3 or UPF3X. SMG1 phosphorylates UPF1 and this hyperphosphorylated UPF1 recruits SMG6 protein (Isken and Maquat 2007). SMG6 performs endonucleolytic cleavage of mRNA. This cleavage occurs between the PTC and EJC sites of the defective mRNA during the last stage of NMD (Eberle et al. 2009; Izaurralde et al. 2008). Activated UPF1 recruits further downstream factors to facilitate decapping and deadenylation of target mRNAs followed by their degradation (Kervestin and Jacobson 2012; Eberle et al. 2009). Image adapted from (Karamyshev and Karamysheva 2018) UPF1: Regulator of nonsense transcripts 1; SMG1: Serine-threonine protein kinase; UPF; Up-frameshift.

Apart from the EJC model, alternative models of NMD pathway based on 3'UTRs also exist: 1) UPF1 sensing/potentiation model and 2) the faux 3'UTR model. According to the sensing/potentiation model, UPF1 detects 3'UTR length and stimulates its degradation (Hogg and Goff 2010; Bühler et al. 2006). 3'UTR isoforms are quite common in mammalian cells (discussed in Introduction section 5.2.4), so it might be difficult to explain a fail-safe pathway just based on the length of the 3'UTR. However, it cannot be ruled out that the secondary structure of the 3'UTR and the factors binding to it might play a role in the NMD pathway. According to the faux model, efficient termination is prevented if the distance between PTC and poly (A) tail is large (Kervestin et al. 2004). Very little is known about the degradation of the truncated polypeptide that is generated before the PTC-containing mRNA degradation occurs. Some studies in yeast points towards the fact that UPF1 might have E3 ubiquitin ligase properties that promote the degradation of truncated polypeptide through proteasome (Kuroha et al. 2009; Takahashi et al. 2008). However, the mechanism of mRNA degradation in mammalian cells remains an open question.

6.1.3 No-go decay (NGD)

No-go decay pathway takes care of translation elongation stalled complexes. The ribosomes might stall on the mRNAs for reasons like encountering a rare codon or secondary structure in the mRNA which might physically block the movement of ribosomes or by specific features of the nascent peptide which might hinder the movement of the peptide through the ribosome exit tunnel (Tsuboi et al. 2012). The proteins that play central roles in NGD pathway are Pelota (mammals)/ Dom34(yeast) and HBS1 (Doma and Parker 2007; Karamyshev and Karamysheva 2018). Dom34/HBS1 mimics the elongation factor/tRNA complex and hence binds to the ribosome A site but they promote dissociation of elongation complex instead of terminating it (Shoemaker and Green 2012; Karamyshev and Karamysheva 2018). Following this, they facilitate the degradation of the substrate mRNA but the endonuclease involved in the process is not yet known. Recently, it was also reported that NGD pathway is triggered by ribosome collision on the mRNA, thereby leading to its degradation (Simms et al. 2017).

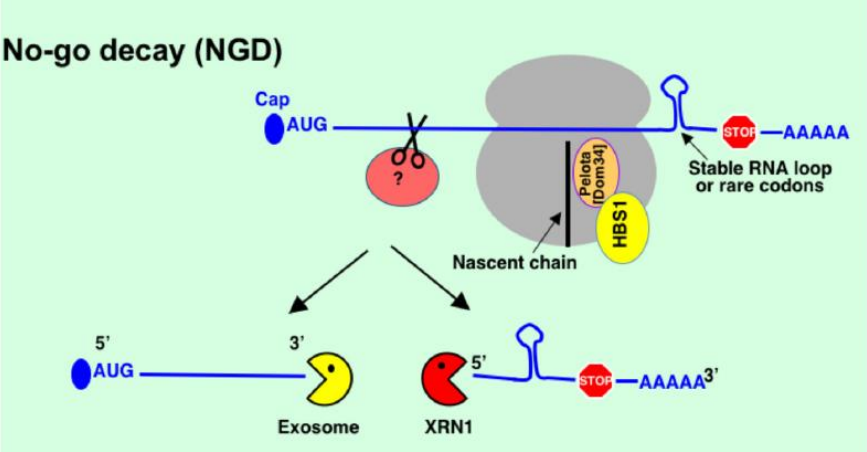


Figure 40: Mechanism of NGD pathway. Proteins Pelota (in mammals; Dom34 in yeast) and HBS1 are structurally related to termination factors eRF1 and eRF3 respectively (Atkinson, Baldauf, and Hauryliuk 2008). They also mimic translation elongation complex and hence bind to ribosome A site thereby promoting dissociation of ribosomes. Reports suggest endonucleolytic cleavage of the mRNA upstream of the stalled ribosome (Tsuboi et al. 2012). Image is from (Karamyshev and Karamysheva 2018).

6.1.4 Non-stop decay (NSD)

This type of decay pathway is activated when the cell has to degrade mRNAs that lack stop codons (non-stop mRNAs). Non-stop mRNAs are generated due to reasons like faulty polyadenylation within the ORF leading to the generation of aberrant mRNA without a stop codon or endonucleotytic cleavage of mRNA generating a non-stop mRNA lacking poly(A) tail (Graille and Séraphin 2012; Ozsolak et al. 2010). Due to the absence of a stop codon, the translation of poly(A) tail might lead to a protein with a poly-lysine stretch. This protein hinders translation elongation due to the interaction of the lysine stretch with the negatively charged ribosomal RNA in the ribosome exit tunnel (Dimitrova et al. 2009). For poly(A) tail lacking non-stop mRNAs, ribosome stalls at the end of the mRNA leading to translational repression and ultimately, degradation of mRNA. It has been shown that Ski7 protein plays a role in NSD pathway in yeast. In the absence of Ski7 protein, Hbs1 and Dom34 proteins function in both NGD and NSD pathways (Tsuboi et al. 2012). The exact mechanism and the endonuclease carrying out the mRNA degradation is not yet known.

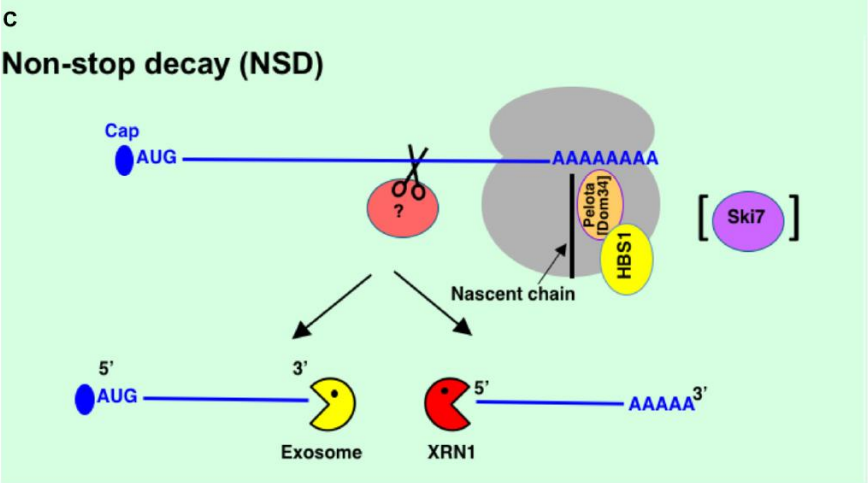


Figure 41: Mechanism of NSD pathway. Ribosome stalls at the end of mRNAs lacking stop codon, leading to translational repression and ultimately, degradation of mRNA. The exact mechanism and the endonuclease carrying out the mRNA degradation is not yet known. It has been suggested that Ski7 protein in yeast or Hbs1 and Dom34 proteins function in both NGD and NSD pathways (Tsuboi et al. 2012). Image is from (Karamyshev and Karamysheva 2018).

6.1.5 Ribosome-associated protein quality control (RAPP)

Nascent chains translated by the ribosome initiates functional interactions with various factors co-translationally, for example chaperones or various organelle-targeting factors. The recent novel type of ribosome-associated protein quality control discovered (Karamyshev and Karamysheva 2018; Karamyshev et al. 2014) involves the identification of nascent chains that are unable to form these functional interactions and directs the mRNA and the nascent protein for degradation. For example, secretory proteins possess specific signal sequence which enable them to interact co-translationally with the ribonucleoprotein complex, signal recognition particle (SRP) which in turn targets them to the ER membrane for translocation. In absence of this interaction with SRP due to either aberrant signal sequence or mutated SRP, degradation of the specific mRNA encoding secretory protein is triggered (Pinarbasi et al. 2018).

6.2 Chaperone, the key player assisting protein folding

Following the exit of a nascent polypeptide chain from the ribosome exit tunnel, it encounters a plethora of molecular chaperones and other peptide chain modifying enzymes that act on the emerging protein (Buchner 2019). These proteins are together called ribosome-bound protein biogenesis factors (RPBs). Molecular chaperones are a dedicated group of protein factors that play important role in protein folding and maintenance of protein homeostasis in the cell. By definition, molecular chaperones are not part of the final protein structure but bind to the protein folding intermediates to aid in folding and assembly. They assist in the *de novo* folding of proteins and maintains pre-existing proteins in the native state thereby preventing non-specific aggregation in the very crowded cellular environment.

Molecular chaperones are also essential to prevent aggregation under cellular stress conditions and play important in protein quality control processes by targeting misfolded proteins for proteolytic degradation. Protein aggregation caused by protein misfolding have been reported in several neurological diseases (Chiti and Dobson 2006). Examples include aggregated α-synuclein in Parkinson's disease, huntingtin in Huntington's disease, as well as the extracellular β-amyloid plaques in Alzheimer's disease (Chiti and Dobson 2006). In addition to that, there are reports of cellular homeostasis studies using model organisms demonstrating a gradual decline in cellular proteostasis capacity occurring with aging in several age-related diseases like type II diabetes, peripheral amyloidosis, cancer, and cardiovascular diseases (Morimoto, 2008). Chaperones are essential players in preventing these diseased conditions.

In the following sections, we will be focussing mainly on the various chaperones localised in the cytoplasm and their functions.

6.2.1 Cytosolic Chaperones

The first set of chaperones that start acting on the nascent polypeptides in the cytoplasm are ribosome binding chaperones, namely trigger factor (TF) in prokaryotes and nascent-chain-associated complex (NAC) and specialised Hsp70s (including yeast Ssb chaperone) in eukaryotes. A second set of chaperones act downstream without directly interacting with the ribosome. A comparative account of cytoplasmic chaperones in bacteria and eukarya are shown below in ***Figure 42***.

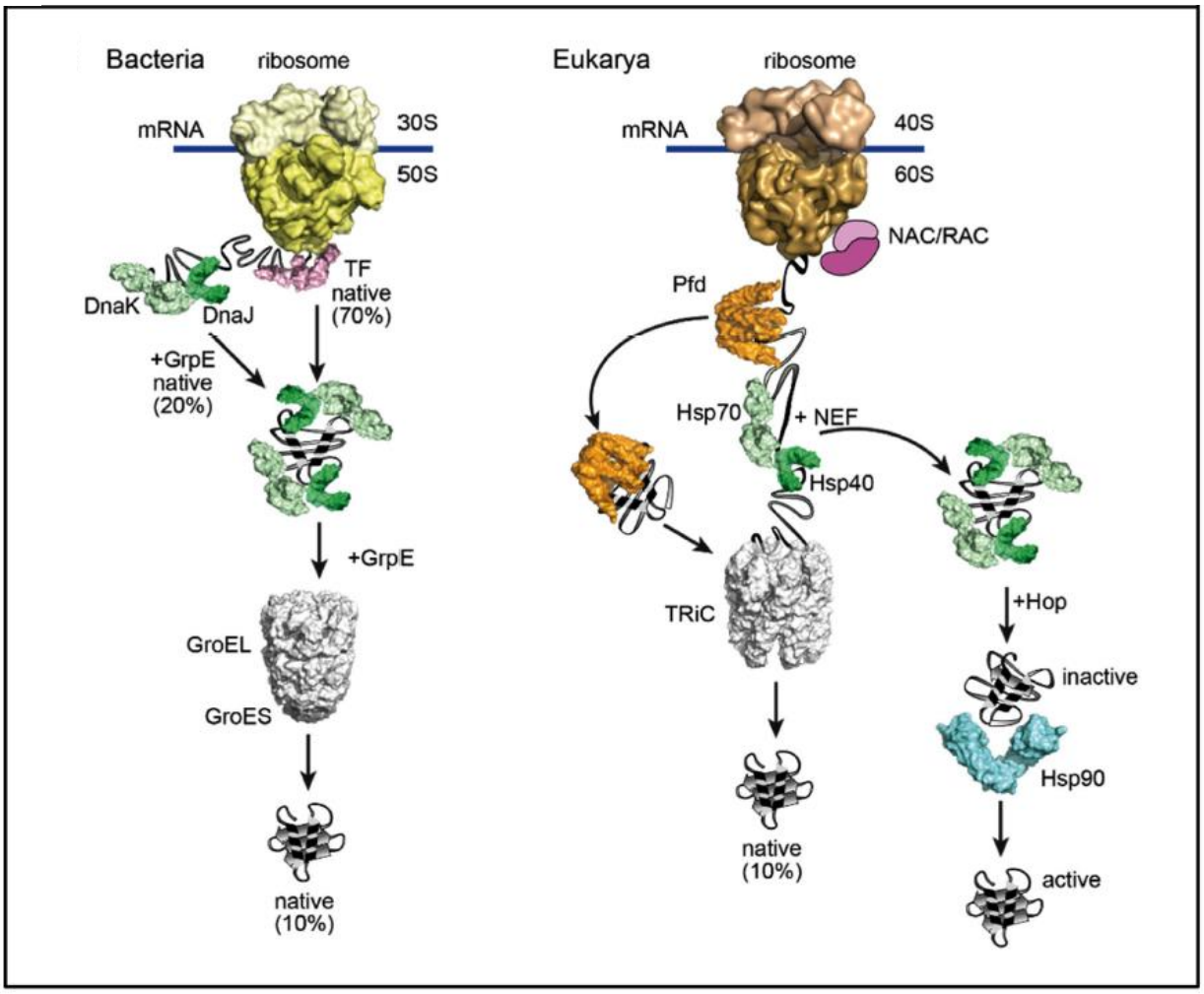


Figure 42: Cytoplasmic chaperone pathways in (a) bacteria and (c) eukarya. Folding of a nascent protein begins co-translationally and finish post-translationally following transfer to downstream chaperones. The different cytosolic chaperones, both ribosome-associated and acting downstream of ribosomes, across different domains of life are shown. NAC/RAC (Nascent chain associated complex/ Ribosome associated complex) are eukaryotic ribosome-associated chaperones and functionally similar to bacterial trigger factor (TF). Prefoldin (Pfd) recruits TRiC to certain nascent chains. Like TRiC, Hsp90 chaperone system acts downstream

of Hsp70. Hop (Hsp70-90 organizing protein) mediates contacts between Hsp70 and Hsp90. Image is from (Hartl 2017).

6.2.1.1 Ribosome-associated chaperones and their mechanism of action

The ribosome-associated molecular chaperones include TF (in prokaryotes), ribosome-associated complex (RAC) and Ssb, the specialised Hsp70 chaperone (in *Saccharomyces cerevisiae*), MPP11 and Hsp70L1 (mRAC in mammals), and nascent chain-associated complex (NAC in archaea and eukaryotes) (***Figure 42 and 43***) ((Hartl et al. 2011; Preissler and Deuerling 2012; Bukau et al. 2000).

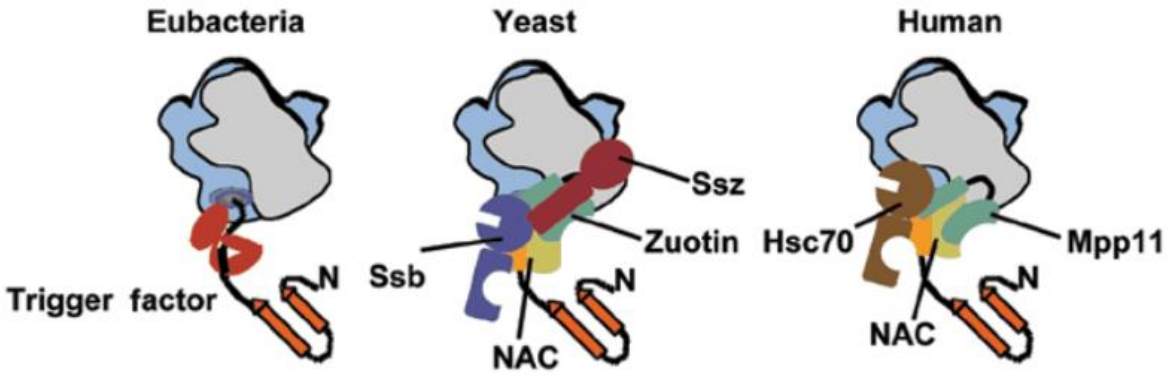


Figure 43: Composition of ribosome-associated chaperones from bacteria to human. The schematic depicts bacterial trigger factor, yeast Zuotin, Ssb and Ssz chaperones, human Mpp11 and Hsc70 chaperones in association with ribosomes and nascent chain associated complex (NAC) with eukaryotic ribosomes. Hsc70 mentioned here is referred to as Hsp70L1 in majority of publications. Image is from (Bukau, 2005).

TF is closely associated with the large ribosomal subunit at the ribosomal exit tunnel (Kramer et al 2002; Ferbitz et al 2004; Merz et al. 2008). This enables it to interact with the most newly synthesizing polypeptide chains thereby assisting in their folding. It has been reported that TF binds ribosomes following translation of at least first 100 amino acids of a polypeptide chain, thus allowing the prior interactions of other ribosome-binding targeting factors and peptide modifying enzymes with the nascent chain (Eisner et al. 2003; Ullers et al. 2003; Bingel-Erlenmeyer et al. 2008). RAC and NAC are two such complexes in eukaryotes might carry out similar functions like TF in prokaryotes. The yeast version of RAC complex is comprised of the Hsp70-like protein Ssz1 and the Hsp70 cochaperone zuotin (Hsp40) (Koplin et al. 2010; Kotani et al. 2009b; Raue et al. 2007; Gautschi et al. 2002; Preissler and Deuerling 2012;

Bukau et al. 2000). In mammals, the J-domain protein MPP11 and the atypical Hsp70 homolog Hsp70L1 forms the mammalian RAC (mRAC) complex (Jaiswal et al. 2011; Otto et al. 2005) ***(Figure 44)***.

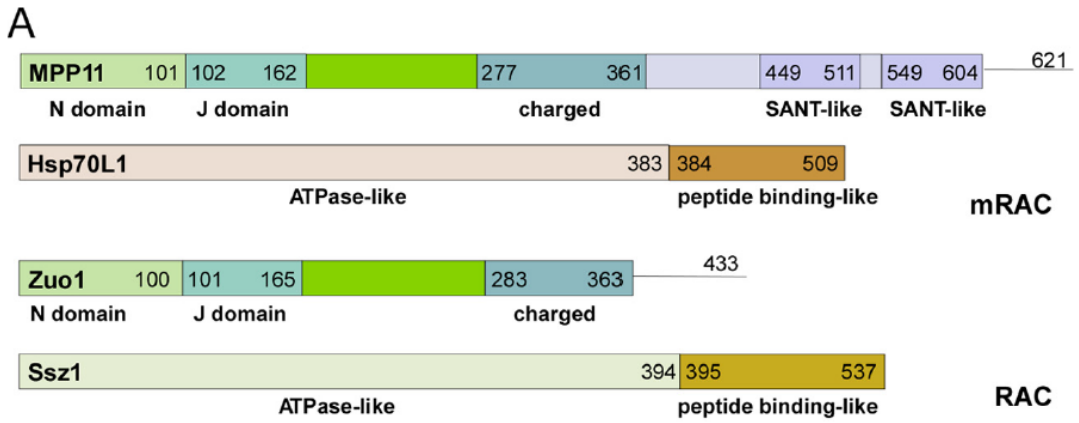


Figure 44: Schematic representation of yeast and human (mRAC) RAC complex. Human MPP11 is an Hsp40 homolog and the only similarity to the yeast apparatus is homology to the unusual N-terminal domain of zuotin (Shoji et al. 1995). Hsp70L1 is a distantly related homolog of yeast Ssz1. The 621 amino acid long C-terminal extension of MPP11 is a two-repeat domain similar to the SANT domain family (Resto et al. 2000; Shoji et al. 1995; Otto et al. 2005).

Despite a low degree of homology on the amino acid level, it was shown by complementation experiments that mRAC is functional in yeast (Otto et al. 2005). So, the cooperation of ribosome-associated chaperones with the translational apparatus is well conserved from yeast to human.

It has been shown that the binding of chaperone Ssb (specialised Hsp70 chaperone in yeast) to the ribosomes is multilayared and involves the co-chaperone RAC and two specific basic regions characterized by positively charged amino acids. The ribosome binding of Ssb is not essential for its functionality and is necessary only when RAC is absent, thereby suggesting a dual mode of Ssb interaction with the nascent chain. The interactions are necessary to position Ssb close to the ribosome exit tunnel for efficient interaction with the nascent polypeptide (Hanebuth et al. 2016).

The second ribosome associated chaperone system is a heterodimer called nascent polypeptide-associated complex (NAC) (Genevaux et al. 2004). Unlike RAC, the structure of NAC is highly conserved from Archaea to human. However, in Archaea, it is a homodimer formed by two α-subunits and in yeast and higher eukaryotes, it is a heterodimer formed by α and β subunits (Preissler and Deuerling 2012). ***(Figure 45)***.

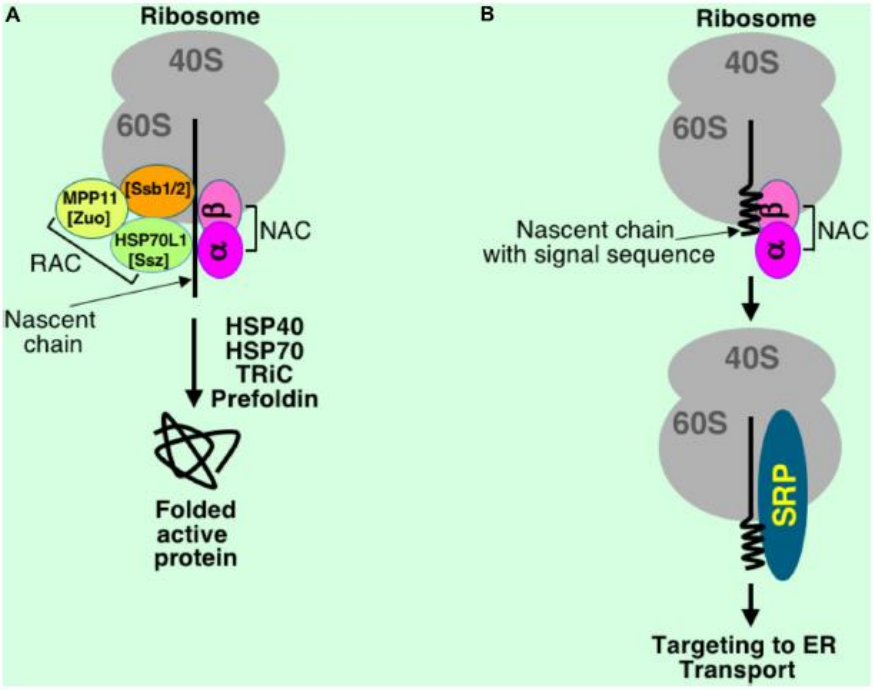


Figure 45: Schematic representation of chaperone interactions with nascent polypeptide of cytosolic proteins (A) and secretory proteins (B) at ribosome exit site. While nascent chains of cytosolic proteins (A) are synthesized in presence of RAC, NAC, Ssb (HSP70) and further folded with assistance of chaperones and chaperonins, the secretory proteins (B) briefly interact with NAC before full exposure of the signal sequence, and when signal sequence is emerged from the ribosome tunnel, SRP binds it leading to temporary elongation arrest and targeting to the ER membrane for further transport through translocon into ER lumen, then to Golgi, and finally outside of the cell. Mammalian proteins are shown, their yeast counterparts are in square brackets. Image is from (Karamyshev and Karamysheva 2018).

6.2.1.2 Chaperones acting downstream of ribosomes

The classical Hsp70 family of chaperones binds to nascent polypeptide chains, but have no direct affinity for the ribosome, unlike ribosome-associated chaperones (Niwa et al. 2012; Calloni et al. 2012; Hartl et al. 2011; Frydman 2001), with the exception of Hsp70L1 which a part of the ribosome-associated RAC complex in mammals is (as mentioned above in ***section 6.1.1***). The basic mechanism of action for this family of chaperones is conserved and it includes DnaK in bacteria and some archaea, Ssa1-4 in yeast and the constitutively expressed heat shock cognate70 (Hsc70) in metazoan and mammalian cells (Reynaud 2010; Bukau et al. 2000). Hsp70 works together with cofactors J protein and nucleotide exchange factors (NEFs). J proteins, which are also sometimes referred to as Hsp40 class of cochaperones, play important role in

attributing functional diversity to the Hsp70 chaperone complex. Depending on the type J protein binding to the Hsp70, it might play role in diverse biological processes like modulating polypeptide folding, degradation and translocation across membranes, as well as protein–protein interactions. The basic function of J protein is to capture substrates for Hsp70 to act and it does so by localising itself to different cellular locations (for example, interaction with membranes or ribosomes) thereby bringing Hsp70 in close proximity to the substrate. Even though J proteins are primarily responsible for the functional diversity of Hsp70 machinery, NEFs play important part as well. Four different NEFs have been identified with no sequence similarity among them. Although they all interact with the ATPase domain of Hsp70, they differ in the mechanism by which destabilization of nucleotide binding is accomplished (Schuermann et al. 2008; Douglas 2016; Polier et al. 2008). It is not clear though how other domains of NEF play role in Hsp70 machinery (Kampinga and Craig 2010). The Hsp70 chaperone system is replaced by the chaperone prefoldin in most species of archaea. Prefoldin is a 200 kDa hexameric complex also termed as Gim complex (GimC) and is made up of two related classes of subunits giving the appearance of a jellyfish, body consisting of a double beta barrel assembly with long α-helical coiled coil structure protruding like tentacles. Prefoldin captures unfolded protein substrate through the tentacles and transfers it to chaperonins in the cytosol (Sahlan et al. 2018). It has been shown that prefoldin participates in chaperone-assisted folding of actin and tubulin in *Saccharomyces cerevisiae* (Hartl 2002; Frydman 2001). However, a prefoldin like complex has also been discovered in mammals and it is essential for maturation and assembly of multisubunit protein complexes. Recent discoveries about the complex are described below in ***section 6.2.2***.

6.2.2 PAQosome (particle for arrangement of quaternary structure)

The R2TP (Rvb1–Rvb2–Tah1–Pih1) complex associates with the PFDL (prefoldin-like) module in mammals to form the R2TP/PFDL complex (***Figure 46***), which has been recently named as the PAQosome because of its functional role in the assembly of quaternary structure. The R2TP complex was first discovered in *Saccharomyces cerevisiae* (Zhao et al. 2005) but it is conserved in eukaryotes (Boulon et al. 2008; Jeronimo et al. 2007). RTP2 consists of the AAA+ ATPases Rvb1 (RuvB-like protein 1) and Rvb2 (RuvB-like protein 2), Pih1 (Protein interacting with Hsp90 1), Tah1 (TPR repeat containing protein associated with Hsp90) in yeast. Human R2TP contains

orthologous proteins, named RUVBL1 (RuvB-like AAA ATPases 1), RUVBL2 (RuvB-like AAA ATPases 2), RPAP3 (RNA Polymerase II Associated Protein 3), and PIH1D1 (PIH1 domain containing 1). PFDL consists of two α subunits and four β subunits and includes two additional components, the RNA polymerase subunit POLR2E (RNA Polymerase II subunit E) and WDR92 (WD Repeat domain 92). ***(Figure 46)***. The PAQosome complex has been suggested to play a role in the assembly of protein complexes related to protein synthesis, cell growth and metabolism, as well as gene expression and genome stability. Both direct binding and indirect binding of PAQosome complex is possible with its substrates; several adaptors have been identified that facilitate the binding of PAQosome complex to its substrates (Houry et al. 2018).

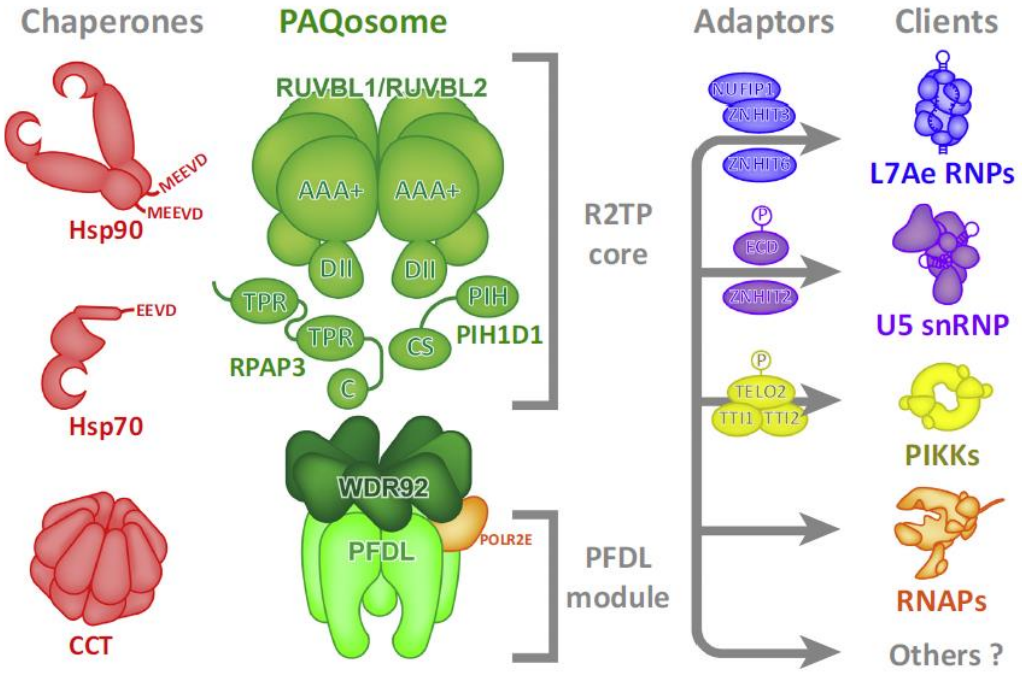


Figure 46: Schematic representation of PAQosome structure, Adaptors and Substrates/Clients. The subunits forming PAQosome are shown in green (described more in text), associated chaperones are shown in red, adaptors are shown to the right and substrates or clients are shown towards far right. Note, however, that not all interactions are well characterised. Image is from (Houry, Bertrand, and Coulombe 2018).

6.2.3 Chaperonins

Chaperonins are a subset of molecular chaperones that act on proteins that are utilize Hsp70 for folding (Kim et al. 2013). They are also referred to as Hsp60s. They have a large oligomeric ring-like structure with central cavity that allows them to fold proteins in an isolated hydrophobic environment protected from non-specific aggregation prone cytosol (Fenton and Horwich 1997). The chaperonins can be structurally classified into two groups: Type I and Type II (Tang et al. 2007; Ansari and Mande 2018; Horwich et al. 2007). Recently, a third group known as Type III chaperonins was reported which are structurally similar to Type II chaperonins but mechanistically and phylogenetically distinct from both the other groups (Techtmann and Robb 2010). Chaperonins are known to exist in almost all prokaryotes and eukaryotes; the only organisms that are shown to lack chaperonins are some parasites like microsporidia and mycoplasma (Glass et al. 2000; Katinka et al. 2001).

6.2.3.1 Type I Chaperonins

This group of chaperonins are found in the prokaryotes and in the mitochondrion and chloroplast of eukaryotic cells. The well-studied GroEL-GroES system is the Type I chaperonin in *Escherichia coli*. Its homologs are Cpn60/Cpn20 in chloroplasts, and mtHsp60/mtHsp10 in mitochondrion (Cheng et al. 1989; Hayer-hartl et al. 2009; Dickson et al. 2000). Type I chaperonins consist of two components: a tetradecameric Hsp60 and a heptameric co-chaperone Hsp10 which acts as a cap on the Hsp60 ring structure. GroEL is the Hsp60 of *E. coli* consisting of two 7-fold symmetric rings which a central cavity in which substrate proteins are caged for folding. GroES is the Hsp10 of the *E. coli,* which acts as the lid of the structure to prevent the exit of the substrate (Jaenicke 1991; Frydman 2001).

6.2.3.2 Type II Chaperonins

Group II chaperonins are found in archaea and eukaryotes. The archaeal type II chaperonin α/β-thermosome was the first structural model of this family of chaperonins (Klumpp et al. 1997). Its eukaryotic homolog is the tailless complex polypeptide-1 (TCP-1) ring complex (TRiC), also known as chaperonin-containing TCP-1 (CCT). This group of chaperonins also consists of two stacked rings but unlike type I chaperonins, they have a built-in lid and hence they do not have an obligate co-chaperone. The built-in lid enables them to close the folding chamber and thus aid in

the folding of substrates. This does not mean that TRiC/CCT carries out its function independently of other chaperones. In fact, TRiC/CCT has been shown to cooperate with Hsp70 in the cotranslational folding of multidomain proteins (Cuéllar et al. 2008; Etchells et al. 2005). TRiC is absolutely required for the folding of many essential proteins (about 10% of proteome), like the cytoskeletal proteins, actin and tubulin, several proteins with ß-propellers/WD40 repeats as well as cell cycle regulators such as CDC20, CDH1 (Ho et al. 2002; Heng et al. 2001; Zhao and Fang 2005; Dekker et al. 2008; Yam et al. 2008). Additionally, TRiC/CCT facilitates the folding of the C-terminal WD40-repeat domain of TAF5 and its subsequent binding to TAF6-TAF9 heterodimer, thereby allowing the efficient formation of holo-TFIID complex (Antonova et al. 2018) . It has been reported that TRiC substrate selection is not solely based on intrinsic determinants in vivo, but specificity is dictated by factors present during protein biogenesis. Multidomain proteins, which are prone to aggregation, have been reported to be one of the substrates for TRiC by bioinformatic analysis. In fact two school of thoughts exist concerning the substrate specificity of TRiC. One of them believes the TRiC interacts with a broader range of substrates, around 10% of cytosolic proteins (Thulasiraman et al. 1999; Yam et al. 2008). The other group believes that TRiC has a highly restricted set of substrates, not more than 1% of the cytosolic proteins (Kramer et al 2002; Preissler and Deuerling 2012; Willison 2018). It would be interesting to unravel whether TRiC is a general cytosolic chaperone.

6.2.4 Chaperone dysregulation under disease condition

Various stress conditions like high temperature, oxidative stress or heavy metals causes protein to misfold or aggregate. These conditions shift the conformational equilibrium towards more aggregation-prone states where aberrant protein-protein interactions are formed between exposed hydrophobic regions of two proteins (Weids et al. 2016). Upregulation of chaperones under these conditions help the cell to prevent aggregation. Age-related neurodegenerative diseases involve protein aggregation (Chiti and Dobson 2006). Aggregation could be of two different types: amyloid and amorphous. Amyloid aggregation, which is associated with many neurodegenerative diseases like Alzheimer’s, Parkinson’s and Huntington’s disease, is a highly structured, insoluble, fibrillar deposit, usually consisting of many repeats of the same protein (Greenwald and Riek 2010; Tipping et al. 2015). On the other hand,

amorphous aggregation is the unordered aggregation of proteins where each protein is not generally associated with disease (Weids et al. 2016).

Cells respond to the increase in the burden of misfolded or unfolded proteins by the activation of molecular chaperones through the cytosolic stress pathway, or heat shock response. Stress-activated chaperones play essential role as modulators of protein homeostasis and promote productive folding or degradation of misfolded proteins. Though actively involved in the task of refolding, Hsp70 is ultimately unable to refold disease proteins thereby affecting protein homeostasis under diseased conditions. This could be due the fact that the cellular system is overburdened by the increased amounts of misfolded proteins (Barral et al. 2004; Gidalevitz et al. 2006). Slowing the ageing process might prevent disease onset as shown by many studies in *C. elegans* models of Huntington's and Alzheimer's diseases (Parker et al. 2005; Morley et al. 2002). Hsp70 is sometimes assisted by its co-chaperone Hsp40 to prevent protein aggregation under diseased condition. For example, mutant huntingtin accumulates in benign amorphous aggregates by the action of Hsp70/Hsp40 chaperones rather than toxic aggregates (Behrends et al. 2006; Wacker et al. 2004; Muchowski et al. 2000). Several studies also pointed towards the fact that the chaperonin TRiC/CCT in its fully assembled state modulates Huntington aggregates, thereby exerting neuroprotective effects in the cell (Kramer et al 2002; Young et al. 2003; Haslberger et al. 2010). A very new approach to treat protein misfolding disease is "chaperone therapy". Small molecule induction of heat shock proteins has been experimentally tried to resolve abnormally accumulated proteins. Apart from this, low molecular competitive inhibitors (chemical chaperones) as well as non-competitive chaperones without inhibitory bioactivity have also been developed (Suzuki 2014).

7. Overview of techniques used to study co-translational assembly

7.1 Biochemical Approach

7.1.1 Indirect Approach

The earliest reports on co-translational assembly were mainly based on indirect approaches which enabled the authors to detect the fully translated functional version of the complex immediately after translation. For example, beta-galactosidase activity was observed in the polysome fractions (Kiho and Rich 1964), the protein tenascin was only observed in its hexameric form by pulse chase labelling followed by protein detection by non-denaturing polyacrylamide gel electrophoresis. No monomeric, dimeric or trimeric precursors were observed (Redick et al 1995). Co-translational assembly of a protein complex into any cellular structure could be followed by immunoprecipitating the nascent protein from the specific cellular structure by using an antibody specific to its N-terminal region, as observed for D1 protein insertion into membrane associated photosystem II or myosin heavy chain insertion into cytoskeleton (Zhang et al. 1999). These molecular mass-based indirect approaches of studying co-translational assembly is mainly useful for homomeric protein complexes, where the only difference between monomer and multimer is molecular mass and functional activity of the structure.

7.1.2 Direct Approach

A commonly used method to study RNA-protein interactions is RNA Imuunoprecipitation (RIP) followed by microarray (RIP-Chip) or qPCR (RIP-qPCR) (Keene et al. 2006). Though this method is mostly used to study direct interactions between RNA-binding proteins and RNA, co-translational assembly can also be studied with addition of proper negative controls. For co-translationally assembling protein partners, the interaction between a protein and its interaction partner's mRNA is indirect and occurs through the interaction partner (***Figure 17***). In that case, though RIP-Chip/ RIP-qPCR would enable us to obtain information about the same, a negative control with puromycin treatment or a construct without a start codon would enable us to rule out the possibility of direct RNA-protein interaction (Duncan and Mata 2011; Halbach et al. 2009; Kassem et al. 2017; Panasenko et al. 2019; Kamenova,

Mukherjee et al. 2019). RIP experiments can be carried out from cell extracts (Kamenova, Mukherjee et al. 2019) or polysome fractions (Kassem et al. 2017).

Another approach to study co-translational assembly is selective ribosome profiling (SeRP) (Becker et al. 2013). This method is based on ribosome profiling developed by Ingolia *et al.*, 2012 where mRNA fragments protected by translating ribosomes are sequenced. SeRP consists of immunoprecipitation of the factor of interest associated with ribosome-nascent chain complex followed by the sequencing of mRNA fragments by ribosome profiling (Shieh et al. 2015; Shiber et al. 2018). For instance, if protein A is co-translationally associating with its ribosome-associated nascent interacting partner B, immunoprecipitating protein A would pull down its ribosome-associated partner B and therefore, the mRNA region encoding protein B covered by the translating ribosome. Sequencing of this mRNA fragment covered by the ribosome would throw light on the region of protein B co-translationally associating with protein A as the mRNA region sequenced would be the region encoding protein sequence immediately downstream of the co-translational interacting region. (***as shown Figure 47***).

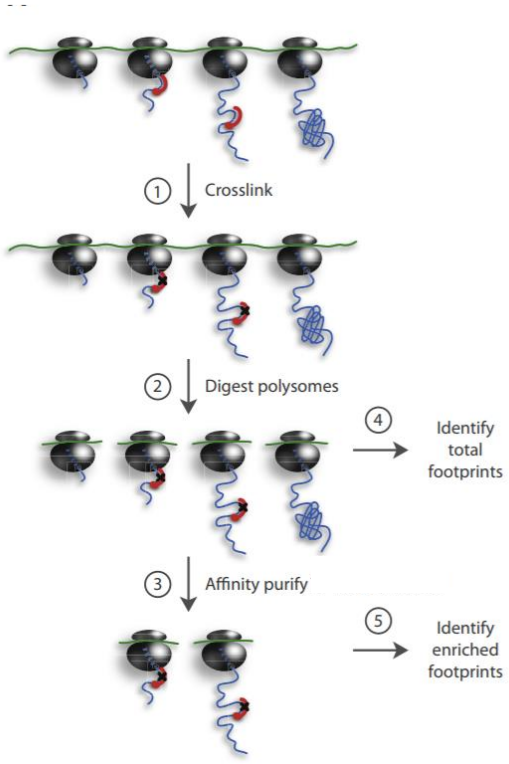


Figure 47: Major steps of selective ribosome profiling (SeRP). A protein (in red) is associated co-translationally with its ribosome-associated translating interacting partner (in blue). Cells are lysed and crosslinked (Step 1), polysomes digested with MNase and yield footprint-containing monosomes (Step 2), digested monosomes are isolated by sucrose cushion and the protein of interest in affinity purified with an antibody against it (Step 3) and the mRNA footprint fragments derived from both total and affinity purified monosomes are cloned into cDNA libraries for deep sequencing (Step 4 and 5 respectively). Image adapted from (Becker et al. 2013).

7.2 Imaging approaches

With the advent of powerful single molecule imaging techniques, it has been possible to study co-translational assembly. The proximity of two mRNA molecules encoding two different genes can be studied by dual colour single molecule RNA FISH (Panasenko et al. 2019; Kamenova, Mukherjee et al. 2019) and the proximity of a protein and its interacting partner's mRNA by immunofluorescence combined with single molecule RNA FISH (Kamenova, Mukherjee et al. 2019). It would be also interesting to follow co-translational assembly by endogenous labelling of mRNAs and carrying out live imaging (Haimovich et al. 2017).

THESIS OBJECTIVES

Thesis Objectives

Multisubunit protein complexes carrying out various biological processes form an integral part of our cells. Despite numerous reports on their structure and function, very few studies focussed on their order and mechanism of assembly so far. In this respect, two models of protein complex assembly have been put forward, post-translational and co-translational. For many years, protein complexes were known to form post-translationally by random collision in the cytoplasm. However, the extreme crowded environment of the cell cytoplasm makes it impossible for the protein subunits to find their partners. Recently, co-translational formation of protein complexes has been put forward in bacteria and yeast to circumvent this limitation.

Reports in mammals suggest the presence of subassemblies of multisubunit protein complexes in the cytoplasm including transcription complexes. The general transcription factor TFIID nucleates the assembly of other transcription factors in most of the expressed protein coding gene promoters and loss of TFIID was shown to be embryonic lethal. TFIID is composed of TBP (TATA-binding protein) and 13 TAFs (TBP associated factors). Majority of the TAFs contain a common structural motif called the histone fold domain (HFD), which facilitates pairwise interaction between specific TAFs, specifically, TAF10-TAF8, TAF3-TAF10, TAF6-TAF9, TAF4-TAF12, TAF11-TAF13 heterodimers. It has been reported previously that the assembly of TFIID takes place in a stepwise manner before it embarks on its function in the nucleus. Several submodules of TFIID have been shown to assemble in the cytoplasm before it forms the holo-TFIID complex. Stable heterotrimers of TAF5-TAF6-TAF9, TAF2-TAF8-TAF10 and TAF7-TAF11-TAF13 have been shown to form in the cytoplasm. In this context, the main aim of my thesis was to study the mechanism of assembly of TFIID submodules in the cytoplasm and specifically answer the following questions:

1. Do TFIID histone-fold domain pairs assemble co-translationally?
2. What drives the co-translational assembly of TFIID HFD pairs?
3. Do non-HFD pairs also assemble co-translationally?
4. Do other multisubunit complexes also assemble co-translationally?
5. Study the co-localisation of co-translationally assembling proteins and RNAs?
6. Study the role of chaperones in co-translational assembly?

RESULTS

Results

1. Co-translational assembly of mammalian nuclear multisubunit complexes (Kamenova[#], Mukherjee[#], et al., Nature Communications, 2019, [#]equal first authors, names appear in alphabetical order)

Authors' contributions:

Pooja Mukherjee: Designed the study, carried out molecular biology and imaging experiments (Figures 2a-d, 3a-b, 4e, 5a-d, 6a-d, 7c-e, Supplementary Figures 2a-d, 3a-b, 4a,c, 5a-d, 6a-c, 7b-c), interpreted and analysed data, wrote manuscript.

Ivanka Kamenova: Designed the study, carried out molecular biology and imaging experiments (Figures 1b-e, 7a,b,f, 8a-c, Supplementary Figures 1, 7a), interpreted and analysed data, wrote the manuscript.

Sascha Conic: Helped in the confocal imaging and preparing imaging figure panels.

Florian Mueller: Analysed imaging experiment data, IF-smiFISH and smiFISH.

Farrah El-Saafin, Paul Bardot: Carried out mouse knockout experiments (Figure 4a-d).

Stéphane D. Vincent: Carried out mouse knockout experiments (Figure 4a-d) and helped in preparing figure panel with R.

Jean-Marie Garnier: Performed all the cloning experiments.

Doulaye Dembele: Analysed the microarray data.

Simona Capponi: Generated a stable cell line.

H.T.Marc Timmers: Generated a stable cell line and wrote the manuscript.

László Tora: Conceived and designed the study, wrote the manuscript.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-09749-y OPEN

Co-translational assembly of mammalian nuclear multisubunit complexes

Ivanka Kamenova[1,2,3,4,7], Pooja Mukherjee[1,2,3,4,7], Sascha Conic[1,2,3,4], Florian Mueller[5], Farrah El-Saafin[1,2,3,4], Paul Bardot[1,2,3,4], Jean-Marie Garnier[1,2,3,4], Doulaye Dembele[1,2,3,4], Simona Capponi[6], H.T.Marc Timmers[6], Stéphane D. Vincent[1,2,3,4] & László Tora[1,2,3,4]

Cells dedicate significant energy to build proteins often organized in multiprotein assemblies with tightly regulated stoichiometries. As genes encoding subunits assembling in a multi-subunit complex are dispersed in the genome of eukaryotes, it is unclear how these protein complexes assemble. Here, we show that mammalian nuclear transcription complexes (TFIID, TREX-2 and SAGA) composed of a large number of subunits, but lacking precise architectural details are built co-translationally. We demonstrate that dimerization domains and their positions in the interacting subunits determine the co-translational assembly pathway (simultaneous or sequential). The lack of co-translational interaction can lead to degradation of the partner protein. Thus, protein synthesis and complex assembly are linked in building mammalian multisubunit complexes, suggesting that co-translational assembly is a general principle in mammalian cells to avoid non-specific interactions and protein aggregation. These findings will also advance structural biology by defining endogenous co-translational building blocks in the architecture of multisubunit complexes.

[1] Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France. [2] Centre National de la Recherche Scientifique (UMR7104), 67404 Illkirch, France. [3] Institut National de la Santé et de la Recherche Médicale (U1258), 67404 Illkirch, France. [4] Université de Strasbourg, Illkirch 67404, France. [5] Computational Imaging & Modeling Unit, Institut Pasteur, Département Biologie Cellulaire et Infections, 25-28 rue du Docteur Roux, 75015 Paris, France. [6] German Cancer Consortium (DKTK) partner site Freiburg, German Cancer Research, Center (DKFZ) and Department of Urology, Medical Center-University of Freiburg, 79106 Freiburg, Germany. [7] These authors contributed equally: Ivanka Kamenova, Pooja Mukherjee. Correspondence and requests for materials should be addressed to L.T. (email: laszlo@igbmc.fr)

Often proteins do not act alone, instead they function as components of large multisubunit complexes in a cell. To better understand cellular functions, investigating the precise mechanism that guide the formation of these multisubunit assemblies is of key importance. A cell uses hundreds of different protein complexes that vary with respect to their complexity. Some complexes require the association of multiple copies of the same subunit, while others are constituted of many different subunits. The latter group includes many transcription regulatory and chromatin remodelling complexes (see below). In order to achieve the efficient formation of protein complexes in eukaryotes, the genes coding for all the subunits (dispersed in the eukaryotic genome) have to be transcribed in the nucleus, their corresponding mRNAs transported to the cytoplasm, translated into proteins, and the formation of correct interactions among the subunits must be orchestrated. A polysome is a cluster of ribosomes acting on a single mRNA to translate its information into polypeptides. Appropriate translation-based mechanisms may exist in the cell to regulate the interactions between specific subunits in order to avoid incorrect non-specific interactions or subunit aggregations in the absence of the correct partner. Currently, it is not well understood how functional subunit interactions are regulated in eukaryotic cells. Protein complex formation is often studied in vitro using purified subunits, assuming that individually translated subunits assemble stochastically by diffusion, and thus favouring the idea that these multisubunit complexes assemble post-translationally[1]. However, in the crowded environment of an eukaryotic cell such simple diffusion-dependent models may not work, as subunits may engage in non-specific interactions or form aggregates. Recent studies in bacteria demonstrated that co-translational building of a functional protein dimer is more efficient than the post-translational assembly of its individual subunits[2,3], and also in yeast co-translation has been shown to be an efficient assembly pathway to assemble multiprotein complexes[4–8]. Consequently, two co-translational models have been put forward: (i) the simultaneous model which suggests that two polysomes in close physical proximity synthesise subunits, which interact while being translated and (ii) the sequential model implies that a mature fully translated subunit interacts co-translationally with its polysome-bound nascent interaction partner[9].

One of the key regulatory steps in the expression of mRNAs is transcription initiation. Co-activators act together to establish a chromatin structure favourable for transcription by facilitating the formation of the preinitiation complex (PIC). PIC is comprised of RNA polymerase II (Pol II) and general transcription factors (GTFs). Many GTFs and co-activators are multisubunit complexes, in which individual subunits are organised into several distinct modules carrying out specific functions. In mammalian cells the TFIID GTF nucleates the assembly of the Pol II preinitiation complex on all protein-coding gene promoters [refs [10,11] and references therein]. Metazoan TFIID is composed of the TATA-binding protein (TBP) and 13 TBP-associated factors (TAFs) (Fig. 1a). SAGA (Spt Ada Gcn5 Acetyltrasferase) is a multisubunit transcriptional coactivator complex, composed of 19 subunits (including a subset of TAFs), required for the transcription of all active genes in yeast[12]. Moreover, the mammalian Transcription and mRNA Export 2 complex (TREX-2) is composed of five subunits, including the subunit ENY2, which is shared with the SAGA complex[13].

The majority of TAFs dimerise via their histone-fold domains (HFDs), which are structurally homologous to histone pairs. In TFIID, TAFs form five HF pairs (TAF4-12, TAF6-9, TAF8-10, TAF3-10 and TAF11-13) [ref. [10] and references therein] (Fig. 1a). Importantly, individual HFD-containing TAFs cannot be expressed in a soluble form in bacteria. However, HFD-containing TAFs become soluble when co-expressed with their corresponding specific interaction partner[14], suggesting that individual HFD-containing TAFs aggregate without their specific partners.

To test how mammalian cells can avoid the aggregation of individual subunits following translation and whether co-translational interactions guide the assembly of transcription complexes, in this study, we investigate pairwise assembly of TFIID subunits between TAF8 and TAF10, TAF6 and TAF9 and TAF1 and TBP in polysome-containing mammalian cell extracts. By using a large series of complementary experiments, we show that TAF8-TAF10 and TAF1-TBP assemble co-translationally according to the sequential assembly pathway, while TAF6-TAF9 assembles co-translationally according to the simultaneous model. We also demonstrate that the ENY2 subunit assembles co-translationally with its interaction partner, GANP, in TREX-2, and with ATXN7L3 in the deubiquitination (DUB) module of SAGA. Furthermore, our experiments show that the interaction domain (ID) and the position of the ID in the given subunit solely drives the co-translational assembly in these complexes. Thus, our results uncover mechanistic principles in the understanding of co-translational control of protein complex formation in mammalian cells.

Results

TAF10 and TAF8 assemble co-translationally. To test whether HFD-containing TAFs assemble co-translationally, we used a monoclonal antibody against the N-terminus of the HFD-containing TAF10 to immunoprecipitate (IP) endogenous TAF10 from human HeLa cell cytosolic polysome extracts (Fig. 1b). Protein–protein interactions between nascent proteins still associated with translating ribosomes would be revealed by enrichment of mRNAs coding for the interacting partners in the IPs. Global microarray analysis of mRNAs precipitated by the anti-TAF10 RNA IPs (RIPs) revealed enrichment of *TAF8* mRNA, suggesting that the well-characterised TAF8-10 HFD dimer[15] forms co-translationally (Fig. 1c). Anti-TAF10 RIP of cytosolic polysome extracts coupled to RT-qPCR validation confirmed our microarray results and showed strong enrichment of the *TAF8* mRNA (Fig. 1d). The absence of significant *TAF10* mRNA signal in the microarray experiments was due to poor quality and the high GC-content of the *TAF10* probe sets present on the commercial microarray. Nevertheless, RT-qPCR validation also revealed the presence of *TAF10* mRNA in the nascent anti-TAF10 RIP. Importantly, cycloheximide, which freezes translating ribosomes on the mRNA[16], stabilised the TAF10-TAF8-*TAF8* mRNA interactions, while puromycin, which causes release of nascent peptides from ribosomes[17], resulted in the loss of co-purified mRNA. Endogenous anti-TAF10 RIP-RT-qPCR from polysome extracts prepared from mouse embryonic stem cells (mESCs) gave nearly identical results, which emphasises the generality of the co-translational pathway for assembly of the mammalian TAF8-TAF10 heterodimer (Fig. 1e; Supplementary Fig. 1). Quantification of the *TAF8* mRNA in the anti-TAF10 RIP normalised to the protein IP efficiency indicated that the enrichment was between 7 and 25%, depending on the cell line and the antibody used. In contrast, to *TAF8*, mRNAs encoding other potential TAF10 dimerization partners, TAF3 and SUPT7L[18], were not enriched in the RT-qPCR validation experiments, in good agreement with the microarray analysis and indicating the specificity of the co-translational assembly of the TA8-TAF10 heterodimer (Fig. 1d, e). Together these results indicate that TAF10 protein is associated with ribosomes which are actively translating *TAF8* mRNA via the nascent TAF8 protein.

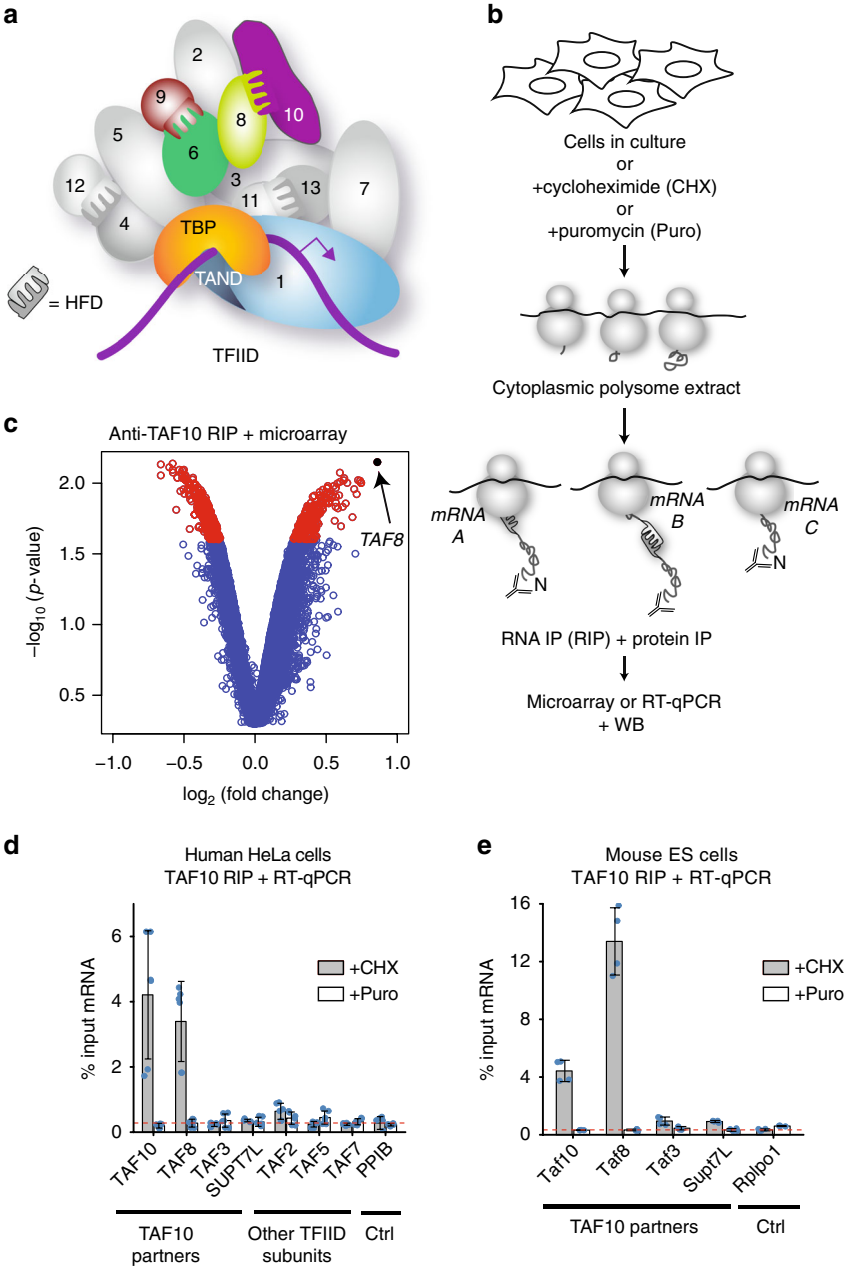


Fig. 1 Co-translational assembly of mammalian TFIID. **a** TFIID bound to promoter DNA. TFIID is composed of TBP and 13 TAFs (indicated by numbers). Subunits analysed in this study are highlighted in colour. The histone fold domain (HFD) interactions and the TBP-TAF1 TAND domain interaction are highlighted. **b** Schematic representation of polysome RIP assay. **c** Endogenous TAF10 was immunoprecipitated from HeLa polysome-containing extract using an antibody targeting the N-terminus of the protein. The enrichment of the precipitated RNAs was assessed globally by microarray. Volcano plot depicting microarray results as log2 of the fold change of IP over a mock IP. A *p*-value cut-off ≤ 0.025 was applied and corresponding transcripts are in red. *TAF8* transcript is highlighted in black. **d**, **e** RIP-qPCR validation of the microarray results in HeLa (**d**) and mouse ES (**e**) cells. Error bars are ±SD from three (HeLa) or two (mESC) biological replicates and two technical replicates (represented by blue dots). *Ctrl* = negative control mRNA. *PPIB* and *Rplp0* were used as unrelated control mRNAs. Source data provided as a Source Data File

HFD drives the co-translational assembly of TAF10-TAF8. The fact that TAF8 has its dimerization HFD at an N-terminal position, and that the TAF10 HFD is at the very C-terminus of the protein, allows the direct testing of the sequential assembly model, as TAF8 and TAF10 may be expected to only heterodimerise if the TAF10 protein is fully synthesised and freed from the ribosome. To examine the two assembly models (see Introduction) and to distinguish between the nascent and mature forms of the TAF8 and TAF10 proteins, we added FLAG-, or HA-tags to either N- (to carry out nascent IPs) or C termini (to carry out mature IPs) of these proteins, respectively. Importantly, exogenous co-expression of N-terminally tagged TAF8 and TAF10 in HeLa cells followed by nascent anti-HA-TAF10 RIP from cytosolic polysome extracts recapitulated the findings obtained with endogenous proteins (Fig. 2a). In contrast, nascent anti-FLAG-TAF8 RIP resulted in high enrichment of its own encoding mRNA, but not that of *TAF10* (Fig. 2b). Immunoprecipitation of mature TAF10-HA protein resulted in *TAF8* mRNA,

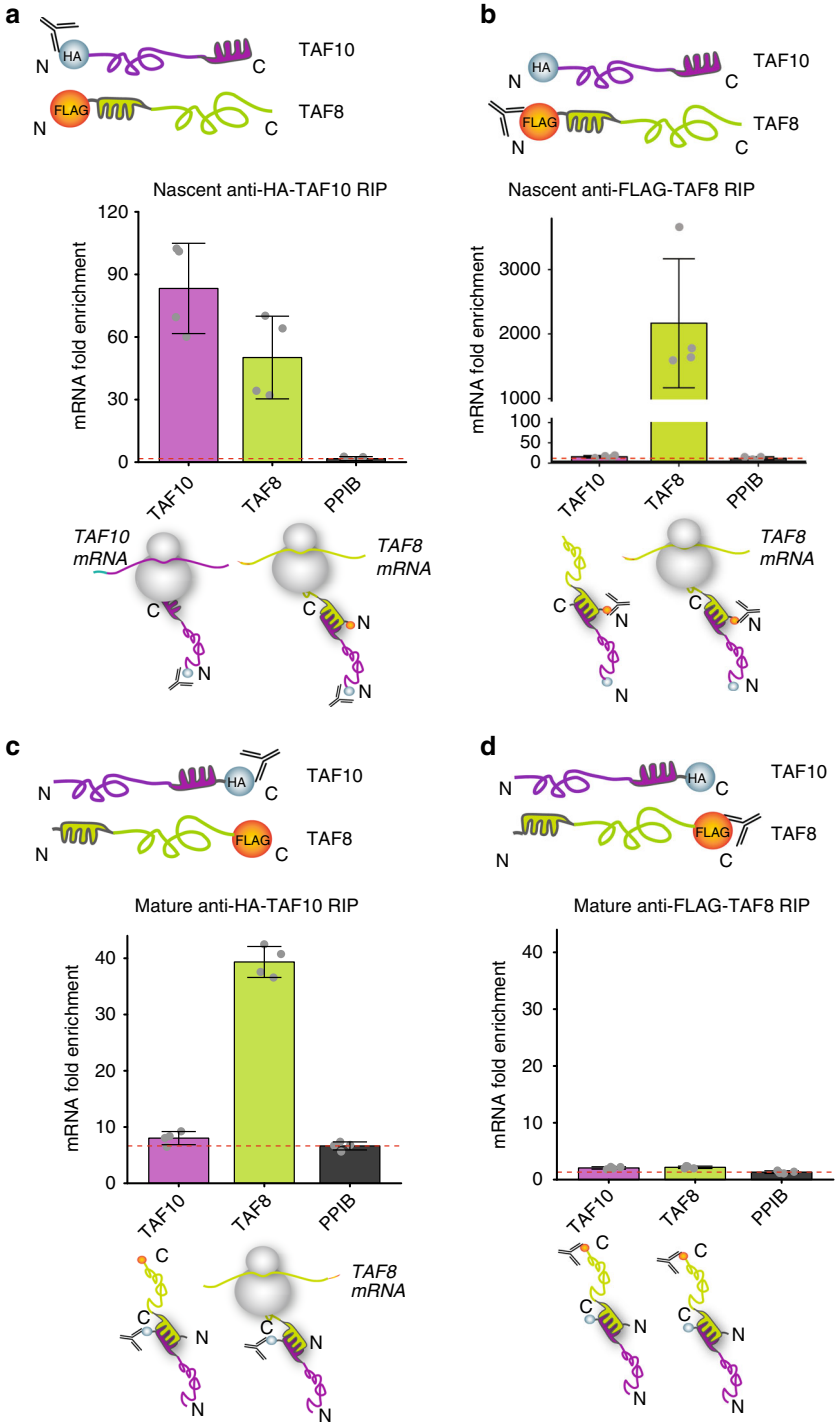


Fig. 2 Sequential assembly of TAF10 and TAF8. N-terminal (**a**) and C-terminal (**c**) anti-HA RIP-qPCR of HeLa cell polysome extracts, co-transfected with the corresponding TAF8 and TAF10 expression plasmids (as indicated). N-terminal (**b**) and C-terminal (**d**) anti-FLAG RIP-qPCR of HeLa cell polysome extracts, co-transfected with the indicated expression plasmids. *PPIB* (**a**–**d**) was used as negative control mRNA. In panels (**a**–**d**) mRNA fold enrichment is expressed as fold change with respect to the mock IP calculated by the formula ΔΔCp [IP/mock] and error bars represent ±SD from two biological and two technical replicates (represented by grey dots). Source data provided as a Source Data File

but not *TAF10* mRNA enrichment (Fig. 2c), supporting the sequential co-assembly model of mature TAF10 interacting with nascent TAF8 exiting from ribosomes translating TAF8 mRNA. In addition, the mature TAF8-FLAG protein did not bring down any of the tested mRNAs (Fig. 2d). In all cases, protein partners were co-immunoprecipitated successfully (Supplementary Fig. 2). Taken together, these results suggest that mature TAF10 binds to the polysome-bound nascent TAF8 protein, and that the respective N- (in TAF8) and C-terminal (in TAF10) HFDs are driving co-translational dimerization.

To test whether the observed co-translational TAF8-TAF10 assembly is specific to the dimerization of their HFDs, we

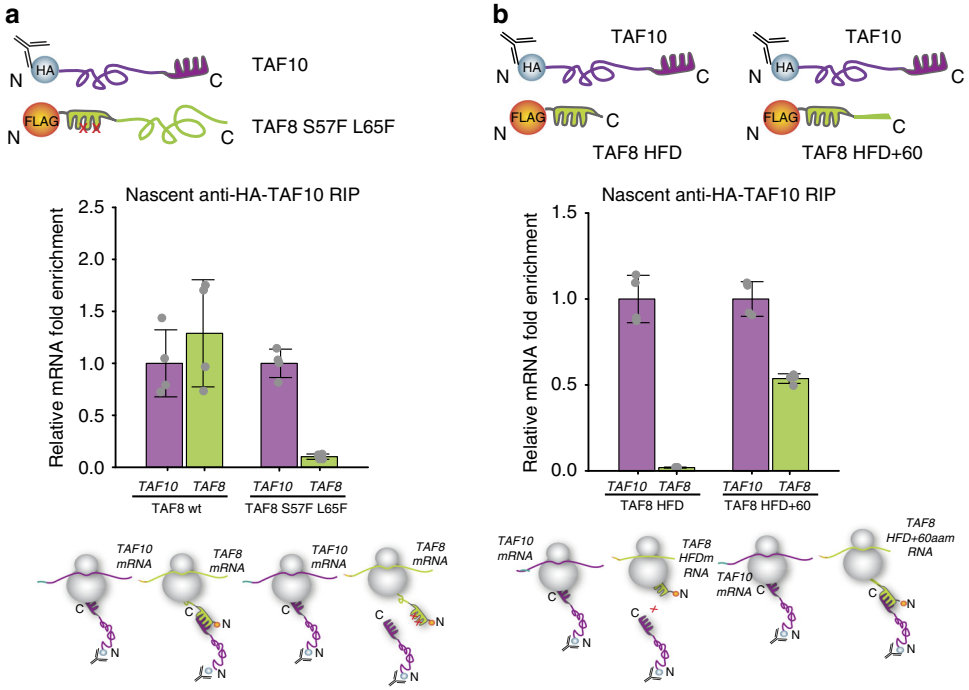


Fig. 3 Protein-protein ID drives co-translational assembly of TAF10 and TAF8. **a** anti-HA-TAF10 RIP-qPCR of HeLa cells co-transfected with HA-TAF10 and either wild type FLAG-TAF8 (TAF8 wt) or mutant FLAG-TAF8 (TAF8 S57F L65F). **b** anti-HA-TAF10 RIP-qPCR of HeLa cells co-transfected with HA-TAF10 and either a minimal TAF8 HFD or TAF8 HFD extended with 60 amino acids (TAF8 HFD + 60aa). In panels (**a**, **b**) relative mRNA fold enrichment is expressed as fold change of *TAF8* mRNA (with respect to the mock IP calculated by the formula ΔΔCp [IP/mock]) relative to *TAF10* mRNA and error bars represent ±SD from two biological and two technical replicates (represented by grey dots). Source data provided as a Source Data File

engineered a mutation disrupting the dimerization ability of the TAF8 HFD (see Methods). Anti-TAF10 RIP from cells co-transfected with *TAF10* cDNA and mutant HFD expressing *TAF8* cDNA (mtTAF8) resulted in a nearly complete loss of the co-precipitated *TAF8* mRNA and TAF8 protein, as compared with the wild-type controls (Fig. 3a and Supplementary Fig. 3a), indicating that the dimerization of TAF8 and TAF10 through their HFDs is crucial for co-translational assembly.

Next, we tested whether the full exposure of the nascent HF interaction domain at the ribosomal exit tunnel would be necessary for co-translational assembly. The ribosome exit tunnel can accommodate up to 60 amino acids [ref. [19] and references therein]. Thus, we constructed two truncated versions of TAF8: one encoding only the TAF8 HFD that would be partially buried in the ribosome exit tunnel during translation, and a second encoding the TAF8 HFD and an additional 60 amino acids of TAF8 (TAF8 HFD + 60) that would allow the appearance of the nascent TAF8 HFD from the ribosomal tunnel. Next a TAF10 expressing plasmid was co-transfected either with TAF8 HFD, or with TAF8 HFD + 60 expressing plasmids and anti-TAF10 RIPs were carried out. Importantly, our results show that the *TAF8 HFD* mRNA is not enriched in the anti-TAF10 RIP, indicating that the minimal TAF8 HFD protein is released immediately from translating polysomes without co-translational binding to TAF10 protein. On the other hand, the *TAF8 HFD + 60* mRNA was enriched in the anti-TAF10 RIP demonstrating that the additional 60 amino acids in the longer TAF8 HFD + 60 protein kept the nascent protein anchored in polysomes allowing for co-translational interaction with TAF10 (Fig. 3b and Supplementary Fig. 3b). Together, our results indicate that TAF8-TAF10 co-translational assembly is driven by dimerization with nascent TAF8 protein upon emergence of its entire HFD from actively translating polysomes. Consequently, these results together demonstrate the sequential co-translational assembly pathway where the fully synthesised TAF10 interacts uni-directionally with the nascent TAF8 polypeptide.

TAF8 is prone to degradation in the absence of TAF10. In the sequential assembly pathway, if nascent chains of a subunit cannot co-translationally interact with its partner, it may become prone to misfolding and degradation by the proteasome, but the fully translated partner should stay stable. To test this hypothesis, we used mouse embryonic stem cells (ESCs) in which either the endogenous *Taf10*, or *Taf8* genes can be conditionally knocked out[20,21]. By using these mouse ESCs we observed that the deletion of *Taf10* not only ablated *Taf10* mRNA and TAF10 protein levels, but significantly reduced both *Taf8* mRNA and TAF8 protein expression (Fig. 4a, c). These results were also confirmed in *Taf10* KO mouse embryos[20]. In contrast, the deletion of *Taf8*, decreased only its own mRNA and protein levels, without affecting the *Taf10* mRNA expression and TAF10 protein levels (Fig. 4b, d). Furthermore, in both KO mESCs other tested TFIID subunits remained unchanged[20].

Next we tested whether TAF10 re-expression would rescue TAF8 from degradation. To this end we used our *Taf10*$^{-/-}$*:R* mouse F9 cells, where the endogenous *Taf10* alleles are inactivated and the cells are viable due to the doxycyclin (Dox) inducible expression of the human TAF10 protein[22]. In this system cells were grown for 5 days without Dox. As a result TAF10 was completely depleted and consequently endogenous TAF8 expression was also abolished (Fig. 4e), in agreement with our above mESC results. Importantly, however, when after 5 days Dox was re-added to the cells for 1 or 2 days, the neosynthesised TAF10 expression re-stabilised the expression of endogenous TAF8 as both TAF10 and TAF8 proteins could again be detected by western blot analysis (Fig. 4e).

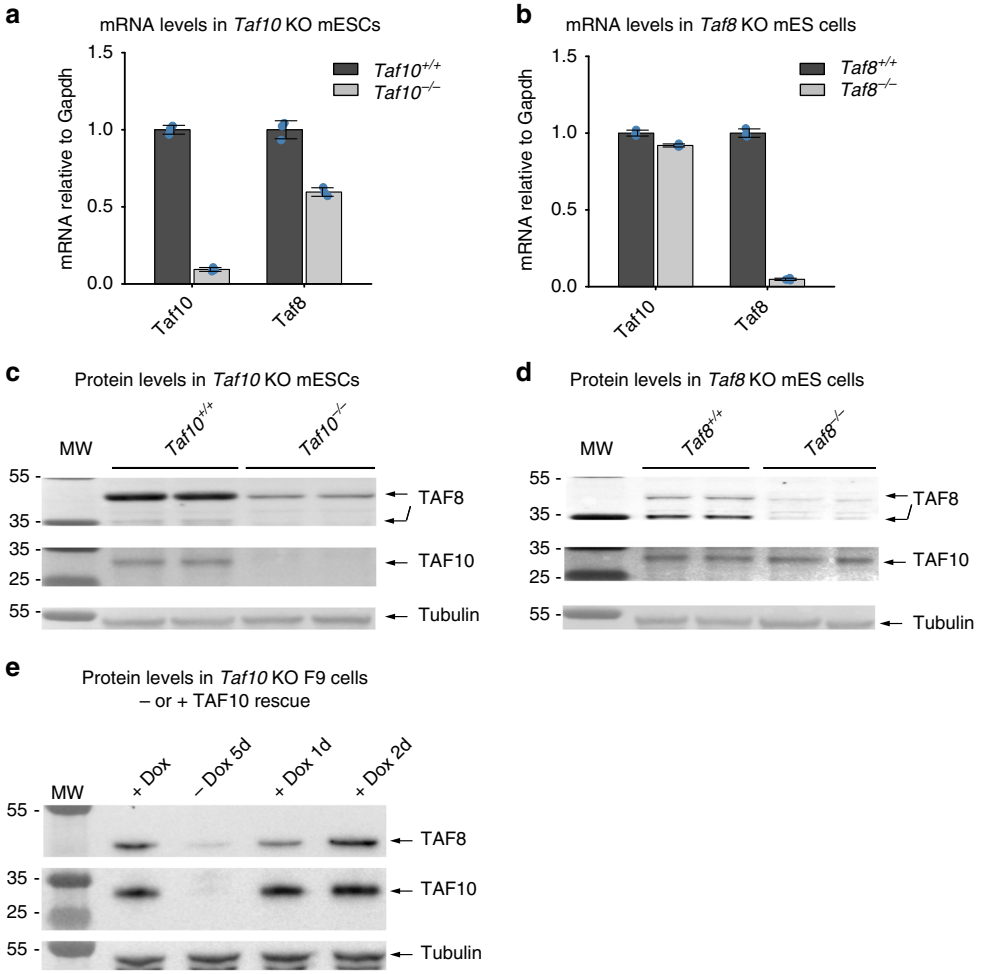


Fig. 4 TAF10 depletion causes degradation of TAF8 in KO mESCs. **a**, **b** RT-qPCR of TAF10 (**a**) and TAF8 (**b**) depleted mESCs. **c**, **d** Western blot analyses from TAF10 (**c**) and TAF8 (**d**) depleted mESCs whole-cell extract using anti-TAF8 and anti-TAF10 antibodies. **e** Western blot analysis of whole cell extracts prepared from the mouse *Taf10*$^{-/-}$ F9 cells with or without Doxycycline (Dox) for the indicated number of days (**d**) using anti-TAF8 and anti-TAF10 antibodies. In panels (**a**, **b**), mRNA levels were normalised to *Gapdh* mRNA and relative enrichment was calculated using the ΔΔCp method and error bars represent ±SD from three technical replicates (represented by blue dots). In panels (**c**–**e**), molecular weight (MW) markers are shown in kDa and an anti-Tubulin was used as a loading control. In panels (**c**–**d**), the two protein isoforms of mTAF8 are indicated. Source data provided as a Source Data File

Together, these results further indicate that TAF10 interacts co-translationally with nascent TAF8 and when TAF10 is not present both TAF8 protein and mRNA could be prone to degradation. Thus, the nascent TAF8 HFD, in the absence of its interaction partner TAF10, may serve as a signal for both protein and mRNA degradation, while TAF10 is stable in the absence of TAF8. However, the reduction of *TAF8* mRNA in the absence of TAF10 protein due to primary transcriptional response cannot be ruled out.

TAF10 protein co-localises with *TAF8* mRNA in the cytoplasm. To visualise the co-localisation of TAF10 protein with *TAF8* mRNA in the cytoplasm, we set out to detect TAF10 protein and *TAF8* mRNA in the cytoplasm of fixed human HeLa cells. To this end we combined protein detection by immunofluorescence (IF) with RNA detection by single molecule inexpensive FISH (smiFISH)[23]. Co-localization of protein and mRNA was then observed by confocal microscopy and quantified. Surprisingly, we observed a large difference between the number of total (nuclear and cytoplasmic) endogenous *TAF8* and *TAF10* mRNAs, showing that there are about four times less *TAF8* mRNAs than those of *TAF10* (Supplementary Fig. 4a, b). In good agreement with our above endogenous anti-TAF10 RIP results (Fig. 1d, e), these IF-smiFISH experiments showed an about 10% co-localization between *TAF8* mRNA and TAF10 protein in the cytoplasm of HeLa cells (Supplementary Fig. 4c). To increase the number of *TAF8* mRNA molecules in the cytoplasm of HeLa cells and to be able to carry out analyses with-wild type (wt) and mutant (mt) TAF8 proteins, we carried out IF-smiFISH detections in HeLa cells exogenously expressing TAF8 protein. The IF-smiFISH co-localization experiments in fixed HeLa cells showed significant co-localisation between TAF10 protein and *TAF8* mRNA in the cytoplasm (Fig. 5a, e; note that to observe only the cytoplasmic IF signals the nuclear signal in the green channel was removed). Importantly, the co-localisation between *mtTAF8* mRNA (Fig. 3a) and TAF10 protein was lost (Fig. 5b, e). In addition, TAF8 protein detection by IF and *TAF10* mRNA by smiFISH, showed no significant co-localisation (Fig. 5c, e). Moreover, we could not detect any co-localisation between *CTNNB1* (*catenin beta-1*) mRNA and TAF10 protein (Fig. 5d, e), which further rules out any non-specific co-localisation of TAF10 protein with wt *TAF8* mRNA. Importantly, the statistical analysis of the co-localization enrichment ratio of TAF10 protein-wt *TAF8* mRNA measured in cells was significantly higher compared with all the

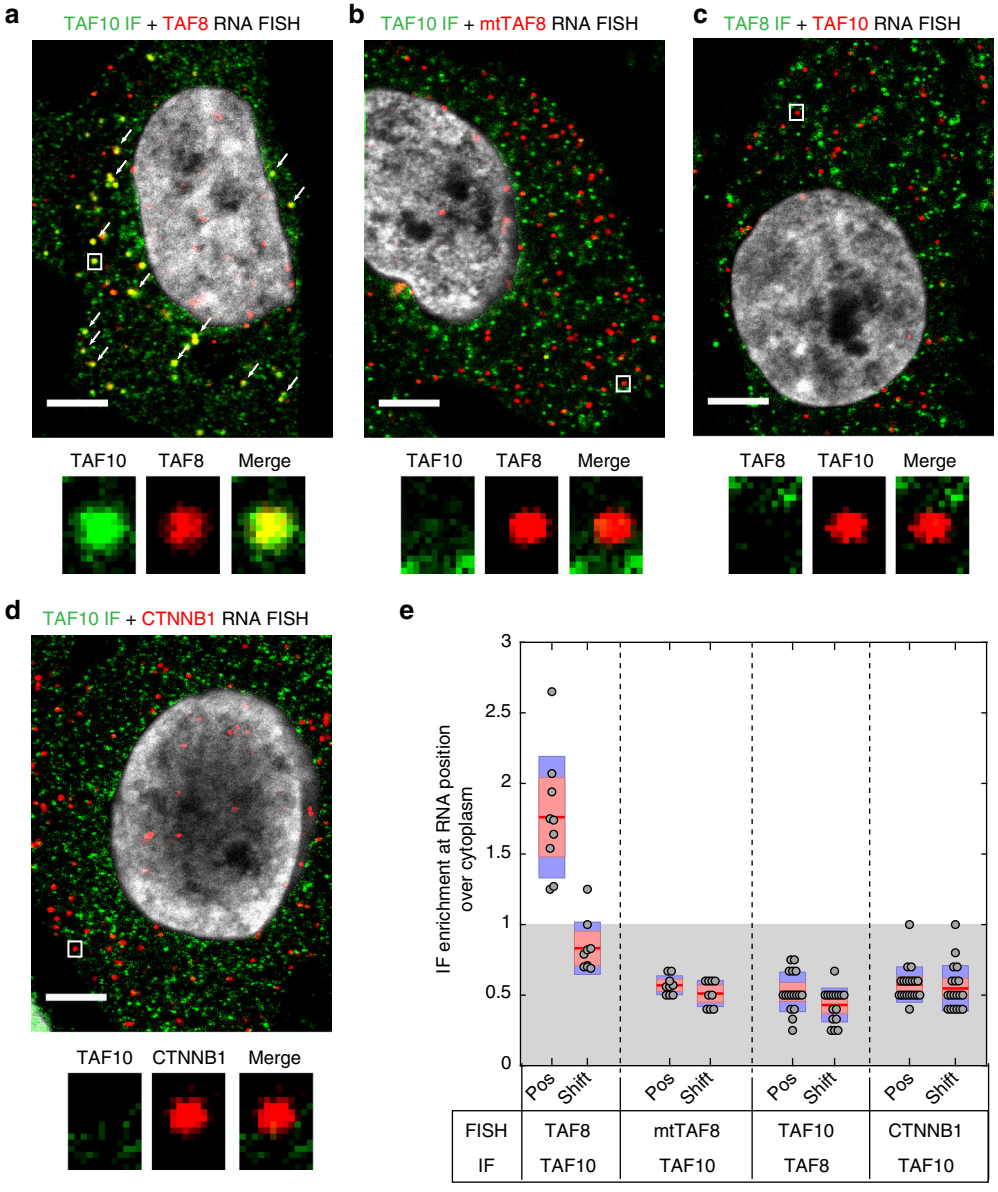


Fig. 5 Co-localization of TAF10 protein and *TAF8* mRNA in cytoplasm. **a**, **b** IF-smiFISH images in HeLa cells expressing either wild-type FLAG-TAF8 or mutant (mt) FLAG-TAF8 (TAF8 S57F L65F). Labels: red, Cy3-labelled *TAF8* probes; green, Alexa-488 labelled secondary antibody for TAF10 protein; co-localizing spots are indicated with white arrows. **c**, **d** Representative IF-smiFISH images of endogenous *TAF10* mRNA and TAF8 protein (**c**) and *CTNNB1* mRNA and TAF10 protein (**d**) in HeLa cells. Labels: red, Cy3-tagged *TAF10* FISH probes (**c**) and *CTNNB1* probes (**d**); green, Alexa-488 labelled secondary antibody for TAF8 (**c**) and TAF10 (**d**) protein. A typical cell recorded in each case and after counterstaining the nucleus with DAPI (grey) is shown. The nuclear signal in the green channel (TAF10 or TAF8 IF) was removed by masking the nucleus and using the "clear" option in ImageJ. Zoom-in regions shown under every image are indicated with a white rectangle. Scale bar (5 μm). **e** Boxplot showing enrichment ratios of IF signal at each RNA position over mean cytoplasmic intensity under all the conditions tested. Each grey dot represents one cell. Red horizontal lines are mean values, 95% confidence interval is shown in pink, and standard deviation in blue

other conditions tested (Fig. 5e). These imaging experiments demonstrate the physical proximity of TAF10 protein to *TAF8* mRNA in the cytoplasm. Moreover, this proximity is dependent on the ability of the two proteins to interact, lending further support to the sequential assembly model.

Position of IDs define the co-translational assembly pathway. To further test whether domain position guides co-translational assembly of HFD pairs in TFIID, TAF8 and TAF10 expression vectors were constructed in which the respective HFDs were exchanged. Our nascent RIP experiments from cells co-transfected with these swapped cDNA constructs (*TAF10-HFD8* and *TAF8-HFD10*) resulted in comparable *TAF8-HFD10* mRNA and protein enrichments (Fig. 6a, b and Supplementary Fig. 5a, b); as observed with the corresponding wt constructs (Fig. 2a, b), indicating that the origin of the HFD does not influence the sequential order of co-translational assembly. This experiment also suggested that the position of the HFD (N- or C-terminal), but not its sequence, determines the co-translational pathway by which the protein partners interact. Thus, next we tested whether

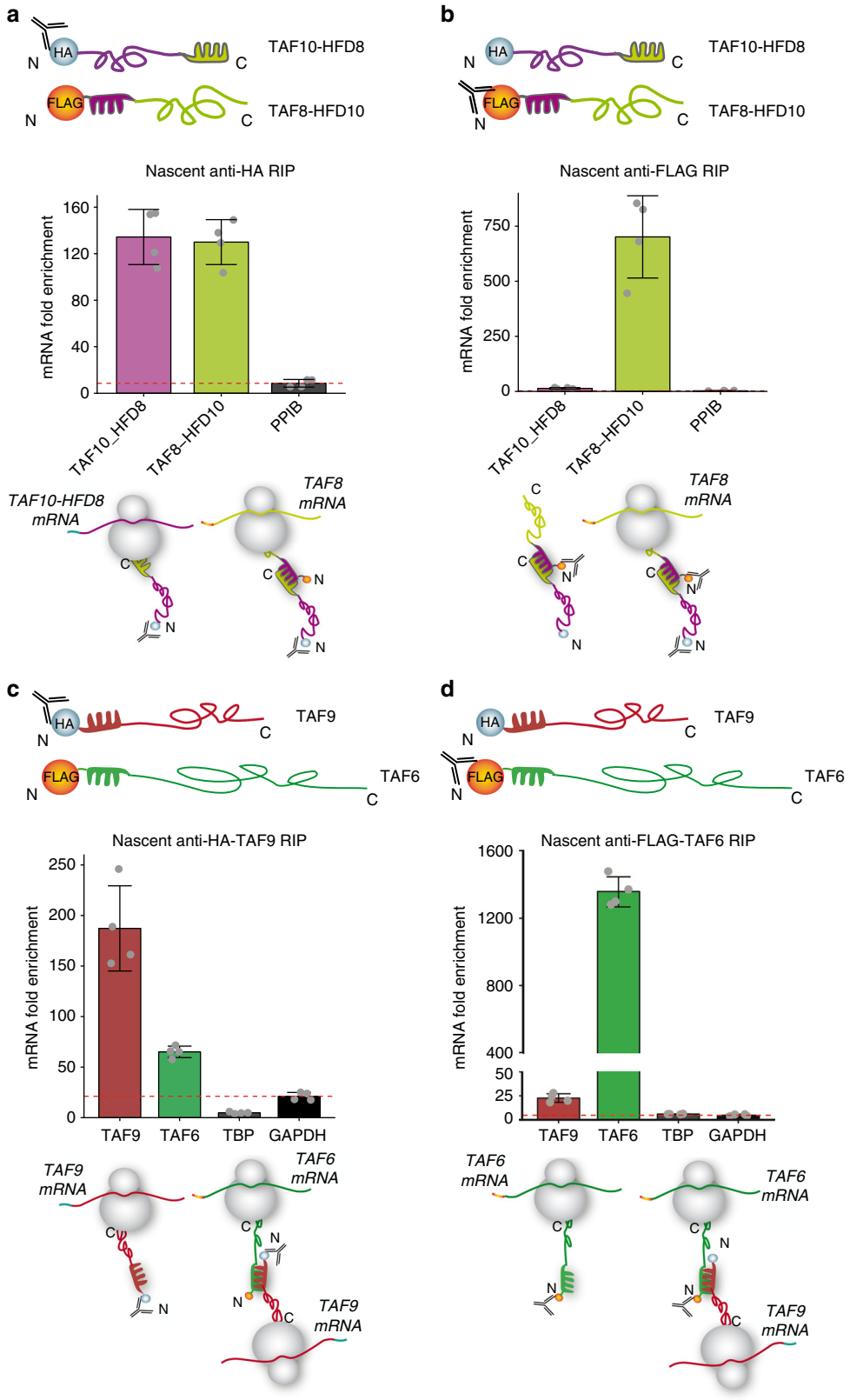


Fig. 6 ID position determines the order of co-translational assembly. RIP-qPCR of HFD domain-swapped TAF10 and TAF8 expression constructs using anti-HA (**a**) or anti-FLAG (**b**) antibodies against the respective N-terminal tags. **c**, **d** RIP-qPCR of anti- HA-TAF9 IP (**c**) and anti-FLAG-TAF6 IP (**d**) from HeLa cell polysome extracts co-transfected with TAF9 and TAF6 expression constructs as indicated. *PPIB* (**a**, **b**) and *GAPDH* (**c**, **d**) were used as negative control mRNAs. In all the graphs error bars represent ±SD from two biological replicates and two technical replicates (represented by grey dots). Source data provided as a Source Data File

the co-translational assembly of TAF6-TAF9 HFD pair would follow the simultaneous pathway, as they interact through their N-terminal HFDs (Fig. 6c, d). Our nascent RIPs revealed that both TAF6 and TAF9 co-IP their partners' mRNA (Fig. 6c, d and Supplementary Fig. 5c, d), suggesting that they assemble through the simultaneous assembly pathway, presumably as the neo-synthesised interaction domains of both proteins are exposed early during their synthesis on the ribosomes. Such a model would further suggest that *TAF6* and *TAF9* mRNAs could be found in close vicinity in the cytoplasm. To test the simultaneous co-translational assembly of TAF6-TAF9 HFD pair we have carried two colour smiFISH co-localization experiments to detect *TAF6* or *TAF9* mRNAs in fixed HeLa cells. These experiments showed a significantly higher co-localisation of the *TAF6* and *TAF9* mRNAs in the cytoplasm than several unrelated negative control mRNAs (Supplementary Fig. 6). These results show that the simultaneous co-translational assembly of TAF6-TAF9 HFD is detectable in the cytoplasm, however, with a relatively low frequency. This can be potentially explained by the fact that TAF6 can interact with TAF9b, and TAF9 with TAF6L[13], but the corresponding *TAF6L* and *TAF9b* mRNA co-localization combinations were not tested. Moreover, we cannot rule out the possibility that the fully synthesised TAF6 or TAF9 could also find their respective nascent partners still bound to the ribosomes through the sequential assembly pathway. Thus, it seems that the position of the dimerization domain may play a critical role in defining the order of co-translational assembly pathway of the corresponding interacting factors.

TBP and TAF1 interact also co-translationally. In TFIID, the evolutionary conserved core domain of TBP interacts with TAF1 via N-terminal TAND region of TAF1 and this interaction modulates the DNA-binding activity of TBP within TFIID[24,25]. To investigate co-translational assembly of other non-HFD-dependent interactions, we carried out genome-wide microarray analysis of TBP-associated RNAs from HeLa cell polysome extracts using a monoclonal antibody against the N-terminus of endogenous human TBP. In addition to TBP mRNA, we detected strong enrichment of 19 coding and non-coding RNAs. Among these, we found mRNAs coding for known TBP-interacting proteins: *BRF1* coding for a factor important for Pol III transcription[26], *BTAF1* coding for a B-TFIID subunit[27], as well as *TAF1*, whose enrichment on the microarray was somewhat weaker (Fig. 7a). Nevertheless, RIP-qPCR analysis in human HeLa cells (Fig. 7b) and mouse ESCs (Supplementary Fig. 7a) confirmed the microarray data and revealed a strong enrichment of the *TAF1* mRNA. Quantification of the *TAF1* mRNA in the anti-TBP RIP normalised to the protein IP efficiency indicated that the *TAF1* mRNA enrichment was around 62%. Consistent with the need for active translation, enrichment of all specific mRNAs was lost, or greatly decreased, upon puromycin treatment (Fig. 7b).

To further investigate the specificity of TBP-TAF1 interaction, we co-transfected expression vectors coding for the full-length human TBP with a ΔTAF1 expression vector, in which sequences coding for the first 168 residues containing the TAND region were deleted. Anti-TBP RIPs from cells expressing ΔTAF1 resulted in complete loss of *TAF1* mRNA enrichment and a reduction of the co-immunoprecipitated protein (Fig. 7c, d, Supplementary Fig. 7b, c). These results are consistent with a requirement of the N-terminal TAF1 domain to recruit TBP to the nascent TAF1 polypeptide. As the protein interface is formed by the C-terminal portion of TBP and the very N-terminus of TAF1[25,28], we predicted that similarly to TAF8-TAF10 assembly, a sequential assembly is also involved in the TBP-TAF1 interaction. Indeed, nascent anti-TAF1 RIP from an engineered GFP-TAF1 HeLa cell line (Fig. 7e, f) resulted in the enrichment of *TAF1* mRNA, but not that of *TBP*, thus supporting the co-translational assembly of TBP-TAF1 by the sequential pathway.

TREX-2 and SAGA DUB complexes assemble co-translationally. To extend our findings beyond TFIID, we examined co-translational assembly of ENY2 subunit with its respective partners. ENY2 is subunit of the TREX-2 mRNA-export complex and the DUB module of the SAGA transcription coactivator[13]. In TREX-2, two ENY2 proteins wrap around the central portion of the large GANP helical scaffold[29]. Similarly, human ENY2 wraps around the N-terminal helix of human ATXN7L3 in the highly intertwined SAGA DUB module[30] (Fig. 8a). To test whether the co-translational model is generally applicable to multisubunit complexes, we analysed ENY2-associated mRNAs from HeLa cells stably expressing ENY2 with an N-terminal GFP-tag[31]. Interestingly, we found that an anti-GFP-ENY2 RIP co-immunoprecipitates predominantly endogenous *GANP* mRNA and protein (the partner of ENY2 in TREX-2), and also endogenous *ATXN7L3* mRNA and protein (the binding partner of ENY2 in the SAGA DUB module) (Fig. 8b, c). Together, these results demonstrate that co-translational assembly is involved in the assembly of mammalian transcription complexes of diverse architecture and function.

Discussion

A functional protein must fold, translocate to its site of action and assemble with the right partners to carry out its function in the cell. The folding and assembly should be a well-regulated process in the cell to avoid non-specific interactions, and also because a single protein might interact with various partners depending on its interaction domain. Most eukaryotic proteins have more than one domain, which enables them to associate with their interaction partners. The building of multi-protein complexes in eukaryotes necessitates co-translational protein folding, the folding of a particular ID while still attached to translating ribosomes, to increase the efficiency of protein synthesis and prevent non-productive interactions[32]. Importantly, co-translational folding is aided by the ribosome, which stabilises specific folding intermediates of a protein[33–35]. Our results further demonstrate that the co-translational dimerization of protein interaction domains directs the assembly of mammalian nuclear multisubunit complexes. The cytoplasmic IF-smiFISH experiments indicate that the described co-translational assembly is clearly occurring in the cytoplasm of human cells and together with the mRNA enrichment calculations show that co-translational assembly is not a minor event. We also show that the position of the heterodimerization domain in a protein could guide its co-translational assembly either by sequential or simultaneous pathways. These mechanisms could play an important role in maintaining cellular health as excess orphan protein subunits can overburden protein folding and quality control machineries[36]. There is a strong correlation between the amino acid sequence of a protein, its translation rate and co-translational folding[37]. Rare codons in the mRNAs decrease the rate of translation, thereby allowing the protein to fold co-translationally[33]. Interestingly, translation pause sites are located downstream of the ID boundaries in order to regulate proper folding of multi-domain proteins[38], probably by assuring enough time for the co-translational interaction between the interacting subunits. In good agreement, our *Taf10* and *Taf8* KO mESCs, as well as F9 TAF10 ablation/re-expression experiments suggest that if the nascent ID exiting from the synthesizing ribosome cannot bind with its partner, the lack of interaction will lead to its

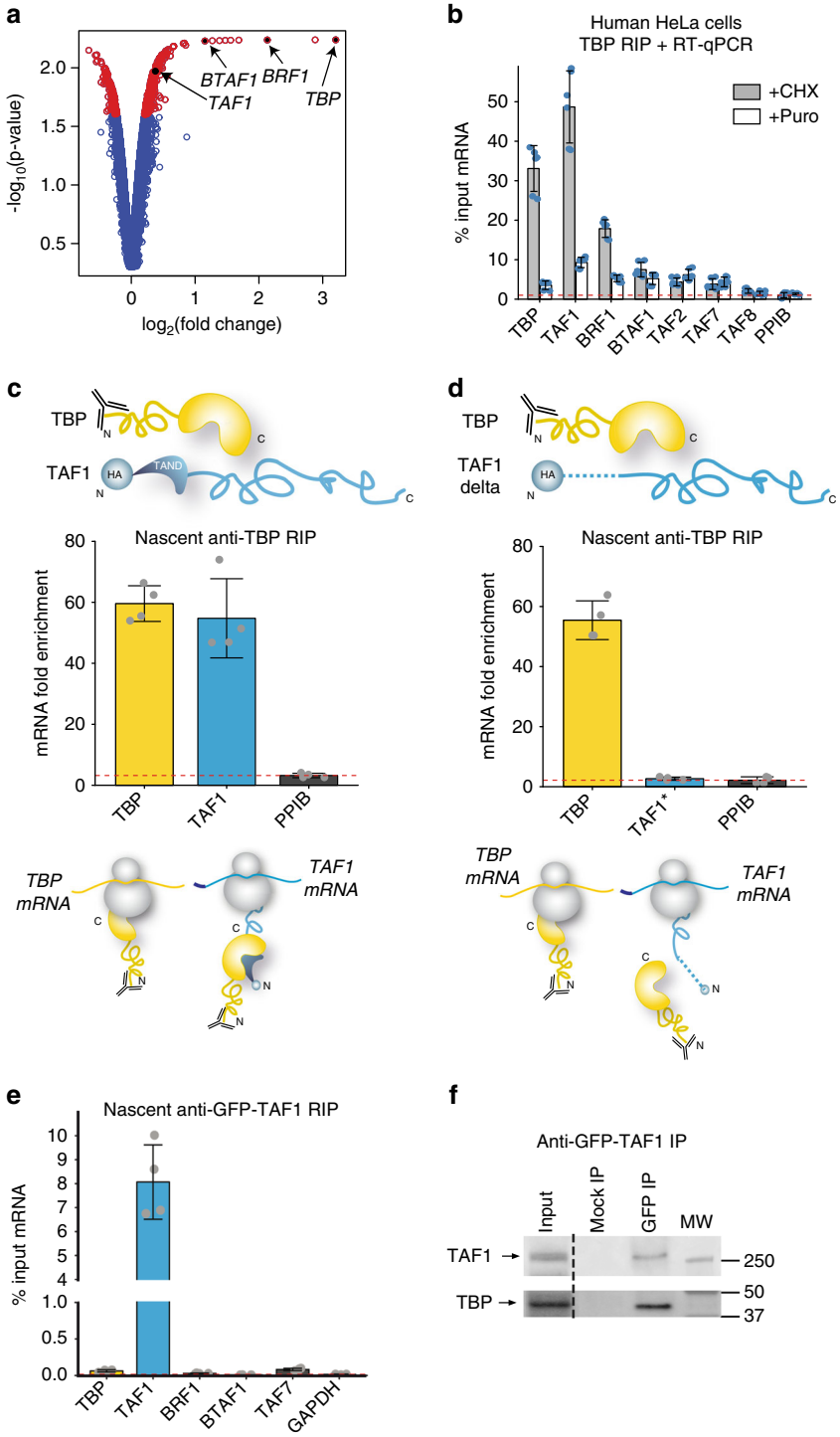


Fig. 7 Co-translational assembly of TBP and TAF1. **a** Microarray results of RIP with an antibody targeting the N-terminus of the endogenous TBP protein. Volcano plot depicting microarray results as log2 of the fold change of IP over mock. A *p*-value cut-off ≤ 0.025 was applied and corresponding transcripts are in red. Transcripts of interest are highlighted in black. **b** RIP-qPCR validation of microarray results in HeLa polysome extracts. **c**, **d** RIP-qPCR using anti-TBP antibody in HeLa cells transfected with TBP and HA-TAF1 expression constructs (**c**) or with TBP and HA-TAF1 with N-terminal deletion of the first 168 amino acids (**d**). **e** Anti-GFP RIP-qPCR from polysomes of HeLa cells stably expressing GFP-TAF1. **f** Western blot analysis (WB) of GFP IP from polysome extracts prepared from GFP-TAF1 cell line. In (**f**) the dotted line indicates the cutting out of unnecessary lanes. All error bars are ±SD from three (**b**) or two (**c**–**e**) biological replicates and two technical replicates [represented by blue dots in (**b**) or grey dots (**c**–**e**)]. Molecular weight (MW) markers are shown in kDa. *PPIB* (**b**–**d**) and *GAPDH* (**e**) were used as unrelated control mRNAs. Source data provided as a Source Data File

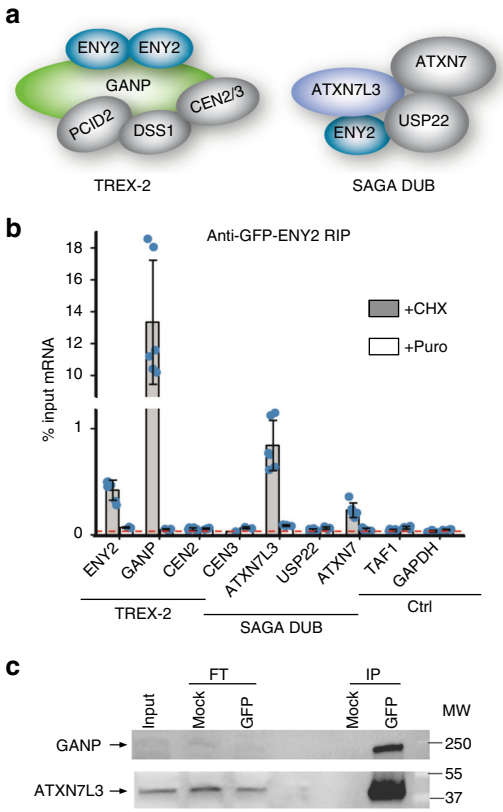


Fig. 8 TREX-2 and SAGA DUB complexes also assemble co-translationally. **a** ENY2 is shared between TREX-2 and SAGA complex. Two protein molecules of ENY2 binds to the large GANP subunit of TREX-2 complex and ENY2 is also a part of the deubiquitination module of the SAGA complex. **b** RIP-qPCR of GFP IP from HeLa cell polysome extracts stably expressing GFP-ENY2 tagged at the N-terminus. Error bars represent ±SD from two biological replicates and two technical replicates (represented by blue dots). **c** Western blot analysis of GFP IP from polysome extract prepared from GFP-ENY2 cell line. Molecular weight (MW) markers are shown in kDa. FT flow-through. IP immunoprecipitation. *GAPDH* was used as unrelated control mRNA. Source data provided as a Source Data File

translational arrest and consequent degradation of both the nascent protein and possibly the mRNA coding it. Note that the translational pausing causing mRNA destabilization could be an attractive model, however, primary transcriptional instead of posttranscriptional response cannot be ruled out. Nevertheless, it is conceivable that when nascent IDs are translated, the ribosome may pause or slow down until the interaction partner would arrive and bind, and thus stabilise the nascent ID. However, further systematic studies need to be carried out in order to study the role of translational pausing in co-translational protein assembly.

Co-translational assembly in homomeric proteins can also cause premature assembly of protein complexes, if two interacting nascent chains are in close proximity. It has been suggested that homomeric protein IDs are enriched toward the C termini of polypeptide chains across diverse proteomes[39] and this ID localisation is essential to prevent the assembly of homomeric proteins before proper folding. In contrast, our preliminary bioinformatics analyses using a limited curated interaction database[39] suggest that in heterodimeric proteins the N-terminal interaction regions are enriched, further underlining the idea that co-translational protein assembly in heterodimeric proteins is beneficial for assembling cellular machineries.

The role of chaperones in ribosome-associated nascent protein folding is well studied. Hsp70 family of proteins (such as, e.g., yeast Ssb) protects the nascent polypeptide from misfolding and aggregation in eukaryotes[39,40]. In bacteria and yeast, the ribosome-associated chaperones have been shown to interact with the nascent polypeptide chain emerging from the ribosome aiding in its folding[8,41–43]. Moreover, recently it has been suggested that upon emergence of a complete ID, the nascent chain interacts with its partner subunit and dissociates the chaperone complex from the nascent chain[8].

Our results reveal a systemic co-translational building of complexes in mammalian cells, but a thorough proteomic approach is necessary to identify chaperones necessary for these assembly pathways. It is possible that some of the chromatin regulatory complexes assemble through other chaperone-based mechanisms in the cytoplasm or directly in the nucleus.

In summary, we show that building blocks of mammalian nuclear transcription complexes, such as TFIID, SAGA and TREX-2, are assembled during translation and the way in which assembly occurs is consistent with the current knowledge of the preliminary structural organization of the complexes. Similar results from yeast, mouse, and human cells demonstrate that co-translational assembly is a general mechanism in eukaryotes [ref. [8] and this study]. Thus, the co-translational assembly of multi-protein complexes pathways seems to be a common regulatory mechanism in all eukaryotic cells to ensure efficient solutions to avoid non-specific protein interactions, protein aggregation and probably also to control the correct stoichiometry of subunits belonging to distinct complexes. In addition, our findings will significantly advance structural biology studies, because in the future extensive screening experiments will not be required to identify a real interaction partner(s) of a given subunit in a multi-protein complex. It will be enough to make an anti-subunit RIP from polysome extracts coupled to microarray analyses (or to RT-qPCRs) and the real endogenous interacting partner(s) can be taken immediately with high confidence for structural determinations and for building the architecture of multi-protein complexes.

Methods

Antibodies. Sources, catalogue numbers and concentrations of antibodies used for RIP, protein IP and western blotting are summarised in Supplementary Table 2.

Preparation of polysome-containing extracts and RIP. Polysome-containing extracts were prepared from adherent cells harvested at ~90% confluence by adapting a method for the isolation of ribosomes translating cell type-specific RNAs[44]. Briefly, 10 cm plates were treated with cycloheximide (100 µg/ml final) or puromycin (50 µg/ml final) and returned to the 37 °C incubator for 15 or 30 min, respectively. Subsequently, plates were placed on ice, washed twice with ice-cold PBS and scraped in 500 µl lysis buffer (20 mM HEPES KOH pH 7.5, 150 mM KCl, 10 mM $MgCl_2$ and 0.5% (vol/vol) NP-40), supplemented with complete EDTA-free protease inhibitor cocktail (Roche), 0.5 µM DTT, 40 U/ml RNasin (Promega), and cycloheximide or puromycin as needed. Extracts were prepared by homogenizing cells by 10 strokes of a B-type dounce and centrifugation at 17,000 × g. Clarified extracts were used to start immunoprecipitations, after saving 10% total RNA for input measurement. For TAF10 and TBP IPs, 20 µl of Protein G Dynabeads (ThermoFisher Scientific) were equilibrated by washing three times in lysis buffer, resuspended in 400 µl of lysis buffer and 2 µl of antibody, and incubated for 1 h at room temperature with end-to-end mixing. Beads were washed twice with IP500 buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 10% glycerol (v/v) and 0.1% NP-40 (v/v)) and three times in lysis buffer. Antibody-bound beads were thus used to perform RIP with polysome extracts overnight at 4 °C with end-over-end mixing. Mock RIP was carried out with equal amount of anti-GST antibody. The next day, beads were washed four times for 10 min at 4 °C with high salt-containing wash buffer (20 mM HEPES-KOH pH 7.5, 350 mM KCl, 10 mM $MgCl_2$ and 0.1% (vol/vol) NP-40) and subsequently eluted in 350 µl RA1 Lysis buffer and 7 µl 1 M DTT. RNAs were purified according to the manufacturer's instructions of the Macherey-Nagel total RNA purification kit, including the optional on-column DNase digestion step, and eluted twice in the same 60 µl of RNAse-free water. In the case of FLAG, HA, or GFP RIPs, 50 µl packed anti-FLAG M2 affinity gel (Sigma), 50 µl

packed EZviewTM Red Protein A affinity gel (Sigma) or 30 µl GFP-TRAP (Chromotek) slurry were equilibrated in lysis buffer and used for RIP.

cDNA preparation and RT-qPCR. For cDNA synthesis, 5 µl of purified RIP-RNA and 5 µl of 1:10 diluted input RNA samples were used. cDNA was synthesised using random hexamers and SuperScript IV (ThermoFischer Scientific) according to the manufacturer's instructions. For RIP performed on transfected cells, RNA was additionally treated with Turbo DNase (Ambion) according to the manufacturer's instructions in order to ensure complete plasmid removal before cDNA synthesis. Quantitative PCR was performed with primers (listed in Supplementary Table 1) on a Roche LightCycler 480 instrument with 45 cycles. In all cases, control cDNAs prepared without reverse transcriptase (−RT) were at least over 10 Cp values of the +RT cDNAs. Enrichment relative to input RNA was calculated using the formula $100 \times 2^{[(Cp(Input) - 6.644) - Cp(IP)]}$ and expressed as "% input RNA". In the case of RIPs performed on transfected cells, enrichment values were expressed as "mRNA fold enrichment" relative to the mock IP using the formula ΔΔCp [IP/mock], to account for the variability of transient transfections. "Relative mRNA fold enrichment" is expressed as mRNA fold enrichment of *TAF8* relative to mRNA fold enrichment of *TAF10* mRNA. All experiments were performed with a minimum of two biological and two technical replicates and values are represented as mean ±SD. Figures panels were prepared with taking in account all these data points using R (RStudio version 1.1.456 and R version 3.5.1).

Microarray analysis and library preparation. Polysome extracts and RIP from HeLa cells were performed as described above with mouse monoclonal antibodies 1H8 targeting the N-terminus of TAF10, 3G3 targeting the N terminus of TBP, and 1D10 targeting GST as a nonspecific control (see Supplementary Table 2). Protein G Sepharose beads were used (100 µl beads coupled to 14 µl antibody). After quantification and quality controls performed on Agilent's Bioanalyzer, biotinylated single strand cDNA targets were prepared, starting from 200 ng of total RNA, using the Ambion WT Expression Kit (Cat # 4411974) and the Affymetrix GeneChip® WT Terminal Labelling Kit (Cat # 900671) according to Affymetrix recommendations. Following fragmentation and end-labelling, 3 µg of cDNAs were hybridised for 16 h at 45 °C on GeneChip® Human Gene 2.0 ST arrays (Affymetrix) interrogating over 40000 RefSeq transcripts and ~11,000 LncRNAs represented by ~27 probes spread across the full-length of the transcript. The chips were washed and stained in the GeneChip® Fluidics Station 450 (Affymetrix) and scanned with the GeneChip® Scanner 3000 7 G (Affymetrix) at a resolution of 0.7 µm. Raw data (.CEL Intensity files) were extracted from the scanned images using the Affymetrix GeneChip® Command Console (AGCC) version 4.0. CEL files were further processed with Affymetrix Expression Console software version 1.3.1 to calculate probe set signal intensities using Robust Multi-array Average (RMA) algorithms with default settings (Sketch quantile normalization). Statistical analysis was performed using the FCROS package version 1.5.4[45]. Differences are considered significant for p value below 0.025. Volcano plots were performed using RStudio software version 3.3.2. Ribosomal RNA transcripts were filtered out. The microarray results reported in this paper are available in the Gene Expression Omnibus (GEO) under accession number GSE106299.

Cell lines, cell culture and transfections. HeLa cells (ATCC® CCL-2™) grown on adherent plates were obtained from the IGBMC cell culture facility and cultured in a 37 °C humidified/5% CO_2 incubator. Culture media consisted of Dulbecco's modified Eagle's medium (DMEM), supplemented with 1 g/l glucose, 5% fetal calf serum (FCS), and 40 µg/ml Gentamycin. The GFP-TAF1 cell line was generated by transferring full length human TAF1 fused at its N-terminus to EGFP into HeLa Flp-In/T-REx cells following procedures described in ref. [46]. E14 mouse embryonic stem cells [mESCs, ES Parental cell line E14Tg2a.4, obtained from Mutant Mouse Resource and Research Center (MMRRC), Citation ID:RRID:MMRRC_015890-UCD] at passage 29-31 were obtained from the IGBMC cell culture facility and cultured on gelatinised plates in feeder-free conditions in KnockOut DMEM (Gibco) supplemented with the following: 20 mM L-glutamine, Pen/Strep, 100 µM non-essential amino acids, 100 µM β-mercaptoethanol, N-2 supplement, B-27 supplement, 1000 U/ml LIF (Millipore), 15% ESQ FBS (Gibco) and 2i (3 µM CHIR99021, 1 µM PD0325901, Axon MedChem). Cells were passaged approximately every 3 days. The EGFP-ENY2 HeLa cell line was generated in our laboratory by D. Umlauf[31] and maintained at 37 °C in DMEM (1 g/l glucose), 10% FCS and 40 µg/ml Gentamycin[31]. The Dox-inducible hTAF10 expression system in *Taf10*$^{-/-}$ mouse F9 embryonal carcinoma cells was generated in our laboratory by E. Scheer[22]. Cells were cultured at 37 °C with 7% CO_2 in gelatinised plates in a culture media consisting of DMEM (4.5 g/l glucose), 10% FCS, 40 µg/ml Gentamycin in the presence of doxycycline (Sigma). The EGFP-ENY2 HeLa and the *Taf10*$^{-/-}$ mouse F9 embryonal carcinoma cell lines are available upon request.

Transfections were performed on ~90% confluent cells in 10 cm plates in antibiotic-free media using Lipofectamine 2000 (Thermo Fisher Scientific) and 3 µg plasmid DNA, according to the manufacturer's instructions. The medium was replaced with fresh medium containing gentamycin ~5–6 h post transfection and cells were harvested 24 h later. A descriptive summary of the plasmids used is presented in Supplementary Table 3.

Protein IP and western blot. Antibodies used for RIP, protein IP and western blotting are summarised in Supplementary Table 2. For protein IP, the procedure was performed essentially as for RIP. Bound proteins were eluted in 2× Laemmli buffer supplemented with 20 mM DTT and boiled for 5 min. Subsequently, samples were resolved on SDS-PAGE gels and transferred to nitrocellulose membranes using either wet transfer or BioRad's Trans-Turbo Blot semi-dry transfer method. Secondary antibodies (goat anti-mouse or rabbit anti-mouse) coupled to HRP (Jackson ImmunoResearch Laboratories) were used at 1:10,000 dilution. Signal was revealed using chemiluminescence (Pierce) and detected on the ChemiDoc imaging system (BioRad). For immunoprecipitation using whole cell extracts, 10 confluent 10 cm plates were scraped in PBS containing protease inhibitor (Roche) and resuspended in ~1 packed cell volume lysis buffer (20 mM Tris-HCl, pH 7.5, 400 mM KCl, 2 mM DTT, 20% glycerol) supplemented with protease inhibitor and 0.5 mM final concentration of DTT. Extracts were prepared by four cycles of freezing on liquid nitrogen followed by thawing on ice. The concentration of the clarified extract was measured by Bradford assay and the extract was diluted ~1:3 using lysis buffer without salt to achieve a final concentration of ~150 mM KCl. One-milligram extract was added to mock- and antibody-bound beads each and IPs were performed as described above. Proteins were eluted twice for 5 min at room temperature in 50 µl 0.1 M Glycine, pH 2.8 and neutralised with 3.5 µl 1.8 M Tris-HCl, pH 8.8. Ten percent of the pooled eluates were resolved on gels.

Plasmids. The eukaryotic expression plasmid pXJ41 used for all the constructs has been previously described[47]. pXJ41-TAF10-Nter-2HA has been previously described[48]. To generate N- and C-terminally Flag-tagged TAF8, the human TAF8 cDNA was PCR amplified from pACEMam1-CFP-TAF8 (kind gift from Imre Berger, University of Bristol, UK) using primers cotaining EcoR I and Bgl II restriction sites and tags incorporated at the N- or C-terminus, respectively, and digestion by appropriate restriction enzymes. Similarly, C-terminal HA tagged TAF10 was subcloned from pXJ41-TAF10-Nter-2HA by PCR amplification and digestion via restriction enzymes Xho I and Kpn I. The TAF8 mutations, TAF8-HFD and TAF8-HFD-60 amino acids were generated by site-directed mutagenesis using PfuUltra High-Fidelity DNA polymerase (Agilent Technologies), according to the manufacturer's instructions. The histone fold domain swapped TAF10 and TAF8 constructs were generated with several rounds of PCR amplification, using the already-mentioned N-terminal tagged TAF10 and TAF8 constructs as a template with specific primers and cloned into the vector via restriction enzymes EcoR I and Bgl II. pXJ41-hTBP has been previously described[49]. The HA-TAF1 cDNA[50] was inserted in pXJ41. TAF1 N-terminal deletion was carried out by site-directed mutagenesis using PfuUltra High-Fidelity DNA polymerase (Agilent Technologies), according to the manufacturer's instructions. HA tagged TAF9 was subcloned from pSG5-TAF9[51] by PCR amplification and digestion by restriction enzymes EcoR I and Bgl II. FLAG-tagged TAF6 was also subcloned in a similar manner from pXJ41-TAF6[52] via restriction enzymes Xho I and Kpn I. All plasmids have been verified by sequencing. Details on the cloning strategies are available upon request. Plasmids are described in Supplementary Table 3.

Mouse *Taf8* and *Taf10* KO ESC lines. The *Rosa26*$^{Cre\text{-}ERT2/+}$; *Taf8*$^{flox/flox}$ mouse embryonic stem cells (mESCs) were generated previously by F. El Saafin[21]. Briefly, mice carrying the *Taf8*^{lox} allele were bred to mice carrying the *Rosa26*$^{Cre\text{-}ERT2}$ allele to produce *Rosa26*$^{Cre\text{-}ERT2/+}$;*Taf8*$^{flox/flox}$ E3.5 blastocysts and to isolate *Rosa26*$^{Cre\text{-}ERT2/+}$;*Taf8*$^{flox/flox}$ mouse embryonic stem cells (mESCs)[21]. The *Rosa26*$^{Cre\text{-}ERT2/R}$; *Taf10*$^{flox/flox}$ mESCs were generated previously by P. Bardot[20]. Briefly, the ESCs were derived from *Rosa26*$^{Cre\text{-}ERT2/R}$;*Taf10*$^{lox/lox}$ E3.5 blastocysts[20]. mESCs were cultured in DMEM (4.5 g/l glucose) with 2 mM Glutamax-I, 15% ESQ FBS (Gibco), penicillin, streptomycine, 0.1 mM non-essential amino acids, 0.1% ß-mercaptoethanol, 1500 U/mL LIF and two inhibitors (2i; 3 µM CHIR99021 and 1 µM PD0325901, Axon MedChem) on gelatin-coated plates. To induce deletion of *Taf8*, mESCs were treated with 0.5 µM 4-OH tamoxifen (Sigma) for 5–6 days, and to induce deletion of *Taf10*, *Rosa26*$^{Cre\text{-}ERT2/R}$;*Taf10*$^{lox/lox}$ mESCs were treated for 4 days with 0.1 µM 4-OH tamoxifen (Sigma). The above-described mESCs have already been described[20,21] and were derived according to animal welfare regulations and guidelines of the French Ministry of Agriculture and French Ministry of Higher Education and Research, and the Australian Animal Welfare Committee, respectively.

smiFISH. smiFISH primary probes were designed with the R script Oligostan as previously described[23]. Primary probes and secondary probes (Cy3 or digoxigenin conjugated FLAPs) were synthesised and purchased from Integrated DNA Technologies (IDT). Primary probes were ordered at a final concentration of 100 µM, wet and frozen in Tris-EDTA pH 8.0 (TE) buffer. Probe sequences are available in Supplementary Table 4. An equimolar mixture of all the primary probes for a particular RNA was prepared with a final concentration 0.833 µM of individual probes. The secondary probes are resuspended in TE buffer at a final concentration of 100 µM. A total of 10 µl of FLAP hybridization reaction was prepared with 2 µl (for single colour smiFISH) or 4 µl (for dual colour smiFISH) of diluted (0.833 µM) primary probe set, 1 µl of secondary probe, 1 µl of 10X NEB3 and 6 µl of water. The reaction mix was then incubated in a cycler under the following conditions: 85 °C, 3 min, 65 °C, 3 min, 25 °C, 5 min. Two microliters of these FLAP hybridised probes

are necessary for each smiFISH reaction. The volume of the reactions were scaled up according to the number of smiFISH reactions carried out.

smiFISH was carried out as follows as per published protocol[23]. HeLa cells were treated with 100 µg/ml cycloheximide (Merck) for 15 min at 37 °C, fixed with 4% paraformaldehyde (Electron Microscopy Sciences) for 20 min at room temperature (RT) followed by overnight incubation with 70% ethanol at 4 °C. Following overnight incubation, cells were rinsed with 1× PBS twice and incubated with Solution A (freshly prepared 15% formamide in 1× SSC buffer) for 15 min at RT. During incubation, 50 µl Mix 1 (5 µl of 20× SSC, 1.7 µl of 20 µg/µl *E. coli* tRNA, 15 µl of 100% formamide, 2 or 4 µl of FLAP hybridised probes, required amount of water) and 50 µl Mix 2 (1 µl of 20 mg/ml RNAse-free BSA, 1 µl of 200 mM VRC, 27 µl of 40% dextran sulfate, 21 µl of water) was prepared. Mix 1 was added to Mix 2 after proper vortexing. The total 100 µl of Mix1 + Mix2 is sufficient for two coverslips. Each coverslip was then incubated on a spot of 50 µl of the Mix in a 15 cm Petri dish with a proper hydration chamber (3.5 cm Petri dish containing 2 ml of 15% formamide/1× SSC solution) overnight at 37 °C. Following overnight incubation, coverslips were washed twice with Solution A at 37 °C for 30 min each and with 1× PBS twice for 10 min each. Coverslips with only Cy3 conjugated secondary probes are mounted with 5 µl of Vectashield containing DAPI at this step. For DIG-labelled secondary probes, cells were further permeabilised with 0.1% Triton-X100 for 10 min at RT and incubated with 0.25 µg/ml anti-digoxigenin-fluorescein Fab fragments (diluted in 1× PBS) (Roche) for 2 h at RT. Following antibody incubation, cells were mounted as before.

IF-smiFISH. To visualise proteins and mRNA together, we first performed immunofluorescence (IF) followed by smiFISH. Briefly, cells were treated with 100 µg/ml cycloheximide (Merck) for 15 min at 37 °C, fixed with 4% paraformaldehyde (Electron Microscopy Sciences) for 10 min at room temperature (RT), blocked and permeabilised with blocking buffer (10% BSA, 10% Triton-X-100, 200 mM VRC, 2X PBS) for 1 h at 40 °C, incubated for 2 h at RT with either anti-TAF8 (mouse monoclonal antibody (mAb) 1FR-1B6[53]; diluted 1:1000) or anti-TAF10 (mAb 6TA-2B11[53]; diluted 1:1000) antibody mix followed by incubation (RT, 1 h) with secondary antibody mix Alexa-488-labelled goat anti-mouse mAb (Life Technologies, catalogue number A-11001, diluted 1:3000). Following immunofluorescence described above, cells were fixed with 4% paraformaldehyde (Sigma) for 10 min at RT. Cells were washed with 1× PBS and incubated with wash buffer [10% Formamide (Sigma) in 2× SSC] for 10 min at RT. smiFISH was carried out as described above and see ref. [23]. Cells were mounted using Vectashield mounting medium with DAPI (Vector laboratories Inc.).

Imaging and image processing. Confocal imaging of smiFISH and IF-smiFISH samples was performed on an SP8UV microscope (Leica) equipped with a 633-nm HeNe laser, a 561-nm DPSS laser, a 488-nm argon laser and a 405-nm laser diode. A ×63 oil immersion objective (NA 1.4) was used and images were taken by using the hybrid detector photon-counting mode. The laser power for all acquisitions and laser lines was set to 10%. All images acquired have a bit depth of 8 bit and a pixel resolution of 70 nm. The z-stacks were taken with a z-spacing of 300 nm for a total of 4–6 µm. Image processing was performed using the Fiji/Image J software. All images were processed the same way. In detail, the channels of the different images were split and grey values were adjusted to better visualise the spots in the cytoplasm. The nuclear signal in the green channel (TAF10 or TAF8 IF) was removed by masking the nucleus and using the "clear" option. Finally, the processed channels were merged again. For IF-smiFISH, one cell of an image was cropped and one representing z-slice per cell was chosen. For smiFISH, maximum intensity Z-projections of individual images were made and one cell per resulting image was cropped as the representative image. In addition, one single IF or smiFISH spot from the corresponding cells was cropped as well.

Image analysis of IF-smiFISH data. To measure the degree of spatial overlap of smiFISH (mRNA) and IF (protein) signal, an enrichment ratio was calculated as described below. Such quantification was chosen in order to take into account the variability of IF signal between cells, making single object detection in this channel difficult. Cells and nuclei were outlined manually in 2D based on the GFP and DAPI image, respectively. Subsequent analyses were restricted to the cytoplasm. mRNAs were detected in 3D with FISH-quant[23]. Identical detection settings were used when different experimental conditions were compared with the same gene. Each cell was post-processed separately. First, the median pixel intensity in the IF image at the identified RNA positions was calculated. Second, a normalization factor was estimated as the median IF intensity of the outlined cytoplasm within the z-range of the detected mRNAs. The enrichment ratio of the cell was then calculated as the ratio of the median IF intensity at the RNA positions divided by the mean cytoplasmic intensity. Boxplots of enrichment ratios were generated with the Matlab function notBoxPlot. Each dot corresponds to the estimation of one cell. Horizontal lines are mean values, 95% confidence interval is shown in red, and standard deviation in blue. Statistical comparison between different experimental conditions was performed with two-sample Kolmogorov–Smirnov test (Matlab function kstest2). The Matlab script is available upon request.

Image analysis of smiFISH co-localization data. Segmentation of nuclei and cells was performed with the DAPI and smiFISH channel channels, respectively. 2D images were obtained with a previously described projection approach based on local and global focus measurements[23]. Segmentation was implemented with the open-source software CellProfiler[54] using a standard workflow: Otsu and watershed separation for nuclei in the DAPI channel. Each nucleus then serves as a seed for a watershed segmentation to obtain the cells in the smiFISH channel. Individual RNA molecules were localised with FISH-quant in 3D and can be treated as point clouds[55]. Co-localization analysis between detected RNAs in two colours was solved as a linear assignment problem (LAP) with the Hungarian algorithm (Matlab function hungarianlinker and munkres from Matlab FileExchange). In short, this algorithm finds the best possible global assignment between these two points-clouds such that for each point in the first colour the closest point in the second channel is found. We implemented a user interface for this analysis tool (FQ_DualColor), which is distributed together with a dedicated user manual with FISH-quant: https://bitbucket.org/muellerflorian/fish_quant

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The microarray data corresponding to Figs. 1a and 7a are available in the Gene Expression Omnibus (GEO) under accession number GSE106299. The source data corresponding to Figs. 1d–e, 2a–d, 3a–b, 4a–d, 6a–d, 7b–f, 8a–c and Supplementary Figs. 1, 2a–d, 3a–b, 5a–d, 7a–c are provided as a Source Data file. A reporting summary for this Article is available as a Supplementary Information file. Raw image files (~800), their corresponding analyses, and all other data supporting the findings of the study are available from the corresponding author upon request.

Code availability

The Matlab script (Kamenova_NatComm__rna_protein_coloc.m) concerning the RNA co-localization and IF-smiFISH analyses is available on the FISH-quant repository [https://bitbucket.org/muellerflorian/fish_quant]. The custom R scripts for dot plot overlaid bar charts are available upon request.

Received: 21 September 2018 Accepted: 29 March 2019
Published online: 15 April 2019

References

1. Chen, J. L., Attardi, L. D., Verrijzer, C. P., Yokomori, K. & Tjian, R. Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* **79**, 93–105 (1994).
2. Shieh, Y. W. et al. Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science* **350**, 678–680 (2015).
3. Wells, J. N., Bergendahl, L. T. & Marsh, J. A. Operon gene order is optimized for ordered protein complex assembly. *Cell Rep.* **14**, 679–685 (2016).
4. Halbach, A. et al. Cotranslational assembly of the yeast SET1C histone methyltransferase complex. *EMBO J.* **28**, 2959–2970 (2009).
5. Duncan, C. D. & Mata, J. Widespread cotranslational formation of protein complexes. *PLoS Genet.* **7**, e1002398 (2011).
6. Kassem, S., Villanyi, Z. & Collart, M. A. Not5-dependent co-translational assembly of Ada2 and Spt20 is essential for functional integrity of SAGA. *Nucleic Acids Res.* **45**, 7539 (2017).
7. Williams, N. K. & Dichtl, B. Co-translational control of protein complex formation: a fundamental pathway of cellular organization? *Biochem Soc. Trans.* **46**, 197–206 (2018).
8. Shiber, A. et al. Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature* **561**, 268 (2018).
9. Natan, E., Wells, J. N., Teichmann, S. A. & Marsh, J. A. Regulation, evolution and consequences of cotranslational protein complex assembly. *Curr. Opin. Struct. Biol.* **42**, 90–97 (2017).
10. Muller, F., Zaucker, A. & Tora, L. Developmental regulation of transcription initiation: more than just changing the actors. *Curr. Opin. Genet. Dev.* **20**, 533–540 (2010).
11. Warfield, L. et al. Transcription of nearly all yeast RNA polymerase II-transcribed genes is dependent on transcription factor TFIID. *Mol. Cell* **68**, 118–129 e5 (2017).
12. Baptista, T. et al. SAGA is a general cofactor for RNA polymerase II transcription. *Mol. Cell* **68**, 130–143 e5 (2017).
13. Helmlinger, D. & Tora, L. Sharing the SAGA. *Trends Biochem. Sci.* **42**, 850–861 (2017).
14. Fribourg, S. et al. Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in *E. coli*. *J. Mol. Biol.* **306**, 363–373 (2001).

15. Trowitzsch, S. et al. Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat. Commun.* **6**, 6011 (2015).
16. Garreau de Loubresse, N. et al. Structural basis for the inhibition of the eukaryotic ribosome. *Nature* **513**, 517–522 (2014).
17. Pestka, S. Inhibitors of ribosome functions. *Annu. Rev. Microbiol.* **25**, 487–562 (1971).
18. Soutoglou, E. et al. The nuclear import of TAF10 is regulated by one of its three histone fold domain-containing interaction partners. *Mol. Cell Biol.* **25**, 4092–4104 (2005).
19. Kramer, G., Boehringer, D., Ban, N. & Bukau, B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* **16**, 589–597 (2009).
20. Bardot, P. et al. The TAF10-containing TFIID and SAGA transcriptional complexes are dispensable for early somitogenesis in the mouse embryo. *Development* **144**, 3808–3818 (2017).
21. El-Saafin, F. et al. Homozygous TAF8 mutation in a patient with intellectual disability results in undetectable TAF8 protein, but preserved RNA polymerase II transcription. *Hum. Mol. Genet.* **27**, 2171–2186 (2018).
22. Metzger, D., Scheer, E., Soldatov, A. & Tora, L. Mammalian TAF(II)30 is required for cell cycle progression and specific cellular differentiation programmes. *EMBO J.* **18**, 4823–4834 (1999).
23. Tsanov, N. et al. smiFISH and FISH-quant—a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* **44**, e165 (2016).
24. Kotani, T. et al. Identification of highly conserved amino-terminal segments of dTAF$_{II}$230 and yTAF$_{II}$145 that are functionally interchangeable for inhibiting TBP-DNA interactions in vitro and in promoting yeast cell growth in vivo. *J. Biol. Chem.* **273**, 32254–32264 (1998).
25. Anandapadamanaban, M. et al. High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nat. Struct. Mol. Biol.* **20**, 1008–1014 (2013).
26. Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* **45**, 439–446 (2012).
27. Pereira, L. A., van der Knaap, J. A., van den Boom, V., van den Heuvel, F. A. & Timmers, H. T. TAF(II)170 interacts with the concave surface of TATA-binding protein to inhibit its DNA binding activity. *Mol. Cell Biol.* **21**, 7523–7534 (2001).
28. Liu, D. et al. Solution structure of a TBP–TAF$_{II}$230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell* **94**, 573–583 (1998).
29. Jani, D. et al. Functional and structural characterization of the mammalian TREX-2 complex that links transcription with nuclear messenger RNA export. *Nucleic Acids Res.* **40**, 4562–4573 (2012).
30. Samara, N. L. et al. Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science* **328**, 1025–1029 (2010).
31. Umlauf, D. et al. The human TREX-2 complex is stably associated with the nuclear pore basket. *J. cell Sci.* **126**, 2656–2667 (2013).
32. Gloge, F., Becker, A. H., Kramer, G. & Bukau, B. Co-translational mechanisms of protein maturation. *Curr. Opin. Struct. Biol.* **24**, 24–33 (2014).
33. Thommen, M., Holtkamp, W. & Rodnina, M. V. Co-translational protein folding: progress and methods. *Curr. Opin. Struct. Biol.* **42**, 83–89 (2017).
34. Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I. Jr. & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 (2011).
35. Holtkamp, W. et al. Cotranslational protein folding on the ribosome monitored in real time. *Science* **350**, 1104–1107 (2015).
36. Harper, J. W. & Bennett, E. J. Proteome complexity and the forces that drive proteome imbalance. *Nature* **537**, 328–338 (2016).
37. Wruck, F., Katranidis, A., Nierhaus, K. H., Buldt, G. & Hegner, M. Translation and folding of single proteins in real time. *Proc. Natl Acad. Sci. USA* **114**, E4399–E4407 (2017).
38. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **16**, 274–280 (2009).
39. Willmund, F. et al. The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. *Cell* **152**, 196–209 (2013).
40. Doring, K. et al. Profiling Ssb-nascent chain interactions reveals principles of Hsp70-assisted folding. *Cell* **170**, 298–311 (2017).
41. Haldar, S., Tapia-Rojo, R., Eckels, E. C., Valle-Orero, J. & Fernandez, J. M. Trigger factor chaperone acts as a mechanical foldase. *Nat. Commun.* **8**, 668 (2017).
42. Ferbitz, L. et al. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590–596 (2004).
43. Hoffmann, A. et al. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J. Biol. Chem.* **281**, 6539–6545 (2006).
44. Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P. & Heintz, N. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protoc.* **9**, 1282–1291 (2014).
45. Dembele, D. & Kastner, P. Fold change rank ordering statistics: a new method for detecting differentially expressed genes. *BMC Bioinforma.* **15**, 14 (2014).
46. van Nuland, R. et al. Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Mol. Cell Biol.* **33**, 2067–2077 (2013).
47. Xiao, J. H., Davidson, I., Matthes, H., Garnier, J. M. & Chambon, P. Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. *Cell* **65**, 551–568 (1991).
48. Jacq, X. et al. Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* **79**, 107–117 (1994).
49. Brou, C. et al. Distinct TFIID complexes mediate the effect of different transcriptional activators. *EMBO J.* **12**, 489–499 (1993).
50. Ruppert, S., Wang, E. H. & Tjian, R. Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation. *Nature* **362**, 175–179 (1993).
51. Frontini, M. et al. TAF9b (formerly TAF9L) is a bona fide TAF that has unique and overlapping roles with TAF9. *Mol. Cell Biol.* **25**, 4638–4649 (2005).
52. Bell, B., Scheer, E. & Tora, L. Identification of hTAF$_{II}$80δ links apoptotic signaling pathways to transcription factor TFIID function. *Mol. Cell* **8**, 591–600 (2001).
53. Mohan, I. W., Scheer, E., Wendling, O., Metzger, D. & Tora, L. TAF10 (TAF$_{II}$30) is necessary for TFIID stability and early embryogenesis in mice. *Mol. Cell Biol.* **23**, 4307–4318 (2003).
54. Dao, D. et al. CellProfiler Analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinformatics* **32**, 3210–3212 (2016).
55. Mueller, F. et al. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013).

Acknowledgements

We would like to thank all members of the Tora lab for thoughtful discussions and suggestions throughout the course of the work, E. Scheer for help with F9 cell experiment, and G. Caliskan for help with the ATAC experiments. In addition, we are grateful to E. Bertrand and his group for smiRNA FISH training, T. Sexton for advice and carefully reading the manuscript, V. Alunni for microarray sample preparation, S. Bour for illustrations, K. Gupta and I. Berger for suggesting TAF deletions and mutations, D. Van Essen for providing constructs, P. Jane Palli and G. Travé for help with protein domain analyses, the IGBMC cell culture facility for cells and media. PM was supported by a fellowship (DOC20180507393) from Fondation ARC pour la Recherche sur le Cancer. This work was supported by funds from CNRS, INSERM, and Strasbourg University. This study was also supported by the European Research Council (ERC) Advanced grant (ERC-2013-340551, Birtoaction, to L.T.) and grant ANR-10-LABX-0030-INRT (to L.T.), a French State fund managed by the Agence Nationale de la Recherche under the frame programme Investissements d'Avenir ANR-10-IDEX-0002-02 (to IGBMC) and a grant from the Collaborative Center for X-linked Dystonia and Parkinsonism (to H.T.M.T.).

Author contributions

I.K., P.M. and L.T. designed the study; I.K. and P.M. performed all the molecular lab work (their names appear in alphabetical order in author list), J.M.G. performed all the cloning experiments. P.M. and S. Conic carried out the imaging experiments, which were analysed by F.M., S. Caponi created a stable cell line, F.E.S., P.B. and S.D.V. carried out the mouse ESC knock-out experiments; I.K. and P.M. analysed data, D.D. analysed the microarray data. All authors contributed to the text and figure panels. I.K., P.M., H.T.M.T., S.D.V. and L.T. wrote the manuscript. All authors gave final approval for publication.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-09749-y.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: *Nature Communications* thanks P. Weil and the other anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information (Figures and Tables)

Co-translational assembly of mammalian nuclear multisubunit complexes

Ivanka Kamenova[1,2,3,4,#], Pooja Mukherjee[1,2,3,4,#], Sascha Conic[1,2,3,4,], Florian Mueller[5], Farrah El-Saafin[1,2,3,4,], Paul Bardot[1,2,3,4,], Jean-Marie Garnier[1,2,3,4], Doulaye Dembele[1,2,3,4,], Simona Capponi[6], H.T. Marc Timmers[6], Stéphane D. Vincent[1,2,3,4,] and László Tora[1,2,3,4, *]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France;

[2]Centre National de la Recherche Scientifique, UMR7104, Illkirch, France;

[3]Institut National de la Santé et de la Recherche Médicale, U1258, Illkirch, France;

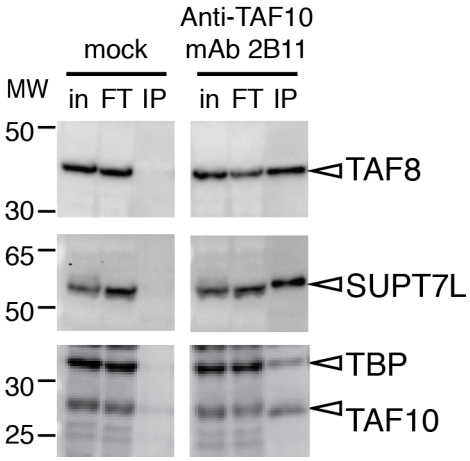
[4]Université de Strasbourg, Illkirch, France;

[5]Computational Imaging & Modeling Unit, Institut Pasteur, Département Biologie Cellulaire et Infections, 25-28 rue du Docteur Roux, 75015 Paris, France;

[6]German Cancer Consortium (DKTK) partner site Freiburg, German Cancer Research, Center (DKFZ) and Department of Urology, Medical Center-University of Freiburg, Germany.

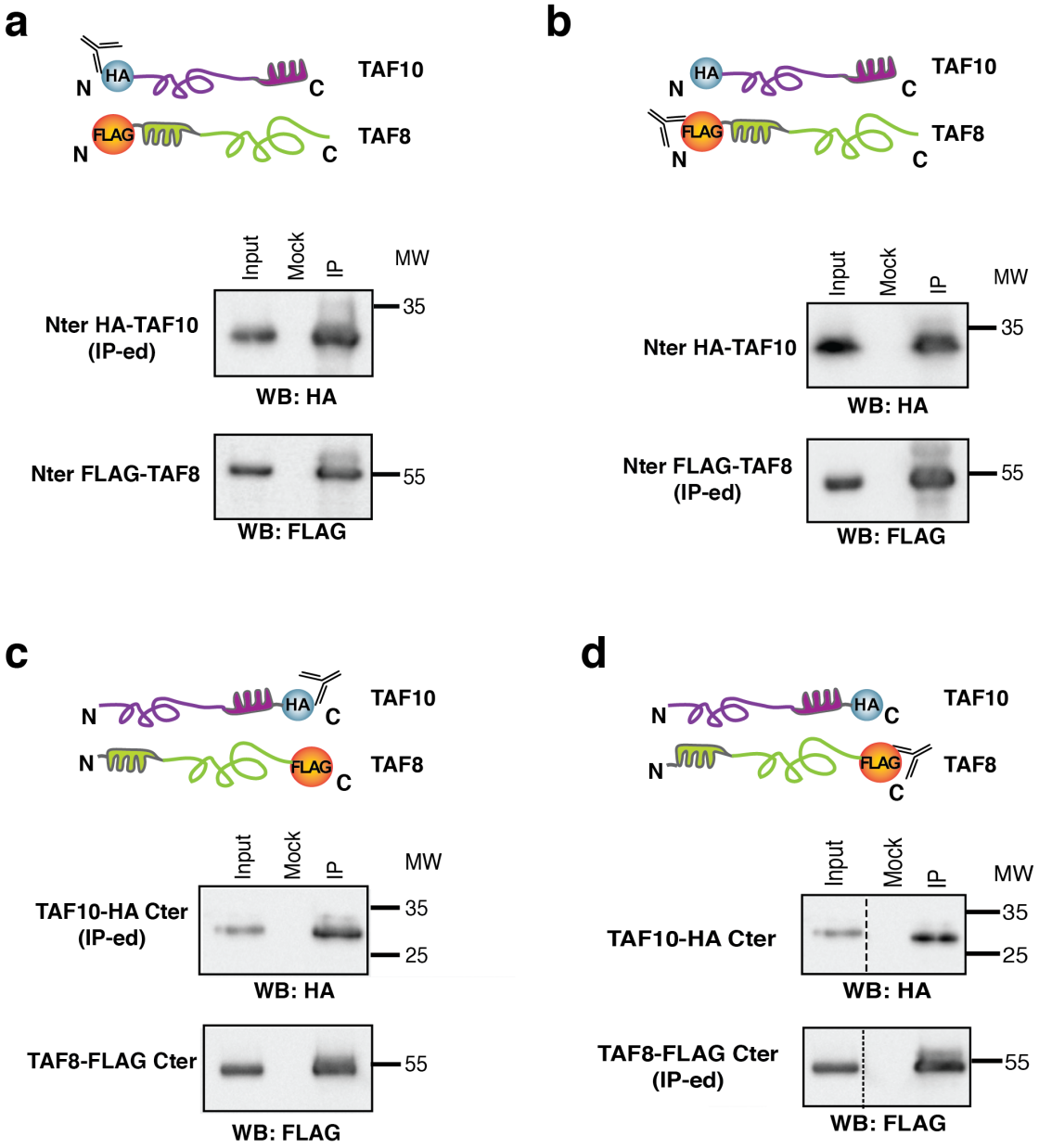
[#]These authors contributed equally

[*]Correspondence should be addressed to L.T. laszlo@igbmc.fr



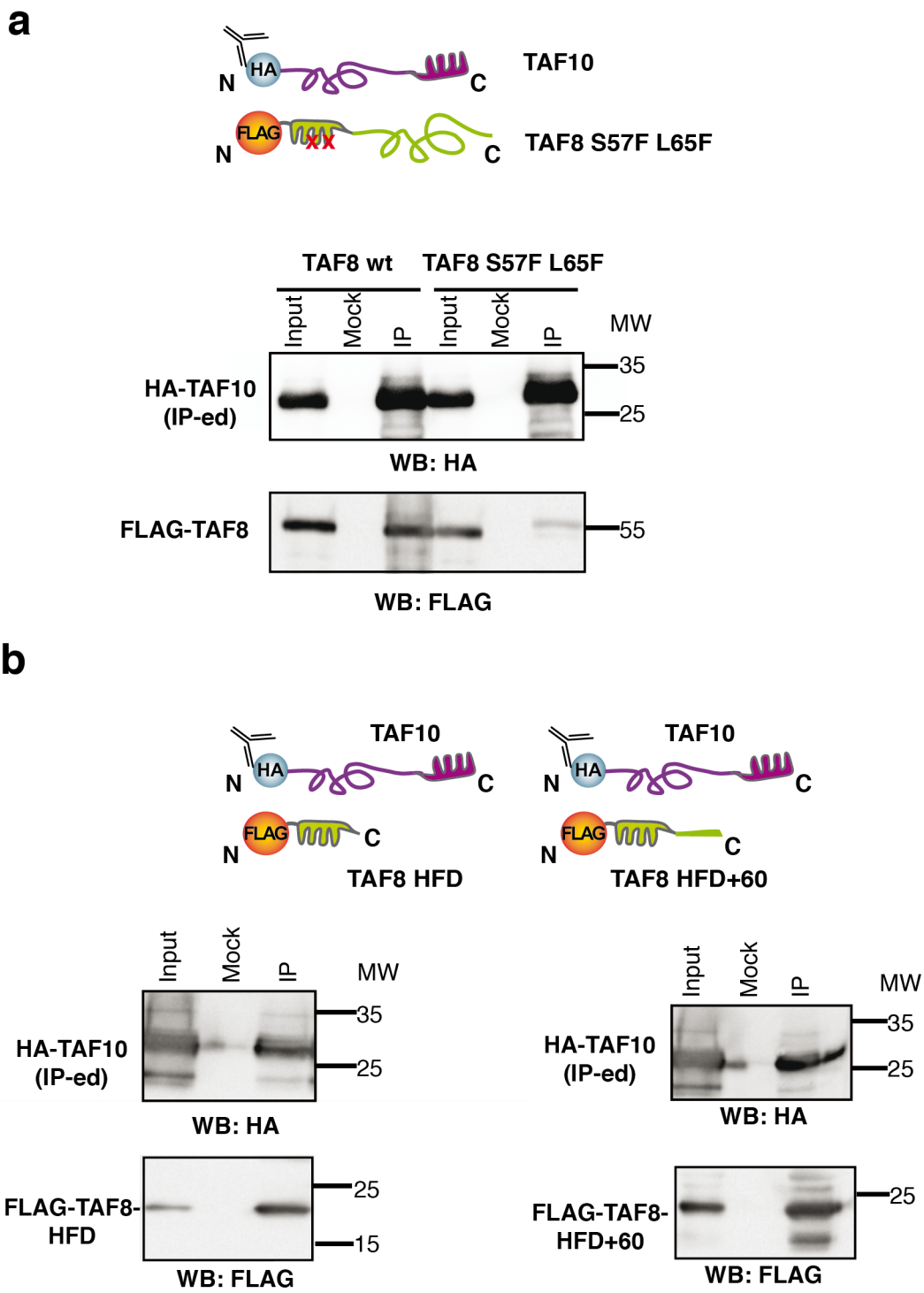
Supplementary Figure 1

Western blot analysis of immunoprecipitations from E14 mESC whole cell extract using mock (anti-GST) or anti-TAF10 (6TA 2B11) antibodies. 6TA 2B11 recognizes the N-terminal unstructured region of TAF10. Blots were probed with antibodies against the indicated proteins. In = input, FT = flow-through, IP = eluate from IP. Molecular weight (MW) markers are shown in kDa. Source data provided as a Source Data File.



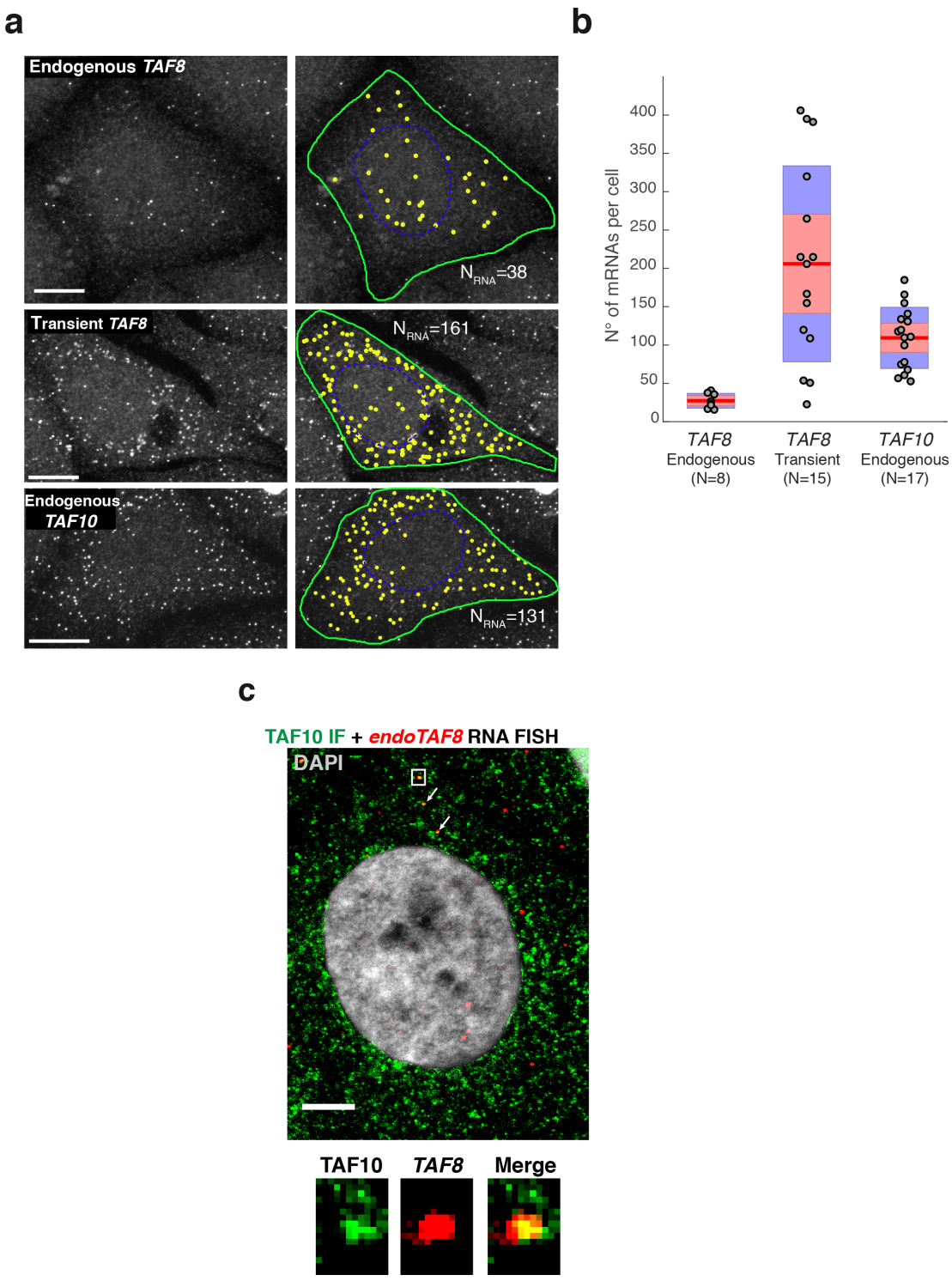
Supplementary Figure 2

(a-d) Western blot analyses of RIPs analysed in Fig. 2. HeLa cells transfected with expression constructs for N terminal (**a-b**) or C terminal (**c-d**) tagged HA-TAF10 and FLAG-TAF8 proteins. Blots were probed with anti-FLAG M2 and anti-HA antibodies as indicated. Molecular weight (MW) markers are shown in kDa. In = input, FT = flow-through, IP = eluate from IP. Molecular weight (MW) markers are shown in kDa. Source data provided as a Source Data File.



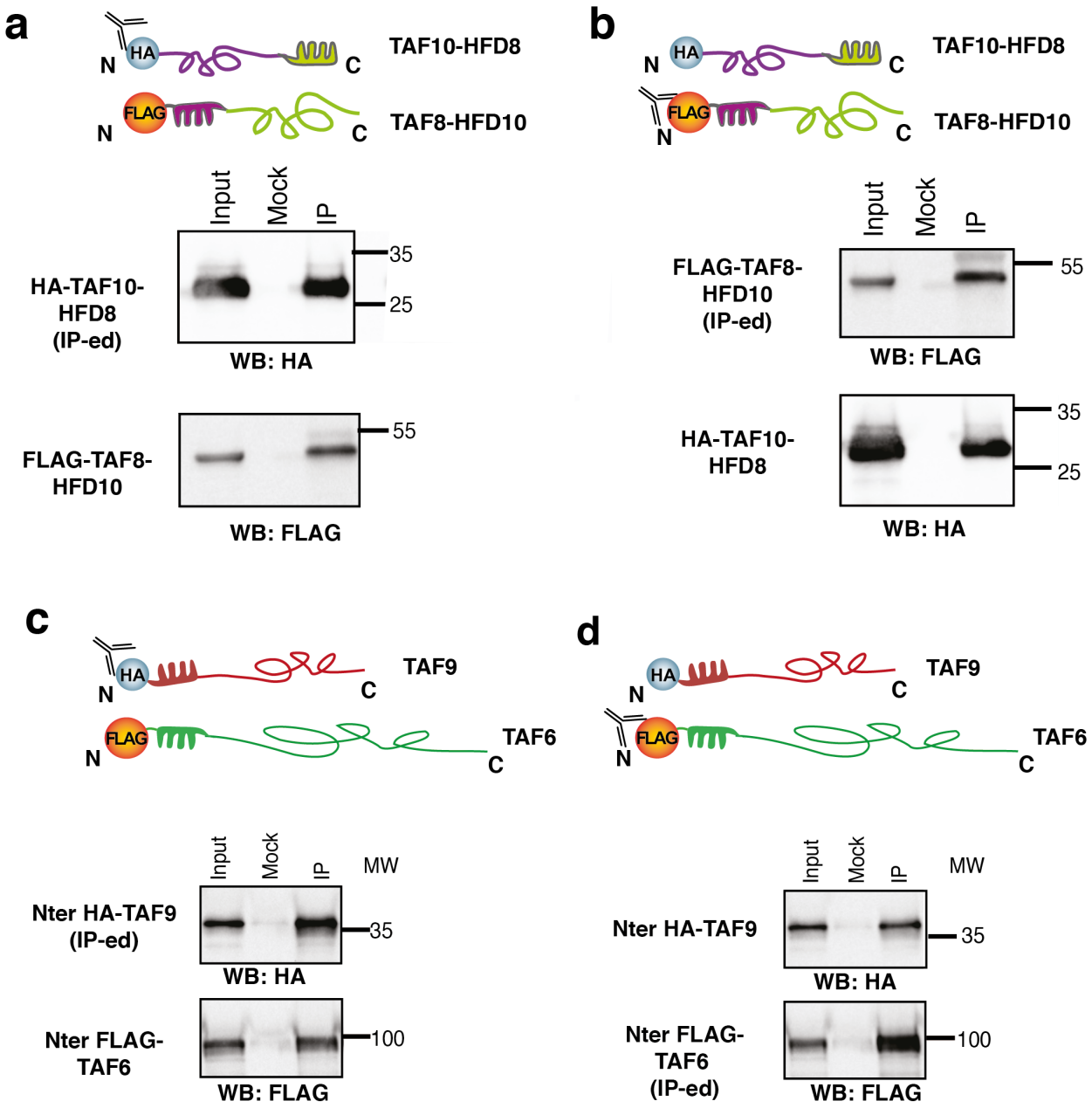
Supplementary Figure 3

Western blot analyses of RIPs analysed in Fig. 3. Blots were probed with anti-HA and anti-FLAG M2 antibodies as indicated. Hela cells transfected with expression constructs coding for HA-TAF10 and mutant FLAG-TAF8 S57F L65F proteins (**a**) and coding for HA-TAF10 and minimal TAF8 HFD (left) or TAF8 HFD extended with 60 amino acids (TAF8 HFD+60, right) (**b**). Molecular weight (MW) markers are shown in kDa. In = input, FT = flow-through, IP = eluate from IP. Molecular weight (MW) markers are shown in kDa. Source data provided as a Source Data File.



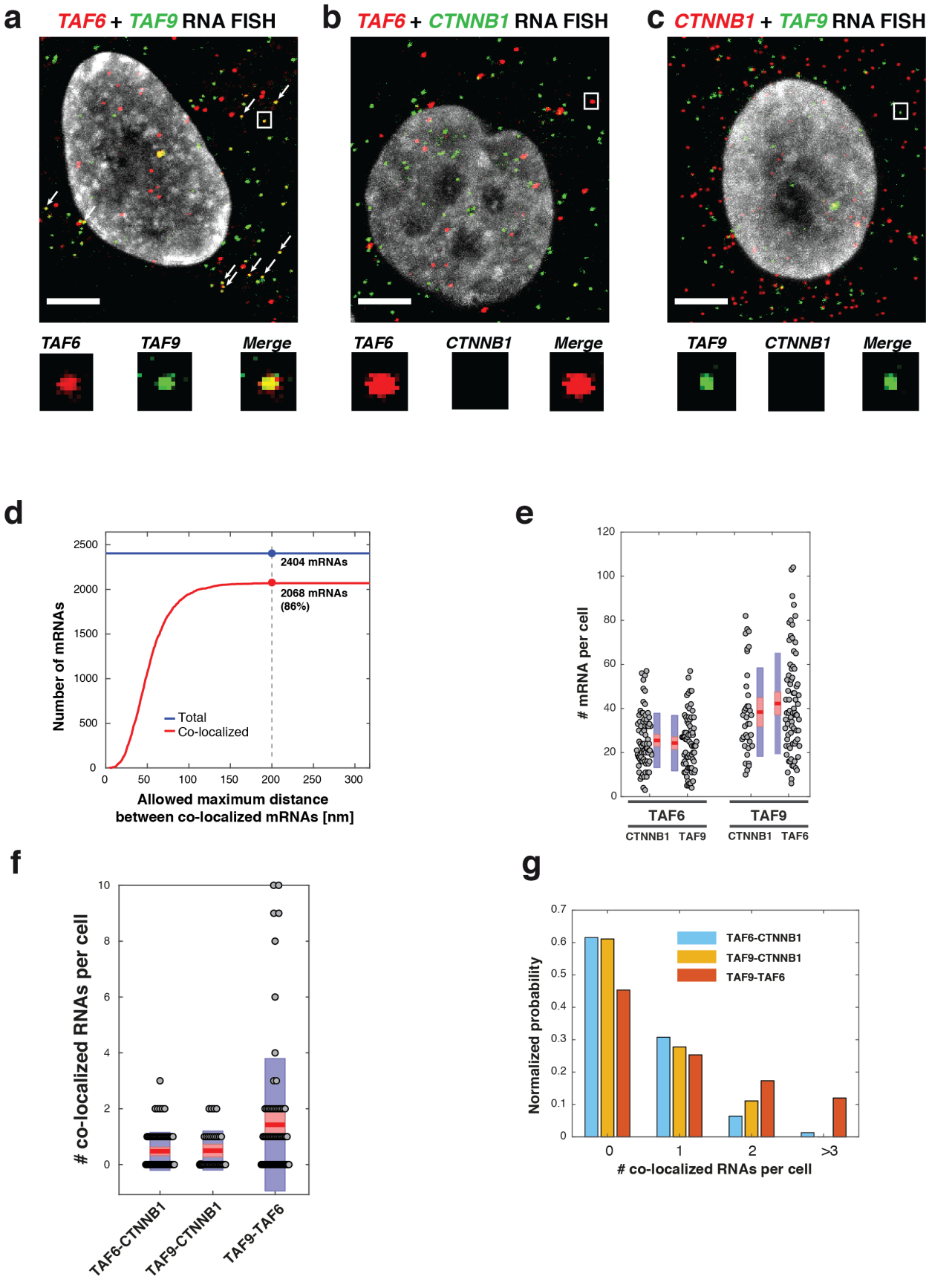
Supplementary Figure 4

(**a**) Representative single colour smiFISH images with Cy3-labelled probes recognizing the indicated mRNAs. Z projections of confocal images are shown. On the right, the cell boundaries are shown in green and the nuclei in blue. Scale bars are 10 µm in each case. RNAs were detected with FISH-quant. N indicates the number of mRNA molecules in the cytoplasm of each individual cell. **(b)** Boxplot showing the comparison of the number (N°) of mRNAs under the conditions tested (as indicated). Each grey dot represents one cell. Red horizontal lines are mean values, 95% confidence interval is shown in pink, and standard deviation in blue. N = number of cells counted. (**c**) IF-smiFISH images of HeLa cells expressing either endogenous (endo) *TAF8* mRNA or TAF10 protein. Labels: red, Cy3-labelled *TAF8* probes; green, Alexa-488 labelled secondary antibody for TAF10 protein; co-localizing spots are indicated with white arrows. A typical cell counterstained with DAPI (grey) is shown. The nuclear signal in the green channel (TAF10 IF) was removed by masking the nucleus and using the "clear" option in ImageJ. Zoom-in regions shown are indicated with a white rectangle. Scale bar (5 µm).



Supplementary Figure 5

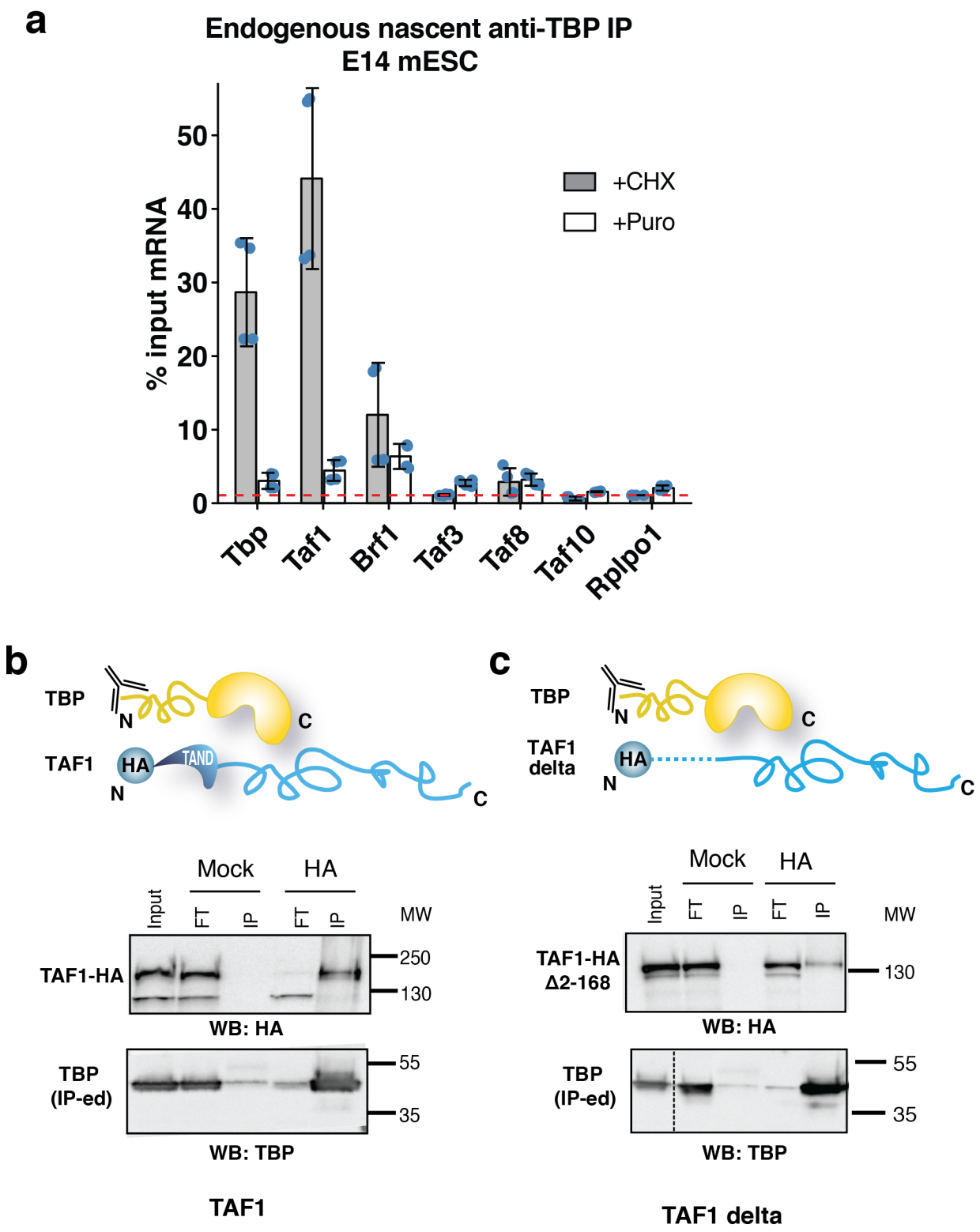
Western blot analyses of RIPs from polysome extracts corresponding to Fig. 6 (**a-d**). HeLa cells transfected with expression constructs coding for HFD domain-swapped TAF10 and TAF8 (**a-b**) and with expression constructs coding for TAF6 and TAF9 (**c-d**). Blots were probed with anti-FLAG M2 and anti-HA antibodies as indicated. Molecular weight (MW) markers are shown in kDa. In = input, FT = flow-through, IP = eluate from IP. Source data provided as a Source Data File.



Supplementary Figure 6

Representative dual colour smiFISH images with (**a**) Cy3-labelled *TAF6* and digoxigenin (DIG)-labelled *TAF9* probes, (**b**) Cy3-labelled *TAF6* and DIG-labelled *CTNNB1* probes, (**c**) Cy3-labelled *CTNNB1* and DIG-labelled TAF9 probes. The DIG labelled probes were then detected with fluorescein-labelled anti-digoxigenin antibodies. Co-localizing spots are indicated with white arrows.

A typical cell recorded in each case and after counterstaining the nucleus with DAPI (grey) is shown. The nuclear signal was masked using the "clear" option in ImageJ. Zoom-in regions shown under every image are indicated with a white rectangle. Scale bar: 5 μm. (**d**) Analysis to determine allowed distance threshold for two RNAs to be considered co-localized. Probe-pool for *TAF9* was split in two and labelled with two different colours. Co-localization between these two channels was determined for different distance threshold. Co-localization reaches a plateau at around 200 nm, with co-localization percentage of 85%. Both, distance threshold and co-localization percentage are in similar to earlier studies (Tsanov et al. 2016). (**e**) Comparison of RNA expression levels per cell for TAF6 and TAF9 in either the respective negative control experiments (against *CTNNB1*), or the *TAF6-TAF9* co-localization experiment. Plot shows that detected expression levels are similar. (**f**) Number of co-localized RNAs per cell. Shows significant increase of co-localization in *TAF6-TAF9* experiments compared to negative controls. In panels (e-f), each grey dot represents one cell. Red horizontal lines are mean values, 95% confidence interval is shown in pink, and standard deviation in blue. (**g**) Same data as in (f) but shown as a histogram. Only in the *TAF6-TAF9* experiments a substantial number of cells have more than 2 co-localized RNAs per cell.



Supplementary Figure 7

(a) RIP-qPCR using TBP antibody from mESC polysome extracts. Values are expressed mean ±S.D. from 2 biological replicates and two technical replicates (represented by blue dots). **(b-c)** Western blot analyses of RIPs with TBP antibody from polysome extracts corresponding to Fig. 7(**c-d**). **(b)** HeLa cells transfected with expression constructs coding for TBP and wild type HA-TAF1, **(c)** HeLa cells transfected with expression constructs coding for TBP and HA-TAF1 with N-terminal deletion of the first 168 amino acids. Molecular weight (MW) markers are shown in kDa. In = input, FT = flow-through, IP = eluate from IP. Source data provided as a Source Data File.

Supplementary Tables

Supplementary Table 1: primers used in this study

gene	organism	name	sequence
TBP	human	hTBP1-F hTBP1-R	TCATACCGTGCTGCTATCT CTCCCTCAAACCAACTTGTC
TAF1	human	hTAF1-F hTAF1-R	TTTGTACCTGCCTTGTTCC GCCCATCTTTCAGTCTCATC
TAF2	human	hTAF2-F hTAF2-R	CATGTGTACCGCCAAAGT GCAGTTGCTTCTGTGTAAATC
TAF3	human	hTAF3-2F hTAF3-2R	GACGACTGCGATGACTGGTA CTTCTTGTTCGCACACTTGG
TAF5	human	hTAF5_F_558 hTAF5_R_627	AGTTGGAAGTGTTGCTGTGG TCCTTGTTGGTTGTAGGCTGAC
TAF6	human	hTAF6_264_352_F hTAF6_264_352_R	CCAGGAGTTCATTCCTTTCC TGATGTCGCTCAGATCAACC
TAF7	human	hTAF7_F_76 hTAF7_R_164	TCTACTGTGAGAAGGGCAGTAC ATTCCATGACGCCCATCAGG
TAF8	human	hTAF8-3F hTAF8-3R	ACAGAGGCAGGGTTTGAGAGT AGACTTGGCACTTCTCCCAAT
TAF9	human	hTAF9_F hTAF9_R	GGAGTTTGCCTTCCGATATG CGCACATCATCTGCATCAAC
TAF10	human/mouse	TAF10s TAF10as	TGCCAATGATGCCCTACAGC AGGGCAGGGGTCAAGTCCTC
SPT7L	human	hSPT7L-3F hSPT7L-3R	AGAATCCCAATGCACCATTC GCCAGCTGAGTTCAGTCACA
BRF1	human	BRF1-2F BRF1-2R	GAGGTGCAGTTCGTGGAGAG CTCTCGACTCCTTCCCCAGA
BTAF1	human	BTAF1-1F BTAF1-1R	CCCTCAGGTCCAACAATGCA GGCCAGACACGTGGAACTAA
*TAF8*_deletion	human	FLAG_F TAF8_HFD_R	GACATCGATTACAAGGATG AGGGTTAGTGGACTGTTTA
*TAF1*_deletion	human	TAF1_HA_F TAF1_HA_R	CTACGACGTGCCCGACTAC GGCAATGATGGAGGGCAAGA
PPIB	human	Cyclopillin B s Cyclopillin B as	CCGAACGCAACATGAAGGTG ACCAAAGATCACCCGGCCTA
GAPDH	human	GAPDH-2F GAPDH-2R	TCGACAGTCAGCCGCATCTTCTTT ACCAAATCCGTTGACTCCGACCTT
ENY2	human	ENY2_fw ENY2_rev	GGAGAAAGAGAACGCCTCAAA AGTGATTTCAGCCACCAAGTCA
GANP	human	GANP_fw GANP_rev	CACGAGCCAGCAGCAGAAGTTC CATCCTGTATCGTCCGACCA
CETN2	human	CENTR2_fw CENTR2_rev	GGACAGGAAAAATGAACTTTGGTGA GGCCACGCGTTTCAGATTTT
CETN3	human	CENTR3_fw	AGAGCCTTGGGGTTTCATGTAA

		CENTR3_rev	TTCTTCATGGGGATCTCTCTTTCC
ATXN7L3	human	ATXN7L3_fw ATXN7L3_rev	CTGGGAATGGGTCGGAACAG CCGAGCCATAGGACCAGTCG
ATXN7	human	ATXN7_fw ATXN7_rev	GCGAAGTCATGGGGCTCTGT TTGAAATGCCTGCGGTTTGA
USP22	human	USP22_fw USP22_rev	TTGCAGATGCCTTTCTGTTG TAGAAAACCGCGAGATGCTT
Tbp	mouse	TBP_mouse_F1 TBP_mouse_R1	AGCAACAAAGACAGCAGCAG CTGTGTGGGTTGCTGAGATG
Taf1	mouse	TAF1_mouse_F1 TAF1_mouse_R1	TGGAGATGGTGATCTTGCAG TCCTCATCATCTTCGCCTTC
Taf3	mouse	TAF3_mouse_F1 TAF3_mouse_R1	TGCTGGCTCCATTTGCAAAG TTTTCTGACCTGGAGAGCTAGC
Taf8	mouse	TAF8_mouse_F1 TAF8_mouse_R1	ATATCAGCACGGACGATTCC GGTTATCGATGACGCTCTCC
Taf10	human/mouse	TAF10s TAF10as	TGCCAATGATGCCCTACAGC AGGGCAGGGGTCAAGTCCTC
Spt7l	mouse	Spt7 NMD F Spt7 NMD R	GGAGCATTGGGATTTTTACAGT TGTGAAGGCTGAAGAGAGTGAA
Brf1	mouse	BRF1_mouse_F1 BRF1_mouse_R1	AGTATCCATGACAGCCTTGAGG TGCAACCAAAAGTGCTGCTC
Rplp0[1]	mouse	RPLP0_mouse_F1 RPLP0_mouse_R1	TTCTGAGTGATGTGCAGCTG GGAGATGTTCAGCATGTTCAGC
Gapdh	mouse	mGapdh_F mGapdh_R	TTCACCACCATGGAGAAGGC CCCTTTTGGCTCCACCCT
Taf8_2	mouse	mTaf8_F mTaf8_R	GAGCTCCTTGCTGACAGAGG GCACTTCTCCCGATTTCTGA
Taf10_2[1]	mouse	mTaf10_F mTaf10_R	CCACGCATAATTCGGCTCAT CCTCCATGGTTAGGTGTACT

Supplementary Table 2: antibodies used in this study

Name	Type	Antigen	Purpose	Dilution	Source
23TA-1H8	mouse monoclonal, ascites	hTAF10 aa 1-20	RIP (human)	2 µl ascites for 20 µl Protein G Dynabeads	Wieczorek *et al*, 1998[2]
6TA-2B11	mouse monoclonal, ascites	hTAF10 aa 89-100	RIP (mouse), protein IP, western blot	2 µl ascites for 20 µl Protein G Dynabeads (RIP, protein IP) 1:1000 (WB)	Mohan *et al*, MCB, 2003[3]
3TF1-3G3	mouse monoclonal, ascites	hTBP aa 1-18	RIP (human, mouse), western blot	2 µl ascites for 20 µl Protein G Dynabeads (RIP) 1:1000 (WB)	Brou *et al*, EMBO J, 1993[4]
3F10	rat monoclonal	HA	western blot	1:500	Roche (Sigma) cat #: 11867423001
anti-FLAG M2	mouse monoclonal, affinity purified	FLAG	western blot	1:2000	Sigma, F1804
#2440	rabbit polyclonal, antigen affinity purified	hTAF1	western blot	1:1000	Choukrallah *et al*, 2011[5]
#2325	rabbit polyclonal, antigen affinity purified	ATXN7L3	western blot	1:1000	Zhao *et al* 2008[6]
ab113295	rabbit polyclonal	hGANP	western blot	1:500	Abcam
SUPT7L	rabbit polyclonal, antigen affinity purified	SUPT7L	western blot	1:5000	Bethyl, A302-803A
#3478	rabbit polyclonal, antigen affinity purified	TAF8	western blot	1:1000	Bardot *et al*, 2017[1]
15-TF2-1D10	mouse monoclonal, ascites	GST	RIP (human, mouse), protein IP	2 µl ascites for 20 µl Protein G Dynabeads	Nagy *et al*., 2010[7]
TUBULIN	mouse monoclonal	TUBULIN	western blot	1:20000	SIGMA-T6557

Supplementary Table 3: plasmids used in this study

Plasmid	*Description*	*Source*
pXJ41-TAF10-Nter-2HA	Eukaryotic expression plasmid containing 2HA-hTAF10 for N-terminal tagging of hTAF10	Jacq et al,1994[8]
pXJ41-TAF8-Nter-3FLAG	Eukaryotic expression plasmid containing 3FLAG-hTAF8 for N-terminal tagging of hTAF8	This study
pXJ41-TAF10-Cter-2HA	Eukaryotic expression plasmid containing 2HA-hTAF10 for C-terminal tagging of hTAF10	This study
pXJ41-TAF8-Cter-3FLAG	Eukaryotic expression plasmid containing 3FLAG-hTAF8 for C-terminal tagging of hTAF8	This study
pXJ41-TAF9-Nter-2HA	Eukaryotic expression plasmid containing 2HA-hTAF9 for N-terminal tagging of hTAF9	This study
pXJ41-TAF6-Nter-3FLAG	Eukaryotic expression plasmid containing 3FLAG-hTAF6 for N-terminal tagging of hTAF6	This study
pXJ41-TAF8-L65F-S57F-Nter-2HA	Eukaryotic expression plasmid containing N-terminally 3FLAG tagged hTAF8 with two point mutations L65F S57F	This study
pXJ41-TAF8HFD-Nter-3FLAG	Eukaryotic expression plasmid containing N-terminally 3FLAG tagged histone fold domain of hTAF8	This study
pXJ41-TAF8HFD60aa-Nter-3FLAG	Eukaryotic expression plasmid containing N-terminally 3FLAG tagged hTAF8 histone fold domain with its 60 aa extension	This study
pXJ41-TAF10(HFD)TAF8-Nter-2HA	Eukaryotic expression plasmid containing N-terminally 2HA tagged hTAF10 with the histone fold domain replaced with that of hTAF8	This study
pXJ41-TAF8(HFD)TAF10-Nter-3FLAG	Eukaryotic expression plasmid containing N-terminally 3FLAG tagged hTAF8 with the histone fold domain replaced with that of hTAF10	This study
pXJ41-TBP	Eukaryotic expression plasmid containing hTBP	May et al,1996[9]
pXJ41-TAF1-Nter-HA	Eukaryotic expression plasmid containing	This study

	1HA-hTAF1 for N-terminal tagging of hTAF1	
pXJ41-ΔTAF1-Nter-HA	Eukaryotic expression plasmid containing N-terminally tagged 1HA-hTAF1, amino acid 2-168 deleted	This study

Supplementary Table 4: smiFISH probes

TAF8 probes

Name	**Probe Sequence (including FLAPY Sequence)**
hTAF8_YDG30_001	TCAGGAGAGGGACTTCTTCCTGCGGATTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_002	TGCAGGACAGAGGTGTTCTCCTTCTCTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_003	CTCCAGAATCCTCCATGCTGATATGATTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_004	AGCAAGGTTCTCTGTGTCTGTCTGTTCATTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_005	CTGCTCCGAGGAATCTGTCTCTTCCATTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_006	TGTTGCATCTCCAGTTCAGACGGAAGAAGTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_007	CTGTCAGGTAGGGGATGGTGAAAGGTCTGTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_008	AGCAATCAATGGAAATGTGCTGACGTCATTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_009	TTGGGCTTCTTCACCGGCCGCAGATATTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_010	TTGAAAAGACTCTGAGTCTCGCCTGTCTTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_011	ACGGGCTCACGGTACGTCGGAGTTTTGTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_012	GTAGGTGTGGGGATCAGGGAACTCAGGATTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_013	TGATTGGTCACCGGAGGAGCAGTGATGTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_014	GGTTATCGATGATGTTCTCCTCCCCATTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_015	CATCCTCTGAGACCGTTTTGCATAAGCTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_016	GGAGAGTGTCCACATTGAAACCCATCTTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_017	ACAAGTGTGACCACGATATCGGACAGTGTTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_018	TGGGTCCTGGCTGTGTGCTCACAGTATTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_019	GCTCTGCAGCATCTCTGTCAGCGTTTTTACACTCGGACCTCGTCGACATGCATT
hTAF8_YDG30_020	ACGGATGCTTTCTCGGCACTCTCAAATTACACTCGGACCTCGTCGACATGCATT

hTAF8_YDG30_021	TGCCTCTGTCAGCAAGGAGCTCACAATTACACTCGGACCTCGTC GACATGCATT
hTAF8_YDG30_022	AGGGTTCTCCTCCGGGCCAGATGATATTACACTCGGACCTCGTC GACATGCATT
hTAF8_YDG30_023	TATCGGCAGGGTTAGTGGACTGTTTACTTCTTACACTCGGACCTC GTCGACATGCATT
hTAF8_YDG30_024	GAAAGTGGGTGATGGGTACATAGGATCTCTTTACACTCGGACCT CGTCGACATGCATT

TAF10 probes

Name	**Probe Sequence (including FLAPY Sequence)**
hTAF10_YDG32_001	CCATACTCGCTGAGGGCAGGGGTCAATTACACTCGGACCTCGT CGACATGCATT
hTAF10_YDG32_002	AAGTTTATTATGAAAACAGGCTGGTGTGGGGATTACACTCGGA CCTCGTCGACATGCATT
hTAF10_YDG32_003	CCTCCATGGTTAGAGTGTACTTGCGGTCTTACACTCGGACCTCG TCGACATGCATT
hTAF10_YDG32_004	CTGCCGGAGGCCGTGCCCTTCATTTTTTACACTCGGACCTCGTC GACATGCATT
hTAF10_YDG32_005	AGTGCTGTAGGGCATCATTGGCAATATCTTTACACTCGGACCTC GTCGACATGCATT
hTAF10_YDG32_006	GATGAATTTCTGGGCAGCTAAGGAGATGAGCTTACACTCGGAC CTCGTCGACATGCATT
hTAF10_YDG32_007	AATTATGCGTGGGTCTGAGGCCTCAAAGCTTACACTCGGACC TCGTCGACATGCATT
hTAF10_YDG32_008	GCACGGTTCAGGTAGTAACCAGTCACTGCTTACACTCGGACC TCGTCGACATGCATT
hTAF10_YDG32_009	CTGGGATCGTAGGCGTGTAATCTTCCAGCTTACACTCGGACCTC GTCGACATGCATT
hTAF10_YDG32_010	CATCAAGAAGTCCACCAAAGGCGTGCTGGTTACACTCGGACCT CGTCGACATGCATT
hTAF10_YDG32_011	ACCACGGGCTTCACGTCTCCGTTGGCTTACACTCGGACCTCGTC GACATGCATT
hTAF10_YDG32_012	CGCTCGGCAGTACGTAAACCCCGTTAGATTACACTCGGACCTC GTCGACATGCATT
hTAF10_YDG32_013	TGGGGACAGATAAGTACATTTAGGTTGGGTGGTTACACTCGGA CCTCGTCGACATGCATT
hTAF10_YDG32_014	CCTTGTTCTCCGCGGCGGTGCTGGAGTTACACTCGGACCTCGTC GACATGCATT

hTAF10_YDG32_015	CAGGTGAAGTAGTGCGGCTTCTTCACATTGATTACACTCGGACC TCGTCGACATGCATT

TAF6 probes

Name	Probe Sequence (including FLAPY Sequence)
TAF6_1	ATGTACTTCTGGACACTCCCAGGACCTTACACTCGGACCT CGTCGACATGCATT
TAF6_2	CCAAGGGACCCGAATTCTGCCCGATATTACACTCGGACCT CGTCGACATGCATT
TAF6_3	CACATGGTCTGCTCCAATCCGGTCAATTTACACTCGGACC TCGTCGACATGCATT
TAF6_4	TGATAACATCGTGTCCCAGCTCAGCCTTACACTCGGACCT CGTCGACATGCATT
TAF6_5	AACTGTCTGCTCACGATGCAGGTCATCTTACACTCGGACC TCGTCGACATGCATT
TAF6_6	AGCTGGAATCAGCTCATGGACGTATTTTTCTTTACACTCG GACCTCGTCGACATGCATT
TAF6_7	AGGTTGTTCTGAACCACGTTCACACGGACTTACACTCGGACCTCGTCGACA TGCATT
TAF6_8	CTGCTGCTCCACAGACAACTCGTGGATTACACTCGGACCTCGTCGACATGC ATT
TAF6_9	CTCAGATCAACCTCCTTCTCCTCATAGTTACACTCGGACCTCGTCGACATGC ATT
TAF6_10	TGAACTCCTGGGCGTGGAAGCCATAGATTACACTCGGACCTCGTCGACATG CATT
TAF6_11	GCTGGGAGCCATTGGCTTTTGGAGTCCTTACACTCGGACCTCGTCGACATG CATT
TAF6_12	TTCTCGTCCACCCAGCTCTTGGTGAATTACACTCGGACCTCGTCGACATGCA TT
TAF6_13	TCTTGGTGATCCGGGACTGGATGTTGTTAGTTTACACTCGGACCTCGTCGAC ATGCATT
TAF6_14	AGTGATTGTCCACATCTGGTCGCAGGCTTACACTCGGACCTCGTCGACATG CATT
TAF6_15	ATAGAGCGTGGGGTTGTCCATCAGCGCTTACACTCGGACCTCGTCGACATG CATT
TAF6_16	CCTCCGAGATAAAGGTACTGAACCGTGGCTTACACTCGGACCTCGTCGACA TGCATT
TAF6_17	CATCTGATACAGTCCAGGGTCCGTGGCTTACACTCGGACCTCGTCGACATG CATT

TAF6_18	GGCTTGGCTGACTTCAGGGGTTCTGTGTTACACTCGGACCTCGTCGACATG CATT
TAF6_19	TTCAGCCTTCTGTTGCTCTTTGGGAGCTTACACTCGGACCTCGTCGACATGC ATT
TAF6_20	CTCGATGCTCAGCCAATGAGCTTTGAGGTTACACTCGGACCTCGTCGACAT GCATT
TAF6_21	GGCTCGACATTCTTTAGCTTCAAGGCGTATTACACTCGGACCTCGTCGACA TGCATT
TAF6_22	CAATGTCACTGGTGGTGAGCTTCTGCCTTACACTCGGACCTCGTCGACATG CATT
TAF6_23	CTGTGCGATCTCTTTGATGCGGTAGCTGTTACACTCGGACCTCGTCGACATG CATT
TAF6_24	CTCATCCGTTAGCAGCTGGCAGGTCTTTACACTCGGACCTCGTCGACATGC ATT

TAF9 probes

Name	**Probe Sequence (including FLAPY Sequence)**
TAF9_1	CCAACACTTAACCGCGGGACTGTTATTCTTTACACTCGGACCTCGTCGACA TGCATT
TAF9_2	CGGCACTGGATTGCCAATCGCACATCTTACACTCGGACCTCGTCGACATGC ATT
TAF9_3	CTGCATCAACAGTAGCTTTCTTAGCATGGTTACACTCGGACCTCGTCGACA TGCATT
TAF9_4	TTTGCATCATCTAGAATTGTGGTCACATATCGTTACACTCGGACCTCGTCG ACATGCATT
TAF9_5	ACTCCATGATATCCGATGATCAGACTTCTTTACACTCGGACCTCGTCGACA TGCATT
TAF9_6	TGAGGGACATGGGAGTCCCTACTTTAGTTGTTACACTCGGACCTCGTCGAC ATGCATT
TAF9_7	ACAGACATGGTCTGTGGGGTTGGTGTGCTTACACTCGGACCTCGTCGACAT GCATT
TAF9_8	AGTGTGGGAGTACTTGGTCTGCTAGTAACTGTTACACTCGGACCTCGTCGA CATGCATT
TAF9_9	GTTGATGCCTTTTTCTGTAAAGATTTCAGCCTTTACACTCGGACCTCGTCGA CATGCATT
TAF9_10	AAGGGGTTTGATTTCTTTGCCTTGCAATATCTTTACACTCGGACCTCGTCG ACATGCATT
TAF9_11	ATTTGTGCCATCATCTGTGCATCTTTCGTTACACTCGGACCTCGTCGACATG CATT

TAF9_12	ATGCTCTTGGGAGAAGCCGTCTTGCCTTACACTCGGACCTCGTCGACATGCATT
TAF9_13	GATTTTGACGCAAGTTCTTTGCCTAGTGTGGTTTACACTCGGACCTCGTCGACATGCATT
TAF9_14	AGACCAAGTATACATGTTACATTCAGCAAGGCTTACACTCGGACCTCGTCGACATGCATT
TAF9_15	TGGATTAATCAGAACATTCTGAACTGCTGAGGTTACACTCGGACCTCGTCGACATGCATT
TAF9_16	GCAGGAATTGAAGCTTTTACAGCTGGAGACTTTACACTCGGACCTCGTCGACATGCATT
TAF9_17	GAAGTAGGCATCTGTACTGTAAACCTTTGACCTTACACTCGGACCTCGTCGACATGCATT
TAF9_18	CAACCTAGGACCTGAATATGGCTTGATCAATTTACACTCGGACCTCGTCGACATGCATT
TAF9_19	AGGCAAACTCCAACATCTGATTTATAACTCTTTTACACTCGGACCTCGTCGACATGCATT
TAF9_20	AGCTAAATCACCCACATTAATGTATTTCAGTCTTACACTCGGACCTCGTCGACATGCATT
TAF9_21	ACTGTTTGTGAAATACTACTTATCACACTGCGTTACACTCGGACCTCGTCGACATGCATT
TAF9_22	CCACCTTCTGCCTTTCCTTTTATTTTTGAGATTACACTCGGACCTCGTCGACATGCATT

CTNNB1 probes[10]

Name	**Probe Sequence (including FLAPY Sequence)**
CTNNB1_P01	CTCATGTTCCATCATGGGGTCCATACCTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P02	GCATCCTGGCCATATCCACCAGAGTGTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P03	TGTTCTGAAGAGAGAGCTGGTCAGCTCAACTTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P04	GCCGTTTCTTGTAATCTTGTGGCTTGTCCTTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P05	AGCTGTGGCTCCCTCAGCTTCAATAGTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P06	TGCAGCTTCCTTGTCCTGAGCAAGTTCATTACACTCGGACCTCGTCGACATGCATT

CTNNB1_P07	GAGCTAGGATGTGAAGGGCTCCGGTACAACTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P08	AAATTGCTGCTGTGTCCCACCCATGGTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P09	GGCCAGTGGGATGGTGGGTGTAAGAGCTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P10	TGGGCCATCTCTGCTTCTTGGTGTCGTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P11	TGATGTCTTCCCTGTCACCAGCCCGATTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P12	GTCCCAAGGAGACCTTCCATCCCTTCTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P13	AGCACCTTCAGCACTCTGCTTGTGGTTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P14	ACCACTAGCCAGTATGATGAGCTTGCTTTTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P15	TTGTTTTGTTGAGCAAGGCAACCATTTTCTGCTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P16	TGGGAAAGGTTATGCAAGGTCCCAGCGGTATTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P17	ATAGCGTGTCTGGAAGCTTCCTTTTTAGAAAGTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P18	TGGTCCTCGTCATTTAGCAGTTTTGTCAGTTCTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P19	ATTGCACGTGTGGCAAGTTCTGCATCATCTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P20	ATGGTTCAGCCAAACGCTGGACATTAGTGGTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P21	GTCCATCAATATCAGCTACTTGTTCTTGAGTGTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P22	CTTGGGAGGTATCCACATCCTCTTCCTTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P23	ATTGCCTTTACCACTCAGAGAAGGAGCTGTTTACACTCGGACCTCGTCGACATGCATT
CTNNB1_P24	GTGGCACCAGAATGGATTCCAGAGTCCAGTTACACTCGGACCTCGTCGACATGCATT

Supplementary References

1.Bardot, P., et al., The TAF10-containing TFIID and SAGA transcriptional complexes are dispensable for early somitogenesis in the mouse embryo. *Development* **2017,** *144* (20), 3808-3818.
2.Wieczorek, E., et al., Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II. *Nature* **1998,** *393* (6681), 187-91.
3.Mohan, W. S., Jr., et al., TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice. *Mol Cell Biol* **2003,** *23* (12), 4307-18.
4.Brou, C., et al., Distinct TFIID complexes mediate the effect of different transcriptional activators. *Embo j* **1993,** *12* (2), 489-99.
5.Choukrallah, M. A., et al., Interconversion between active and inactive TATA-binding protein transcription complexes in the mouse genome. *Nucleic Acids Res* **2012,** *40* (4), 1446-59.
6.Zhao, Y., et al., A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Mol Cell* **2008,** *29* (1), 92-101.
7.Nagy, Z., et al., The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. *Cellular and molecular life sciences : CMLS* **2010,** *67* (4), 611-28.
8.Jacq, X., et al., Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* **1994,** *79* (1), 107-17.
9.May, M., et al., Human TAF(II28) promotes transcriptional stimulation by activation function 2 of the retinoid X receptors. *Embo j* **1996,** *15* (12), 3093-104.
10.Tsanov, N., et al., smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res* **2016,** *44* (22), e165.

2. Unpublished Results

2.1 Co-translational assembly of TFIID submodules

2.1.1 TAF2-TAF8-TAF10 co-translational assembly

Quantitative mass spectrometry analysis revealed the presence of TAF2-TAF8-TAF10 heterotrimer in the cytoplasm of mammalian cells (Trowitzsch et al. 2015, unpublished data of our laboratory). This heterotrimer was shown to regulate the incorporation of TAF2 into TFIID complex. It was previously reported that TAF8 transports NLS-lacking TAF10 to the nucleus via Importin α/β (Soutoglou et al. 2005). Along the same lines a co-import mechanism was put forward for TAF2-TAF8-TAF10 with Importin α/β by *in vitro* studies and *in vivo* knockdown studies (Soutoglou et al. 2005). It was also proposed that the incorporation of this heterotrimer to the already assembled symmetric TFIID-core complex breaks its symmetry, thereby facilitating the incorporation of the remaining TAFs to the complex (Trowitzsch et al. 2015 Introduction section 2.2.4). Thus, we studied the co-translational assembly of TAF2-TAF8-TAF10 heterotrimer by RIP assay (described in Methods section 4). To this end, we carried out RIP assays in TAF2 and TAF8 overexpressing cell lines obtained from Marc Timmer's lab (described in methods section 1.2.4). We observed enrichment of *TAF8* mRNA with TAF2 protein and enrichment of *TAF2* mRNA with TAF8 protein, compared with negative controls tested. Note, however, in both the cases we did not observe any enrichment of *TAF10* mRNA. We also showed that *TAF8* but not *TAF2* mRNA was enriched in TAF10 RIPs (Kamenova, Mukherjee et al. 2019) and full-length TAF10 protein interacts with ribosome-associated nascent TAF8 protein. This order of assembly is established by the localisation of the interaction domains of TAF10 and TAF8. The interaction between TAF10 and TAF8 are structurally and functionally well-characterised unlike that of TAF2-TAF8. It has been put forward that TAF2 binds to several motifs in the C-terminal region of TAF8 (Trowitzsch et al. 2015). But the N-terminal protein interaction between TAF2 and TAF8 cannot be ruled out because our RIP assays reveal a simultaneous co-translational assembly model between TAF2 and TAF8 (***Figure 48 A-B***). In this scenario, it would be interesting to understand the temporal and spatial regulation of full-length TAF10 protein co-translational assembly to nascent TAF8. A model for TAF2-TAF8-TAF10 co-translational assembly is shown in figure 43C.

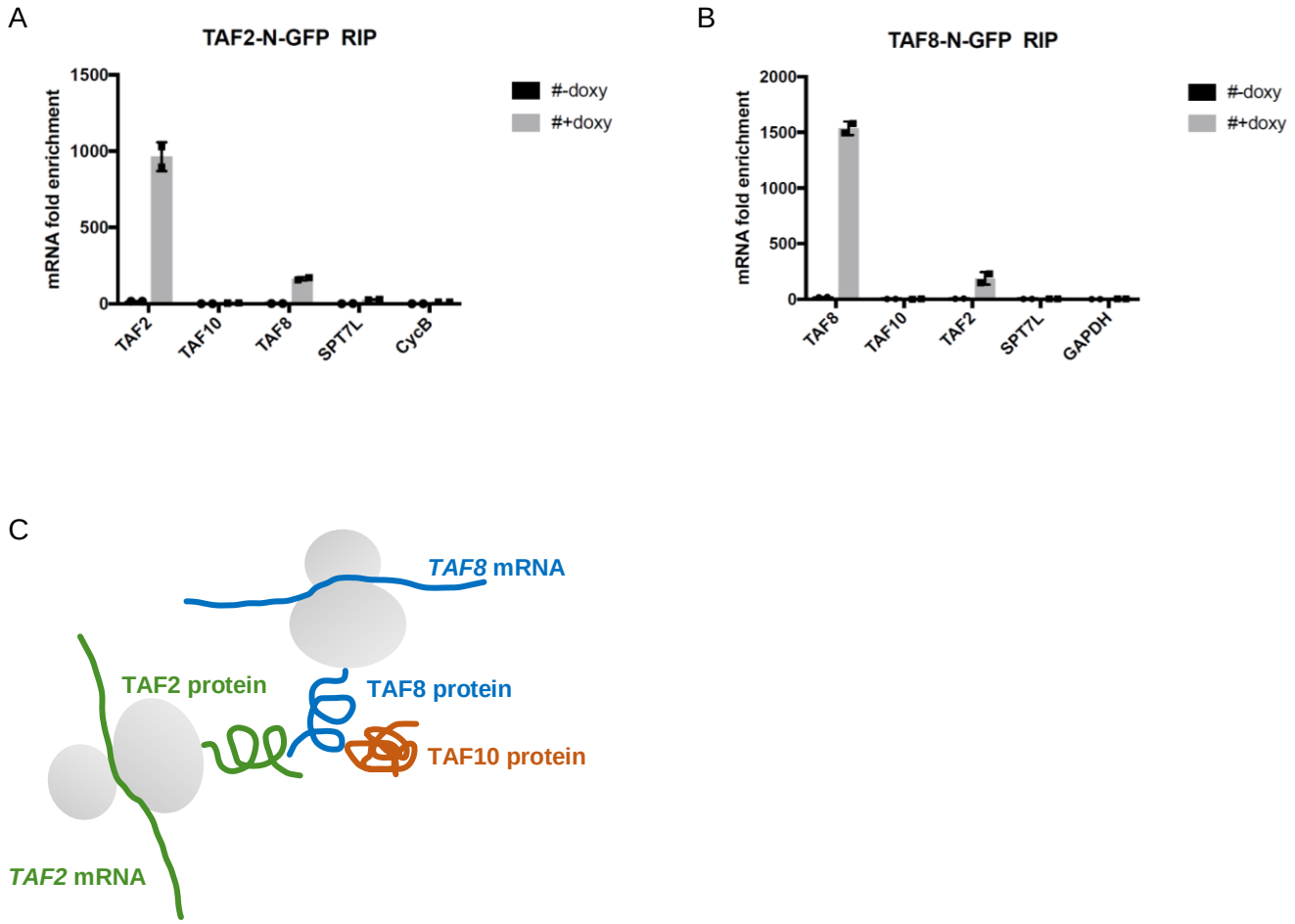


Figure 48: Co-translational assembly of TFIID heterotrimer TAF2-TAF8-TAF10. (A-B) Anti-GFP RIP-qPCR from polysomes of HeLa cells stably expressing GFP-TAF2 (A) and GFP-TAF8 (B). Error bars are ±SD from two technical replicates (represented in black dots). CycB and GAPDH were used as unrelated control mRNAs. #-doxy and #+doxy indicates without or with doxycycline treatment of cells respectively. (C) Schematic representation of TAF2-TAF8-TAF10 co-translation assembly model. Green-TAF2, Blue-TAF8, Brown-TAF10.

2.1.2 Co-translational assembly of TAF5-TAF6-TAF9

Another TFIID heterotrimeric module that exists in the cytoplasm is TAF5-TAF6-TAF9 (Antonova et al. 2018). We showed that TAF6 and TAF9 assemble with each other co-translationally by simultaneous assembly pathway and this is guided by their interaction domain at the N-terminus. However, we did not test the enrichment of TAF5 mRNA in the above case. Instead, we carried out RIP assay with N-terminally GFP-tagged TAF5 (ectopically expressing) cell line and observed the enrichment of *TAF6* mRNA but not *TAF9* mRNA (***Figure 49***). Recent cryo-EM data has shown TAF6-TAF9 heterodimer sandwiched between NTD and WD-40 domains of TAF5 (Antonova et al. 2018). It would be interesting to study the temporal and spatial regulation of association of TAF5 to TAF6-TAF9 co-translationally assembling unit, which forms an essential part of the core-TFIID complex.

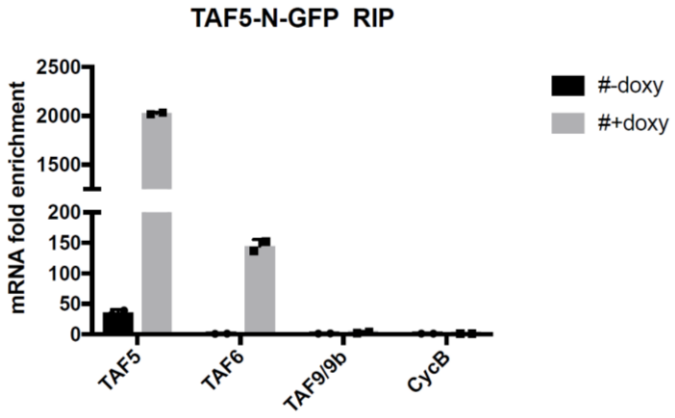


Figure 49: Co-translational assembly of TFIID heterotrimer TAF5-TAF6-TAF9. Anti-GFP RIP-qPCR from polysomes of HeLa cells stably expressing GFP-TAF5. Error bars are ±SD from two technical replicates (represented by black dots). CycB is used as unrelated control mRNA. #-doxy and #+doxy indicate without or with doxycycline treatment of cells respectively.

2.1.3 Co-translational assembly of TBP and its interacting partners in mouse and human cell lines

We showed co-translational assembly of TBP and TAF1 by sequential assembly pathway in HeLa and mES cells (Kamenova, Mukherjee et al. 2019). Next, we wanted to study if co-translational assembly of multisubunit complexes occurs across diverse cell types. So we set out to detect co-translational assembly of TBP (we selected TBP as an example to study TFIID complex) by carrying out TBP RIP assay in HEK293T (human) and NIH3T3 (mouse) cell lines. The data observed in HeLa and mES cells were recapitulated perfectly in both HEK293T and NIH3T3 cell lines. We observed enrichment of *TAF1*, *BRF1* and *BTAF1* mRNAs in TBP RIPs. Hence co-translational assembly is conserved across different cell types.

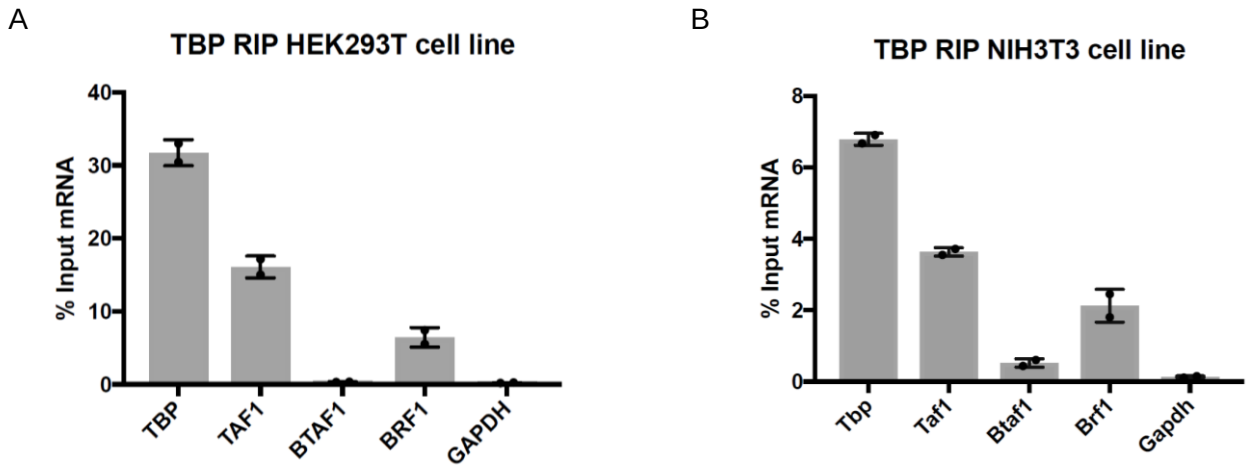


Figure 50: Co-translational assembly of TBP and TAF1 in (A) HEK293T and (B) NIH3T3 cell lines. RIP-qPCR using anti-TBP antibody in HEK293T (A) and NIH3T3 (B) polysome containing extracts. Error bars are ±SD from two technical replicates (represented by black dots). GAPDH is used as unrelated control mRNA.

2.2 Co-translational assembly of SAGA and ATAC complex

Coactivators SAGA complex is organized into separate functional modules: a histoneacetyltransferase (HAT) module, a histone deubiquitinase module (DUB), an activator-binding module and a structural core module. Among them it shares the HAT module with another coactivator ATAC complex. Having studied TFIID co-translational assembly, we set out to study the mechanism in SAGA and ATAC complexes. We carried out RIP assay with one subunit specific to each of the two complexes, TAF6L of SAGA complex and YEATS2 of ATAC complex. TAF6L is a paralogue of TAF6 subunit of TFIID complex and is a subunit of SAGA core structural module. It interacts directly with TAF5L, TADA, TAF9 and TAF10. However, our RIP assay with N-terminally GFP tagged TAF6L did not show significant enrichment of any of the interacting subunits except *TAF9* mRNA, which was also not very high as compared with negative control and other genes tested (Figure 51A). YEATS2 RIP did not yield *NC2beta* or *ZZZ3* mRNAs (Figure 51B). Note, however we did not study enrichment of any HAT module subunits. Although the two subunits studied of each of the two complexes did not yield any of the expected mRNAs, a thorough study needs to be carried out on other subunits to rule out the possibility of only post-translational interaction of the subunits of the complexes.

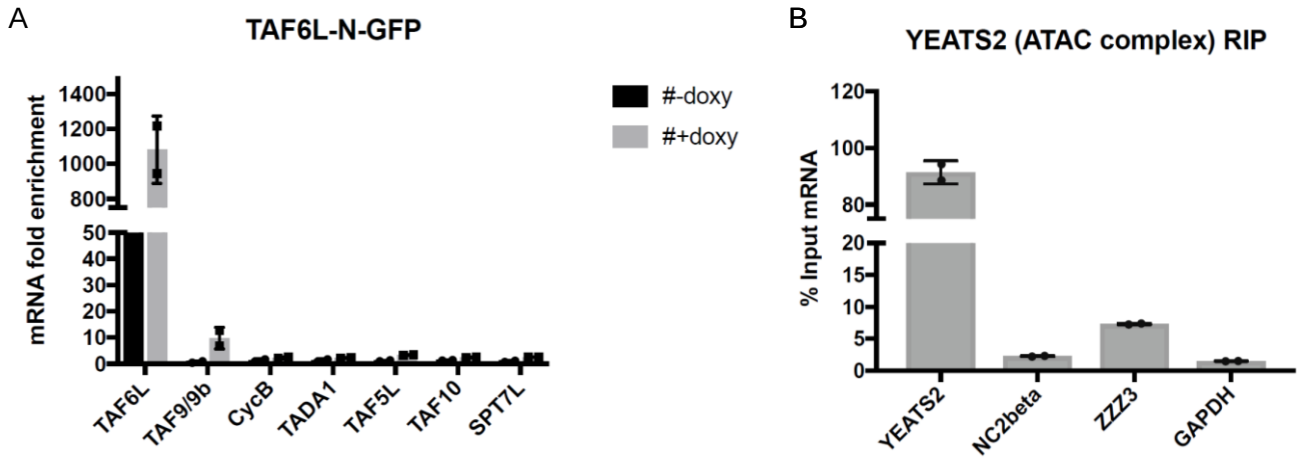


Figure 51: Co-translational assembly of (A) SAGA and (B) ATAC complex. (A) Anti-GFP RIP-qPCR from polysomes of HeLa cells stably expressing GFP-TAF6L. (B) RIP-qPCR using anti-YEATS2 antibody in polysome containing extracts of Hela cells. Error bars are ±SD from two technical replicates (represented by black dots). CycB and GAPDH are used as unrelated control mRNA (A and B respectively). #-doxy and #+doxy indicate without or with doxycycline treatment of cells respectively (A).

2.3 Identification of chaperones guiding co-translational assembly of TFIID complex

In order to identify any chaperone associated with TFIID co-translational assembly, we investigated the proteins associated with TBP and TAF10 polysome extracts from Hela cells by mass spectrometry analysis. Alternatively, we also designed strategies to carry out pulldown of translating *TAF8* mRNA polysomes by MS2-TRAP (MS2-tagged RNA affinity purification) (Yoon and Gorospe 2016) and study the proteins associated with it.

2.3.1 Polysome extract analysis by Western blot

In order to study the nature of the polysome extract (nuclear or cytoplasmic), we compared it with nuclear and cytoplasmic extracts prepared from HeLa cells by established protocols in our laboratory. The detection of more alpha-tubulin and less H3 suggests that our polysome extract is cytoplasmic (Figure 52).

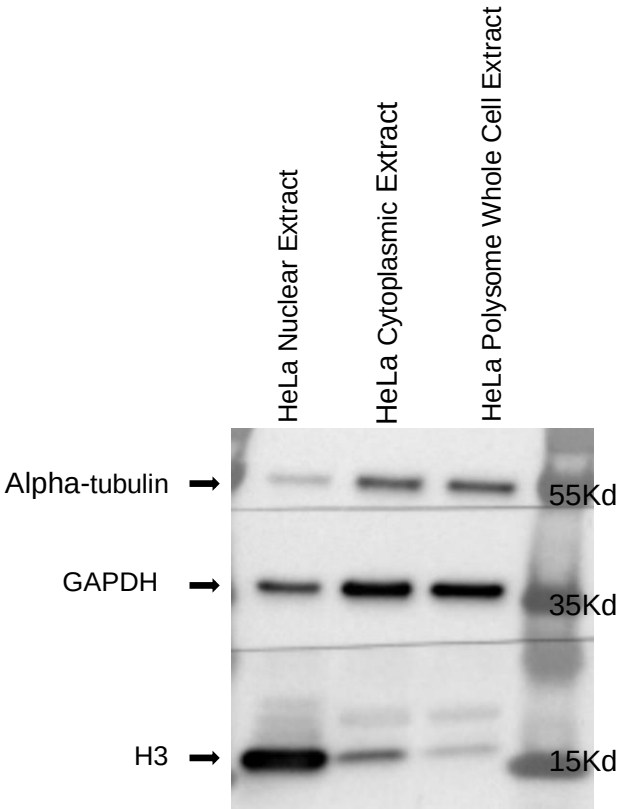


Figure 52: Western Blot analysis of HeLa polysome extract. Western blot analysis of HeLa polysome extract in comparison to HeLa nuclear and cytoplasmic extracts. Blots were probed with anti α-tubulin, anti-GAPDH and anti-H3 antibodies as indicated. Molecular weight (MW) markers are shown in kDa.

2.3.2 Mass spectrometric analysis of TAF10 and TBP-polysome associated factors

Next, we set out to analyse the proteins associated with TAF10 and TBP polysome extracts by mass spectrometry. To purify these complexes, we prepared polysome extract from HeLa after treatment with cycloheximide and immunoprecipitated TAF10 and TBP with antibodies specific to its N-terminal region. GST was used as mock. After elution, the samples were submitted to Orbitrap mass spectrometry. Successful polysome-IP was assessed by Western blot and qPCR analysis (data not shown). The workflow for data analysis is described in (Methods section 16). We argued that a common chaperone might be involved in both TAF10-TAF8 and TBP-TAF1 co-translational assembly and hence for TFIID co-translational assembly. So, we considered only the common proteins enriched in both the IPs. Interestingly only one chaperone, heat shock protein beta-1 or Hsp27 was enriched in both TAF10 and TBP IPs. Several mRNA processing factors were enriched (Figure 53). The presence of mRNA processing factors could be because mRNAs are normally coated with proteins to form the mRNP particles and the factors are removed during the first round of translation (Katahira 2015; Maquat et al. 2010). Nevertheless any specific role of the factors in co-translational assembly cannot be ruled out. Further knockdown and depletion studies are necessary to assess the processing factors' and chaperones' role in TFIID co-translational assembly.

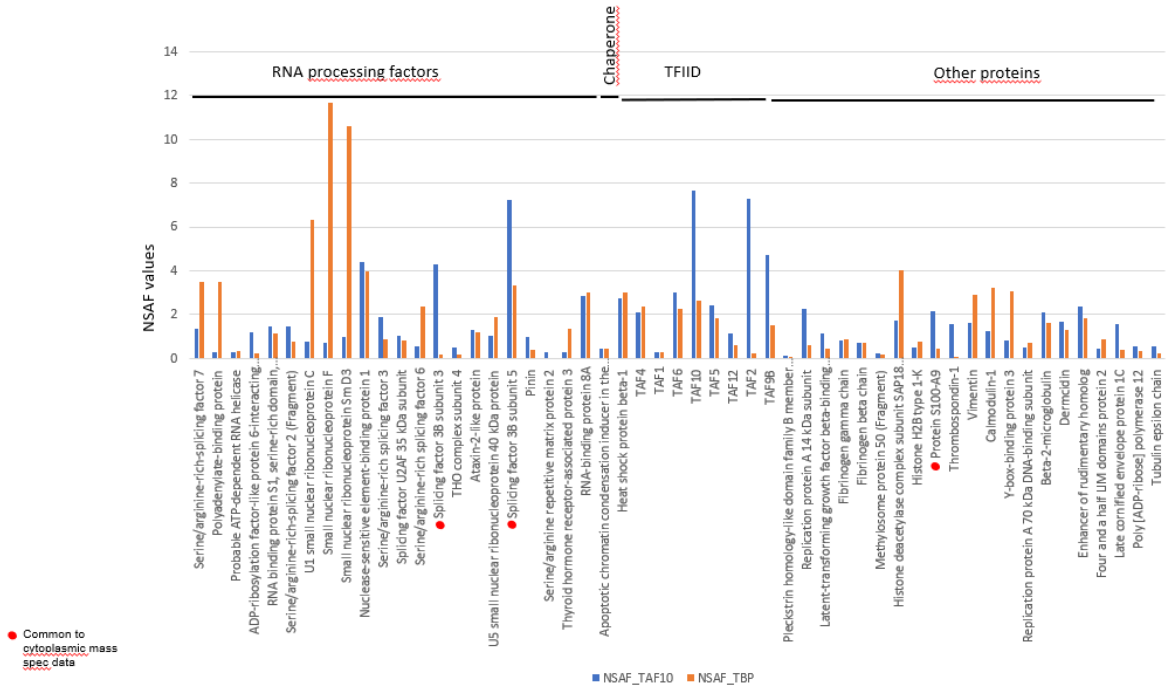


Figure 53: Potential TAF10 and TBP interacting factors from HeLa polysome containing cell cytoplasmic extracts. Common interactors of TAF10 and TBP in cytoplasmic polysome extracts are shown. TAF10 interactors are shown in blue, TBP interactors are shown in orange. X-axis shows the NSAF values (calculation described in methods section). Red dots denote proteins common to HeLa cytoplasmic extract mass spectrometry data.

2.3.3 TAF8 MS2-TRAP and identification of associated factors

The MS2 hairpin sequences of bacteriophage MS2 and its affinity to its binding protein MCP (MS2-coat protein) have long been utilized for studying mRNAs in eukaryotic cells. Tagging an mRNA of interest with MS2 and expressing MCP have been used to visualise and purify ribonucleoprotein (RNP) complexes without introducing purified proteins and RNAs into cells. MS2-TRAP is a technique by which the proteins associated with our mRNA of interest can be studied (Yoon and Gorospe 2016). We introduced 24 MS2 repeats in the 3'UTR of N-terminally FLAG-tagged *TAF8* mRNA and transiently expressed the construct in stably integrated MCP expressing cells (Figure 54A). Next, we studied the proteins associated with it by pulling down MCP bound to MS2-tagged *TAF8* mRNA. Although we could successfully pull-down HA and GFP tagged MCP and *TAF8* mRNA associated with it (Figure 54B-C) but we could not detect nascent TAF8 protein and full-length TAF10 protein by western blot analysis (Figure 54C). Note, however that we could detect FLAG-tagged TAF8 protein in our input sample which rules out the possibility of TAF8 protein expression inhibition by the introduction of MS2 repeats in its 3'UTRs. Nevertheless, a more sensitive detection method like mass spectrometry analysis could be useful to study proteins associated in lower amounts which are not possible by western blot detection.

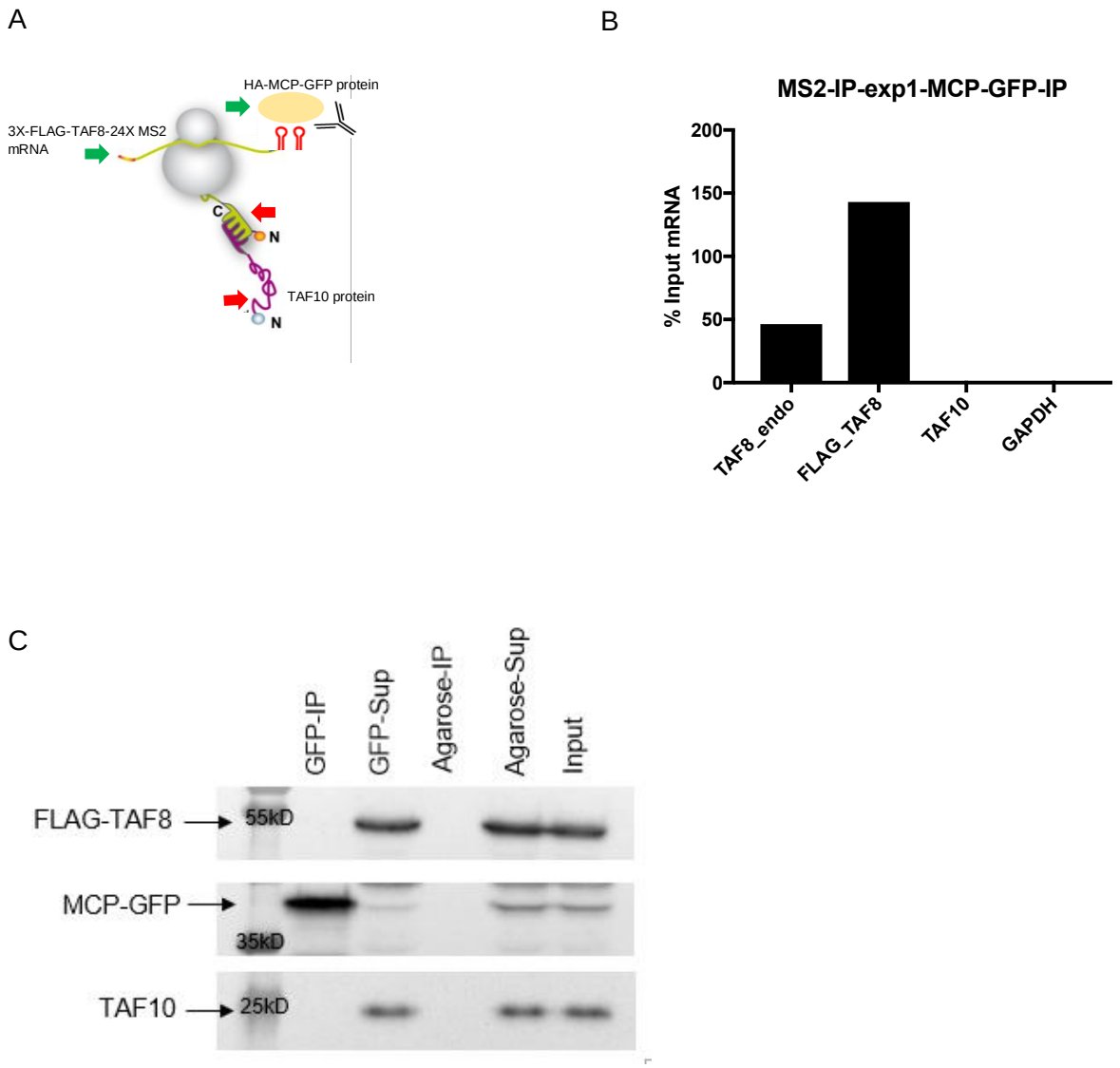


Figure 54: Identification of *TAF8* mRNA associated factors by TAF8-MS2-TRAP. (A) Schematic representation of *TAF8*-MS2-TRAP experiment. Green arrow indicates the factors successfully detected in our experiments and red arrows indicates the factors that were not detected. (B) qPCR analysis of *FLAG-TAF8-MS2* mRNA. TAF8_endo represents primers designed endogenous TAF8 mRNA and FLAG_TAF8 represents primers designed against *FLAG-TAF8-MS2* mRNA. GAPDH is used as negative control. (C) Western blot analysis of GFP immunoprecipitates from HeLa polysome extract transfected with FLAG-TAF8-MS2 construct. Blots were probed with anti TAF10, anti-FLAG and anti-GFP antibodies as indicated. Molecular weight (MW) markers are shown in kDa. Sup indicates IP supernatant.

DISCUSSION

Discussion

1. General transcription factor TFIID assembles co-translationally

Our results have demonstrated a novel paradigm shifting mechanism for the assembly of TFIID subunits. It has been reported earlier by our laboratory as well as by our collaborators that the huge 14-subunit containing TFIID complex is assembled in submodules (made up of two to three subunits) in the cytoplasm of cells. (Trowitzsch et al. 2015; Antonova et al. 2018; Gupta et al. 2017) . Quantitative mass spectrometry analysis has earlier shown the presence of heterotrimeric TFIID submodules TAF2-TAF8-TAF10 and TAF5-TAF6-TAF9 in the cytoplasm (Trowitzsch et al. 2015; Antonova et al. 2018). We show that these submodules are assembled co-translationally, instead of finding their interacting partners post-translationally in the cytoplasm. We hypothesize that this form of assembly is important for the cell to prevent non-specific interactions and aggregations in the crowded environment of the cell cytoplasm. Both the two heterotrimers mentioned above contain HFD pair TAF8-TAF10 and TAF6-TAF9 respectively. Our biochemical analyses provide evidence for co-translational assembly of the two above HFD pairs. We also provide direct evidence of co-localization of TAF10 protein and *TAF8* mRNA by imaging experiments, lending further support to our biochemical observations. Additionally, we show co-translational assembly of a non-HFD pair in TFIID complex, TBP-TAF1, thereby ruling out the possibility that this mechanism is specific to HFD pairs of TFIID complex. Co-translational assembly has been shown to be a prevalent mechanism in assembling a wide range of complexes in yeast (Duncan and Mata 2011; Shiber et al. 2018). There have been some reports about co-translational assembly of proteins to specific subcellular structure (Chang et al. 2006; Singer 1992; Képès 1996; Liu et al. 2016). However, no report so far focussed on the study of co-translational assembly of soluble protein complexes in the cytoplasm and nucleus of mammalian cells. We and Panasenko et al. 2019 reported parallelly for the first time co-translational assembly of soluble multi-subunit protein complexes in mammalian cells (Kamenova, Mukherjee et al. 2019; Panasenko et al. 2019). Nevertheless, the formation of TFIID complex co-translationally in building blocks might serve as a regulatory step in its assembly thereby serving as a quality control check of unwanted interactions. No solid evidence of chaperone involvement in TFIID assembly have been reported so far. However, a recent study points towards the role of chaperonin CCT complex (Antonova et al.

2018)in TAF5 folding and its assembly to TAF6-TAF9 heterodimer. But we did not observe any significant enrichment of CCT complex in our preliminary mass spectrometry results from TAF10 and TBP immunoprecipitated polysome extracts, (Figure 53) thereby giving less credence to the involvement of this complex in TFIID co-translational assembly. Alternately, the association of this chaperone/ chaperonin with TFIID complex could be very transient and other methods might be necessary to study their association with co-translationally assembling proteins in polysomes.

2. Position of dimerization domain drives co-translationally assembly

One of the intriguing questions about co-translational assembly is the mechanism behind their directionality or order of assembly. We put forward two types of models of co-translational assembly: sequential and simultaneous. Sequential assembly occurs where a fully translated protein finds its interacting partner still attached to the ribosome while in the case of simultaneous assembly, the two protein partners engage with each other while still attached to the ribosome. In this respect, we investigated three protein heterodimers, each interacting with their partners through domains localised at either N- or C- terminus as shown below. We observed sequential assembly for TAF10-TAF8 and TBP-TAF1 protein pair due to their C-N domain localisation and simultaneous assembly for TAF6-TAF9 protein pair due to their N-N domain localisation (Figure 2a-b, 6c-d, 7c-d of (Kamenova, Mukherjee et al. 2019)). Additionally, we also carried out an experiment where we swapped the HFDs of TAF10 and TAF8 to observe the effects on co-translational assembly. Indeed, it was the position of the HFD and not the sequence that was driving the assembly, since swapping the domain sequences and keeping the position unchanged did not affect the sequential co-translational assembly of TAF10 and TAF8 (Figure 6a-b of (Kamenova, Mukherjee et al. 2019)). Moreover, shifting the HFD of TAF8 to the C-terminus abolished co-translational assembly of TAF10 and TAF8 (data not shown). Taken together, these results indicate that the position of the interaction domain is solely responsible for regulating the directionality of co-translational assembly of interaction partner proteins. However, it is still unknown how the subunits are brought in close proximity to each other to enable co-translational assembly. Our single molecule RNA FISH experiments point towards the fact that two mRNAs are in close proximity only when they assemble via simultaneous assembly model (Supplementary Figure 6 of (Kamenova, Mukherjee et al. 2019)), thereby ruling out the fact that all co-

translationally assembling protein encoding mRNAs are tethered together in the cytoplasm by RNA-binding protein or in any RNA granule. Our results with cDNA constructs also ruled out the fact that 3'UTRs might be guide their localisation and thus co-translational assembly in the cell as suggested before (Lawrence and Singer 1986; Mayr 2016). However, the dynamics of the process could be fast enough to be undetectable with the current techniques. It would be interesting to carry out live imaging experiments in this respect (Haimovich et al. 2017; Chang et al. 2006).

Few reports across different domains of life point towards the presence of specific factors or chaperones (Shieh et al. 2015; Kassem et al. 2017; Panasenko et al. 2019) or RNA granules (Panasenko et al. 2019) facilitating co-translational assembly. In contrast a study carried out in yeast on many different complexes showed that nine out of the twelve complexes studied by them assembled co-translationally and the three complexes that assembled post-translationally were assisted by dedicated chaperones or inhibitors that prevented their non-specific aggregation (Shiber et al. 2018). The authors hypothesised that co-translational assembly serves a similar purpose as chaperones for preventing protein misfolding and random aggregation. They also showed that the Hsp70 chaperone (Ssb in yeast) associates with the nascent proteins until the full-length partner protein is available to bind to it co-translationally (Shiber et al. 2018). Nevertheless, a conclusive study needs to be carried out to shed more light on the factors guiding co-translational assembly specifically in mammals.

3. Co-translational assembly is essential for the cell

While several evidences point towards the prevalence of co-translational assembly of a diverse range of protein complexes, an open question is whether this form of assembly is essential for the building up of complexes in the cell. With existing technologies, this question is difficult to answer directly as no method has been discovered so far to selectively block co-translational assembly. Testing the functionality of a protein complex would only shed light on the correct assembly of a protein complex, co-translationally or post-translationally. Instead we hypothesized that co-translational assembly might be necessary for the stability of ribosome-assisted nascent peptides. Supporting this theory, when we knocked out Taf10 protein from mESCs, we observed decrease in Taf8 mRNA and protein levels. Knocking out

Taf8 instead did not affect Taf10 levels (Figure 4 of (Kamenova et al. 2019)). Hence, the co-translationally assembling partner protein is required to bind and stabilise the nascent partner synthesising from the ribosome. This was also supported by another study in yeast where the authors studied different cytoplasmic complexes and the absence of the full-length protein partners was making the nascent proteins associated with the ribosome prone to aggregation and degradation (Shiber et al. 2018). In parallel, another study observed ribosome pausing following the emergence of the interaction domain out of the ribosome (Panasenko et al. 2019). Together, this study suggests that following the synthesis and emergence of the interaction domain out of the ribosomes, there is a brief pause to allow co-translational assembly with its partner. The authors suggested the role of a DP codon pair present upstream of the interaction domain of the studied protein pair in stalling the ribosome (Panasenko et al. 2019). However, it would be interesting to study the ribosome pause and the factors driving it across a wide variety of protein complexes. Nevertheless, absence of co-translational interaction might lead to prolonged ribosome pause and subsequent de-stabilisation and degradation of the partner mRNA and nascent protein by ribosome quality control machinery.

4. Co-translational assembly is a general mechanism of complex assembly

Over the years, studies on many different protein complexes across different kingdoms of life point towards the fact that co-translational assembly is a prevalent form of mechanism for complex formation. Earlier reports in both budding and fission yeast (Duncan and Mata 2011; Kassem et al. 2017; Shiber et al. 2018) showed this form of assembly in a wide range of complexes. We and Panasenko et al. 2019 showed that this is also true for mammalian cells. While Panasenko et al. 2019 studied proteasome complex, we studied three functionally distinct transcription complexes (TFIID, SAGA, TREX-2). Additionally, despite the fact that TAF10 is shared between TFIID and SAGA, but it only co-translationally assembles to TAF8, which is a subunit of TFIID and not to SPT7L, which is a component of SAGA complex (Figure 1d-e of (Kamenova, Mukherjee et al. 2019)). Reducing the level of TAF8 also did not facilitate co-translational assembly of TAF10 and SPT7L (data not shown). ENY2 on the other

hand co-translationally assembles to subunits of both SAGA and TREX-2 complex (Figure 8b of (Kamenova, Mukherjee et al. 2019). This points towards the existence of a mechanism in the cell that guides systematic assembly.

If this form of assembly is true for the majority of complexes, then a simple RIP assay followed by RNA detection can also be used to discover novel interacting protein partners of a given protein in a stable complex. There have been no reports so far on co-translational assembly of two proteins associating transiently to carry out functions in the cell. Hence, identification of a novel partner would most likely mean they are part of a stable complex.

PERSPECTIVES

Perspectives

1. Comprehensive study of the assembly of TFIID submodules.

The multisubunit complex TFIID assembles in submodules existing in the cytoplasm of cells. Previous results showed the presence of at least two TFIID heterotrimers in the cytoplasm (TAF2-TAF8-TAF10 (Trowitzsch et al. 2015) and TAF5-TAF6-TAF9 (Antonova et al. 2018)). In the present study, we extensively investigated the co-translational assembly of three cytoplasmic heterodimers, TAF10-TAF8, TBP-TAF1 and TAF6-TAF9. Our unpublished results also show co-translationally assembly of TAF2 and TAF8 (Figure 48). It would be interesting to understand the dynamics of co-translational assembly of higher order complexes. To this end, a comprehensive RIP assay (described in Methods Section 4) of all TFIID subunits and RNA analysis of all TAFs would shed light on how the building blocks of TFIID are co-translationally built. It would also be interesting to study ribosomal pausing and destabilisation of individual TAFs following the translation of their interaction domains. However, carrying out RIP assays with all TAFs might be limited by the unavailability of suitable antibodies against N-terminal region of all TAFs to immunoprecipitate (IP) nascent proteins. TAF cell lines with a tag (for example, GFP) at its N-terminus generated by Flp-FRT recombination system (Marc Timmers' lab) is an attractive alternative to pulldown all TAFs without any difference in efficiency between each IP (Antonova et al. 2018). However, conditions might be optimised to induce overexpression of TAFs to a level comparable to the endogenous amounts. The best method to address this question is generating CRISPR-Cas9 knockin TAF cell lines with N-terminal tags (Koch et al. 2018). Our published results show the co-localisation of *TAF8* mRNA and TAF10 protein in the cytoplasm. It would be interesting to see the localisation of *TAF2* mRNA and protein with respect to TAF8 and TAF10. In this respect multiplexed error robust FISH (MERFISH) is a useful technique (Chen et al. 2015) to study the localisation and copy number of all TAF mRNAs in the cytoplasm simultaneously. MERFISH is an imaging method capable of simultaneously measuring the copy number and spatial distribution of hundreds to thousands of RNA species in single cells (Chen et al. 2015). Another interesting factor is whether the number of mRNA molecules play a role in guiding co-translational assembly. We observed a striking difference in the number of endogenous *TAF8* and *TAF10* mRNA molecules, with the number of *TAF8* mRNAs being much less than that of *TAF10.* But so far, our results did not indicate an effect

on co-translational assembly upon changing the number *TAF8* molecules (Figure 48B). Moreover, the number of *TAF6* and *TAF9* mRNAs are comparable in the cell and they are also assembled co-translationally. Thus, we can propose that the copy number of mRNAs does not play any role in guiding their assembly. A comprehensive study of the copy number of all TAFs is necessary to have a comparable account of all the TAFs and this is possible by MER FISH (Chen et al. 2015).

2. Is co-translational assembly guided by chaperone(s)?

Most of the studies on co-translational assembly thus far have mainly been done in yeast. Only one report has addressed the mechanism in bacteria and two reports in mammals. Nevertheless, as mentioned in the discussion, no report so far has shown any conclusive evidence on the involvement on chaperones in co-translational assembly. In order to find out the involvement of chaperones in TFIID co-translational assembly, we prepared polysome containing cytoplasmic extracts from HeLa cells and carried out mass spectrometry analysis of TAF10 and TBP immunoprecipitated extracts (Figure 53). We observed enrichment of some chaperones like Hsp70 and Hsp90 in the TAF10-IPed mass spec results. Next, we compared the results to detect a common chaperone that might be playing in the co-translational assembly of TAF10, TAF8 and TBP, TAF1 and therefore TFIID assembly. Only the small heat shock protein HSPB1 (heat shock protein beta-1 or Hsp27) was enriched to some extent in both the cases. Though several roles of HSPB1 in different biological processes have been addressed (Rogalla et al. 1999; Holmgren et al. 2013; Kostenko et al. 2009; Almeida-Souza et al. 2010), there is no evidence so far on its role in protein assembly. siRNA knockdown and its effect on TFIID co-translational assembly would throw some light on its role. Further results employing alternate techniques is necessary to conclusively show the role of an identified chaperone in co-translational assembly. For example, pulling down the mRNA in question and then studying the associated proteins with it can be an alternate technique to identify any associated chaperone. To this end, pulling down the required mRNA with either biotinylated oligos or tagging the mRNA with MS2 and then pulling it down by MCP (MS2 binding protein) associated with it could also help to study the associated factors/chaperones. However, it cannot be ruled out that the association with chaperones can be transient and hence may not be possible to detect by the standard techniques described above. A well-suited method to study transient interactions is proximity-dependent biotin identification (BioID) which

enables the identification of proteins that are in the same subcellular neighborhood to the protein of interest in living cells. The protein of interest is fused to BirA* (mutant form of the biotin ligase enzyme BirA) which is capable of promiscuously biotinylating proximal proteins existing in the near vicinity of the protein of interest, regardless of whether they are directly or indirectly interacting with the fusion protein. The covalent modification of the nearby proteins by biotin allows them to be purified by streptavidin affinity purification and subsequent identification by mass spectrometry (Kim and Roux 2016). Another straightforward but lengthy way to identify co-translation assembly associated chaperones is to knockdown each of the major ribosome- associated or downstream chaperones and study their effect on co-translational assembly.

3. Mechanism of subunit distribution between two complexes?

An intriguing observation that came out from our RIP assays is that co-translational assembly is specific to certain protein heterodimers. As mentioned in Introduction (Section 2.1), TAF10 is a subunit of both general transcription factor TFIID and general coactivator SAGA complex. Our data suggests that TAF10 assembles co-translationally with TAF8 which is part of the TFIID complex but not with SPT7L which is part of the SAGA complex. Additionally, knocking down TAF8 does not enable TAF10 to interact co-translationally with SPT7L (data not shown). It would be very interesting to understand the mechanism guiding the distribution of common subunits into different complexes. We speculate that alternate 3'UTR isoforms of mRNAs of common protein subunits might exist, consisting of the same coding sequence but differing in 3' UTR lengths. More than half of the human and mouse genes generate alternative mRNA isoforms that differ in their 3'UTRs but encode proteins with identical amino acid sequences. This is exemplified by the protein CD47 existing in two different isoforms CD47-Long UTR (CD47-LU) and CD47-Short UTR (CD47-SU). Due to the differences in the length of 3'UTR, CD47-LU and CD47-SU reside in different cellular compartments (Berkovits and Mayr 2015). The difference in the length of the 3'UTR is due to the presence of multiple poly(A) sites in the pre-mRNA (Figure 35). We hypothesize that TAF10, TBP, ENY2 (each of these proteins are shared between different complexes) each exist in two/three different 3'UTR isoforms in the cell, with one isoform assembling to one complex and the other going to another complex. To this end, we analyzed the published 3' end sequencing data (Lianoglou et al. 2013) (for our genes of interest) which is normally used to detect the presence of 3'UTR

isoforms in the cell. We observed two distinct peaks for ENY2 and TBP, which point towards the possibility of the presence of two different isoforms of ENY2 and TBP. But unexpectedly, for TAF10, I observed only one peak. Further studies of ENY2 and TBP would be necessary to explain the presence of two 3'UTR isoforms. Northern blot analysis would provide a definitive answer to the existence of mRNA isoforms of our proteins of interest.

4. What is the mechanism behind *TAF8* mRNA degradation in the absence of its co-translationally assembling partner TAF10?

We observed significant reduction in the level of *TAF8* mRNA in the absence of TAF10 protein (Figure 4a,c of (Kamenova et al. 2019)).But the exact mechanism of *TAF8* mRNA degradation still remains unknown. Ribosome quality control mechanisms exist in the cell that degrade translationally defective mRNAs and nascent peptides (Introduction Section 6). It has been reported that cells employ no go decay (NGD) pathway to degrade translation elongation stalled ribosome complexes (Introduction Section 6). Knocking down key players of NGD pathway and studying its effect on TAF8 mRNA degradation would throw some light into the mechanism. Nevertheless, it is very interesting to understand the mechanism by which the cell senses the absence of co-translational assembly and subsequently degrades the stalled polysome complex and further studies with several alternate strategies are necessary to explain the mechanism completely.

CONCLUSIONS

Conclusions

We studied the co-translational assembly mechanism of histone fold domain (HFD) pairs TAF10-TAF8 and TAF6-TAF9 and non-HFD pair TBP-TAF1 of general transcription factor TFIID. Nascent and full-length protein RNA Immunoprecipitation (RIP), mutation and domain swapping approaches show that the position of the interaction domain drives the mechanism by which two proteins are assembled co-translationally. If two proteins interact with each other through N- and C-terminal domains respectively, they assemble co-translationally by sequential model (fully folded protein partner with C-terminal interaction domain binds to the nascent protein partner with N-terminal interaction domain). N-N terminal interacting protein partners assemble co-translationally by simultaneous model (where two nascent protein partners engage with each other while they are still attached to the ribosome). Disrupting the interaction domain by point mutation or deletion abolished co-translational assembly. Additionally, swapping the positions of the interaction domains reversed the order of sequential assembly, thereby lending further support to the role of domain position in co-translational assembly. Co-translational assembly would also require the co-localisation of assembling proteins and their respective encoding mRNAs in the cytoplasm depending on the assembly model. Hence, we supported our observations by imaging approaches (smFISH and IF-smFISH) which showed the co-localisation of *TAF8* mRNA and TAF10 protein following sequential assembly and co-localisation of *TAF6* and *TAF9* mRNAs following simultaneous assembly in the cytoplasm. Interestingly absence of the fully folded co-translationally partner (TAF10) reduced the level of partner protein (TAF8) and its encoding mRNA (*TAF8*). Finally, we studied two other transcription complexes (SAGA and TREX-2) and showed that co-translational assembly is a general mechanism and is true for most, if not all, complexes in mammalian cells. We published these results together recently in Nature Communications (Kamenova, Mukherjee et al., 2019). Following the publication, we initiated a systematic study on the co-translational assembly of all TFIID subunits and some of SAGA and ATAC complex subunits. The project is still in its preliminary stage and needs more evidence to develop a conclusive model. In parallel, we carried out mass spectrometry analysis to identify any chaperone(s) associated with TFIID subunits undergoing co-translational assembly. Although we detected enrichment of a small heat shock protein (HSPβ1) chaperone in both TAF10 and TBP RNA

immunoprecipitation from polysome extracts, further knockdown experiments are needed to confirm its role in co-translational assembly of TFIID complex.

METHODS

1. Cell lines and Cell culture

HeLa cells (ATCC® CCL-2TM) grown on adherent plates were obtained from the IGBMC cell culture facility and cultured in a 37°C humidified/5% CO2 incubator. Culture media consisted of Dulbecco's modified Eagle's medium (DMEM), supplemented with 1 g/l glucose, 5% fetal calf serum (FCS), and 40 µg/ml Gentamycin. The GFP-TAF1 cell line was generated by transferring full length human TAF1 fused at its N-terminus to EGFP into HeLa Flp-In/T-REx cells following procedures described in (Nuland et al. 2013). E14 mouse embryonic stem cells [mESCs, ES Parental cell line E14Tg2a.4, obtained from Mutant Mouse Resource and Research Center (MMRRC), Citation ID:RRID:MMRRC_015890- UCD] at passage 29-31 were obtained from the IGBMC cell culture facility and cultured on gelatinised plates in feeder-free conditions in KnockOut DMEM (Gibco) supplemented with the following: 20 mM L-glutamine, Pen/Strep, 100 µM non-essential amino acids, 100 µM β-mercaptoethanol, N-2 supplement, B-27 supplement, 1000 U/ml LIF (Millipore), 15% ESQ FBS (Gibco) and 2i (3 µM CHIR99021, 1 µM PD0325901, Axon MedChem). Cells were passaged approximately every 3 days. The EGFP-ENY2 HeLa cell line was generated in our laboratory by D. Umlauf (Umlauf et al. 2013) and maintained at 37 °C in DMEM (1 g/l glucose), 10% FCS and 40 µg/ml Gentamycin (Umlauf et al. 2013). The Dox-inducible hTAF10 expression system in Taf10−/− mouse F9 embryonal carcinoma cells was generated in our laboratory by E. Scheer (Metzger et al. 1999). Cells were cultured at 37 °C with 7% CO2 in gelatinised plates in a culture media consisting of DMEM (4.5 g/l glucose), 10% FCS, 40 µg/ml Gentamycin in the presence of doxycycline (Sigma).

2. Cell treatments

For polysome-IP assays, 10 or 15 cm plates were treated with cycloheximide (100 l final) or puromycin (50 µg/ml final) and returned to the 37 °C incubator for 15 or 30 min, respectively. Doxycycline inducible GFP-TAF1 cell lines were treated with doxycycline (1 µg/ml final (Sigma)) overnight at 37 °C (incubator) and dox-inducible Taf10−/− mouse F9 embryonal carcinoma cells were treated with doxycycline (1 µg/ml final) for 1 or 2 days for TAF10 re-expression following depletion of the same.

3. Transfections

Transfections were performed on ~90% confluent cells in 10 cm plates in antibiotic-free media using Lipofectamine 2000 (Thermo Fisher Scientific) and 3 µg plasmid DNA, according to the manufacturer's instructions. The medium was replaced with fresh medium containing gentamycin ~5–6 h post transfection and cells were harvested 24 h later. A descriptive summary of the plasmids used in this study is presented in Table 3 of Supplementary file of our paper ***(Kamenova, Mukherjee et al. 2019; Results Section 1).***

4. Preparation of polysome-containing extracts and RNA Immunoprecipitation (RIP)

Polysome-containing extracts were prepared from adherent cells harvested at ~90% confluence by adapting a method for the isolation of ribosomes translating cell type-specific RNAs (Heiman et al. 2014). Briefly, 10 cm plates were treated with cycloheximide (100 µg/ml final) or puromycin (50 µg/ml final) and returned to the 37 °C incubator for 15 or 30 min, respectively. Subsequently, plates were placed on ice, washed twice with ice-cold PBS and scraped in 500 µl lysis buffer (20mM HEPES KOH pH 7.5, 150mM KCl, 10mM MgCl2 and 0.5% (vol/vol) NP-40), supplemented with complete EDTA-free protease inhibitor cocktail (Roche), 0.5 µM DTT, 40 U/ml RNasin (Promega), and cycloheximide or puromycin as needed. Extracts were prepared by homogenizing cells by 10 strokes of a B-type dounce and centrifugation at 17,000 × g. Clarified extracts were used to start immunoprecipitations, after saving 10% total RNA for input measurement. For TAF10 and TBP IPs, 20 µl of Protein G Dynabeads (ThermoFisher Scientific) were equilibrated by washing three times in lysis buffer, resuspended in 400 µl of lysis buffer and 2 µl of antibody, and incubated for 1 h at room temperature with end-to-end mixing. Beads were washed twice with IP500 buffer (20mM Tris-HCl, pH 7.5, 500mM KCl, 10% glycerol (v/v) and 0.1% NP-40 (v/v)) and three times in lysis buffer. Antibody-bound beads were thus used to perform RIP with polysome extracts overnight at 4 °C with end-over-end mixing. Mock RIP was carried out with equal amount of anti-GST antibody. The next day, beads were washed four times for 10 min at 4 °C with high salt-containing wash buffer (20mM HEPES-KOH pH 7.5, 350mM KCl, 10mM MgCl2 and 0.1% (vol/ vol) NP-40) and subsequently eluted in 350 µl RA1 Lysis buffer (Macherey-Nagel total RNA purification kit) and 7 µl 1M DTT.

RNAs were purified according to the manufacturer's instructions of the Macherey-Nagel total RNA purification kit, including the optional on-column DNase digestion step, and eluted twice in the same 60 µl of RNAse-free water. In the case of FLAG, HA, or GFP RIPs, 50 µl packed anti-FLAG M2 affinity gel (Sigma), 50 µl packed EZviewTM Red Protein A affinity gel (Sigma) or 30 µl GFP-TRAP (Chromotek) slurry were equilibrated in lysis buffer and used for RIP.

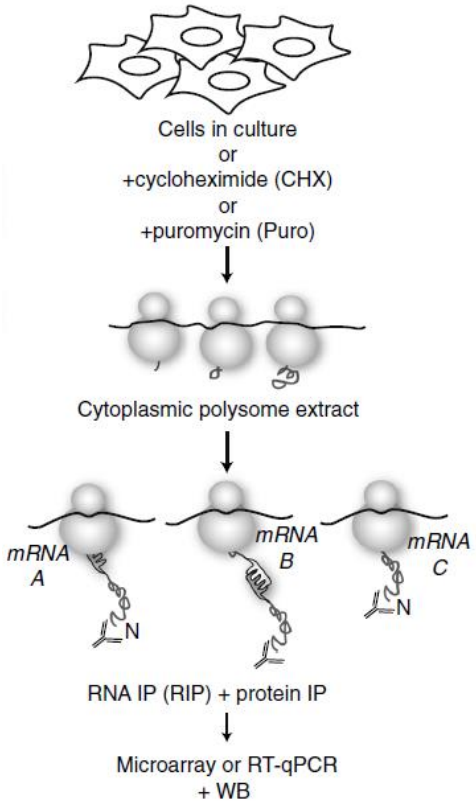


Figure 55: **Schematic representation of polysome RIP assay**

5. cDNA preparation and RT-qPCR

For cDNA synthesis, 5µl of purified RIP-RNA and 5µl of 1:10 diluted input RNA samples were used. cDNA was synthesised using random hexamers and SuperScript IV (ThermoFischer Scientific) according to the manufacturer's instructions. For RIP performed on transfected cells, RNA was additionally treated with Turbo DNase (Ambion) according to the manufacturer's instructions in order to ensure complete plasmid removal before cDNA synthesis. Quantitative PCR was performed with

primers (listed in Supplementary Table 1 our our paper, Results Section 1) on a Roche LightCycler 480 instrument with 45 cycles. In all cases, control cDNAs prepared without reverse transcriptase (−RT) were at least over 10 Cp values of the +RT cDNAs. Enrichment relative to input RNA was calculated using the formula 100 × 2[(Cp(Input) − 6.644) − Cp(IP)] and expressed as "% input RNA". In the case of RIPs performed on transfected cells, enrichment values were expressed as "mRNA fold enrichment" relative to the mock IP using the formula ΔΔCp [IP/mock], to account for the variability of transient transfections. "Relative mRNA fold enrichment" is expressed as mRNA fold enrichment of TAF8 relative to mRNA fold enrichment of TAF10 mRNA. All experiments were performed with a minimum of two biological and two technical replicates and values are represented as mean ±SD. Figures panels were prepared with taking in account all these data points using R (RStudio version 1.1.456 and R version 3.5.1).

6. Microarray analysis and library preparation

Polysome extracts and RIP from HeLa cells were performed as described above with mouse monoclonal antibodies 1H8 targeting the N-terminus of TAF10, 3G3 targeting the N terminus of TBP, and 1D10 targeting GST as a nonspecific control (see Supplementary Table 2). Protein G Sepharose beads were used (100 µl beads coupled to 14 µl antibody). After quantification and quality controls performed on Agilent's Bioanalyzer, biotinylated single strand cDNA targets were prepared, starting from 200 ng of total RNA, using the Ambion WT Expression Kit (Cat # 4411974) and the Affymetrix GeneChip® WT Terminal Labelling Kit (Cat # 900671) according to Affymetrix recommendations. Following fragmentation and end-labelling, 3 µg of cDNAs were hybridised for 16 h at 45 °C on GeneChip® Human Gene 2.0 ST arrays (Affymetrix) interrogating over 40000 RefSeq transcripts and ~11,000 LncRNAs represented by ~27 probes spread across the full-length of the transcript. The chips were washed and stained in the GeneChip® Fluidics Station 450 (Affymetrix) and scanned with the GeneChip® Scanner 3000 7 G (Affymetrix) at a resolution of 0.7 µm. Raw data (.CEL Intensity files) were extracted from the scanned images using the Affymetrix GeneChip® Command Console (AGCC) version 4.0. CEL files were further processed with Affymetrix Expression Console software version 1.3.1 to calculate probe set signal intensities using Robust Multi-array Average (RMA) algorithms with default settings (Sketch quantile normalization). Statistical analysis was performed

using the FCROS package version 1.5.4 (Dembélé and Kastner 2014). Differences are considered significant for p value below 0.025. Volcano plots were performed using RStudio software version 3.3.2. Ribosomal RNA transcripts were filtered out. The microarray results reported in this paper are available in the Gene Expression Omnibus (GEO) under accession number GSE106299

7. Protein IP and Western blot

Antibodies used for RIP, protein IP and western blotting are summarised in Supplementary Table 2. For protein IP, the procedure was performed essentially as for RIP. Bound proteins were eluted in 2× Laemmli buffer supplemented with 20mM DTT and boiled for 5 min. Subsequently, samples were resolved on SDS-PAGE gels and transferred to nitrocellulose membranes using either wet transfer or BioRad's Trans-Turbo Blot semi-dry transfer method. Secondary antibodies (goat anti-mouse or rabbit anti-mouse) coupled to HRP (Jackson ImmunoResearch Laboratories) were used at 1:10,000 dilution. Signal was revealed using chemiluminescence (Pierce) and detected on the ChemiDoc imaging system (BioRad). For immunoprecipitation using whole cell extracts, 10 confluent 10 cm plates were scraped in PBS containing protease inhibitor (Roche) and resuspended in ~1 packed cell volume lysis buffer (20mM Tris-HCl, pH 7.5, 400 mM KCl, 2 mM DTT, 20% glycerol) supplemented with protease inhibitor and 0.5 mM final concentration of DTT. Extracts were prepared by four cycles of freezing on liquid nitrogen followed by thawing on ice. The concentration of the clarified extract was measured by Bradford assay and the extract was diluted ~1:3 using lysis buffer without salt to achieve a final concentration of ~150mM KCl. One milligram extract was added to mock- and antibody-bound beads each and IPs were performed as described above. Proteins were eluted twice for 5 min at room temperature in 50 µl 0.1M Glycine, pH 2.8 and neutralised with 3.5 µl 1.8M Tris- HCl, pH 8.8. Ten percent of the pooled eluates were resolved on gels.

8. Plasmids

The eukaryotic expression plasmid pXJ41 used for all the constructs has been previously described (Xiao et al. 1991). pXJ41-TAF10-Nter-2HA has been previously described (Jacq et al. 1994). To generate N- and C-terminally Flag-tagged TAF8, the human TAF8 cDNA was PCR amplified from pACEMam1-CFP-TAF8 (kind gift from

Imre Berger, University of Bristol, UK) using primers cotaining EcoR I and Bgl II restriction sites and tags incorporated at the N- or C-terminus, respectively, and digestion by appropriate restriction enzymes. Similarly, C-terminal HA tagged TAF10 was subcloned from pXJ41-TAF10-Nter-2HA by PCR amplification and digestion via restriction enzymes Xho I and Kpn I. The TAF8 mutations, TAF8-HFD and TAF8-HFD-60 amino acids were generated by site-directed mutagenesis using PfuUltra High-Fidelity DNA polymerase (Agilent Technologies), according to the manufacturer's instructions. The histone fold domain swapped TAF10 and TAF8 constructs were generated with several rounds of PCR amplification, using the already-mentioned N-terminal tagged TAF10 and TAF8 constructs as a template with specific primers and cloned into the vector via restriction enzymes EcoR I and Bgl II. pXJ41-hTBP has been previously described (Brou et al. 1993). The HA-TAF1 cDNA (Ruppert, Wang, and Tjian 1993) was inserted in pXJ41. TAF1 N-terminal deletion was carried out by site-directed mutagenesis using Pfu Ultra High-Fidelity DNA polymerase (Agilent Technologies), according to the manufacturer's instructions. HA tagged TAF9 was subcloned from pSG5-TAF9 (Frontini et al. 2005) by PCR amplification and digestion by restriction enzymes EcoR I and Bgl II. FLAG-tagged TAF6 was also subcloned in a similar manner from pXJ41-TAF6 (Bell, Scheer, and Tora 2001) via restriction enzymes Xho I and Kpn I. All plasmids have been verified by sequencing. Plasmids are described in Supplementary Table 3 of our paper ***(Kamenova, Mukherjee et al. 2019; Results Section 1).***

9. Mouse Taf8 and Taf10 KO ESC lines

The Rosa26Cre-ERT2/+; Taf8flox/flox mouse embryonic stem cells (mESCs) were generated previously by F. El Saafin (El-Saafin et al. 2018). Briefly, mice carrying the Taf8lox allele were bred to mice carrying the Rosa26Cre-ERT2 allele to produce Rosa26Cre-ERT2/+;Taf8flox/flox E3.5 blastocysts and to isolate Rosa26Cre-ERT2/+;Taf8flox/flox mouse embryonic stem cells (mESCs) (El-Saafin et al. 2018). The Rosa26Cre-ERT2/R; Taf10flox/flox mESCs were generated previously by P. Bardot (Bardot et al. 2017). Briefly, the ESCs were derived from Rosa26Cre-ERT2/R;Taf10lox/lox E3.5 blastocysts (Bardot et al. 2017). mESCs were cultured in DMEM (4.5 g/l glucose) with 2mM Glutamax-I, 15% ESQ FBS (Gibco), penicillin, streptomycine, 0.1mM non-essential amino acids, 0.1% ß-mercaptoethanol, 1500

U/mL LIF and two inhibitors (2i; 3 µM CHIR99021 and 1 µM PD0325901, Axon MedChem) on gelatin-coated plates. To induce deletion of Taf8, mESCs were treated with 0.5 µM 4-OH tamoxifen (Sigma) for 5–6 days, and to induce deletion of Taf10, Rosa26Cre-ERT2/R;Taf10lox/lox mESCs were treated for 4 days with 0.1 µM 4-OH tamoxifen (Sigma). The above-described mESCs have already been described (Bardot et al. 2017; El-Saafin et al. 2018) and were derived according to animal welfare regulations and guidelines of the French Ministry of Agriculture and French Ministry of Higher Education and Research, and the Australian Animal Welfare Committee, respectively.

10. Single molecule inexpensive RNA FISH (smiFISH)

smiFISH primary probes were designed with the R script Oligostan as previously described (Tsanov et al. 2016). Primary probes and secondary probes (Cy3 or digoxigenin conjugated FLAPs) were synthesised and purchased from Integrated DNA Technologies (IDT). Primary probes were ordered at a final concentration of 100 µM, wet and frozen in Tris-EDTA pH 8.0 (TE) buffer. Probe sequences are available in Supplementary Table 4. An equimolar mixture of all the primary probes for a particular RNA was prepared with a final concentration 0.833 µM of individual probes. The secondary probes are resuspended in TE buffer at a final concentration of 100 µM. A total of 10 µl of FLAP hybridization reaction was prepared with 2 µl (for single colour smiFISH) or 4 µl (for dual colour smiFISH) of diluted (0.833 µM) primary probe set, 1 µl of secondary probe, 1 µl of 10X NEB3 and 6 µl of water. The reaction mix was then incubated in a cycler under the following conditions: 85 °C, 3 min, 65 °C, 3 min, 25 °C, 5 min. Two microliters of these FLAP hybridised probes are necessary for each smiFISH reaction. The volume of the reactions were scaled up according to the number of smiFISH reactions carried out. smiFISH was carried out as follows as per published protocol (Tsanov et al. 2016). HeLa cells were treated with 100 µg/ml cycloheximide (Merck) for 15 min at 37 °C, fixed with 4% paraformaldehyde (Electron Microscopy Sciences) for 20 min at room temperature(RT) followed by overnight incubation with 70% ethanol at 4 °C. Following overnight incubation, cells were rinsed with 1× PBS twice and incubated with Solution A (freshly prepared 15% formamide in 1× SSC buffer) for 15 min at RT. During incubation, 50 µl Mix 1 (5 µl of 20× SSC, 1.7 µl of 20 µg/µl E. coli tRNA, 15 µl of 100% formamide, 2 or 4 µl of FLAP hybridised

probes, required amount of water) and 50 µl Mix 2 (1 µl of 20 mg/ml RNAse-free BSA, 1 µl of 200mM VRC, 27 µl of 40% dextran sulfate, 21 µl of water) was prepared. Mix 1 was added to Mix 2 after proper vortexing. The total 100 µl of Mix1 + Mix2 is sufficient for two coverslips. Each coverslip was then incubated on a spot of 50 µl of the Mix in a 15 cm Petri dish with a proper hydration chamber (3.5 cm Petri dish containing 2ml of 15% formamide/1× SSC solution) overnight at 37 °C. Following overnight incubation, coverslips were washed twice with Solution A at 37 °C for 30 min each and with 1× PBS twice for 10 min each. Coverslips with only Cy3 conjugated secondary probes are mounted with 5 µl of Vectashield containing DAPI at this step. For DIG-labelled secondary probes, cells were further permeabilised with 0.1% Triton-X100 for 10 min at RT and incubated with 0.25 µg/ml antidigoxigenin- fluorescein Fab fragments (diluted in 1× PBS) (Roche) for 2 h at RT. Following antibody incubation, cells were mounted as before.

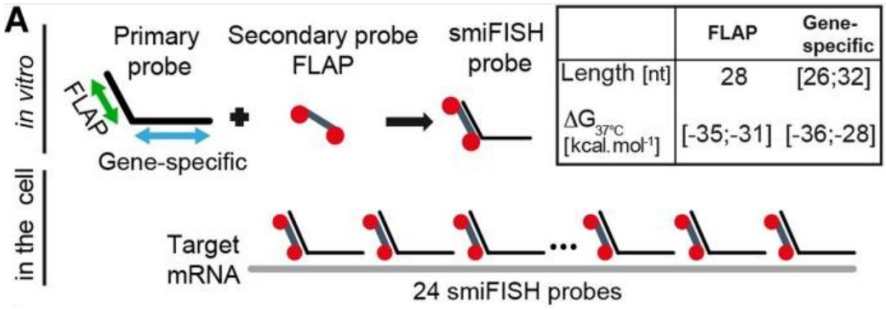


Figure 56: Schematic representation of working principle of smiFISH. ~24 primary probes are pre-hybridized with secondary probes via FLAP sequence *in vitro.* The resulting duplexes are subsequently hybridized in cells. Length (nt: nucleotides) and red circles are Cy3 moieties. Image is from (Tsanov et al. 2016).

11. Immunofluorescence-single molecule RNA FISH (IF-smiFISH)

To visualise proteins and mRNA together, we first performed immunofluorescence (IF) followed by smiFISH. Briefly, cells were treated with 100 µg/ml cycloheximide (Merck) for 15 min at 37 °C, fixed with 4% paraformaldehyde (Electron Microscopy Sciences) for 10 min at room temperature (RT), blocked and permeabilised with blocking buffer (10% BSA, 10% Triton-X-100, 200mM VRC, 2X PBS) for 1 h at 40 °C, incubated for 2 h at RT with either anti-TAF8 (mouse monoclonal antibody (mAb) 1FR-1B6 (William S Mohan et al. 2003); diluted 1:1000) or anti-TAF10 (mAb 6TA-2B11 (William S Mohan

et al. 2003); diluted 1:1000) antibody mix followed by incubation (RT, 1 h) with secondary antibody mix Alexa-488-labelled goat anti-mouse mAb (Life Technologies, catalogue number A-11001, diluted 1:3000). Following immunofluorescence described above, cells were fixed with 4% paraformaldehyde (Sigma) for 10 min at RT. Cells were washed with 1× PBS and incubated with wash buffer [10% Formamide (Sigma) in 2× SSC] for 10 min at RT. smiFISH was carried out as described above and see ref. (Tsanov et al. 2016). Cells were mounted using Vectashield mounting medium with DAPI (Vector laboratories Inc.).

12. Imaging and image processing

Confocal imaging of smiFISH and IF-smiFISH samples was performed on an SP8UV microscope (Leica) equipped with a 633-nm HeNe laser, a 561-nm DPSS laser, a 488-nm argon laser and a 405-nm laser diode. A ×63 oil immersion objective (NA 1.4) was used and images were taken by using the hybrid detector photon-counting mode. The laser power for all acquisitions and laser lines was set to 10%. All images acquired have a bit depth of 8 bit and a pixel resolution of 70 nm. The z-stacks were taken with a z-spacing of 300 nm for a total of 4–6 µm. Image processing was performed using the Fiji/Image J software. All images were processed the same way. In detail, the channels of the different images were split and grey values were adjusted to better visualise the spots in the cytoplasm. The nuclear signal in the green channel (TAF10 or TAF8 IF) was removed by masking the nucleus and using the “clear” option. Finally, the processed channels were merged again. For IF-smiFISH, one cell of an image was cropped and one representing z-slice per cell was chosen. For smiFISH, maximum intensity Z-projections of individual images were made and one cell per resulting image was cropped as the representative image. In addition, one single IF or smiFISH spot from the corresponding cells was cropped as well.

13. Image analysis of IF-smiFISH data

To measure the degree of spatial overlap of smiFISH (mRNA) and IF (protein) signal, an enrichment ratio was calculated as described below. Such quantification was chosen in order to take into account the variability of IF signal between cells, making single object detection in this channel difficult. Cells and nuclei were outlined manually in 2D based on the GFP and DAPI image, respectively. Subsequent analyses were

restricted to the cytoplasm. mRNAs were detected in 3D with FISH-quant (Tsanov et al. 2016). Identical detection settings were used when different experimental conditions were compared with the same gene. Each cell was post-processed separately. First, the median pixel intensity in the IF image at the identified RNA positions was calculated. Second, a normalization factor was estimated as the median IF intensity of the outlined cytoplasm within the z-range of the detected mRNAs. The enrichment ratio of the cell was then calculated as the ratio of the median IF intensity at the RNA positions divided by the mean cytoplasmic intensity. Boxplots of enrichment ratios were generated with the Matlab function notBoxPlot. Each dot corresponds to the estimation of one cell. Horizontal lines are mean values, 95% confidence interval is shown in red, and standard deviation in blue. Statistical comparison between different experimental conditions was performed with two-sample Kolmogorov–Smirnov test (Matlab function kstest2). The Matlab script is available upon request.

14. Image analysis of smiFISH co-localization data

Segmentation of nuclei and cells was performed with the DAPI and smiFISH channel channels, respectively. 2D images were obtained with a previously described projection approach based on local and global focus measurements (Tsanov et al. 2016). Segmentation was implemented with the open-source software CellProfiler (Dao et al. 2016) using a standard workflow: Otsu and watershed separation for nuclei in the DAPI channel. Each nucleus then serves as a seed for a watershed segmentation to obtain the cells in the smiFISH channel. Individual RNA molecules were localised with FISH-quant in 3D and can be treated as point clouds (Mueller et al. 2013). Co-localization analysis between detected RNAs in two colours was solved as a linear assignment problem (LAP) with the Hungarian algorithm (Matlab function hungarianlinker and munkres from Matlab FileExchange). In short, this algorithm finds the best possible global assignment between these two points-clouds such that for each point in the first colour the closest point in the second channel is found. We implemented a user interface for this analysis tool (FQ_DualColor), which is distributed together with a dedicated user manual with FISH-quant: https://bitbucket.org/muellerflorian/fish_quant.

15. Mass Spectrometry analysis

Samples were analyzed using an UltiMate 3000 RSLCnano (Thermo Scientific) coupled in line with a linear trap Quadrupole (LTQ)-Orbitrap ELITE mass spectrometer via a nano-electrospray ionization source. In detail, samples were TCA precipitated, reduced, alkylated and digested with LysC and Trypsin at 37°C overnight. After C18 desalting, samples were analyzed using an Ultimate 3000 nano-RSLC (Thermo Scientific, San Jose, California) coupled in line with a linear trap Quadrupole (LTQ)-Orbitrap ELITE mass spectrometer via a nano-electrospray ionization source (Thermo Scientific). Peptide mixtures were loaded on a C18 Acclaim PepMap100 trap column (75 µm inner diameter × 2 cm, 3 µm, 100 Å; Thermo Fisher Scientific) for 3.5 min at 5 µl/min with 2% acetonitrile (ACN), 0.1% formic acid in H2O and then separated on a C18 Accucore nano-column (75 µm inner diameter × 50 cm, 2.6 µm, 150 Å; Thermo Fisher Scientific) with a 240-min linear gradient from 5% to 50% buffer B (A: 0.1% FA in H2O; B: 80% ACN, 0.08% FA in H2O) followed with 10 min at 99% B. The total duration was set to 280 min at a flow rate of 200 nL/min. Peptides were analyzed by high resolution full MS scan (R240K, from 300 to 1650 m/z range) followed by 20 MS/MS events using data-dependent CID (collision induced dissociation) acquisition. Proteins were identified by database searching using SequestHT (Thermo Fisher Scientific) with Proteome Discoverer 1.4 software (Thermo Fisher Scientific) a combined Homo sapiens database (Swissprot, release 2015_11, 16730 entries) where 5 sequences of protein of interest (TrEMBL entries) were added. Precursor and fragment mass tolerances were set at 7 ppm and 0.5 Da respectively, and up to 2 missed cleavages were allowed. Oxidation (M) was set as variable modification, and carbamidomethylation (C) as fixed modification. Peptides were filtered with a false discovery rate (FDR) and rank 1: FDR at 5 %, rank 1 and proteins were identified with 1 unique peptide.

16. Mass spectrometry data analysis

Normalized spectral abundance factor (NSAF) (Zybailov et al. 2006) normalized to the bait ($NSAF_{bait}$) were obtained as followed (PSM*; peptide spectrum match, SAF; spectral abundance factor, x; protein of interest):

$$SAF_{(x)} = \frac{PSM^{*}_{x(IP)} - PSM^{*}_{x(IPmock)}}{length(x)}$$

$$NSAF_{(x)} = \frac{SAF_{(x)}}{\sum_{i=1}^{n} SAF_{(xi)}} \times 100$$

$$NSAF_{bait(x)} = \frac{NSAF_{(x)}}{NSAF_{(bait)}}$$

We only considered proteins with positive SAF values.

17. TAF8-MS2-TRAP (MS2-tagged RNA affinity purification)

A pcDNA3 plasmid with lacZ sequence followed by 24 MS2 repeats in its 3'UTR and a stable HA-MCP-GFP cell line was obtained from Edouard Bertrand's lab as a kind gift. We replaced the lacZ sequence in the plasmid with our gene of interest, FLAG-TAF8. MS2 repeat containing plasmid was transformed in XL1-Blue (rec- bacteria) and grown at 30°C to preserve integrity of MS2 repeats. Several minipreps were prepared from single colonies, and a clone with intact MS2 repeat was used to carry out a maxiprep. The entire 24xMS2 repeat is about 1,5 kb long, thus every step of growing bacteria during cloning was carried out at 30°C and using rec- strain. The HA-MCP-GFP cell line was generated by transferring MCP fused with HA and GFP at N- and C-terminus respectively into HeLa Flp-In/T-REx cells following procedures described in (van Nuland et al. 2013). Cells were cultured in a 37°C humidified/5% CO2 incubator. Culture media consisted of Dulbecco's modified Eagle's medium (DMEM), supplemented with 1 g/l glucose, 10% fetal calf serum (FCS), and 40 µg/ml Gentamycin.

TAF8-MS2 plasmid was transfected into HeLa cells by Lipofectamine 2000 as per manufacturer's protocol. Following that, RIP assay was carried out as described in Methods Section 4 with anti-GFP beads (Chromotek) and finally RNA isolation, cDNA preparation and qPCR for RNA analysis and Western blot for protein anaysis was carried out as described in Methods Section 5 and 7 respectively.

BIBLIOGRAPHY

Bibliography

Ackerman, Sharon H., and Alexander Tzagoloff. 2005. “Function, Structure, and Biogenesis of Mitochondrial ATP Synthase.” *Progress in Nucleic Acid Research and Molecular Biology* 80 (05): 95–133. https://doi.org/10.1016/S0079-6603(05)80003-0.

Adelman, Karen, and John T. Lis. 2012. “Promoter-Proximal Pausing of RNA Polymerase II: Emerging Roles in Metazoans.” *Nature Reviews. Genetics* 13 (10): 720–31. https://doi.org/10.1038/nrg3293.

Aizer, Adva, Alon Kalo, Pinhas Kafri, Amit Shraga, Rakefet Ben-Yishay, Avi Jacob, Noa Kinor, and Yaron Shav-Tal. 2014. “Quantifying MRNA Targeting to P-Bodies in Living Human Cells Reveals Their Dual Role in MRNA Decay and Storage.” *Journal of Cell Science* 127 (20): 4443–56. https://doi.org/10.1242/jcs.152975.

Al-Ahmadi, Wijdan, Maha Al-Ghamdi, Latifa Al-Haj, Maher Al-Saif, and K. S A Khabar. 2009. “Alternative Polyadenylation Variants of the RNA Binding Protein, HuR: Abundance, Role of AU-Rich Elements and Auto-Regulation.” *Nucleic Acids Research* 37 (11): 3612–24. https://doi.org/10.1093/nar/gkp223.

Alberti, Simon. 2017. “Phase Separation in Biology.” *Current Biology* 27 (20): R1097–1102. https://doi.org/10.1016/j.cub.2017.08.069.

Alkalaeva, Elena Z., Andrey V. Pisarev, Lyudmila Y. Frolova, Lev L. Kisselev, and Tatyana V. Pestova. 2006. “In Vitro Reconstitution of Eukaryotic Translation Reveals Cooperativity between Release Factors ERF1 and ERF3.” *Cell* 125 (6): 1125–36. https://doi.org/10.1016/j.cell.2006.04.035.

Almeida-Souza, Leonardo, Sofie Goethals, Vicky De Winter, Ines Dierick, Rodrigo Gallardo, Joost Van Durme, Joy Irobi, et al. 2010. “Increased Monomerization of Mutant HSPB1 Leads to Protein Hyperactivity in Charcot-Marie-Tooth Neuropathy.” *Journal of Biological Chemistry* 285 (17): 12778–86. https://doi.org/10.1074/jbc.M109.082644.

Anandapadamanaban, Madhanagopal, Cecilia Andresen, Sara Helander, Yoshifumi Ohyama, Marina I. Siponen, Patrik Lundström, Tetsuro Kokubo, Mitsuhiko Ikura, Martin Moche, and Maria Sunnerhagen. 2013. “High-Resolution Structure of TBP with TAF1 Reveals Anchoring Patterns in Transcriptional Regulation.” *Nature Structural and Molecular Biology* 20 (8): 1008–14. https://doi.org/10.1038/nsmb.2611.

Anderson, Paul, and Nancy Kedersha. 2009. “RNA Granules: Post-Transcriptional and Epigenetic Modulators of Gene Expression.” *Nature Reviews Molecular Cell Biology* 10 (6): 430–36. https://doi.org/10.1038/nrm2694.

Ansari, Mohammed Y., and Shekhar C. Mande. 2018. “A Glimpse Into the Structure and Function of Atypical Type I Chaperonins.” *Frontiers in Molecular Biosciences* 5 (April): 1–8. https://doi.org/10.3389/fmolb.2018.00031.

Antonova, Simona V., Matthias Haffke, Eleonora Corradini, Mykolas Mikuciunas, Teck Y. Low, Luca Signor, Robert M. van Es, et al. 2018. “Chaperonin CCT Checkpoint Function in Basal Transcription Factor TFIID Assembly.” *Nature*

Structural and Molecular Biology 25 (12): 1119–27. https://doi.org/10.1038/s41594-018-0156-z.

Arents, Gina. 1993. “Topography of the Histone Octamer Surface : Repeating Structural Motifs Utilized in the Docking of Nucleosomal DNA” 90 (November): 10489–93.

Arents, Gina, Rufus W Burlingame, Bi-cheng Wangt, Warner E Love, and Evangelos N Moudrianakis. 1991. “The Nucleosomal Core Histone Octamer at 3 . 1 A Resolution : A Tripartite Protein Assembly and a Left-Handed Superhelix” 88 (November): 10148–52.

Arkov, Alexey L., and Andres Ramos. 2010. “Building RNA-Protein Granules: Insight from the Germline.” *Trends in Cell Biology* 20 (8): 482–90. https://doi.org/10.1016/j.tcb.2010.05.004.

Atkinson, Gemma C., Sandra L. Baldauf, and Vasili Hauryliuk. 2008. “Evolution of Nonstop, No-Go and Nonsense-Mediated MRNA Decay and Their Termination Factor-Derived Components.” *BMC Evolutionary Biology* 8 (1): 1–18. https://doi.org/10.1186/1471-2148-8-290.

Attikum, Haico Van, and Susan M. Gasser. 2005. “The Histone Code at DNA Breaks: A Guide to Repair?” *Nature Reviews Molecular Cell Biology* 6 (10): 757–65. https://doi.org/10.1038/nrm1737.

Aziz, Muhammad. 2011. “Scholarship at UWindsor PREDICTION OF OBLIGATE AND NON- OBLIGATE PROTEIN-PROTEIN INTERACTIONS.”

Bakheet, T. 2001. “ARED: Human AU-Rich Element-Containing MRNA Database Reveals an Unexpectedly Diverse Functional Repertoire of Encoded Proteins.” *Nucleic Acids Research* 29 (1): 246–54. https://doi.org/10.1093/nar/29.1.246.

Baltimore David. 1970. “Viral RNA-Dependent DNA Polymerase” 226 (Table 2): 1209–11.

Bao, Yunhe, and Xuetong Shen. 2007. “SnapShot: Chromatin Remodeling Complexes.” *Cell* 129 (3): 632.e1-632.e2. https://doi.org/10.1016/j.cell.2007.04.018.

Baptista, Tiago, Sebastian Grünberg, Nadège Minoungou, Maria J.E. Koster, H. T.Marc Timmers, Steve Hahn, Didier Devys, and László Tora. 2017. “SAGA Is a General Cofactor for RNA Polymerase II Transcription.” *Molecular Cell* 68 (1): 130-143.e5. https://doi.org/10.1016/j.molcel.2017.08.016.

Bardot, Paul, Stéphane D. Vincent, Marjorie Fournier, Alexis Hubaud, Mathilde Joint, László Tora, and Olivier Pourquié. 2017. “The TAF10-Containing TFIID and SAGA Transcriptional Complexes Are Dispensable for Early Somitogenesis in the Mouse Embryo.” *Development* 144 (20): 3808–18. https://doi.org/10.1242/dev.146902.

Barral, José M., Sarah A. Broadley, Gregor Schaffar, and F. Ulrich Hartl. 2004. “Roles of Molecular Chaperones in Protein Misfolding Diseases.” *Seminars in Cell and Developmental Biology* 15 (1): 17–29. https://doi.org/10.1016/j.semcdb.2003.12.010.

Barreau, Carine, Luc Paillard, and H. Beverley Osborne. 2005. “AU-Rich Elements

and Associated Factors: Are There Unifying Principles?" *Nucleic Acids Research* 33 (22): 7138–50. https://doi.org/10.1093/nar/gki1012.

Bartel, David P. 2009. "MicroRNAs: Target Recognition and Regulatory Functions." *Cell* 136 (2): 215–33. https://doi.org/10.1016/j.cell.2009.01.002.

Beaudoin, Jean Denis, and Jean Pierre Perreault. 2010. "5′-UTR G-Quadruplex Structures Acting as Translational Repressors." *Nucleic Acids Research* 38 (20): 7022–36. https://doi.org/10.1093/nar/gkq557.

Becker, Annemarie H., Eugene Oh, Jonathan S. Weissman, Günter Kramer, and Bernd Bukau. 2013. "Selective Ribosome Profiling as a Tool for Studying the Interaction of Chaperones and Targeting Factors with Nascent Polypeptide Chains and Ribosomes." *Nature Protocols* 8 (11): 2212–39. https://doi.org/10.1038/nprot.2013.133.

Becker, Peter B., and Jerry L. Workman. 2013. "Nucleosome Remodeling and Epigenetics." *Cold Spring Harbor Perspectives in Biology* 5 (9). https://doi.org/10.1101/cshperspect.a017905.

Behrends, Christian, Carola A. Langer, Raina Boteva, Ulrike M. Böttcher, Markus J. Stemp, Gregor Schaffar, Bharathi Vasudeva Rao, et al. 2006. "Chaperonin TRiC Promotes the Assembly of PolyQ Expansion Proteins into Nontoxic Oligomers." *Molecular Cell* 23 (6): 887–97. https://doi.org/10.1016/j.molcel.2006.08.017.

Bell, Brendan, Elisabeth Scheer, and Làszlò Tora. 2001. "Identification of HTAFII80δ Links Apoptotic Signaling Pathways to Transcription Factor TFIID Function." *Molecular Cell* 8 (3): 591–600. https://doi.org/10.1016/S1097-2765(01)00325-2.

Ben-Asouli, Yitzhak, Yona Banai, Yehuda Pel-Or, Alexei Shir, and Raymond Kaempfer. 2002. "Human Interferon-γ MRNA Autoregulates Its Translation through a Pseudoknot That Activates the Interferon-Inducible Protein Kinase PKR." *Cell* 108 (2): 221–32. https://doi.org/10.1016/S0092-8674(02)00616-5.

Bergendahl, L. Therese, and Joseph A. Marsh. 2017. "Functional Determinants of Protein Assembly into Homomeric Complexes." *Scientific Reports* 7 (1): 1–10. https://doi.org/10.1038/s41598-017-05084-8.

Berkovits, Binyamin D., and Christine Mayr. 2015. "Alternative 3′ UTRs Act as Scaffolds to Regulate Membrane Protein Localization." *Nature* 522 (7556): 363–67. https://doi.org/10.1038/nature14321.

Bernecky, Carrie, Jürgen M. Plitzko, and Patrick Cramer. 2017. "Structure of a Transcribing RNA Polymerase II-DSIF Complex Reveals a Multidentate DNA-RNA Clamp." *Nature Structural and Molecular Biology* 24 (10): 809–15. https://doi.org/10.1038/nsmb.3465.

Bernstein, Bradley E., Tarjei S. Mikkelsen, Xiaohui Xie, Michael Kamal, Dana J. Huebert, James Cuff, Ben Fry, et al. 2006. "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells." *Cell* 125 (2): 315–26. https://doi.org/10.1016/j.cell.2006.02.041.

Bhattacharya, S., X. Lou, P. Hwang, K. R. Rajashankar, X. Wang, J.-A. Gustafsson, R. J. Fletterick, R. H. Jacobson, and P. Webb. 2014. "Structural and Functional Insight into TAF1-TAF7, a Subcomplex of Transcription Factor II D."

Proceedings of the National Academy of Sciences 111 (25): 9103–8. https://doi.org/10.1073/pnas.1408293111.

Bhaumik, Sukesh R., Tamal Raha, David P. Aiello, and Michael R. Green. 2004. “In Vivo Target of a Transcriptional Activator Revealed by Fluorescence Resonance Energy Transfer.” *Genes and Development* 18 (3): 333–43. https://doi.org/10.1101/gad.1148404.

Bieniossek, Christoph, Gabor Papai, Christiane Schaffitzel, Frederic Garzoni, Maxime Chaillet, Elisabeth Scheer, Petros Papadopoulos, Laszlo Tora, Patrick Schultz, and Imre Berger. 2013. “The Architecture of Human General Transcription Factor TFIID Core Complex.” *Nature* 493 (7434): 699–702. https://doi.org/10.1038/nature11791.

Bingel-Erlenmeyer, Rouven, Rebecca Kohler, Günter Kramer, Arzu Sandikci, Snježana Antolić, Timm Maier, Christiane Schaffitzel, Brigitte Wiedmann, Bernd Bukau, and Nenad Ban. 2008. “A Peptide Deformylase-Ribosome Complex Reveals Mechanism of Nascent Chain Processing.” *Nature* 452 (7183): 108–11. https://doi.org/10.1038/nature06683.

Birck, Catherine, Yinghua Chen, F. Marion Hulett, and Jean Pierre Samama. 2003. “The Crystal Structure of the Phosphorylation Domain in PhoP Reveals a Functional Tandem Association Mediated by an Asymmetric Interface.” *Journal of Bacteriology* 185 (1): 254–61. https://doi.org/10.1128/JB.185.1.254-261.2003.

Birck, Catherine, Olivier Poch, Christophe Romier, Marc Ruff, Gabrielle Mengus, Anne Claire Lavigne, Irwin Davidson, and Dino Moras. 1998. “Human TAF(II)28 and TAF(II)18 Interact through a Histone Fold Encoded by Atypical Evolutionary Conserved Motifs Also Found in the SPT3 Family.” *Cell* 94 (2): 239–49. https://doi.org/10.1016/S0092-8674(00)81423-3.

Boeynaems, Steven, Elke Bogaert, Denes Kovacs, Albert Konijnenberg, Evy Timmerman, Alex Volkov, Mainak Guharoy, et al. 2017. “Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics.” *Molecular Cell* 65 (6): 1044-1055.e5. https://doi.org/10.1016/j.molcel.2017.02.013.

Bolduc, François, Jean Michel Garant, Félix Allard, and Jean Pierre Perreault. 2016. “Irregular G-Quadruplexes Found in the Untranslated Regions of Human MRNAs Influence Translation.” *Journal of Biological Chemistry* 291 (41): 21751–60. https://doi.org/10.1074/jbc.M116.744839.

Bonnet, Jacques, Chen Yi Wang, Tiago Baptista, Stéphane D. Vincent, Wei Chun Hsiao, Matthieu Stierle, Cheng Fu Kao, László Tora, and Didier Devys. 2014. “The SAGA Coactivator Complex Acts on the Whole Transcribed Genome and Is Required for RNA Polymerase II Transcription.” *Genes and Development* 28 (18): 1999–2012. https://doi.org/10.1101/gad.250225.114.

Boulay, Gaylor, Gabriel J. Sandoval, Nicolo Riggi, Sowmya Iyer, Rémi Buisson, Beverly Naigles, Mary E. Awad, et al. 2017. “Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain.” *Cell* 171 (1): 163-178.e19. https://doi.org/10.1016/j.cell.2017.07.036.

Boulon, Séverine, Nathalie Marmier-Gourrier, Bérengère Pradet-Balade, Laurence Wurth, Céline Verheggen, Beáta E. Jády, Benjamin Rothé, et al. 2008. “The

Hsp90 Chaperone Controls the Biogenesis of L7Ae RNPs through Conserved Machinery." *Journal of Cell Biology* 180 (3): 579–95. https://doi.org/10.1083/jcb.200708110.

Boulon, Séverine, Bérengère Pradet-Balade, Céline Verheggen, Dorothée Molle, Stéphanie Boireau, Marya Georgieva, Karim Azzag, et al. 2010. "HSP90 and Its R2TP/Prefoldin-like Cochaperone Are Involved in the Cytoplasmic Assembly of RNA Polymerase II." *Molecular Cell* 39 (6): 912–24. https://doi.org/10.1016/j.molcel.2010.08.023.

Brou, C., S. Chaudhary, I. Davidson, Y. Lutz, J. Wu, J.M. Egly, L. Tora, and P. Chambon. 1993. "Distinct TFIID Complexes Mediate the Effect of Different Transcriptional Activators." *The EMBO Journal* 12 (2): 489–99. https://doi.org/10.1002/j.1460-2075.1993.tb05681.x.

Brownlee, Proudfoot and. 1976. "3′ Non-Coding Region Sequences in Eukaryotic Messenger RNA." *Nature* 263 (5574): 211–14. https://doi.org/10.1038/263211a0.

Buchan, J. Ross, and Roy Parker. 2009. "Eukaryotic Stress Granules: The Ins and Outs of Translation." *Molecular Cell* 36 (6): 932–41. https://doi.org/10.1016/j.molcel.2009.11.020.

Buchner, Johannes. 2019. "Molecular Chaperones and Protein Quality Control: An Introduction to the JBC Reviews Thematic Series." *Journal of Biological Chemistry* 294 (6): 2074–75. https://doi.org/10.1074/jbc.REV118.006739.

Bugaut, Anthony, and Shankar Balasubramanian. 2012. "5′-UTR RNA G-Quadruplexes: Translation Regulation and Targeting." *Nucleic Acids Research* 40 (11): 4727–41. https://doi.org/10.1093/nar/gks068.

Bühler, Marc, Silvia Steiner, Fabio Mohn, Alexandra Paillusson, and Oliver Mühlemann. 2006. "EJC-Independent Degradation of Nonsense Immunoglobulin-μ MRNA Depends on 3′ UTR Length." *Nature Structural and Molecular Biology* 13 (5): 462–64. https://doi.org/10.1038/nsmb1081.

Bukau, Bernd. 2005. "Ribosomes Catch Hsp70s." *Nature Structural and Molecular Biology News and Vies* 12: 472–73. https://doi.org/10.1126/science.1109247.

Bukau, Bernd, Elke Deuerling, Christine Pfund, and Elizabeth A. Craig. 2000. "Getting Newly Synthesized Proteins into Shape." *Cell* 101 (2): 119–22. https://doi.org/10.1016/S0092-8674(00)80806-5.

Buratowski, Stephen. 1994. "The Basics of Basal Transcription by RNA Polymerase II." *Cell* 77 (1): 1–3. https://doi.org/10.1016/0092-8674(94)90226-7.

Buratowski, Stephen. 2003. "The CTD Code." *Nature Structural Biology* 10 (9): 679–80. https://doi.org/10.1038/nsb0903-679.

Burke, Kathleen A., Abigail M. Janke, Christy L. Rhine, and Nicolas L. Fawzi. 2015. "Residue-by-Residue View of In Vitro FUS Granules That Bind the C-Terminal Domain of RNA Polymerase II." *Molecular Cell* 60 (2): 231–41. https://doi.org/10.1016/j.molcel.2015.09.006.

Burley, S. K. 1996. "Biochemistry and Structural Biology of Transcription Factor IID (TFIID)." *Annual Review of Biochemistry* 65 (1): 769–99.

https://doi.org/10.1146/annurev.biochem.65.1.769.

Calloni, Giulia, Taotao Chen, Sonya M. Schermann, Hung Chun Chang, Pierre Genevaux, Federico Agostini, Gian Gaetano Tartaglia, Manajit Hayer-Hartl, and F. Ulrich Hartl. 2012. “DnaK Functions as a Central Hub in the E. Coli Chaperone Network.” *Cell Reports* 1 (3): 251–64. https://doi.org/10.1016/j.celrep.2011.12.007.

Campbell, Zachary T., Devesh Bhimsaria, Cary T. Valley, Jose A. Rodriguez-Martinez, Elena Menichelli, James R. Williamson, Aseem Z. Ansari, and Marvin Wickens. 2012. “Cooperativity in RNA-Protein Interactions: Global Analysis of RNA Binding Specificity.” *Cell Reports* 1 (5): 570–81. https://doi.org/10.1016/j.celrep.2012.04.003.

Chang, Lynne, Yaron Shav-Tal, Tatjana Trcek, Robert H. Singer, and Robert D. Goldman. 2006. “Assembling an Intermediate Filament Network by Dynamic Cotranslation.” *Journal of Cell Biology* 172 (5): 747–58. https://doi.org/10.1083/jcb.200511033.

Chantry, Andrew. 1995. “The Kinase Domain and Membrane Localization Determine Intracellular Interactions between Epidermal Growth Factor Receptors.” *Journal of Biological Chemistry*, 3068–73.

Chao, Yanjie, and Jörg Vogel. 2016. “A 3’ UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response.” *Molecular Cell* 61 (3): 352–63. https://doi.org/10.1016/j.molcel.2015.12.023.

Chartron, Justin W., Katherine C.L. Hunt, and Judith Frydman. 2016. “Cotranslational Signal-Independent SRP Preloading during Membrane Targeting.” *Nature* 536 (7615): 224–28. https://doi.org/10.1038/nature19309.

Chen, C Y, N Xu, and A B Shyu. 1995. “MRNA Decay Mediated by Two Distinct AU-Rich Elements from c-Fos and Granulocyte-Macrophage Colony-Stimulating Factor Transcripts: Different Deadenylation Kinetics and Uncoupling from Translation.” *Molecular and Cellular Biology* 15 (10): 5777–88. https://doi.org/10.1128/mcb.15.10.5777.

Chen, Cho Yi, Shui Tein Chen, Hsueh Fen Juan, and Hsuan Cheng Huang. 2012. “Lengthening of 3′UTR Increases with Morphological Complexity in Animal Evolution.” *Bioinformatics* 28 (24): 3178–81. https://doi.org/10.1093/bioinformatics/bts623.

Chen, Chyi-ying A, and Ann-gin Shyu. 1995. “C AUorieh d e Ts : Cterizatioa J Import Ce Ia MRNA Degradation.” *Science*.

Chen, Kok Hao, Alistair N. Boettiger, Jeffrey R. Moffitt, Siyuan Wang, and Xiaowei Zhuang. 2015. “Spatially Resolved, Highly Multiplexed RNA Profiling in Single Cells.” *Science* 348 (6233): 1360–63. https://doi.org/10.1126/science.aaa6090.

Chen, Ying, Andreas Boland, Duygu Kuzuoğlu-Öztürk, Praveen Bawankar, Belinda Loh, Chung Te Chang, Oliver Weichenrieder, and Elisa Izaurralde. 2014. “A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between MiRNA Target Recognition and Silencing.” *Molecular Cell* 54 (5): 737–50. https://doi.org/10.1016/j.molcel.2014.03.034.

Cheng, Ming Y., Robert A. Pollock, Frantisek Kalousek, Arthur L. Horwich, F.-Ulrich Hartl, Jörg Martin, Walter Neuper, Elizabeth M. Hallberg, and Richard L. Hallberg. 1989. “Mitochondrial Heat-Shock Protein Hsp60 Is Essential for Assembly of Proteins Imported into Yeast Mitochondria.” *Nature* 337 (6208): 620–25. https://doi.org/10.1038/337620a0.

Chiti, Fabrizio, and Christopher M. Dobson. 2006. “Protein Misfolding, Functional Amyloid, and Human Disease.” *Annual Review of Biochemistry* 75 (1): 333–66. https://doi.org/10.1146/annurev.biochem.75.101304.123901.

Choi and Hynes. 1979. “Biosynthesis and Processing of Fibronectin in NIL.8 Hamster Cells.” *Journal of Biological Chemistry* 254 (23): 12050–55.

Cirillo, Lisa Ann, Frank Robert Lin, Isabel Cuesta, Dara Friedman, Michal Jarnik, and Kenneth S Zaret. 2002. “FoxA1-Gata4-Epigenetic.Pdf” 9: 279–89. https://doi.org/10.1016/S1097-2765(02)00459-8.

Citterio, E., V. Van Den Boom, G. Schnitzler, R. Kanaar, E. Bonte, R. E. Kingston, J. H. J. Hoeijmakers, and W. Vermeulen. 2000. “ATP-Dependent Chromatin Remodeling by the Cockayne Syndrome B DNA Repair-Transcription-Coupling Factor.” *Molecular and Cellular Biology* 20 (20): 7643–53. https://doi.org/10.1128/mcb.20.20.7643-7653.2000.

Clapier, Cedric R., Janet Iwasa, Bradley R. Cairns, and Craig L. Peterson. 2017. “Mechanisms of Action and Regulation of ATP-Dependent Chromatin-Remodelling Complexes.” *Nature Reviews Molecular Cell Biology* 18 (7): 407–22. https://doi.org/10.1038/nrm.2017.26.

Claxton, Derek P., and Eric Gouaux. 2018. “Expression and Purification of a Functional Heteromeric GABAA Receptor for Structural Studies.” *PLoS ONE* 13 (7): 1–25. https://doi.org/10.1371/journal.pone.0201210.

Cler, Emilie, Gabor Papai, Patrick Schultz, and Irwin Davidson. 2009. “Recent Advances in Understanding the Structure and Function of General Transcription Factor TFIID.” *Cellular and Molecular Life Sciences* 66 (13): 2123–34. https://doi.org/10.1007/s00018-009-0009-3.

Cohen-Chalamish, Smadar, Anat Hasson, Dahlia Weinberg, Lise Sarah Namer, Yona Banai, Farhat Osman, and Raymond Kaempfer. 2009. “Dynamic Refolding of IFN-γ MRNA Enables It to Function as PKR Activator and Translation Template.” *Nature Chemical Biology* 5 (12): 896–903. https://doi.org/10.1038/nchembio.234.

Conicella, Alexander E., Gül H. Zerze, Jeetain Mittal, and Nicolas L. Fawzi. 2016. “ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain.” *Structure* 24 (9): 1537–49. https://doi.org/10.1016/j.str.2016.07.007.

Counts et al. 1978. “Disulfide Bonds and the Quaternary Structure of Factor VIII/von Willebrand Factor.” *Journal of Clinical Investigation* 62 (3): 702–9. https://doi.org/10.1172/JCI109178.

Crick, Francis. 1970. “Central Dogma of Molecular Biology.” *Nature*: 561-563.

Cuéllar, Jorge, Jaime Martín-Benito, Sjors H.W. Scheres, Rui Sousa, Fernando

Moro, Eduardo López-Vĩas, Paulino Gómez-Puertas, Arturo Muga, José L. Carrascosa, and José M. Valpuesta. 2008. “The Structure of CCT-Hsc70NBD Suggests a Mechanism for Hsp70 Delivery of Substrates to the Chaperonin.” *Nature Structural and Molecular Biology* 15 (8): 858–64. https://doi.org/10.1038/nsmb.1464.

Dao, David, Adam N. Fraser, Jane Hung, Vebjorn Ljosa, Shantanu Singh, and Anne E. Carpenter. 2016. “CellProfiler Analyst: Interactive Data Exploration, Analysis and Classification of Large Biological Image Sets.” *Bioinformatics* 32 (20): 3210–12. https://doi.org/10.1093/bioinformatics/btw390.

Dean, Kimberly, Jonathan Heidt, Ali Shilatifard, Owen W. Ryan, Mark Johnston, Jim Dover, Nevan J. Krogan, et al. 2004. “The Paf1 Complex Is Required for Histone H3 Methylation by COMPASS and Dot1p: Linking Transcriptional Elongation to Histone Methylation.” *Molecular Cell* 11 (3): 721–29. https://doi.org/10.1016/s1097-2765(03)00091-1.

Decker, Carolyn J., Daniela Teixeira, and Roy Parker. 2007. “Edc3p and a Glutamine/Asparagine-Rich Domain of Lsm4p Function in Processing Body Assembly in Saccharomyces Cerevisiae.” *Journal of Cell Biology* 179 (3): 437–49. https://doi.org/10.1083/jcb.200704147.

Dekker, Carien, Peter C. Stirling, Elizabeth A. McCormack, Heather Filmore, Angela Paul, Renee L. Brost, Michael Costanzo, Charles Boone, Michel R. Leroux, and Keith R. Willison. 2008. “The Interaction Network of the Chaperonin CCT.” *EMBO Journal* 27 (13): 1827–39. https://doi.org/10.1038/emboj.2008.108.

Dembélé, Doulaye, and Philippe Kastner. 2014. “Fold Change Rank Ordering Statistics: A New Method for Detecting Differentially Expressed Genes.” *File:///Home/Sbagew/PhD/Workspace/Publications/Co-ExpressionNetworks/Afroza/Fold_change/Gene_Category/GSE19677/Ddq018.Pdf* 15 (1): 14. https://doi.org/10.1186/1471-2105-15-14.

Demény, Màté A., Evi Soutoglou, Zita Nagy, Elisabeth Scheer, Àgnes Jànoshàzi, Magalie Richardot, Manuela Argentini, Pascal Kessler, and Laszlo Tora. 2007. “Identification of a Small TAF Complex and Its Role in the Assembly of TAF-Containing Complexes.” *PLoS ONE* 2 (3). https://doi.org/10.1371/journal.pone.0000316.

Derti, Adnan, Shreedharan Sriram, Ronghua Chen, Carol A. Rohl, Jason M. Johnson, Philip Garrett-Engele, Richard C. Stevens, Kenzie D. MacIsaac, and Tomas Babak. 2012. “A Quantitative Atlas of Polyadenylation in Five Mammals.” *Genome Research* 22 (6): 1173–83. https://doi.org/10.1101/gr.132563.111.

Dickson, Ramona, Richard J. Howard, Paul V. Viitanen, Celeste Weiss, Abdussalam Azem, Sharon P. Alldrick, R. John Ellis, and George Lorimer. 2000. “Reconstitution of Higher Plant Chloroplast Chaperonin 60 Tetradecamers Active in Protein Folding.” *Journal of Biological Chemistry* 275 (16): 11829–35. https://doi.org/10.1074/jbc.275.16.11829.

Dikstein, Rivka, Sharleen Zhou, and Robert Tjian. 1996. “Human TAF(II)105 Is a Cell Type-Specific TFIID Subunit Related to HTAF(II)130.” *Cell* 87 (1): 137–46. https://doi.org/10.1016/S0092-8674(00)81330-6.

Dimitrova, Lyudmila N., Kazushige Kuroha, Tsuyako Tatematsu, and Toshifumi Inada. 2009. “Nascent Peptide-Dependent Translation Arrest Leads to Not4p-Mediated Protein Degradation by the Proteasome.” *Journal of Biological Chemistry* 284 (16): 10343–52. https://doi.org/10.1074/jbc.M808840200.

Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. “Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions.” *Nature* 485 (7398): 376–80. https://doi.org/10.1038/nature11082.

Doma, Meenakshi K., and Roy Parker. 2007. “RNA Quality Control in Eukaryotes.” *Cell* 131 (4): 660–68. https://doi.org/10.1016/j.cell.2007.10.041.

Dostie, Josée, and Gideon Dreyfuss. 2002. “Translation Is Required to Remove Y14 from MRNAs in the Cytoplasm.” *Current Biology* 12 (13): 1060–67. https://doi.org/10.1016/S0960-9822(02)00902-8.

Douglas. 2016. “Swapping Nucleotides, Tuning Hsp70.” *Cell* 25 (3): 289–313. https://doi.org/110.1016/j.bbi.2017.04.008.

Duncan, Caia D.S., and Juan Mata. 2011. “Widespread Cotranslational Formation of Protein Complexes.” *PLoS Genetics* 7 (12). https://doi.org/10.1371/journal.pgen.1002398.

Eberle, Andrea B., Søren Lykke-Andersen, Oliver Mühlemann, and Torben Heick Jensen. 2009. “SMG6 Promotes Endonucleolytic Cleavage of Nonsense MRNA in Human Cells.” *Nature Structural and Molecular Biology* 16 (1): 49–55. https://doi.org/10.1038/nsmb.1530.

Eisner, Gottfried, Hans Georg Koch, Konstanze Beck, Joseph Brunner, and Matthias Müller. 2003. “Ligand Crowding at a Nascent Signal Sequence.” *Journal of Cell Biology* 163 (1): 35–44. https://doi.org/10.1083/jcb.200306069.

El-Saafin, Farrah, Cynthia Curry, Tao Ye, Jean Marie Garnier, Isabelle Kolb-Cheynel, Matthieu Stierle, Natalie L. Downer, et al. 2018. “Homozygous TAF8 Mutation in a Patient with Intellectual Disability Results in Undetectable TAF8 Protein, but Preserved RNA Polymerase II Transcription.” *Human Molecular Genetics* 27 (12): 2171–86. https://doi.org/10.1093/hmg/ddy126.

Erickson, Harold P., and José L. Inglesias. 1984. “A Six-Armed Oligomer Isolated from Cell Surface Fibronectin Preparations.” *Nature* 311 (5983): 267–69. https://doi.org/10.1038/311267a0.

Ernst, Jason, and Manolis Kellis. 2010. “Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome HHS Public Access Author Manuscript.” *Nat Biotechnol* 28 (8): 817–25. https://doi.org/10.1038/nbt.1662.

Etchells, Stephanie A., Anne S. Meyer, Alice Y. Yam, Anne Roobol, Yiwei Miao, Yuanlong Shao, Martin J. Carden, William R. Skach, Judith Frydman, and Arthur E. Johnson. 2005. “The Cotranslational Contacts between Ribosome-Bound Nascent Polypeptides and the Subunits of the Hetero-Oligomeric Chaperonin TRiC Probed by Photocross-Linking.” *Journal of Biological Chemistry* 280 (30): 28118–26. https://doi.org/10.1074/jbc.M504110200.

Eulalio, A., I. Behm-Ansmant, D. Schweizer, and E. Izaurralde. 2007. “P-Body Formation Is a Consequence, Not the Cause, of RNA-Mediated Gene Silencing.” *Molecular and Cellular Biology* 27 (11): 3970–81. https://doi.org/10.1128/mcb.00128-07.

Evangelista, Federica M., Anne Maglott-Roth, Matthieu Stierle, Laurent Brino, Evi Soutoglou, and László Tora. 2018. “Transcription and MRNA Export Machineries SAGA and TREX-2 Maintain Monoubiquitinated H2B Balance Required for DNA Repair.” *Journal of Cell Biology* 217 (10): 3382–97. https://doi.org/10.1083/jcb.201803074.

Fa, F L. 1973. “Chromatin Structure : Oligomers of the Histones F2AI the Molecular an F2A1 NaCI-25 Dium > H Shown to Be Homogeneous By” 637: 0–3.

Falender, Allison E., Masayuki Shimada, Yuet K. Lo, and Jo Anne S. Richards. 2005. “TAF4b, a TBP Associated Factor, Is Required for Oocyte Development and Function.” *Developmental Biology* 288 (2): 405–19. https://doi.org/10.1016/j.ydbio.2005.09.038.

Fei, Jia, Sharon E. Torigoe, Christopher R. Brown, Mai T. Khuong, George A. Kassavetis, Hinrich Boeger, and James T. Kadonaga. 2015. “The Prenucleosome, a Stable Conformational Isomer of the Nucleosome.” *Genes and Development* 29 (24): 2563–75. https://doi.org/10.1101/gad.272633.

Fenton, Wayne A, and Arthur L Horwich. 1997. “GroEL-Mediated Protein Folding.” *Protein Science*.

Ferbitz et al. 2004. “Trigger Factor in Complex with the Ribosome Forms a Molecular Cradle for Nascent Proteins.” *Nature* 431 (September): 9396–9401. https://doi.org/10.1038/nature02944.1.

Ferreira, J. P., K. W. Overton, and C. L. Wang. 2013. “Tuning Gene Expression with Synthetic Upstream Open Reading Frames.” *Proceedings of the National Academy of Sciences* 110 (28): 11284–89. https://doi.org/10.1073/pnas.1305590110.

Fiebiger, Edda, Domenico Tortorella, Marie-Helene Jouvin, Jean-Pierre Kinet, and Hidde L. Ploegh. 2005. “Cotranslational Endoplasmic Reticulum Assembly of FcεRI Controls the Formation of Functional IgE-Binding Receptors.” *The Journal of Experimental Medicine* 201 (2): 267–77. https://doi.org/10.1084/jem.20041384.

Finch, J. T., and A. Klug. 1976. “Solenoidal Model for Superstructure in Chromatin.” *Proceedings of the National Academy of Sciences* 73 (6): 1897–1901. https://doi.org/10.1073/pnas.73.6.1897.

Fischer, Tamás, Susana Rodríguez-Navarro, Gislene Pereira, Attila Rácz, Elmar Schiebel, and Ed Hurt. 2004. “Yeast Centrin Cdc31 Is Linked to the Nuclear MRNA Export Machinery.” *Nature Cell Biology* 6 (9): 840–48. https://doi.org/10.1038/ncb1163.

Forget, Diane, Andrée-Anne Lacombe, Philippe Cloutier, Racha Al-Khoury, Annie Bouchard, Mathieu Lavallée-Adam, Denis Faubert, Célia Jeronimo, Mathieu Blanchette, and Benoit Coulombe. 2010. “The Protein Interaction Network of the Human Transcription Machinery Reveals a Role for the Conserved GTPase

RPAP4/GPN1 and Microtubule Assembly in Nuclear Import and Biogenesis of RNA Polymerase II." *Molecular & Cellular Proteomics* 9 (12): 2827–39. https://doi.org/10.1074/mcp.m110.003616.

Franks, Tobias M., and Jens Lykke-Andersen. 2008. "The Control of MRNA Decapping and P-Body Formation." *Molecular Cell* 32 (5): 605–15. https://doi.org/10.1016/j.molcel.2008.11.001.

Freiman et al. 2001. "Requirement of Tissue-Selective TBP-Associated Factor TAFII105 in Ovarian Development." *Science* 293 (5537): 2084–87. https://doi.org/10.1126/science.1061935.

Fribourg, Sébastien, Christophe Romier, Sebastiaan Werten, Yann Gaël Gangloff, Arnaud Poterszman, and Dino Moras. 2001. "Dissecting the Interaction Network of Multiprotein Complexes by Pairwise Coexpression of Subunits in E. Coli." *Journal of Molecular Biology* 306 (2): 363–73. https://doi.org/10.1006/jmbi.2000.4376.

Frontini, M., E. Soutoglou, M. Argentini, C. Bole-Feysot, B. Jost, E. Scheer, and L. Tora. 2005. "TAF9b (Formerly TAF9L) Is a Bona Fide TAF That Has Unique and Overlapping Roles with TAF9." *Molecular and Cellular Biology* 25 (11): 4638–49. https://doi.org/10.1128/mcb.25.11.4638-4649.2005.

Frydman, Judith. 2001. "FOLDING OF NEWLY TRANSLATED PROTEINS IN VIVO : The Role of Molecular Chaperones."

Fuda, Nicholas J., M. Behfar Ardehali, and John T. Lis. 2009. "Defining Mechanisms That Regulate RNA Polymerase II Transcription in Vivo." *Nature* 461 (7261): 186–92. https://doi.org/10.1038/nature08449.

Fyodorov, Dmitry V., Bing Rui Zhou, Arthur I. Skoultchi, and Yawen Bai. 2018. "Emerging Roles of Linker Histones in Regulating Chromatin Structure and Function." *Nature Reviews Molecular Cell Biology* 19 (3): 192–206. https://doi.org/10.1038/nrm.2017.94.

Gallardo, Mercedes, Rosa Luna, Hediye Erdjument-Bromage, Paul Tempst, and Andrés Aguilera. 2003. "Nab2p and the Thp1p-Sac3p Complex Functionally Interact at the Interface between Transcription and MRNA Metabolism." *Journal of Biological Chemistry* 278 (26): 24225–32. https://doi.org/10.1074/jbc.M302900200.

Galy, Vincent, Olivier Gadal, Micheline Fromont-Racine, Alper Romano, Alain Jacquier, and Ulf Nehrbass. 2004. "Nuclear Retention of Unspliced MRNAs in Yeast Is Mediated by Perinuclear Mlp1." *Cell* 116 (1): 63–73. https://doi.org/10.1016/S0092-8674(03)01026-2.

Gangloff, Yann-gaël, Christophe Romier, Sylvie Thuault, Sebastiaan Werten, Irwin Davidson, and Irwin Davidson. 2001. "TAF II 250: Three Enzymatic Activities in a Single Protein" 26 (4): 250–57. https://ac.els-cdn.com/S0968000400017412/1-s2.0-S0968000400017412-main.pdf?_tid=217ef13e-1658-11e8-850e-00000aacb35f&acdnat=1519143021_b0f761910f7269d7e4153b167f42da93.

Gautschi, M., A. Mun, S. Ross, and S. Rospert. 2002. "A Functional Chaperone Triad on the Yeast Ribosome." *Proceedings of the National Academy of Sciences* 99 (7): 4209–14. https://doi.org/10.1073/pnas.062048599.

Gazdag, Emese, Aleksandar Rajkovic, Maria Elena Torres-Padilla, and Làszlò Tora. 2007. “Analysis of TATA-Binding Protein 2 (TBP2) and TBP Expression Suggests Different Roles for the Two Proteins in Regulation of Gene Expression during Oogenesis and Early Mouse Development.” *Reproduction* 134 (1): 51–62. https://doi.org/10.1530/REP-06-0337.

Gehring, Niels H., Styliani Lamprinaki, Matthias W. Hentze, and Andreas E. Kulozik. 2009. “The Hierarchy of Exon-Junction Complex Assembly by the Spliceosome Explains Key Features of Mammalian Nonsense-Mediated MRNA Decay.” *PLoS Biology* 7 (5). https://doi.org/10.1371/journal.pbio.1000120.

Genevaux, Pierre, France Keppel, Françoise Schwager, Petra S. Langendijk-Genevaux, F. Ulrich Hartl, and Costa Georgopoulos. 2004. “In Vivo Analysis of the Overlapping Functions of DnaK and Trigger Factor.” *EMBO Reports* 5 (2): 195–200. https://doi.org/10.1038/sj.embor.7400067.

Ghosh, Sankar, Ann M. Gifford, Lise R. Riviere, Paul Tempst, Garry P. Nolan, and David Baltimore. 1990. “Cloning of the P50 DNA Binding Subunit of NF-KB: Homology to Rel and Dorsal.” *Cell* 62 (5): 1019–29. https://doi.org/10.1016/0092-8674(90)90276-K.

Gidalevitz, Tali, Anat Ben-Zvi, Kim H. Ho, Heather R. Brignull, and Richard Morimoto. 2006. “Progressive Disruption of Cellular Protein Folding in Models of Polyglutamine Diseases.” *Science Reports* 347 (March): 1471–75.

Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, and Cassell GH. 2000. “The Complete Sequence of the Mucosal Pathogen Ureaplasma Urealyticum.” *Nature* 407 (6805): 757–62.

Goodrich, James A., and Robert Tjian. 1994. “Transcription Factors IIE and IIH and ATP Hydrolysis Direct Promoter Clearance by RNA Polymerase II.” *Cell* 77 (1): 145–56. https://doi.org/10.1016/0092-8674(94)90242-9.

Graille, Marc, and Bertrand Séraphin. 2012. “Surveillance Pathways Rescuing Eukaryotic Ribosomes Lost in Translation.” *Nature Reviews Molecular Cell Biology* 13 (11): 727–35. https://doi.org/10.1038/nrm3457.

Grant P. et al. 1997. “Yeast Gcn5 Functions in Two Multisubunit Complexes to Acetylate Nucleosomal Histones: Characterization of an Ada Complex and the SAGA (Spt/Ada) Complex,” 1640–50.

Greenwald, Jason, and Roland Riek. 2010. “Biology of Amyloid: Structure, Function, and Regulation.” *Structure* 18 (10): 1244–60. https://doi.org/10.1016/j.str.2010.08.009.

Grewal, Shiv I.S., and Songtao Jia. 2007. “Heterochromatin Revisited.” *Nature Reviews Genetics* 8 (1): 35–46. https://doi.org/10.1038/nrg2008.

Gupta et al. 2017. “Architecture of TAF11 / TAF13 / TBP Complex Suggests Novel Regulation Properties of General Transcription Factor,” 1–31.

Haberle, Vanja, and Alexander Stark. 2018. “Eukaryotic Core Promoters and the Functional Basis of Transcription Initiation.” *Nature Reviews Molecular Cell Biology* 19 (10): 621–37. https://doi.org/10.1038/s41580-018-0028-8.

Hahn, Steven. 2004. “Structure and Mechanism of the RNA Polymerase II

Transcription Machinery." *Nature Structural and Molecular Biology* 11 (5): 394–403. https://doi.org/10.1038/nsmb763.

Haimov, Ora, Hadar Sinvani, and Rivka Dikstein. 2015. "Cap-Dependent, Scanning-Free Translation Initiation Mechanisms." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1849 (11): 1313–18. https://doi.org/10.1016/j.bbagrm.2015.09.006.

Haimovich, Gal, Christopher M. Ecker, Margaret C. Dunagin, Elliott Eggan, Arjun Raj, Jeffrey E. Gerst, and Robert H. Singer. 2017. "Intercellular MRNA Trafficking via Membrane Nanotube-like Extensions in Mammalian Cells." *Proceedings of the National Academy of Sciences* 114 (46): E9873–82. https://doi.org/10.1073/pnas.1706365114.

Halbach, André, Haidi Zhang, Agnieszka Wengi, Zofia Jablonska, Isabel M.L. Gruber, Regula E. Halbeisen, Pierre Marie Dehé, et al. 2009. "Cotranslational Assembly of the Yeast SET1C Histone Methyltransferase Complex." *EMBO Journal* 28 (19): 2959–70. https://doi.org/10.1038/emboj.2009.240.

Halder, Kangkan, Markus Wieland, and Jörg S. Hartig. 2009. "Predictable Suppression of Gene Expression by 5′-UTR-Based RNA Quadruplexes." *Nucleic Acids Research* 37 (20): 6811–17. https://doi.org/10.1093/nar/gkp696.

Halees, Anason S., Edward Hitti, Maher Al-Saif, Linah Mahmoud, Irina A. Vlasova-St Louis, Daniel J. Beisang, Paul R. Bohjanen, and Khalid S.A. Khabar. 2011. "Global Assessment of GU-Rich Regulatory Content and Function in the Human Transcriptome." *RNA Biology* 8 (4). https://doi.org/10.4161/rna.8.4.16283.

Hall, Zoe, Helena Hernández, Joseph A. Marsh, Sarah A. Teichmann, and Carol V. Robinson. 2013. "The Role of Salt Bridges, Charge Density, and Subunit Flexibility in Determining Disassembly Routes of Protein Complexes." *Structure* 21 (8): 1325–37. https://doi.org/10.1016/j.str.2013.06.004.

Hanebuth, Marie A., Roman Kityk, Sandra J. Fries, Alok Jain, Allison Kriel, Veronique Albanese, Tancred Frickey, et al. 2016. "Multivalent Contacts of the Hsp70 Ssb Contribute to Its Architecture on Ribosomes and Nascent Chain Interaction." *Nature Communications* 7: 1–13. https://doi.org/10.1038/ncomms13695.

Hansen, Stig K., Shinako Takada, Raymond H. Jacobson, John T. Lis, and Robert Tjian. 1997. "Transcription Properties of a Cell Type-Specific TATA-Binding Protein, TRF." *Cell* 91 (1): 71–83. https://doi.org/10.1016/S0092-8674(01)80010-6.

Harrison, Alice Ford, and James Shorter. 2017. "RNA-Binding Proteins with Prion-like Domains in Health and Disease." *Biochemical Journal* 474 (8): 1417–38. https://doi.org/10.1042/bcj20160499.

Hartl, F. Ulrich. 2017. "Unfolding the Chaperone Story." *Molecular Biology of the Cell* 28 (22): 2919–23. https://doi.org/10.1091/mbc.e17-07-0480.

Hartl, F. Ulrich, Andreas Bracher, and Manajit Hayer-Hartl. 2011. "Molecular Chaperones in Protein Folding and Proteostasis." *Nature* 475 (7356): 324–32. https://doi.org/10.1038/nature10317.

Hartl, F. 2002. “Molecular Chaperones in the Cytosol: From Nascent Chainto Folded Protein.” *Science* 295 (5561): 1852. https://doi.org/10.1126/science.1068408.

Haslberger, Tobias, Bernd Bukau, and Axel Mogk. 2010. “Towards a Unifying Mechanism for ClpB/Hsp104-Mediated Protein Disaggregation and Prion PropagationThis Paper Is One of a Selection of Papers Published in This Special Issue Entitled 8th International Conference on AAA Proteins and Has Undergone the Journa.” *Biochemistry and Cell Biology* 88 (1): 63–75. https://doi.org/10.1139/o09-118.

Hayer-hartl, Author Manajit K, Jörg Martin, and F Ulrich Hartl. 2009. “Asymmetrical Interaction of GroEL and GroES in the ATPase Cycle of Assisted Protein Folding Published by : American Association for the Advancement of Science Stable URL : Http://Www.Jstor.Org/Stable/2888494.” *Advancement Of Science* 269 (5225): 836–41.

Heiman, Myriam, Ruth Kulicke, Robert J. Fenster, Paul Greengard, and Nathaniel Heintz. 2014. “Cell Type-Specific MRNA Purification by Translating Ribosome Affinity Purification (TRAP).” *Nature Protocols* 9 (6): 1282–91. https://doi.org/10.1038/nprot.2014.085.

Helmlinger, Dominique, Samuel Marguerat, Judit Villén, Danielle L. Swaney, Steven P. Gygi, Jürg Bähler, and Fred Winston. 2011. “Tra1 Has Specific Regulatory Roles, Rather than Global Functions, within the SAGA Co-Activator Complex.” *EMBO Journal* 30 (14): 2843–52. https://doi.org/10.1038/emboj.2011.181.

Helmlinger, Dominique, and László Tora. 2017. “Sharing the SAGA.” *Trends in Biochemical Sciences* 42 (11): 850–61. https://doi.org/10.1016/j.tibs.2017.09.001.

Heng, Zhu, Metin Bilgin, Rhonda Bangham, David Hall, Antonio Casamayor, Paul Bertone, Ning Lan, et al. 2001. “Global Analysis of Protein Activities Using Proteome Chips.” *Science* 293 (5537): 2101–5. https://doi.org/10.1126/science.1062191.

Henras, A. K., J. Soudet, M. Gérus, S. Lebaron, M. Caizergues-Ferrer, A. Mougin, and Y. Henry. 2008. “The Post-Transcriptional Steps of Eukaryotic Ribosome Biogenesis.” *Cellular and Molecular Life Sciences* 65 (15): 2334–59. https://doi.org/10.1007/s00018-008-8027-0.

Hentze, Matthias W., S. Wright Caughman, Tracey A. Rouault, Javier G. Barriocanal, Andrew Dancis, Joe B. Harford, and Richard D. Klausner. 1987. “Identification of the Iron-Responsive Element for the Translational Regulation of Human Ferritin MRNA.” *Science* 238 (4833): 1570–73. https://doi.org/10.1126/science.3685996.

Herl, Lauren, Alberto Lleo, Anne V. Thomas, Andrew C. Nyborg, Karen Jansen, Todd E. Golde, Bradley T. Hyman, and Oksana Berezovska. 2006. “Detection of Presenilin-1 Homodimer Formation in Intact Cells Using Fluorescent Lifetime Imaging Microscopy.” *Biochemical and Biophysical Research Communications* 340 (2): 668–74. https://doi.org/10.1016/j.bbrc.2005.12.063.

Hernández, Greco, Michael Altmann, and Paul Lasko. 2010. “Origins and Evolution of the Mechanisms Regulating Translation Initiation in Eukaryotes.” *Trends in Biochemical Sciences* 35 (2): 63–73. https://doi.org/10.1016/j.tibs.2009.10.009.

Hernández, Helena, and Carol V. Robinson. 2007. “Determining the Stoichiometry and Interactions of Macromolecular Assemblies from Mass Spectrometry.” *Nature Protocols* 2 (3): 715–26. https://doi.org/10.1038/nprot.2007.73.

Herrera, Francisco J., Teppei Yamaguchi, Henk Roelink, and Robert Tjian. 2014. “Core Promoter Factor TAF9B Regulates Neuronal Gene Expression.” *ELife* 2014 (3): 1–17. https://doi.org/10.7554/eLife.02559.001.

Ho, Y, Y Ho, A Gruhler, A Gruhler, A Heilbut, A Heilbut, G D Bader, et al. 2002. “Systematic Identification of Protein Complexes in Saccharomyces Cerevisiae by Mass Spectrometry.” *Nature* 415 (6868): 180–83. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11805837%5Cnpapers3://publication/uuid/BB1850BC-BC4C-4AF7-9352-F1FF171C4B68.

Hoffmann, Alexander, and Robert G. Roeder. 2002. “Cloning and Characterization of Human TAF20/15.” *Journal of Biological Chemistry* 271 (30): 18194–202. https://doi.org/10.1074/jbc.271.30.18194.

Hogg, J. Robert, and Stephen P. Goff. 2010. “Upf1 Senses 3’UTR Length to Potentiate MRNA Decay.” *Cell* 143 (3): 379–89. https://doi.org/10.1016/j.cell.2010.10.005.

Holmes, M C, and R Tjian. 2000. “Promoter-Selective Properties of the TBP-Related Factor TRF1.” *Science (New York, N.Y.)* 288 (5467): 867–70. https://doi.org/10.1126/science.288.5467.867.

Holmgren, Anne, Delphine Bouhy, Vicky De Winter, Bob Asselbergh, Jean Pierre Timmermans, Joy Irobi, and Vincent Timmerman. 2013. “Charcot-Marie-Tooth Causing HSPB1 Mutations Increase Cdk5-Mediated Phosphorylation of Neurofilaments.” *Acta Neuropathologica* 126 (1): 93–108. https://doi.org/10.1007/s00401-013-1133-6.

Hoque, Mainul, Zhe Ji, Dinghai Zheng, Wenting Luo, Wencheng Li, Bei You, Ji Yeon Park, Ghassan Yehia, and Bin Tian. 2013. “Analysis of Alternative Cleavage and Polyadenylation by 3′ Region Extraction and Deep Sequencing.” *Nature Methods* 10 (2): 133–39. https://doi.org/10.1038/nmeth.2288.

Horwich, Arthur L., Wayne A. Fenton, Eli Chapman, and George W. Farr. 2007. “Two Families of Chaperonin: Physiology and Mechanism.” *Annual Review of Cell and Developmental Biology* 23 (1): 115–45. https://doi.org/10.1146/annurev.cellbio.23.090506.123555.

Hou, Chunhui, Li Li, Zhaohui S. Qin, and Victor G. Corces. 2012. “Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains.” *Molecular Cell* 48 (3): 471–84. https://doi.org/10.1016/j.molcel.2012.08.031.

Houry, Walid A, Edouard Bertrand, and Benoit Coulombe. 2018. “The PAQosome , an Chaperone for Quaternary Structure Formation R2TP Core PIH1D1 POLR2E PFDL Module.” *Trends in Biochemical Sciences* xx (Figure 1): 1–5. https://doi.org/10.1016/j.tibs.2017.11.001.

Hu, Shi, Yuna Sun, Yanchun Meng, Xiaoze Wang, Weili Yang, Wenyan Fu, Huaizu Guo, et al. 2015. “Molecular Architecture of the ErbB2 Extracellular Domain

Homodimer." *Oncotarget* 6 (3). https://doi.org/10.18632/oncotarget.2713.

Hu, Zhenhua, and Wee-Wei Tee. 2017. "Enhancers and Chromatin Structures: Regulatory Hubs in Gene Expression and Diseases." *Bioscience Reports* 37 (2): BSR20160183. https://doi.org/10.1042/bsr20160183.

Huch, Susanne, and Tracy Nissan. 2017. "An MRNA Decapping Mutant Deficient in P Body Assembly Limits MRNA Stabilization in Response to Osmotic Stress." *Scientific Reports* 7 (March): 1–13. https://doi.org/10.1038/srep44395.

Huisinga, Kathryn L., Brent Brower-Toland, and Sarah C R Elgin. 2006. "The Contradictory Definitions of Heterochromatin: Transcription and Silencing." *Chromosoma* 115 (2): 110–22. https://doi.org/10.1007/s00412-006-0052-x.

HURWITZ, J, J J FURTH, M ANDERS, P J ORTIZ, and J T AUGUST. 1961. "The Enzymatic Incorporation of Ribonucleotides into RNA and the Role of DNA." *Cold Spring Harbor Symposia on Quantitative Biology* 26: 91–100.

Hwang, Jungwook, Hanae Sato, Yalan Tang, Daiki Matsuda, and Lynne E. Maquat. 2010. "UPF1 Association with the Cap-Binding Protein, CBP80, Promotes Nonsense-Mediated MRNA Decay at Two Distinct Steps." *Molecular Cell* 39 (3): 396–409. https://doi.org/10.1016/j.molcel.2010.07.004.

Ingen, Hugo van, Frederik M.A. van Schaik, Hans Wienk, Joost Ballering, Holger Rehmann, Annemarie C. Dechesne, John A.W. Kruijzer, Rob M.J. Liskamp, H. Th Marc Timmers, and Rolf Boelens. 2008. "Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3." *Structure* 16 (8): 1245–56. https://doi.org/10.1016/j.str.2008.04.015.

Ingolia, Nicholas T., Gloria A. Brar, Silvia Rouskin, Anna M. McGeachy, and Jonathan S. Weissman. 2012. "The Ribosome Profiling Strategy for Monitoring Translation in Vivo by Deep Sequencing of Ribosome-Protected MRNA Fragments." *Nature Protocols* 7 (8): 1534–50. https://doi.org/10.1038/nprot.2012.086.

Irina Vlasova-St. Louis and Calandra Sagarsky. 2018. "Irina Vlasova-St. Louis and Calandra Sagarsky." *Long-Haul Travel Motivation by International Tourist to Penang* i (tourism): 13.

Isaacs, W. B., and A. B. Fulton. 1987. "Cotranslational Assembly of Myosin Heavy Chain in Developing Cultured Skeletal Muscle." *Proceedings of the National Academy of Sciences* 84 (17): 6174–78. https://doi.org/10.1073/pnas.84.17.6174.

Isken, Olaf, and Lynne E. Maquat. 2007. "Quality Control of Eukaryotic MRNA: Safeguarding Cells from Abnormal MRNA Function." *Genes and Development* 21 (15): 1833–56. https://doi.org/10.1101/gad.1566807.

Iwafuchi-Doi, Makiko, and Kenneth S. Zaret. 2016. "Cell Fate Control by Pioneer Transcription Factors." *Development* 143 (11): 1833–37. https://doi.org/10.1242/dev.133900.

Iwre, David Balti. 1970. "RNA-Dependent DNA Polymerase in INFECTION of Sensitive Cells by RNA Sarcoma Viruses" 226: 1211–13.

Izaurralde, E., J. Sauliere, E. Huntzinger, M. Fauser, and I. Kashima. 2008. "SMG6

Is the Catalytic Endonuclease That Cleaves MRNAs Containing Nonsense Codons in Metazoan." *Rna* 14 (12): 2609–17. https://doi.org/10.1261/rna.1386208.

Jacobson, Raymond H, Andreas G Ladurner, David S King, and Robert Tjian. 2000. "Structure and Function of a Human TAF." *Science* 288 (May): 1422–25.

Jacq, Xavier, Christel Brou, Yves Lutz, Irwin Davidson, Pierre Chambon, and Laszlo Tora. 1994. "Human TAFII30 Is Present in a Distinct TFIID Complex and Is Required for Transcriptional Activation by the Estrogen Receptor." *Cell* 79 (1): 107–17. https://doi.org/10.1016/0092-8674(94)90404-9.

Jaenicke, Rainer. 1991. "Perspectives in Biochemistry Protein Folding : Local Structures , Domains , Subunits , and Assemblies + ,*." *Society* 30 (13).

Jaiswal, H., C. Conz, H. Otto, T. Wolfle, E. Fitzke, M. P. Mayer, and S. Rospert. 2011. "The Chaperone Network Connected to Human Ribosome-Associated Complex." *Molecular and Cellular Biology* 31 (6): 1160–73. https://doi.org/10.1128/mcb.00986-10.

Jens, Marvin, and Nikolaus Rajewsky. 2015. "Competition between Target Sites of Regulators Shapes Post-Transcriptional Gene Regulation." *Nature Reviews Genetics* 16 (2): 113–26. https://doi.org/10.1038/nrg3853.

Jeronimo, Célia, Diane Forget, Annie Bouchard, Qintong Li, Gordon Chua, Christian Poitras, Cynthia Thérien, et al. 2007. "Systematic Analysis of the Protein Interaction Network for the Human Transcription Machinery Reveals the Identity of the 7SK Capping Enzyme." *Molecular Cell* 27 (2): 262–74. https://doi.org/10.1016/j.molcel.2007.06.027.

Jin, Fulai, Yan Li, Jesse R. Dixon, Siddarth Selvaraj, Zhen Ye, Ah Young Lee, Chia An Yen, Anthony D. Schmitt, Celso A. Espinoza, and Bing Ren. 2013. "A High-Resolution Map of the Three-Dimensional Chromatin Interactome in Human Cells." *Nature* 503 (7475): 290–94. https://doi.org/10.1038/nature12644.

Jones, Susan, and Janet M Thornton. 1995. "PROTEIN-PROTEIN INTERACTIONS: A REVIEW OF PROTEIN DIMER STRUCTURES." *Prog. Biophys. Molec. Biol.* 63 (94): 31–65. https://doi.org/10.1016/0079-6107(94)00008-W.

Jonkers, Iris, and John T. Lis. 2015. "Getting up to Speed with Transcription Elongation by RNA Polymerase II." *Nature Reviews Molecular Cell Biology* 16 (3): 167–77. https://doi.org/10.1038/nrm3953.

Kamenova, Ivanka, Pooja Mukherjee, Sascha Conic, Florian Mueller, Farrah El-Saafin, Paul Bardot, Jean Marie Garnier, et al. 2019. "Co-Translational Assembly of Mammalian Nuclear Multisubunit Complexes." *Nature Communications* 10 (1): 25–28. https://doi.org/10.1038/s41467-019-09749-y.

Kampinga, Harm H., and Elizabeth A. Craig. 2010. "The HSP70 Chaperone Machinery: J Proteins as Drivers of Functional Specificity." *Nature Reviews Molecular Cell Biology* 11 (8): 579–92. https://doi.org/10.1038/nrm2941.

Kanno, Tatsuo, Bruno Huettel, M. Florian Mette, Werner Aufsatz, Estelle Jaligot, Lucia Daxinger, David P. Kreil, Marjori Matzke, and Antonius J.M. Matzke. 2005. "Atypical RNA Polymerase Subunits Required for RNA-Directed DNA

Methylation." *Nature Genetics* 37 (7): 761–65. https://doi.org/10.1038/ng1580.

Kaplan, Craig D., Michael J. Holland, and Fred Winston. 2005. "Interaction between Transcription Elongation Factors and MRNA 3′-End Formation at the Saccharomyces Cerevisiae GAL10-GAL7 Locus." *Journal of Biological Chemistry* 280 (2): 913–22. https://doi.org/10.1074/jbc.M411108200.

Karamyshev, Andrey L., and Zemfira N. Karamysheva. 2018. "Lost in Translation: Ribosome-Associated MRNA and Protein Quality Controls." *Frontiers in Genetics* 9 (OCT): 1–13. https://doi.org/10.3389/fgene.2018.00431.

Karamyshev, Andrey L., Anna E. Patrick, Zemfira N. Karamysheva, Dustin S. Griesemer, Henry Hudson, Sandra Tjon-Kon-Sang, Ingmarie Nilsson, et al. 2014. "Inefficient SRP Interaction with a Nascent Chain Triggers a MRNA Quality Control Pathway." *Cell* 156 (1–2): 146–57. https://doi.org/10.1016/j.cell.2013.12.017.

Kashima, Isao, Akio Yamashita, Natsuko Izumi, Naoyuki Kataoka, Ryo Morishita, Shinichi Hoshino, Mutsuhito Ohno, Gideon Dreyfuss, and Shigeo Ohno. 2006. "Nonsense-Mediated MRNA Decay (NMD) Is a Surveillance Mechanism That Degrades MRNA Containing Premature Termination Codons (PTCs). In Mammalian Cells, Recognition of PTCs Requires Translation and Depends on the Presence on the MRNA with the Splicing-Depend." *Genes & Development*, 355–67. https://doi.org/10.1101/gad.1389006.phenotype.

Kassem, Sari, Zoltan Villanyi, and Martine A. Collart. 2017. "Not5-Dependent Co-Translational Assembly of Ada2 and Spt20 Is Essential for Functional Integrity of SAGA." *Nucleic Acids Research* 45 (3): 1186–99. https://doi.org/10.1093/nar/gkw1059.

Katahira, Jun. 2015. "Nuclear Export of Messenger RNA." *Genes* 6 (2): 163–84. https://doi.org/10.3390/genes6020163.

Katinka, Michaël D., Simone Duprat, Emmanuel Cornillott, Guy Méténler, Fabienne Thomarat, Gérard Prensier, Valérie Barbe, et al. 2001. "Genome Sequence and Gene Compaction of the Eukaryote Parasite Encephalitozoon Cuniculi." *Nature* 414 (6862): 450–53. https://doi.org/10.1038/35106579.

Kaufmann, Jörg, C. Peter Verrijzer, John Shao, and Stephen T. Smale. 1996. "CIF, an Essential Cofactor for TFIID-Dependent Initiator Function." *Genes and Development* 10 (7): 873–86. https://doi.org/10.1101/gad.10.7.873.

Kebede, Adam F., Robert Schneider, and Sylvain Daujat. 2015. "Novel Types and Sites of Histone Modifications Emerge as Players in the Transcriptional Regulation Contest." *FEBS Journal* 282 (9): 1658–74. https://doi.org/10.1111/febs.13047.

Kedinger, C., M. Gniazdowski, J. L. Mandel, F. Gissinger, and P. Chambon. 1970. "α-Amanitin: A Specific Inhibitor of One of Two DNA-Dependent RNA Polymerase Activities from Calf Thymus." *Biochemical and Biophysical Research Communications* 38 (1): 165–71. https://doi.org/10.1016/0006-291X(70)91099-5.

Keene, Jack D., Jordan M. Komisarow, and Matthew B. Friedersdorf. 2006. "RIP-Chip: The Isolation and Identification of MRNAs, MicroRNAs and Protein

Components of Ribonucleoprotein Complexes from Cell Extracts." *Nature Protocols* 1 (1): 302–7. https://doi.org/10.1038/nprot.2006.47.

Képès, François. 1996. "The '+70 Pause': Hypothesis of a Translational Control of Membrane Protein Assembly." *Journal of Molecular Biology* 262 (2): 77–86. https://doi.org/10.1006/jmbi.1996.0500.

Kervestin, Stephanie, and Allan Jacobson. 2012. "NMD: A Multifaceted Response to Premature Translational Termination." *Nature Reviews Molecular Cell Biology* 13 (11): 700–712. https://doi.org/10.1038/nrm3454.

Kervestin, Stephanie, David a D.A. Mangus, Shubhendu Ghosh, Allan Jacobson, Nadia Amrani, and Robin Ganesan. 2004. "A Faux 3′-UTR Promotes Aberrant Termination and Triggers Nonsense-Mediated MRNA Decay." *Nature* 432 (7013): 112–118. https://doi.org/10.1038/nature03060.

Kevin G.McLure, PatrickW.K.Lee. 1998. "How P53 Binds DNA as a Tetramer." *The EMBO Journal* 17 (12): 3342–50. https://doi.org/10.1093/emboj/17.12.3342.

Kieran, Mark, Volker Blank, Frédérique Logeat, Joël Vandekerckhove, Frledrich Lottspeich, Odile Le Bail, Manuela B. Urban, Philippe Kourilsky, Patrick A. Baeuerle, and Alain Israël. 1990. "The DNA Binding Subunit of NF-KB Is Identical to Factor KBF1 and Homologous to the Rel Oncogene Product." *Cell* 62 (5): 1007–18. https://doi.org/10.1016/0092-8674(90)90275-J.

Kiho, Y., and A. Rich. 1964. "Induced Enzyme Formed on Bacterial Polyribosomes." *Proceedings of the National Academy of Sciences* 51 (1): 111–18. https://doi.org/10.1073/pnas.51.1.111.

Kim, Dae In, and Kyle J. Roux. 2016. "Filling the Void: Proximity-Based Labeling of Proteins in Living Cells." *Trends in Cell Biology* 26 (11): 804–17. https://doi.org/10.1016/j.tcb.2016.09.004.

Kim, Jungmook, Patricia Gamble Klein, and John E. Mullet. 1991. "Ribosomes Pause at Specific Sites during Synthesis of Membrane-Bound Chloroplast Reaction Center Protein D1." *Journal of Biological Chemistry* 266 (23): 14931–38.

Kim, Tae Kyung, Richard H. Ebright, and Danny Reinberg. 2000. "Mechanism of ATP-Dependent Promoter Melting by Transcription Factor IIH." *Science* 288 (5470): 1418–21. https://doi.org/10.1126/science.288.5470.1418.

Kim, Yujin E., Mark S. Hipp, Andreas Bracher, Manajit Hayer-Hartl, and F. Ulrich Hartl. 2013. *Molecular Chaperone Functions in Protein Folding and Proteostasis*. *Annual Review of Biochemistry*. Vol. 82. https://doi.org/10.1146/annurev-biochem-060208-092442.

Klumpp, Martin, Wolfgang Baumeister, and Lars Oliver Essen. 1997. "Structure of the Substrate Binding Domain of the Thermosome, an Archaeal Group II Chaperonin." *Cell* 91 (2): 263–70. https://doi.org/10.1016/S0092-8674(00)80408-0.

Kocabas, Arif, Terence Duarte, Saranya Kumar, and Mary A. Hynes. 2015. "Widespread Differential Expression of Coding Region and 3' UTR Sequences in Neurons and Other Tissues." *Neuron* 88 (6): 1149–56.

https://doi.org/10.1016/j.neuron.2015.10.048.

Koch, Birgit, Bianca Nijmeijer, Moritz Kueblbeck, Yin Cai, Nike Walther, and Jan Ellenberg. 2018. “Generation and Validation of Homozygous Fluorescent Knock-in Cells Using CRISPR-Cas9 Genome Editing.” *Nature Protocols* 13 (6): 1465–87. https://doi.org/10.1038/nprot.2018.042.

Kolesnikova, Olga, Adam Ben-Shem, Jie Luo, Jeff Ranish, Patrick Schultz, and Gabor Papai. 2018. “Molecular Structure of Promoter-Bound Yeast TFIID.” *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-07096-y.

Koplin, Ansgar, Steffen Preissler, Yulia Llina, Miriam Koch, Annika Scior, Marc Erhardt, and Elke Deuerling. 2010. “A Dual Function for Chaperones SSB-RAC and the NAC Nascent Polypeptide-Associated Complex on Ribosomes.” *Journal of Cell Biology* 189 (1): 57–68. https://doi.org/10.1083/jcb.200910074.

Korkhin, Yakov, Ulug M. Unligil, Otis Littlefield, Pamlea J. Nelson, David I. Stuart, Paul B. Sigler, Stephen D. Bell, and Nicola G.A. Abrescia. 2009. “Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure.” *PLoS Biology* 7 (5). https://doi.org/10.1371/journal.pbio.1000102.

Kornberg, Arthur. 1961. “ACID BY AN ENZYME FROM ESCHERICHIA COLI * Synthesis of Polyribonucleotides , Does Not by Itself Provide a Mechanism for The.”

Kornberg, Rd. 1974. “Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin Structure Is Based on a Repeating Unit of Eight.” *Science* 184: 868–71.

Kornberg, Roger D. 1987. “From Saccharomyces” 84 (December): 8839–43.

Kostenko, Sergiy, Mona Johannessen, and Ugo Moens. 2009. “PKA-Induced F-Actin Rearrangement Requires Phosphorylation of Hsp27 by the MAPKAP Kinase MK5.” *Cellular Signalling* 21 (5): 712–18. https://doi.org/10.1016/j.cellsig.2009.01.009.

Kotani, Hirokazu, Yoshinobu Toda, Krystyna Mazan-mamczarz, Eva-christina Mueller, Albrecht Otto, Lena Disch, Mark Norman, et al. 2009a. “Novel Roles of Hakai in Cell Proliferation and Oncogenesis.” *Molecular Biology of the Cell* 20 (October): 3533–42. https://doi.org/10.1091/mbc.E08.

Kotani. 2009b. “Novel Roles of Hakai in Cell Proliferation and Oncogenesis.” *Molecular Biology of the Cell* 20 (December): 3533–42. https://doi.org/10.1091/mbc.E08.

Koutelou, Evangelia, Calley L. Hirsch, and Sharon Y R Dent. 2010. “Multiple Faces of the SAGA Complex.” *Current Opinion in Cell Biology* 22 (3): 374–82. https://doi.org/10.1016/j.ceb.2010.03.005.

Kozak, Marilyn. 1978. “How Do Eucaryotic Ribosomes Select Initiation Regions in Messenger RNA?” *Cell* 15 (4): 1109–23. https://doi.org/10.1016/0092-8674(78)90039-9.

Kozak. 1986. “Influences of MRNA Secondary Structure on Initiation by Eukaryotic Ribosomes (Translational Control/Insertion Mutagenesl/Scanning Model/RNA Stem-and-Loop Structures/Preproinsulin).” *Proc. Nati. Acad. Sci. USA* 83 (May):

2850–54. http://www.pnas.org/content/pnas/83/9/2850.full.pdf.

Kramer et al. 2002. “L23 Protein Functions as a Chaperone Docking Site on the Ribosome.” *Nature*, 171–74.

Krasnov, Aleksey N., Marina Yu Mazina, Julia V. Nikolenko, and Nadezhda E. Vorobyeva. 2016. “On the Way of Revealing Coactivator Complexes Cross-Talk during Transcriptional Activation.” *Cell and Bioscience* 6 (1): 1–14. https://doi.org/10.1186/s13578-016-0081-y.

Kristjánsdóttir, Katla, Elizabeth A. Fogarty, and Andrew Grimson. 2015. “Systematic Analysis of the Hmga2 3’ UTR Identifies Many Independent Regulatory Sequences and a Novel Interaction between Distal Sites.” *Rna* 21 (7): 1346–60. https://doi.org/10.1261/rna.051177.115.

Kuehner, Jason N., Erika L. Pearson, and Claire Moore. 2011. “Unravelling the Means to an End: RNA Polymerase II Transcription Termination.” *Nature Reviews Molecular Cell Biology* 12 (5): 283–94. https://doi.org/10.1038/nrm3098.

Kuroha, Kazushige, Tsuyako Tatematsu, and Toshifumi Inada. 2009. “Upf1 Stimulates Degradation of the Product Derived from Aberrant Messenger RNA Containing a Specific Nonsense Mutation by the Proteasome.” *EMBO Reports* 10 (11): 1265–71. https://doi.org/10.1038/embor.2009.200.

Kwak, Hojoong. 2013. “Initiation and Pausing.” *Science* 339: 950. https://doi.org/10.1126/science.1229386.

Lai, William K.M., and B. Franklin Pugh. 2017. “Understanding Nucleosome Dynamics and Their Links to Gene Expression and DNA Replication.” *Nature Reviews Molecular Cell Biology* 18 (9): 548–62. https://doi.org/10.1038/nrm.2017.47.

Lang, G., J. Bonnet, D. Umlauf, K. Karmodiya, J. Koffler, M. Stierle, D. Devys, and L. Tora. 2011. “The Tightly Controlled Deubiquitination Activity of the Human SAGA Complex Differentially Modifies Distinct Gene Regulatory Elements.” *Molecular and Cellular Biology* 31 (18): 3734–44. https://doi.org/10.1128/mcb.05231-11.

Langdon, Erin M, Yupeng Qiu, Amirhossein Ghanbari Niaki, Grace A Mclaughlin, Chase A Weidmann, Therese M Gerbich, Jean A Smith, et al. 2018. “Phase Separation” 1 (May): 922–27.

Langemeyer, Lars, and Christian Ungermann. 2015. “BORC and BLOC-1: Shared Subunits in Trafficking Complexes.” *Developmental Cell* 33 (2): 121–22. https://doi.org/10.1016/j.devcel.2015.04.008.

Law, Martin J., Karen M. Lower, Hsiao P.J. Voon, Jim R. Hughes, David Garrick, Vip Viprakasit, Matthew Mitson, et al. 2010. “ATR-X Syndrome Protein Targets Tandem Repeats and Influences Allele-Specific Expression in a Size-Dependent Manner.” *Cell* 143 (3): 367–78. https://doi.org/10.1016/j.cell.2010.09.023.

Lawrence and Singer. 1986. “Intracellular Localization of Messenger RNAs for Cytoskeletal Proteins.” *Cell* 45 (3): 407–15. https://doi.org/10.1016/0092-8674(86)90326-0.

Lawrence, Moyra, Sylvain Daujat, and Robert Schneider. 2016. “Lateral Thinking: How Histone Modifications Regulate Gene Expression.” *Trends in Genetics* 32 (1): 42–56. https://doi.org/10.1016/j.tig.2015.10.007.

Lawrence W. Bergman and W. Michael Kuehl. 1979. “Formation of an Intrachain Disulfide Bond on Nascent Immunoglobulin Light Chains” 254 (18): 8869–76.

Lebedeva, Svetlana, Marvin Jens, Kathrin Theil, Björn Schwanhäusser, Matthias Selbach, Markus Landthaler, and Nikolaus Rajewsky. 2011. “Transcriptome-Wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR.” *Molecular Cell* 43 (3): 340–52. https://doi.org/10.1016/j.molcel.2011.06.008.

Lee, Amy S.Y., Philip J. Kranzusch, Jennifer A. Doudna, and Jamie H.D. Cate. 2016. “EIF3d Is an MRNA Cap-Binding Protein That Is Required for Specialized Translation Initiation.” *Nature* 536 (7614): 96–99. https://doi.org/10.1038/nature18954.

Lee, Jung Shin, Abhijit Shukla, Jessica Schneider, Selene K. Swanson, Michael P. Washburn, Laurence Florens, Sukesh R. Bhaumik, and Ali Shilatifard. 2007. “Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS.” *Cell* 131 (6): 1084–96. https://doi.org/10.1016/j.cell.2007.09.046.

Lee, Sung Haeng, and Roberto Dominguez. 2010. “Regulation of Actin Cytoskeleton Dynamics in Cells.” *Molecules and Cells* 29 (4): 311–25. https://doi.org/10.1007/s10059-010-0053-8.

Lehmann, J M, X K Zhang, G Graupner, M O Lee, T Hermann, B Hoffmann, and M Pfahl. 1993. “Formation of Retinoid X Receptor Homodimers Leads to Repression of T3 Response: Hormonal Cross Talk by Ligand-Induced Squelching.” *Molecular and Cellular Biology* 13 (12): 7698–7707. https://doi.org/10.1128/mcb.13.12.7698.

Leone, Gustavo, Matthew C. Coffey, Ross Gilmore, Roy Duncan, Lloyd Maybaum, and Patrick W.K. Lee. 1996. “C-Terminal Trimerization, but Not N-Terminal Trimerization, of the Reovirus Cell Attachment Protein Is a Posttranslational and Hsp70/ATP-Dependent Process.” *Journal of Biological Chemistry* 271 (14): 8466–71. https://doi.org/10.1074/jbc.271.14.8466.

Leppek, Kathrin, Rhiju Das, and Maria Barna. 2018. “Functional 5′ UTR MRNA Structures in Eukaryotic Translation Regulation and How to Find Them.” *Nature Reviews Molecular Cell Biology* 19 (3): 158–74. https://doi.org/10.1038/nrm.2017.103.

Lessard, Julie A., and Gerald R. Crabtree. 2010. “Chromatin Regulatory Mechanisms in Pluripotency.” *Annual Review of Cell and Developmental Biology* 26 (1): 503–32. https://doi.org/10.1146/annurev-cellbio-051809-102012.

Leurent, Claire, Christine Ruhlmann, Véronique Mallouh, Doris B. Kirschner, Laszlo Tora, Patrick Schultz, Steven Sanders, and P. Anthony Weil. 2002. “Mapping Histone Fold TAFs within Yeast TFIID.” *EMBO Journal* 21 (13): 3424–33. https://doi.org/10.1093/emboj/cdf342.

Levy, Emmanuel D., Elisabetta Boeri Erba, Carol V. Robinson, and Sarah A. Teichmann. 2008. “Assembly Reflects Evolution of Protein Complexes.” *Nature*

453 (7199): 1262–65. https://doi.org/10.1038/nature06942.

Levy, Emmanuel D., Jose B. Pereira-Leal, Cyrus Chothia, and Sarah A. Teichmann. 2006. “3D Complex: A Structural Classification of Protein Complexes.” *PLoS Computational Biology* 2 (11): 1395–1406. https://doi.org/10.1371/journal.pcbi.0020155.

Lewis, P. W., S. J. Elsaesser, K.-M. Noh, S. C. Stadler, and C. D. Allis. 2010. “Daxx Is an H3.3-Specific Histone Chaperone and Cooperates with ATRX in Replication-Independent Chromatin Assembly at Telomeres.” *Proceedings of the National Academy of Sciences* 107 (32): 14075–80. https://doi.org/10.1073/pnas.1008850107.

Li, Bing, Michael Carey, and Jerry L. Workman. 2007. “The Role of Chromatin during Transcription.” *Cell* 128 (4): 707–19. https://doi.org/10.1016/j.cell.2007.01.015.

Li, Pilong, Sudeep Banjade, Hui Chun Cheng, Soyeon Kim, Baoyu Chen, Liang Guo, Marc Llaguno, et al. 2012. “Phase Transitions in the Assembly of Multivalent Signalling Proteins.” *Nature* 483 (7389): 336–40. https://doi.org/10.1038/nature10879.

Li, Shulin, and Miles F Wilkinson. 1998. “Nonsense Surveillance in Lymphocytes ? Minireview Programmed Gene Rearrangements : Good.” *Cell* 8: 135–41.

Li, Yun R., Oliver D. King, James Shorter, and Aaron D. Gitler. 2013. “Stress Granules as Crucibles of ALS Pathogenesis.” *Journal of Cell Biology* 201 (3): 361–72. https://doi.org/10.1083/jcb.201302044.

Lianoglou, S, V Garg, JL Yang, CS Leslie, and Christine Mayr. 2013. “Ubiquitously Transcribed Genes Use Alternative Polyadenylation to Achieve Tissue-Specific Expression.” *Genes*, 2380–96. https://doi.org/10.1101/gad.229328.113.Freely.

Lin, L. 2000. “Cotranslational Dimerization of the Rel Homology Domain of NF-KappaB1 Generates P50-P105 Heterodimers and Is Required for Effective P50 Production.” *The EMBO Journal* 19 (17): 4712–22. https://doi.org/10.1093/emboj/19.17.4712.

Lingala, Shilpa MD, and Marc G MD MHSc Ghany. 2016. “**可**乐定和右美托嗪啶产生抗伤害协同作用 HHS Public Access” 25 (3): 289–313. https://doi.org/110.1016/j.bbi.2017.04.008.

Liu, Dingjiang, Rieko Ishima, Kit I. Tong, Stefan Bagby, Tetsuro Kokubo, D. R. Muhandiram, Lewis E. Kay, Yoshihiro Nakatani, and Mitsuhiko Ikura. 1998. “Solution Structure of a TBP-TAF(II)230 Complex: Protein Mimicry of the Minor Groove Surface of the TATA Box Unwound by TBP.” *Cell* 94 (5): 573–83. https://doi.org/10.1016/S0092-8674(00)81599-8.

Liu, Fang, David K. Jones, Willem J. de Lange, and Gail A. Robertson. 2016. “Cotranslational Association of MRNA Encoding Subunits of Heteromeric Ion Channels.” *Proceedings of the National Academy of Sciences* 113 (17): 4859–64. https://doi.org/10.1073/pnas.1521577113.

Lu, J., J. M. Robinson, D. Edwards, and C. Deutsch. 2001. “T1-T1 Interactions Occur in ER Membranes While Nascent KV Peptides Are Still Attached to Ribosomes.”

Biochemistry 40 (37): 10934–46. https://doi.org/10.1021/bi010763e.

Luo, Yang, Zhenkun Na, and Sarah A. Slavoff. 2018. “P-Bodies: Composition, Properties, and Functions.” *Biochemistry* 57 (17): 2424–31. https://doi.org/10.1021/acs.biochem.7b01162.

Lynch, Michael, Douglas G. Scofield, and Xin Hong. 2005. “The Evolution of Transcription-Initiation Sites.” *Molecular Biology and Evolution* 22 (4): 1137–46. https://doi.org/10.1093/molbev/msi100.

Ma, Weirui, and Christine Mayr. 2018. “A Membraneless Organelle Associated with the Endoplasmic Reticulum Enables 3′UTR-Mediated Protein-Protein Interactions.” *Cell* 175 (6): 1492-1506.e19. https://doi.org/10.1016/j.cell.2018.10.007.

Maharana, Shovamayee, Jie Wang, Dimitrios K Papadopoulos, Doris Richter, Andrey Pozniakovsky, Ina Poser, Marc Bickle, et al. 2018. “Binding Proteins” 921 (March): 918–21.

Mal, Tapas K., Yutaka Masutomi, Le Zheng, Yasuto Nakata, Hiroshi Ohta, Yoshihiro Nakatani, Tetsuro Kokubo, and Mitsuhiko Ikura. 2004. “Structural and Functional Characterization on the Interaction of Yeast TFIID Subunit TAF1 with TATA-Binding Protein.” *Journal of Molecular Biology* 339 (4): 681–93. https://doi.org/10.1016/j.jmb.2004.04.020.

Mangelsdorf, David, and Ronald Evans. 1995. “The RXR Heterodimers and Orphan Receptors.” *Cell* 83: 841–50. https://doi.org/10.1088/0305-4470/23/3/006.

Maquat, Lynne E., and Gordon G. Carmichael. 2001. “Quality Control of MRNA Function.” *Cell* 104 (2): 173–76. https://doi.org/10.1016/S0092-8674(01)00202-1.

Maquat, Lynne E., Woan Yuh Tarn, and Olaf Isken. 2010. “The Pioneer Round of Translation: Features and Functions.” *Cell* 142 (3): 368–74. https://doi.org/10.1016/j.cell.2010.07.022.

Marsh, Joseph A., and Sarah A. Teichmann. 2014a. “Protein Flexibility Facilitates Quaternary Structure Assembly and Evolution.” *PLoS Biology* 12 (5). https://doi.org/10.1371/journal.pbio.1001870.

Marsh, Joseph A., and Sarah A. Teichmann. 2014b. “Structure, Dynamics, Assembly, and Evolution of Protein Complexes.” *Annual Review of Biochemistry* 84 (1): 551–75. https://doi.org/10.1146/annurev-biochem-060614-034142.

Martianov, Igor, Stefano Brancorsini, Anne Gansmuller, Martti Parvinen, Irwin Davidson, and Paolo Sassone-Corsi. 2002. “Distinct Functions of TBP and TLF/TRF2 during Spermatogenesis: Requirement of TLF for Heterochromatic Chromocenter Formation in Haploid Round Spermatids.” *Development (Cambridge, England)* 129 (4): 945–55. http://www.ncbi.nlm.nih.gov/pubmed/11861477.

Martin, Kelsey C., and Anne Ephrussi. 2009. “MRNA Localization: Gene Expression in the Spatial Dimension.” *Cell* 136 (4): 719–30. https://doi.org/10.1016/j.cell.2009.01.044.

Marzluff, William F., Eric J. Wagner, and Robert J. Duronio. 2008. “Metabolism and

Regulation of Canonical Histone MRNAs: Life without a Poly(A) Tail." *Nature Reviews Genetics* 9 (11): 843–54. https://doi.org/10.1038/nrg2438.

Maston, Glenn A., Lihua Julie Zhu, Lynn Chamberlain, Ling Lin, Minggang Fang, and Michael R. Green. 2012. "Non-Canonical TAF Complexes Regulate Active Promoters in Human Embryonic Stem Cells." *ELife* 2012 (1): 1–19. https://doi.org/10.7554/eLife.00068.

Mayr, Christine. 2016. "Evolution and Biological Roles of Alternative 3'UTRs." *Trends in Cell Biology* 26 (3): 227–37. https://doi.org/10.1016/j.tcb.2015.10.012.

Mayr, 2018. "Proteins Assemble as They Are Being Made An." *Nature Research News and Views* 561 (13 September 2018): 6–7.

Mayr, Christine, and David P. Bartel. 2009. "Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells." *Cell* 138 (4): 673–84. https://doi.org/10.1016/j.cell.2009.06.016.

Mayr, Christine, S Lianoglou, V Garg, JL Yang, and CS Leslie. 2013. "F1000Prime Recommendation of: Ubiquitously Transcribed Genes Use Alternative Polyadenylation to Achieve Tissue-Specific Expression." *Genes*, 2380–96. https://doi.org/10.1101/gad.229328.113.Freely.

McDaniel, Stephen L., and Brian D. Strahl. 2017. "Shaping the Cellular Landscape with Set2/SETD2 Methylation." *Cellular and Molecular Life Sciences* 74 (18): 3317–34. https://doi.org/10.1007/s00018-017-2517-x.

McMahon, Steven B., Heather A. Van Buskirk, Kerri A. Dugan, Terry D. Copeland, and Michael D. Cole. 1998. "The Novel ATM-Related Protein TRRAP Is an Essential Cofactor for the c- Myc and E2F Oncoproteins." *Cell* 94 (3): 363–74. https://doi.org/10.1016/S0092-8674(00)81479-8.

Mercer, Tim R., Dagmar Wilhelm, Marcel E. Dinger, Giulia Soldà, Darren J. Korbie, Evgeny A. Glazov, Vy Truong, et al. 2011. "Expression of Distinct RNAs from 3′ Untranslated Regions." *Nucleic Acids Research* 39 (6): 2393–2403. https://doi.org/10.1093/nar/gkq1158.

Merz, Frieder, Daniel Boehringer, Christiane Schaffitzel, Steffen Preissler, Anja Hoffmann, Timm Maier, Anna Rutkowska, et al. 2008. "Molecular Mechanism and Structure of Trigger Factor Bound to the Translating Ribosome." *EMBO Journal* 27 (11): 1622–32. https://doi.org/10.1038/emboj.2008.89.

Metzger, Daniel, Elisabeth Scheer, Aleksey Soldatov, and Làszlò Tora. 1999. "Mammalian TAF(II)30 Is Required for Cell Cycle Progression and Specific Cellular Differentiation Programmes." *EMBO Journal* 18 (17): 4823–34. https://doi.org/10.1093/emboj/18.17.4823.

Meurisse, Julie, Agathe Bacquin, Nicolas Richet, Jean Baptiste Charbonnier, Françoise Ochsenbein, and Anne Peyroche. 2014. "Hug1 Is an Intrinsically Disordered Protein That Inhibits Ribonucleotide Reductase Activity by Directly Binding Rnr2 Subunit." *Nucleic Acids Research* 42 (21): 13174–85. https://doi.org/10.1093/nar/gku1095.

Mignone, Flavio, Carmela Gissi, Sabino Liuni, and Graziano Pesole. 2002. "Untranslated Regions of MRNAs." *Genome Biology* 3 (3): REVIEWS0004.

http://www.ncbi.nlm.nih.gov/pubmed/11897027%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC139023.

Mital, Jeffrey, Markus Meissner, Dominique Soldati, and Gary E Ward. 2005. “Conditional Expression Of.” *Molecular Biology of the Cell* 16 (September): 4341–49. https://doi.org/10.1091/mbc.E05.

Mitrea, Diana M., and Richard W. Kriwacki. 2016. “Phase Separation in Biology; Functional Organization of a Higher Order Short Linear Motifs - The Unexplored Frontier of the Eukaryotic Proteome.” *Cell Communication and Signaling* 14 (1): 1–20. https://doi.org/10.1186/s12964-015-0125-7.

Molliex, Amandine, Jamshid Temirov, Jihun Lee, Maura Coughlin, Anderson P. Kanagaraj, Hong Joo Kim, Tanja Mittag, and J. Paul Taylor. 2015. “Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization.” *Cell* 163 (1): 123–33. https://doi.org/10.1016/j.cell.2015.09.015.

Moriarty, Patrick M., C. Channa Reddy, and Lynne E. Maquat. 1998. “Selenium Deficiency Reduces the Abundance of MRNA for Se-Dependent Glutathione Peroxidase 1 by a UGA-Dependent Mechanism Likely To Be Nonsense Codon-Mediated Decay of Cytoplasmic MRNA.” *Molecular and Cellular Biology* 18 (5): 2932–39. https://doi.org/10.1128/mcb.18.5.2932.

Morimoto. 2008. “Proteotoxic Stress and Inducible Chaperone Networks in Neurodegenerative Disease and Aging.” *Genes and Development* 22 (11): 1427–38. https://doi.org/10.1101/gad.1657108.

Morley, J. F., H. R. Brignull, J. J. Weyers, and R. I. Morimoto. 2002. “The Threshold for Polyglutamine-Expansion Protein Aggregation and Cellular Toxicity Is Dynamic and Influenced by Aging in Caenorhabditis Elegans.” *Proceedings of the National Academy of Sciences* 99 (16): 10417–22. https://doi.org/10.1073/pnas.152161099.

Muchowski, P. J., G. Schaffar, A. Sittler, E. E. Wanker, M. K. Hayer-Hartl, and F. U. Hartl. 2000. “Hsp70 and Hsp40 Chaperones Can Inhibit Self-Assembly of Polyglutamine Proteins into Amyloid-like Fibrils.” *Proceedings of the National Academy of Sciences* 97 (14): 7841–46. https://doi.org/10.1073/pnas.140202897.

Mueller, Florian, Adrien Senecal, Katjana Tantale, Hervé Marie-Nelly, Nathalie Ly, Olivier Collin, Eugenia Basyuk, Edouard Bertrand, Xavier Darzacq, and Christophe Zimmer. 2013. “FISH-Quant: Automatic Counting of Transcripts in 3D FISH Images.” *Nature Methods* 10 (4): 277–78. https://doi.org/10.1038/nmeth.2406.

Mukherjee, Neelanjan, David L. Corcoran, Jeffrey D. Nusbaum, David W. Reid, Stoyan Georgiev, Markus Hafner, Manuel Ascano, Thomas Tuschl, Uwe Ohler, and Jack D. Keene. 2011. “Integrative Regulatory Mapping Indicates That the RNA-Binding Protein HuR Couples Pre-MRNA Processing and MRNA Stability.” *Molecular Cell* 43 (3): 327–39. https://doi.org/10.1016/j.molcel.2011.06.007.

Müller, Ferenc, and Làszlò Tora. 2004. “The Multicoloured World of Promoter Recognition Complexes.” *EMBO Journal* 23 (1): 2–8.

https://doi.org/10.1038/sj.emboj.7600027.

Murakami, Tetsuro, Seema Qamar, Julie Qiaojin Lin, Gabriele S.Kaminski Schierle, Eric Rees, Akinori Miyashita, Ana R. Costa, et al. 2015. “ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function.” *Neuron* 88 (4): 678–90. https://doi.org/10.1016/j.neuron.2015.10.030.

Murray, Dylan T., Masato Kato, Yi Lin, Kent R. Thurber, Ivan Hung, Steven L. McKnight, and Robert Tycko. 2017. “Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains.” *Cell* 171 (3): 615-627.e16. https://doi.org/10.1016/j.cell.2017.08.048.

Nakabayashi, Y., S. Kawashima, T. Enomoto, M. Seki, and M. Horikoshi. 2014. “Roles of Common Subunits within Distinct Multisubunit Complexes.” *Proceedings of the National Academy of Sciences* 111 (2): 699–704. https://doi.org/10.1073/pnas.1316433111.

Nakanishi, Shima, Shin Lee Jung, Kathryn E. Gardner, Jennifer M. Gardner, Yoh Hei Takahashi, Mahesh B. Chandrasekharan, Zu Wen Sun, et al. 2009. “Histone H2BK123 Monoubiquitination Is the Critical Determinant for H3K4 and H3K79 Trimethylation by COMPASS and Dot1.” *Journal of Cell Biology* 186 (3): 371–77. https://doi.org/10.1083/jcb.200906005.

Narlikar, Geeta J., Ramasubramanian Sundaramoorthy, and Tom Owen-Hughes. 2013. “Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes.” *Cell* 154 (3): 490–503. https://doi.org/10.1016/j.cell.2013.07.011.

Naryzhny, Stanislav N., Helen Zhao, and Hoyun Lee. 2005. “Proliferating Cell Nuclear Antigen (PCNA) May Function as a Double Homotrimer Complex in the Mammalian Cell.” *Journal of Biological Chemistry* 280 (14): 13888–94. https://doi.org/10.1074/jbc.M500304200.

Natan, Eviatar, Tamaki Endoh, Liora Haim-Vilmovsky, Tilman Flock, Guilhem Chalancon, Jonathan T.S. Hopper, Bálint Kintses, et al. 2018. “Cotranslational Protein Assembly Imposes Evolutionary Constraints on Homomeric Proteins.” *Nature Structural and Molecular Biology* 25 (3): 279–88. https://doi.org/10.1038/s41594-018-0029-5.

Nguyen, Trung Van, Federico Giannoni, Marie Françoise Dubois, Sook Jae Seo, Marc Vigneron, Claude Kédinger, and Olivier Bensaude. 1996. “In Vivo Degradation of RNA Polymerase II Largest Subunit Triggered by α-Amanitin.” *Nucleic Acids Research* 24 (15): 2924–29. https://doi.org/10.1093/nar/24.15.2924.

Nicholls, Chris D., Kevin G. McLure, Michael A. Shields, and Patrick W.K. Lee. 2002. “Biogenesis of P53 Involves Cotranslational Dimerization of Monomers and Posttranslational Dimerization of Dimers. Implications on the Dominant Negative Effect.” *Journal of Biological Chemistry* 277 (15): 12937–45. https://doi.org/10.1074/jbc.M108815200.

Niednery, Annika, Franziska T. Edelmanny, and Dierk Niessing. 2014. “Of Social Molecules: The Interactive Assembly of ASH1 MRNA-Transport Complexes in Yeast.” *RNA Biology* 11 (8): 998–1009. https://doi.org/10.4161/rna.29946.

Niwa, T., T. Kanamori, T. Ueda, and H. Taguchi. 2012. “Global Analysis of Chaperone Effects Using a Reconstituted Cell-Free Translation System.” *Proceedings of the National Academy of Sciences* 109 (23): 8937–42. https://doi.org/10.1073/pnas.1201380109.

Nooren, M.A. Irene, and M. Janet Thornton. 2003. “Diversity of Protein-Protein Interactions.” *The EMBO Journal* 22 (14): 3486–92. https://doi.org/10.1093/emboj/cdg359.

Nuland, R. van, A. H. Smits, P. Pallaki, P. W. T. C. Jansen, M. Vermeulen, and H. T. M. Timmers. 2013. “Quantitative Dissection and Stoichiometry Determination of the Human SET1/MLL Histone Methyltransferase Complexes.” *Molecular and Cellular Biology* 33 (10): 2067–77. https://doi.org/10.1128/mcb.01742-12.

Ohbayashi, Tetsuya, Toshihiko Kishimoto, Yasutaka Makino, Miho Shimada, Tomoyoshi Nakadai, Tsutomu Aoki, Takefumi Kawata, Shin Ichiro Niwa, and Taka Aki Tamura. 1999. “Isolation of CDNA, Chromosome Mapping, and Expression of the Human TBP-like Protein.” *Biochemical and Biophysical Research Communications* 255 (1): 137–42. https://doi.org/10.1006/bbrc.1999.0159.

Onodera, Yasuyuki, Jeremy R. Haag, Thomas Ream, Pedro Costa Nunes, Olga Pontes, and Craig S. Pikaard. 2005. “Plant Nuclear RNA Polymerase IV Mediates SiRNA and DNA Methylation- Dependent Heterochromatin Formation.” *Cell* 120 (5): 613–22. https://doi.org/10.1016/j.cell.2005.02.007.

Orphanides, George, Thierry Lagrange, and Danny Reinberg. 1996. “The General Transcription Machinery of RNA Polymerase II.” *Genes & Dev.* 10 (7): 2657–83. http://genesdev.cshlp.org/content/10/21/2657.full.pdf.

Orphanides, George, and Danny Reinberg. 2016. “The General Transcription Factors of RNA Polymerase II The General Transcription Factors of RNA Polymerase II,” no. December 1996: 1479–90. https://doi.org/10.1101/gad.10.21.2657.

Ossipow, Vincent, Jean Pierre Tassan, Erich A. Nigg, and Ueli Schibler. 1995. “A Mammalian RNA Polymerase II Holoenzyme Containing All Components Required for Promoter-Specific Transcription Initiation.” *Cell* 83 (1): 137–46. https://doi.org/10.1016/0092-8674(95)90242-2.

Otto, H., C. Conz, P. Maier, T. Wolfle, C. K. Suzuki, P. Jeno, P. Rucknagel, J. Stahl, and S. Rospert. 2005. “The Chaperones MPP11 and Hsp70L1 Form the Mammalian Ribosome-Associated Complex.” *Proceedings of the National Academy of Sciences* 102 (29): 10064–69. https://doi.org/10.1073/pnas.0504400102.

Ozer, Josef, Arthur H. Bolden, Arianna Lee, Paul M. Lieberman, Paul A. Moore, and Craig A. Rosen. 1994. “Molecular Cloning of the Small (γ) Subunit of Human TFIIA Reveals Functions Critical for Activated Transcription.” *Genes and Development* 8 (19): 2324–35. https://doi.org/10.1101/gad.8.19.2324.

Ozsolak, Fatih, Philipp Kapranov, Sylvain Foissac, Sang Woo Kim, Elane Fishilevich, A. Paula Monaghan, Bino John, and Patrice M. Milos. 2010. “Comprehensive Polyadenylation Site Maps in Yeast and Human Reveal Pervasive Alternative

Polyadenylation." *Cell* 143 (6): 1018–29. https://doi.org/10.1016/j.cell.2010.11.020.

P. Anthony Weil, Jacqueline Segall, Barry Harris, Sun-Yu Ng, II and Robert G. Roeder. 1979. "Faithful Transcription of Eukaryotic Genes by RNA Polymerase III in Systems Reconstituted with Purified DNA Templates." *Journal of Biological Chemistry* 254 (13): 6163–73.

Panasenko, Olesya O., Syam Prakash Somasekharan, Zoltan Villanyi, Marina Zagatti, Fedor Bezrukov, Ravish Rashpa, Julien Cornut, et al. 2019. "Co-Translational Assembly of Proteasome Subunits in NOT1-Containing Assemblysomes." *Nature Structural and Molecular Biology* 26 (2): 110–20. https://doi.org/10.1038/s41594-018-0179-5.

Pandit, Shashi Bhushan, and Jeffrey Skolnick. 2008. "Fr-TM-Align: A New Protein Structural Alignment Method Based on Fragment Alignments and the TM-Score." *BMC Bioinformatics* 9: 1–11. https://doi.org/10.1186/1471-2105-9-531.

Parker, Rex A, Oliver P Flint, Ruth Mulvey, Carolina Elosua, Faye Wang, William Fenderson, Shulin Wang, Wen-pin Yang, and Mustafa A Noor. 2005. "Endoplasmic Reticulum Stress Links Dyslipidemia to Inhibition of Proteasome Activity and Glucose Transport by HIV Protease Inhibitors." *Molecular Pharmacology* 67 (6): 1909–19. https://doi.org/10.1124/mol.104.010165.cellular.

Parker, Roy, Muriel Brengues, and Daniela Teixeira. 2005. "Movement of Eukaryotic MRNAs Between Polysomes and Cytoplasmic Processing Bodies." *Science* 310 (5747): 486–89. http://www.sciencemag.org/cgi/content/abstract/310/5747/486.

Parvin, Jeffrey D., and Phillip A. Sharp. 1993. "DNA Topology and a Minimal Set of Basal Factors for Transcription by RNA Polymerase II." *Cell* 73 (3): 533–40. https://doi.org/10.1016/0092-8674(93)90140-L.

Parvin, Jeffrey D., Benjamin M. Shykind, Rachel E. Meyers, Jaesang Kim, and Phillip A. Sharp. 1994. "Multiple Sets of Basal Factors Initiate Transcription by RNA Polymerase II." *Journal of Biological Chemistry* 269 (28): 18414–21.

Passarge, E. 1979. "Emil Heitz and the Concept of Heterochromatin: Longitudinal Chromosome Differentiation Was Recognized Fifty Years Ago." *American Journal of Human Genetics* 31 (2): 106–15. http://www.ncbi.nlm.nih.gov/pubmed/377956%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1685768.

Patel, Avinash B., Robert K. Louder, Basil J. Greber, Sebastian Grünberg, Jie Luo, Jie Fang, Yutong Liu, Jeff Ranish, Steve Hahn, and Eva Nogales. 2018. "Structure of Human TFIID and Mechanism of TBP Loading onto Promoter DNA." *Science* 362 (6421). https://doi.org/10.1126/science.aau8872.

Pavri, Rushad, Bing Zhu, Guohong Li, Patrick Trojer, Subhrangsu Mandal, Ali Shilatifard, and Danny Reinberg. 2006. "Histone H2B Monoubiquitination Functions Cooperatively with FACT to Regulate Elongation by RNA Polymerase II." *Cell* 125 (4): 703–17. https://doi.org/10.1016/j.cell.2006.04.029.

Pelletier, Jerry, and Nahum Sonenberg. 1985. "Insertion Mutagenesis to Increase Secondary Structure within the 5′ Noncoding Region of a Eukaryotic MRNA Reduces Translational Efficiency." *Cell* 40 (3): 515–26.

https://doi.org/10.1016/0092-8674(85)90200-4.

Penuel, Elicia, Robert W. Akita, and Mark X. Sliwkowski. 2002. “Identification of a Region within the ErbB2/HER2 Intracellular Domain That Is Necessary for Ligand-Independent Association.” *Journal of Biological Chemistry* 277 (32): 28468–73. https://doi.org/10.1074/jbc.M202510200.

Perica, Tina, Joseph A. Marsh, Filipa L. Sousa, Eviatar Natan, Lucy J. Colwell, Sebastian E. Ahnert, and Sarah A. Teichmann. 2012. “The Emergence of Protein Complexes: Quaternary Structure, Dynamics and Allostery.” *Biochemical Society Transactions* 40 (3): 475–91. https://doi.org/10.1042/bst20120056.

Pesole, Graziano, Flavio Mignone, Carmela Gissi, Giorgio Grillo, Flavio Licciulli, and Sabino Liuni. 2001. “Structural and Functional Features of Eukaryotic MRNA Untranslated Regions.” *Gene* 276 (1–2): 73–81. https://doi.org/10.1016/S0378-1119(01)00674-6.

Petesch, Steven J., and John T. Lis. 2012. “Overcoming the Nucleosome Barrier during Transcript Elongation.” *Trends in Genetics* 28 (6): 285–94. https://doi.org/10.1016/j.tig.2012.02.005.

Phartiyal, Pallavi, Eugenia M C Jones, and Gail A. Robertson. 2007. “Heteromeric Assembly of Human Ether-à-Go-Go-Related Gene (HERG) 1a/1b Channels Occurs Cotranslationally via N-Terminal Interactions.” *Journal of Biological Chemistry* 282 (13): 9874–82. https://doi.org/10.1074/jbc.M610875200.

Pianka, E R, K J Gaston, H Hillebrand, M Cardillo, M Cardillo, C D L Orme, I P F Owens, et al. 2007. “No Title,” no. March: 1576–80.

Pinarbasi, Emile S., Andrey L. Karamyshev, Elena B. Tikhonova, I. Hui Wu, Henry Hudson, and Philip J. Thomas. 2018. “Pathogenic Signal Sequence Mutations in Progranulin Disrupt SRP Interactions Required for MRNA Stability.” *Cell Reports* 23 (10): 2844–51. https://doi.org/10.1016/j.celrep.2018.05.003.

Pointud, J.-C. 2003. “The Intracellular Localisation of TAF7L, a Paralogue of Transcription Factor TFIID Subunit TAF7, Is Developmentally Regulated during Male Germ-Cell Differentiation.” *Journal of Cell Science* 116 (9): 1847–58. https://doi.org/10.1242/jcs.00391.

Polier, Sigrun, Zdravko Dragovic, F. Ulrich Hartl, and Andreas Bracher. 2008. “Structural Basis for the Cooperation of Hsp70 and Hsp110 Chaperones in Protein Folding.” *Cell* 133 (6): 1068–79. https://doi.org/10.1016/j.cell.2008.05.022.

Polymenidou, Magdalini. 2018. “The RNA Face of Phase Separation.” *Science* 360 (6391): 859–60. https://doi.org/10.1126/science.aat8028.

Popp, Maximilian Wei-Lin, and Lynne E. Maquat. 2013. “Organizing Principles of Mammalian Nonsense-Mediated MRNA Decay.” *Annual Review of Genetics* 47 (1): 139–65. https://doi.org/10.1146/annurev-genet-111212-133424.

Porrua, Odil, and Domenico Libri. 2015. “Transcription Termination and the Control of the Transcriptome: Why, Where and How to Stop.” *Nature Reviews Molecular Cell Biology* 16 (3): 190–202. https://doi.org/10.1038/nrm3943.

Pot, The. 2008. “(Kb) 9.42 4.36” 320 (June): 1344–50.

Preissler, Steffen, and Elke Deuerling. 2012. “Ribosome-Associated Chaperones as Key Players in Proteostasis.” *Trends in Biochemical Sciences* 37 (7): 274–83. https://doi.org/10.1016/j.tibs.2012.03.002.

Protter, David S.W., and Roy Parker. 2016. “Principles and Properties of Stress Granules.” *Trends in Cell Biology* 26 (9): 668–79. https://doi.org/10.1016/j.tcb.2016.05.004.

Pruneski, Justin A., Sarah J. Hainer, Kostadin O. Petrov, and Joseph A. Martens. 2011. “The Paf1 Complex Represses SER3 Transcription in Saccharomyces Cerevisiae by Facilitating Intergenic Transcription-Dependent Nucleosome Occupancy of the SER3 Promoter ∇.” *Eukaryotic Cell* 10 (10): 1283–94. https://doi.org/10.1128/EC.05141-11.

Pu, Jing, Christina Schindler, Rui Jia, Michal Jarnik, Peter Backlund, and Juan S. Bonifacino. 2015. “BORC, a Multisubunit Complex That Regulates Lysosome Positioning.” *Developmental Cell* 33 (2): 176–88. https://doi.org/10.1016/j.devcel.2015.02.011.

Pueschel, Ringo, Francesca Coraggio, and Peter Meister. 2016. “From Single Genes to Entire Genomes: The Search for a Function of Nuclear Organization.” *Development* 143 (6): 910–23. https://doi.org/10.1242/dev.129007.

Rada-Iglesias, Alvaro, Ruchi Bajpai, Sara Prescott, Samantha A. Brugmann, Tomek Swigut, and Joanna Wysocka. 2012. “Epigenomic Annotation of Enhancers Predicts Transcriptional Regulators of Human Neural Crest.” *Cell Stem Cell* 11 (5): 633–48. https://doi.org/10.1016/j.stem.2012.07.006.

Rattenbacher, B., D. Beisang, D. L. Wiesner, J. C. Jeschke, M. von Hohenberg, I. A. St. Louis-Vlasova, and P. R. Bohjanen. 2010. “Analysis of CUGBP1 Targets Identifies GU-Repeat Sequences That Mediate Rapid MRNA Decay.” *Molecular and Cellular Biology* 30 (16): 3970–80. https://doi.org/10.1128/mcb.00624-10.

Raue, Uta, Stefan Oellerer, and Sabine Rospert. 2007. “Association of Protein Biogenesis Factors at the Yeast Ribosomal Tunnel Exit Is Affected by the Translational Status and Nascent Polypeptide Sequence.” *Journal of Biological Chemistry* 282 (11): 7809–16. https://doi.org/10.1074/jbc.M611436200.

Redick et al. 1995. “Rapid Intracellular Assembly of Tenascin Hexabrachions Suggests a Novel Co-Translational Process.” *Journal of Cell Science* 108 (4): 1761–69.

Reinberg, Danny, Robert J. Sims III, and Rimma Belotserkovskaya. 2004. “Elongation by RNA Polymerase II: The Short and Long of It.” *Genes and Development* 18 (20): 2437–68. https://doi.org/10.1101/gad.1235904.

Resto, A. Vicente, L. Otavia Caballero, R. Martin Buta, H. William Westra, LI Wu, M. Joanne Westendorf, Jen Jin, Philip Hieter, and David Sidransky. 2000. “A Putative Oncogenic Role for MPP11 in Head and Neck Squamous Cell Cancer.” *Cancer Research* 60 (19): 5529–35.

Reynaud, E. 2010. “Molecular Chaperones in Cellular Protein Folding.” *Nature Education* 3 (9): 571–80.

Richmond, Robin K, David F Sargent, Timothy J Richmond, Karolin Luger, and Armin W Ma. 1997. “Crystal Structure of the Nucleosome ˚ Resolution Core Particle at 2 . 8 A” 7: 251–60.

Richmond, Timothy J, and Curt A Davey. 2003. “The Structure of DNA in the Nucleosome Core,” 145–50.

Robinson, John M., and Carol Deutsch. 2005. “Coupled Tertiary Folding and Oligomerization of the T1 Domain of Kv Channels.” *Neuron* 45 (2): 223–32. https://doi.org/10.1016/j.neuron.2004.12.043.

Roeder, R. G., and W. J. Rutter. 1970. “Specific Nucleolar and Nucleoplasmic RNA Polymerases.” *Proceedings of the National Academy of Sciences* 65 (3): 675–82. https://doi.org/10.1073/pnas.65.3.675.

Roeder, Robert G. 1996. “The Role of General Initiation Factors in Transcription by RNA Polymerase II.” *Trends in Biochemical Sciences* 21 (9): 327–35. https://doi.org/10.1016/0968-0004(96)10050-5.

Rogalla, Thorsten, Gudrun Lutsch, Martin Wieske, Matthias Gaestel, Monika Ehrnsperger, Johannes Buchner, Xavier Preville, et al. 1999. “Regulation of Hsp27 Oligomerization, Chaperone Function, and Protective Activity against Oxidative Stress/Tumor Necrosis Factor by Phosphorylation.” *Journal of Biological Chemistry* 274 (27): 18947–56. https://doi.org/10.1074/jbc.274.27.18947.

Rohs, Remo, Xiangshu Jin, Sean M. West, Rohit Joshi, Barry Honig, and Richard S. Mann. 2010. “Origins of Specificity in Protein-DNA Recognition.” *Annual Review of Biochemistry* 79 (1): 233–69. https://doi.org/10.1146/annurev-biochem-060408-091030.

Rui Kamadaa, Yu Toguchia, Takao Nomurab, Toshiaki Imagawaa, Kazuyasu Sakaguchia. 2007. “Tetramer Formation of Tumor Suppressor Protein P53: Structure, Function, and Applications.” *Biopolymers* 89 (1): 40–51. https://doi.org/10.1002/bip.

Ruotolo, Brandon T., Justin L.P. Benesch, Alan M. Sandercock, Suk Joon Hyung, and Carol V. Robinson. 2008. “Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes.” *Nature Protocols* 3 (7): 1139–52. https://doi.org/10.1038/nprot.2008.78.

Ruppert, Siegfried, Edith H. Wang, and Robert Tjian. 1993. “Cloning and Expression of Human TAFII250: A TBP-Associated Factor Implicated in Cell-Cycle Regulation.” *Nature* 362 (6416): 175–79. https://doi.org/10.1038/362175a0.

Saha, Anjanabha, Jacqueline Wittmeyer, and Bradley R. Cairns. 2006. “Chromatin Remodelling: The Industrial Revolution of DNA around Histones.” *Nature Reviews Molecular Cell Biology* 7 (6): 437–47. https://doi.org/10.1038/nrm1945.

Sahlan, Muhamad, Tamotsu Zako, and Masafumi Yohda. 2018. “Prefoldin, a Jellyfish-like Molecular Chaperone: Functional Cooperation with a Group II Chaperonin and Beyond.” *Biophysical Reviews* 10 (2): 339–45. https://doi.org/10.1007/s12551-018-0400-0.

Sainsbury, Sarah, Carrie Bernecky, and Patrick Cramer. 2015. “Structural Basis of

Transcription Initiation by RNA Polymerase II." *Nature Reviews Molecular Cell Biology* 16 (3): 129–43. https://doi.org/10.1038/nrm3952.

Sandaltzopoulos, Raphael, and Peter B. Becker. 1998. "Heat Shock Factor Increases the Reinitiation Rate from Potentiated Chromatin Templates." *Molecular and Cellular Biology* 18 (1): 361–67. https://doi.org/10.1128/mcb.18.1.361.

Sanders, S. L., K. A. Garbett, and P. A. Weil. 2002. "Molecular Characterization of Saccharomyces Cerevisiae TFIID." *Molecular and Cellular Biology* 22 (16): 6000–6013. https://doi.org/10.1128/mcb.22.16.6000-6013.2002.

Saunders, Abbie, Leighton J. Core, and John T. Lis. 2006. "Breaking Barriers to Transcription Elongation." *Nature Reviews Molecular Cell Biology* 7 (8): 557–67. https://doi.org/10.1038/nrm1981.

Scheer, Elisabeth, Frédéric Delbac, Laszlo Tora, Dino Moras, and Christophe Romier. 2012. "TFIID TAF6-TAF9 Complex Formation Involves the HEAT Repeat-Containing C-Terminal Domain of TAF6 and Is Modulated by TAF5 Protein." *Journal of Biological Chemistry* 287 (33): 27580–92. https://doi.org/10.1074/jbc.M112.379206.

Schneider, Maren, Doris Hellerschmied, Tobias Schubert, Stefan Amlacher, Vinesh Vinayachandran, Rohit Reja, B. Franklin Pugh, Tim Clausen, and Alwin Köhler. 2015. "The Nuclear Pore-Associated TREX-2 Complex Employs Mediator to Regulate Gene Expression." *Cell* 162 (5): 1016–28. https://doi.org/10.1016/j.cell.2015.07.059.

Schuermann, Jonathan P., Jianwen Jiang, Jorge Cuellar, Oscar Llorca, Liping Wang, Luis E. Gimenez, Suping Jin, et al. 2008. "Structure of the Hsp110:Hsc70 Nucleotide Exchange Machine." *Molecular Cell* 31 (2): 232–43. https://doi.org/10.1016/j.molcel.2008.05.006.

Sexton, Tom, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. 2012. "Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome." *Cell* 148 (3): 458–72. https://doi.org/10.1016/j.cell.2012.01.010.

Sharma, Urvashi, Loïc Carrique, Sandrine Vadon-Le Goff, Natacha Mariano, Rainier Numa Georges, Frederic Delolme, Peppi Koivunen, et al. 2017. "Structural Basis of Homo- and Heterotrimerization of Collagen I." *Nature Communications* 8. https://doi.org/10.1038/ncomms14671.

Shen, Kuang, and Shu-ou Shan. 2010. "Transient Tether between the SRP RNA and SRP Receptor Ensures Efficient Cargo Delivery during Cotranslational Protein Targeting." *Proceedings of the National Academy of Sciences* 107 (17): 7698–7703. https://doi.org/10.1073/pnas.1002968107.

Sherwood, Richard I, Tatsunori Hashimoto, Charles W O Donnell, Sophia Lewis, Amira A Barkal, John Peter Van Hoff, Vivek Karun, Tommi Jaakkola, and K David. 2014. "Discovery of Non-Directional and Directional Pioneer Transcription Factors by Modeling DNase Profile Magnitude and Shape" 32 (2): 171–78. https://doi.org/10.1038/nbt.2798.Discovery.

Shi, Y., K. J. Hertel, E.-A. Choi, P. J. Shepard, J. Lu, and L. A. Flanagan. 2011. “Complex and Dynamic Landscape of RNA Polyadenylation Revealed by PAS-Seq.” *Rna* 17 (4): 761–72. https://doi.org/10.1261/rna.2581711.

Shiber, Ayala, Kristina Döring, Ulrike Friedrich, Kevin Klann, Dorina Merker, Mostafa Zedan, Frank Tippmann, Günter Kramer, and Bernd Bukau. 2018. “Cotranslational Assembly of Protein Complexes in Eukaryotes Revealed by Ribosome Profiling.” *Nature* 561 (7722): 268–72. https://doi.org/10.1038/s41586-018-0462-y.

Shieh, Yu-Wei, Josef J. Auburger, D. Lys Guilbride, Günter Kramer, Bernd Bukau, Pablo Minguez, and Peer Bork. 2015. “Operon Structure and Cotranslational Subunit Association Direct Protein Assembly in Bacteria.” *Science* 350 (6261): 678–80. https://doi.org/10.1126/science.aac8171.

Shoemaker, Christopher J., and Rachel Green. 2012. “Translation Drives MRNA Quality Control.” *Nature Structural and Molecular Biology* 19 (6): 594–601. https://doi.org/10.1038/nsmb.2301.

Shoji, Wataru, Toshiaki Inoue, Tohru Yamamoto, and Masuo Obinata. 1995. “MIDA1, a Protein Associated with Id, Regulates Cell Growth.” *Journal of Biological Chemistry* 270 (42): 24818–25. https://doi.org/10.1074/jbc.270.42.24818.

Siepel, Adam, Gill Bejerano, Jakob S. Pedersen, Angie S. Hinrichs, Minmei Hou, Kate Rosenbloom, Hiram Clawson, et al. 2005. “Evolutionarily Conserved Elements in Vertebrate, Insect, Worm, and Yeast Genomes.” *Genome Research* 15 (8): 1034–50. https://doi.org/10.1101/gr.3715005.

Simic, Rajna, Derek L Lindstrom, Hien G Tran, Kelli L Roinick, Patrick J Costa, Alexander D Johnson, Grant A Hartzog, and Karen M Arndt. 2003. “Simic 2003” 22 (8).

Simms, Carrie L., Erica N. Thomas, and Hani S. Zaher. 2017. “Ribosome-Based Quality Control of MRNA and Nascent Peptides.” *Wiley Interdisciplinary Reviews: RNA* 8 (1): 1–27. https://doi.org/10.1002/wrna.1366.

Simpson, Robert T. 1978. “Structure of the Chromatosome, a Chromatin Particle Containing 160 Base Pairs of DNA and All the Histones.” *Biochemistry* 17 (25): 5524–31. https://doi.org/10.1021/bi00618a030.

Singer, Robert H. 1992. “The Cytoskeleton and MRNA Localization.” *Current Opinion in Cell Biology* 4 (1): 15–19. https://doi.org/10.1016/0955-0674(92)90053-F.

Singh, Irtisha, Shih Han Lee, Adam S. Sperling, Mehmet K. Samur, Yu Tzu Tai, Mariateresa Fulciniti, Nikhil C. Munshi, Christine Mayr, and Christina S. Leslie. 2018. “Widespread Intronic Polyadenylation Diversifies Immune Cell Transcriptomes.” *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-04112-z.

Slattery, Matthew, Tianyin Zhou, Lin Yang, Ana Carolina Dantas Machado, Raluca Gordân, and Remo Rohs. 2014. “Absence of a Simple Code: How Transcription Factors Read the Genome.” *Trends in Biochemical Sciences* 39 (9): 381–99. https://doi.org/10.1016/j.tibs.2014.07.002.

Smardon, Anne M., Maureen Tarsio, and Patricia M. Kane. 2002. “The RAVE Complex Is Essential for Stable Assembly of the Yeast V-ATpase.” *Journal of Biological Chemistry* 277 (16): 13831–39. https://doi.org/10.1074/jbc.M200682200.

Smith, E L, G Zubay, H R Wilson, V Luzzati, J F Pardon, M H F Wilkins, K V Shooter, et al. n.d. “Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin Structure Is Based on a Repeating Unit of Eight” 184.

Smith, Edwin, and Ali Shilatifard. 2010. “The Chromatin Signaling Pathway: Diverse Mechanisms of Recruitment of Histone-Modifying Enzymes and Varied Biological Outcomes.” *Molecular Cell* 40 (5): 689–701. https://doi.org/10.1016/j.molcel.2010.11.031.

Song, Jingwen, Jean-Pierre Perreault, Ivan Topisirovic, and Stéphane Richard. 2016. “RNA G-Quadruplexes and Their Potential Regulatory Roles in Translation.” *Translation* 4 (2): e1244031. https://doi.org/10.1080/21690731.2016.1244031.

Soutoglou, E., M. A. Demeny, E. Scheer, G. Fienga, P. Sassone-Corsi, and L. Tora. 2005. “The Nuclear Import of TAF10 Is Regulated by One of Its Three Histone Fold Domain-Containing Interaction Partners.” *Molecular and Cellular Biology* 25 (10): 4092–4104. https://doi.org/10.1128/mcb.25.10.4092-4104.2005.

Spedale G, Timmers H.T.M., and Pijnappel W.W.M.P. 2012. “ATAC-King the Complexity of SAGA during Evolution.” *Genes and Development* 26 (6): 527–41. https://doi.org/10.1101/gad.184705.111.DNA-mediated.

Spriggs, Keith A., Mark Stoneley, Martin Bushell, and Anne E. Willis. 2008. “Re-Programming of Translation Following Cell Stress Allows IRES-Mediated Translation to Predominate.” *Biology of the Cell* 100 (1): 27–38. https://doi.org/10.1042/bc20070098.

Stella, Stefano, Duilio Cascio, and Reid C. Johnson. 2010. “The Shape of the DNA Minor Groove Directs Binding by the DNA-Bending Protein Fis.” *Genes and Development* 24 (8): 814–26. https://doi.org/10.1101/gad.1900610.

Sun, X., X. Li, P. M. Moriarty, T. Henics, J. P. LaDuca, and L. E. Maquat. 2001. “Nonsense-Mediated Decay of MRNA for the Selenoprotein Phospholipid Hydroperoxide Glutathione Peroxidase Is Detectable in Cultured Cells but Masked or Inhibited in Rat Tissues.” *Molecular Biology of the Cell* 12 (4): 1009–17. https://doi.org/10.1091/mbc.12.4.1009.

Sun, Xiaoqing, Dongmin Ma, Michael Sheldon, Kam Yeung, and Danny Reinberg. 1994. “Reconstitution of Human TFIIA Activity from Recombinant Polypeptides: A Role in TFIID-Mediated Transcription.” *Genes and Development* 8 (19): 2336–48. https://doi.org/10.1101/gad.8.19.2336.

SUZUKI, Yoshiyuki. 2014. “Emerging Novel Concept of Chaperone Therapies for Protein Misfolding Diseases.” *Proceedings of the Japan Academy, Series B* 90 (5): 145–62. https://doi.org/10.2183/pjab.90.145.

Swapna, Lakshmipuram Seshadri, Kuchi Srikeerthana, and Narayanaswamy Srinivasan. 2012. “Extent of Structural Asymmetry in Homodimeric Proteins: Prevalence and Relevance.” *PLoS ONE* 7 (5).

https://doi.org/10.1371/journal.pone.0036688.

Takada, Shinako, John T Lis, Sharleen Zhou, and Robert Tjian. 2000. “A TRF1:BRF Complex Directs Drosophila RNA Polymerase III Transcription Universal ‘Basal Transcriptional Apparatus.’ In Fact, Mul-Tiple Cell Type-Selective TBP-Associated Factors (TAFs) That Make up Variant TFIID Complexes Have Been Discov.” *Cell* 101: 459–69.

Takahashi, S., Y. Ohya, Y. Araki, T. Katada, T. Sakuno, S.-I. Hoshino, H. Nishina, and K. Kontani. 2008. “Upf1 Potentially Serves as a RING-Related E3 Ubiquitin Ligase via Its Association with Upf3 in Yeast.” *Rna* 14 (9): 1950–58. https://doi.org/10.1261/rna.536308.

Talbert, Paul B., and Steven Henikoff. 2017. “Histone Variants on the Move: Substrates for Chromatin Dynamics.” *Nature Reviews Molecular Cell Biology* 18 (2): 115–26. https://doi.org/10.1038/nrm.2016.148.

Tallec, Benoît Le, Marie Bénédicte Barrault, Régis Courbeyrette, Raphaël Guérois, Marie Claude Marsolier-Kergoat, and Anne Peyroche. 2007. “20S Proteasome Assembly Is Orchestrated by Two Distinct Pairs of Chaperones in Yeast and in Mammals.” *Molecular Cell* 27 (4): 660–74. https://doi.org/10.1016/j.molcel.2007.06.025.

Tang, Yun Chi, Hung Chun Chang, Manajit Hayer-Hartl, and F. Ulrich Hartl. 2007. “SnapShot: Molecular Chaperones, Part II.” *Cell* 128 (2): 412.e1-412.e2. https://doi.org/10.1016/j.cell.2007.01.013.

Tebo, Julie, Sandy Der, Mathias Frevel, Khalid S.A. Khabar, Bryan R.G. Williams, and Thomas A. Hamilton. 2003. “Heterogeneity in Control of MRNA Stability by AU-Rich Elements.” *Journal of Biological Chemistry* 278 (14): 12085–93. https://doi.org/10.1074/jbc.M212992200.

Techtmann, S. M., and F. T. Robb. 2010. “Archaeal-like Chaperonins in Bacteria.” *Proceedings of the National Academy of Sciences* 107 (47): 20269–74. https://doi.org/10.1073/pnas.1004783107.

Tee, W.-W., and D. Reinberg. 2014. “Chromatin Features and the Epigenetic Regulation of Pluripotency States in ESCs.” *Development* 141 (12): 2376–90. https://doi.org/10.1242/dev.096982.

Teichmann, Martin, Zhengxin Wang, Ernest Martinez, Agneta Tjernberg, Di Zhang, Frank Vollmer, Brian T Chait, and Robert G Roeder. 1999. “Stably Associates with HTFIIA in HeLa Cells” 2: 2–7.

Teves, Sheila S., Christopher M. Weber, and Steven Henikoff. 2014. “Transcribing through the Nucleosome.” *Trends in Biochemical Sciences* 39 (12): 577–86. https://doi.org/10.1016/j.tibs.2014.10.004.

Thomas, Mary C., and Cheng Ming Chiang. 2006. “The General Transcription Machinery and General Cofactors.” *Critical Reviews in Biochemistry and Molecular Biology* 41 (3): 105–78. https://doi.org/10.1080/10409230600648736.

Thulasiraman, Vanitha, Chin Fen Yang, and Judith Frydman. 1999. “In Vivo Newly Translated Polypeptides Are Sequestered in a Protected Folding Environment.” *EMBO Journal* 18 (1): 85–95. https://doi.org/10.1093/emboj/18.1.85.

Tian Bin and James L. Manley. 2008. “Alternative cleavage and polyadenylation: the long and short of it.” *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2013.03.005

Tian, Bin, Jun Hu, Haibo Zhang, and Carol S. Lutz. 2005. “A Large-Scale Analysis of MRNA Polyadenylation of Human and Mouse Genes.” *Nucleic Acids Research* 33 (1): 201–12. https://doi.org/10.1093/nar/gki158.

Timmers, H.T. 1994. “Transcription Initiation by RNA Polymerase II Does Not Require Hydrolysis of the Beta-Gamma Phosphoanhydride Bond of ATP.” *The EMBO Journal* 13 (2): 391–99. https://doi.org/10.1002/j.1460-2075.1994.tb06273.x.

Tipping, Kevin W., Patricija van Oosten-Hawle, Eric W. Hewitt, and Sheena E. Radford. 2015. “Amyloid Fibres: Inert End-Stage Aggregates or Key Players in Disease?” *Trends in Biochemical Sciences* 40 (12): 719–27. https://doi.org/10.1016/j.tibs.2015.10.002.

Tora, László. 2002. “A Unified Nomenclature for TATA Box Binding Protein (TBP)-Associated Factors (TAFs) Involved in RNA Polymerase II Transcription.” *Genes and Development*, no. 4: 19–23. https://doi.org/10.1101/gad.976402.GENES.

Trowitzsch, Simon, Cristina Viola, Elisabeth Scheer, Sascha Conic, Virginie Chavant, Marjorie Fournier, Gabor Papai, et al. 2015. “Cytoplasmic TAF2-TAF8-TAF10 Complex Provides Evidence for Nuclear Holo-TFIID Assembly from Preformed Submodules.” *Nature Communications* 6: 1–14. https://doi.org/10.1038/ncomms7011.

Tsanov, Nikolay, Aubin Samacoits, Racha Chouaib, Abdel Meneem Traboulsi, Thierry Gostan, Christian Weber, Christophe Zimmer, et al. 2016. “SmiFISH and FISH-Quant - A Flexible Single RNA Detection Approach with Super-Resolution Capability.” *Nucleic Acids Research* 44 (22). https://doi.org/10.1093/nar/gkw784.

Tsuboi, Tatsuhisa, Kazushige Kuroha, Kazuhei Kudo, Shiho Makino, Eri Inoue, Isao Kashima, and Toshifumi Inada. 2012. “Dom34: Hbs1 Plays a General Role in Quality-Control Systems by Dissociation of a Stalled Ribosome at the 3’ End of Aberrant MRNA.” *Molecular Cell* 46 (4): 518–29. https://doi.org/10.1016/j.molcel.2012.03.013.

Tu, Liwei, Jing Wang, Andrew Helm, William R. Skach, and Carol Deutsch. 2000. “Transmembrane Biogenesis of Kv1.3.” *Biochemistry* 39 (4): 824–36. https://doi.org/10.1021/bi991740r.

Tutucci, Evelina, and Françoise Stutz. 2011. “Keeping MRNPs in Check during Assembly and Nuclear Export.” *Nature Reviews Molecular Cell Biology* 12 (6): 377–84. https://doi.org/10.1038/nrm3119.

Ulitsky, Igor, Alena Shkumatava, Calvin H. Jan, Alexander O. Subtelny, David Koppstein, George W. Bell, Hazel Sive, and David P. Bartel. 2012. “Extensive Alternative Polyadenylation during Zebrafish Development.” *Genome Research* 22 (10): 2054–66. https://doi.org/10.1101/gr.139733.112.

Ullers, Ronald S., Edith N.G. Houben, Amanda Raine, Corinne M. Ten Hagen-Jongman, Måns Ehrenberg, Joseph Brunner, Bauke Oudega, Nellie Harms, and Joen Luirink. 2003. “Interplay of Signal Recognition Particle and Trigger Factor

at L23 near the Nascent Chain Exit Site on the Escherichia Coli Ribosome." *Journal of Cell Biology* 161 (4): 679–84. https://doi.org/10.1083/jcb.200302130.

Umlauf, D., J. Bonnet, F. Waharte, M. Fournier, M. Stierle, B. Fischer, L. Brino, D. Devys, and L. Tora. 2013. "The Human TREX-2 Complex Is Stably Associated with the Nuclear Pore Basket." *Journal of Cell Science* 126 (12): 2656–67. https://doi.org/10.1242/jcs.118000.

Uversky, Vladimir N. 2017. "Intrinsically Disordered Proteins in Overcrowded Milieu: Membrane-Less Organelles, Phase Separation, and Intrinsic Disorder." *Current Opinion in Structural Biology* 44: 18–30. https://doi.org/10.1016/j.sbi.2016.10.015.

Vannini, Alessandro, and Patrick Cramer. 2012. "Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries." *Molecular Cell* 45 (4): 439–46. https://doi.org/10.1016/j.molcel.2012.01.023.

Venkatesh, Swaminathan, and Jerry L. Workman. 2015. "Histone Exchange, Chromatin Structure and the Regulation of Transcription." *Nature Reviews Molecular Cell Biology* 16 (3): 178–89. https://doi.org/10.1038/nrm3941.

Venter, J C, Myers E W Adams MD Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Kodira C D Zhang Q Zheng XQH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang JH, Miklos GLG, Nelson C, Broder S, Clark AG, Nadeau C, McKusick VA, Zinder, Levine A J N Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Reinert K Mobarry C Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng ZM, Di, Dunn P Francesco V Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge WM, Gong FC, Gu ZP, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke ZX, Ketchum KA, Lai ZW, Lei, Li Z Y YD Li JY, Liang Y, Lin XY, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue BX, Sun, et al. 2001. "The Sequence of the Human Genome." *Science* 291 (5507): 1304. https://doi.org/10.1126/science.1058040.

Verma, Neha, Ko Hsuan Hung, Jin Joo Kang, Nermeen H. Barakat, and William E. Stumph. 2013. "Differential Utilization of TATA Box-Binding Protein (TBP) and TBP-Related Factor 1 (TRF1) at Different Classes of RNA Polymerase III Promoters." *Journal of Biological Chemistry* 288 (38): 27564–70. https://doi.org/10.1074/jbc.C113.503094.

Verrijzer, C. Peter, Jin Long Chen, Kyoko Yokomori, and Robert Tjian. 1995. "Binding of TAFs to Core Elements Directs Promoter Selectivity by RNA Polymerase II." *Cell* 81 (7): 1115–25. https://doi.org/10.1016/S0092-8674(05)80016-9.

Vlasova, Irina A., and Paul R. Bohjanen. 2008. "Posttranscriptional Regulation of Gene Networks by GU-Rich Elements and CELF Proteins." *RNA Biology* 5 (4): 201–7. https://doi.org/10.4161/rna.7056.

Vlasova, Irina A., Jennifer McNabb, Arvind Raghavan, Cavan Reilly, Darlisha A. Williams, Kimberly A. Bohjanen, and Paul R. Bohjanen. 2005. "Coordinate

Stabilization of Growth-Regulatory Transcripts in T Cell Malignancies." *Genomics* 86 (2): 159–71. https://doi.org/10.1016/j.ygeno.2005.04.013.

Vlasova, Irina A., Nuzha M. Tahoe, Danhua Fan, Ola Larsson, Bernd Rattenbacher, Julius R. SternJohn, Jayprakash Vasdewani, et al. 2008. "Conserved GU-Rich Elements Mediate MRNA Decay by Binding to CUG-Binding Protein 1." *Molecular Cell* 29 (2): 263–70. https://doi.org/10.1016/j.molcel.2007.11.024.

Vogel, Katharina U., Lewis S. Bell, Alison Galloway, Helena Ahlfors, and Martin Turner. 2016. "The RNA-Binding Proteins Zfp36l1 and Zfp36l2 Enforce the Thymic β-Selection Checkpoint by Limiting DNA Damage Response Signaling and Cell Cycle Progression." *The Journal of Immunology* 197 (7): 2673–85. https://doi.org/10.4049/jimmunol.1600854.

Voges, D, P Zwickl, and W Baumeister. 1999. "The 26S P Roteasome : A Molecular M Achine Designed for Controlled." *Annual Review of Biochemistry* 68: 1015–68. https://doi.org/10.1146/annurev.biochem.68.1.1015.

Voigts, J, J H Siegle, D L Pritchett, C I Moore, A G Siapas, M A Wilson, R Urioste, et al. 2016. "Re s Ear Ch | r e p o r t S" 352 (6285): 595–600.

Vos, Seychelle M., Lucas Farnung, Marc Boehning, Christoph Wigge, Andreas Linden, Henning Urlaub, and Patrick Cramer. 2018. "Structure of Activated Transcription Complex Pol II–DSIF–PAF–SPT6." *Nature* 560 (7720): 607–12. https://doi.org/10.1038/s41586-018-0440-4.

Vuorio et al. 1990. "The Family Of Collagen Genes." *Annual Review of Biochemistry* 59 (1): 837–72. https://doi.org/10.1146/annurev.biochem.59.1.837.

Wacker, Jennifer L., M. Hadi Zareie, Hanson Fong, Mehmet Sarikaya, and Paul J. Muchowski. 2004. "Hsp70 and Hsp40 Attenuate Formation of Spherical and Annular Polyglutamine Oligomers by Partitioning Monomer." *Nature Structural and Molecular Biology* 11 (12): 1215–22. https://doi.org/10.1038/nsmb860.

Wall, M A, D E Coleman, E Lee, J A Iñiguez-Lluhi, B A Posner, A G Gilman, and S R Sprang. 1995. "The Structure of the G Protein Heterotrimer Gi Alpha 1 Beta 1 Gamma 2." *Cell* 83 (6): 1047–10458. https://doi.org/10.2210/pdb1gg2/pdb.

Wang, E., S. Kawaoka, M. Yu, J. Shi, T. Ni, W. Yang, J. Zhu, R. G. Roeder, and C. R. Vakoc. 2013. "Histone H2B Ubiquitin Ligase RNF20 Is Required for MLL-Rearranged Leukemia." *Proceedings of the National Academy of Sciences* 110 (10): 3901–6. https://doi.org/10.1073/pnas.1301045110.

Wang et al. 1992. "The Acidic Activator GAL4-AH Can Stimulate Polymerase II Transcription by Promoting Assembly of a Closed Complex Requiring TFIID and TFIIA." *Genes and Development* 6 (9): 1716–27. https://doi.org/10.1101/gad.6.9.1716.

Wang, Hongyang, Rui Li, Xiang Zhou, Liyao Xue, Xuewen Xu, and Bang Liu. 2016. "Genome-Wide Analysis and Functional Characterization of the Polyadenylation Site in Pigs Using RNAseq Data." *Scientific Reports* 6 (May): 1–11. https://doi.org/10.1038/srep36388.

Wang, P. J. 2002. "Functional Substitution for TAFII250 by a Retroposed Homolog That Is Expressed in Human Spermatogenesis." *Human Molecular Genetics* 11

(19): 2341–46. https://doi.org/10.1093/hmg/11.19.2341.

Warfel, Alwin Harry. 1970. “Specific Inhibition,” no. October: 447–49.

Weake and Workman. 2012. “SAGA Function in Tissue-Specific Gene Expression” 22 (4): 177–84. https://doi.org/10.1016/j.tcb.2011.11.005.SAGA.

Weids, Alan J., Sebastian Ibstedt, Markus J. Tamás, and Chris M. Grant. 2016. “Distinct Stress Conditions Result in Aggregation of Proteins with Similar Properties.” *Scientific Reports* 6 (March): 1–12. https://doi.org/10.1038/srep24554.

Weil, P. Anthony, and Stanley P. Blatti. 1976. “HeLa Cell Deoxyribonucleic Acid Dependent RNA Polymerases: Function and Properties of the Class III Enzymes.” *Biochemistry* 15 (7): 1500–1509. https://doi.org/10.1021/bi00652a022.

Weinmann, R., H. J. Raskas, and R. G. Roeder. 1974. “Role of DNA-Dependent RNA Polymerases II and III in Transcription of the Adenovirus Genome Late in Productive Infection.” *Proceedings of the National Academy of Sciences* 71 (9): 3426–30. https://doi.org/10.1073/pnas.71.9.3426.

Weiss, Samuel B., and Leonard Gladstone. 1959. “A Mammalian System for The Incorporation of Cytidine Triphosphate into Ribonucleic Acid.” *Journal of the American Chemical Society* 81 (15): 4118–19. https://doi.org/10.1021/ja01524a087.

Welch, Ellen M., and Allan Jacobson. 1999. “An Internal Open Reading Frame Triggers Nonsense-Mediated Decay of the Yeast SPT10 MRNA.” *EMBO Journal* 18 (21): 6134–45. https://doi.org/10.1093/emboj/18.21.6134.

Weng, Lingjie, Y I Li, Xiaohui Xie, and Yongsheng Shi. 2016. “Poly (A) Code Analyses Reveal Key Determinants for Tissue-Specific MRNA Alternative Polyadenylation,” 813–21. https://doi.org/10.1261/rna.055681.115.4.

Wertent, Sebastiaan, André Mitschler, Christophe Romier, Yann Gaël Gangloff, Sylvie Thuault, Irwin Davidson, and Dino Moras. 2002. “Crystal Structure of a Subcomplex of Human Transcription Factor TFIID Formed by TATA Binding Protein-Associated Factors HTAF4 (HTAFII135) and HTAF12 (HTAFII20).” *Journal of Biological Chemistry* 277 (47): 45502–9. https://doi.org/10.1074/jbc.M206587200.

Wheeler, Joshua R., Tyler Matheny, Saumya Jain, Robert Abrisch, and Roy Parker. 2016. “Distinct Stages in Stress Granule Assembly and Disassembly.” *ELife* 5 (Se): 1–25. https://doi.org/10.7554/eLife.18413.

Wierzbicki, Andrzej T., Thomas S. Ream, Jeremy R. Haag, and Craig S. Pikaard. 2009. “RNA Polymerase v Transcription Guides ARGONAUTE4 to Chromatin.” *Nature Genetics* 41 (5): 630–34. https://doi.org/10.1038/ng.365.

Wild, Thomas, and Patrick Cramer. 2012. “Biogenesis of Multisubunit RNA Polymerases.” *Trends in Biochemical Sciences* 37 (3): 99–105. https://doi.org/10.1016/j.tibs.2011.12.001.

William S Mohan, Elisabeth Scheer, Olivia Wendling, and Daniel Metzger. 2003. “TAF10 (TAFII30) Is Necessary for TFIID Stability and Early Embryogenesis in

Mice William." *American Society for Microbiology* 23 (12): 4307–18. https://doi.org/10.1128/MCB.23.12.4307.

Willison, Keith R. 2018. "The Substrate Specificity of Eukaryotic Cytosolic Chaperonin CCT." *Philosophical Transactions of the Royal Society B: Biological Sciences* 373 (1749). https://doi.org/10.1098/rstb.2017.0192.

Winn, S. I., H. C. Watson, R. N. Harkins, and L. A. Fothergill. 1981. "Structure and Activity of Phosphoglycerate Mutase." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 293 (1063): 121–30. https://doi.org/10.1098/rstb.1981.0066.

Wissink, Erin M., Elizabeth A. Fogarty, and Andrew Grimson. 2016. "High-Throughput Discovery of Post-Transcriptional Cis-Regulatory Elements." *BMC Genomics* 17 (1): 1–14. https://doi.org/10.1186/s12864-016-2479-7.

Wittmann, J., E. M. Hol, and H.-M. Jack. 2006. "HUPF2 Silencing Identifies Physiologic Substrates of Mammalian Nonsense-Mediated MRNA Decay." *Molecular and Cellular Biology* 26 (4): 1272–87. https://doi.org/10.1128/mcb.26.4.1272-1287.2006.

Wood, Adam, Jessica Schneider, Jim Dover, Mark Johnston, and Ali Shilatifard. 2003. "The Paf1 Complex Is Essential for Histone Monoubiquitination by the Rad6-Bre1 Complex, Which Signals for Histone Methylation by COMPASS and Dot1p." *Journal of Biological Chemistry* 278 (37): 34739–42. https://doi.org/10.1074/jbc.C300269200.

Woodcock, C L F, J P Safer, and J E Stanchfield. 1976. "STRUCTURAL REPEATING UNITS IN CHROMATIN I . Evidence for Their General Occurrence It Has Recently Been Shown That under Cer- Tain Preparative Conditions , Chromatin Fibers Exhibit a Distinctive Particulate Sub- Structure [l-3]. This Is in Contrast to Th" 97 (1975).

Woodruff, Jeffrey B., Beatriz Ferreira Gomes, Per O. Widlund, Julia Mahamid, Alf Honigmann, and Anthony A. Hyman. 2017. "The Centrosome Is a Selective Condensate That Nucleates Microtubules by Concentrating Tubulin." *Cell* 169 (6): 1066-1077.e10. https://doi.org/10.1016/j.cell.2017.05.028.

Wright, K. J., M. T. Marr, and R. Tjian. 2006. "TAF4 Nucleates a Core Subcomplex of TFIID and Mediates Activated Transcription from a TATA-Less Promoter." *Proceedings of the National Academy of Sciences* 103 (33): 12347–52. https://doi.org/10.1073/pnas.0605499103.

Xiao, Jia Hao, Irwin Davidson, Hans Matthes, Jean Marie Garnier, and Pierre Chambon. 1991. "Cloning, Expression, and Transcriptional Properties of the Human Enhancer Factor TEF-1." *Cell* 65 (4): 551–68. https://doi.org/10.1016/0092-8674(91)90088-G.

Xiao, Mei-Sheng, Bin Zhang, Yi-Sheng Li, Qingsong Gao, Wei Sun, and Wei Chen. 2016. "Global Analysis of Regulatory Divergence in the Evolution of Mouse Alternative Polyadenylation." *Molecular Systems Biology* 12 (12): 890. https://doi.org/10.15252/msb.20167375.

Xie, Xiaohui, Lu Jun, Kulbokas E J, Golub Todd R, Mootha Vamsi, Lindblad-Toh Kerstin, Lander Eric S, and Kellis Manolis. 2005. "Systematic Discovery of

Regulatory Motifs in Human Promoters and 3' UTRs by Comparison of Several Mammals." *Nature* 434 (March): 338–45. https://doi.org/10.1038/nature03441.

Xie, Xiaoling, Tetsuro Kokubo, Steven L. Cohen, Urooj A. Mirza, Alexander Hoffmann, Brian T. Chait, Robert G. Roeder, Yoshihiro Nakatani, and Stephen K. Burley. 1996. "Structural Similarity between TAFs and the Heterotetrameric Core of the Histone Octamer." *Nature* 380 (6572): 316–22. https://doi.org/10.1038/380316a0.

Xu, Yali, Joseph P. Milazzo, Tim D.D. Somerville, Yusuke Tarumoto, Yu Han Huang, Elizabeth L. Ostrander, John E. Wilkinson, Grant A. Challen, and Christopher R. Vakoc. 2018. "A TFIID-SAGA Perturbation That Targets MYB and Suppresses Acute Myeloid Leukemia." *Cancer Cell* 33 (1): 13-28.e8. https://doi.org/10.1016/j.ccell.2017.12.002.

Yam, Alice Y., Yu Xia, Hen Tzu Jill Lin, Alma Burlingame, Mark Gerstein, and Judith Frydman. 2008. "Defining the TRiC/CCT Interactome Links Chaperonin Function to Stabilization of Newly Made Proteins with Complex Topologies." *Nature Structural and Molecular Biology* 15 (12): 1255–62. https://doi.org/10.1038/nsmb.1515.

Yokomori, Kyoko, Jin-Long Chen, C. Peter Verrijzer, Robert Tjian, Martin P. Zeidler, and Marek Mlodzik. 1994. "Drosophila TFIIA Directs Cooperative DNA Binding with TBP and Mediates Transcriptional Activation." *Genes and Development* 8 (19): 2313–23. https://doi.org/10.1101/gad.8.19.2313.

Yokoyama, Takeshi, Mikako Shirouzu, Haruhiko Ehara, Hideki Shigematsu, Shigeyuki Yokoyama, and Shun-ichi Sekine. 2017. "Structure of the Complete Elongation Complex of RNA Polymerase II with Basal Factors." *Science* 357 (6354): 921–24. https://doi.org/10.1126/science.aan8552.

Yoon, Je-hyun, and Myriam Gorospe. 2016. "RNA-Protein Complexes and Interactions" 1421: 15–22. https://doi.org/10.1007/978-1-4939-3591-8.

Yoon, Je Hyun, Supriyo De, Subramanya Srikantan, Kotb Abdelmohsen, Ioannis Grammatikakis, Jiyoung Kim, Kyoung Mi Kim, et al. 2014. "PAR-CLIP Analysis Uncovers AUF1 Impact on Target RNA Fate and Genome Integrity." *Nature Communications* 5: 1–15. https://doi.org/10.1038/ncomms6248.

Young, Jason C., José M. Barral, and F. Ulrich Hartl. 2003. "More than Folding: Localized Functions of Cytosolic Chaperones." *Trends in Biochemical Sciences* 28 (10): 541–47. https://doi.org/10.1016/j.tibs.2003.08.009.

Young, Jason C, and David W Andrews. 1996. "Assembles Co-Translationally on the Endoplasmic Translation Pause in Vitro" 15 (1): 172–81.

Yu et al. 1983. "Intracellular Assembly of Human Fibrinogen." *Journal of Biological Chemistry* 258 (22): 13407–10.

Yudkovsky, Natalya, Jeffrey A Ranish, and Steven Hahn. 2000. "A Transcription Reinitiation Intermediate That Is Stabilized by Activator." *Nature* 408 (6809): 225–29. https://doi.org/10.1038/35041603.

Zaessinger, S., I. Busseau, and M. Simonelig. 2006. "Oskar Allows Nanos MRNA Translation in Drosophila Embryos by Preventing Its Deadenylation by

Smaug/CCR4." *Development* 133 (22): 4573–83. https://doi.org/10.1242/dev.02649.

Zeitelhofer, M., D. Karra, P. Macchi, M. Tolino, S. Thomas, M. Schwarz, M. Kiebler, and R. Dahm. 2008. "Dynamic Interaction between P-Bodies and Transport Ribonucleoprotein Particles in Dendrites of Mature Hippocampal Neurons." *Journal of Neuroscience* 28 (30): 7555–62. https://doi.org/10.1523/jneurosci.0104-08.2008.

Zhang, Lixin, Virpi Paakkarinen, Klaas Jan, Van Wijk, and Eva-mari Aro. 1999. "Co-Translational Assembly of the D1 Protein into Photosystem II Co-Translational Assembly of the D1 Protein into Photosystem II *." *The Journal of Biological Chemistryournal of Biological Chemistry* 274 (23): 16062–67.

Zhang, Xiao Yong, Maya Varthi, Stephen M. Sykes, Charles Phillips, Claude Warzecha, Wenting Zhu, Anastasia Wyce, Alan W. Thorne, Shelley L. Berger, and Steven B. McMahon. 2008. "The Putative Cancer Stem Cell Marker USP22 Is a Subunit of the Human SAGA Complex Required for Activated Transcription and Cell-Cycle Progression." *Molecular Cell* 29 (1): 102–11. https://doi.org/10.1016/j.molcel.2007.12.015.

Zhao, Rongmin, Mike Davey, Ya Chieh Hsu, Pia Kaplanek, Amy Tong, Ainslie B. Parsons, Nevan Krogan, et al. 2005. "Navigating the Chaperone Network: An Integrative Map of Physical and Genetic Interactions Mediated by the Hsp90 Chaperone." *Cell* 120 (5): 715–27. https://doi.org/10.1016/j.cell.2004.12.024.

Zhao, Wei Meng, and Guowei Fang. 2005. "Anillin Is a Substrate of Anaphase-Promoting Complex/Cyclosome (APC/C) That Controls Spatial Contractility of Myosin during Late Cytokinesis." *Journal of Biological Chemistry* 280 (39): 33516–24. https://doi.org/10.1074/jbc.M504657200.

Zhao, Yue, Guillaume Lang, Saya Ito, Jacques Bonnet, Eric Metzger, Shun Sawatsubashi, Eriko Suzuki, et al. 2008. "A TFTC/STAGA Module Mediates Histone H2A and H2B Deubiquitination, Coactivates Nuclear Receptors, and Counteracts Heterochromatin Silencing." *Molecular Cell* 29 (1): 92–101. https://doi.org/10.1016/j.molcel.2007.12.011.

Zhou, Bing-Rui, and Yawen Bai. 2019. "Chromatin Structures Condensed by Linker Histones." *Essays In Biochemistry* 63 (1): 75–87. https://doi.org/10.1042/ebc20180056.

Zhu, Bing, Subhrangsu S. Mandal, Anh Dung Pham, Yong Zheng, Hediye Erdjument-Bromage, Surinder K. Batra, Paul Tempst, and Danny Reinberg. 2005. "The Human PAF Complex Coordinates Transcription with Events Downstream of RNA Synthesis." *Genes and Development* 19 (14): 1668–73. https://doi.org/10.1101/gad.1292105.

Zybailov, Boris, Amber L. Mosley, Mihaela E. Sardiu, Michael K. Coleman, Laurence Florens, and Michael P. Washburn. 2006. "Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces Cerevisiae." *Journal of Proteome Research* 5 (9): 2339–47. https://doi.org/10.1021/pr060161n.

Zylber, E. A., and S. Penman. 1971. "Products of RNA Polymerases in HeLa Cell Nuclei." *Proceedings of the National Academy of Sciences* 68 (11): 2861–65.

https://doi.org/10.1073/pnas.68.11.2861.

Pooja MUKHERJEE

Etude des mécanismes d'assemblage co-traductionnel des complexes protéiques impliqués dans la transcription dans les cellules de mammifère

Résumé en français

La majorité des processus biologiques sont réalisés par des complexes protéiques multisubunités dans les cellules et une quantité importante d'énergie est requise par les cellules pour construire ces énormes complexes. Contrairement aux bactéries, les gènes codant pour les protéines sont dispersés dans le génome des eucaryotes, ce qui complique la compréhension de l'assemblage des complexes protéiques. En utilisant l'immunoprécipitation d'ARN suivie par la détection des ARNm à l'échelle du génome par analyse par micropuce, ARN molécule unique, FISH, immunofluorescence, cellules souches embryonnaires knock-out de souris et approches de permutation de domaines, nous montrons que les complexes de transcription multisubunit de mammifère s'assemblent de manière co-traductionnelle. Nous démontrons que les domaines de dimérisation et leurs positions dans les sous-unités en interaction déterminent la voie d'assemblage de co-traduction (simultanée ou séquentielle). En outre, les expériences cytoplasmiques IF-smFISH et bicolores smFISH indiquent que l'assemblage de co-traduction décrit se produit clairement dans le cytoplasme de cellules humaines. Des résultats identiques dans les cellules de levure, de souris et humaine suggèrent que l'assemblage par co-traduction est un mécanisme général chez les eucaryotes, qui pourrait être nécessaire pour éviter les interactions non spécifiques et l'agrégation de protéines dans la cellule.

Mots-clés: Transcription, Complexes de transcription, TFIID, Assemblage co-traductionnelle, Complexes protéiques composé de plusieurs sous-unités, Immunoprécipitation d'ARN, FISH contre molécule simple

Resume in English

Majority of the biological processes are carried out by multisubunit protein complexes in cells and a significant amount of energy is required by the cells to build these huge complexes. Unlike bacteria, genes encoding proteins are dispersed in the genome of eukaryotes and this makes the assembly of protein complexes more complicated to understand. By using RNA immunoprecipitation followed by genome-wide detection of mRNAs by microarray analysis, single molecule RNA FISH, immunofluoresence, mouse knock-out embryonic stem cells and domain swapping approaches, we show that the mammalian multisubunit transcription complexes assemble co-translationally. We demonstrate that the dimerization domains and their positions in the interacting subunits determine the co-translational assembly pathway (simultaneous or sequential). Furthermore, cytoplasmic IF-smFISH and two-colour smFISH experiments indicate that the described co-translational assembly is clearly occurring in the cytoplasm of human cells. Identical results in yeast, mouse and human cells suggests that co-translational assembly is a general mechanism in eukaryotes which might be necessary to avoid non-specific interactions and protein aggregation in the cell.

Keywords: Transcription, Transcription complexes, TFIID, Co-translational assembly, multisubunit protein complexes, RNA Immunoprecipitation, single molecule FISH