Biomolecular condensation of intrinsically disordered regions in *P. falciparum* and *H. sapiens* transcription factors

By Leonidas Marthinus Karamanof

Dissertation presented for the degree of

Master of Science

Supervisor: Dr Thomas Oelgeschläger

Department of Molecular and Cell Biology

Faculty of Science

University of Cape Town

South Africa

February 2022

Plagiarism Declaration

I am presenting this dissertation in full fulfilment of the requirements for my degree.

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

I have used the *Nature* convention for citation and referencing. Each contribution to, and quotation in, this *Dissertation* from the work(s) of other people has been attributed and has been cited and referenced.

I declare that all the work in this *Dissertation*, save for that which is properly acknowledged, is my own.

I have not allowed and will not allow anyone to copy my work with the intention of passing it off as his or her own work.

Name: Leonidas Marthinus Karamanof

Signed: Signed by candidate

Date: 13 February 2022

Acknowledgements

Thank you to the University of Cape Town (UCT), the National Research Foundation (NRF) and the Harry Crossley Foundation (HCF) for financial support in the form of grant money and scholarships. The opinions and conclusions reached in this study are those of the author and are not attributed to UCT, NRF or HCF.

Thank you to Dr Thomas Oelgeschläger and all members of the noodle lab, both past and present. Thomas' supervision, expansive knowledge of eukaryotic transcription and keen editor's eye have been invaluable in putting this thesis together. Thank you to Jasmin, Joanna and Keso for being great lab partners, always willing to lend a helping hand. You guys made pandemic-era MCB manageable! I would also like to acknowledge Lize van der Linden for starting the investigation into the IDRs of *Pf*TLP; without her initial curiosity, this project would not have come together.

I would like to thank Dr Richard Young, Dr Nancy Hannett and Dr Jonathan Henninger at the Whitehead Institute for sharing reagents and image analysis code. They patiently communicated with me across the Atlantic, overcoming logistical issues and my limited knowledge of Python.

A special thanks to Prof Dirk Lang and Carla van Niekerk at the confocal microscopy unit. The mixed droplet images and FRAP experiments were captured with their expert assistance.

A big thank you to Jeremy Wilkinson and Claude Formanek for their help getting to grips with Python and Linux and making me realise I probably should have done computer science. I also appreciate the welcome distractions along the way.

Tessa Botha, thank you for your precise copy-editing skills, hunting down stray commas and semi-colons.

Finally, thank you to my parents, Michele and Michael Karamanof, and my girlfriend, Jess Bourn for their support over the years.

Abstract

Plasmodium falciparum is still a major cause of disease and death, especially in sub-Saharan Africa. The organism's complex life cycle is tightly coupled to a carefully controlled gene expression programme, and its highly divergent transcriptional machinery suggests unique mechanisms of transcriptional control. Biomolecular condensation, driven by intrinsically disordered regions (IDRs) in proteins, has emerged as a previously underappreciated mechanism of cellular compartmentalisation and regulation. The study of eukaryotic transcription has been revolutionised by the understanding that biomolecular condensates formed by transcription factors play a key role in transcriptional regulation. The study these of so-called transcriptional condensates has so far not addressed the role of the general transcription factors (GTFs) and has not investigated the phenomenon in non-model eukaryotes such as *P. falciparum.*

Here we report on the biochemical purification and functional characterisation of an ensemble of TF-IDR fluorescent protein fusion constructs, using fluorescence microscopy. During the course of this work, we established a standardised data acquisition, analysis and visualisation pipeline for fluorescence microscopy images. Bioinformatic analysis uncovered unique sequence characteristics of the *P. falciparum* RPB1 C-terminal domain (*Pf*RPB1-CTD) that suggest a functionally divergent role in biomolecular condensation. This work demonstrates that the *Pf*RPB1-CTD, as well as a panel of GTF-IDR fusion proteins, including *P. falciparum* (*Pf*) and *Homo sapiens* (*Hs*) TATA-binding protein-IDR (TBP), TFIIAαß-IDR and TFIIFß-IDR can drive biomolecular condensation *in vitro*. Comparative analyses show that *Pf*RPB1-CTD and *Pf*TBP-IDR drive biomolecular condensation at much lower protein concentrations than homologous regions from *H. sapiens,* suggesting a potential role in organising *P. falciparum* transcriptional condensates. This work further expands on the characterisation of the IDR from the largest subunit of the Mediator complex, MED1-IDR. Our results underscore its status as a strong driver of biomolecular condensation and further show that the material properties of MED1-IDR condensates are affected by a range of environmental conditions.

Finally, we report on the compatibility of *P. falciparum* and *H. sapiens* proteins to form mixed condensates. We demonstrate that both *Pf*RPB1-CTD and MED1-IDR exhibit wide-ranging compatibility with GTF-IDR fusion proteins, that are recruited to and concentrated in mixed assemblies. However, while MED1-IDR recruits *Hs*RPB1-CTD to form homogenously mixed transcriptional condensates, *Pf*RPB1-CTD and MED1-IDR form heterogenous mixed condensates, in which the two protein components form separate compartments.

This result provides first evidence that transcriptional condensates formed in human cells and in the malaria parasite may have distinct properties and thus may provide a highly promising target for the much-needed development of new antimalarial drugs.

Common abbreviations

A – adenine

AMP – ampicillin

ApiAP2 – *Apicomplexan* AP2

BSA – bovine serum albumin

C – cytosine

CAM – chloramphenicol

CDK – cyclin dependent kinase

CFP – cyan fluorescent protein

CTD – carboxyl terminal domain

DNA – deoxyribonucleic acid

ds – double stranded

DTT – dithiothreitol

E. coli – Escherichia coli

EDTA – ethylenediaminetetraacetic acid

Em – emission

Ex – excitation

FRAP – fluorescence recovery after photobleaching

G – guanine

GFP – green fluorescent protein

GTF – general transcription factor

HRP – horseradish peroxidase

Hs – Homo sapiens

HSV1 – Herpes Simplex Virus 1

IDC – intra-erythrocyte development cycle

IDR – intrinsically disordered region

IPTG – isopropyl β-D-1-thiogalactopyranoside

KAN – kanamycin

LB – lysogeny broth

LCR – low complexity region

LLPS – liquid-liquid phase separation

MED1 – Mediator subunit 1

MLO – membrane-less organelle

mRNA – messenger RNA

MUSCLE – Multiple Sequence Comparison by Log-Expectation

MW – molecular weight

NE – nuclear extract

NP-40 – Nonidet P-40 Substitute

NTD – amino terminal domain

OCT4 – octamer-binding transcription factor 4

OD – optical density

PAGE – polyacrylamide gel electrophoresis

PCR – polymerase chain reaction

PCV – packed column volume

PEG – polyethylene glycol

Pf – Plasmodium falciparum

PIC – pre-initiation complex

PMSF – phenylmethylsulphonyl fluoride

PONDR – Predictor of Natural Disordered Regions

PTEFb – positive transcription elongation factor

PTM – post-translational modification

PVDF – polyvinylidene difluoride

RBD – RNA binding domain

RPB1 – RNAPII subunit 1

RNA – ribonucleic acid

RNAPII – RNA polymerase II

SD – standard deviation

SDS – sodium dodecyl sulphate

SP-motif – serine proline motif

T – thymine

TAF – TBP associated factor

TBP – TATA-binding protein

TF – transcription factor

Tris – trisaminomethane

VP16 – viral protein 16

YFP – yellow fluorescent protein

Contents

Introduction

Plasmodium falciparum: impact and life cycle

Plasmodium falciparum is a eukaryotic, protozoan parasite belonging to the phylum *Apicomplexa*[1]. It is the causative agent of the deadliest form of malaria[2], causing around 200 million cases annually and leading to half a million deaths as a result of infection[3]. The disease is a particular burden on young children in sub-Saharan Africa, where it is the cause of 20% of infant mortalities[4]. Whilst major strides in combatting the disease have been made over the past decades, growing resistance to current antimalarial drugs threatens to overturn hard-fought gains[5]. The large reductions in cases and deaths seen in the early 21st century have stalled in recent years, highlighting the need for innovative approaches to understanding and tackling the threat of *P. falciparum*[3]. Whilst ongoing trials on the use of the antimalarial RTS,S/AS01 vaccine show some promise[6], the elimination of the disease will require a multipronged approach, combining public policy, vector control and novel anti-malarial therapies informed by insights into the molecular mechanisms governing parasite biology.

The malaria pathogen moves through a complex, multistage life cycle in which changing hosts and environments are accompanied by drastic morphological changes (Fig. I1)[2,7]. *P. falciparum* is transmitted by female *Anopheles* mosquito vectors and undergoes zoite life-stages in human host cells[1]. *P. falciparum* sporozoites, injected into a human host during a mosquito bloodmeal, travel to the liver where they invade hepatocytes[4]. Through a form of asexual reproduction, termed schizogony, tens of thousands of merozoites are produced and released from hepatocytes[1]. Merozoites infect red blood cells to begin the intra-erythrocytic phase of parasite reproduction. In the 48 hour intra-erythrocyte development cycle (IDC) merozoites develop through ring and trophozoite stages and reproduce by schizogony[1]. The release of merozoites from rupturing red blood cells every 48 hours produces the typical clinical symptoms of malaria infection, in particular the sharp onset of high fever, accompanied by headaches, nausea and muscle pains[1,7]. During the IDC, a fraction of merozoites are reprogrammed to differentiate into female and male gametocytes which are ingested by feeding mosquitoes through the bloodstream[4]. Sexual reproduction occurs in the mosquito midgut, through formation of an oocyst and finally the production of motile sporozoites to complete the parasitic cycle. These numerous changes in host cell-type and parasite morphology are accompanied by tightly controlled gene expression programmes[4]. Certain mechanisms of gene regulation in the parasite, in particular chromatin-mediated regulation[8], are well studied, but regulation at the level of transcription initiation, mediated by DNA-binding transcription regulatory proteins and the general RNA polymerase transcription machinery, remains poorly understood.

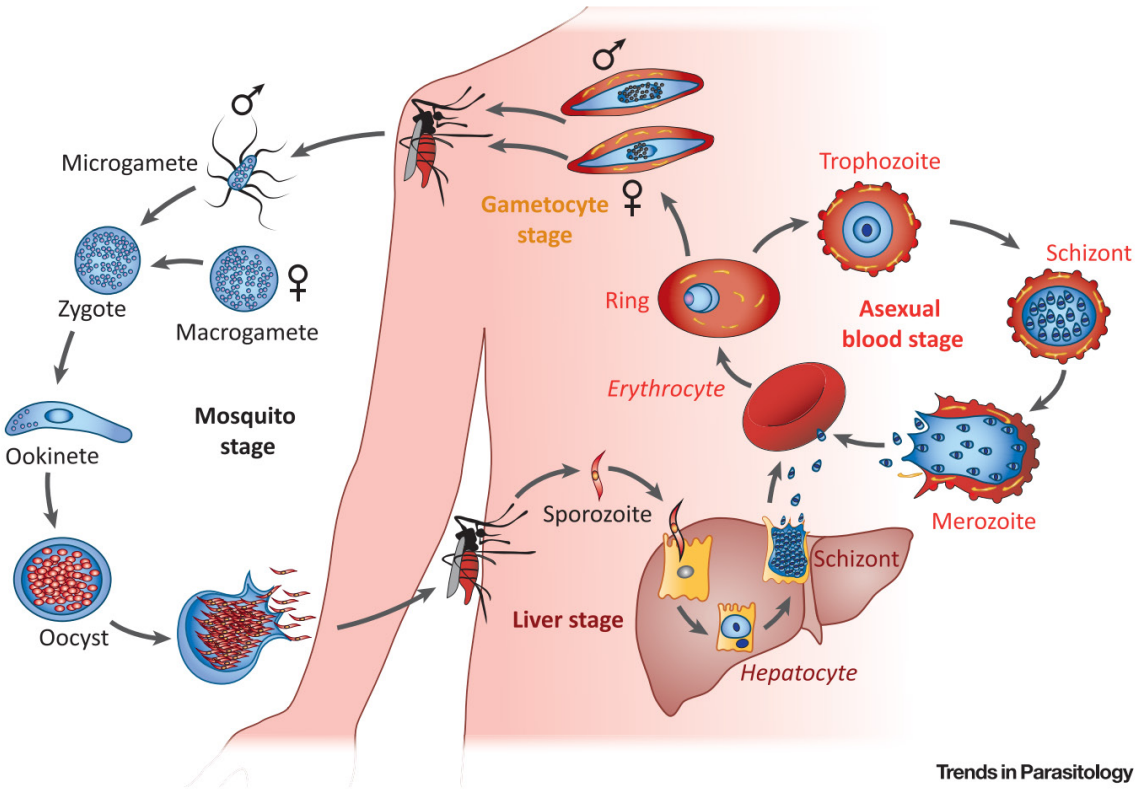


Fig. I1. *Plasmodium falciparum* exhibits a complex, multi-stage life cycle. The parasite experiences many morphological changes as it moves through different hosts and environments. These morphological changes are tightly coupled to a precisely regulated gene expression programme. *Adapted from Maier et al., 2019* [2].

Gene regulation in *Plasmodium falciparum*

P. falciparum exhibits uncommon genomic and proteomic features that suggest unique mechanisms of transcriptional control. Complete sequencing of the organism's genome revealed an extreme compositional bias, with adenine (A) and thymine (T) making up more than 80% of genomic sequences[9,10]. This bias in nucleotide content likely increases the difficulty of differentiating DNA regulatory elements from a homogenous background[11] and also skews the amino acid composition of the *P. falciparum* proteome[9,12].

Past work by our research group has sought to understand how *P. falciparum* transcription factors might recognise core-promoter elements in this context[13–15]. We have focused on proteins of the *P. falciparum* general transcription machinery, which display significant sequence divergence compared to other eukaryotes[16,17]. Many components of the general transcription machinery could initially not be identified by homology searches of the organism's predicted proteome[10]. Later investigations, using hydrophobic cluster analysis, uncovered many of the general transcription factors (GTFs), and revealed extreme sequence divergence of these generally highly conserved

eukaryotic proteins[17]. Other transcriptional proteins seem to have been completely lost in the parasite or have diverged so dramatically in sequence that current bioinformatic tools are unable to uncover them. For example, none of the TBP-associated factors (TAFs) containing histone fold domains have thus far been identified, suggesting a unique arrangement of the parasite's TFIID complex[17]. Furthermore, only a single protein of the 25 subunit Mediator complex, a central regulator of eukaryotic transcription, has been identified in the parasite[16]. The crucial role of Mediator in canonical eukaryotic transcription[18], described below, points to critical differences in the structure of the *P. falciparum* transcription machinery and mechanisms of transcription regulation.

Compared to other eukaryotes, *P. falciparum* is further characterised by an apparent paucity of DNA-binding transcription regulatory factors[8,16]. So far, only an *Apicomplexan*-specific set of regulatory proteins, the ApiAP2 family of transcription factors, could be identified and studied in greater detail [8,19]. It has been demonstrated that sub-families of ApiAP2 transcription factors are expressed during different parasitic stages, suggesting that they may play a role as master regulators in differentiation and life-cycle control[19].

A further unusual feature of *P. falciparum* is that the organism's proteome is marked by an extremely high prevalence of low complexity regions (LCRs)[9]. These are regions of proteins that do not exploit the full suite of amino acids but are highly enriched in a subset of residues. Around 30% of the *P. falciparum* proteome consists of long, uninterrupted stretches of the amino acids asparagine[20]. Whilst the proximate cause of these repeats is likely the codon bias imposed by the A/T-rich genome, their prevalence in well-conserved proteins suggests some functional relevance for these sequences[12,20].

Whilst the exact molecular mechanisms with which *P. falciparum* accomplishes its gene regulation are not fully understood, gene expression data clearly demonstrate the precise control of mRNA dynamics throughout the parasite's life cycle, with distinct sets of genes transcribed at precise points in the various life stages[7,21,22]. Research in the last decade has largely focused on the role of epigenetic regulation, long non-coding RNAs and chromatin architecture in *P. falciparum* gene regulation[8]. In contrast, next to nothing is known about the assembly, structure, and function of the ultimate target of gene regulatory pathways – the general transcription machinery, composed of GTFs and RNA polymerase.

The general model of eukaryotic transcription regulation

In contrast to *P. falciparum,* the basics of transcription initiation are well understood in model eukaryotic organisms such as *Saccharomyces cerevisiae* and *Homo sapiens*. The canonical model of eukaryotic transcriptional control describes a stepwise process, whereby GTFs assemble at the core promoter to form the pre-initiation complex (PIC)[23]. RNA Polymerase II (RNAPII), responsible for transcription from protein-coding genes, cannot directly recognise DNA sequence elements and relies on the GTFs to be recruited to sites of transcription[24]. The first of these GTFs to bind to the core promoter is the TATA-binding protein (TBP), which in conjunction with TAFs forms the TFIID complex[25]. Following TBP/TFIID binding, TFIIA and TFIIB are recruited and stabilise the TBP-DNA interaction[23]. TFIIB is also responsible for the subsequent recruitment of the RNAPII/TFIIF complex[24]. The remaining GTFs, including TFIIE and TFIIH, play important roles during the processes of open complex formation and the initiation of transcription by RNAPII [26].

Sequence-specific transcriptional activators bind DNA at regulatory promoter regions, including enhancers, which can be hundreds of thousands of base pairs away from the genes they influence[18]. One of the key regulators of eukaryotic RNAPII transcription is Mediator, a large, multi-subunit protein complex that integrates and mediates the signals from transcriptional activators to the general transcription machinery[26]. Initially characterised as an important co-activator, the Mediator is now considered a GTF, since it is capable of modulating the activity of the PIC on its own and participates in all transcription initiation events[27,28].

All transcription associated proteins, including general and specific transcription factors, the Mediator complex as well as histone-modifying proteins, ultimately serve to affect the activity of RNAPII and its occupancy at gene promoters. In addition to the mechanisms described above, RNAPII itself contains a protein region central to eukaryotic transcriptional regulation: the RPB1-CTD.

The RPB1-CTD is a nexus of transcriptional control

RNAPII is a 12-subunit multiprotein complex, with the constitutive components designated RPB1-12. The largest subunit, RPB1, is characterised by a highly repetitive C-terminal domain. The sequence patterning of the RPB1-CTD is so striking, and its role in transcription regulation so central, that it is often referred to simply as “the CTD”. The CTD does not play a catalytic role in the synthesis of RNA transcripts, but rather as an important regulator of transcription initiation and co-transcriptional processes[29]. It is largely unstructured and extends away from the catalytic, globular core of RNAPII[29]. Despite its low-complexity nature and lack of structure, the CTD is remarkably well conserved across higher eukaryotes[30,31]. Almost all species of plants, animals and fungi possess between 25 and 52 repeats of the characteristic amino acid heptad “YSPTSPS”[30,32]. In all mammals, including *H. sapiens,*

the CTD is made up of 52 near-perfect repeats of the heptad amino acid sequence[31]. Other eukaryotic organisms, especially unicellular species, may not possess canonical CTD sequences, but their RBP1 sequences often still display repetitive, unstructured C-termini[33,34].

Post-translational modification of CTD residues is critical in the regulation of transcription[35]. Five of the seven heptad repeat residues are phosphor-acceptors and subject to phosphorylation by kinases including CDK7, a component of TFIIH, CDK8, a component of Mediator and CDK9, a component of the transcription elongation factor PTEFb, during different stages of the transcription cycle. Phosphorylation, along with glycosylation of serine residues and cis-trans isomerisation of proline residues, create a range of combinatorial possibilities that make up the CTD code[32]. Various sets of modification states are associated with different stages of mRNA production, including transcription initiation, elongation and termination as well as mRNA capping and co-transcriptional splicing.

The CTD is an inherently unstructured protein region that does not fold into a well-conserved three-dimensional conformation. Labile, unstructured protein regions such as the CTD are known to engage in types of weak intra- and intermolecular interactions that can drive biomolecular condensation[36], which has emerged in recent years as an important mechanism in the regulation of transcription[37].

Biomolecular condensation is an important general regulatory mechanism in cells

The past decade has seen the field of cellular biochemistry revolutionised by investigations into biomolecular condensates and an understanding of the regulatory functions they play in cells[36,38]. Biomolecular condensation describes the spontaneous formation of high-density assemblies of biological molecules, usually proteins and/or nucleic acids. In biology, the term describes a range of phenomena in which homogenous solutions spontaneously de-mix into multi-phase systems, where different concentrations of single components can co-exist. In cells, condensates can selectively concentrate specific biological molecules, as well as exclude others, and therefore function as non-membrane-bound assemblies, often termed membrane-less organelles (MLOs)[36,39]. Prominent examples of MLOs include the nucleolus and P-bodies in the nucleus and stress granules in the cytoplasm[40–43].

Biomolecular condensation is often driven by weak intermolecular interactions between intrinsically disordered regions (IDRs) of proteins[44]. Intrinsic disorder in protein structure is the result of low complexity amino acid sequences. Low complexity regions are protein regions with a skewed amino acid composition, heavily enriched in certain residues and lacking others[45]. As shown in Fig. I2A, the canonical RPB1-CTD is a clear example: a protein region hundreds of amino acids long with an extreme prevalence of tyrosine, serine, proline and threonine residues and an almost complete absence of charged residues[33,46]. However, there are many other types of protein sequences that form IDRs and

are capable of engaging in the weak, multivalent interactions necessary to drive biomolecular condensation. For example, protein regions consisting of alternating blocks of positively and negatively charged residues, separated by flexible linker sequences have been shown to possess a strong disposition to drive condensation[47–49]. Furthermore, protein regions such as RNA binding domains (RBDs) can promote strong binding to RNAs and subsequently drive condensation of protein-RNA complexes through a range of interactions between different chemical groups, including the RNA backbone and bases[50,51]. The condensation state of an IDR-containing protein, or any condensation-prone biomolecule, is primarily determined by its concentration[52]. At low concentrations, molecules will exist in a single-phase state, dissolved in the surrounding aqueous solution. At higher concentrations, increased favourable interactions cause the spontaneous assembly of biomolecules into high-density condensates (Fig. I2).

The term biomolecular condensation encompasses a wide range of biophysical processes that result in the formation of high-density biomolecular assemblies. It is impartial with respect to the process by which the assemblies are formed as well as their final material properties[53]. The assemblies formed through biomolecular condensation may exhibit a range of material states, ranging from viscous liquids to gels and crystalline structures[54]. In contrast, the term liquid-liquid phase separation (LLPS) refers to a specific physical process, whereby a homogenous solution of two (or more) components spontaneously separates into distinct phases, which are both characterised by liquid properties[52]. Many biomolecular condensates also exhibit an "aging" phenomenon, whereby they initially form through LLPS and transition to gel- and solid-like states over time[50,55].

Biomolecular condensation is harnessed by the cell to spatiotemporally regulate cellular processes[53]. Condensates can selectively enrich certain components through favourable interactions or conversely exclude others by regulating mesh size, phase-states and dominant interaction types[56]. In cells and *in vitro,* condensates can be formed by and contain a mixture of many different biomolecules. In complex, mixed condensates, one distinguishes between homotypic and heterotypic interactions[57,58]. Homotypic interactions are ones between molecules of the same kind, for example a disordered region of a specific protein interacting with neighbouring IDRs of the same protein. The term heterotypic interaction describes interactions between different molecules, for example different proteins and nucleic acids[57,59]. Similarly, condensates containing only a single type of biomolecule are often referred to as homotypic condensates, whereas complex assemblies of multiple components are referred to as heterotypic condensates.

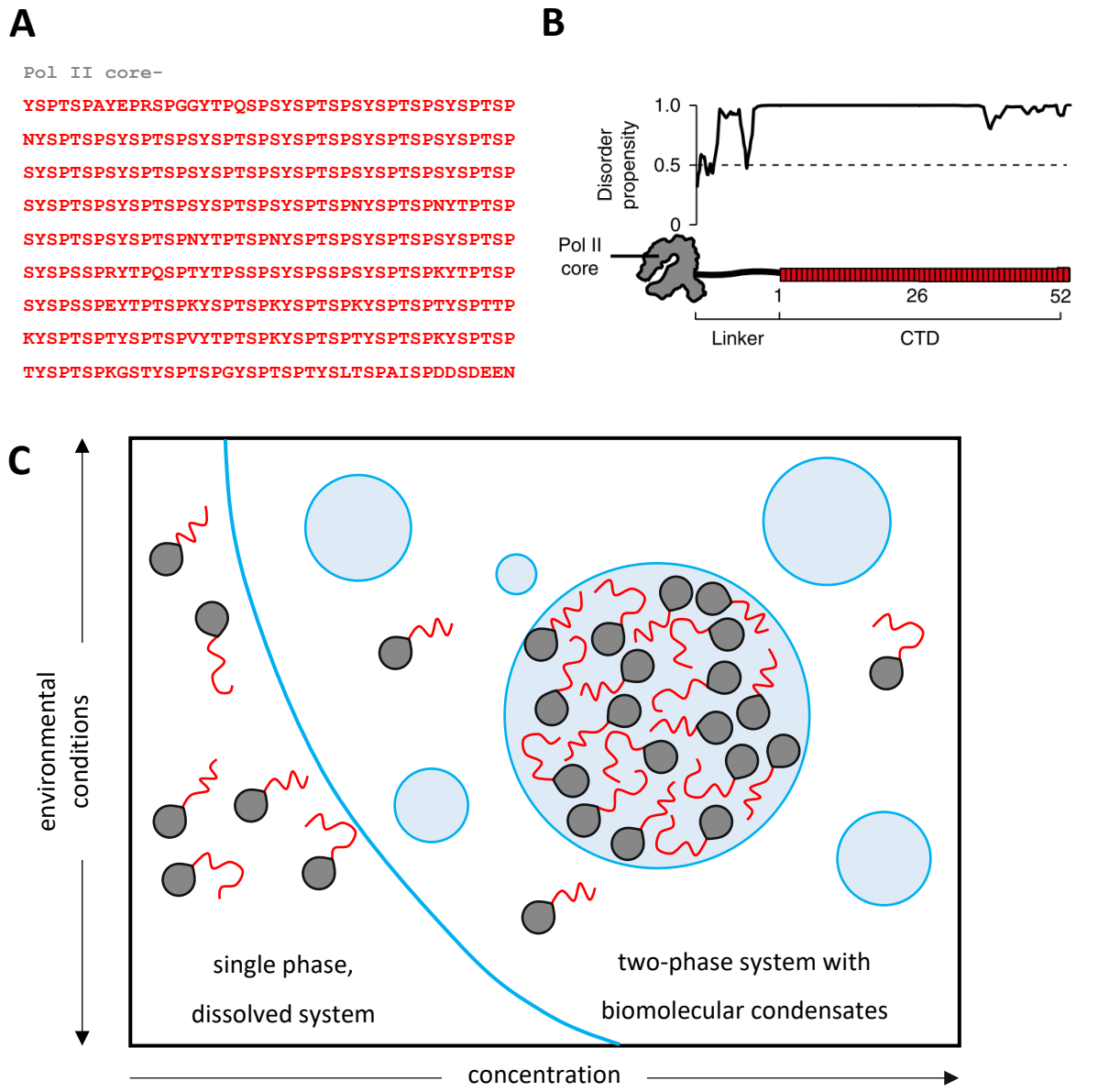


Fig. I2. Biomolecular condensation of intrinsically disordered protein regions. (A) The *Hs*RPB1-CTD, a striking example of a low complexity sequence. The almost 400 residue-long sequence contains repeating units of the amino acid heptad "YSPTSPS" and an almost complete absence of hydrophobic and charged residues. (B) Low amino acid sequence complexity results in a high propensity to form IDRs. The disordered CTD extends away from the globular core of RNAPII. (C) Intrinsically disordered regions drive the formation of biomolecular condensates in a protein concentration-dependent manner. At low concentrations, IDR-containing proteins may exist in a single phase, dissolved system but at higher concentrations, multivalent interactions between IDRs drive the formation of high-density molecular assemblies. In addition to protein concentration, environmental factors such as ionic strength and pH can promote or inhibit the formation of condensates. *Adapted from Boehning et al., 2018* [46] and *Franzmann and Alberti, 2019* [49].

In addition to protein concentration, IDR-driven biomolecular condensation is heavily influenced and regulated by environmental conditions[52]. For example, different temperatures can promote or inhibit condensation based on the sequence composition of protein IDRs[52]. Important for the cellular regulation of condensates, salt concentration and pH can greatly affect the condensation propensity of proteins[60]. Salt concentration can modulate multivalent interactions by shielding ionic interactions between charged IDR residues[48]. In addition to environmental factors, protein modification plays a large role in the regulation of condensates[61]. For example, negative charges introduced through phosphorylation can alter the charge balance of a system, and consequently promote or inhibit condensation[43,61]. Additionally, some phase separating proteins may only form condensates in the presence of RNAs[50]. Stress granules, for example, are a type of MLO that reversibly form in the presence of RNA and in response to stress stimuli[42,50].

It has been proposed that all proteins may be able to undergo condensation-like behaviours, given the right protein concentration and set of environmental conditions[62]. To differentiate between different types of condensate-forming molecules, a useful distinction between "drivers" or "scaffolds" and "clients" has been proposed[63]. Drivers of biomolecular condensation are molecules that readily phase separate into condensates, at low concentrations and under a range of environmental conditions[64], whereas client molecules do not readily phase separate on their own but are recruited into and enriched in complex mixed condensates, such as MLOs[65].

Biomolecular condensation in eukaryotic transcription regulation

Recent evidence has suggested that the multivalent interactions driving biomolecular condensation may play a key role in the regulation of transcription[37].

Some of the first evidence that biomolecular condensation may play a role in the organisation and regulation of transcription came from the observation that RNAPII, Mediator and the co-activator BRD4 form clusters in the nuclei of mouse embryonic stem cells[66]. These clusters exhibit dynamic properties as they form in a reversible manner, coalesce into larger assemblies and recover from photobleaching, typical of condensates with liquid-like properties[67]. It was further shown that clusters of RNAPII and Mediator co-localise and that RNAPII clusters dissolve when phosphorylated by CDK9[66]. CDK9 itself was also shown to cluster in foci overlapping with the sites of nascent RNA transcription[68].

These results were subsequently supported by the observation that the IDRs of MED1, the largest Mediator subunit, and BRD4 form assemblies *in vitro*, which show hallmarks of LLPS[69]. These assemblies form in the presence of biomolecular crowders and their size scales in a concentration-dependent manner[69]. Further work by the same group demonstrated that intrinsically disordered activation domains from transcription activators such as OCT4 and the estrogen receptor form

condensates with Mediator and that the residues important for gene activation are also required for biomolecular condensation[70]. Mediator condensates were further shown to recruit and concentrate other key transcription factors, including regulators of the WNT, TGF-ß and JAK/STAT pathways[71].

The so-called activation domains of transcription factors have long been known to be intrinsically disordered, but the mechanism of action of these labile regions had until recently remained elusive[72]. Investigations into the IDRs of transcription activators in live cells have revealed that they form condensate hubs at sites of active transcription[72]. The TFIID component TAF14 has also been proposed to function as a nuclear hub, forming clusters that interact with and recruit a wide array of transcriptional proteins to transcription foci[73].

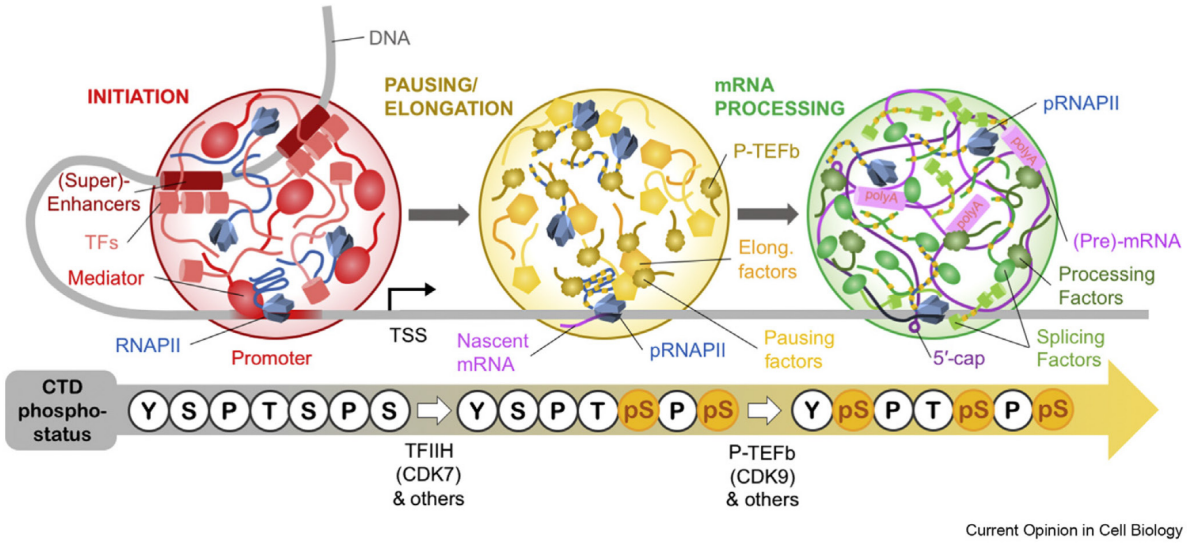


Fig. I3. A model for the phosphorylation dependent regulation of transcriptional condensates. RNAPII, Mediator and transcription factors form phase-separated condensates, through the action of multivalent interactions between their IDRs, at gene promoters and distal regulatory sequences. Phosphorylation of the RNAPII-CTD alters the condensate preference of RNAPII and releases it into active transcription. Further phosphorylation promotes RNAPII separation into condensates enriched in co-transcriptional processing and splicing factors. Abbreviations: RNAPII, unphosphorylated RNA polymerase II; pRNAPII, CTD-phosphorylated RNAPII; TFs, transcription factors; TFIIH, transcription factor II H; CDK, cyclin-dependent kinase; P-TEFb, positive transcription elongating factor; TSS, transcription start site; CTD, RNAPII C-terminal domain. Adapted from *Nosella and Forman-Kay, 2021* [43].

Other work demonstrated that the CTD of human and yeast RPB1, the largest subunit of RNAPII, is capable of driving concentration-dependent condensation *in vitro*[46]. It was observed that the length of the CTD influenced RNAPII cluster dynamics in cells, with the 26 heptad-repeat of the *Saccharomyces cerevisiae* CTD forming smaller, more fluid condensates and the 52 heptad-repeat *H. sapiens* CTD generating larger, more stable clusters[46]. CTD phosphorylation by the TFIIH subunit CDK7, which coincides with RNAPII release and transcription initiation *in vivo*, leads to the dissolution of CTD condensates *in vitro*[46]. Further transcription regulatory proteins that interact with and modify the CTD have also been shown to form biomolecular condensates and to reversibly associate with CTD clusters. Cyclin T1 of the PTEFb complex, which together with CDK9 has been demonstrated to phosphorylate the CTD and release RNAPII into transcription elongation, forms phase-separated condensates through the action of a disordered histidine-rich domain[74]. These cyclin T1 condensates concentrate RPB1-CTD and form larger assemblies when the CTD is pre-phosphorylated with CDK7. Unphosphorylated RPB1-CTD also preferentially localises to Mediator condensates *in vitro*. Importantly, CTD phosphorylation diminishes its association with Mediator and, in turn, promotes association with condensate-forming splicing factors, such as SRSF1/2 and U2AF2[75]. Thus, the CTD phosphorylation status appears to regulate preferential recruitment of RNAPII into specialised transcription condensates promoting either transcription initiation or transcription elongation coupled to co-transcriptional RNA processing[46,74,75].

Finally, very recent work showed that the activity of transcriptional condensates at enhancer sites is regulated by RNA concentration via a feedback mechanism. At low concentrations, RNA stimulates the formation of transcription condensates, but inhibits condensate formation when RNA levels surpass a certain threshold[76].

These observations can be combined into a model in which unphosphorylated CTD is incorporated into Mediator condensates that promote PIC assembly and transcription initiation (Fig. I3). The formation of these condensates may be further promoted and stabilised by the action of various transcription activators, influencing condensation through the action of their intrinsically disordered activation domains. Subsequent phosphorylation of the CTD by CDK7 promotes association with PTEFb condensates and further phosphorylation by the PTEFb kinase subunit, CDK9 (Fig. I3). CTD phosphorylation by CDK9 eventually leads to RNAPII release from the promoter and preferential localisation of the CTD to nuclear speckles, MLOs, which concentrate splicing factors responsible for co-transcriptional RNA processing[37,43,77].

While the data support an important role of biomolecular condensation in the regulation of transcription, many questions remain outstanding[78–80]. The role of GTFs in this emerging model of transcription regulation is unclear, as investigations have focused on RNAPII, Mediator and specific transcription activators[53]. Furthermore, the majority of work in this field has focused on human and yeast proteins[46,73]. Whether the role of biomolecular condensation is a general feature of eukaryotic transcription or a derived feature in higher eukaryotes, remains an important question that needs to be further investigated. Relevant to this study, the role of biomolecular condensates in non-model eukaryotes like *P. falciparum* remains largely unexplored. Given the abundance of LCRs in the parasite's proteome[9], IDR-mediated biomolecular condensation may play a significant role in its biology.

Aims of this study

This study seeks to supplement our current understanding of TF-IDR-driven biomolecular condensation by investigating a panel of model fusion proteins by fluorescent microscopy. The first goal was to further investigate whether and under which conditions GTF IDRs from both *H. sapiens* and *P. falciparum* drive biomolecular condensation *in vitro*, including the CTD of *P. falciparum* RPB1, the well characterised IDR from *H. sapiens* MED1 as well as previously generated *H. sapiens* and *P. falciparum* GTF-IDRs[81]. The primary objectives were to understand if IDR-driven condensation is a feature of transcription factors in non-model eukaryotes, like *P. falciparum,* and how IDRs from general, TFII-type transcription factors may be involved in the formation of transcriptional condensates. Following characterisation of individual protein regions, further studies investigated interactions between various TF-IDR regions. We sought to classify IDRs as drivers or clients of biomolecular condensates and to examine the compatibility of different proteins in forming mixed condensates. Of particular interest here was testing for compatibility between human protein condensates and *P. falciparum* components to identify *P. falciparum*-specific features that could aid the development of novel antimalarial drugs.

Materials and Methods

Multiple sequence alignments and disorder predictions

Amino acid sequences of proteins of interest were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/) or PlasmoDB (https://plasmodb.org/plasmo/app) where applicable. A full list of species is detailed in Figs. S1, S3-4. Multiple sequence alignments were performed using the MUSCLE algorithm and displayed in UGene[82,83]. Protein sequence information was extracted using ProtParam on the ExPASy Server (https://web.expasy.org/protparam/). Protein disorder predictions were performed using the Predictor of Natural Disordered Regions (PONDR) server and the VSL2 setting[84].

Generation of transcription factor-IDR fluorescent protein fusion constructs

Three new protein expression vectors were created for the purposes of this study, including pET11d-6His-mC-*Pf*RPB1-CTD, pET11d-6His-MED1-IDR-YFP and pET11d-6His-mC-VP16-AD. cDNA coding for fusion proteins was generated by PCR-driven overlap extension[85] (see Supplementary materials, Table 2 for primer sequences) and resultant fragments inserted into pET11d expression vectors. The *Pf*RPB1-CTD sequence was amplified from a *P. falciparum* unsynchronous cDNA library[86]. The MED1-IDR sequence was amplified from a human gDNA library. The VP16-AD sequence was amplified from a vector containing the GAL4-VP16-AD fusion protein sequence. Fluorescent protein sequences were similarly cloned from vector templates of already cloned fusion proteins (pET11d-6His-*Hs*TBP-IDR-YFP and pET11d-6His-*Pf*TFIIB-IDR-mCherry, Leonidas Karamanof and Jasmin Knopp, Oelgeschläger lab). All plasmid constructs containing the coding sequences for the various fusion proteins were verified by sequencing.

pET11d expression vectors for the *Pf*- and *Hs*TBP-IDR-YFP, *Pf*- and *Hs*TFIIAαβ-IDR-GFP and *Pf*- and *Hs*TFIIFß-IDR-CFP fusion proteins as well as a YFP control were generated in a previous study in a similar fashion[81].

Expression vectors with a pETEC backbone for the mEGFP-MED1-IDR, mCherry-MED1-IDR and mEGFP-*Hs*RPB1-CTD fusion proteins were shared by the laboratory of Prof Richard Young[69,75].

Expression of transcription factor-IDR fluorescent protein fusion constructs in *E. coli*

Proteins were expressed in BL21-CodonPlus (DE3)-RIL cells (Agilent Technologies). A single transformed colony was used to inoculate 100 mL overnight lysogeny broth (LB) starter cultures supplemented with selective antibiotics (34 µg/mL Cam and 100µg/mL Amp for pET11d constructs OR

50 µg/µL Kan for pETEC constructs). Starter cultures were grown at 37 °C with shaking for 16 h. Cells were collected by centrifugation (3000 x g for 10 min at 25 °C) and resuspended in fresh media. 50 mL starter culture was used to inoculate 1 L LB (34 µg/mL Cam and 100µg/mL Amp OR 50 µg/µL Kan), divided into 500 mL across two 2 L flasks. Expression cultures were grown for 1.5 hours with shaking at 16 °C prior to inducing expression with the addition of 1 mM IPTG. Cultures were returned to a shaking incubator at 16 °C for 22 h. Cells were then pelleted by centrifugation (3800 x g for 15 min at 4 °C) and used directly for protein purification.

For expression of mC-VP16-AD, 10 mL of overnight culture was diluted in 1L LB pre warmed to 37 °C (34 µg/mL Cam and 100µg/mL Amp) and grown at 37 °C with shaking for 2 h. Protein expression was induced with the addition of 1 mM IPTG and cultures were returned to shaking at 37 °C for 3 h. Cells were then pelleted by centrifugation (3800 x g for 15 min at 4 °C) and stored overnight at 4 °C prior to protein purification.

Nickel-affinity purification of recombinant fusion protein

Cells were resuspended in 5 mL per g wet cell pellet in sonication buffer (50 mM Tris-Cl pH 7.2; 500 mM NaCl; 10 mM Imidazole pH 8; 0.1% [v/v] NP-40; 3.3 mM β-mercaptoethanol, 1 mM phenylmethylsulphonyl fluoride (PMSF)). Following 15x 15 s rounds of sonication on ice at power 10 (MSE Soniprep 150) with 60 s cooling between cycles, cell lysates were incubated with 250 U benzonase (Sigma-Aldrich, E1014). Cleared lysates were generated by centrifugation of the cell slurry at 26,000 x g at 4 ˚C for 35 min. All subsequent purification steps were performed at 4 ˚C.

Cleared lysates were added to 1 mL of nickel-affinity chromatography resins (Protino Ni-NTA Agarose, 745400) pre-equilibrated with 10X packed column volume (PCV) sonication buffer and rotated for 1.5 h at 4 ˚C. The resin with bound protein was separated from unbound material by centrifugation at 1000 x g and transferred to a disposable column (Poly-Prep, BioRad). Resins were washed with 5X PCV wash buffer (Tris-Cl pH7.9@4˚C; 20 mM imidazole pH8; 0.1% [v/v] NP40; 3.3 mM β-mercaptoethanol) containing either 100 mM KCl in the case of mC-PfRPB1-CTD and mC-VP16-AD or 300 mM KCl in the case of MED1-IDR-YFP, mEGFP- and mC-MED1-IDR and mEGFP-HsRPB1-CTD. Proteins were eluted as 1 mL fractions in BC-buffer (20 mM Tris-Cl pH7.9 @4 ˚C; 0.2 mM EDTA pH 8.0; 20% [v/v] glycerol; 0.1% [v/v] NP-40; 3.3mM β-mercaptoethanol) containing 100 mM or 300 mM KCl and 250 mM imidazole pH8, with the KCl concentration of the elution buffer matching that used in the wash step. Fractions containing the highest concentrations of protein, judged by the brightness of the fluorescent sample (generally elution two and three), were aliquoted, snap frozen in liquid nitrogen and stored at -80 °C.

mC-VP16-AD was purified in a similar fashion, except that the cleared lysate was added directly to a column with pre-packed and pre-equilibrated resin and purified by gravity flow.

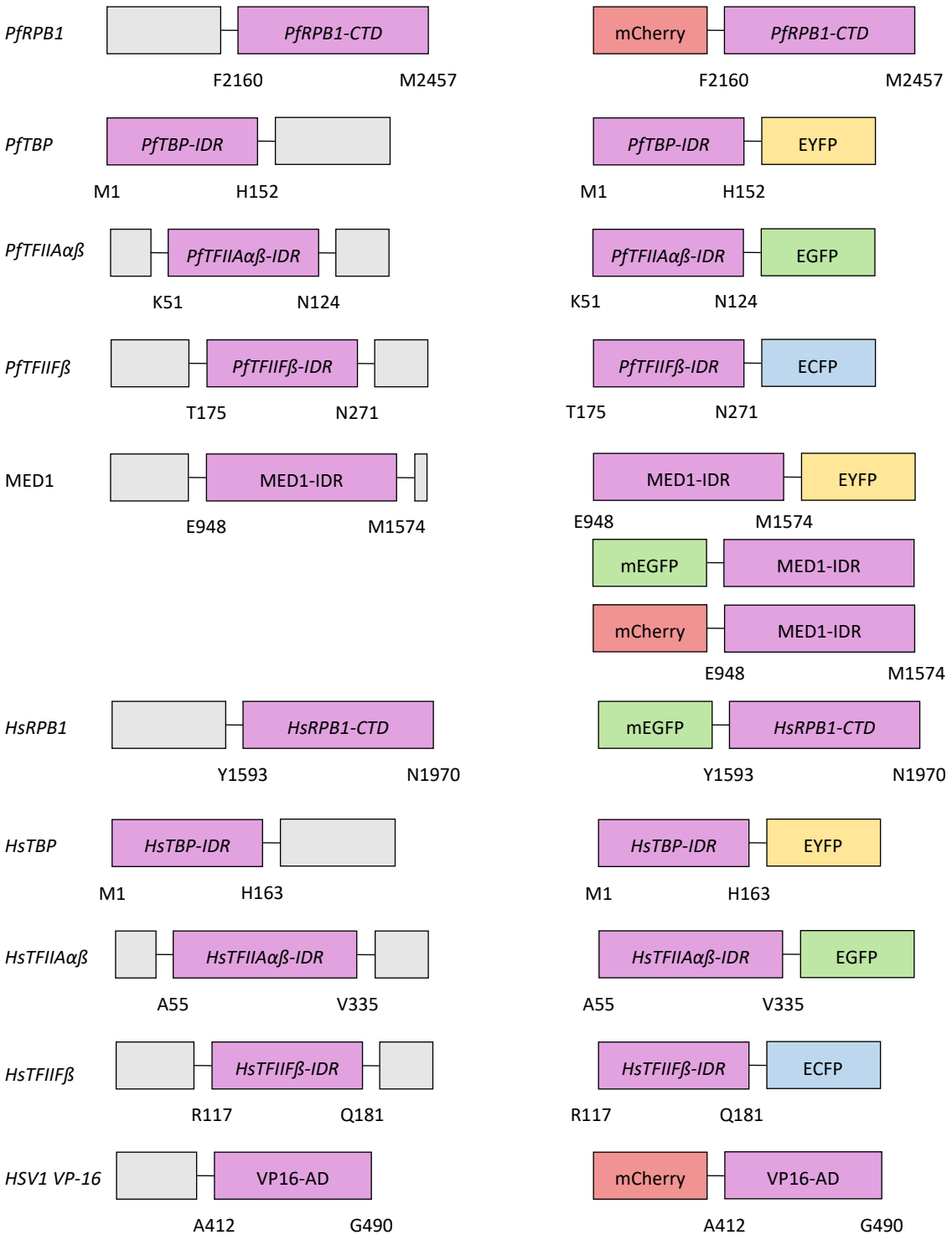


Fig. M1. *P. falciparum* and *H. sapiens* transcription factor-IDR fluorescent protein fusion constructs. *Left:* Depiction of full-length transcription factors, structured regions in grey and IDRs in pink with amino acid positions given underneath; *right:* depiction of TF-IDR fusion proteins, which indicate the type of fluorescent protein and position relative to IDR. Abbreviations: *Pf*, *Plasmodium falciparum*; *Hs*, *Homo sapiens*; RPB1, RNA polymerase II subunit 1; TBP, TATA-binding protein; TFIIAαß, transcription factor IIA αß-subunit; TFIIFß, transcription factor IIF ß-subunit; MED1, Mediator subunit 1; IDR, intrinsically disordered region; EYFP/EGFP/ECFP, enhanced yellow/green/cyan fluorescent protein; m, monomeric; VP, viral protein; AD, activation domain.

Purity and integrity of protein preparations was assessed by SDS-polyacrylamide gel electrophoresis (PAGE), followed by visualisation of protein bands by Coomassie staining[87].

Total protein concentration of samples was determined using Bradford reagent (BioRad) by measuring absorbance at 590nm of a range of dilutions against standards of BSA in BC-100 buffer.

Pf- and *Hs*TBP-IDR-YFP, *Pf*- and *Hs*TFIIAαβ-IDR-GFP and *Pf*- and *Hs*TFIIFß-IDR-CFP fusion proteins as well as a YFP-control were previously expressed and purified in a similar fashion to mC-VP16-AD and stored as aliquots at -80 °C[81]. A notable difference was the addition of 5 µL/g protease inhibitor cocktail (Sigma-Aldrich, P8849) to cell pellets prior to sonication, rather than 1 mM PMSF.

General investigation of biomolecular condensation by widefield fluorescence microscopy

Purified protein preparations, stored at -80 °C in either BC-100 or BC-300 buffer, were thawed at room temperature and centrifuged (17000 x g for 5 min at room temperature) to clear any accumulated aggregates. Samples were transferred to fresh Eppendorf tubes and diluted to appropriate concentrations with their respective storage buffers. Protein dilutions were then mixed 1:1 with biomolecular condensation buffer (20 mM Tris-Cl pH 7.9 @ 4 °C; 1 mM DTT; 0-1000 mM KCl; 0-40% [v/v] polyethylene glycol (PEG) 20,000 or 3000) to induce the formation of condensates. The majority of reactions took place at final conditions of 20 mM Tris-Cl, 0.5 mM DTT, 150 mM KCl and 12% PEG-3000 (Sigma-Aldrich, P1458) but where indicated, conditions were adjusted to encourage the formation of condensates or to test alternative conditions. Additional components, such as 1,6-hexanediol or poly(dG-dC) dsDNA, were incorporated into the final reaction by including them at twofold of the desired final concentration in modified condensation buffer.

Samples of condensation reactions were visualised by fluorescence microscopy, generally using a Nikon Ti-E inverted widefield fluorescence microscope or, as indicated, a Zeiss LSM 880 confocal fluorescent microscope. Glass chambers for containing droplet reactions were created by pre-treating glass slides and coverslips with silanization solution (Sigmacote, Sigma-Aldrich) and washing them with RoH_2O. Double-sided adhesive tape was fixed to glass slides to hold coverslips in place. 2–5 µL samples of condensation reactions were transferred to slides and covered with a coverslip. The slide-sample-coverslip sandwich was then inverted and incubated for 30 min prior to imaging.

Imaging was managed using macro scripts written in Nikon NIS-Elements AR software. Fusion proteins were visualised under appropriate filter sets for CFP (425-445 nm Ex, 455-500 nm Em); GFP (460-500 nm Ex, 505-560 nm Em); YFP (490-510 nm Ex, 520-550 nm) and mCherry (540-590 nm Ex, 550-650 nm

Em) fluorescence. Imaging generally proceeded by focussing on the layer of droplets settled on the coverslip. Ten images of the same field of view were taken in rapid succession after which the slide stage was moved a fixed distance along a single axis to an independent field of view where the process was repeated. Where possible, fluorescence exposure was kept constant across conditions for a single protein but had to be increased in certain cases where low protein concentrations resulted in very small droplets and low fluorescence readings. To capture dynamic events, like the fusion of two droplets, reactions were immediately placed under the microscope and a field of view was chosen. The region was recorded for around a minute, taking images at the highest rate allowed by the chosen exposure.

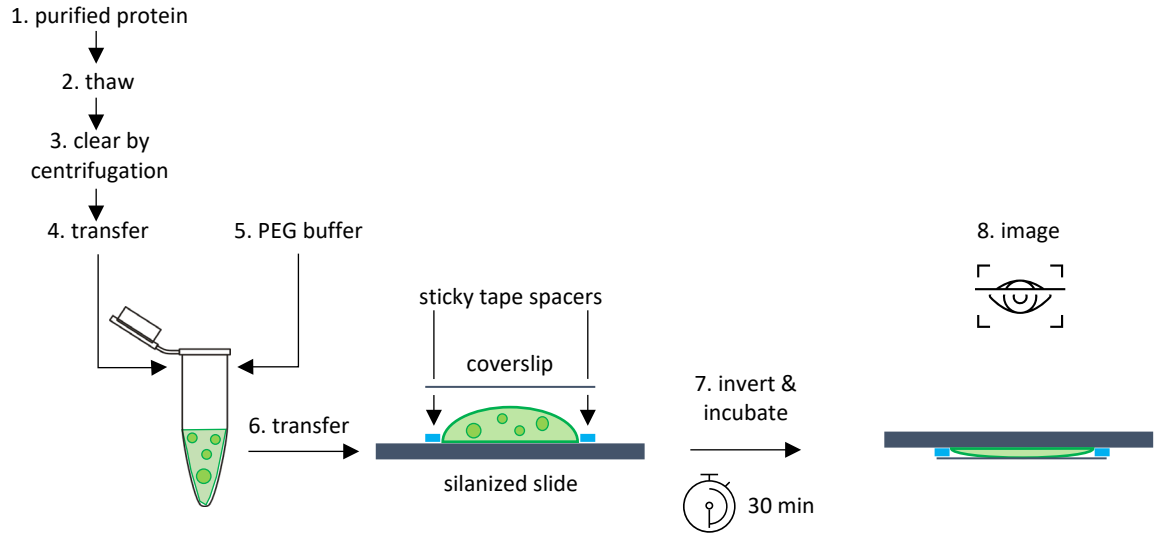


Fig. M2. Finalised workflow for microscopy analysis of fluorescent protein-IDR fusion constructs. Protein preparations are stored at -80 °C, thawed at room temperature and cleared by centrifugation at maximum speed. Proteins are transferred to a fresh tube and combined with a condensation reaction buffer containing the molecular crowder PEG and an appropriate KCl concentration. A sample of the protein and PEG mixture is then transferred to a silanized microscopy slide with pre-mounted double-sided sticky tape strips. A silanized coverslip is placed on top of the sticky tape strips. This "chamber" is inverted and incubated for 30 min, to allow the condensation processes to take place and droplets to settle onto the surface of the coverslip. Following incubation, the imaging chamber is transferred to an inverted fluorescent microscope and imaged with the appropriate filter set(s) for the fluorescent protein(s) of the fusion constructs under investigation.

Salt reversibility assays

To investigate the reversibility of condensates, an assay was adapted based on published methods[69]. Droplets were formed in condensation buffer containing a permissive KCl concentration (e.g. 150 mM KCl). A fraction of this initial reaction was then diluted in a separate tube either with the same volume (1:2) of an identical condensation buffer, thereby halving the concentration of protein in solution but keeping the concentration of all other reaction components constant, or diluted with the same volume (1:2) of a buffer containing a higher concentration of KCl, appropriate to achieve the desired target concentration (e.g. 500 mM). Dilutions were performed immediately after initiating condensation (<1 min) or after 30 min, as indicated. Dilutions, along with the initial reaction, were visualised by widefield fluorescence microscopy after allowing droplets to settle onto the coverslips.

Confocal microscopy and fluorescence recovery after photobleaching assays

To investigate the internal diffusion dynamics of condensates, fluorescence recovery after photobleaching (FRAP) assays[88], coupled to confocal fluorescence microscopy, were employed. Condensation reactions were transferred to imaging chambers and visualised immediately. A field of view with two large (5–10 µm) droplets in proximity was first identified, with one droplet to serve as the target of bleaching and the other to control for background bleaching. The field of view was monitored for 30 s to ensure stable fluorescence signals. Following this pre-bleaching period, the selected droplet was bleached with 100% laser power to 10% of its initial fluorescence brightness. The field of view was monitored for a total of 5 min to capture droplet recovery. For FRAP assays containing two differentially labelled protein components, bleaching was performed on both laser channels simultaneously.

Two-component mixing assays

To interrogate putative interactions between GTF-IDR fusion proteins, mixed condensation assays with two different proteins were undertaken. Mixing experiments generally proceeded in one of two ways. In the “pre-mix” scenario, protein preparations were first combined and mixed before addition of condensation buffer. Alternatively, in the “pre-form” scenario, condensation of each protein component was first induced separately, and the reactions were combined.

Microscopy image analysis

The analysis of fluorescence microscopy images was achieved using custom scripts written in Python3, based on scripts shared by Henninger et al.[76]. Annotated scripts for image analysis and data display are freely available at https://github.com/leetheflee/droplet-analysis. Scripts use the scikit-image[89] and SciPy[90] packages for image processing, the NumPy[91] package for mathematical operations and the Matplotlib[92] package for the display of data.

A Python3 script was written to merge the brightness information of multiple images taken in rapid succession at the same location. This was done to minimise noise present in background, non-droplet regions of images. Briefly, the script loads a series of images, converting each image into an array of brightness values between 0 and 1. The brightness value at each pixel coordinate is then summed across the series of images and divided by the number of images used in the series. The output is a single image, representing the average brightness of a field of view across a short time period (generally ten images in 1 s).

A further Python3 script was written to identify droplets from images processed as described above. Briefly, the script works by loading images as arrays of pixel brightness values. Given that protein condensates contain a higher concentration of fluorescent protein, than the surrounding dilute phase, they are much brighter objects than the surrounding solution. Pixels residing in putative droplet regions are identified as pixels more than two times the standard deviation brighter than the mean brightness of an image. This value is lower than the commonly used value of three times the standard deviation used in the analysis of images generated by confocal microscopy[76], to account for the increased background brightness in widefield microscopy. Clusters of pixels which supersede this brightness threshold are then further screened to determine condensate areas. A minimum size threshold is used to reject any clusters consisting of less than five pixels and a circularity filter can be used to reject non-circular objects, where circularity is defined as $4\pi*(\text{area})/(\text{perimeter}^2)$. A graphical representation of this process is depicted in Fig. M3.

The properties of individual condensates, including area and brightness are then extracted and displayed as a function of the experimentally adjusted variable, such as protein, KCl or 1,6-hexanediol concentration. Alternatively, data figures can be generated that present the average condensed fraction of an image, which is defined as the total area of condensate droplets as a fraction of the total image area.

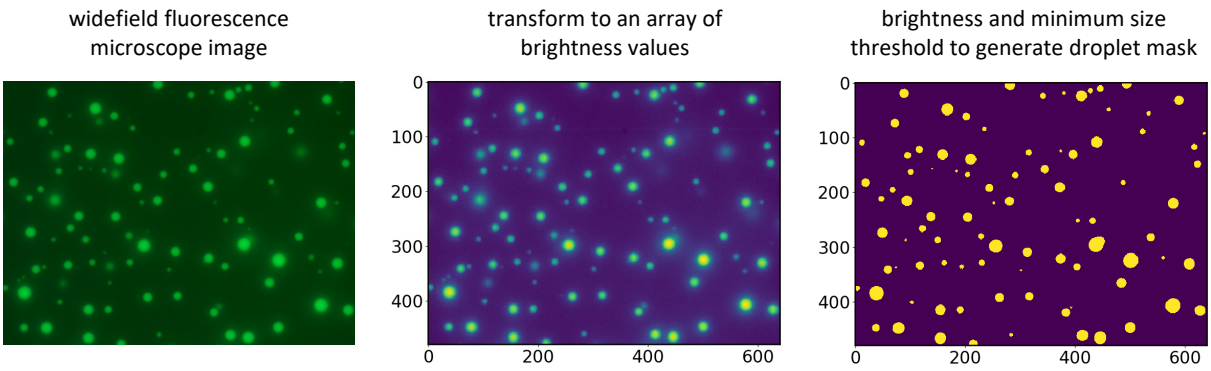


Fig. M3. The droplet analysis script identifies condensates based on fluorescence brightness values. Scripts for the analysis and quantification of widefield fluorescence microscopy images were written in Python3. Script inputs are widefield fluorescence microscopy images (*left*), always in replicate, which are read by the script as an array of brightness values (*middle*). Each pixel coordinate is assigned a brightness value between 1 and 0. Droplets are initially identified as pixels with an average brightness greater than two standard deviations above the mean of the image. Further minimum size thresholds (usually 5 pixels) and optional circularity filters are applied to generate the final droplet mask (*right*). Using the droplet mask, droplet features can subsequently be extracted, including size, number and brightness of droplets, as well as the condensed fraction of a field.

Statistical analysis of microscope images

Condensation reactions were imaged in at least ten independent fields of view. In data figures, where individual droplet data are depicted, each data point represents a single condensate observed within these ten images. The median, upper and lower quartiles are superimposed as boxplots to show the distribution of droplet properties. Measurements of the condensed fraction are given as the mean ± standard deviation for each condition.

FRAP bleaching experiments were repeated three times at the same condition. Recovery curves display the mean ± standard deviation across these three replicates.

Standardising and quantifying images generated for mixed condensation reactions containing two different fluorescent protein constructs was beyond the scope of this study. In this instance, images depicted in data figures give a representative view of condensate properties across individual fields of view.

Spin-down assays

A spin-down assay protocol was developed based on previously published methods[69] in order to assay for recruitment of HeLa nuclear extract (HeLa NE) proteins into condensates formed by IDRs fused to fluorescent protein. IDR-fluorescent protein or an identical concentration of the fluorescent protein control without an IDR were combined with diluted HeLa NE[93] and mixed. Reactions were incubated at room temperature for ten min to allow for condensate formation, followed by centrifugation (17000 x g for 10 min at 21 °C) to pellet condensates. Supernatants were carefully removed and transferred to fresh Eppendorf tubes and the presence of specific proteins present in pellets and supernatants determined by SDS-PAGE and immunoblotting.

Following SDS-PAGE, protein bands were transferred to a Millipore Immobilon PVDF membrane using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) according to manufacturer's instructions. Following transfer, the membrane was washed three times in TBST buffer (150 mM NaCl; 10 mM Tris pH 8; 0.05 % [v/v] Tween 20) and then incubated in blocking buffer (5% [w/v] skim milk powder in TBST) for 1 h. Following blocking, membrane was cut into strips covering specific molecular mass ranges and individual strips probed with specific antibodies overnight at 4 °C.

Membranes were washed three times in TBST and incubated with secondary detection antibodies, either anti-mouse IgG-peroxidase conjugate (Sigma-Aldrich) or anti-rabbit IgG-peroxidase conjugate (Sigma-Aldrich) diluted 1:3000 in TBST. Immunoreactive protein bands were detected by chemiluminescence (WesternBright Sirius kit Advansta) and visualised on X-ray film. *Hs*RPB1 was detected using mouse raw ascites fluid 8WG16 (Covance) diluted 1:3000 in blocking buffer. *Hs*TBP was detected with a rabbit α-TBP antibody, a kind gift from Robert G. Roeder (The Rockefeller University, USA) to Thomas Oelgeschläger.

Results

The *Pf*RPB1-CTD has a unique sequence structure and patterning compared to other eukaryotes

To uncover potential unique features of *Pf*RPB1, its protein sequence was aligned to homologs from a diverse range of eukaryotic species, including *H. sapiens, A. thaliana, C. elegans* and *S. cerevisiae* (Fig. S1). All alignments were performed in UGENE using the MUSCLE algorithm[82,83]. Comparing the alignment over the complete length of RPB1, it is notable that the *Pf*RPB1 protein is around 500 amino acid residues larger than its human counterpart, despite having a shorter CTD. This is explained by the presence of three large *P. falciparum*-specific insertions in the structured, catalytic domain of the protein (Fig. 1A). These regions are highly enriched in asparagine as well as aspartic- and glutamic acid and lysine (Fig. S2). Insertions with this sequence makeup are not uncommon in *P. falciparum* and are often attributed to its highly AT-rich genome[9,12]. It is still notable that these large, predicted IDRs regions are tolerated in a protein as central and conserved in transcription as RNAPII. The *Hs*RPB1-CTD is predicted to be highly disordered, likely due its extreme sequence composition bias (Figs. 1A, S1–2). The *Pf*RPB1-CTD, containing a more diverse set of residues, has a slightly less distinct disorder prediction, but is nonetheless also classified as disordered, starting around residue 2100 (Fig. 1A). These high confidence disorder predictions are an indication that these regions might engage in the weak, multivalent interactions necessary to drive biomolecular condensation[62].

Closer investigation of the *Pf*RPB1-CTD reveals a unique structure and amino acid sequence patterning. RPB1-CTDs from higher eukaryotes adhere quite strictly to a canonical heptad repeat structure, consisting of variable numbers of the repeating sequence YSPTSPS (Fig. S1). The N-terminal portion of the *Pf*RPB1-CTD is characterised by many asparagine and glutamine insertions, whilst still containing multiple SP-motif repeats (Fig. 1B). The “SP-motif” – a serine residue followed by a proline – appears twice in the canonical heptad sequence, and its repetition throughout the *Pf*RPB1-CTD hints at shared ancestry and functional conservation with other eukaryotic CTDs. The middle section of the *Pf*RPB1-CTD stands out due to its striking adherence to the canonical heptad repeat sequence, bar the consistent substitution of Ser_7 with a lysine residue. Ser_7 is notably a phosphor-acceptor residue in RPB1-CTD phosphorylation[31].

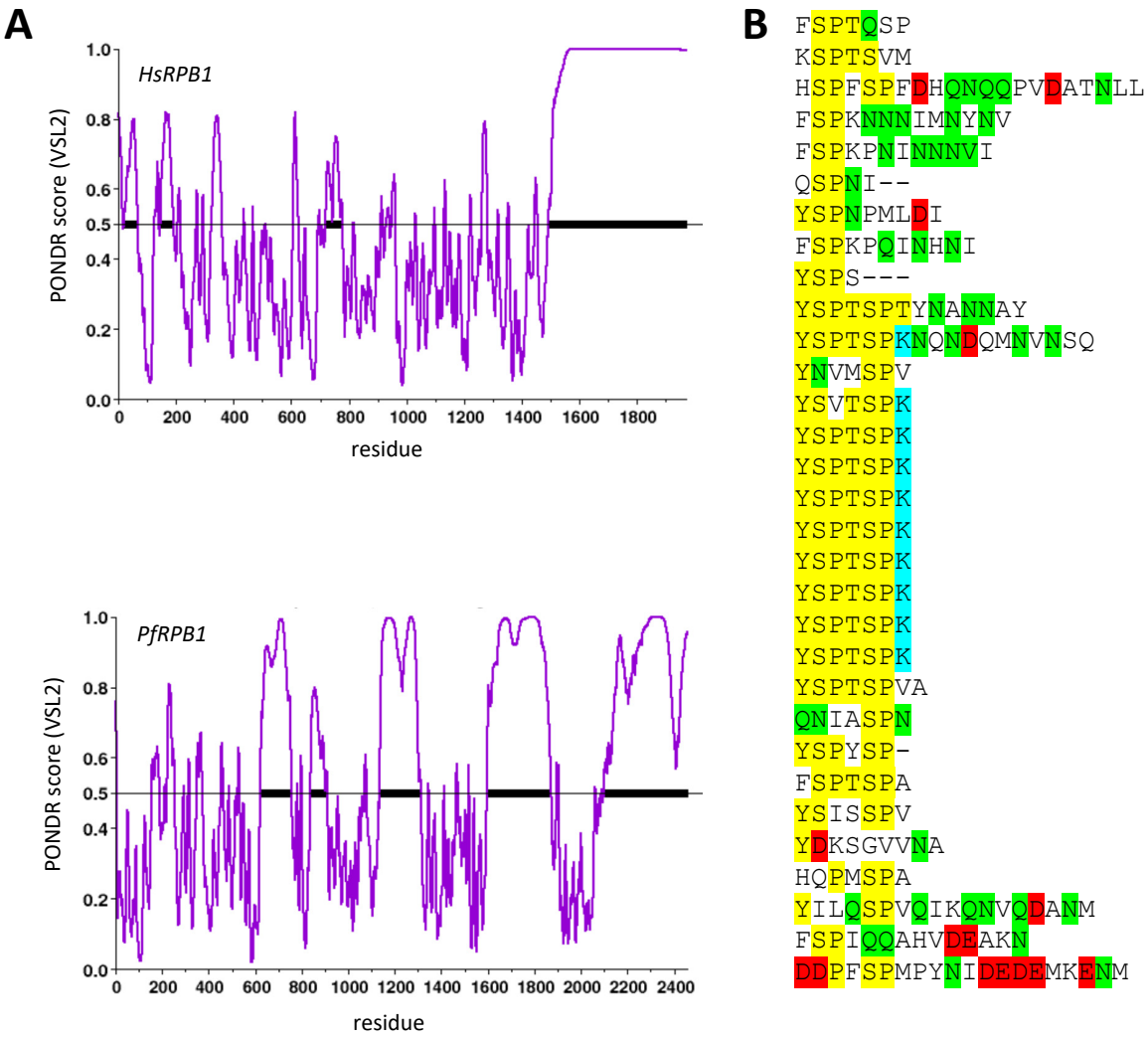


Fig. 1. *P. falciparum* RPB1 possesses unique sequence features and does not contain a canonical eukaryotic CTD. (**A**) Predictor of natural intrinsic disorder (PONDR) score for *H. sapiens* and *P. falciparum* RPB1. *Pf*RPB1 contains three large intrinsically disordered insertions in its structured core domain, not present in *Hs*RPB1. The CTDs of both proteins are predicted to be highly disordered. A PONDR score of >0.5 is a commonly used benchmark for likely IDRs and significantly long regions that exceed this threshold are highlighted by black bars. (**B**) The *Pf*RPB1-CTD as determined by multiple sequence alignment and protein domain prediction. N- and C-terminal portions are characterised by a loss of heptad repeat structure and insertions of D, E, Q and N residues. A central region displays strong resemblance to the canonical eukaryotic heptad motif YSPTSPS, with Ser_7 substituted by K. Conservation of heptad structure highlighted in *yellow*, residues substituted by K marked in *cyan,* negatively charged residues coloured *red* and N/Q in *green*.

This central region contains a block of nine repetitions of the heptad sequence YSPTSPK. The extreme C-terminus is in contrast characterised by an almost complete absence of heptad repeat structure. This region is also marked by the presence of asparagine and glutamine insertions as well as negatively charged aspartic and glutamic acid residues (Fig. 1B). A breakdown of heptad sequence structure towards the extreme C-terminus of RPB1 is however common in eukaryotes (Fig. S1). In general, the *Pf*RPB1-CTD contains a more diverse set of amino acid residues compared to the homologous *H. sapiens* region (Fig. S2).

Following the alignment to RPB1-CTD sequences of diverse eukaryotes, the goal was to determine how the *Pf*RPB1-CTD compares to homologous proteins within the *Apicomplexan* clade as well as the *Plasmodium* genus. Alignment to other *Plasmodium* RPB1 sequences reveals that not all *Plasmodium* species possess a central, nine-heptad repeat block in their CTDs. It is notably absent in the species that infect rodents (*P. chabaudi, P. berghei, P. yoelii*) and birds (*P. gallinaceum, P. relictum*) as well as *P. ovale*, which all only contain three repeats of the heptad sequence YSPTSPK (Fig. S3). In contrast, the other human- and monkey-infecting strains included in this analysis all do possess the nine-heptad repeat block. This includes *P. vivax*, as well as *P. reichenowi, P. knowlesi* and *P. cynomolgi*. These species all share the Ser_7 to lysine substitution across most heptads (Fig. S3). N- and C-terminal regions of the RPB1-CTDs, which diverge from the canonical heptad repeat sequence, are quite well conserved across *Plasmodium* species (Fig. S3). SP-repeat motifs are present throughout the CTDs of RPB1 subunits of all *Plasmodium* species.

Multiple sequence alignment of *Pf*RPB1 with homologs from diverse *Apicomplexan* parasites reveals a much greater degree of divergence between species, with an almost complete loss of a recognisable heptad consensus sequence in many representatives (Fig. S4). CTD length also varies greatly, from 139 amino acids in *T. annulata* to 321 residues in *B. besnoiti.* Where repetitive heptads can be identified, they diverge from the typical eukaryotic consensus and substitution of Ser_7 is common (Fig. S3). In *B. bigemina*, the only other species with an appreciable number of heptad repeats, Ser_7 is generally replaced with an alanine residue. Despite the loss of clear heptad consensus repeats, SP-motifs are conserved throughout *Apicomplexan* RPB1-CTDs (Fig. S4).

Combining the results of the eukaryotic, *Apicomplexan* and *Plasmodium* alignments, as well as the disorder prediction, enabled the determination of the C-terminal end of the structured region and N-terminal end of the *Pf*RPB1-CTD. This was set at residue F2160, where the first hallmarks of a heptad repeat structure and SP-motifs become apparent.

Cloning, protein expression and purification of constructs

Following the bioinformatic sequence investigation of *Pf*RPB1, a suite of bacterial expression vectors was used to express the IDRs of several *P. falciparum* and *H. sapiens* GTFs fused to fluorescent proteins (Fig. M1). This included previously generated constructs containing IDRs from TBP, TFIIAαß and TFIIFß, from both *H. sapiens* and *P. falciparum* homologs[81] as well as newly generated constructs for the expression of *Pf*RPB1-CTD, MED1-IDR and VP16-AD (Fig. M1). Further constructs provided by the Young laboratory (MIT Whitehead Institute, Boston, USA) enabled the expression of MED1-IDR and *Hs*RPB1-CTD fusion proteins featured in published data[69,75].

Three new protein expression constructs were generated by PCR-driven overlap extension[85] and cloned into pET11d expression vectors with N-terminal 6his tags for purification. The first additional construct consists of the CTD of *Pf*RPB1, as determined through the sequence alignments described above, fused C-terminally to the fluorescent protein mCherry. This construct is henceforth referred to as mC-*Pf*RPB1-CTD. The next construct expresses the literature-defined activation domain of the VP16 protein from *Herpes Simplex Virus 1* (HSV-1), fused C-terminally to mCherry. The VP16 activation domain is one of the most powerful transactivators known[94] and was chosen as a representative activation domain to study in the context of condensate formation and transcription regulation. This construct is henceforth referred to as mC-VP16-AD. The orientation of these IDRs relative to the fluorescent proteins in these constructs was chosen to match the positions of the structured and intrinsically disordered domains in the native proteins (Fig. M1).

These IDR-fluorescent protein fusions represent entirely new protein constructs, whose bimolecular condensation capabilities have hitherto not been studied. A further construct was also generated, to act as a positive control – a transcription factor-IDR that has been previously characterised and demonstrated to readily form biomolecular condensates. The IDR of *H. sapiens* MED1, the largest subunit of the Mediator complex, was chosen and fused N-terminally to YFP. This construct is henceforth referred to as MED1-IDR-YFP. The Mediator complex has long been known to coordinate signals from transcriptional activators to the PIC at the proximal promoter[18], but more recent work has suggested that it may also regulate transcription through the action of its subunit IDRs driving biomolecular condensation[69,70]. It has been demonstrated to be a potent driver of biomolecular condensation and is a central player in the emerging hypothesis of condensate regulated transcription. The role of MED1-IDR condensation in transcription is also of particular interest, given the distinct lack of identified Mediator subunits in *P. falciparum*[16].

As described in the Materials and Methods section, further constructs were shared by the Richard Young research group, including MED1-IDR fused C-terminally to mCherry and mEGFP as well as the CTD of *H. sapiens* RPB1 fused to mEGFP. These constructs are henceforth referred to as mCherry-MED1-IDR, mEGFP-MED1-IDR and mEGFP-*Hs*RPB1-CTD respectively (Fig. M1).

The recombinant proteins were expressed and purified from *E. coli* as described in the Materials and Methods section and the integrity and purity of protein preparations was determined by SDS PAGE analysis. Each final protein preparation shows a distinct band around the expected molecular mass (Figs. S5A–B), calculated from its primary amino acid sequence. Many preparations are however affected by a common set of issues: contaminant bands around 80 kDa, and in addition to the full-length recombinant protein, truncated species which distribute around the 32 kDa marker. The 80 kDa contaminant bands are most prominent in the *Pf*RPB1-CTD protein preparation and further purification techniques, including ion-exchange chromatography, were investigated to separate these contaminants from the full-length proteins. These investigations did not yield fruitful results as the contaminant bands co-purified with the dominant *Pf*RPB1-CTD band and large amounts of protein were lost during further purification (data not shown). The collection of bands around 32 kDa, most prominent in the *Pf*- and *Hs*RPB1-CTD fusion protein preparations are likely C-terminal truncations of the full-length protein constructs, containing the 6His-tag, the fluorescent protein and a distribution of fragments of the fused IDR (Figs. S5A–B). The highly repetitive nature of the CTD sequences may result in premature translation termination. Furthermore, because highly disordered protein sequences are unstructured, they are particularly protease sensitive. Despite these issues, the protein preparations were deemed of sufficient quality to investigate the condensation properties of the IDR fusion proteins.

The TBP-, TFIIAαβ- and TFIIFß-IDR fusion proteins were previously expressed and purified in a similar fashion[81] and these preparations were reused in the study described here. These constructs generally contain shorter IDR regions and are, likely for this reason, of higher purity, with exception of the HsTFIIA-IDR-GFP fusion protein preparation, which contains some truncations/breakdown products (Fig. S5C).

Establishing a standardised data acquisition and analysis pipeline for biomolecular condensation studies using fluorescence microscopy

During exploratory investigations of the biomolecular condensation properties of the generated fusion proteins, it became apparent that a standardised data acquisition and analysis pipeline would be necessary for their precise characterisation using fluorescence microscopy. Phase separation was previously monitored using turbidity measurements of protein solutions in 384-well plates[81]. This method comes with numerous drawbacks. Firstly, the data are often difficult to reproduce due to the finicky nature of pipetting in small wells and the frequent introduction of bubbles. Furthermore, the method is very sample-hungry, requiring multiple tens of microlitres of purified protein preparations per replicated condition. Finally, increases in turbidity cannot be directly attributed to liquid-like condensates, as protein aggregates will also scatter light in solution. Increases in the number as opposed to size of droplets can also not be distinguished. This means that results of turbidity measurements must be independently verified by microscopy[95,96].

The goal was thus to establish a start-to-finish pipeline for assessing the biomolecular condensation of transcription factor-IDR fusion proteins using microscopy data. For this purpose, groups in the field of LLPS research use ImageJ[97] or custom scripts to extract information about droplet size, quantity and distribution in response to changing protein concentration or environmental conditions[98,99] and the goal was to develop a similar system.

Python scripts written by Henninger et al.[76], shared upon request, were consulted for inspiration and information on useful packages to use for biological image processing. Based on this information, a custom Python script was written to extract droplet data from the images generated with the widefield fluorescence microscope setup used in this study. Further scripts were written to display various variables generated from the image extraction, including droplet number, brightness, size distribution and the condensed fraction of a field. Briefly, the script works by segmenting an image field into droplet areas and non-droplet areas, where droplet areas are classified as pixel clusters with a brightness two times higher than the image average. The droplet areas are then further screened with a minimum size boundary and optional circularity filter, resulting in a list of regions classified as droplets (Fig. M3). Information about each region can then be extracted and displayed, including the droplet areas or total brightness of the pixel values (Fig. 2A). A further useful measure, the condensed fraction, can be calculated by dividing the total area of droplets by the total area of the respective field of view.

Lessons learnt throughout the iterative process of data generation and data analysis were used to finetune both the data acquisition techniques as well as the analysis scripts. Specific artefacts and aberrations resulting from widefield microscopy had to be accounted for. Firstly, imaged droplets often possessed a halo of light surrounding the actual droplet region. This halo is the result of the point spread function of light and is compounded in widefield microscopy as many z-stacks are represented at once. During the thresholding process, these halos are interpreted as a spiky surface on the droplets (Fig. 2B) and distort measurements of a droplet's circumference. Furthermore, background regions display highly irregular and noisy distribution of fluorescence, making the distinction between droplet and dilute phase more difficult to delineate and connecting areas of droplets that are in fact separate (Fig. 2B). These issues can be overcome by repeatedly imaging the same region over a short space of time (10 images over 1 s) and generating an average image over this timespan (Fig. 2B). Rapid and consistent imaging was accomplished with the help of imaging macros in the NIS-elements microscopy software and image averaging was done with the help of an additional Python script.

Results of initial time course experiments revealed that the number of droplets observed in a given field is highly dependent on the time at which fields are imaged. This phenomenon has been previously noted[69,100]: following the transfer of the droplet assay to the imaging slide (Fig. M2), protein droplets, which are heavier than the surrounding solution, begin to settle onto the coverslip (Figs. 2C–D). To ensure consistency across replicates and comparability across experimental conditions, droplet reactions were allowed to settle for 30 min following transfer to the slide chamber before image acquisition (Fig. M2). A circularly-reinforcing, iterative process of finetuning both data acquisition and analysis methodologies resulted in a standardised workflow for the investigation of biomolecular condensation by transcription factor-IDR fusion proteins by widefield fluorescent microscopy.

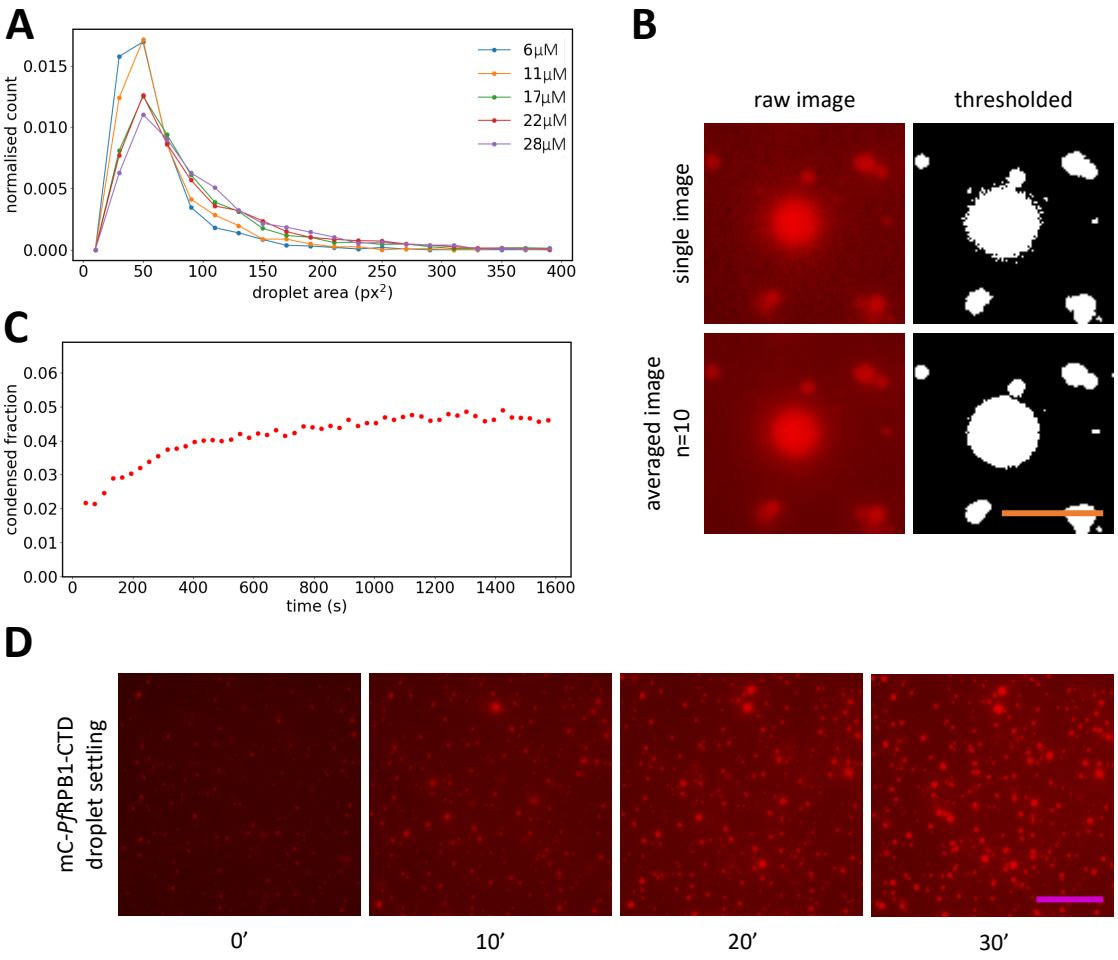


Fig. 2. Establishing a standardised microscope image data acquisition and analysis pipeline. (A) Python script output displaying normalised distribution of mC-*Pf*RPB1-CTD droplet areas across multiple fields of view (*n=10*) observed at different mC-*Pf*RPB1-CTD concentrations (6-28 μM). Lower protein concentrations result in a more left-skewed distribution, indicating a greater proportion of droplets are smaller in area; higher protein concentrations result in a fatter-tailed distribution, indicating a greater proportion of droplets are larger in area. These histograms were used to validate the functionality of the analysis scripts as well as ensure reproducibility of results. (B) The circularity of a thresholded droplet and background brightness values are affected by noisy microscopy data. This can be partially overcome by averaging images at a single location over a short period of time (<1 s). (C - D) Droplets settle onto the coverslip over time, plateauing after 30 min. (C) Condensed fraction, defined as the proportion of the entire image area occupied by droplets, plotted over a period of 30 min (D) Widefield microscopy images of the same field of view over a 30 min period. *Scale bars*, *orange*: 10 μm; *pink*: 50 μm.

*Pf*RPB1-CTD promotes biomolecular condensation *in vitro*

With a standardised workflow in hand, the goal was to analyse individual fusion proteins using the pipeline to understand their basic properties. In light of recent literature on *Hs*RPB1-CTD LLPS and the observation of distinct sequence features of the *Pf*RPB1-CTD (Fig. 1B), a significant amount of focus was dedicated to examining the properties of the mC-*Pf*RPB1-CTD fusion protein.

As shown in Fig. 3, *Pf*RPB1-CTD forms largely spherical assemblies in the presence of the molecular crowding agent PEG (MW 3000 Da) at physiological salt concentrations[62] (150 mM KCl; Fig. 3B). PEG is commonly used in the field of LLPS research as a molecular crowder in order to mimic the crowded cellular environment. Molecular crowders effectively increase the concentration of proteins in solution and thus facilitate protein condensation under *in vitro* conditions[62]. At the highest protein concentrations of 28 µM, mC-*Pf*RPB1-CTD condensates are multiple µm in diameter. The condensed fraction of *Pf*RPB1-CTD condensates decreases at lower protein concentrations and condensates can no longer be observed at a concentration of 1 µM *Pf*RPB1-CTD (Fig. 3A). In contrast, a fluorescent protein-only control does not form condensates at high protein and molecular crowder concentrations (Fig. S6A). The droplets formed by mC-*Pf*RPB1-CTD are thus driven by the presence of the fused IDR.

Further assays were carried out to investigate *Pf*RPB1-CTD LLPS under a variety of conditions as well as the material properties of resultant condensates. The effect of ionic buffer strength on biomolecular condensation of *Pf*RPB1-CTD was tested by incubating *Pf*RPB1-CTD protein at different concentrations in the presence of varying KCl concentrations. mC-*Pf*RPB1-CTD condensation occurs at a wide range of salt conditions, and spherical assemblies are observed between 50 mM and 300 mM KCl (Fig. 3). Assembly formation is only inhibited in the presence of 500 mM KCl – under these conditions spherical condensates are no longer visible and microscopy fields show an even distribution of fluorescence (Fig. 3B). mC-*Pf*RPB1-CTD condensate formation is further a function of the PEG concentration in the assay (Fig. 4A). *Pf*RPB1-CTD does not form droplets in the absence of PEG (Fig. 4A). At 6% PEG concentration, irregular aggregates are detected, and distinct spherical droplets are formed at 9 and 12% PEG (Fig. 4A).

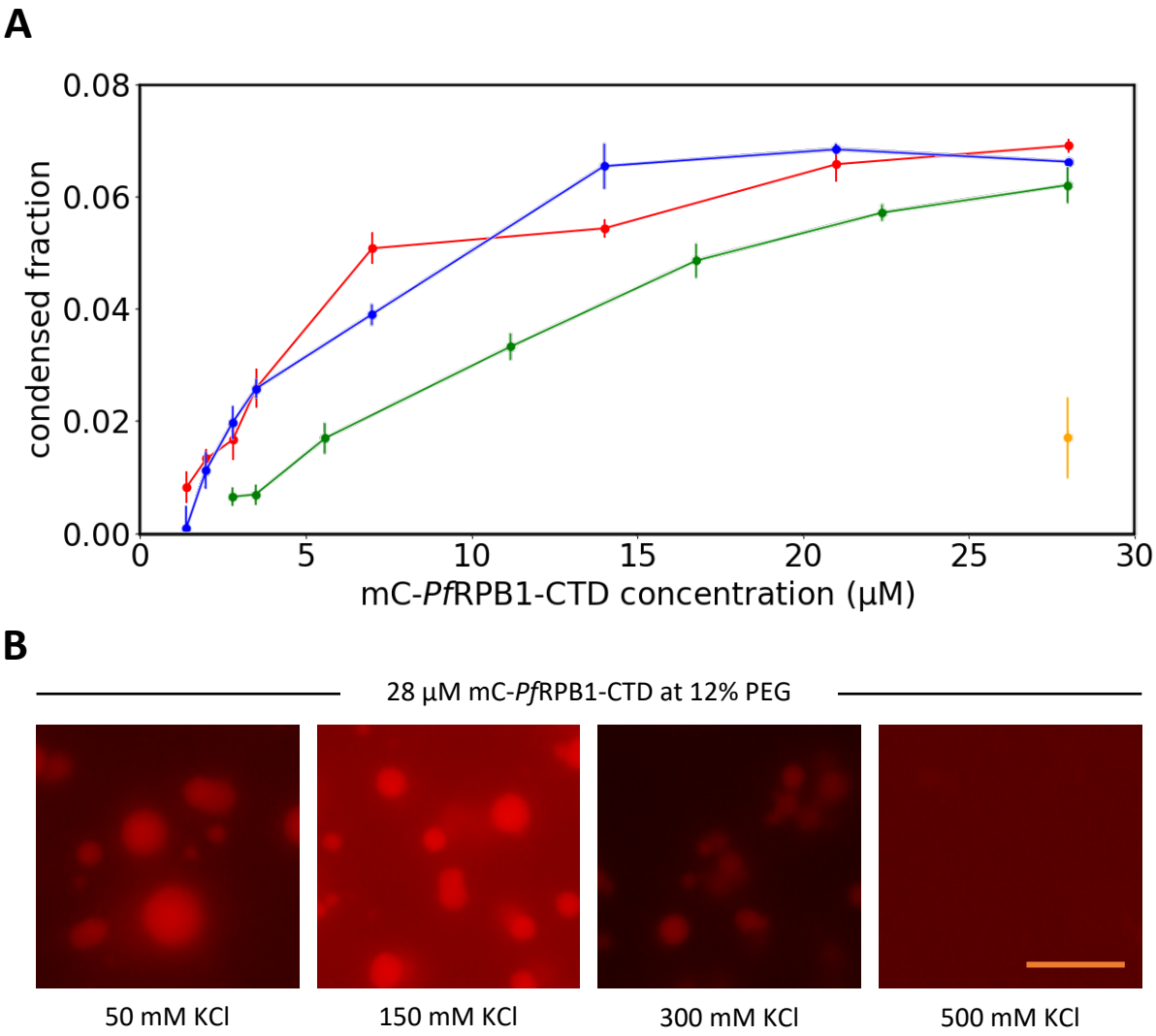


Fig. 3. mC-*Pf*RPB1-CTD forms spherical biomolecular condensates under a range of salt concentrations conditions *in vitro*. (A) Condensed fraction curves for mC-*Pf*RPB1-CTD in response to increasing protein concentration at 12% PEG and 150 mM (*red*), 150 mM (*blue*), 300 mM (*green*) and 500 mM KCl (*yellow*, single point). Condensed fraction points represent the mean ± SD of 10 images for each condition. (B) Representative images of condensate formation at a protein concentration of 28 µM and indicated KCl concentrations. *Scale bar*: 10 µm.

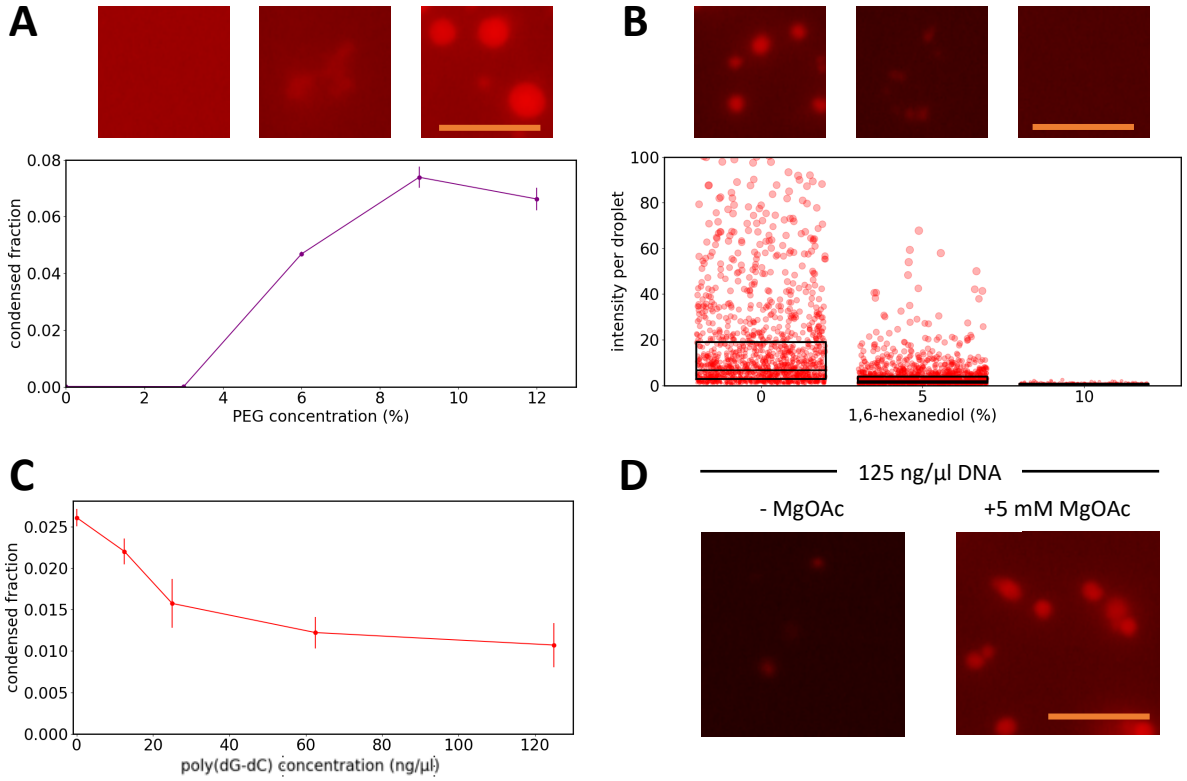


Fig. 4. mC-*Pf*RPB1-CTD biomolecular condensation is induced by PEG and negatively affected by 1,6-hexanediol and poly(dG-dC) dsDNA. (**A**) mC-*Pf*RPB1-CTD droplet formation in response to PEG concentration at 28 µM protein and 150 mM KCl. *Top*: representative images of assemblies at 0, 6 and 12% PEG concentration. Note indistinct and irregular shape of assemblies at 6% PEG; *bottom*: condensed fraction curve in response to PEG concentration. (**B**) mC-*Pf*RPB1-CTD droplet formation is sensitive to 1,6-hexanediol. Condensates were formed at 10 µM protein, 150 mM KCl and 12% PEG, in the absence and presence of 5% and 10% (w/v) 1,6-hexanediol. *Top*: representative fluorescent microscopy images; *bottom*: integrated intensity of pixel brightness per droplet across 10 replicate fields of view per condition. Box limits denote upper and lower quartiles and the centre line denotes the median value. (**C**) Condensed fraction curve for mC-*Pf*RPB1-CTD at 10 µM protein, 75 mM KCl and 12% PEG in response to increasing concentrations of poly(dG-dC) dsDNA. (**D**) Inhibition of mC-*Pf*RPB1-CTD droplet formation by DNA is not observed in the presence of 5 mM MgOAc. Condensed fraction points represent the mean ± SD of 10 images for each condition. *Scale bars:* 10 µm.

Next, mC-*Pf*RPB1-CTD was incubated in the presence of increasing concentrations of the aliphatic alcohol 1,6-hexanediol. This compound is frequently used in the study of biomolecular condensates and is thought to disrupt hydrophobic interactions that may play a role in assembly formation[101]. The presence of 5% 1,6-hexanediol reduces the size and number of mC-*Pf*RPB1-CTD droplets and at a concentration of 10%, droplets are no longer detectable (Fig 4B).

Finally, nucleic acids are known to affect condensation properties of various phase separating proteins[50,76,102]. Given that in the physiological context of the cell, the *Pf*RPB1-CTD functions in the presence of high concentrations of DNA and RNA, the effect of DNA on condensate formation was investigated. Incubating a fixed concentration of mC-*Pf*RPB1-CTD in the presence of increasing concentrations of poly(dG-dC) double stranded DNA (dsDNA), reduces the formation of assemblies, and droplet formation of 10 µM mC-PfRPB1-CTD was almost completely inhibited in the presence of 125 ng/µl dsDNA (Fig. 4C). Magnesium ions are known to influence protein–nucleic acid interactions driving condensation[103]. Interestingly, inhibition of mC-*Pf*RPB1-CTD condensation by dsDNA is not observed in the presence of 5mM magnesium acetate (MgOAc) (Fig. 4D).

*Pf*RPB1 assemblies rapidly mature and obtain gel-like properties

Observations made during the initial characterisation of mC-*Pf*RPB1-CTD condensation warranted a more careful investigation of the material properties of the resulting droplets. It was noted that assemblies are generally spherical, distinct and separated from one another when visualised immediately after formation (Fig. 5A). The spherical appearance of the assemblies is a strong indication that LLPS is driving their formation, given that surface tension in liquids drives them to adopt the smallest volume-to-surface-area ratio[40]. However, when condensation reactions were incubated for tens of minutes to allow condensates to settle onto the coverslip for imaging (Figs. M2, 2D), droplet clusters are observed, consisting of up to ten individual droplets that appear to have partially fused (Fig. 5). This observation indicates a change in material properties over time, from an initial liquid-like state to a more gel- or solid-like state that prevents complete fusion of individual droplets. (Fig. 5B).

To investigate when this hardening occurs, an attempt was made to capture the fusion of two droplets. Fusion of droplets, a characteristic property of liquids, is often used as evidence for LLPS driving the formation of the assemblies under investigation[62]. This proved difficult, even immediately after the induction of condensation, suggesting that hardening sets in very soon after condensate formation. Many partial fusion events could however be recorded (Fig 5C). These involve two droplets coming together and partially merging compartments, but not fully fusing into a larger spherical droplet. The changing material properties of mC-*Pf*RPB1-CTD condensates were further investigated using a salt-reversibility assay. Briefly, droplet formation was induced in a 12% PEG, 150 mM KCl buffer.

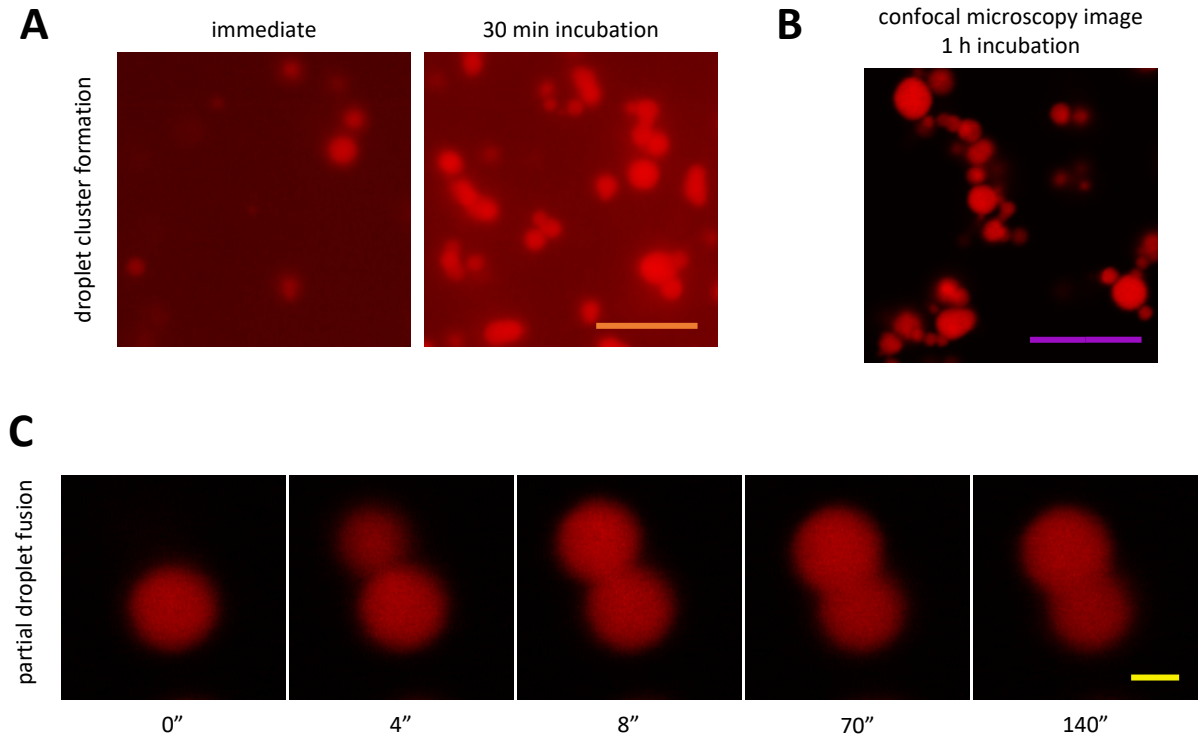


Fig. 5. mC-*Pf*RPB1-CTD condensates form droplet clusters through partial fusion. (**A**) mC-*Pf*RPB1-CTD at 28 µM protein in the presence of 150 mM KCl and 12% PEG. The same field of view imaged by widefield microscopy immediately after droplet formation and after 30 min incubation (**B**) Same conditions as described in (A) but imaged after 1 h incubation by confocal fluorescent microscopy. (**C**) Partial droplet fusion of mC-*Pf*RPB1-CTD condensates captured by confocal fluorescent microscopy. Droplets rapidly join upon contact, but coalescence stalls and the resulting assembly is no longer spherical. *Scale bars*, *orange:* 10 µm*, pink:* 20 µM, *yellow*: 1 µm.

Samples of this solution were then either diluted with the same buffer or with a high-salt buffer, resulting in a final concentration of 500 mM KCl. In order to test for changes in droplet material properties over time, dilution was carried out either immediately after droplet formation or after a 30 min incubation period. If droplets are fully dynamic, they should dissolve after dilution to 500 mM KCl, as mC-*Pf*RPB1-CTD cannot form condensates under these conditions (Fig. 3). In contrast, dilution of the reaction without changing salt concentration should only moderately impact the size and distribution of droplets, as the protein concentration is halved but conditions remain permissive to condensate formation. As shown in Fig. 6A, droplets disperse when challenged by 500 mM KCl immediately following droplet formation (Fig. 6A). In contrast, the majority of droplets are resistant to 500 mM KCl after 30 min incubation, indicating a hardening of material properties.

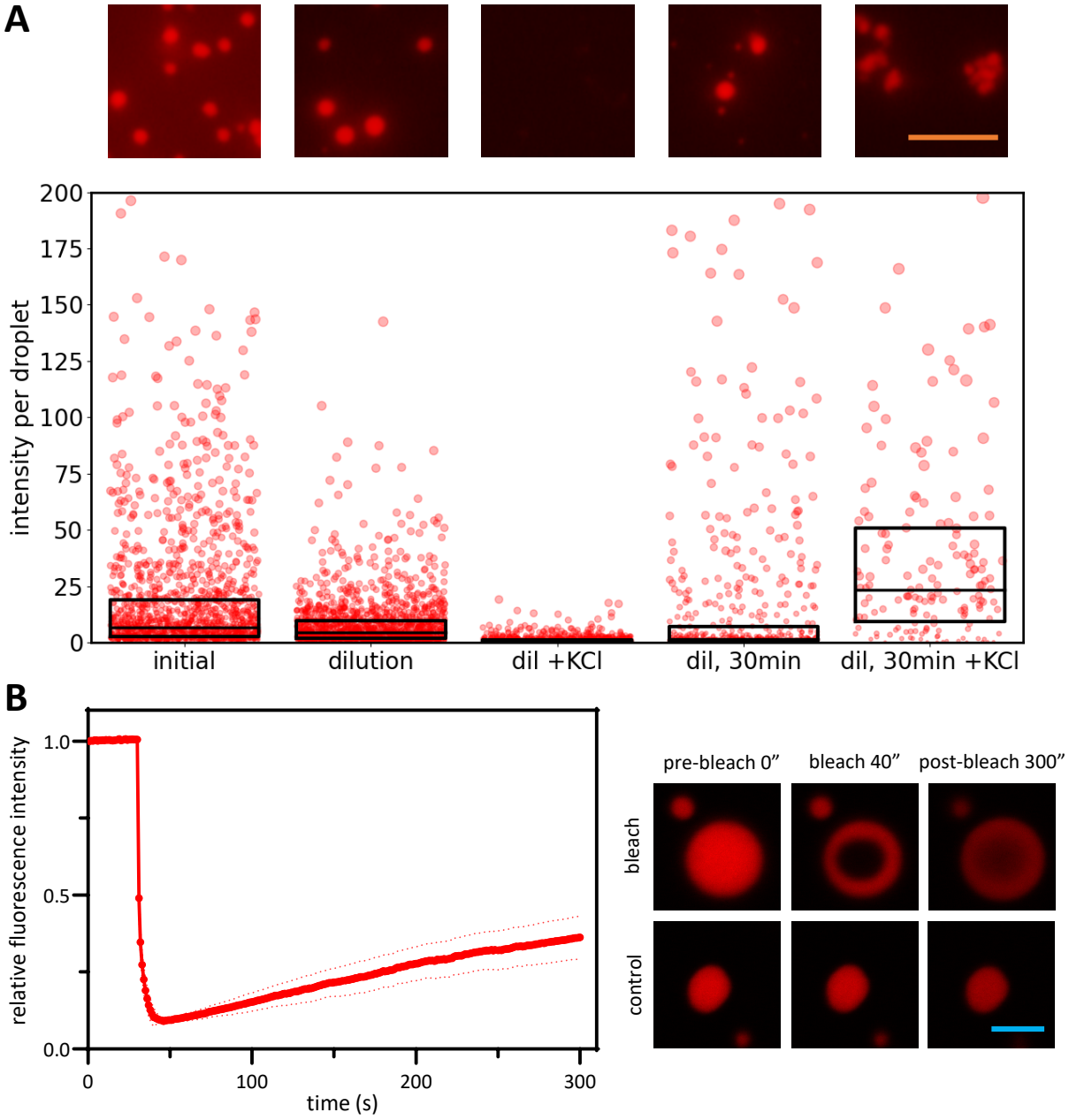


Fig. 6. mC-*Pf*RPB1-CTD condensates exhibit gel-like properties and "harden" over time. (**A**) Salt reversibility assay for mC-*Pf*RPB1-CTD. Condensates are formed at 10 µM protein, 150 mM KCl, 12% PEG (*initial*), followed by a 1:1 dilution with the same buffer (*dil*) or with an increase to 500 mM KCl (*dil +KCl*), either immediately or after a 30 min incubation (*30min*). *Top*: representative images of droplets at each indicated condition; *bottom*: total pixel intensity per droplet, box limits denote upper and lower quartiles and the centre line denotes the median value. (**B**) Fluorescence recovery after photobleaching of mC-*Pf*RPB1-CTD condensates at 28 µM protein, 150 mM KCl and 12% PEG. Reactions were visualised, and bleaching was carried out, immediately after condensate formation. Following 30 s of a pre-bleaching period, regions were bleached to 10% of their original intensity and monitored for a total of 5 min. *Left*: Relative fluorescence intensity defined as average pixel intensity of bleached region normalised to region of similar size in unbleached droplet to account for background bleaching. Each point is normalised as a fraction of the initial reading; *right*: representative images of bleached and unbleached droplets at indicated time points. All points represent the mean ± SD (dotted line) for 3 replicate bleaching experiments. *Scale bars*, *orange:* 10 µm*, cyan*: 5 µm.

These observations support a model in which spherical mC-*Pf*RPB1-CTD assemblies form through LLPS and then rapidly transition from a liquid-like to a gel- or solid-like material state. This transition is largely completed after 30 min (Fig 6A).

The material properties of mC-*Pf*RPB1-CTD condensates were further investigated by FRAP[62] using a confocal fluorescent microscope. This method involves bleaching a region of a droplet with a high intensity light burst and monitoring the rate at which photobleached molecules are replaced with unbleached, fluorescent ones from surrounding regions. As shown in Fig. 6B, bleached regions of mC-*Pf*RPB1-CTD droplets, assessed immediately after formation, exhibit partial recovery from photobleaching, indicating both static and dynamic fractions co-existing within droplets (Fig. 6B). These results are consistent with a rapid change from liquid-like to gel- or solid-like properties of mC-*Pf*RPB1-CTD condensates.

MED1-IDR condensates transition to gel-like states

Multiple MED1-IDR fusion proteins were expressed and purified during this study. This includes expression of an independently constructed MED1-IDR-YFP fusion protein as well as expression of mC- and mEGFP-MED1-IDR fusion constructs provided by the Young laboratory (Whitehead Institute, MIT Boston, USA; see Materials and Methods section).

Following an initial characterisation of the self-generated MED1-IDR-YFP construct, it was decided to perform further characterisation and assays using the previously published constructs[69]. Not only were there existing data to act as a benchmark, but they could be purified at higher concentrations and formed larger assemblies (Figs. S5, compare S7 and 7B). Furthermore, having available MED1-IDR fused to either mEGFP or mCherry allowed us to freely combine fusion proteins in mixing studies performed later. The self-generated construct does show that the orientation of MED1-IDR with respect to the fused fluorescent protein does not significantly affect the condensation properties of the fusion protein. The self-generated MED1-IDR-YFP formed condensates both in the absence and presence of PEG at 150 mM KCl (Fig. S7), in accordance with the behaviour of the mEGFP-MED1-IDR fusion protein (Figs. 7–8).

Previously published work has summarised the biophysical properties of MED1-IDR condensation *in vitro*, but not rigorously investigated its condensate-forming properties across a wide range of conditions[69]. 10 µM MED1-IDR has been demonstrated to form condensates between 50 and 350 mM monovalent salt ion concentration and published data suggest these condensates are reversible upon challenge with a high-salt buffer[69]. This study sought to extend the *in vitro* characterisation of MED1-IDR-driven condensation.

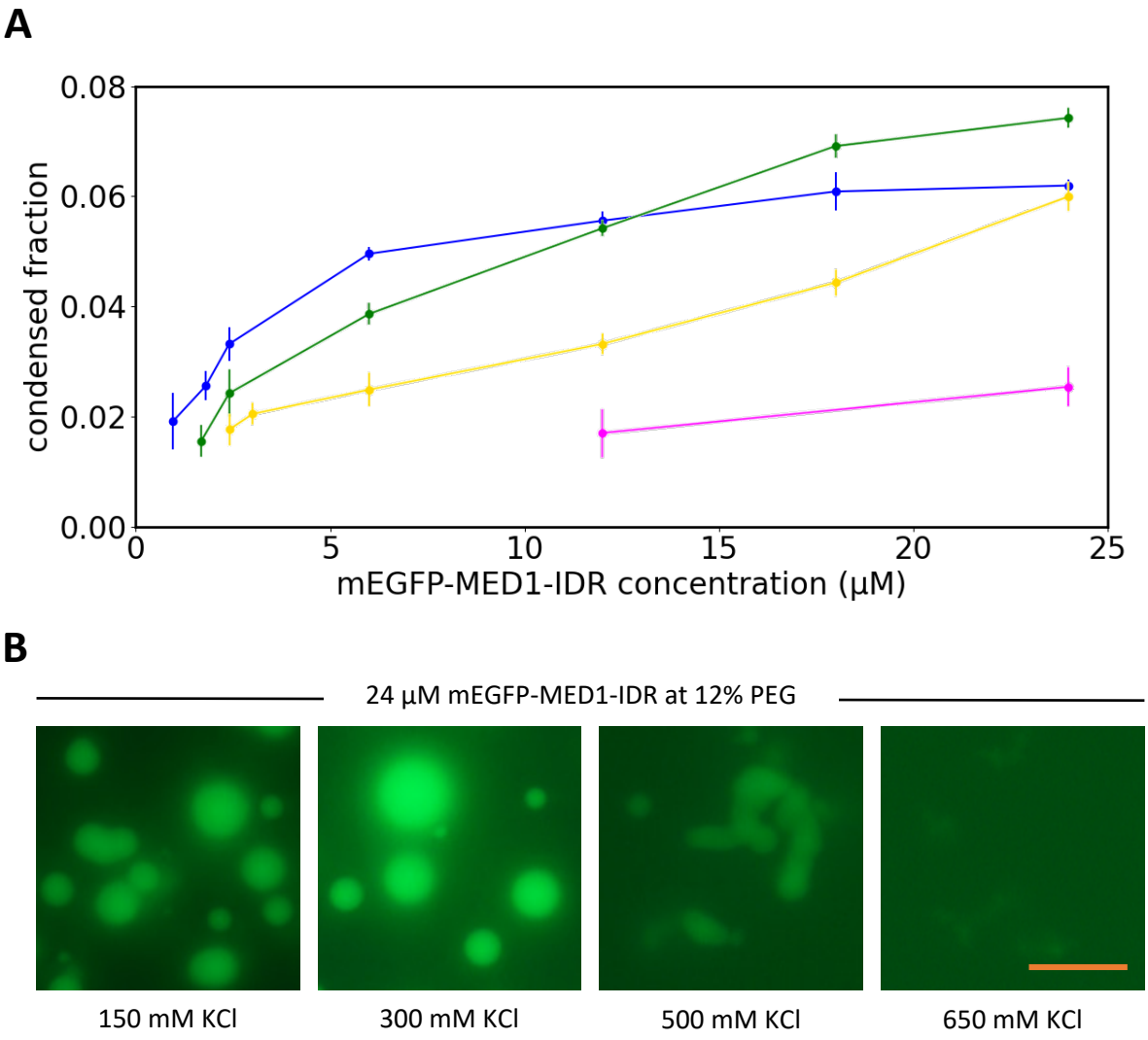


Fig. 7. mEGFP-MED1-IDR forms biomolecular condensates under a range of salt concentrations conditions *in vitro.* (**A**) Condensed fraction curves for mEGFP-MED1-IDR in response to increasing protein concentration at 12% PEG and 150 mM (*blue*), 300 mM (*green*), 500 mM (*yellow*) and 650 mM (*pink*) KCl. Condensed fraction points represent the mean ± SD of 10 images for each condition. (**B**) Representative images of condensate formation at a protein concentration of 24 μM and indicated KCl concentrations. *Scale bar*: 10 μm.

Titration curves generated with mEGFP-MED1-IDR at fixed PEG (12%) and varying KCl concentrations confirm that condensation is tolerant to a wide range of salt conditions, reproducibly forming droplets at up to 500 mM KCl (Fig. 7). At 150 and 300 mM KCl, droplets are up to multiple µm in diameter, spherical and separated (Fig. 7B). At 500 mM KCl, mEGFP-MED1-IDR forms gel-like fusions of multiple smaller droplets. Salt concentrations of 650 mM KCl disrupt spherical assembly formation and result in irregular and indistinct aggregate formation (Fig. 7B). 1,6-hexanediol has been demonstrated to disrupt MED1-IDR puncta in living cells[69] and to moderately affect condensation *in vitro*. The data presented here show that MED1-IDR condensate formation is remarkably similar in the absence and in the presence of 5% and 10% 1,6-hexanediol (Fig. 8A), even at the low protein concentration of 6 µM used. Thus, MED1-IDR condensates are highly resistant to challenge with 1,6-hexanediol. This discrepancy between published results and the data generated here may be the result of divergent buffer conditions (PEG-3000 used here compared to PEG-8000[69]) or slight differences in handling techniques.

Notably, MED1-IDR forms droplets in the absence of the molecular crowding agent PEG at 150 mM KCl (Fig. 8B), a property not exhibited by any other fusion proteins studied here. However, increasing the PEG concentration up to 12% results in increasingly higher measures of the phase separated fraction (Fig. 8B). The ability to form condensates in the absence of PEG provided a useful opportunity to investigate whether the presence of PEG affects droplet properties. Salt-reversibility assays were performed with mEGFP-MED1-IDR droplets, pre-formed at different PEG concentrations. These experiments show a clear effect of PEG on droplet properties. In the absence of PEG, droplets remain dynamic and reversible in the absence of PEG, even after a 30 min incubation (Fig. 9A). In contrast, the presence of only 6% PEG, a relatively low concentration, promotes hardening of a fraction of the phase-separated assemblies, which cannot be dissolved by 500 mM KCl (Fig. 9B). Strikingly, MED1-IDR droplets in 6% PEG challenged with the higher salt buffer, adopt a hollow appearance, with an outer layer more hardened and resistant than internal regions (Fig. 9B).

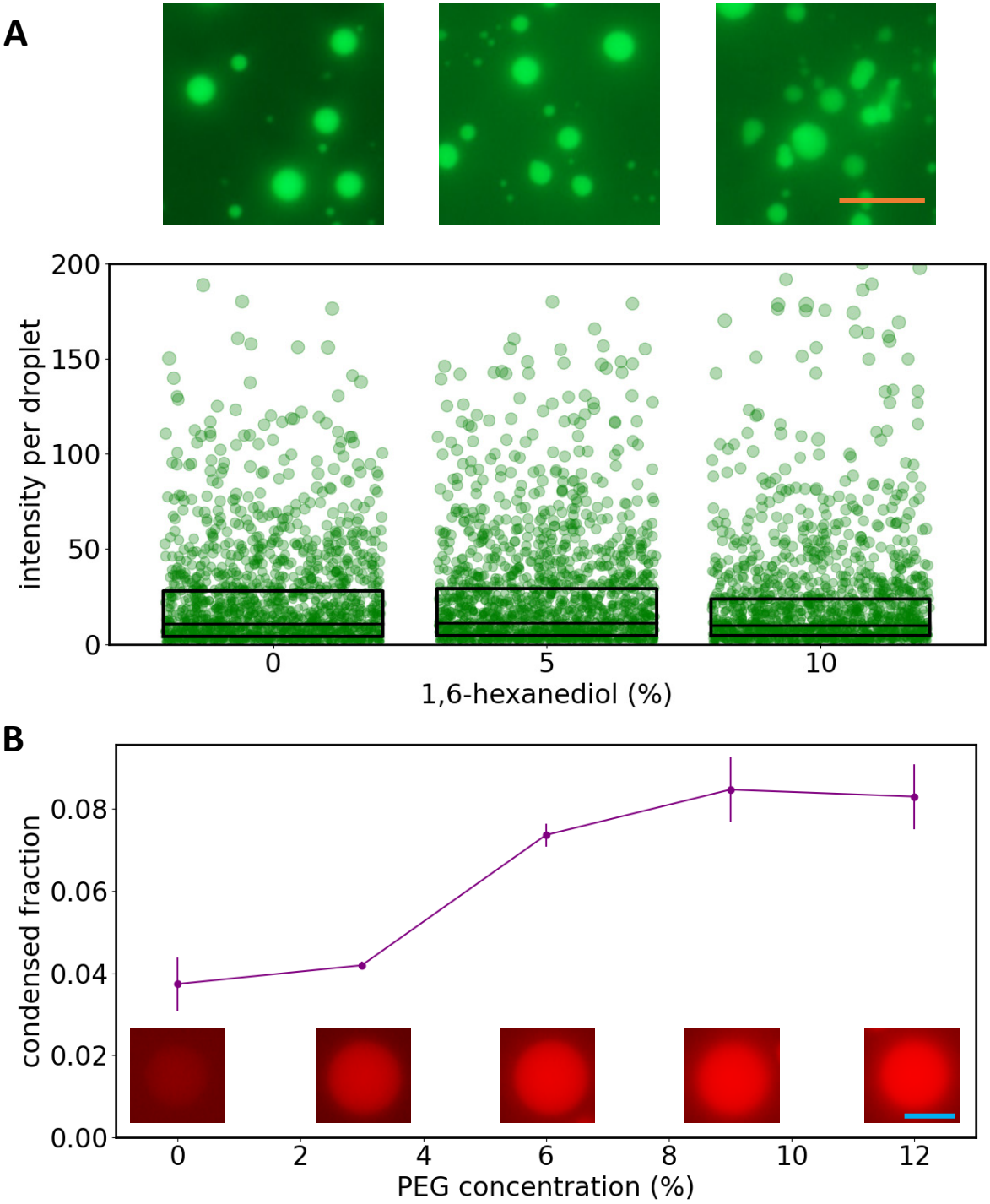


Fig. 8. MED1-IDR droplets are resistant to hexanediol and form in the absence of PEG. (**A**) mEGFP-MED1-IDR droplet formation at 6 µM protein, 150 mM KCl and 12% PEG, in the absence and presence of 1,6-hexanediol. *Top*: representative fluorescence microscopy; *bottom*: integrated intensity of pixel brightness per droplet across 10 replicate fields of view per condition; box limits denote upper and lower quartiles and the centre line denotes the median value. (**B**) Condensed fraction curve for mC-MED1-IDR in response to increasing PEG concentration at 24 µM protein and 150 mM KCl. Condensed fraction points represent the mean ± SD of 10 images for each condition. *Inset:* representative images of single droplets in condensation reaction buffer, 0–12% PEG. *Scale bars*, *orange*: 10 µm; *cyan*: 5 µm.

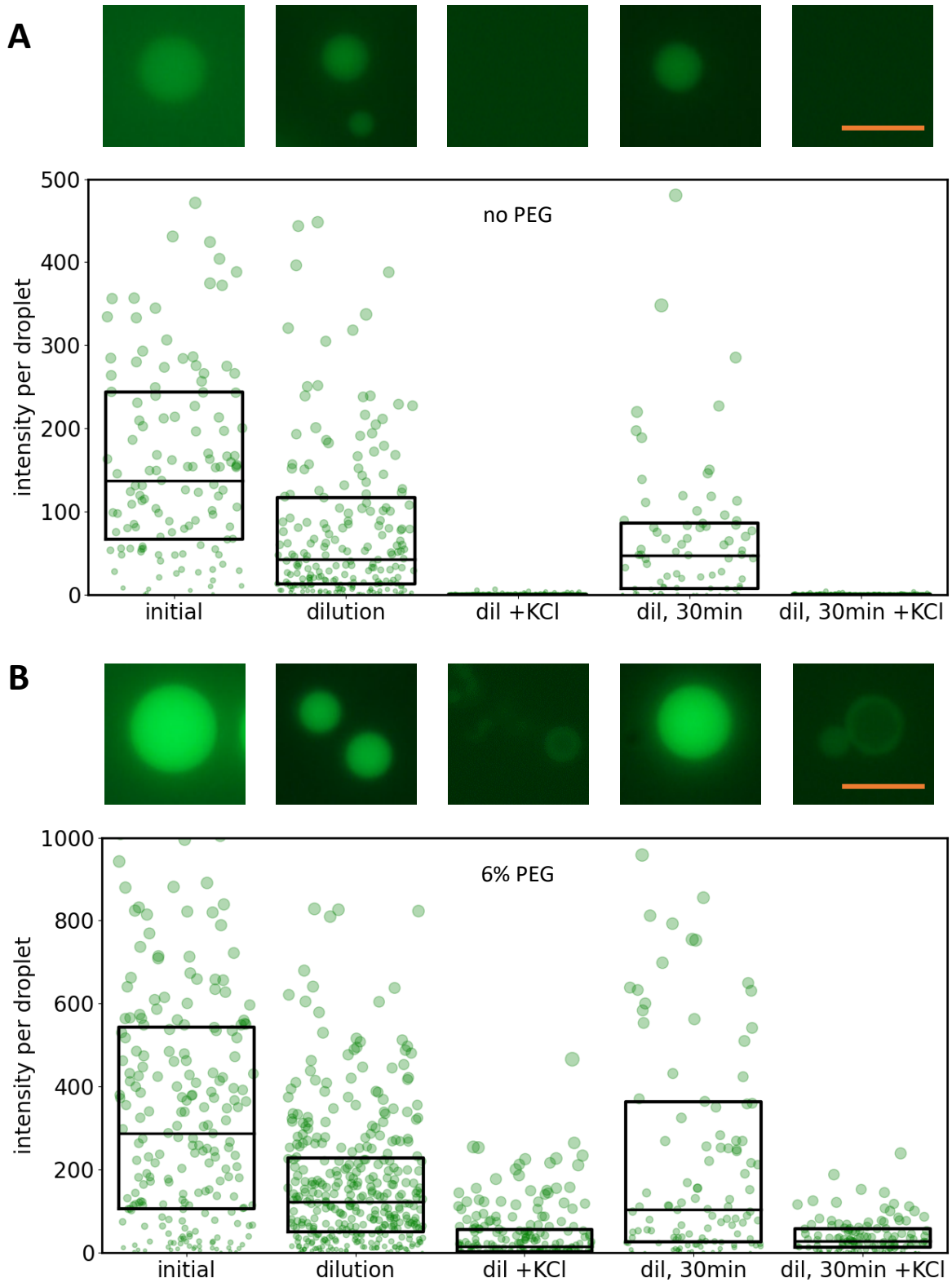


Fig. 9. mEGFP-MED1-IDR condensate formation is reversible, but droplet properties are hardened by the addition of PEG. (**A**) Salt reversibility assay for mEGFP-MED1-IDR. Condensates were formed with 24 µM protein, 150 mM KCl, 0% PEG (*initial*), followed by 1:1 dilution in the same buffer (*dil*) or to 500 mM KCl (*dil +KCl*), either immediately or after a 30 min incubation (*30min*). *Top*: representative fluorescence microscopy images of droplets at each indicated condition; *bottom*: total pixel intensity per droplet; box limits denote upper and lower quartiles and the centre line denotes the median value. (**B**) Same as in (A) but in the presence of 6% PEG. *Scale bars*: 10 µm.

Consistent with these observations, different concentrations of PEG also moderately affect the ability of bleached mEGFP-MED1-IDR droplets to recover in FRAP experiments, with assemblies formed in 12% PEG recovering more slowly than those formed in 6% PEG (Fig. 10A). Unfortunately, reliable FRAP data of mEGFP-MED1-IDR in the absence of PEG could not be obtained, due to the much larger degree of droplet movement in the less viscous environment. Partial FRAP recovery of mEGFP-MED1-IDR droplets, to only 50% of pre-bleaching levels, indicates that a fraction of constituent molecules is immobile over the time periods tested here (Fig. 10A). Despite this, MED1-IDR droplets also display characteristic liquid-like properties. Droplets are spherical and fuse upon contact, forming larger bodies, which also adopt spherical shapes (Fig. 10B). A qualitative assessment suggests that PEG also affects droplet fusion, as two fused droplets take longer to relax to a spherical shape in higher PEG conditions. Taken together, these results suggest that MED1-IDR condensates display a spectrum of material properties, from dynamic liquids to immobile gels. These properties are significantly affected by the molecular crowding agent PEG.

P. falciparum and *H. sapiens* TBP N-terminal IDRs drive biomolecular condensation

Having established the workflow for acquisition and quantitative analysis of microscopy data, protein constructs generated in earlier work[81], which could so far only be characterised in a preliminary manner, were assessed.

The N-terminal regions of *H. sapiens* and *P. falciparum* TBP (Fig. M1) are predicted to be highly disordered and were previously shown to generate spherical condensates in the presence of 20% PEG-20000[81]. Most, if not all, proteins may be induced to undergo LLPS, given the right set of environmental conditions[62]. For this reason, LLPS properties of these protein constructs were re-assessed using the more stringent conditions of 12% PEG-3000 and 150 mM KCl. As shown in Fig. 11A, *Pf*TBP-IDR condensation can be recapitulated under these conditions, evidenced by the appearance of spherical assemblies. In contrast, HsTBP-IDR condensate formation could not be observed under these conditions and required the presence of 20% PEG to observe distinct spherical condensates (Fig. 13A).

*Pf*TBP-IDR condensation is protein concentration-dependent, forming small speckles at a concentration of 2.8 µM and µm-sized spherical droplets at a concentration of 28 µM. Similar to *Pf*RPB1-CTD, these droplets partially fuse to form droplet clusters during a 30 min incubation period (Fig. 11A). Condensate formation is sensitive to increased salt concentration, with droplets not forming at 300 mM KCl (Fig. 11A).

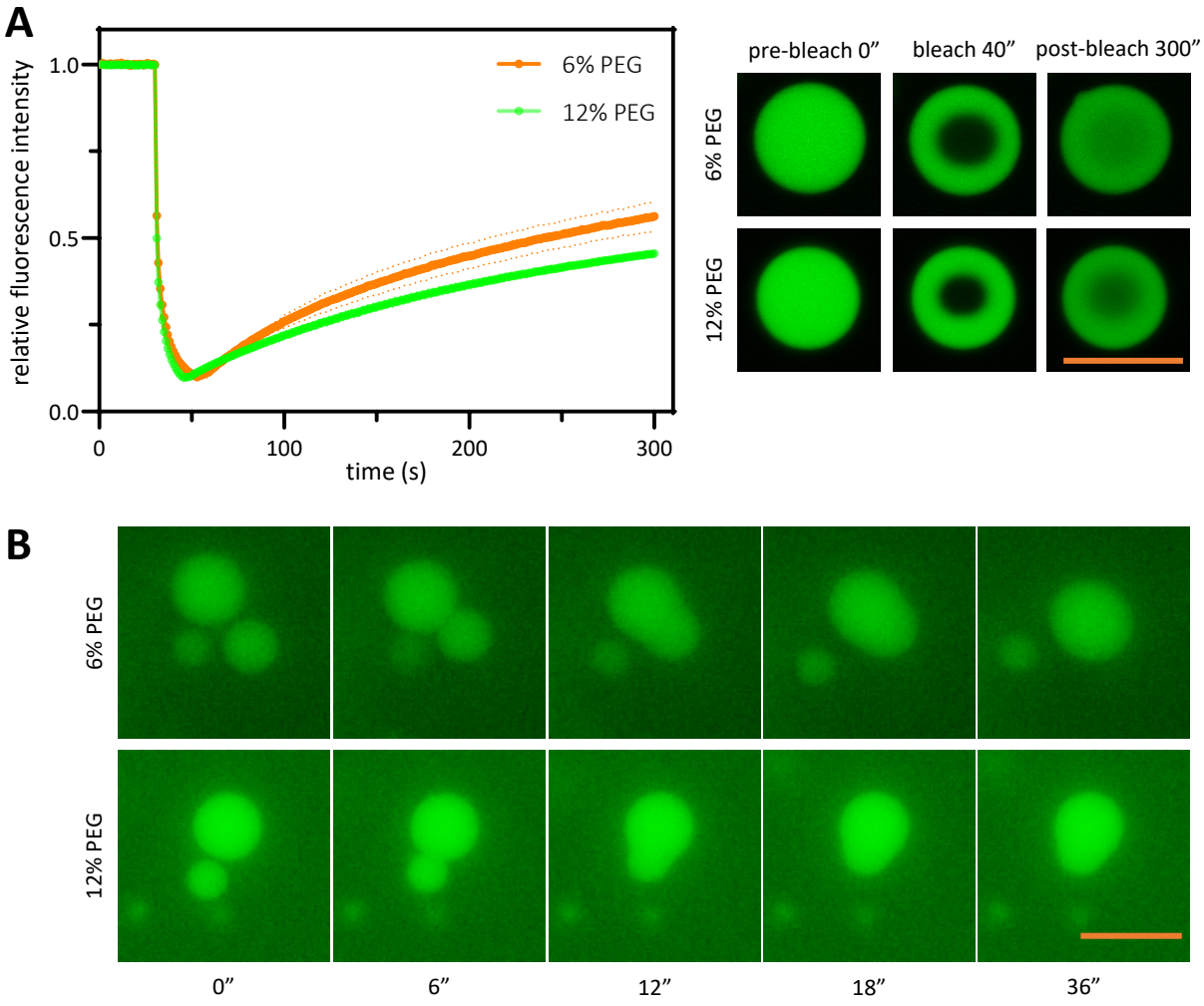


Fig. 10. mEGFP-MED1-IDR condensates partially recover from photobleaching and fuse to form larger condensates. (A) Fluorescence recovery after photobleaching of mEGFP-MED1-IDR droplets at 24 µM protein, 150 mM KCl and PEG concentrations indicated. Condensation reactions were visualised immediately after condensate formation. Following 30 s of a pre-bleaching period, regions were bleached to 10% of their original intensity and monitored for a total of 5 min. *Left*: Relative fluorescence intensity, defined as average pixel intensity of bleached region normalised to region of similar size in unbleached droplet to account for background bleaching. Each point is normalised as a fraction of the initial reading; *right*: representative images of bleached droplets under both conditions at indicated time points. All points are the mean ± SD (dotted line) for 3 replicate bleaching experiments. (B) mEGFP-MED1-IDR droplets fuse upon contact, but dynamics are affected by PEG concentration. *Top:* droplet fusion at 6% PEG; *bottom:* droplet fusion at 12% PEG. *Scale bars:* 10 µm.

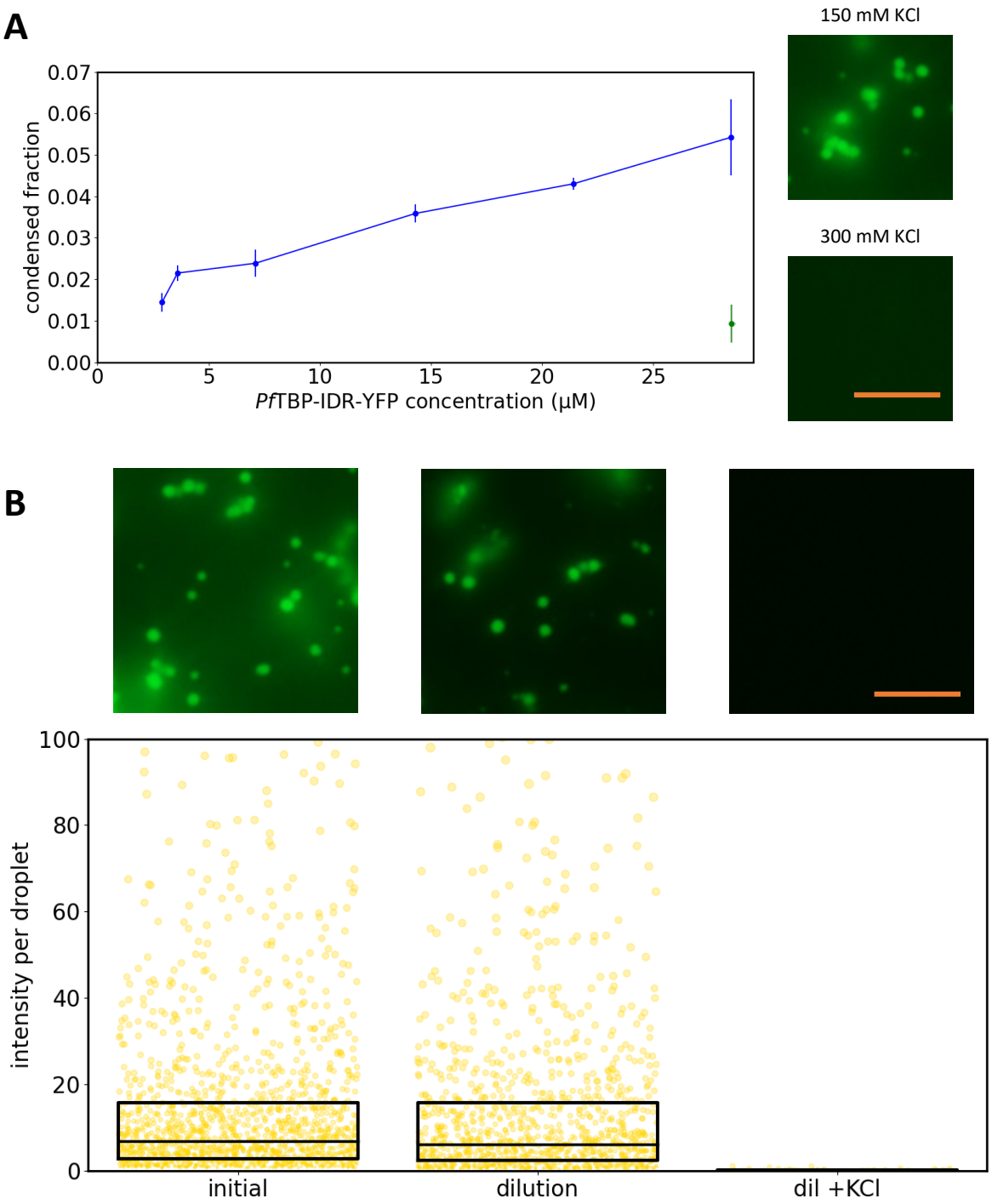


Fig. 11. *Pf*TBP-IDR-YFP forms small, spherical condensates in a concentration-dependent manner which are dissolved at 300 mM KCl. (**A**) *Pf*TBP-IDR-YFP forms condensates at 150 mM, but not 300 mM KCl. *Left:* condensed fraction curves for *Pf*TBP-IDR-YFP in response to increasing to increasing protein concentration at 12% PEG and 150 mM KCl (*blue*) or 300 mM KCl (*green,* single point); *right*: representative images of condensate formation at 28 µM protein and indicated KCl concentrations. Condensed fraction points represent the mean ± SD of 10 images for each condition. (**B**) Salt reversibility assay for *Pf*TBP-IDR-YFP. Condensates were formed at 28 µM protein, 150 mM KCl, 12% PEG (*initial*), followed by a 1:1 dilution with the same buffer (*dil*) or to 500 mM KCl (*dil +KCl*). *Top*: representative images of condensates at the indicated condition; *bottom*: total pixel intensity per droplet; box limits denote upper and lower quartiles and the centre line denotes the median value. *Scale bars*: 10 µm.

Furthermore, condensate formation is reversible, as droplets are dissolved by increasing the buffer salt concentration to 500 mM in a reversibility assay (Fig. 10B). The addition of 10% 1,6-hexanediol reduces droplet size and intensity but does not inhibit droplet formation entirely (Fig. 12A). However, high concentrations of poly(dG-dC) dsDNA completely abolish condensate formation (Fig. 12B). Taken together, these data suggest that *Pf*TBP-IDR condensation is influenced by a variety of environmental conditions, including buffer ionic strength, and the presence of 1,6-hexanediol and dsDNA.

*Hs*TBP-IDR-YFP is unable to form condensates at 12% PEG, even at a higher protein concentration of 60 µM. It forms small droplets at 18% PEG and clear, spherical condensates at 20% PEG (Fig. 13A). Further characterisation of this fusion protein was therefore undertaken at 20% PEG. *Hs*TBP-IDR forms protein concentration-dependent condensates at 150 mM KCl as evidenced by its condensed fraction curve (Fig. 13B). Similar to its *P. falciparum* counterpart, it does not form condensates at 300 mM KCl (Fig. 13B). *Hs*TBP-IDR condensates are mostly dissolved by 500 mM KCl in a salt reversibility assay but become partially resistant after a 30 min incubation period, suggesting a degree of droplet hardening over time (Fig. 14A).

*Hs*TBP-IDR condensate formation is more sensitive to 1,6-hexanediol than *Pf*TBP-IDR, forming significantly smaller assemblies at 5% 1,6-hexanediol (compare Figs. 12A and 14B). At a concentration of 10%, condensate formation is essentially inhibited (Fig. 14B). Poly(dG-dC) dsDNA has a negative effect on *Hs*TBP-IDR-YFP condensation, decreasing the size of droplets with increasing concentration (Fig. 14C). Taken together, these results demonstrate that *Hs*TBP-IDR forms condensates only under the more permissive conditions of 20% PEG and that condensation is negatively influenced by increased KCl concentration and the presence of 1,6-hexanediol and dsDNA.

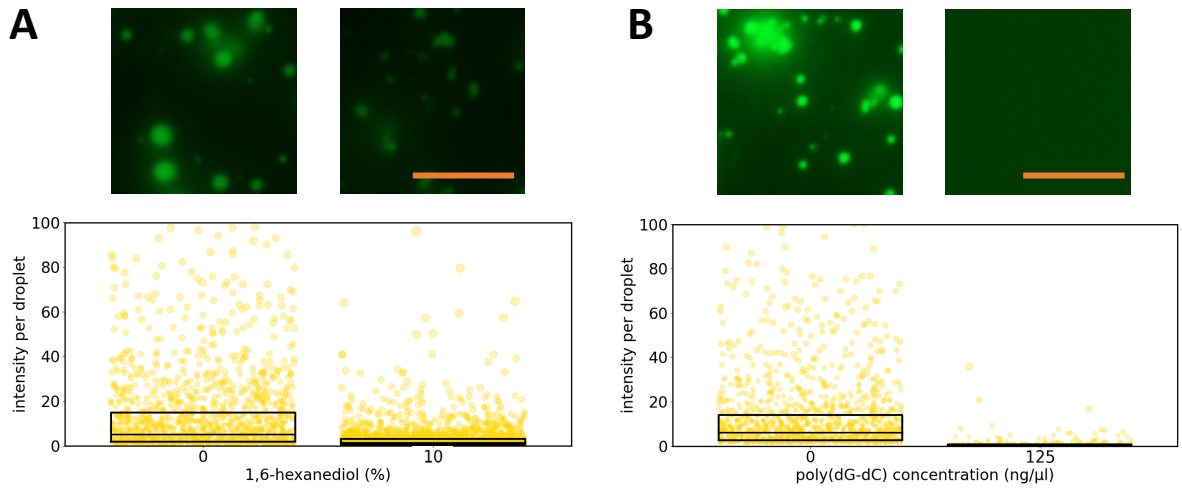


Fig. 12. *Pf*TBP-IDR-YFP condensate formation is moderately resistant to challenge with 1,6-hexanediol but fully inhibited by dsDNA. (**A**) *Pf*TBP-IDR-YFP condensate formation in response to 1,6-hexanediol at 28 µM protein, 150 mM KCl and 12% PEG. (**B**) *Pf*TBP-IDR-YFP condensate formation in response to poly(dG-dC) dsDNA at 28 µM protein, 75 mM KCl and 12% PEG. *Top*: representative images of condensates in the absence or presence of 1,6 hexanediol (A) or 125 ng/µl poly(dG-dC) dsDNA (B); *bottom*: integrated intensity of pixel brightness per droplet across 10 replicate fields of view per condition; box limits denote upper and lower quartiles and the centre line denotes the median value. *Scale bars*: 10 µm.

P. falciparum and *H. sapiens* TFIIA-IDR drive biomolecular condensation

The largest subunits of both *H. sapiens* and *P. falciparum* TFIIA contain internal IDRs, bridging conserved, structured N- and C-terminal regions (Fig. M1). In *Hs*TFIIAαβ this region is ~100 residues larger and dominated by hydrophobic amino acids[81]. *Pf*TFIIAαβ-IDR was previously demonstrated to form condensates at high PEG concentrations[81] but is also able to do so under more stringent conditions (Fig. 15A). Titration experiments demonstrate its ability to form small condensates at protein concentrations between 6 and 30 µM at 50 mM KCl and 12% PEG (Fig. 15A). Below 6 µM the fusion protein remains evenly distributed throughout solution, suggesting a single-phase, dissolved system. Increased salt concentrations also disrupt condensation. At 150 mM KCl, which resembles physiological ionic strength[62], *Pf*TFIIAαβ-IDR forms irregular aggregates (Fig. 15B). In the presence of 300 mM KCl, condensation could not be observed (Fig. 15B). Finally, *Pf*TFIIAαβ-IDR condensate formation is also affected by 1,6-hexanediol. However, addition of 10% 1,6-hexanediol does not fully inhibit droplet formation, but instead significantly reduces droplet size (Fig. 15C).

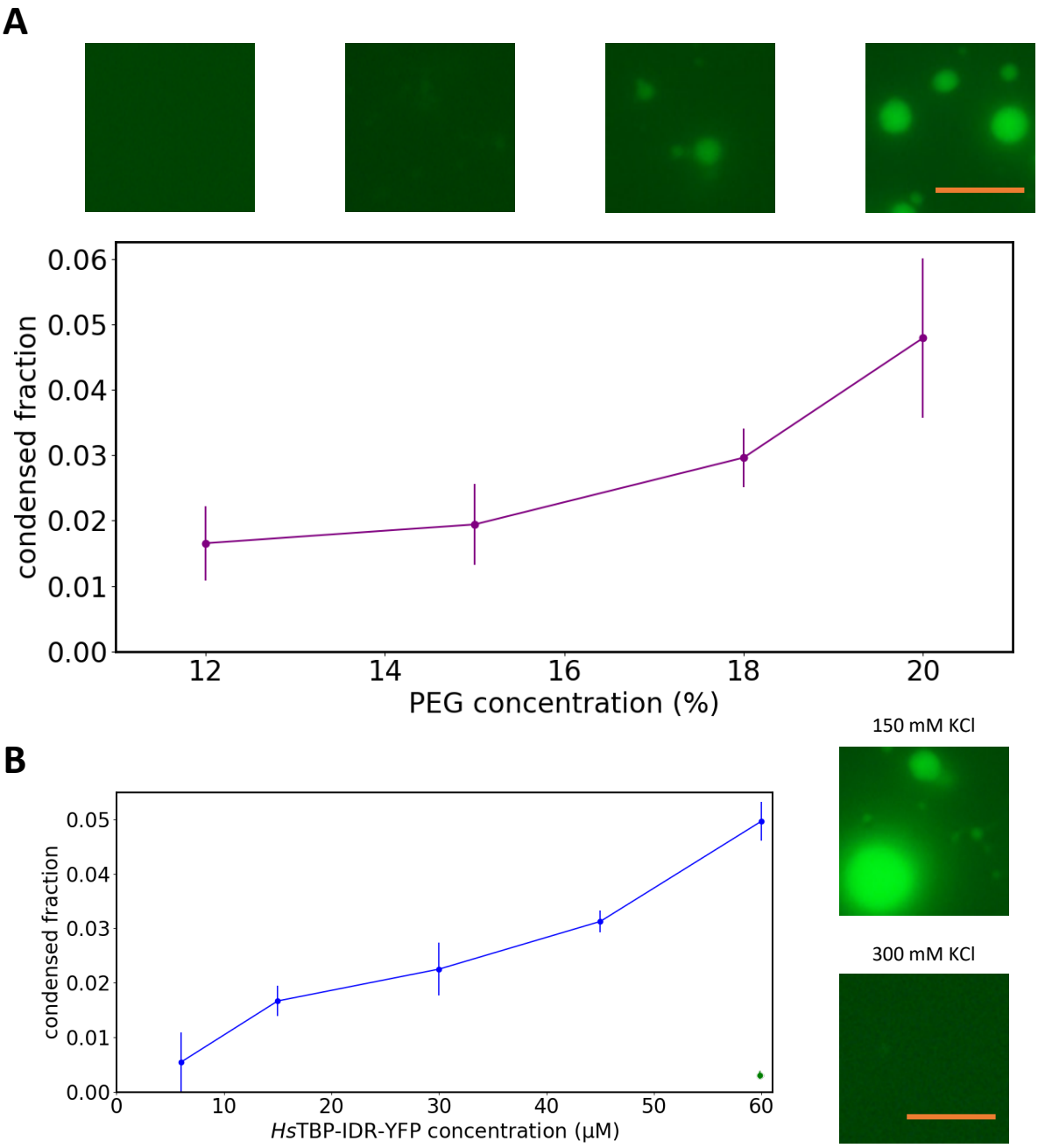


Fig. 13. *Hs*TBP-IDR-YFP forms condensates at high PEG concentrations in a protein concentration-dependent manner. (**A**) *Hs*TBP-IDR-YFP only forms condensates at higher PEG concentrations. *Top:* representative images of condensates formed at 12, 15, 18 and 20% PEG; *bottom:* condensed fraction curves for *Hs*TBP-IDR-YFP in response to increasing PEG concentration at 60 µM protein and 150 mM KCl. (**B**) *Hs*TBP-IDR-YFP forms condensates at 150 mM, but not 300 mM KCl. *Left:* condensed fraction curves for *Hs*TBP-IDR-YFP in response to increasing protein concentration at 20% PEG and 150 mM KCl (*blue*) or 300 mM KCl (*green*, single point); *right*: representative fluorescence microscopy images of condensates at 60 µM protein and indicated KCl concentrations. Condensed fraction points represent the mean ± SD of 10 images for each condition. *Scale bars:* 10 µm.

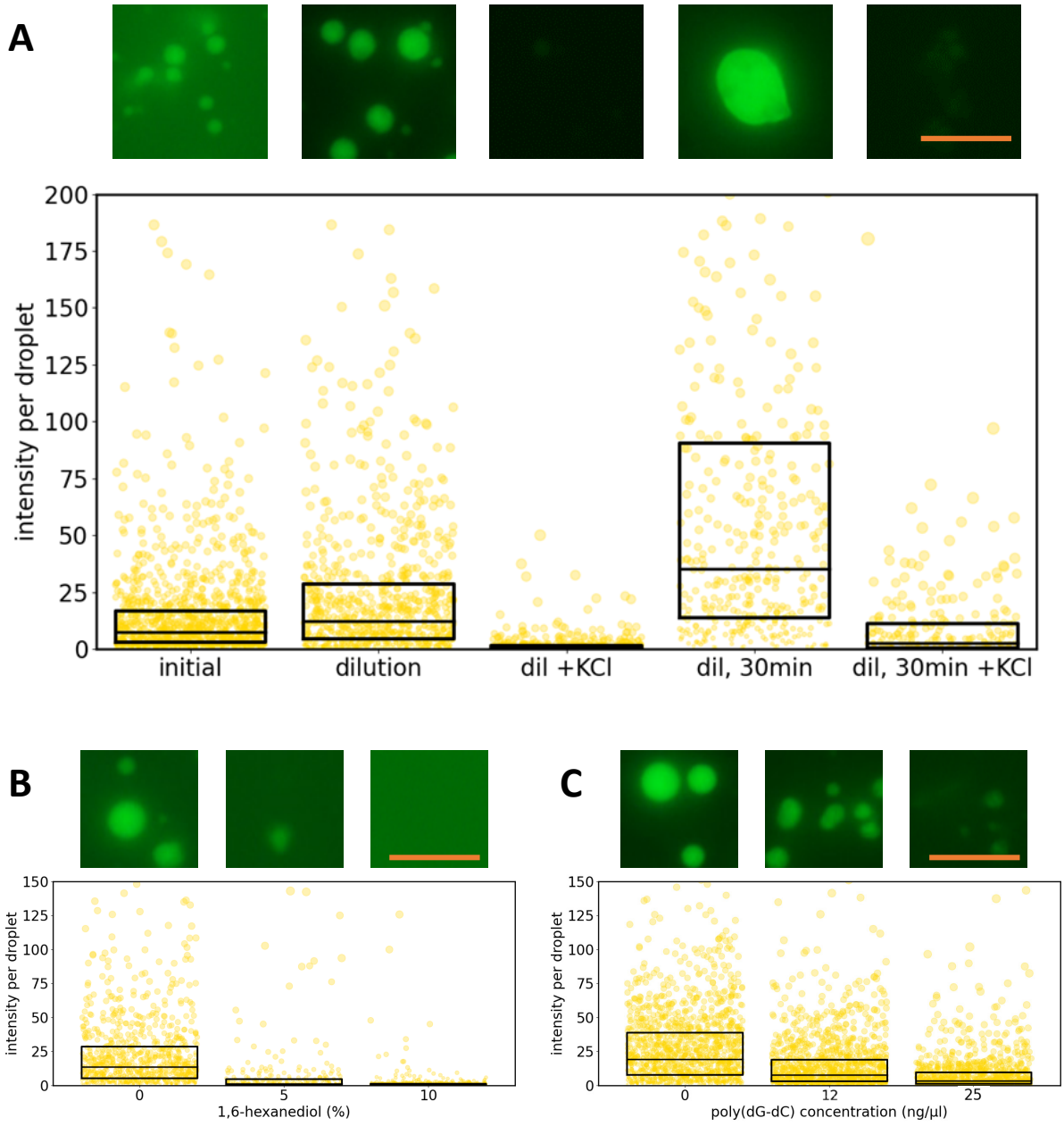


Fig. 14. *Hs*TBP-IDR-YFP condensates are dissolved with high salt concentrations and inhibited by 1,6-hexanediol and dsDNA. (**A**) Salt reversibility assay for *Hs*TBP-IDR-YFP. Condensates were formed at 60 µM protein, 150 mM KCl, 20% PEG (*initial*), followed by a 1:1 dilution in the same buffer (*dil*) or to 500 mM KCl (*dil +KCl*), either immediately or after a 30 min incubation (*30min*). *Top*: representative images of droplets at each indicated condition; *bottom*: integrated intensity of pixel brightness per droplet. (**B**) *Hs*TBP-IDR-YFP droplet formation is sensitive to 1,6-hexanediol. Condensates were formed at 60 µM protein, 150 mM KCl and 20% PEG in the absence or presence of 5% and 10% (w/v) 1,6-hexanediol. (**C**) *Hs*TBP-IDR-YFP droplet formation in response to poly(dG-dC) dsDNA at 60 µM protein, 75 mM KCl and 20% PEG. *Top*: representative images of condensates in the presence of indicated amounts of 1,6-hexanediol (B) or poly(dG-dC) DNA (C); *bottom*: integrated intensity of pixel brightness per droplet across 10 replicate fields of view per condition; box limits denote upper and lower quartiles and the centre line denotes the median value. *Scale bars:* 10 µm.

While *Hs*TFIIAαβ-IDR was initially shown to form irregular aggregates in condensation assays[81], Figs. 16–17 clearly show that it forms spherical assemblies under specific conditions in a concentration-dependent manner (Fig. 16A). Interestingly, these assemblies also exhibit a clustering phenotype upon prolonged incubation (Fig. 16A). *Hs*TFIIAαβ-IDR generates similar condensed fraction curves at 50 and 150 mM KCl, but droplet morphology and clustering is affected by buffer ionic strength (Fig. 16A). In salt reversibility assays, droplet size is somewhat reduced by dilution with isotonic solution, but fully dissolved at 500 mM KCl (Fig. 16B). Thus, *Hs*TFIIAαβ-IDR condensates have clearly liquid-like properties and are dynamic in nature. Furthermore, *Hs*TFIIAαβ-IDR condensate formation is negatively influenced by both 1,6-hexanediol and poly(dG-dC) dsDNA (Fig. 17). Droplet intensity is significantly reduced by 10% 1,6-hexanediol, but it does not completely abolish droplet formation (Fig. 17A). Similarly, addition of 125 ng/µl dsDNA leads to the assembly of significantly smaller and less distinct droplets (Fig. 17B).

Taken together, these results indicate that the IDRs of both *Pf*- and *Hs*TFIIAαβ can drive biomolecular condensation, but at different protein concentrations. Because protein preparations of *Hs*TFIIAαβ-IDR-GFP are of significantly lower concentration than *Pf*TFIIAαβ-IDR-GFP, titration curves cover different ranges. Despite the lower concentrations, *Hs*TFIIAαβ-IDR achieves higher condensed fraction measures, suggesting it drives droplet formation more strongly than *Pf*TFIIAαβ-IDR (compare Figs. 15A and 16A). The condensation of both *Hs*TFIIAαβ-IDR and *Pf*TFIIAαβ-IDR fusion proteins is moderately affected by the aliphatic alcohol 1,6-hexanediol and *Hs*TFIIAαβ-IDR droplet formation is reduced by dsDNA.

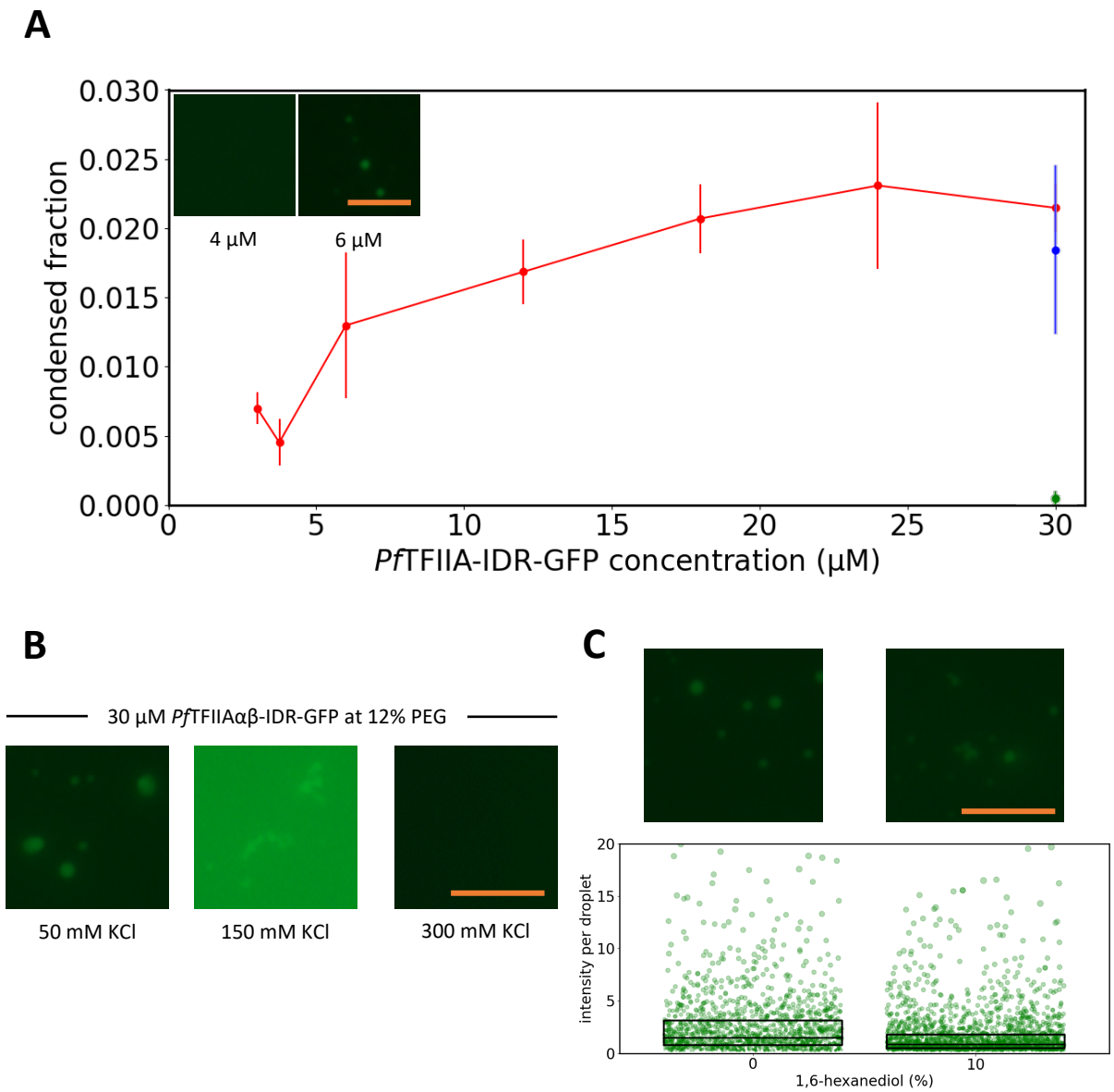


Fig. 15. *Pf*TFIIAαβ-IDR-GFP condensate formation is sensitive to high ionic strength but resistant to 1,6-hexanediol. (**A**) *Pf*TFIIAαβ-IDR-GFP forms condensates at 50 mM, but not 150 or 300 mM KCl. Condensed fraction curves for *Pf*TFIIAαβ-IDR-GFP in response to increasing protein concentration at 12% PEG and 50 mM KCl (*red*), 150 mM KCl (*blue*, single point) or 300 mM KCl (*green*, single point). *Inset: Pf*TFIIAαβ-IDR-GFP at 4 and 6 µM protein and 50 mM KCl. *Pf*TFIIAαβ-IDR-GFP does not form condensates at 4 µM. Condensed fraction points represent the mean ± SD of 10 images for each condition (**B**) Representative images of condensates at 30 µM protein and indicated KCl concentrations. (**C**) *Pf*TFIIAαβ-IDR-GFP droplet formation is resistant to 1,6-hexanediol. Condensates were formed at 30 µM protein, 50 mM KCl, 12% PEG in the absence or presence of 10% (w/v) 1,6-hexanediol. *Top*: representative fluorescence microscopy images; *bottom*: integrated intensity of pixel brightness per droplet; box limits denote upper and lower quartiles and the centre line denotes the median value. *Scale bars:* 10 µm.

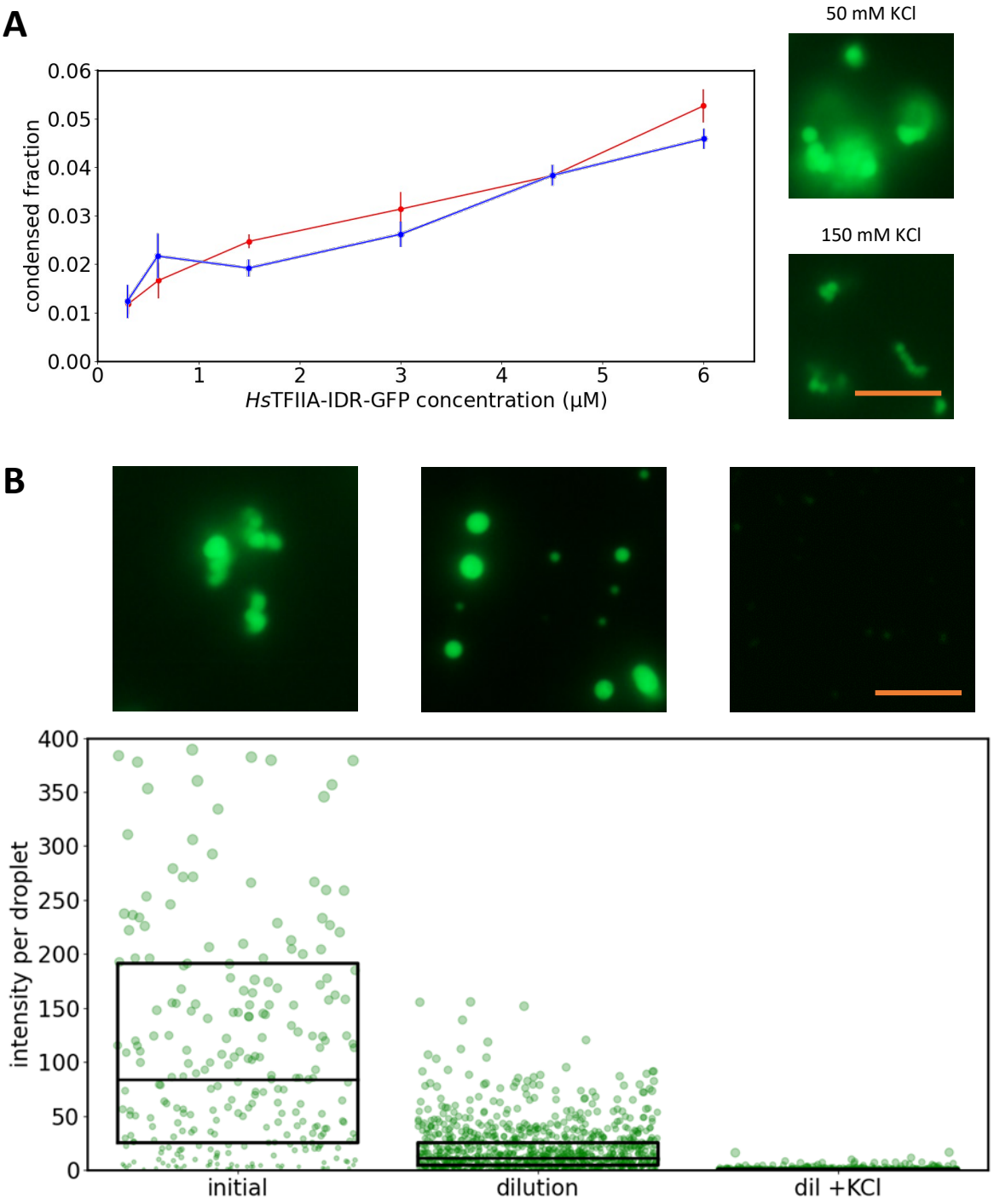


Fig. 16. *Hs*TFIIAαβ-IDR-GFP forms spherical condensates sensitive to high ionic strength. (**A**) *Hs*TFIIAαβ-IDR-GFP condensates form clusters at 50 mM and 150 mM KCl. *Left:* condensed fraction curves for *Hs*TFIIAαβ-IDR-GFP in response to increasing protein concentration at 12% PEG and 50 mM KCl (*red*) or 150 mM KCl *(blue); right*: representative images of condensates at 6 µM protein and indicated KCl concentrations. Condensed fraction points represent the mean ± SD of 10 images for each condition. (**B**) Salt reversibility assay for *Hs*TFIIAαβ-IDR-GFP. Condensates were formed at 6 µM protein, 50 mM KCl, 12% PEG (*initial*), followed by a 1:1 dilution with the same buffer (*dil*) or to 500 mM KCl (*dil +KCl*). *Top*: representative images of droplets at each indicated condition; *bottom*: integrated intensity of pixel brightness per droplet. Box limits denote upper and lower quartiles and the centre line denotes the median value. *Scale bars*: 10 µm.

*Hs*TFIIFß-IDR forms large spherical assemblies with liquid properties, whilst *Pf*TFIIFß-IDR assembles into irregular aggregate

The TFIIFß protein also contains internal IDRs in both *H. sapiens* and *P. falciparum,* which were previously cloned and fused to cyan fluorescent protein (CFP) (Fig. M1)[81]. Previous work demonstrated that *Pf*TFIIFß-IDR forms irregular aggregates in condensation assays, in the presence of 20% PEG-20000[81]. Lower concentrations of the lower molecular mass crowder PEG-3000 were investigated here, but assemblies remain irregular under these conditions (Fig. 18A). It should be noted that the assemblies formed by *Pf*TFIIFß-IDR are often too small to unequivocally determine their shape and statements regarding sphericity cannot be made. Interestingly, these assemblies do not form in the presence of 125 ng/µl poly(dG-dC) DNA, indicating an influence of environmental conditions (Fig. 18B).

In contrast to the *P. falciparum* fusion protein, *Hs*TFIIFß-IDR exhibits properties consistent with liquid-like condensate formation. Condensates form at 50 mM KCl and 12% PEG in a concentration-dependent manner and with a clear phase boundary. Below concentrations of 3 µM, the protein remains evenly distributed in solution, even in the presence of PEG (Fig. 19A). Higher protein concentrations lead to droplet formation and the condensed fraction increases as a function of protein concentration (Fig. 19A). Condensates do not form at 150 mM KCl (Fig. 19A), and droplet size is dramatically reduced by 5 and 10% of 1,6-hexanediol (Fig. 19B). Thus, *Hs*TFIIFß-IDR condensates are sensitive to both buffer ionic strength and the presence of 1,6-hexanediol.

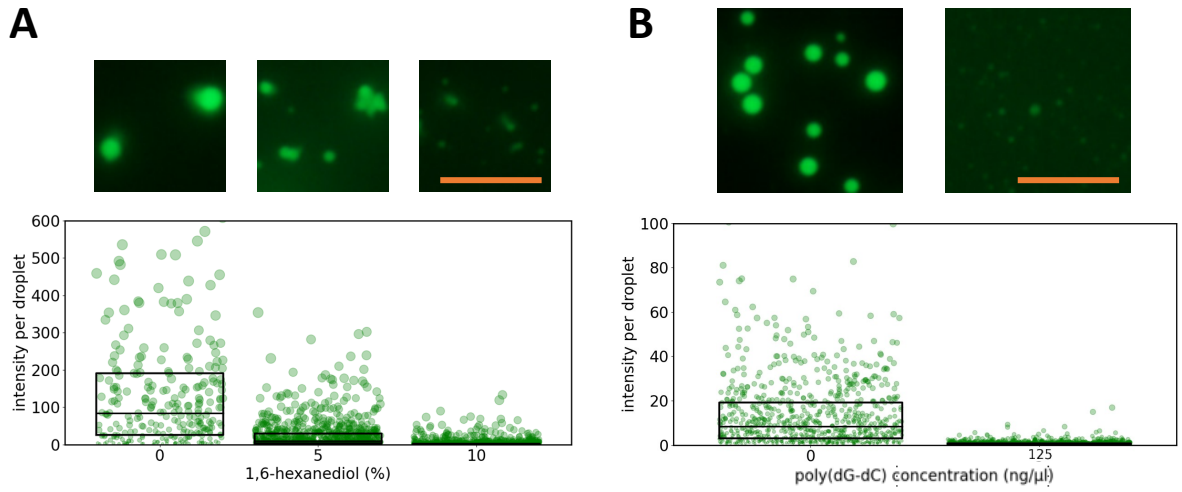


Fig. 17. *Hs*TFIIAαß-IDR-GFP condensate formation is moderately resistant to challenge with 1,6-hexanediol and strongly inhibited by dsDNA. (A) *Hs*TFIIAαβ-IDR-GFP droplet formation in response to 1,6-hexanediol at 6 µM protein, 50 mM KCl and 12%. (B) *Hs*TFIIAαβ-IDR-GFP droplet formation in response to poly(dG-dC) dsDNA at 6 µM protein 25 mM KCl and 12% PEG. *Top*: representative images of condensates in the absence or presence of 5% or 10% hexanediol (A) or 125 ng/µl poly(dG-dC) dsDNA; *bottom*: integrated intensity of pixel brightness per droplet. Box limits denote upper and lower quartiles and the centre line denotes the median value. *Scale bars*: 10 µm.

The results of salt reversibility assays suggest that *Hs*TFIIFß-IDR condensates are highly dynamic, as they dissolve upon challenge with 300 mM KCl after droplet formation (Fig. 20). Furthermore, the data indicate that this dynamism is maintained, even after incubating condensation reactions for 30 min (Fig. 20). *Hs*TFIIFß-IDR condensates fuse quickly upon contact, resulting in the formation of a larger body that relaxes to a spherical shape (Fig. 21B), providing further evidence for their liquid properties. However, droplets bleached in FRAP experiments do not fully recover over a 5 min time period (Fig. 21A). Thus, although droplets appear to be overall liquid in nature, they do not show rapid internal dynamics, indicating static behaviour of at least a proportion of fluorescent molecules. This behaviour is reminiscent of MED1-IDR condensates which also show evidence of liquidity, such as droplet fusing and reversibility of condensate formation, but recover slowly in FRAP assays (Fig. 10A).

Taken together, both *P. falciparum* and *H. sapiens* TFIIFß-IDRs display complex behaviours in condensation assays but clearly exhibit distinct condensation properties. Under the experimental conditions, PfTFIIFß-IDR forms static, irregular aggregates, whilst *Hs*TFIIFß-IDR condenses reversibly into spherical droplets that fuse upon contact.

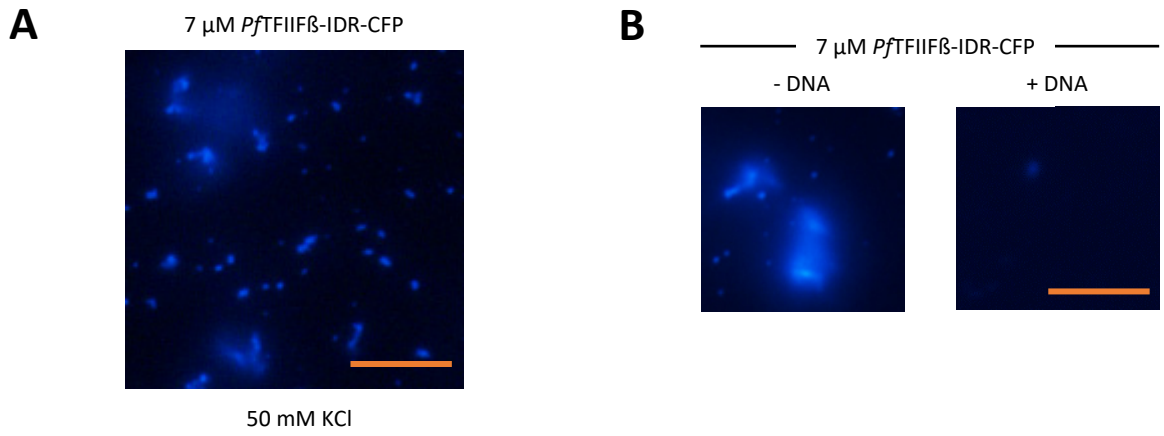


Fig. 18. *Pf*TFIIFß-IDR-CFP forms irregular aggregates in the absence but not in the presence of dsDNA. (**A**) 7 µM *Pf*TFIIFß-IDR-CFP in the presence of 12% PEG, 50 mM KCl forms small, irregular aggregates. (**B**) Representative images of *Pf*TFIIFß-IDR-CFP condensation reactions in the absence or presence of 125 ng/µl poly(dG-dC) dsDNA at 7 µM protein, 25 mM KCl and 12% PEG. *Scale bars:* 10 µm.

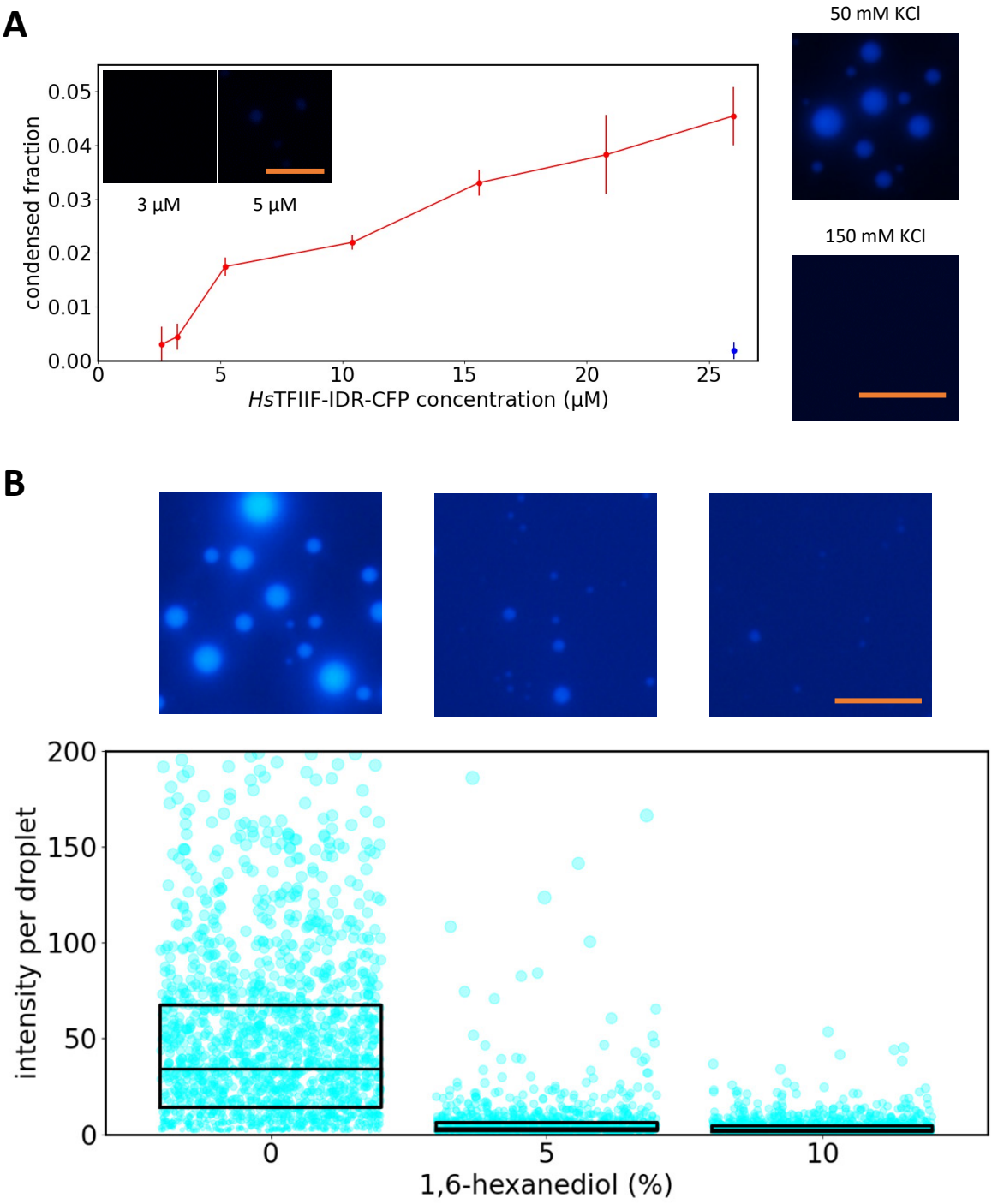


Fig. 19. *Hs*TFIIFß-IDR-CFP forms spherical condensates moderately sensitive to 1,6-hexanediol. (A) *Hs*TFIIFß-IDR-CFP forms condensates at 50 mM but not 150 mM KCl. *Left:* condensed fraction curves for *Hs*TFIIFß-IDR-CFP in response to increasing protein concentration at 12% PEG and 50 mM KCl (*red*) and 150 mM KCl (*blue*). Condensed fraction points represent the mean ± SD of 10 images for each condition; *inset: Hs*TFIIFß-IDR-CFP condensation at 3 and 5 µM protein and 50 mM KCl. *Hs*TFIIFß-IDR-CFP does not form condensates at 3 µM; *right*: representative images of condensates at 26 µM protein and indicated KCl concentrations. (B) *Hs*TFIIFß-IDR-CFP condensate formation at 26 µM protein, 50 mM KCl and 12% PEG in the absence or presence of 5% or 10% 1,6-hexanediol. *Top*: representative fluorescence microscopy images; *bottom*: integrated intensity of pixel brightness per droplet. Box limits denote upper and lower quartiles and the centre line denotes the median value. *Scale bars*: 10 µm.

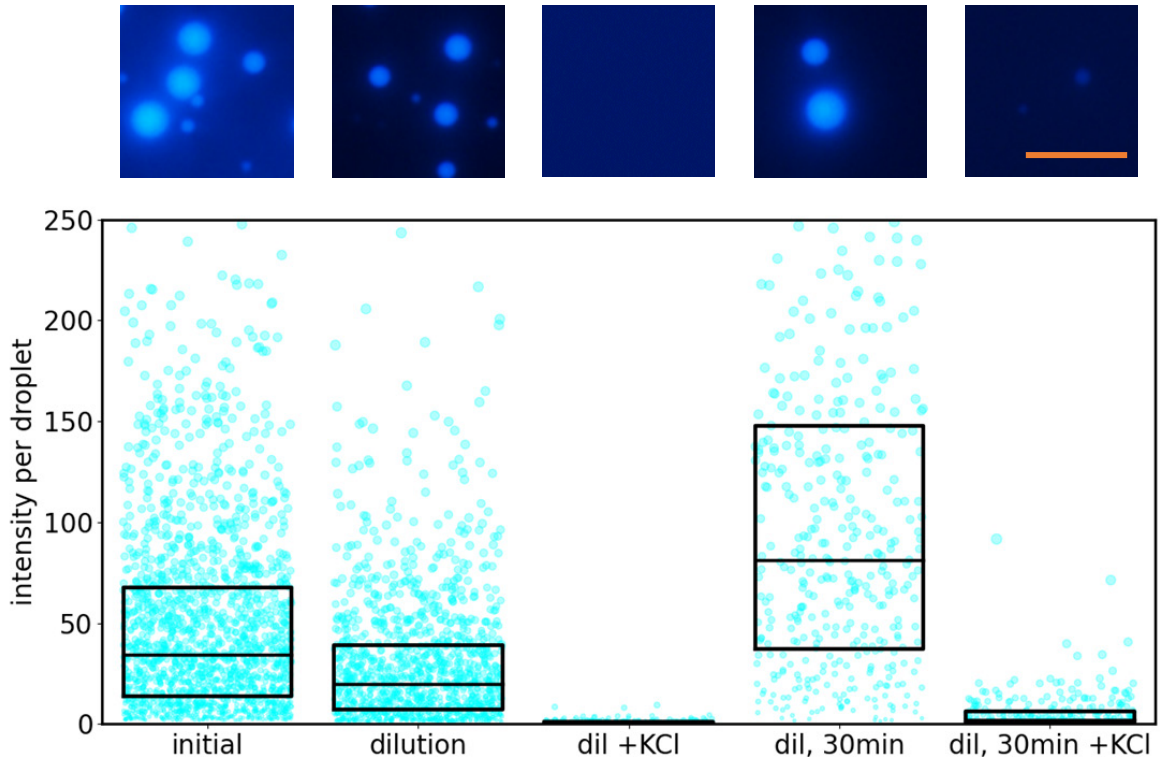


Fig. 20. *Hs*TFIIFß-IDR-CFP droplet formation is reversible. Condensates were formed with 26 µM protein, 50 mM KCl, 20% PEG (*initial*), followed by 1:1 dilution in the same buffer (*dil*) or to 300 mM KCl (*dil +KCl*), either immediately or after a 30 min incubation (*30min*). *Top*: representative images of droplets at each indicated condition; *bottom*: total pixel intensity per droplet; box limits denote upper and lower quartiles and the centre line denotes the median value. *Scale bar:* 10 µm.

VP16-AD forms assemblies only in the presence of high concentrations of PEG

The VP16-AD is known to be one of the most potent activators of transcription ever studied[94,104]. Given the emerging link between biomolecular condensation and transcription regulation, and the observation that IDRs of multiple activators of transcription can form condensates[70,71], it was investigated if the potent VP16-AD transcription activation domain can undergo condensation. Under standard biomolecular condensation conditions (12% PEG-3000, 150 mM KCl), in which most other fusion proteins readily form condensates, mC-VP16-AD remains evenly distributed in solution (Fig. 22B). Even at 46 µM protein concentration and 50 mM KCl the presence of 20% PEG-3000 is required to induce the formation of spherical condensates (Fig. 22B). These data suggest that mC-VP16-AD is not a powerful driver of biomolecular condensation, and only forms notable assemblies at very high concentrations of PEG-20000 or low ionic strength (Fig. 22A). In a similar fashion to other fusion proteins under investigation, mC-VP16-AD condensation is inhibited by high concentrations of poly(dG-dC) dsDNA (Fig. 22C).

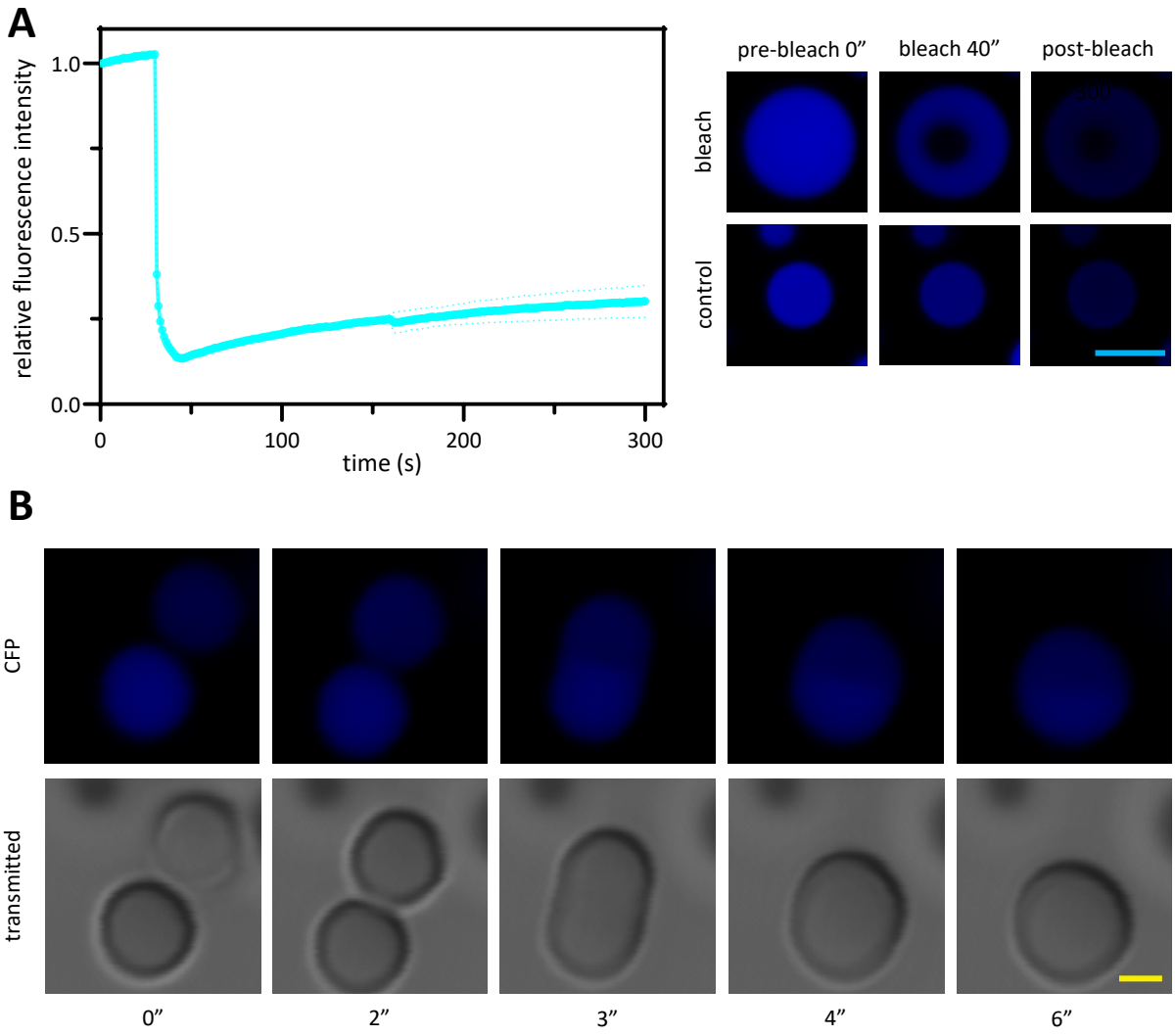


Fig. 21. *Hs*TFIIFß-IDR-CFP condensates partially recover from photobleaching and fuse to form larger condensates. (A) Fluorescence recovery after photobleaching of *Hs*TFIIFß-IDR-CFP droplets at 26 µM protein, 50 mM KCl and 12% PEG. Condensation reactions were visualised immediately after condensate formation. Following 30 s of a pre-bleaching period, regions were bleached to 10% of their original intensity and monitored for a total of 5 min. *Left*: relative fluorescence intensity is defined as average pixel intensity of bleached region normalised to region of similar size in unbleached droplet to account for background bleaching. Each point is normalised as a fraction of the initial reading; *right*: representative images of bleached and control droplets at indicated time points. All points are the mean ± SD (dotted line) for 3 replicate bleaching experiments. (B) *Hs*TFIIFß-IDR-CFP droplets fuse upon contact. *Top:* droplet fusion visualised by confocal fluorescent microscopy; *bottom:* the same fusion event, observed under the transmitted light channel. *Scale bars*, *cyan:* 5 µm; *yellow:* 1 µm.

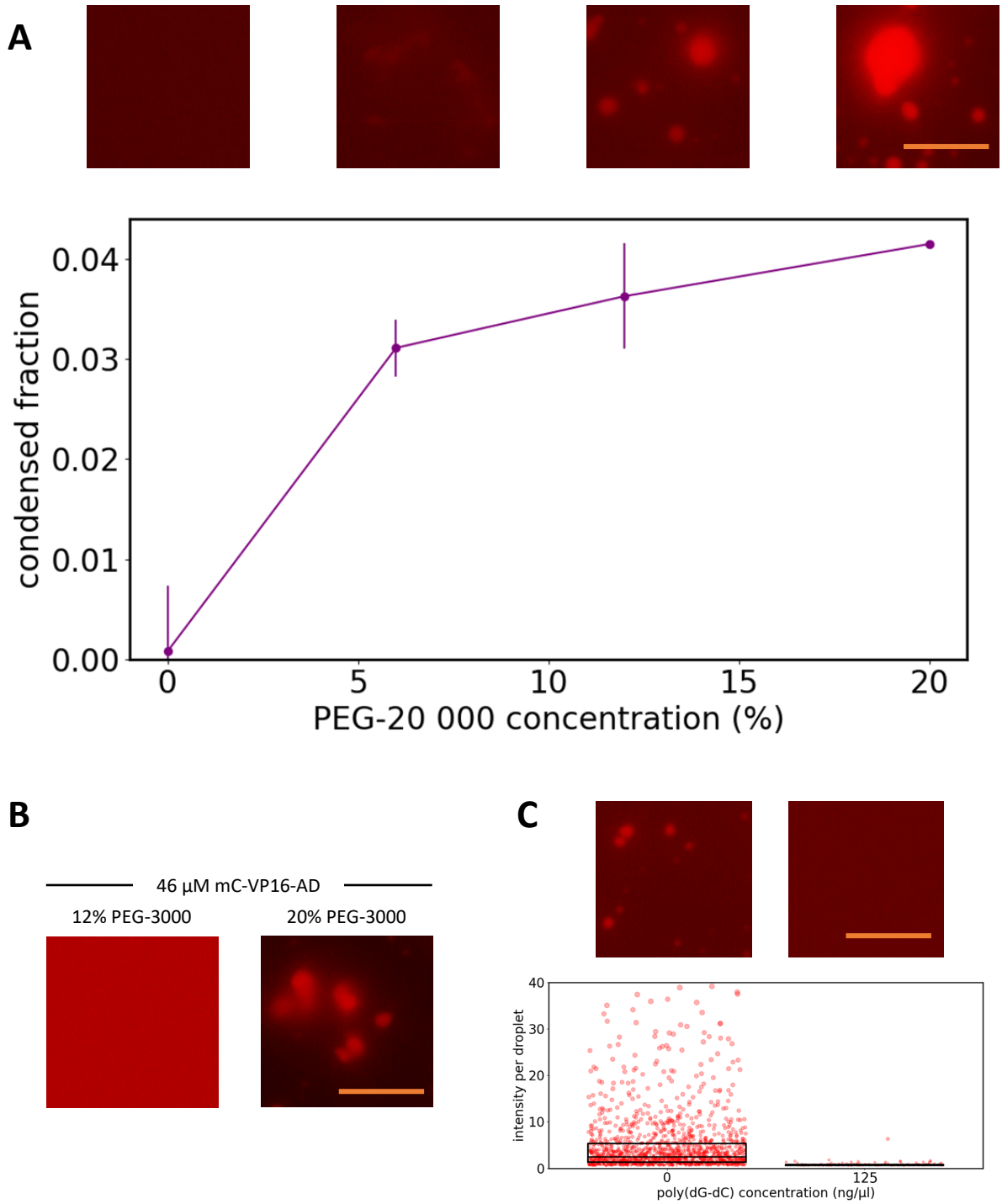


Fig. 22. mC-VP16-AD only forms condensates in the presence of high concentrations of PEG. (**A**) VP16-AD forms condensates with PEG-20000. *Top*: representative images of droplets at 0%, 6%, 12% and 20% PEG-20000; *bottom:* condensed fraction curve for mC-VP16-AD in response to increasing concentrations of PEG-20000 at 46 µM protein and 50 mM KCl. Condensed fraction points represent the mean ± SD of 10 images for each condition. (**B**) mC-VP16-AD forms condensates at 20%, but not at 12% PEG-3000 in the presence of 50 mM KCl. (**C**) mC-VP16-AD droplet formation in the absence and presence of 125 ng/µl poly(dG-dC) dsDNA at 46 µM protein, 25 mM KCl and 12% PEG-20000. *Top*: representative fluorescence microscopy images; *bottom*: integrated intensity of pixel brightness per droplet across 10 replicate fields of view per condition. Box limits denote upper and lower quartiles and the centre line denotes the median value. *Scale bars:* 10 µm.

*Hs*RPB1-CTD forms assemblies only in the presence of high concentrations of PEG

Existing literature on the condensation properties of the *H. sapiens* RPB1-CTD appears inconsistent and somewhat contradictory. Some results suggest that the *Hs*RPB1-CTD is capable of driving biomolecular condensation on its own [46], whereas work from two other research groups suggests that it is only recruited to condensates pre-formed by other transcriptional proteins such as the PTEFb[74] or Mediator complex[75]. This study sought to re-examine whether the *Hs*RPB1-CTD fusion protein is able to form condensates. As shown in Fig. 23A, in the presence of 20% PEG-20000, mEGFP-*Hs*RPB1-CTD forms small speckles at a concentration of 18 µM, and at 70 µM forms spherical condensates. Under the standard conditions with 12% PEG-3000, mEGFP-*Hs*RPB1-CTD remains initially evenly distributed in solution but, over time, forms hazy aggregates that settle onto the coverslip (Fig. 23B). In contrast to the *Pf*RPB1-CTD (Fig. 3), these data suggest that the *Hs*RPB1-CTD is not a potent driver of biomolecular condensation, requiring highly permissive conditions to form spherical condensates. It should be noted that the full-length mEGFP-*Hs*RPB1-CTD fusion protein proved difficult to purify and is only fractionally present in the final protein preparations (Fig. S4).

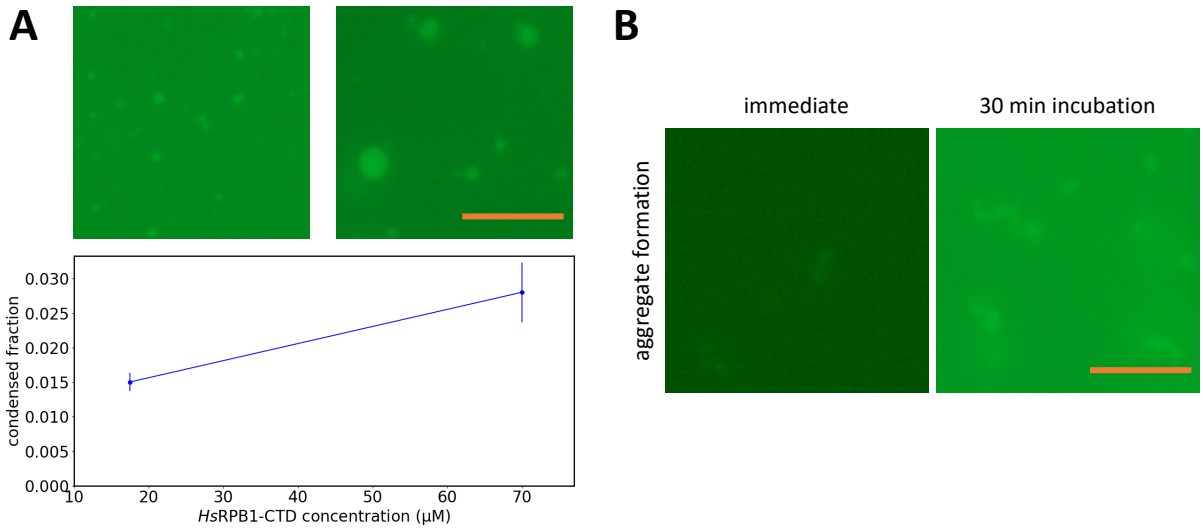


Fig. 23. *Hs*RPB1-CTD forms condensates in the presence of high concentrations of PEG. (A) *Hs*RPB1-CTD forms condensates with PEG-20000. *Top*: representative images of droplets at 18 and 70 µM protein, 150 mM KCl and 20% PEG-20000; *bottom:* condensed fraction curve for *Hs*RPB1-CTD. Condensed fraction points represent the mean ± SD of 10 images for each condition. (B) *Hs*RPB1-CTD at 70 µM does not form spherical condensates under standard condensation reaction conditions (12% PEG-3000, 150 mM KCl). Fusion protein fluorescence is relatively evenly distributed upon immediate visualisation and condenses into irregular aggregates during 30 min incubation. *Scale bars:* 10 µm.

MED1-IDR and *Pf*RPB1-CTD form incompletely mixed assemblies

Following the characterisation of the conditions that promoted biomolecular condensation of individual fusion proteins, further work addressed the potential interactions between different proteins in mixing assays.

Initial trial experiments revealed an interesting interaction between MED1-IDR and *Pf*RPB1-CTD, which was investigated in more detail. Two approaches to mixing phase-separating components were investigated: (i) proteins were first combined, followed by the addition of PEG-containing condensation reaction buffer, or (ii) condensation of protein components was induced separately, and resultant condensates combined. Both approaches were investigated combining a range of fluorescent protein fusion constructs to study their interactions.

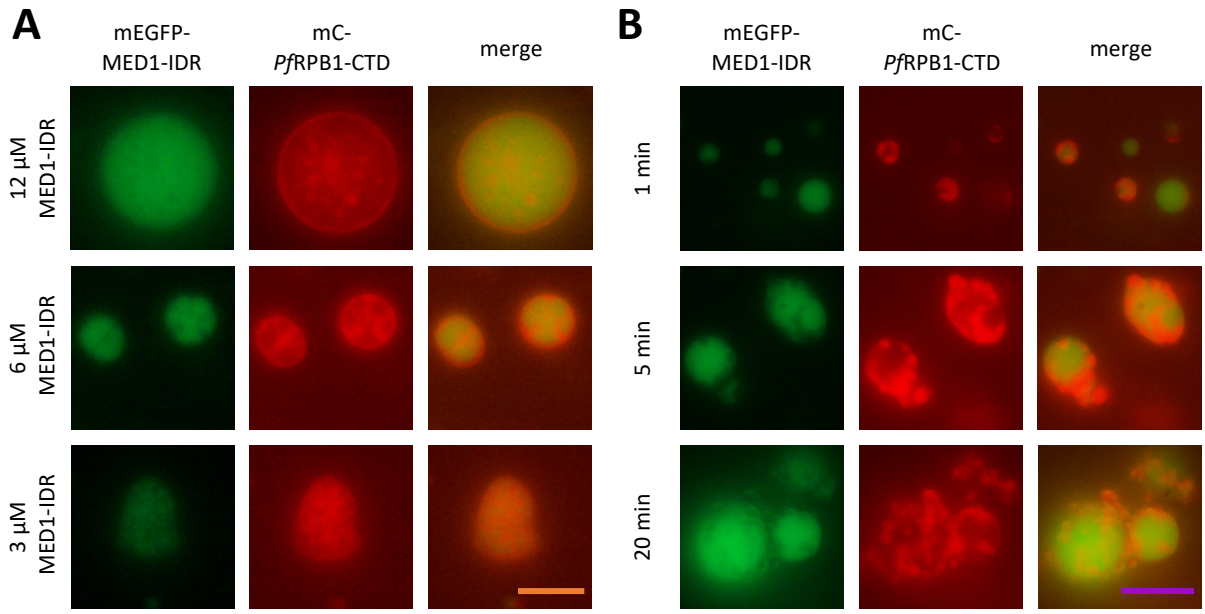


Fig. 24. mEGFP-MED1-IDR and mC-*Pf*RPB1-CTD form incompletely mixed assemblies. (A) Different concentrations of mEGFP-MED1-IDR are mixed with 14 µM mC-*Pf*RPB1-CTD in standard condensation reaction buffer containing 12% PEG-3000 and 150 mM KCl. The two fusion proteins were first thoroughly mixed prior to addition of condensation reaction buffer. (B) mEGFP-MED1-IDR and mC-*Pf*RPB1-CTD were added separately to condensation reaction buffer and incubated for 1, 5 or 20 min to induce biomolecular condensation before being combined and analysed. *Scale bars*, *orange:* 10 µm; *purple:* 20 µm.

As shown in Fig. 24A, mEGFP-MED1-IDR and mC-*Pf*RPB1-CTD form incompletely mixed assemblies with a heterogenous distribution of MED1-IDR and *Pf*RPB1-CTD. Notably, these assemblies show a distinct distribution of the two components: the majority of *Pf*RPB1-CTD appears concentrated on the surface of mixed assemblies, whereas the bulk of the internal volume of the mixed droplets is occupied by MED1-IDR, together with smaller *Pf*RPB1-CTD droplets that are clearly separated from MED1-IDR (Fig. 24A). The distribution of *Pf*RPB1-CTD within these mixed assemblies is influenced by the concentration ratio of the two components (Fig. 24A). At lower MED1-IDR concentrations, the bulk of droplets are occupied by *Pf*RPB1-CTD and smaller MED1-IDR condensates sit within these (Fig. 24A). mC-*Pf*RPB1-CTD brightness peaks at the border of droplets as well as distinct internal locations. mEGFP-MED1-IDR brightness can be measured throughout the internal regions of droplets but drops concurrently with peaks of mC-*Pf*RPB1-CTD (Fig. 25A). The distinct localisation of the two components in mixed assemblies is further confirmed by the analysis of confocal microscopy images (Fig. 25A).

When mEGFP-MED1-IDR and mC-*Pf*RPB1-CTD are mixed after separately inducing condensation of the two components, the heterogenous assemblies take on a fundamentally different appearance (Fig. 24B). mC-*Pf*RPB1-CTD no longer localises to internal regions of mEGFP-MED1-IDR condensates, but rather clusters outside of MED1-IDR containing droplets. These mixed assemblies form, over time, large, multi-droplet clusters, which are static in nature (Fig. 24B). Longer individual incubation times for each component lead to larger resulting droplet assemblies when the reactions are combined (Fig. 24B). A three-dimensional reconstruction of a mixed MED1-IDR/*Pf*RPB1-CTD condensation reaction, shown in Fig. 25B, clearly demonstrates the distinct localisation of both components in mixed assemblies.

The presence of *Pf*RPB1-CTD also affects the material properties of MED1-IDR droplets and interferes with the ability of MED1-IDR droplets to fuse (Fig. 26). Within the 30 min incubation period, *Pf*RPB1-CTD localised on the surface of mixed assemblies results in arrest during fusion events between individual spherical droplets (Fig. 26A), resulting in the formation of large droplet clusters (Fig. 26B). These properties, of stalled droplet fusion and cluster formation, are not generally seen for condensates of just MED1-IDR (Fig. 7), suggesting that the mixture with *Pf*RPB1-CTD changes droplet properties.

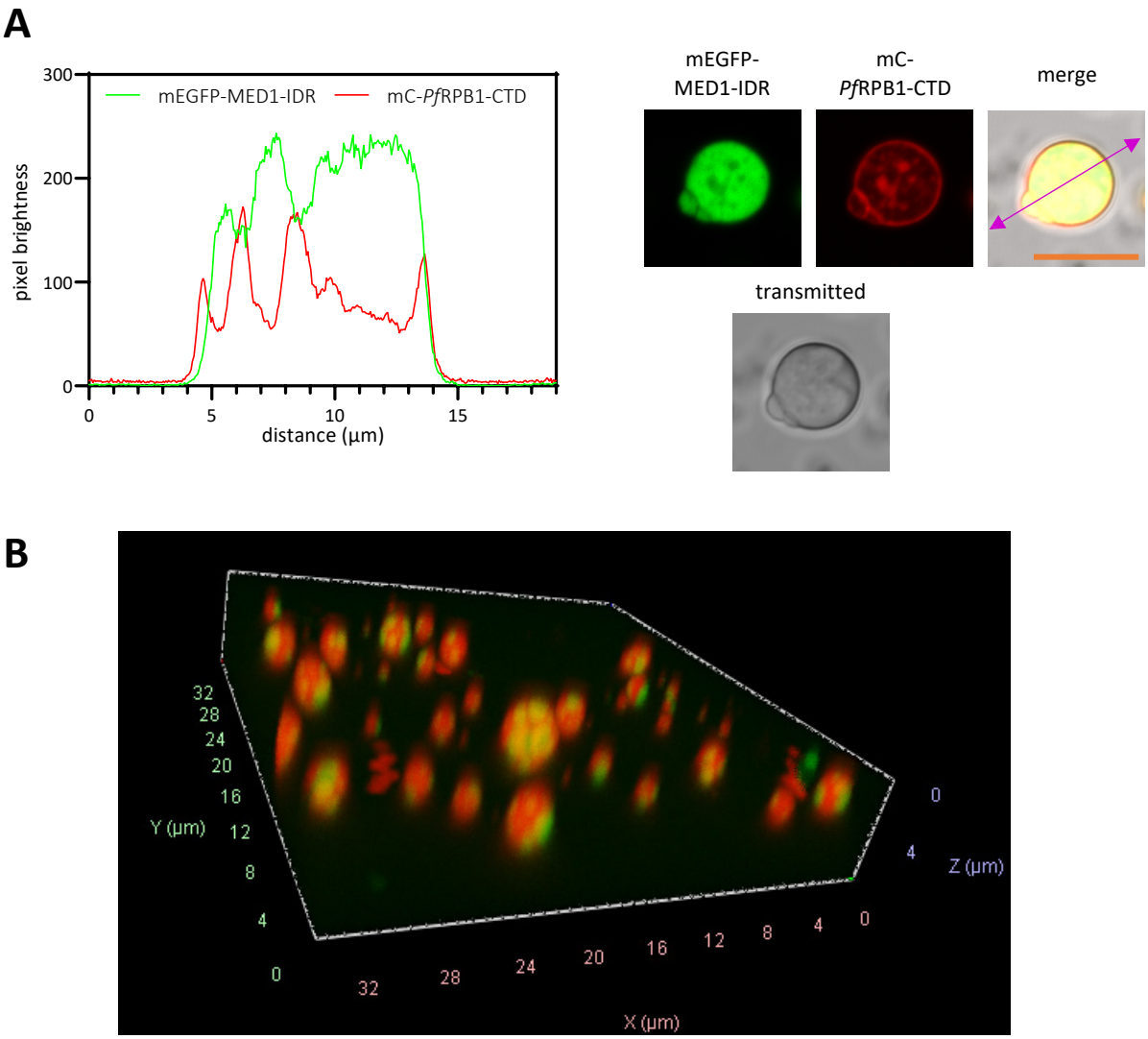


Fig. 25. mEGFP-MED1-IDR and mC-*Pf*RPB1-CTD occupy distinct regions in mixed assemblies. (**A**) mEGFP-MED1-IDR and mC-*Pf*RPB1-CTD occur in mutually exclusive regions of incompletely mixed droplets. *Left:* Brightness values of each pixel along a cross-section of a mixed droplet. Pixel brightness for red (mC-*Pf*RPB1-CTD) and green (mEGFP-MED1IDR) channels was measured using the plot profile function in ImageJ; *right:* mixed mEGFP-MED1-IDR/mC-*Pf*RPB1-CTD droplet imaged by confocal fluorescent microscopy. Pink line in merged image delineates the plotted cross-section *left*. (**B**) Three-dimensional reconstruction of a MED1-IDR/*Pf*RPB1-CTD mixed condensation reaction from a series of z-stack images. *Scale bar:* 10 µm.

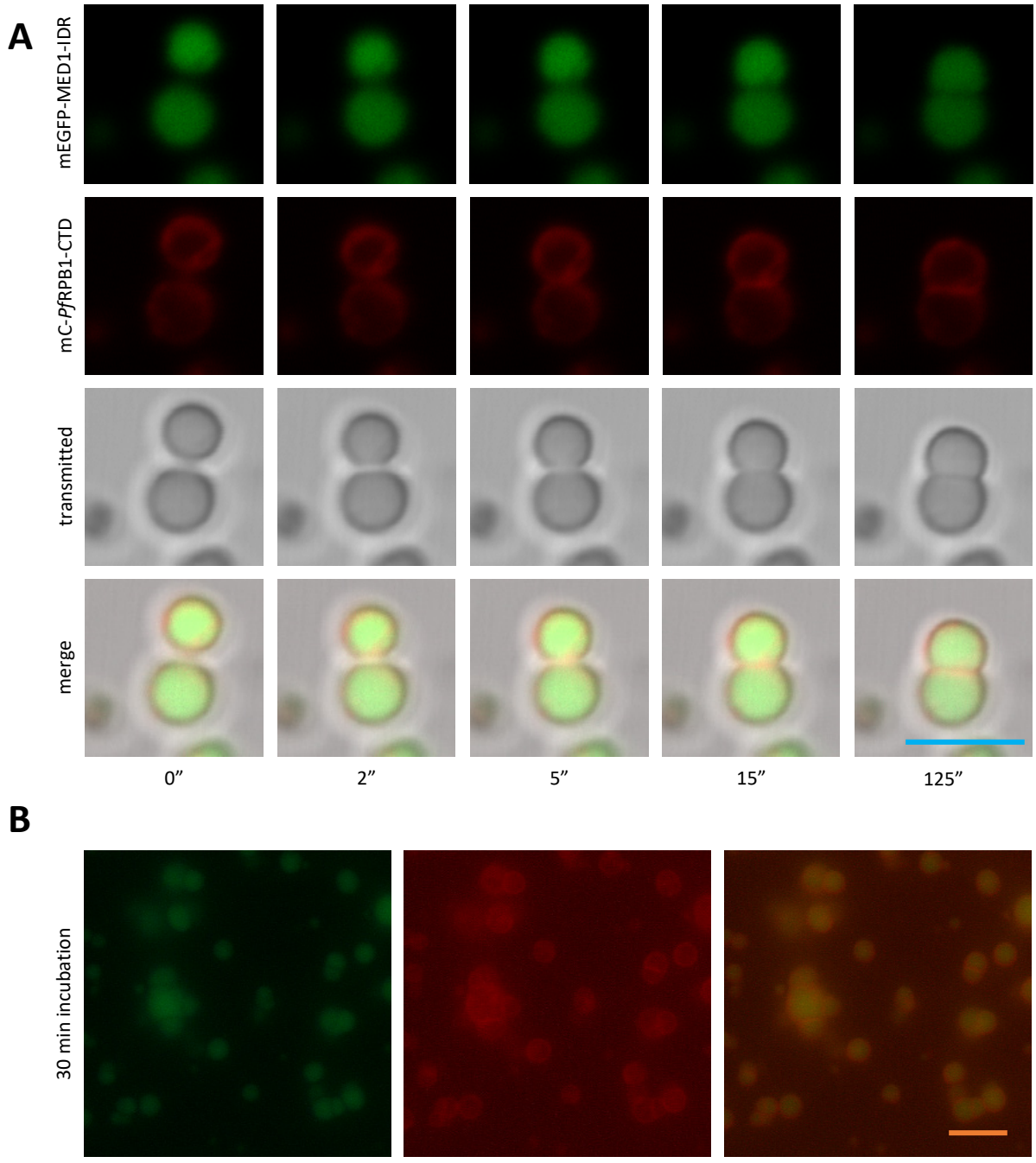


Fig. 26. mC-*Pf*RPB1-CTD coats mEGFP-MED1-IDR in mixed assemblies and prevents fusion. (**A**) Stalled fusion event of droplets containing mEGFP-MED1-IDR and mC-*Pf*RPB1-CTD formed in standard condensation reaction buffer containing 12% PEG-3000 and 150 mM KCl, visualised by confocal fluorescence microscopy. (**B**) Mixed assemblies of mEGFP-MED1-IDR and mC-*Pf*RPB1-CTD after 30 min incubation. Note extensive droplet clusters, generally not seen in MED1-IDR condensation reactions. *Scale bars*, *cyan:* 5µm; *orange:* 10 µm.

MED1-IDR recruits *Hs*RPB1-CTD to homogeneous mixed assemblies

In contrast to the incomplete mixing of MED1-IDR and *Pf*RPB1-CTD, MED1-IDR and *Hs*RPB1-CTD form homogenously mixed condensates (Fig. 27A). These data replicate previously published results that demonstrated the ability of MED1-IDR to recruit *Hs*RPB1-CTD to mixed condensates[75]. Interestingly, mixed condensate formation occurs in the presence of 12% PEG and 150 mM KCl, conditions under which *Hs*RPB1-CTD alone does not form condensates (Fig. 23B). Unlike heterogenous MED1-IDR/*Pf*RPB1-CTD droplet assemblies, these homogeneously mixed droplets retain a dynamic character and fuse upon contact to form larger spherical assemblies (Fig. 27B).

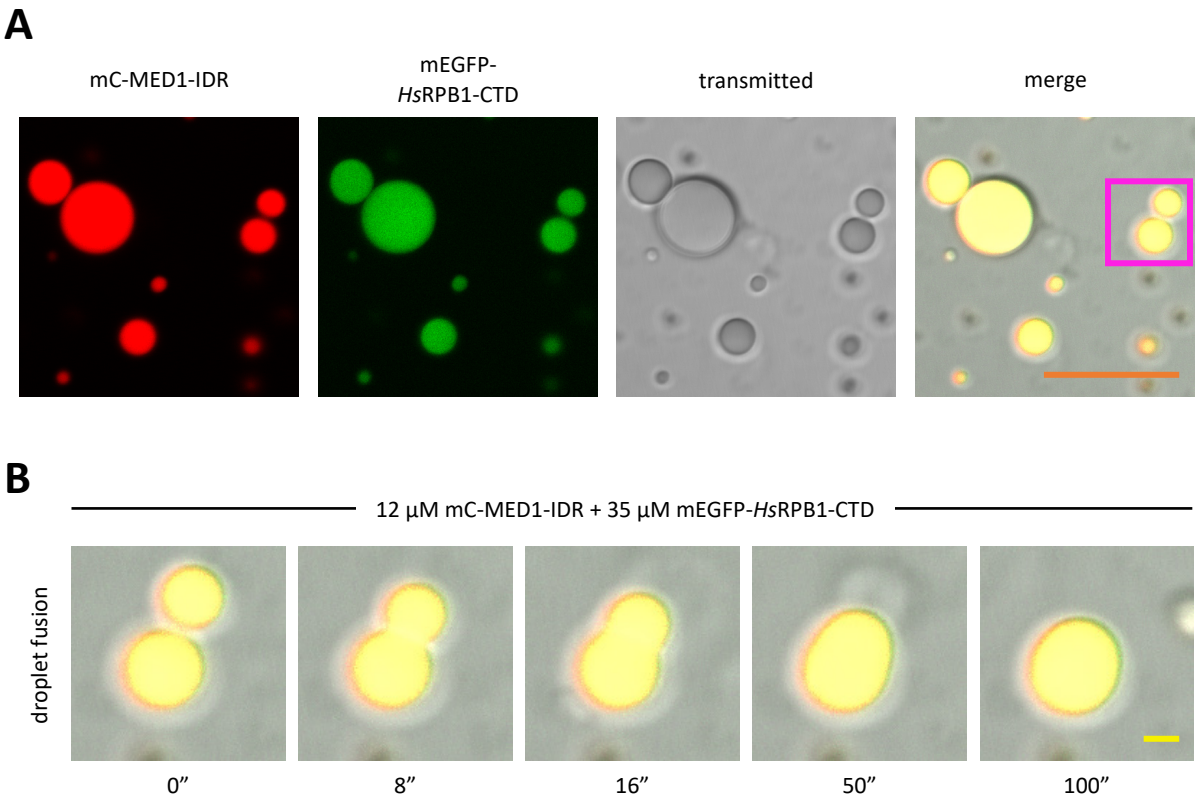


Fig. 27. mC-MED1-IDR recruits mEGFP-*Hs*RPB1-CTD to homogenously mixed assemblies, which maintain liquid properties. (A) mC-MED1-IDR recruits mEGFP-*Hs*RPB1-CTD to mixed droplets in standard condensation reaction buffer containing 12% PEG-3000 and 150 mM KCl. The two fusion proteins were first thoroughly mixed prior to addition of condensation reaction buffer. Mixed droplets visualised by confocal fluorescent microscopy in indicated channels. (B) Fusion event of mixed mC-MED1-IDR/mEGFP-*Hs*RPB1-CTD droplets. First frame corresponds to field highlighted in pink box in (A). *Scale bars*, *orange:* 10 µm; *yellow:* 1 µm.

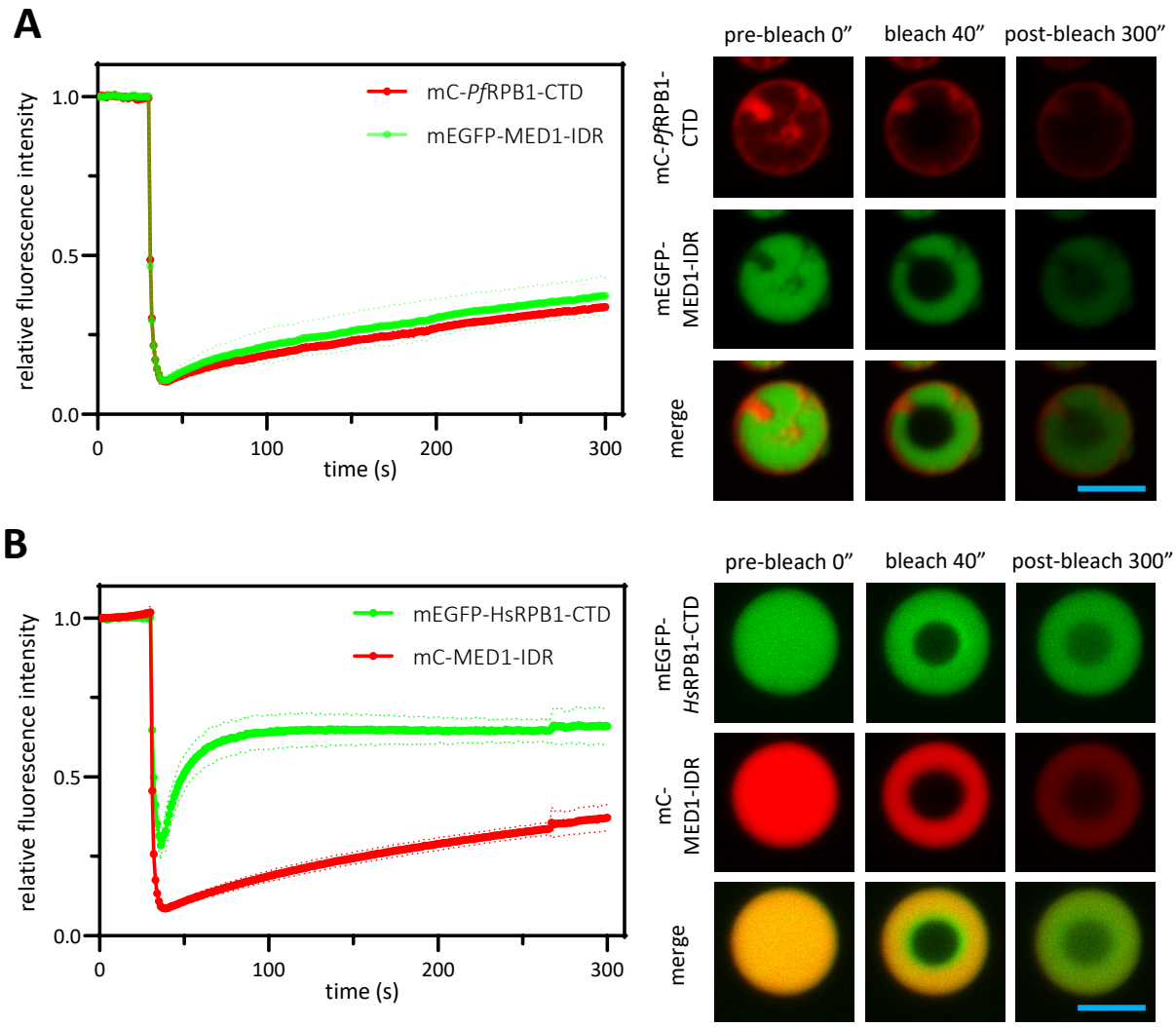


Fig. 28. *Pf*- and *Hs*RPB1-CTD exhibit different mobilities in mixed condensates with MED1-IDR. (**A**) Fluorescence recovery after photobleaching of mixed mC-*Pf*RPB1-CTD (14 µM) and mEGFP-MED1-IDR (12 µM) condensates at 150 mM KCl and 12% PEG. (**B**) Fluorescence recovery after photobleaching of mixed mEGFP-*Hs*RPB1-CTD (30 µM) and mC-MED1-IDR (12 µM) condensates at 150 mM KCl and 12% PEG. *Left*: FRAP curves for mixed assemblies; *right*: representative images of bleached droplets at indicated time points, visualised in red and green fluorescent channels. For both FRAP experiments, following 30 s of a pre-bleaching period, regions were bleached to 10% of their original intensity and monitored for a total of five min. Relative fluorescence intensity is defined as average pixel intensity of bleached region normalised to region of similar size in unbleached droplet to account for background bleaching. Each point is normalised as a fraction of the initial reading. All points are the mean ± SD (dotted line) for 3 replicate bleaching experiments. *Scale bars:* 5 µm.

Mixed assemblies formed with MED1-IDR and either *Pf*RPB1-CTD or *Hs*RPB1-CTD were probed by FRAP to investigate the mobility of individual components (Fig. 28). Both components of incompletely mixed MED1-IDR/*Pf*RPB1-CTD assemblies revealed static internal dynamics and very little recovery from photobleaching (Fig. 28A). This result is not unexpected given that MED1-IDR and *Pf*RPB1-CTD separate in mixed assemblies. Consistent with published results[75], MED1-IDR/*Hs*RPB1-CTD assemblies exhibit a divergent fluorescence recovery. The MED1-IDR fraction recovers slowly to around 40% of pre-bleaching levels, whilst the *Hs*RPB1-CTD fraction recovers much more quickly (Fig. 28B). This indicates that *Hs*RPB1-CTD molecules retain a far more dynamic character in mixed assemblies and can freely diffuse within condensates to a significant extent. On the other hand, *Hs*RPB1-CTD fluorescence recovery does reach a plateau at around 60% of pre-bleaching intensity. This would suggest that a fraction of *Hs*RPB1-CTD molecules engage in tight interactions with MED1-IDR in droplets that prohibit free diffusion.

Given the divergence in mixing behaviour between *Pf*- and *Hs*RPB1-CTD with MED1-IDR, the combability of both RPB1-CTD constructs with one another was briefly investigated. The data suggest that *Pf*RPB1-CTD, similar to MED1-IDR[75] (Fig. 27), is able to recruit *Hs*RPB1-CTD to droplets under conditions where *Hs*RPB1-CTD does not form condensates alone (Fig. 29). Judged by the degree of mEGFP brightness observed within droplets versus background fluorescence, as an estimate of relative concentrations, *Hs*RPB1-CTD recruitment into *Pf*RPB1-CTD droplets appears not as robust as *Hs*RPB1-CTD recruitment into MED1-IDR droplets (compare Fig. 29 with Fig. 27).

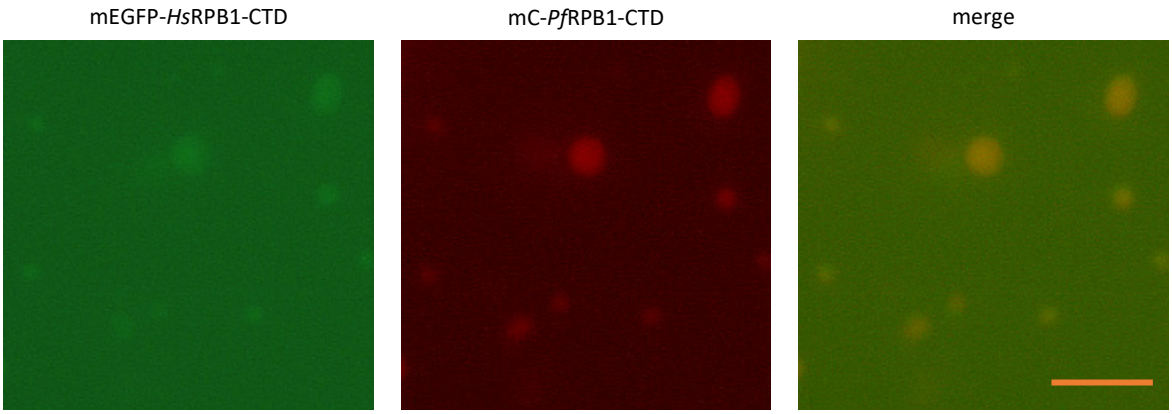


Fig. 29. *Pf*-RPB1-CTD recruits *Hs*RPB1-CTD to homogeneously mixed assemblies. 14 µM mC-*Pf*RPB1-CTD and 30 µM mEGFP-*Hs*RPB1-CTD were thoroughly mixed prior to addition of standard condensation reaction buffer containing 12% PEG-3000 and 150 mM KCl. *Scale bar:* 10 µm

TBP-IDRs form mixed assemblies with MED1-IDR and *Pf*RPB1-CTD

Having established that MED1-IDR and *Pf*RPB1-CTD act as powerful drivers of biomolecular condensation, further work turned to investigating their interaction with GTF-IDRs fused to fluorescent proteins.

As an important control, it was determined if fluorescent protein alone could be recruited to condensates formed by MED1-IDR and *Pf*RPB1-CTD. As shown in Figs. S6B–C, both TF-IDR fusion proteins form condensates whilst YFP remains evenly distributed throughout the field of view. Thus, fluorescent protein alone is not recruited into TF-IDR condensates.

*Pf*TBP-IDR, which forms small droplet clusters by itself (Fig. 11), forms similar clusters together with *Pf*RPB1-CTD (Fig. 30A). Interestingly, these clusters exhibit incomplete mixing of the two components. Furthermore, incomplete mixing of *Pf*TBP-IDR and *Pf*RPB1-CTD is exacerbated when condensation of both components is first induced separately, before combining the resulting condensates (Fig. 30A). This effect is time dependent as both components remained completely segregated when individual condensate reactions were incubated over longer periods of time prior to mixing. These observations suggest that condensates formed by individual *Pf*RPB1-CTD and *Pf*TBP-IDR fusion proteins with liquid-like properties (Figs. 5, 11) harden over time, rendering droplets unable to fuse with one another. The hardening of *Pf*TBP-IDR components in mixed clusters is extensive and the YFP-labelled *Pf*TBP-IDR takes on an irregular, aggregate-like appearance (Fig. 30A).

*Pf*TBP-IDR exhibits fundamentally different mixing properties with MED1-IDR condensates. Mixed condensates containing a homogenous mixture of both components are formed, both when components are mixed prior to the condensation reaction or when condensates pre-formed with individual components are combined (Fig. 30B). The resultant mixed condensates share the qualities of droplets formed by MED1-IDR alone and do not exhibit the clustering phenotype of *Pf*TBP-IDR (Fig. 11). The data suggest that even when MED1-IDR droplets are pre-formed, they are permeable to new components, which can diffuse into the condensates, are homogenously distributed, and are retained at higher concentration compared to the surrounding solution. Despite the extensive hardening of *Pf*TBP-IDR observed when droplets are formed alone or with *Pf*RPB1-CTD (Figs. 11, 29A), favourable interactions with MED1-IDR must be able to overcome this hardening. These interactions are able to reintroduce liquidity into *Pf*TBP-IDR droplets, at least for a short time, such that they are able to homogenously distribute within MED1-IDR condensates.

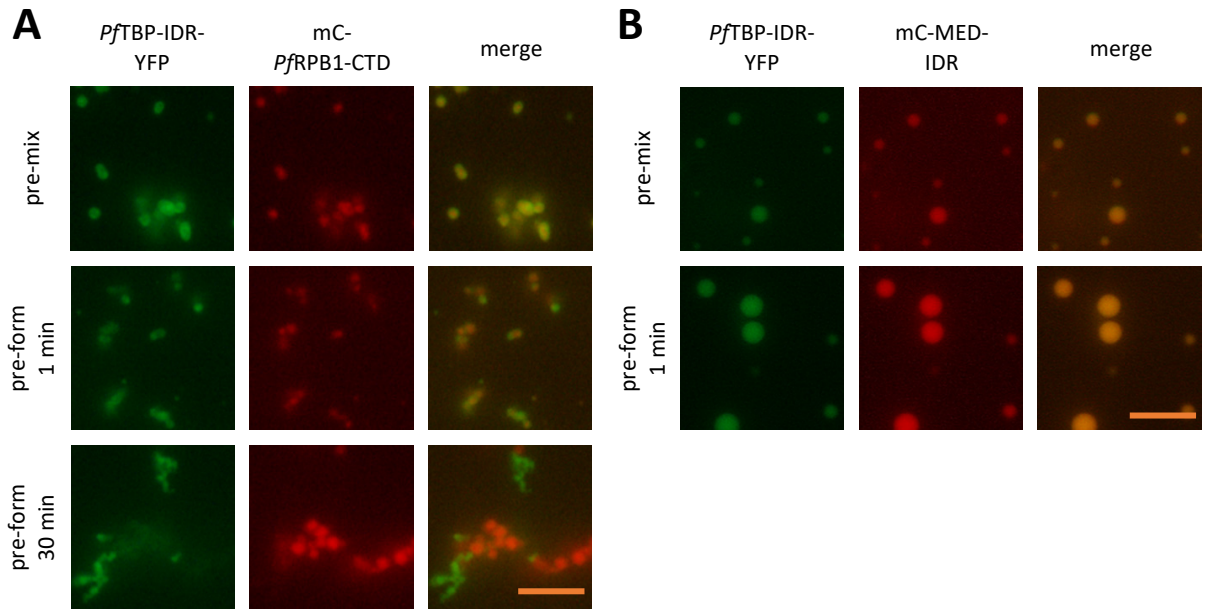


Fig. 30. *Pf*TBP-IDR forms mixed assemblies with *Pf*RPB1-CTD and MED1-IDR. (**A**) *Pf*RPB1-CTD and *Pf*TBP-IDR form mixed condensates when mixed before addition of condensation reaction buffer, but remain separated when individually formed condensates are combined. Biomolecular condensation was induced by addition of standard condensation reaction buffer, containing 12% PEG-3000 and 150 mM KCl to 5 µM mC-*Pf*RPB1-CTD and 14 µM *Pf*TBP-IDR-YFP. *Pre-mix:* protein preparations were combined and thoroughly mixed, before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately either for 1 min or for 30 min and resultant condensates combined (**B**) MED1-IDR and *Pf*TBP-IDR form mixed condensates either when condensation is induced simultaneously or after combining separately formed droplets. Biomolecular condensation was induced by addition of standard condensation reaction buffer, containing 12% PEG-3000 and 150 mM KCl to 12 µM mEGFP-MED1-IDR and 14 µM *Pf*TBP-IDR-YFP. *Pre-mix:* protein preparations were combined and thoroughly mixed, before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately either for 1 min and resultant condensates combined. *Scale bars:* 10 µm

Mixing experiments with *Hs*TBP-IDR were undertaken with a different approach. Given that *Hs*TBP-IDR forms condensates at 20% PEG-3000, but not at 12%, both these conditions were investigated in combination with *Pf*RPB1-CTD and MED1-IDR. At 12% PEG-3000, both *Pf*RPB1-CTD and MED1-IDR form spherical condensates, whilst *Hs*TBP-IDR remains evenly distributed in solution (Figs. 30A, B). This indicates that *Hs*TBP-IDR is not recruited to MED1-IDR condensates in conditions where it does not form droplets by itself. At higher PEG concentrations of 20%, *Pf*RPB1-CTD and MED1-IDR form large spherical condensates that are homogenously mixed with *Hs*TBP-IDR (Figs. 30A, B). This result indicates that, during droplet formation, *Hs*TBP-IDR molecules engage in both homotypic and heterotypic interactions to generate homogenously mixed droplets with *Pf*RPB1-CTD (Fig. 31A) and MED1-IDR (Fig. 31B). In summary, *Hs*TBP-IDR is compatible with *Pf*RPB1-CTD and MED1-IDR condensates and favourable interactions encourage the formation of homogeneously mixed condensates. Notably, we do not observe a mixture of homotypic and heterotypic droplets, but rather exclusively see homogenously mixed, heterotypic droplets (Fig. 31).

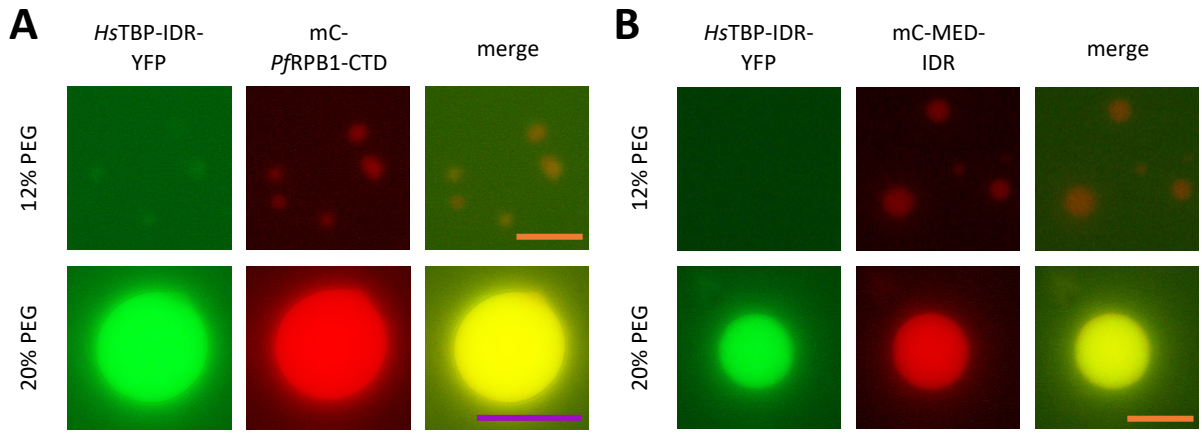


Fig. 31. *Hs*TBP-IDR forms mixed assemblies with *Pf*RPB1-CTD and MED1-IDR in the presence of 20% PEG-3000. (**A**) *Pf*RPB1-CTD and *Hs*TBP-IDR form mixed condensates only at a PEG concentration sufficient to induce *Hs*TBP-IDR condensation alone. Biomolecular condensation was induced by addition of standard condensation reaction buffer, containing 12% or 20% PEG-3000 and 150 mM KCl to 14 µM mC-*Pf*RPB1-CTD and 30 µM *Hs*TBP-IDR. Purified proteins were combined and thoroughly mixed before addition of 2x condensation buffer containing PEG-3000 as indicated. (**B**) Biomolecular condensation was induced by addition of standard condensation reaction buffer, containing 12% or 20% PEG-3000 and 150 mM KCl to 12 µM MED1-IDR and 30 µM *Hs*TBP-IDR. Purified proteins were combined and thoroughly mixed before addition of 2x condensation buffer containing PEG-3000 as indicated. *Scale bars*, *orange:* 10 µm; *purple:* 20 µm.

TFIIA-IDRs form mixed assemblies with MED1-IDR and *Pf*RPB1-CTD

The characteristic condensation properties of *Pf*TFIIA-IDR (Fig. 15) allowed us to investigate the influence of KCl concentration on the formation of mixed condensates. Under conditions at which both *Pf*TFIIA-IDR and *Pf*RPB1-CTD can form condensates individually, mixing of proteins prior to inducing condensate formation as well as mixing of pre-formed condensates results in the formation of homogenous droplets enriched in both proteins (Fig. 32A). This result contrasts with the mixing behaviour of *Pf*RPB1-CTD and *Pf*TBP-IDR, which were unable to form homogenously mixed droplets with one another (Fig. 30A). Taken together, these observations suggest that *Pf*RPB1-CTD droplets, despite their rapid transition to gel-like states (Fig. 5), are able to fuse and homogeneously combine with condensates formed by at least some factors, depending on the nature of their IDR. At KCl concentrations of 150 mM, at which *Pf*TFIIA-IDR does not phase separate alone (Fig. 15), it also does not concentrate in *Pf*RPB1-CTD droplets (Fig. 32A). This suggests that *Pf*RPB1-CTD is unable to recruit *Pf*TFIIA-IDR to mixed condensates at non-permissive salt concentrations.

MED1-IDR condensates can also homogeneously mix with *Pf*TFIIA-IDR at a low salt concentration of 50 mM KCl (Fig. 32B). Spherical, separated condensates contain a uniform mixture of both components, when droplets are pre-formed, or when components are pre-mixed prior to the condensation reaction. Unlike *Pf*RPB1-CTD (Fig. 32A), MED1-IDR is also able to concentrate *Pf*TFIIA-IDR in mixed condensates at 150 mM KCl (Fig. 32B). This suggests that MED1-IDR can recruit *Pf*TFIIA-IDR to mixed condensates under conditions where it does not characteristically phase separate. Whilst the combination of MED1-IDR, a human TF-region, and *Pf*TFIIA-IDR is not a physiologically relevant one, this observation highlights a general ability of MED1-IDR to concentrate transcription factor-IDR regions, an important property for its proposed role as universal co-activator in the regulation of transcriptional condensates.

Whilst *Pf*RPB1-CTD is unable to overcome the sensitivity of *Pf*TFIIA-IDR phase separation to high ionic strength, it is still able to recruit *Pf*TFIIA-IDR. When *Pf*RPB1-CTD is mixed with a sub-saturating concentration of *Pf*TFIIA-IDR at 50 mM KCl, *Pf*TFIIA-IDR moderately localises to mixed droplets (Fig. 32A). At this concentration, *Pf*TFIIA-IDR does form condensates alone (Fig. 15A) but is recruited to *Pf*RPB1-CTD droplets. Thus, in this context, *Pf*RPB1-CTD appears to act as a scaffold protein driving the formation of condensates and recruiting *Pf*TFIIA-IDR as client molecules.

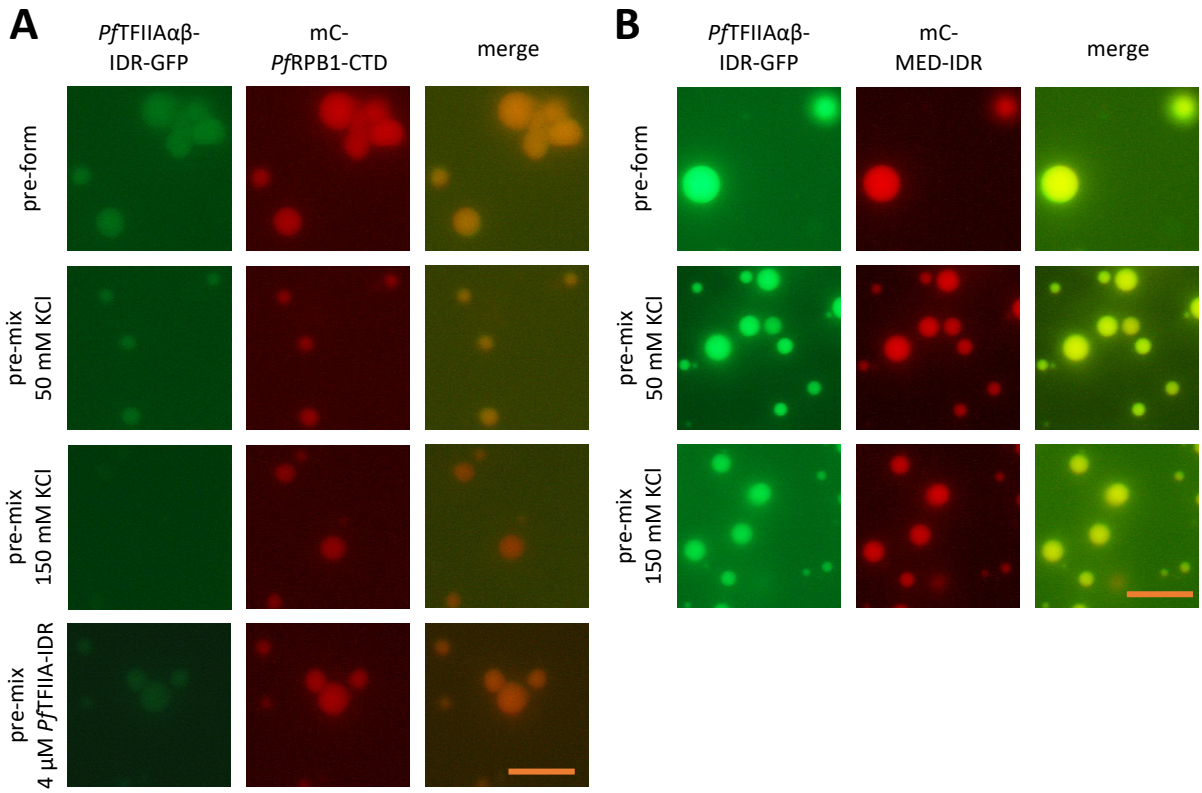


Fig. 32. *Pf*TFIIAαβ-IDR forms mixed assemblies with *Pf*RPB1-CTD and MED1-IDR. (**A**) *Pf*RPB1-CTD and *Pf*TFIIAαß-IDR form mixed condensates under a range of conditions. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 50 mM or 150 mM KCl as indicated to 14 µM mC-*Pf*RPB1-CTD and 15 µM or 4 µM *Pf*TFIIAαß-IDR-GFP as indicated. *Pre-mix:* protein preparations were combined and thoroughly mixed before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately for 1 min and resultant condensates combined. (**B**) MED1-IDR and *Pf*TFIIAαß-IDR form mixed condensates under a range of conditions. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 50 mM or 150 mM KCl as indicated to 12 µM MED-IDR and 15 µM *Pf*TFIIAαß-IDR-GFP. *Pre-mix:* protein preparations were combined and thoroughly mixed before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately for 1 min and resultant condensates combined. *Scale bars:* 10 µm.

In comparison to *Pf*TFIIA-IDR, *Hs*TFIIA-IDR exhibits a more uniform mixing behaviour across the combinations and reaction conditions investigated (Fig. 33). *Hs*TFIIA-IDR demonstrates homogeneous mixing with *Pf*RPB1-CTD when the components are pre-mixed prior to the induction of condensation (Fig. 33A). When condensate formation is first induced and droplets are then combined, the droplets formed by each fusion protein combine in heterogenous mixed assemblies (Fig. 33A). This result suggests that the homotypic interactions within droplets formed by the individual proteins rapidly strengthen, thereby causing a change in material properties that renders these largely immiscible.

*Hs*TFIIA-IDR exhibits homogeneous mixing with MED1-IDR in both pre-mixing and pre-forming assay formats (Fig. 33B). Mixed condensates contain a uniform distribution of both components, indicating that the different IDRs readily engage in heterotypic interactions. The data suggest that droplets remain liquid and dynamic enough to mix with one another, even after a minute of individual incubating (Fig. 33B). These results again highlight the wide range of IDRs compatible with MED1-IDR droplets.

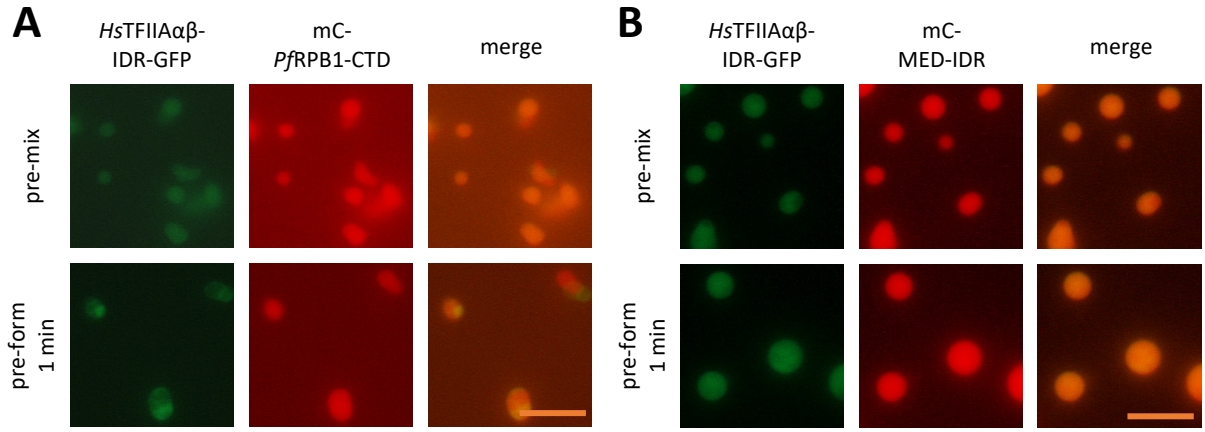


Fig. 33. *Hs*TFIIAαβ-IDR forms mixed assemblies with *Pf*RPB1-CTD and MED1-IDR. (**A**) *Pf*RPB1-CTD and *Hs*TFIIAαβ-IDR form mixed condensates when mixed before addition of condensation reaction buffer and mix partially when individually formed condensates are combined. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 50 mM KCl to 14 µM mC-*Pf*RPB1-CTD and 3 µM *Hs*TFIIAαß-IDR-GFP. *Pre-mix:* protein preparations were combined and thoroughly mixed before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately for 1 min and resultant condensates combined. (**B**) MED1-IDR and *Hs*TFIIAαβ-IDR form mixed condensates under a range of conditions. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 50 mM KCl to 12 µM MED-IDR and 15 µM *Hs*TFIIAαß-IDR-GFP. *Pre-mix:* protein preparations were combined and thoroughly mixed before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately for 1 min and resultant condensates combined. *Scale bars:* 10 µm.

TFIIF-IDRs form mixed assemblies with MED1-IDR and *Pf*RPB1-CTD

The *Pf*TFIIFß-IDR fusion protein presented an interesting case to study in the context of mixing experiments. Given its tendency to form irregular aggregates in condensation buffer (Fig. 18A) and the previous observation that it also forms aggregates in combination with other GTF-IDR fusion proteins[81], this study sought to investigate its mixing behaviour with MED1-IDR and *Pf*RPB1-CTD (Fig. 34). Unexpectedly, *Pf*RPB1-CTD can concentrate *Pf*TFIIFß-IDR within relatively homogenous spherical condensates when the two components are pre-mixed prior to the induction of condensation (Fig. 34A). Even when pre-formed droplets of the two components are mixed, the majority of the resulting condensates contain a mixture of both fusion proteins. In this case, the distribution of *Pf*TFIIFß-IDR is somewhat irregular, and some *Pf*TFIIFß-IDR assemblies do not associate with mCherry-labelled *Pf*RPB1-CTD droplets, but spherical condensates can clearly be identified in both fluorescent channels (Fig. 34A).

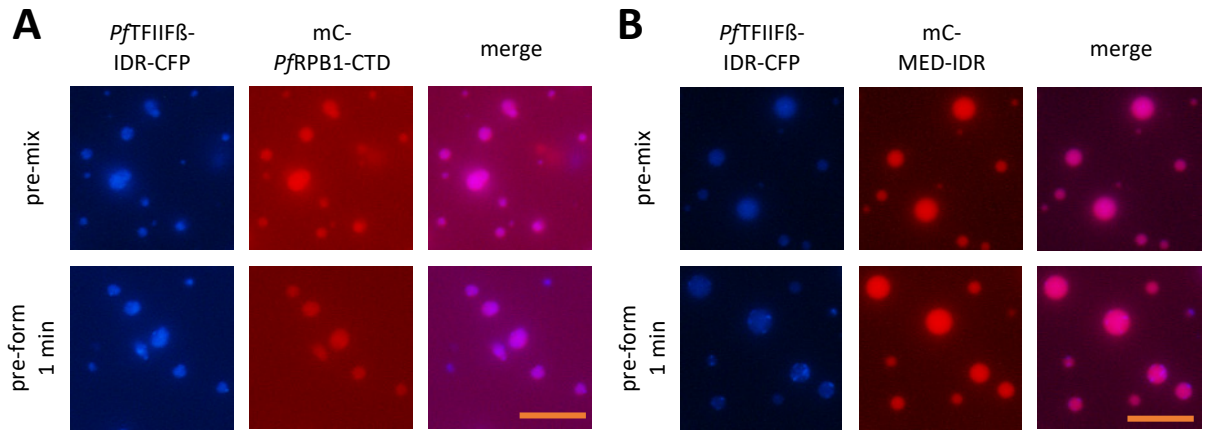


Fig. 34. *Pf*TFIIFß-IDR forms mixed assemblies with *Pf*RPB1-CTD and MED1-IDR. (**A**) *Pf*RPB1-CTD and *Pf*TFIIFß-IDR form mixed condensates when mixed before addition of condensation reaction buffer and mix partially when individually formed condensates are combined. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 50 mM KCl to 14 µM mC-*Pf*RPB1-CTD and 3 µM *Pf*TFIIFß-IDR-CFP. *Pre-mix:* protein preparations were combined and thoroughly mixed before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately for 1 min and resultant condensates combined. (**B**) MED1-IDR and *Pf*TFIIFß-IDR form mixed condensates under a range of conditions. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 50 mM KCl to 12 µM mEGFP-MED1-IDR and 3 µM *Pf*TFIIFß-IDR-CFP. *Pre-mix:* protein preparations were combined and thoroughly mixed before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately for 1 min and resultant condensates combined. *Scale bars:* 10 µm.

*Pf*TFIIFß-IDR and MED1-IDR demonstrate similar mixing behaviour. In the pre-mixing scenario, MED1-IDR homogenously concentrates *Pf*TFIIFß-IDR throughout spherical condensates (Fig. 34B). When pre-formed droplets are combined, the resulting condensates contain a relatively even distribution of both components punctuated by high concentration speckles of *Pf*TFIIFß-IDR (Fig. 34B). Taken together, these results suggest that both MED1-IDR and *Pf*RPB1-CTD overcome the aggregation propensity of the *Pf*TFIIFß-IDR fusion protein. This highlights that, whilst understanding the individual condensation properties of TF-IDRs is important, mixtures of multiple factors can generate condensates with new properties. The observation that fusion proteins do not form condensates under certain conditions, as in the case of *Pf*TFIIAαß-IDR (Fig. 15), or exhibit aggregation-propensity, such as *Pf*TFIIFß-IDR (Fig. 18), does not necessarily hold true when these proteins are combined with strong drivers of condensation. Even pre-formed aggregates seem to be, at least partially, dissolved and homogenously distributed throughout condensates formed by MED1-IDR and *Pf*RPB1-CTD (Fig. 34).

To conclude the investigation into the mixing properties of general TF-IDR fusion proteins, the combination of *Hs*TFIIF-IDR with *Pf*RPB1-CTD and MED1-IDR was assessed. Like *Pf*TFIIA-IDR, *Hs*TFIIF-IDR exhibits the property of forming condensates at 50 mM but not 150 mM KCl (Fig. 19). At permissible salt concentrations, *Hs*TFIIF-IDR forms homogenously mixed assemblies with *Pf*RPB1-CTD and MED1-IDR in both pre-forming and pre-mixing scenarios (Figs. 34A, B). In all cases, the resulting condensates are spherical, separated and contain an even distribution of both components. These results imply that these condensates can fuse and dynamically mix, even when pre-formed, suggesting they maintain a high degree of liquidity during the period of time immediately following droplet formation.

Mixing experiments were repeated at 150 mM KCl to investigate the ability of MED1-IDR and *Pf*RPB1-CTD to recruit *Hs*TFIIF-IDR. Whilst the concentration of *Hs*TFIIF-IDR in mixed condensates is reduced under these conditions (qualitatively judged by the brightness of droplets in the CFP channel), spherical condensates are still clearly visible in the CFP fluorescent channel (Figs. 34A, B). These results suggest that interactions with *Pf*RPB1-CTD or MED1-IDR drive *Hs*TFIIF-IDR molecules into condensates under circumstances where the fusion protein cannot form condensates itself. To further investigate the recruitment potential of *Pf*RPB1-CTD, it was mixed with a lower concentration of *Hs*TFIIF-IDR. As previously demonstrated, 3 µM *Hs*TFIIF-IDR does not form condensates at 12% PEG and 50 mM KCl (Fig. 19). In the presence of *Pf*RPB1-CTD, however, *Hs*TFIIF-IDR is recruited and concentrated in mixed assemblies (Fig. 35A). These data strengthen the conclusion that *Pf*RPB1-CTD condensates can recruit client molecules, acting as a scaffold for a variety of TF-IDRs.

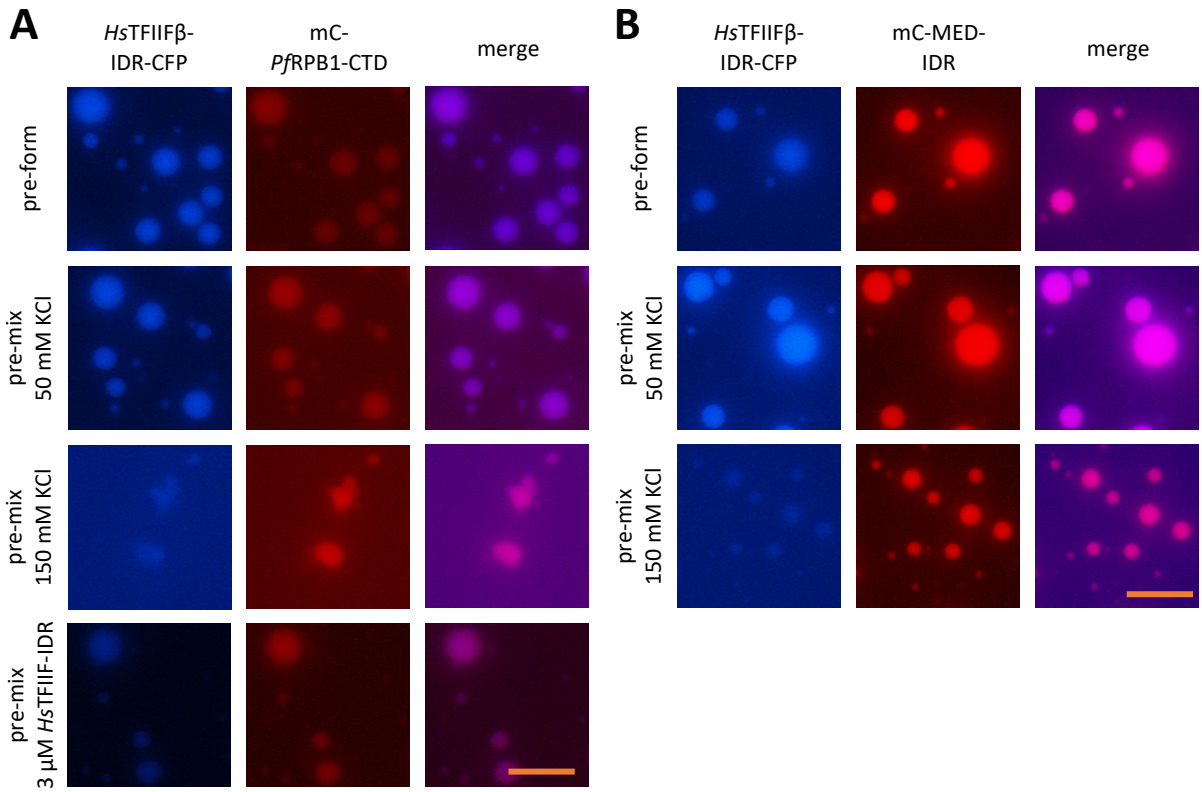


Fig. 35. *Hs*TFIIFβ-IDR forms mixed assemblies with *Pf*RPB1-CTD and MED1-IDR. (A) *Pf*RPB1-CTD and *Hs*TFIIFß-IDR form mixed condensates under a range of conditions. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 50 mM or 150 mM KCl as indicated to 14 µM mC-*Pf*RPB1-CTD and 13 µM or 3 µM *Hs*TFIIFß-IDR-GFP as indicated. *Pre-mix:* protein preparations were combined and thoroughly mixed before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately for 1 min and resultant condensates combined. (B) MED1-IDR and *Hs*TFIIFß-IDR form mixed condensates under a range of conditions. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 50 mM or 150 mM KCl as indicated to 12 µM mEGFP-MED1-IDR and 13 µM *Hs*TFIIFß-IDR-GFP. *Pre-mix:* protein preparations were combined and thoroughly mixed before addition of 2x condensation buffer to induce droplet formation; *pre-form:* biomolecular condensate formation was induced separately for 1 min and resultant condensates combined. *Scale bars:* 10 µm.

During the investigation of *Hs*TFIIF-IDR/*Pf*RPB1-CTD mixed condensates, it became apparent that the mixed droplets exhibit properties not seen for droplets formed by *Pf*RPB1-CTD alone. Specifically, the mixed condensates maintain liquid-like properties for longer periods of time, as evidenced by the frequent capture of fusion events (Fig. 36A). Smaller droplets fuse and relax to larger spherical bodies, unlike *Pf*RPB1-CTD-only droplets, which generally arrest during the fusion process, forming irregular droplet clusters (Fig. 5). This is particularly striking when droplet reactions are observed after 30 min incubation. As previously described, *Pf*RPB1-CTD alone forms strings of smaller droplets stuck together (Figs. 5, 36B). Mixed *Pf*RPB1-CTD/*Hs*TFIIF-IDR droplet reactions, however, generate distributed droplets that remain spherical and separated over longer periods (Fig. 36B). These data confirm that the droplet properties of strong drivers of biomolecular condensation can be modulated by the presence of other molecules in complex condensates.

MED1-IDR condensates recruit the VP16-AD fusion protein

Finally, the interaction between MED1-IDR and the transcription activator VP16-AD was investigated in mixing experiments. Published data have demonstrated the ability of MED1-IDR condensates to concentrate a wide variety of fusion proteins containing IDRs from diverse transcription activators[70,71]. Furthermore, the VP16-AD has long been known to engage in structured interactions with the Mediator complex, important for the activation of HSV gene transcription[105]. Consistent with the results of earlier studies, mCherry-labelled VP16-AD is clearly visible in spherical condensates, overlapping with mEGFP-labelled MED1-IDR (Fig. 37). Importantly, MED1-IDR forms condensates under these conditions (Fig. 7), but VP16-AD does not (Fig. 22). Thus, mC-VP16-AD is recruited and concentrated in MED1-IDR condensates formed in the presence 12% PEG and 150 mM KCl (Fig. 37).

The full range of MED1-IDR mixing experiments presented here greatly expand the catalogue of known factors recruited to MED1-IDR condensates. Published data have demonstrated that the fusion protein is capable of mixing with and recruiting transcription activator IDRs[69–71,99] and *Hs*RPB1-CTD[75], but the results presented here expand this list to include general TF-IDRs, from both *H. sapiens* and *P. falciparum* (Figs. 29-34), as well as the IDR of the virally encoded transcription activator VP16 (Fig. 37). MED1-IDR droplets exhibit a compatibility to interact with and to recruit a wide range of different IDRs, from species as different as *H. sapiens* and *P. falciparum,* and are therefore clearly capable of engaging in a wide range of heterotypic multivalent interactions.

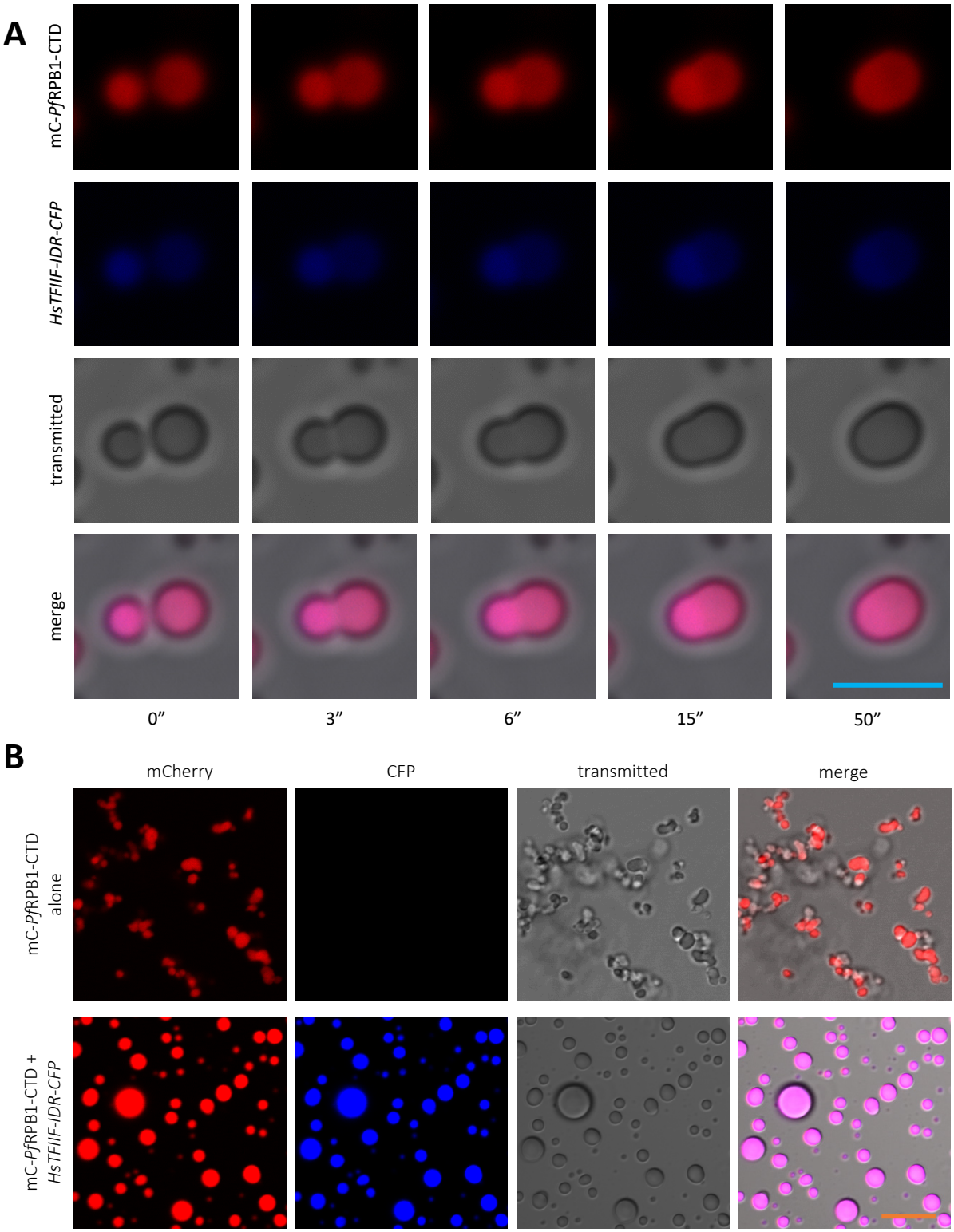


Fig. 36. *Hs*TFIIFβ-IDR/*Pf*RPB1-CTD mixed assemblies have liquid-like properties. (**A**) Complete fusion event of mC-*Pf*RPB1-CTD (14 µM) and *Hs*TFIIFß-IDR-CFP (13 µM) droplets in condensation reaction buffer at 12% PEG-3000 and 50 mM KCl, visualised by confocal fluorescent microscopy. (**B**) *Hs*TFIIFß-IDR-CFP encourages more liquid-like *Pf*RPB1-CTD droplet properties. mC-*Pf*RPB1-CTD assemblies after 30 minutes incubation, either alone or in the presence of *Hs*TFIIFß-IDR-CFP. *Scale bars*, *cyan:* 5µm; *orange:* 10 µm.

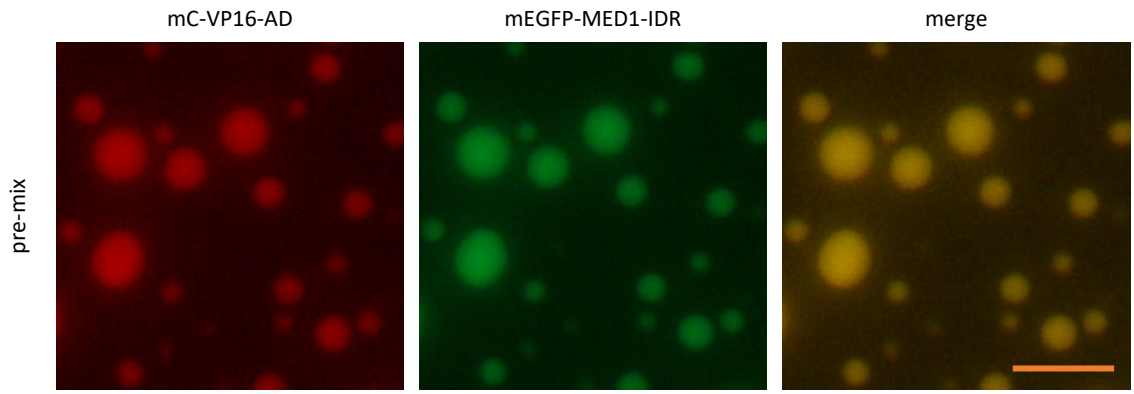


Fig. 37. VP16-AD is recruited to MED1-IDR condensates. MED1-IDR and VP16-AD form mixed condensates under conditions at which VP16-AD does not phase separate by itself. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 150 mM KCl to 12 µM mEGFP-MED1-IDR and 23 µM mC-VP16-AD. Protein preparations were combined and thoroughly mixed, before addition of 2x condensation buffer to induce droplet formation. *Scale bar:* 10 µm.

MED1-IDR condensates may concentrate transcription factors in a complex mixture of nuclear extract

Following the investigation into the interactions of mEGFP-MED1-IDR with other purified protein components, the final section of this study sought to extend these observations to a more complex environment, closer to the physiological context of the cellular nucleus. First, it was determined if droplet formation of mEGFP-MED1-IDR in nuclear extract from cultured human HeLa cells, as previously reported[69], could be recapitulated. As shown in Figs. 38A and S8, droplet formation of mEGFP-MED1-IDR in HeLa NE, in the absence of any crowding agent, could be observed. However, titration experiments revealed that MED1-IDR droplet formation is inhibited by increasing concentrations of HeLa NE (Figs. 38A, S8). Given recently published results, which show that high levels of RNA can inhibit MED1-IDR droplet formation[76], HeLa nuclear extracts were treated with benzonase nuclease, prior to mixing with purified MED1-IDR. Somewhat surprisingly, benzonase treatment of HeLa NE reduces droplet formation, suggesting that nucleic acids present in the complex mixture stimulate MED1-IDR biomolecular condensation (Fig. 38A). A further interesting observation is that in the presence of HeLa NE, MED1-IDR condensates arrange into structures resembling “strings of beads” (Fig. S8). This behaviour is notably different from the behaviour of condensates formed by isolated MED1-IDR, which exhibit no particular arrangement and remain separated (Figs. 7, 8). In summary, MED1-IDR forms condensates in the presence of HeLa nuclear extract, whose size and distribution are dependent on HeLa NE concentration; these differ in their material properties from condensates formed by MED1-IDR alone. In addition, MED1-IDR droplet formation is inhibited by a non-nucleic acid component in the complex HeLa NE mixture.

After establishing conditions in which MED1-IDR forms condensates in HeLa NE, the next goal was to attempt to recapitulate the observation that MED1-IDR droplets can concentrate full-length RPB1 from a complex mixture of human nuclear proteins[69]. Given their higher density, protein droplets can be separated from the more dilute surrounding solution by centrifugation[62,106]. In such a spin-down assay, described in the Materials and Methods section, the supernatant should contain the dilute phase of a condensation reaction, while the pellet should be enriched in the protein driving condensation, as well as any other proteins that are recruited to the protein droplets. The contents of pellets and supernatants were separated by SDS PAGE and probed by immunoblotting with specific antibodies.

Previously published results of spin-down assays show that increasing concentrations of mEGFP-MED1-IDR, but not the mEGFP fluorescent protein control lacking an IDR, recruit increasing amounts of *Hs*RPB1 in HeLa NE to mEGFP-MED1-IDR condensates[69]. This observation could not be convincingly recapitulated here. As shown in Fig. 38B, the amount of *Hs*RPB1 detected in the pellet does not seem to correlate with the concentration of MED1-IDR in the condensation reaction. Furthermore, pellets from condensation reactions containing the YFP control protein also contain *Hs*RPB1, albeit lower amounts compared to the reactions with mEGFP-MED1-IDR. YFP is unable to form protein droplets under these conditions, so the presence of *Hs*RPB1 in pellets, obtained in the presence of YFP, therefore suggests that *Hs*RPB1 in HeLa NE aggregates during the condensation reaction. Previously published data did not show the analysis of contents of supernatants obtained in spin-down assays[69]. The results presented here show that supernatants contained the vast majority of *Hs*RPB1 present in the assay. Thus, *Hs*RPB1 was not efficiently recruited to the condensed phase.

In contrast to *Hs*RPB1, results for *Hs*TBP may indicate MED1-IDR mediated recruitment (Fig. 38B). For example, at 15 µM MED1-IDR, *Hs*TBP can be clearly detected in the pellet, whereas *Hs*TBP is almost undetectable in the pellet obtained in the presence of the YFP-control (Fig. 38B, compare lanes 2 and 4). This pattern is repeated across the concentrations tested, with the amount of *Hs*TBP detected in the pellet fraction increasing with increased MED1-IDR concentrations in the assay (Fig. 38B). However, even in the presence of MED1-IDR, the *Hs*TBP signal is still stronger in the supernatant, suggesting only a weak recruitment of *Hs*TBP to condensates.

The results of the spin-down assays presented here must be viewed as preliminary and need to be confirmed using optimised assay conditions. Nevertheless, they may serve as a basis for further experiments towards understanding how transcriptional condensates may regulate the transcriptional machinery in the complex environment of the nucleus.

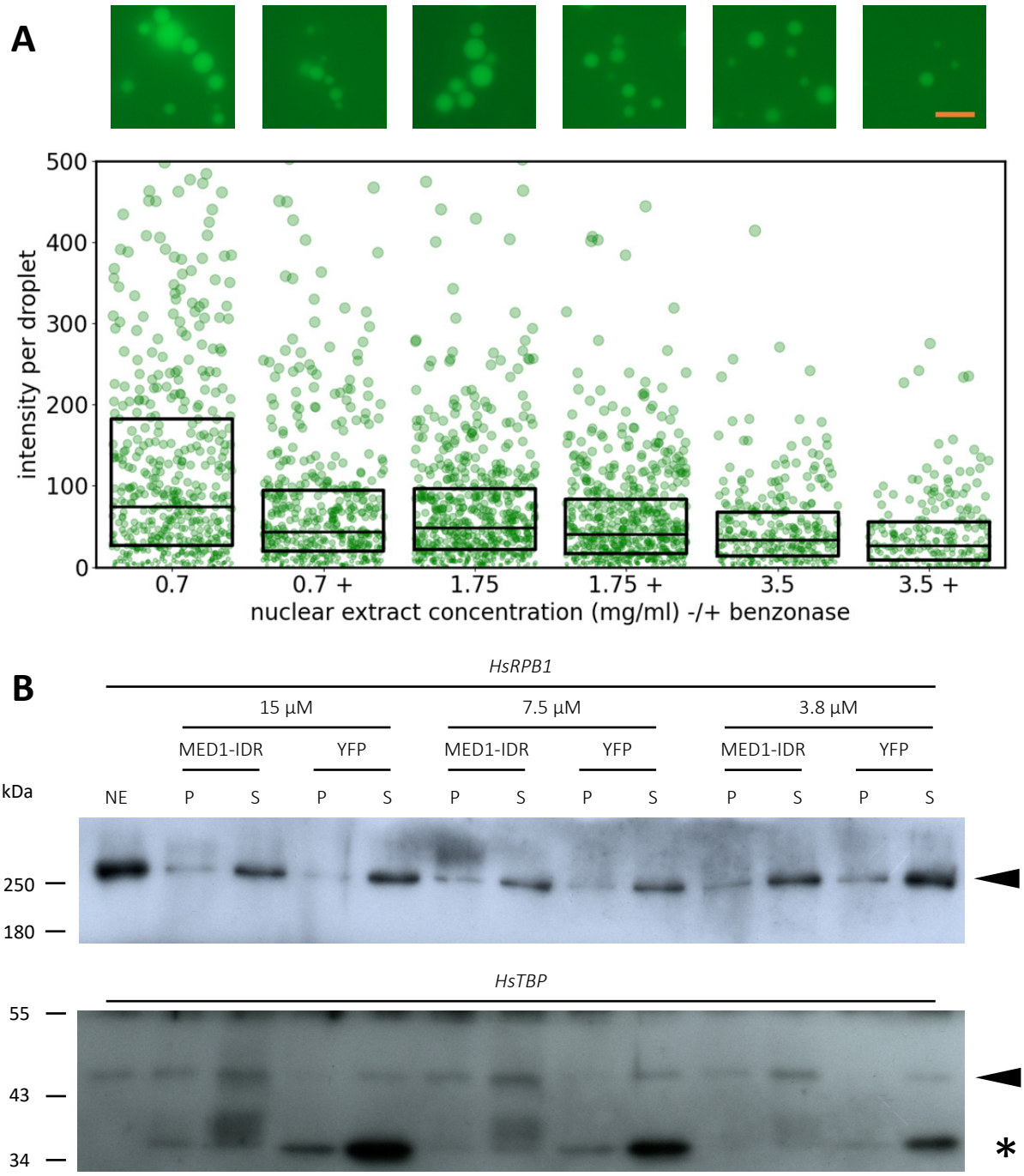


Fig. 38. MED1-IDR condensates in nuclear extract may concentrate RNAPII transcriptional machinery components. (A) Droplet formation observed by fluorescent microscopy in the presence of 15 µM mEGFP-MED1-IDR, 150 mM KCl and increasing concentrations of HeLa NE. Where indicated (*+*), HeLa NE was treated with benzonase nuclease prior to condensate formation. *Top*: representative images of condensates at indicated HeLa NE concentrations; *bottom*: integrated intensity of pixel brightness per droplet across 20 replicate fields of view per condition; box limits denote upper and lower quartiles and the centre line denotes the median value. (B) 1.75 mg/mL HeLa NE protein was mixed with 15 µM, 7.5 µM or 3.8 µM of either mEGFP-MED1-IDR or YFP control protein at 150 mM KCl. After 10 min incubation, droplets were separated from the dilute phase by centrifugation. 100% pellets (P) and 10% supernatants (S) were analysed by SDS PAGE and immunoblotting using antibodies for the largest RNAPII subunit, *Hs*RPB1, and the TATA-binding protein *Hs*TBP. The positions of protein molecular weight markers and bands corresponding to *Hs*RPB1 and *Hs*TBP (arrowhead) and the YFP control protein (asterisk) are indicated. *Top*: immunoblot probed for *Hs*RPB1; *bottom*: immunoblot probed for *Hs*TBP. *Scale bar*: 10 µm.

Discussion

Quantitative analysis of biomolecular condensation by fluorescence microscopy

During the course of this study, a start-to-finish data acquisition and analysis pipeline was established for the quantitative analysis of widefield microscopy images, based on previously published tools[76]. The Python scripts developed can be used to quantify data that could previously only be assessed qualitatively. Condensed fraction curves can be generated as a standard starting point for the analysis of biomolecular phase separation[48,71] (for example Figs. 3, 7). An extremely powerful aspect of the tools is the direct coupling of data analysis to easily interpretable data visualisation. Scatter plots give a representative view of the distribution of droplets across multiple fields of view, whilst overlaid boxplots show the mean and variance across conditions (for example, Figs. 4, 6). Importantly, the tools and protocols developed here can use widefield microscopy images as input data, whereas most available scripts so far were designed to work with confocal microscopy images[76]. The data analysis pipeline allowed us to examine phase separation under different conditions in rapid succession, in order to get a broad overview of condensate properties (see for example *Pf*RPB1-CTD, Figs. 3–6). The tools developed have been packaged and annotated in a user-friendly manner, are routinely used by other members of the research group and are freely available to interested researchers in the field.

GTF-IDRs form biomolecular condensates

With the help of the image analysis pipeline, we were able to demonstrate that a wide variety of GTF-IDR fusion constructs drive the formation of biomolecular condensates and to determine their general properties such as their response to buffer ionic strength and 1,6-hexanediol (Figs. 3–23, Table 1). The results of this study present first evidence that GTF-IDRs of diverse eukaryotes can drive biomolecular condensation and the properties of the resulting condensates are summarised in Table 1.

We were able to replicate the observation that the IDR present in the largest subunit of the Mediator complex readily forms spherical biomolecular condensates *in vitro*[69]. MED1-IDR droplets form under a broad range of environmental conditions (Figs. 7–8) and even in the absence of molecular crowding agents (Fig. 8), consistent with the proposal that MED1 acts as a driver/scaffold protein in the formation of transcriptional condensates[37,43,76].

The central role of the CTD in eukaryotic transcription regulation, as well as the more recent observation that the RPB1-CTDs of *H. sapiens* and *S. cerevisiae* undergo LLPS *in vitro*[46], stimulated an interest in the biochemical characterisation of the *Pf*RPB1-CTD.

Construct	[conc] µM	PEG-3000	KCl (mM)	1,6-hexanediol	dsDNA (ng/µl)	appearance	dynamic/ reversible	FRAP
*Pf*RPB1-CTD	1–28	> 9%	50–300	5%: ~ 10%: ✗	inhib. 15–125	0 min: separated 30 min: clusters	stalled fus. 0 min: ✓ 30 min: ✗	slow recov.
*Pf*TBP-IDR	3–28	> 12%	150	10%: ~	inhib. 125	clusters	0 min: ✓	n.d.
*Pf*TFIIAαß-IDR	6–30	> 12%	50	10%: ~	n.d.	separated	n.d.	n.d.
*Pf*TFIIFß-IDR	7	n.a.	n.a.	n.a.	inhib. aggregates: 125	aggregates	solid aggregates	n.d.
MED1-IDR	0.5–24	0-20%	150–500	not inhib.	n.d.	separated	complete fus. at 0%, 6% PEG: 0 min: ✓, ✗ 30 min: ✓, ✗	mod. recov.
*Hs*RPB1-CTD	18–70 (with PEG-20000)	n.a.	150	n.d.	n.d.	separated	n.d.	fast recov.*
*Hs*TBP-IDR	15–60	20%	150	5%: ~ 10%: ✗	inhib. 12–25	separated	0 min: ✓ 30 min: ✓	n.d.
*Hs*TFIIAαß-IDR	0.5–6	> 12%	50–150	5%: ~ 10%: ~	inhib. 125	clusters	0 min: ✓	n.d.
*Hs*TFIIFß-IDR	5–26	> 12%	50	5%: ~ 10%: ✗	n.d.	separated	complete fus. 0 min: ✓ 30 min: ✓	slow recov.
VP16-AD	46	20%	50	n.d.	inhib. 125	clusters	n.d.	n.d.

Table 1. Summary of GTF-IDR fusion protein properties.

Legend: [conc], concentration (mM); dsDNA, double-stranded poly(dG-dC) DNA; **~**, moderately inhibitory; ✗, completely inhibitory / not observed; ✓, observed; inhib., inhibitory across a given range; fus., (condensate) fusion; recov., (fluorescence) recovery; n.d., not determined; n.a., not applicable; fast recov.*, FRAP for HsRPB1-CTD in the presence of MED1-IDR.

The data presented here show that *Pf*RPB1-CTD can undergo biomolecular condensation in a concentration-dependent manner (Fig. 3), forming spherical droplets that closely resemble those produced by other proteins well-studied in the field of phase separation[46,69]. Published work on the biomolecular condensation of the homologous *Hs*RPB1-CTD by different laboratories reported apparently contradicting results. While some work suggested that *Hs*RPB1-CTD is not able to undergo biomolecular condensation by itself, but is readily recruited into condensates formed by other transcriptional proteins such as MED1[75] or PTEFb[74], other work reported phase separation of *Hs*RPB1-CTD fused to the solubility tag maltose-binding protein in the presence of the molecular crowding agent dextran at concentrations of 16%[46]. Taken together, these observations suggest that *Hs*RPB1-CTD is not a strong driver of biomolecular condensation and should therefore be considered a client protein rather than a driver of transcriptional condensate formation.

Here, we show that both the CTDs from *P. falciparum* and human RPB1 can undergo biomolecular condensation *in vitro*, at least under certain conditions (Figs. 3, 23). However, we find that the CTD of *P. falciparum* is a far stronger driver of biomolecular condensation than its human counterpart. This is clearly demonstrated by the fact that *Pf*RPB1-CTD is able to form condensates at much lower protein concentrations and in the presence of much lower concentrations of the molecular crowder PEG (compare Figs. 3 and 23).

The CTDs of the largest RPB1 subunit of RNAPII in higher eukaryotes feature a highly conserved pattern of the repeated amino acid heptad "YSPTSPS"[31] (Fig. S1). In mammalians, CTD amino acid sequences of RPB1 orthologues are essentially identical, consistent with a highly conserved, functional role in higher eukaryotes[31,107]. In contrast, many protists have lost any semblance of the heptad repeat structure conserved in other eukaryotes. This includes species from the genera *Trypanosoma* and *Leishmania*[31], as well as other *Apicomplexan* parasites, such as *Theileria* and *Toxoplasma*[34] (Fig. S4). The RPB1-CTD of *P. falciparum* sits squarely between these two extremes, displaying both divergent and conserved sequence features (Fig. 1B). The unique sequence structure of the *Pf*RPB1-CTD, as well as large insertions in the conserved, structural domains of RPB1 (Figs. 1A, S2) were already noted when the gene was first sequenced[108]. However, to our knowledge the work presented here provides first evidence for differences in the biomolecular condensation properties between the well conserved *Hs*RPB1-CTD and the uniquely patterned *Pf*RPB1-CTD (Fig. 3).

Whilst the appearance of spherical *Pf*RPB1-CTD condensates is a strong indicator of their liquid property, our data further suggest that *Pf*RPB1-CTD phase behaviour is more complex. *Pf*RPB1-CTD droplets only partially fuse to form static droplet clusters when incubated over longer time periods and, consistent with this observation, show slow FRAP recovery rates (Figs. 5-6). These observations

suggest that *Pf*RPB1-CTD condensates formed under these conditions transition over time from a liquid- to a gel-like state.

The transition of biomolecular condensates from liquid to gel-like or even solid states is not unusual and has been well documented for a number of proteins[109]. In this context, it is important to note that the *in vitro* phase separating properties of proteins are generally studied in isolation or in the presence of a limited number of factors. Other cellular components, present in high concentrations, for example RNAs, have been shown to play a key role in the phase behaviour of protein condensates[55]. For this reason, although simplified *in vitro* systems such as the ones used in this study are very useful in describing the basic biochemical properties of biomolecules, they cannot capture the full physiological range of a protein's phase behaviour.

Taken together, the results presented here provide evidence to suggest that the *Pf*RPB1-CTD is a strong driver of biomolecular condensation and give first insights into the nature and material properties of *Pf*RPB1-CTD condensates. The results provide a solid basis for future work towards a more detailed characterisation of *Pf*RPB1-CTD phase behaviour and the identification of the responsible CTD amino acid sequence elements. An important aspect of future studies will be to assess the effect of phosphorylation on *Pf*RPB1-CTD condensate formation.

Much of the leading work in this field has focused on transcription driven by super-enhancers[37,69,77,110], but biomolecular condensation of GTFs has so far not been comprehensively addressed. Parallel to our own investigations, which demonstrate that the N-terminal IDR of *H. sapiens* and *P. falciparum* TBP can generate spherical assemblies *in vitro*[81] (Figs. 11–14), it was reported that clustering of TBP occurs in human cells[99]. In addition, TAF14 and TAF15, components of the TFIID complex, have also been shown to undergo biomolecular condensation *in vitro* and in cells and to act as regulators of transcription in a condensation-dependent manner[73,111]. Bar these results, no published data exist on the phase separation capacity of central players in transcription initiation such as TFIIA, -B and -F.

Here we show that fusion proteins, containing IDRs from either *P. falciparum* or orthologous *H. sapiens* sequences, form spherical condensates in the presence of the macromolecular crowder PEG-3000 (Figs. 11–21). These results extend the existing model of transcriptional regulation by condensates, which so far has been supported by observations that activators, co-activators and RNAPII form condensates.

Given its central role in every transcriptional event[23], TBP was of particular interest in our investigations. It has long been appreciated that *Hs*TBP contains a 150 amino acid long, unstructured N-terminal region, although its function has remained elusive[23]. *Pf*TBP contains an N-terminal IDR of

similar length (Fig. M1). Our investigations demonstrate that both regions can drive biomolecular condensation *in vitro* (Figs. 11-14). Whilst both regions can form spherical condensates, they clearly show different propensities for biomolecular condensation. *Hs*TBP-IDR only forms condensates in the presence of high concentrations of molecular crowder, whereas *Pf*TBP-IDR readily forms condensates at about half the protein concentration and in the presence of half the concentration of the crowder PEG. These results suggest that the *Pf*TBP-IDR is a much stronger driver of biomolecular condensation compared to its human counterpart.

The observation that the *Pf*TBP-IDR strongly drives condensation is complementary with previous data from our group, demonstrating that an IDR in *P. falciparum* TBP-like protein (*Pf*TLP), a TBP-paralogue, is capable of driving the formation of condensates at low protein concentrations[15,112]. Recent studies suggest that condensates are a general feature of eukaryotic transcription regulation, and that MED1-IDR plays a key role in driving the formation of transcriptional condensates in *H. sapiens*[69–71]. The ability of *Pf*RPB1-CTD, *Pf*TBP-IDR and *Pf*TLP-IDR to drive the formation of biomolecular condensates at low concentrations may thus indicate an evolutionary compensation to the loss of the Mediator complex in *P. falciparum*[16]. Without the central regulation and strong condensation propensity provided by Mediator, it may be the case, that many *P. falciparum* GTFs are jointly responsible for the formation and regulation of transcriptional condensates.

GTF-IDR condensates are responsive to environmental conditions

In agreement with the results of other recent research into the condensation of transcription factor-IDRs[46,69], we have shown that the formation and material states of GTF-IDR fusion protein condensates are sensitive to environmental conditions.

For example, MED1-IDR readily forms condensates at a wide range of ionic conditions, up to 300 mM KCl. These are spherical, clearly separated from one another (Fig. 7) and dynamically fuse into larger spherical droplets (Fig. 10). At very high ionic strength, 500 mM KCl and above, MED1-IDR begins to form condensates clusters and eventually irregular assemblies resembling aggregates (Fig. 7). The observed shift from liquid, spherical droplets to solid, irregular aggregates is likely caused by changes in intermolecular interactions, from transient, dynamic interactions to stronger, more static interactions. Increased KCl concentrations could, for example, shield weak and highly dynamic electrostatic interactions[48] and, at the same time, favour hydrophobic interactions that ultimately lead to protein aggregation.

Buffer ionic strength[48] and 1,6-hexanediol[101] are commonly used tools to probe the underlying intermolecular interactions in condensate formation. Sensitivity to salts such as KCl suggests the involvement of charge–charge interactions, whilst sensitivity to 1,6-hexanediol is thought to indicate

hydrophobic interactions. *Pf*RPB1-CTD exhibits robust condensation across a range of KCl concentrations but is sensitive to concentrations above 500 mM KCl. The highly divergent C-terminus of the *Pf*RPB1-CTD may be responsible for its ionic sensitivity. Breakdowns in repeat structure at the extreme C-terminus of RPB1-CTDs are not uncommon[31], but the size of this region in the *Pf*RPB1-CTD is unique (Figs. 1A, S1). Relative to other regions, and the canonical eukaryotic CTD, it is enriched in the negatively charged amino acids aspartate and glutamate (Fig. S2). These residues may be capable of engaging in ionic interactions with the regularly spaced, positively charged lysine residues in the middle section of the *Pf*RPB1-CTD and could explain the KCl-sensitive condensation properties of the *Pf*RPB1-CTD fusion protein (Fig. 3A).

The majority of condensation data generated here were produced in the presence of molecular crowding agent. Molecular crowders, such as PEG, are useful tools in the study of condensation, as they help to mimic the protein-dense environment of the cell. Molecular crowders essentially raise the effective concentration of biomolecules, and, as a result, promote molecular interactions in solution[52,62]. The ability of MED1-IDR to form biomolecular condensates in the absence of PEG provided the opportunity to investigate how the presence of PEG affects condensate properties.

We found that the brightness of MED1-IDR condensates relative to the surrounding solution increases in the presence of increasing PEG concentrations, suggesting increasing protein concentrations within the condensates (Fig. 8B). Higher concentrations of MED1-IDR within condensates facilitate intermolecular interactions between MED1-IDR molecules, which may explain the changing material properties of MED1-IDR condensates as a function of PEG concentration. Evidence for increased intermolecular interactions is provided by the observations that increased PEG concentrations result in resistance to challenge in salt dilution assays (Fig. 9B) and in slower FRAP recovery rates (Fig. 10). These data may also indicate that the loss of dynamic properties and formation of gel-like clusters by other fusion proteins, such as *Pf*RPB1-CTD (Fig. 5) and *Pf*TBP-IDR (Fig. 11), is caused by the presence of PEG. In native contexts, where extremely high concentrations of biomolecules induce crowding, the effects of crowding on condensate material properties may be different[62].

Environmental conditions do not only affect the formation condensates but were also shown to influence the behaviour of pre-formed condensates. Reversibility is a good indicator of liquid-like properties in condensates as solid aggregated structures are generally irreversible[113]. In addition to the *Pf*RPB1-CTD and MED1-IDR fusion proteins, droplets formed by *Pf*- and *Hs*TBP-IDR, *Hs*TFIIAαβ-IDR and *Hs*TFIIF-IDR are all dissolved upon challenge with higher salt concentrations (Figs. 11B, 14A, 16B, 20). Even after 30 min incubation periods, *Hs*TBP-IDR and *Hs*TFIIF-IDR droplets could be largely dissolved by high salt challenge, indicating that they remain dynamic for long periods of time

(Figs. 14A, 20). In the cellular context, the reversibility of condensates is a primary functional feature, allowing them to dynamically form and disperse in response to cellular signalling. Stress granules, for example, have been shown to form in response to external stress stimuli and protect cytosolic RNAs[42,50]. Upon alleviation of the stress-inducing condition, stress granules dissolve, releasing RNAs and making them available for translation. In the realm of transcriptional condensates, RNAPII and Mediator clusters have been demonstrated to transiently form and disperse throughout the nucleus[66]. The reversibility of condensates formed by general TF-IDRs is therefore an important functional property, which could be utilised by the cell in the dynamic regulation of transcription.

Condensate properties are, however, complex and multi-faceted, and their material states cannot be determined by a single assay. For example, although *Hs*TFIIFß-IDR condensates exhibit dynamic properties in the form of spherical droplets that fuse upon contact, they do not recover quickly from bleaching in FRAP assays (Fig. 21). This suggests that dynamic internal rearrangement of molecules and macroscopic liquid-like properties are somewhat orthogonal. *Hs*TFIIFß-IDR condensates do not behave as simple liquids, but are clearly not solid aggregates either, inhabiting variable and complex spaces between these two material states.

Our data further show that condensation behaviour is modulated by the presence of dsDNA. In our experiments, dsDNA universally suppressed biomolecular condensation of TF-IDRs (Figs. 4–18, Table 1). This observation may be the result of the simplified, synthetic constructs used and not representative of the behaviour of the full-length proteins in their native contexts. Recent results have demonstrated that a synthetic GAL4-VP16-AD fusion protein exerts tensive forces on DNA and drives condensation[114]. Crucially, this construct contains the GAL4 DNA-binding domain, capable of tight DNA-binding. Fully understanding the co-condensation properties of transcription factors in the presence of nucleic acids may thus require the study of full-length proteins or IDRs fused to heterologous DNA-binding domains. The efficient expression and purification of full-length *P. falciparum* proteins is a complex task[13,15] and presents difficulties in obtaining the concentrations necessary for biomolecular condensation assays. TBP, for example, contains a well-characterised DNA-binding domain[23], and it may be the case that the presence of this domain reverses the negative effect of DNA on condensation. Biomolecular condensation is principally governed by the multivalency of interactions between molecules[36] and strong interactions between condensation partners can increase the length of interacting polymers and number of potential interactions.

Interestingly, the presence of MgOAc counteracts the negative effect of dsDNA on condensation of *Pf*RPB1-CTD (Fig. 4D). This observation indicates that the negative effect of DNA on condensation may be explained by its polyanionic properties, affecting interactions between charged residues in a similar

fashion to high ionic strength solutions. Again, in the presence of generally positively charged DNA-binding domains[26], the negative charges present in the phosphate backbone of DNA may in turn have a stimulatory effect on condensation. Older biochemical investigations of the *Hs*RPB1-CTD indicate that its tyrosine residues are able to intercalate with DNA[115], which may also inhibit multivalent interactions between CTD residues. Mg^{2+} may in turn inhibit non-specific DNA interactions allowing the CTD molecules to engage in homotypic interactions and form condensates[103].

A wealth of data suggests that moderate concentrations of RNA generally promote the condensation of IDRs, especially those enriched in positively charged residues[49,55,58,116]. Relevant to the protein regions studied here, MED1-IDR has been shown to co-condense with low levels of RNA[76]. Conversely, high concentrations of RNA disperse condensates formed by MED1-IDR. Given that RNA is likely to be prominent in the vicinity of transcriptional condensates, its effect on the GTF-IDRs studied here should be a priority of future investigations.

*Pf*RPB1-CTD and MED1-IDR mix homogenously with GTF-IDRs, but not with each other

Previous investigations into the function of MED1-IDR have highlighted its ability to recruit a range of additional TF-IDRs to condensates. This includes IDRs from transcription activators[70,71] as well as the *Hs*RPB1-CTD[75]. The data presented here further underscore the ability of MED1-IDR to recruit a wide range of proteins, consistent with its proposed role to act as a central node in the regulation of transcriptional condensates. MED1-IDR condensates were able to homogenously mix with all GTF-IDR fusion proteins (Figs. 30–35). Even after mixing pre-formed droplets of individual proteins, mixed condensates were observed with an even distribution of both components. This suggests that MED1-IDR condensates maintain liquidity following formation and can reverse strong homotypic interactions in droplets formed by other factors. Both *P. falciparum* and *H. sapiens* GTF-IDRs can homogeneously mix with MED1-IDR condensates which does not suggest any species-specific regulation in condensate compatibility, but instead highlights the broad compatibility of MED1-IDR interactions.

In addition to homogenously mixing with varied fusion proteins under permissive conditions, MED1-IDR condensates are able to recruit GTF-IDR fusion proteins as client molecules. Mixing experiments with *Pf*TFIIAαß-IDR (Fig. 32B), *Hs*TFIIFß-IDR (Fig. 35B) and VP16-AD (Fig. 37), which are unable to form spherical condensates at physiological ionic strength[62] (150 mM KCl; Figs. 15,19, 22), show that these fusion proteins are efficiently recruited to mixed MED1-IDR condensates. Therefore, interactions between client GTF-IDRs and MED1-IDR must be more favourable than TF-IDR-solvent interactions, thus encouraging a fraction of molecules to localise to condensates. These data lend further credence to the characterisation of MED1-IDR as a driver or scaffold protein in condensate-regulated transcription[37,53].

Mixing experiments similarly showcase the ability of *Pf*RPB1-CTD to form mixed condensates and to recruit a wide range of GTF-IDR fusion proteins to mixed condensates. Specifically, reactions with *Pf*TFIIAαβ-IDR and *Hs*TFIIFß-IDR showcase the ability of *Pf*RPB1-CTD to recruit proteins to mixed condensates, under conditions where they would normally not phase separate (Figs. 32A, 35A). Individual characterisation of *Pf*TFIIAαβ-IDR and *Hs*TFIIFß-IDR established clear saturation concentrations for each protein at 50 mM KCl. At and below 4 µM and 3 µM respectively, these proteins do not form condensates in solitary condensation reactions (Figs. 15A, 19A). In combination with *Pf*RPB1-CTD, however, *Pf*TFIIAαβ-IDR and *Hs*TFIIFß-IDR are visible as spherical condensates under the relevant fluorescent channels. This clearly demonstrates that *Pf*RPB1-CTD engages in favourable heterotypic interactions with these molecules, encouraging them to separate out of solution and partition into condensates. This behaviour is similar to MED1-IDR, which is able to recruit important transcription factors, such as OCT4 and p53, as well as the GTF-IDR fusion proteins studied here, to mixed condensates *in vitro*[70].

Previously published data[75], replicated here, demonstrated the ability of MED1-IDR to recruit unphosphorylated *Hs*RPB1-CTD to mixed condensates (Fig. 27). These mixed condensates contain a homogeneous distribution of the two components, suggesting no particular preference for homotypic interactions (MED1–MED1 or CTD–CTD) in favour of heterotypic interactions (MED1–CTD). The unique sequence features of the *Pf*RPB1-CTD, possessing some heptad consensus repeats as well as long regions containing *P. falciparum*-specific sequences (Figs. 1, S1), raised the question of how it would behave in combination with MED1-IDR. In contrast to the homogenous mixing observed with *Hs*RPB1-CTD and MED1-IDR, mixing of *Pf*RPB1-CTD with MED1-IDR generated multi-droplet assemblies with a heterogeneous distribution of components (Fig. 24). Thus, in stark contrast to wide compatibility of MED1-IDR and PfRPB1-CTD condensates with GTF-IDR fusion proteins, these proteins are unable to form homogenous assemblies and, instead, form uniquely heterogenous assemblies when mixed with each other. This observation strongly suggests that critical features of transcriptional condensate regulation have diverged in the two organisms.

Although the two molecules resist homogenous mixing, they clearly do possess some level of affinity for one another. First, both MED1-IDR and *Pf*RPB1-CTD are always observed in heterogenous condensates in mixing experiments, and not in a mixture of mixed and homogenous condensates containing only one of the proteins. Therefore, *Pf*RPB1-CTD is partially attracted to MED1-IDR condensates, or *vice versa*.

Secondly, when condensates are pre-formed with individual components and are mixed after long periods of incubation, they adhere to one another in multi-droplet clusters (Fig. 24B). This dual

behaviour can be explained by the partial similarity between *Pf*RPB1-CTD and *Hs*RPB1-CTD sequences (Fig. S1). Given that MED1-IDR readily recruits the *Hs*RPB1-CTD, which is composed almost entirely of consensus heptad repeats, it stands to reason that it would have some affinity for the *Pf*RPB1-CTD, given its central block of near-consensus heptads. The incomplete mixing must, however, be the result of *P. falciparum*-specific regions engaging in homotypic interactions that are favoured over heterotypic interactions with MED1-IDR in mixed condensates.

The properties of mixed *Pf*RPB1-CTD/MED1-IDR condensates stand in stark contrast to the mixture of *Pf*RPB1-CTD and *Hs*TFIIF-IDR. Whilst neither of these combinations are of particular physiological relevance, each highlights important ways in which diverse biomolecules can influence the properties of mixed condensates. Whilst mixed *Pf*RPB1-CTD/MED1-IDR condensates are dominated by *Pf*RPB1-CTD-like properties, evidenced by the stalled fusion of droplets and formation of arrested droplet clusters (Fig. 26), *Pf*RPB1-CTD/*Hs*TFIIFß-IDR condensate properties seem primarily driven by the *Hs*TFIIFß-IDR component (Figs. 35, 36). These droplets remain dynamic for longer periods of time, completely fusing to form larger spherical condensates (Fig. 36).

These results highlight the importance of studying transcription factors-IDRs in combination with one another and in complex systems that more closely contain the physiological complement of nuclear factors, such as nuclear extracts. Following the individual and pairwise characterisation of fusion proteins presented here, future work should investigate the properties of complex, multifactor condensates. The aging of *Pf*RPB1-CTD condensates might be the result of dominating homotypic interactions, and there might exist *P. falciparum*-specific factors that encourage liquidity of *Pf*RPB1-CTD condensates. Secondary molecules affecting droplet properties are a well-documented phenomenon – proteins, which would otherwise form solid aggregates, can form liquid condensates in the presence of RNA and other polyanions[58]. Even small molecules such as Mg^{2+} and ATP can tune the material properties of condensates[103,117]. This phenomenon is further highlighted by the behaviour of *Pf*TFIIFß-IDR in mixed and solitary condensation reactions. We were unable to find reaction conditions in which *Pf*TFIIFß-IDR formed clear and spherical condensates, instead observing irregular, solid-state aggregates (Fig. 18). In combination with MED1-IDR or *Pf*RPB1-CTD however, especially when components are mixed prior to the induction of condensation, *Pf*TFIIFß-IDR is localised and homogenously distributed within spherical condensates. Thus, both MED1-IDR and *Pf*RPB1-CTD are able to solubilize *Pf*TFIIFß-IDR within the context of mixed biomolecular condensates.

Initial attempts to study recruitment of GTFs to MED1-IDR condensates in the context of HeLa nuclear extract

In an effort to study the recruitment of RNAPII and GTFs to transcription factor-IDR condensates under conditions closer to the physiological context, we extended mixed condensation assays to include complex mixtures of nuclear proteins. Unfortunately, we were unable to reproduce previously published results showing recruitment of *Hs*RPB1 in HeLa NE to MED1-IDR condensates[69]. This result is perhaps due to properties inherent to the nuclear extract used and highlights the difficulties of working in more complex systems. The system does, however, show promise for elucidating interactions between transcriptional condensates in complex mixtures, as the results provide some preliminary evidence that *Hs*TBP can be recruited to MED1-IDR condensates in HeLa NE (Fig. 38). Interestingly, MED1-IDR and *Hs*TBP-IDR only formed mixed condensates in the presence of high concentrations of PEG (Fig. 31), conditions where *Hs*TBP-IDR phase-separates independently. Our results may therefore indicate that recruitment of *Hs*TBP in HeLa NE to MED1-IDR condensates is mediated by additional factors in the complex mixture. Finally, nuclease treatment reduced the formation of MED1-IDR condensates in HeLa NE, highlighting an important regulatory role of nucleic acids in transcriptional condensate formation (Fig. 38), consistent with recent literature[76]. Further work is needed to optimise assay conditions in order to get clear results. If conditions can indeed be found in which recruitment of HeLa NE proteins to GTF-IDR condensates can be demonstrated this type of assay could be adopted utilising extracts of *P. falciparum.* Of particular interest will be to test whether the strong drivers of condensate formation, *Pf*RPB1-CTD and MED1-IDR, show distinct ability to recruit GTF orthologues in *P. falciparum* and in HeLa NE.

Key conclusions and future work

The data generated during the course of this study significantly extend the body of research around the role of biomolecular condensates in transcription regulation. We demonstrated that a range of GTF-IDRs with very different amino acid sequence properties, from both *H. sapiens* and *P. falciparum* homologous proteins, can drive biomolecular condensation *in vitro*. It seems now clear that all classes of transcriptional proteins contain IDRs capable of driving biomolecular condensation, including transcription activators, the Mediator complex, RNAPII as well as the GTFs and that biomolecular condensation plays a role non-model eukaryotic organisms such as *P. falciparum.* Furthermore, our results suggest that the *Pf*RPB1-CTD is a strong driver of biomolecular condensation, forming condensates much more readily than the homologous *Hs*RPB1-CTD. Our data further support the proposed role of the MED1-IDR as a central node in the regulation of transcriptional condensates in higher eukaryotes, forming condensates across a range of conditions and through its broad mixing and recruiting capabilities. However, while *Pf*RPB1-CTD and MED1-IDR condensates are compatible with a

wide range of different TF-IDRs, readily forming homogenous mixed condensates with heterotypic interaction partners, they are unable to homogenously mix with one another in mixed condensates. This result clearly demonstrates functional divergence in the formation of complex transcriptional condensates between *H. sapiens* and *P. falciparum*, which may be targetable by future drugs that seek to specifically affect transcriptional condensates in the parasite, but not in the human host.

The insights gained through bioinformatic and biochemical investigations of the *Pf*RPB1-CTD should be combined in the design of future experiments. Given the known importance of phosphorylation in both canonical CTD regulation[31] and the condensation behaviour of *Hs*RPB1[74,75], the effect of phosphorylation on the properties of *Pf*RPB1-CTD are of keen interest. In conjunction with an investigation into the effect of phosphorylation on *Pf*RPB1-CTD condensation, the role of lysine substitutions in CTD heptads warrants careful investigation (Fig. 1B). Through point mutations or the generation of truncation constructs, the relative importance of specific residues and sequences on the condensation properties of *Pf*RPB1-CTD could be determined. Phosphorylation assays could initially be carried out using broad-spectrum serine-kinases that indiscriminately phosphorylate serine residues, but convincing evidence would require the use of the relevant functional kinases such as *Pf*CDK7. Phosphorylation of the *Hs*RPB1-CTD has been demonstrated to negatively regulate condensation[46], but the presence of regularly spaced lysine residues may affect the molecular calculus governing condensate formation. The results presented here give the first evidence that the *Pf*RPB1-CTD is capable of driving biomolecular condensation and future work should seek to carefully dissect the connection between *Pf*RPB1-CTD sequence components and phase behaviour as well as means through with the condensation of *Pf*RPB1-CTD may be regulated in cells.

The functional relevance of the three *P. falciparum* specific insertions in *Pf*RPB1 is also of great interest (Fig. 1A, S2). Given their low complexity and predicted intrinsic disorder, future work should investigate if they are also able to drive biomolecular condensation *in vitro*. If these regions can form assemblies, their compatibility with *Pf*RPB1-CTD condensates should be assessed. As demonstrated here, studying individual protein regions, separated from their full context, can only give preliminary information on the condensation properties of proteins. In a further effort to study transcriptional condensates closer to their *in vivo* context, future biochemical investigations must include the use of cell extracts and, further, assess the impact of nucleic acids on GTF-IDR condensation. RNA has been demonstrated to regulate the properties of condensates generally[42,58,118] and, specifically, the formation and material properties of transcriptional condensates[76].

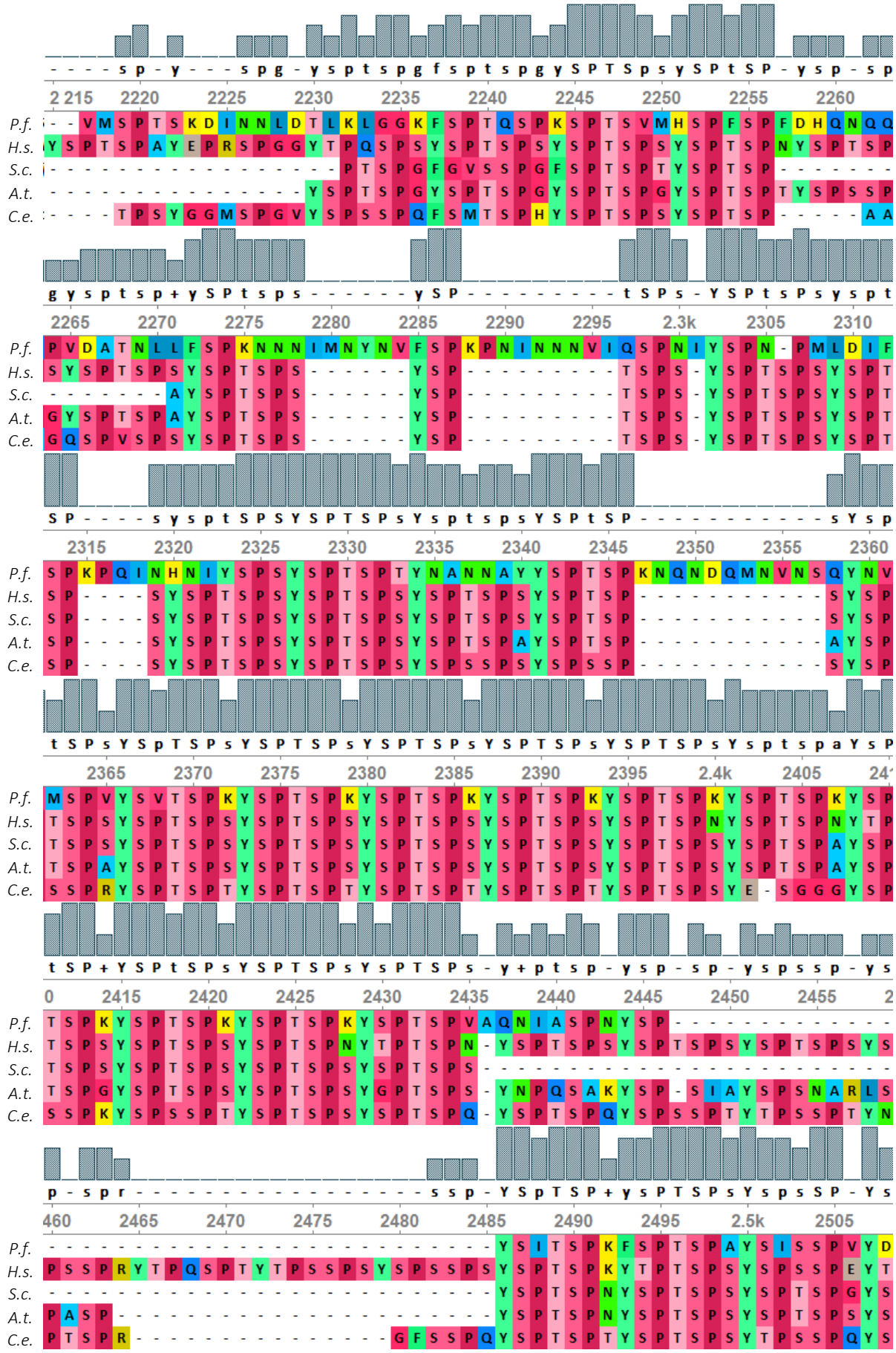
Finally, future efforts should focus on further dissecting key differences in the condensation properties of *H. sapiens* and *P. falciparum* transcription factor-IDRs. The assembly and recruitment properties of transcriptional condensates are already being investigated as potential drug targets[119], and understanding the differences in the formation and properties of human and malaria parasite transcription condensates may provide a unique opportunity for the design of novel, highly effective anti-malarials.

Supplementary material

Table 2. Primers for generation of fusion proteins by overlap extension.

Primer Name	Sequence
MED1-IDR	
MED1 IDR FWD (AseI overhang)	gcggcagcattaatGAGCATCACAGTGGTAGTCAG
MED1 IDR REV (N-term YFP overhang	ctcgcccttgctcaccatCATAAGATCATCATCTTCCTCCC
YFP FWD (C-term MED1 overhang)	gggaggaagatgatgatcttatgATGGTGAGCAAGGGCGAGG
YFP REV (BclI overhang)	tagcagcctgatcattatcaCTTGTACAGCTCGTCCATGC
VP16-AD	
FP-VP16AD-FWD	gcggcagccatatgCTGTCCACCACCGCCCC
FP-VP16AD-REV	ctcgcccttgctcaccatCCCCCCAAAGTCGTCAATG
VP16AD-BamHI-REV3	cattgacgactttgggggATGGTGAGCAAGGGCGAGG
***Pf*RPB1-CTD**	
Pf RPB1 FWD (NdeI overhang)	gcggcagccatatgGGTTTACAATCACCTGTGGCAA
Pf RPB1 REV (N-Term YFP overhang)	ctcgcccttgctcaccatCATATTTTCCTTCATTTCGTCCTC
YFP FWD (C-term PfRPB1-CTD overhang)	gaggacgaaatgaaggaaaatatgATGGTGAGCAAGGGCGAGG

Red: Oligo tail not identical to target sequence Black: Segment identical to target sequence.


P.f.
H.s.
S.c.
A.t.
C.e.

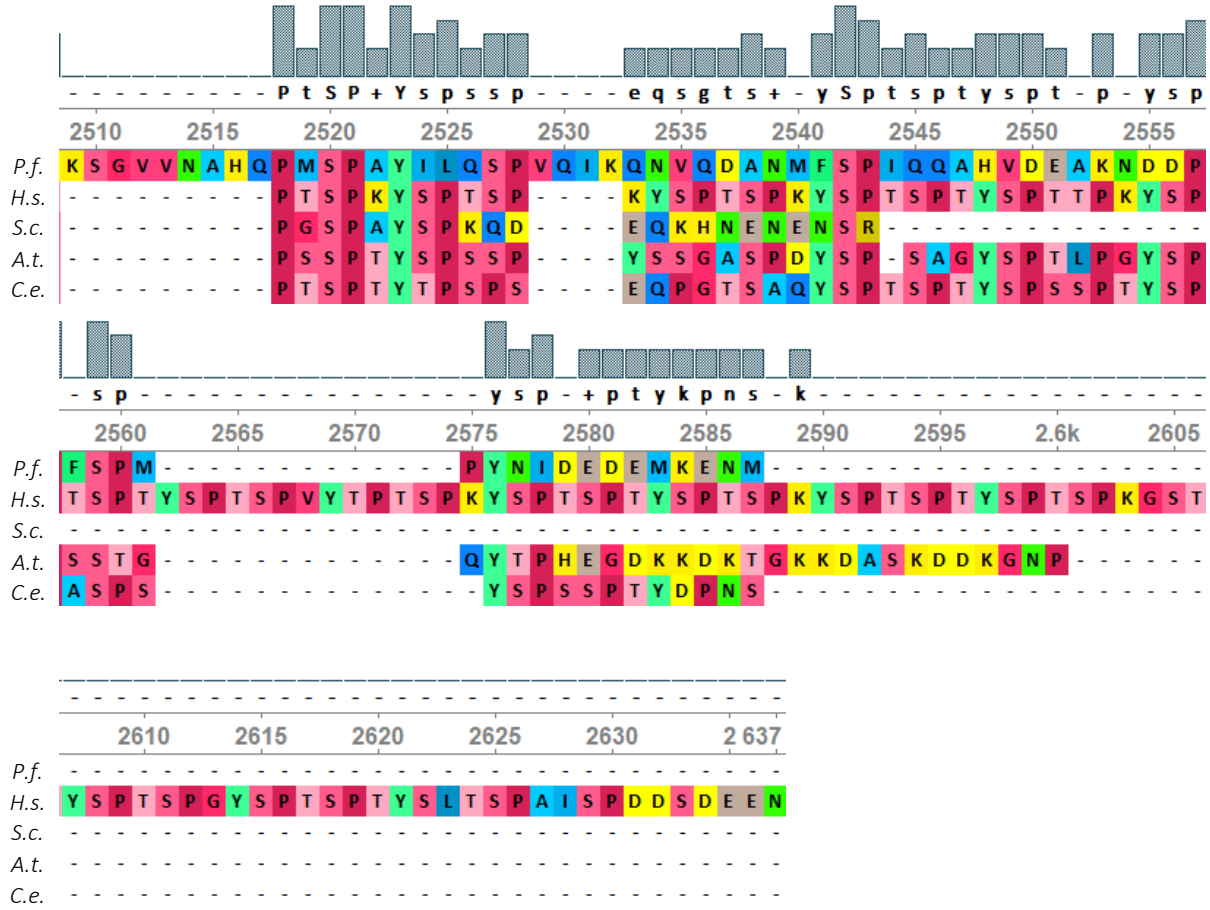


Fig. S1. MUSCLE multiple sequence alignment of the *Pf*RPB1-CTD and homologous protein regions from eukaryotic model organisms. Full length RPB1 protein sequences were aligned, but only the CTD-portion of the alignment is depicted. Blue bars indicate the conservation at a given residue, with a consensus sequence given below. *Species abbreviations, P.f.: Plasmodium falciparum; H.s.: Homo sapiens; S.c.: Saccharomyces cerevisiae; A.t.: Arabadopsis thaliana; C.e.: Caenorhabditis elegans.*

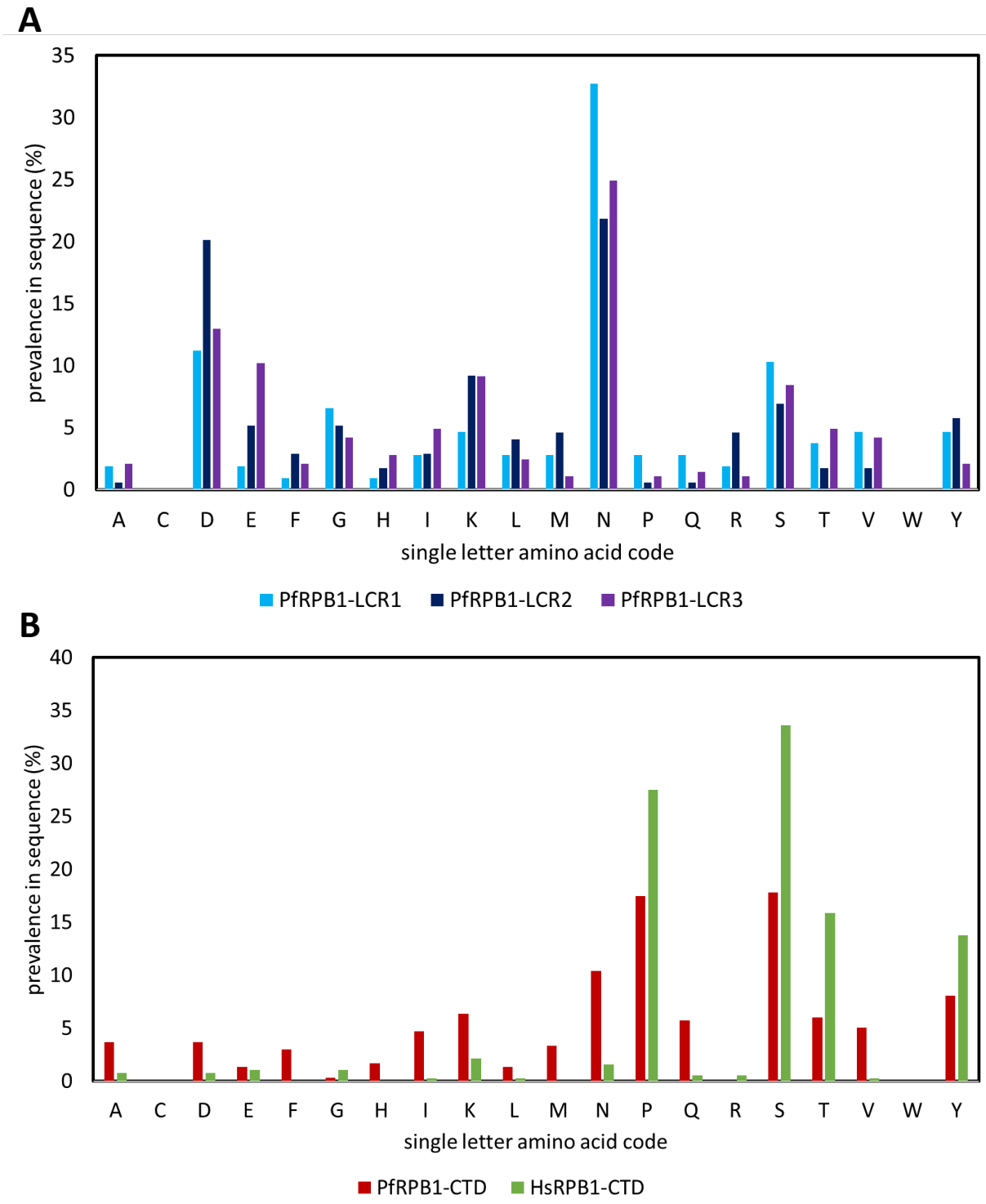
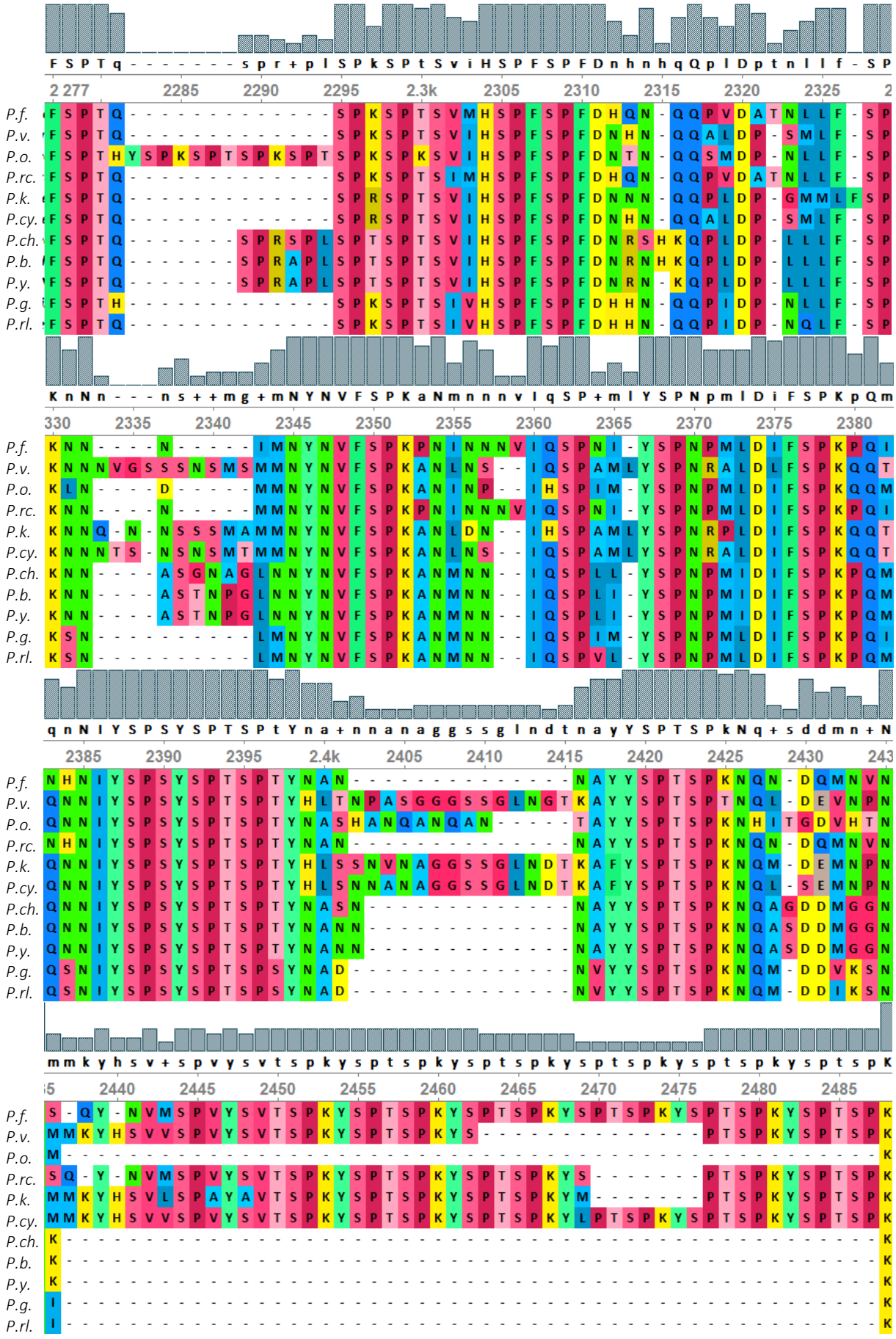


Fig. S2. Protein sequence composition analysis of *Pf*RPB1- and *Hs*RPB1. (A) Amino acid composition of the three inserted low complexity regions (LCRs) in the *Pf*RPB1 sequence. These regions are enriched in asparagine (N), acidic (D,E) and lysine (K) residues. (**B**) Amino acid composition of the C-terminal domains (CTDs) of the *Pf*- and *Hs*RPB1 sequences. The *Hs*RPB1-CTD is enriched in the residues that compose the canonical heptad repeat (S, P, T, Y). These residues are also frequently represented in the *Pf*RPB1-CTD, along with a more diverse set including lysine (K) and asparagine (N). Amino acids are represented by their single letter code on the horizontal axis and their frequencies (% of total sequence) on the vertical axis.


P.f.
P.v.
P.o.
P.rc.
P.k.
P.cy.
P.ch.
P.b.
P.y.
P.g.
P.rl.
2 277
2285
2290
2295
2.3k
2305
2310
2315
2320
2325
330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2.4k
2405
2410
2415
2420
2425
2430
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485

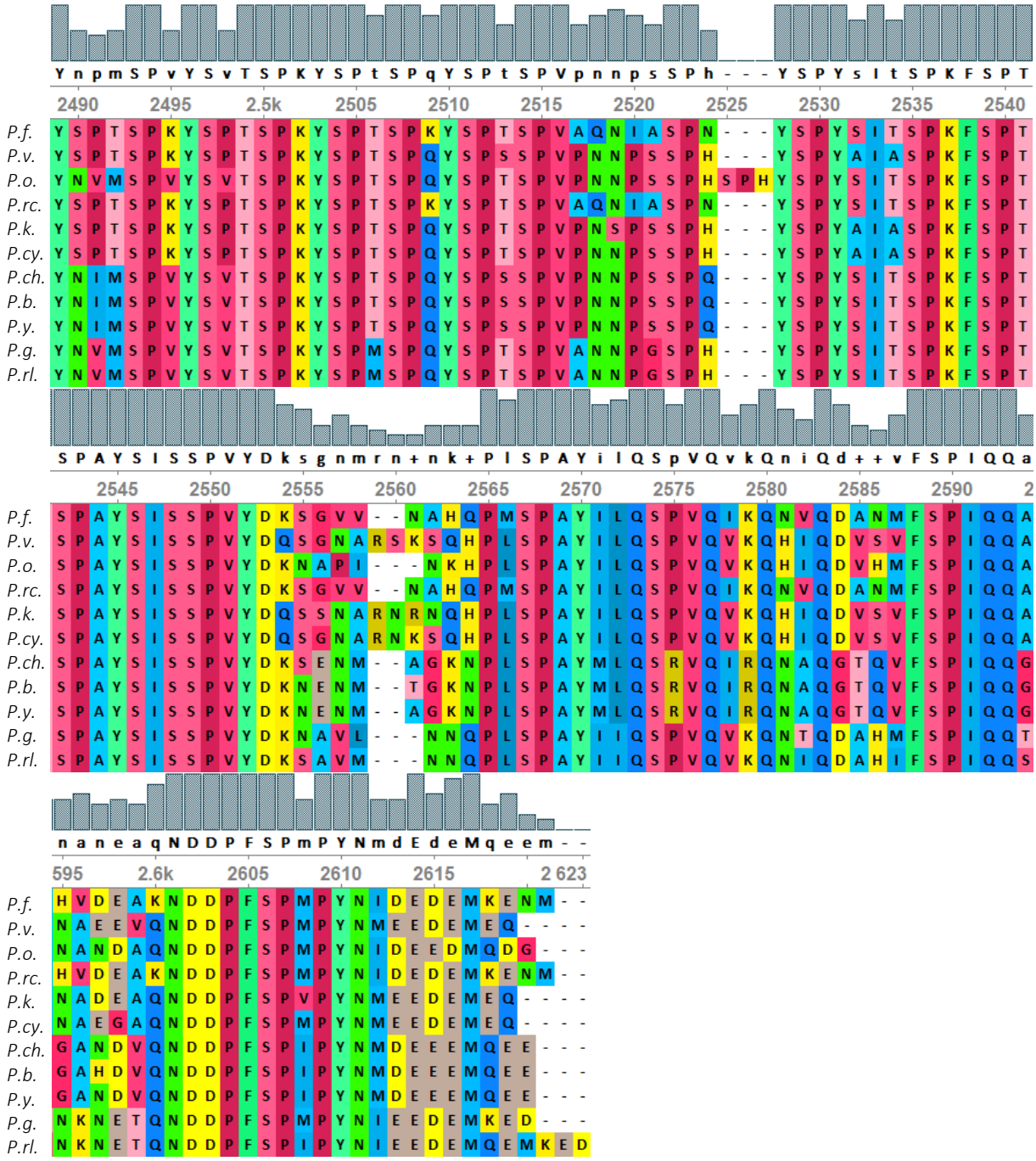


Fig. S3. MUSCLE alignment of the *Pf*RPB1-CTD and homologous proteins from other *Plasmodium* species. Full length RPB1 protein sequences were aligned, but only the CTD-portion of the alignment is depicted. Blue bars indicate the conservation at a given residue, with a consensus sequence given below. *Species abbreviations, P.f.: Plasmodium falciparum; P.v.: P. vivax; P.o.: P. ovale; P.rc.: P. reichenowi; P.k.: P. knowlesi; P.cy.: P. cynomolgi; P.ch.:P. chaubaudi; P.b.: P. berghei; P.y.: P. yoelii; P.g.: P. gallinaceum; P.rl.: P. relictum.*

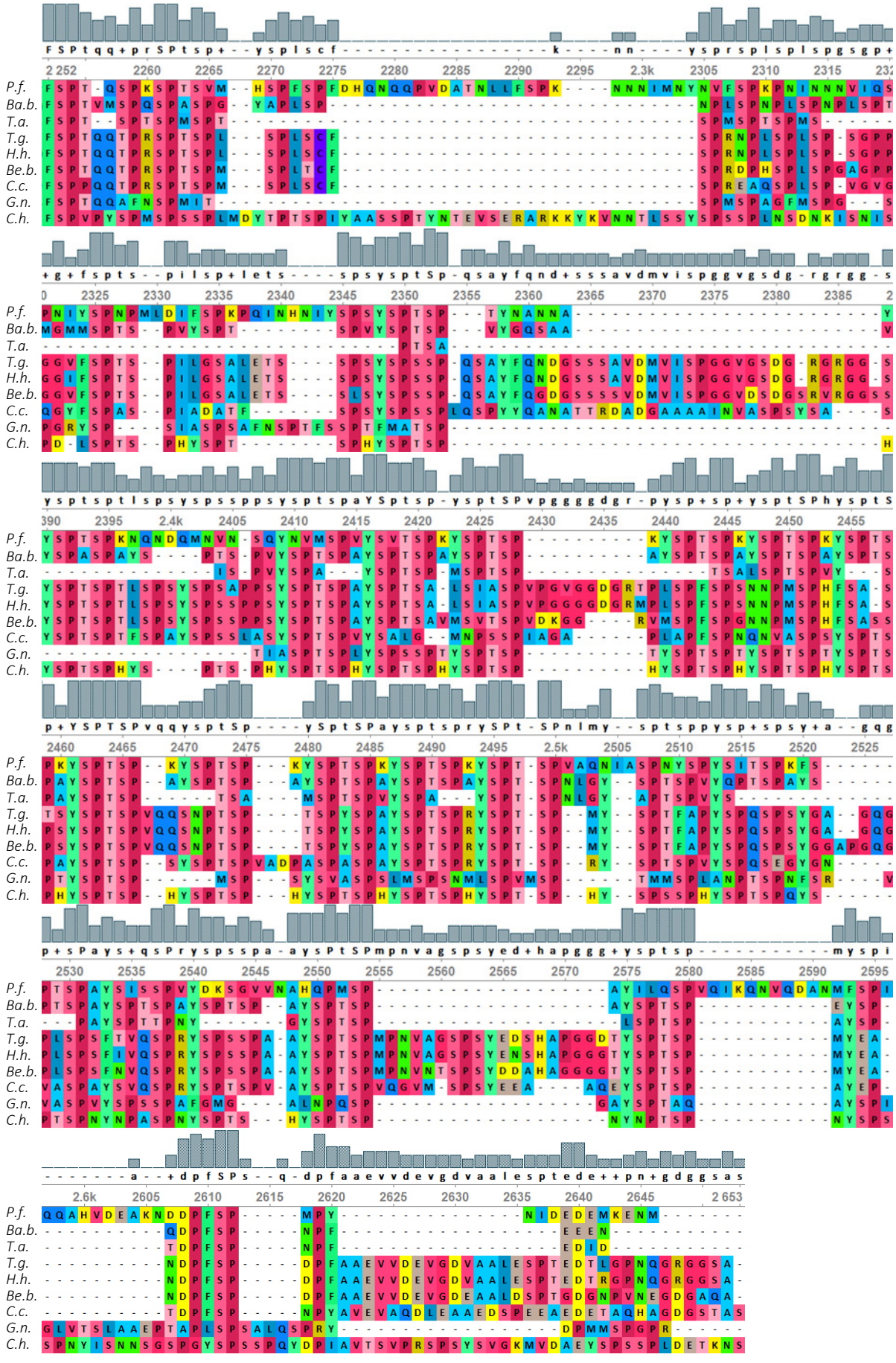

P.f.
Ba.b.
T.a.
T.g.
H.h.
Be.b.
C.c.
G.n.
C.h.

Fig. S4. (Previous page) MUSCLE alignment of the *Pf*RPB1-CTD and homologous proteins from other species in the phylum *Apicomplexa*. Full length RPB1 protein sequences were aligned, but only the CTD-portion of the alignment is depicted. Blue bars indicate the conservation at a given residue, with a consensus sequence given below. *Species abbreviations, P.f.: Plasmodium falciparum; Ba.b.: Babesia bigemina; T.a.: Theileria annulata; T.g.: Toxoplasma gondii; H.h.: Hammondia hammondi; Be.b.: Besnoita besnoiti; C.c.: Cyclospora cayetanensis; G.n.: Gregarina niphandrodes; C.h.: Cryptosporidium hominis.*

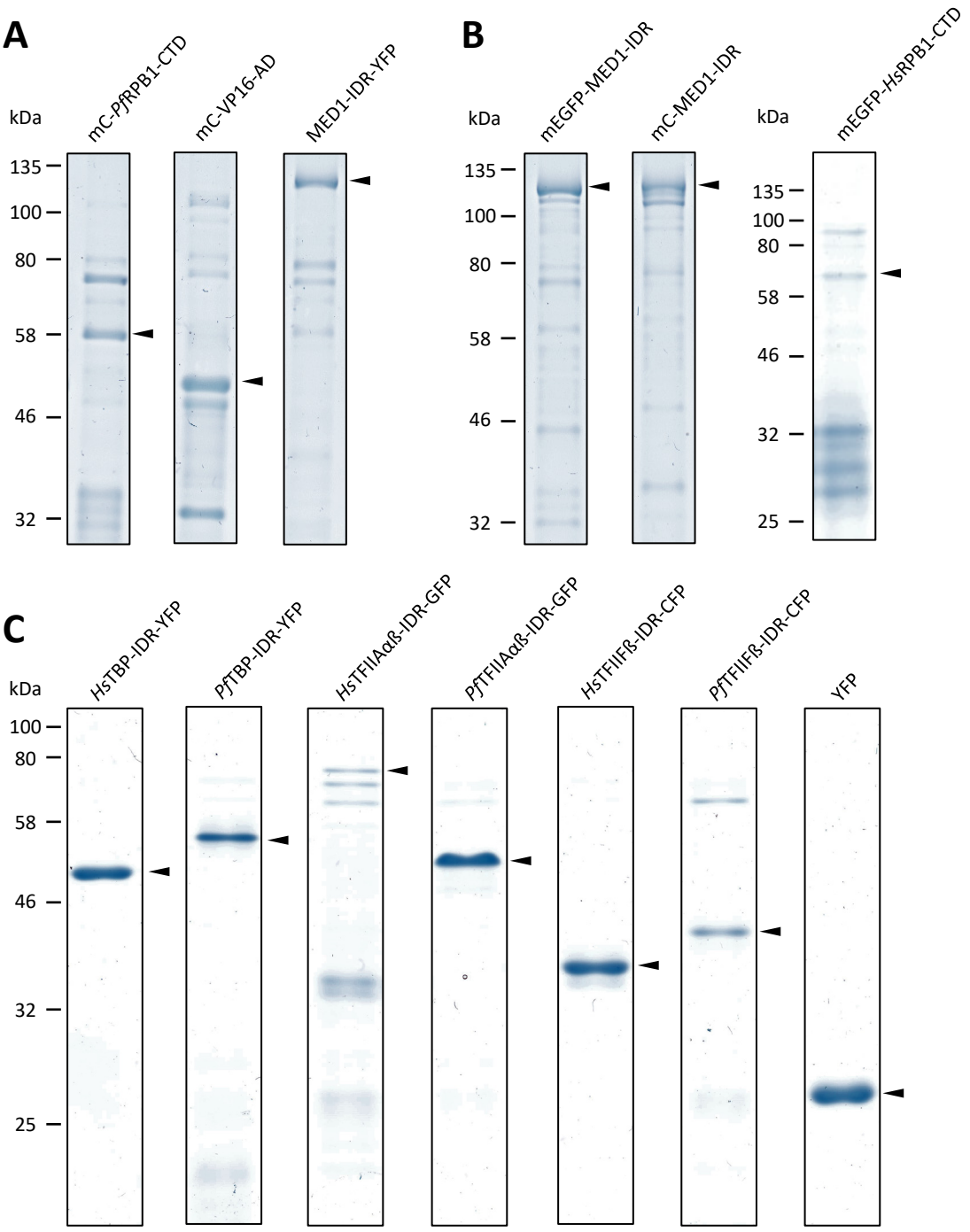


Fig. S5. SDS-PAGE of recombinant general transcription factor IDR-fluorescent protein fusion constructs. (**A**) Newly generated constructs include the CTD of *Pf*RPB1 and the AD of VP16 fused to mCherry as well as the IDR of MED1 fused to YFP. (**B**) Constructs shared by the Richard Young group include the IDR of MED1 fused to mCherry and mEGFP as well as the CTD of *Hs*RPB1 fused to mEGFP. (**C**) Previously generated constructs of *P. falciparum* and *H. sapiens* basal transcription factor IDRs fused to various fluorescent proteins [81]. These include TBP-IDRs fused to YFP, TFIIAαß-IDRs fused to GFP and TFIIFß-IDRs fused to CFP. All proteins expressed in *E. coli* and purified by nickel-affinity chromatography. 500 ng total protein, as determined by Bradford assay, of each preparation loaded on 12% resolving polyacrylamide gel and stained with Coomassie blue. Bands of protein of interest, determined by reasonable match to expected molecular weight, indicated by a black arrow to differentiate them from contaminant and truncation bands.

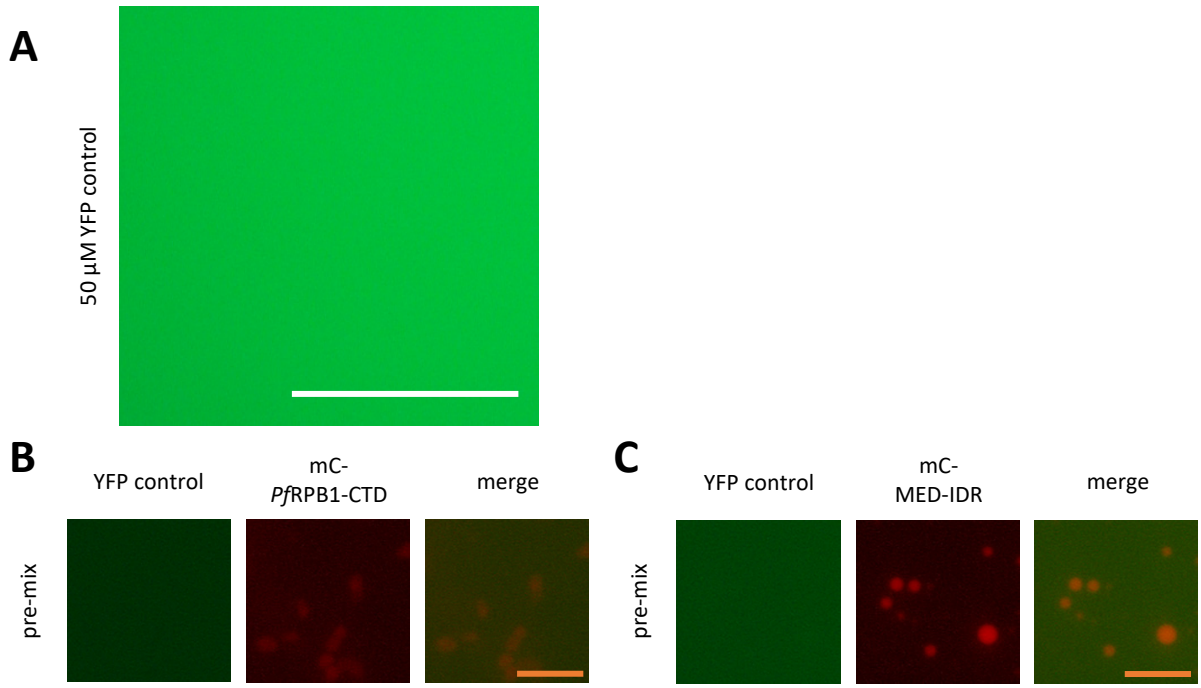


Fig. S6. YFP alone does not form condensates and is not recruited to droplets formed by *Pf*RPB1-CTD or MED1-IDR. (A) 50 µM YFP, incubated in the presence of 20% PEG-20000 and 150 mM KCl remains evenly distributed in solution. (B) YFP is not recruited to *Pf*RPB1-CTD condensates. Biomolecular condensation was induced by addition of condensation reaction buffer, containing 12% PEG-3000 and 150 mM KCl to 14 µM mC-*Pf*RPB1-CTD and 25 µM YFP. Protein preparations were combined and thoroughly mixed before addition of 2x condensation buffer to induce droplet formation. (C) YFP is not recruited to MED1-IDR condensates. As in (B), but YFP was mixed with 12 µM MED1-IDR. *Scale bars*, *white:* 100 µm; *orange:* 10 µm.

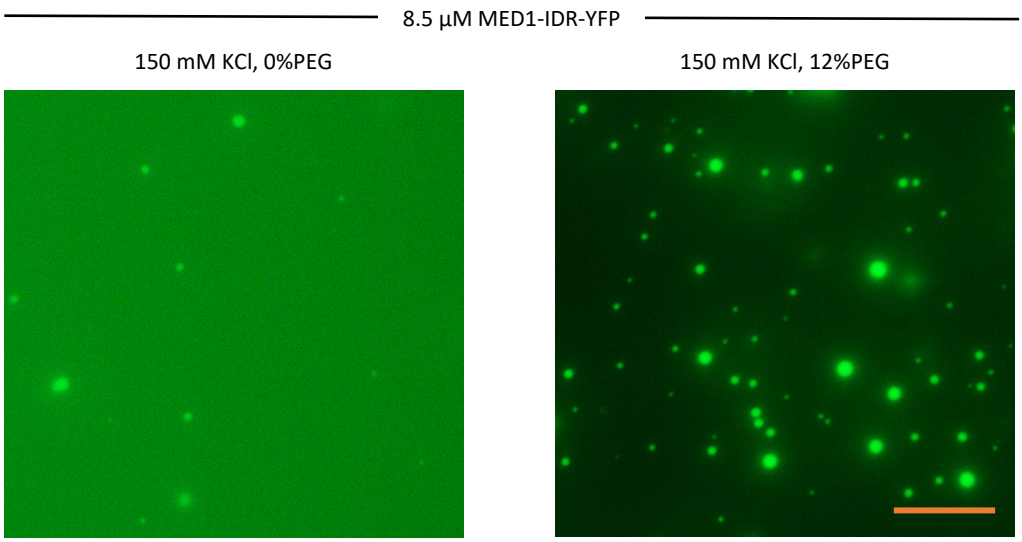


Fig. S7. MED1-IDR-YFP forms spherical droplets in the absence of PEG and condensation is stimulated by PEG. Representative images of MED1-IDR-YFP condensate formation at the maximum protein concentration of 8.5 µM, 150 mM KCl and indicated PEG concentrations. *Scale bar*: 10 µm.

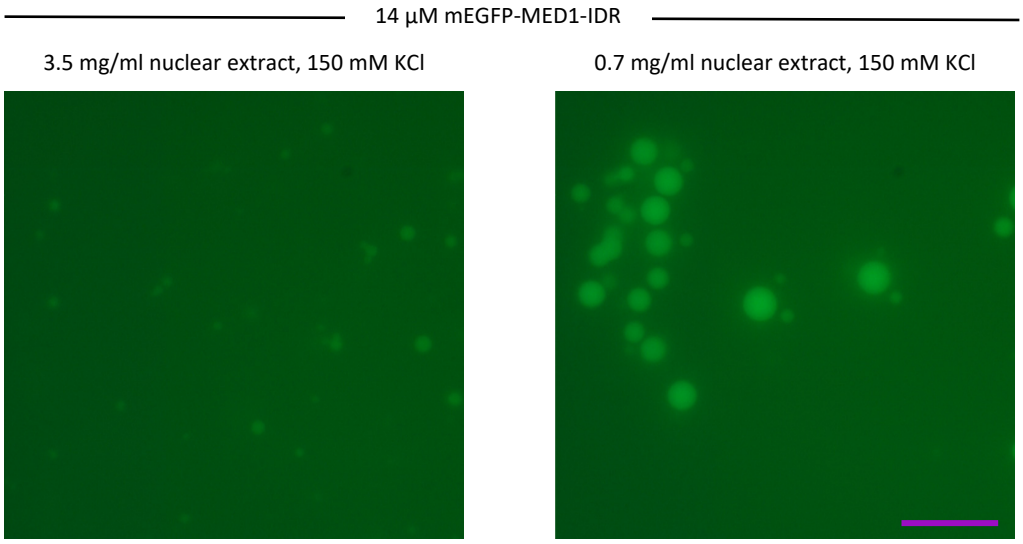


Fig. S8. mEGFP-MED1-IDR forms droplets in HeLa nuclear extract that arrange in strings. Representative images of mEGFP-MED1-IDR condensate formation at 14 µM protein concentration and 150 mM KCl. *Scale bar*: 20 µm.

References

1. Cowman, A. F., Healer, J., Marapana, D. & Marsh, K. Malaria: Biology and Disease. *Cell* **167**, 610–624 (2016).

2. Maier, A. G., Matuschewski, K., Zhang, M. & Rug, M. Plasmodium falciparum. *Trends Parasitol.* **35**, 481–482 (2019).

3. Global Malaria Programme: WHO Global. *World malaria report 2019*. *WHO Regional Office for Africa* (Geneva: World Health Organization, 2019).

4. Hollin, T. & Le Roch, K. G. From Genes to Transcripts, a Tightly Regulated Journey in Plasmodium. *Front. Cell. Infect. Microbiol.* **10**, 1–13 (2020).

5. Tan, Q. W. & Mutwil, M. Malaria.tools—comparative genomic and transcriptomic database for Plasmodium species. *Nucleic Acids Res.* **48**, D768–D775 (2020).

6. Laurens, M. B. RTS,S/AS01 vaccine (Mosquirix™): an overview. *Hum. Vaccin. Immunother.* **16**, 480–489 (2020).

7. Chappell, L. *et al.* Refining the transcriptome of the human malaria parasite Plasmodium falciparum using amplification-free RNA-seq. *BMC Genomics* **21**, 395 (2020).

8. Abel, S. & Le Roch, K. G. The role of epigenetics and chromatin structure in transcriptional regulation in malaria parasites. *Brief. Funct. Genomics* **18**, 302–313 (2019).

9. Zilversmit, M. M. *et al.* Low-Complexity Regions in Plasmodium falciparum: Missing Links in the Evolution of an Extreme Genome. *Mol. Biol. Evol.* **27**, 2198–2209 (2010).

10. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498–511 (2002).

11. Ruvalcaba-Salazar, O. K. *et al.* Recombinant and native Plasmodium falciparum TATA-binding-protein binds to a specific TATA box element in promoter regions. *Mol. Biochem. Parasitol.* **140**, 183–196 (2005).

12. Muralidharan, V. & Goldberg, D. E. Asparagine Repeats in Plasmodium falciparum Proteins: Good for Nothing? *PLoS Pathog.* **9**, e1003488 (2013).

13. Bing, S. Expression and initial characterisation of the Plasmodium falciparum general transcription factors TFIIB and TLP. (University of Cape Town, 2015).

14. Milton, R. A. Regulation of transcription in Plasmodium falciparum, the causative agent of severe malaria: initial characterisation of PfTBP and PfTFIIA. (University of Cape Town, 2015).

15. van der Linden, L. Functional Characterization of Plasmodium falciparum TATA-box binding-like Protein. (University of Cape Town, 2019).

16. Bischoff, E. & Vaquero, C. In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of Plasmodium falciparum. *BMC Genomics* **11**, 34 (2010).

17. Callebaut, I., Prat, K., Meurice, E., Mornon, J.-P. & Tomavo, S. Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: conserved features and differences relative to other eukaryotes. *BMC Genomics* **6**, 100 (2005).

18. Soutourina, J. Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* **19**, 262–274 (2018).

19. Jeninga, M., Quinn, J. & Petter, M. ApiAP2 Transcription Factors in Apicomplexan Parasites. *Pathogens* **8**, 47 (2019).

20. Pallarès, I. *et al.* Discovering Putative Prion-Like Proteins in Plasmodium falciparum: A Computational and Experimental Analysis. *Front. Microbiol.* **9**, 1737 (2018).

21. Toenhake, C. G. & Bártfai, R. What functional genomics has taught us about transcriptional regulation in malaria parasites. *Brief. Funct. Genomics* **00**, 1–12 (2019).

22. Otto, T. D. *et al.* New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Mol. Microbiol.* **76**, 12–24 (2010).

23. Thomas, M. C. & Chiang, C. The General Transcription Machinery and General Cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**, 105–178 (2006).

24. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).

25. Malkowska, M., Kokoszynska, K., Rychlewski, L. & Wyrwicz, L. Structural bioinformatics of the general transcription factor TFIID. *Biochimie* **95**, 680–691 (2013).

26. Schier, A. C. & Taatjes, D. J. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* **34**, 465–488 (2020).

27. Verger, A., Monté, D. & Villeret, V. Take Your PIC. *Trends Biochem. Sci.* **46**, 705–707 (2021).

28. Putlyaev, E. V, Ibragimov, A. N., Lebedeva, L. A., Georgiev, P. G. & Shidlovskii, Y. V. Structure and Functions of the Mediator Complex. **83**, (2018).

29. Harlen, K. M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* **18**, 263–273 (2017).

30. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936 (2006).

31. Eick, D. & Geyer, M. The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chem. Rev.* **113**, 8456–8490 (2013).

32. Dieci, G. Removing quote marks from the RNA polymerase II CTD 'code'. *Biosystems* **207**, 104468 (2021).

33. Chapman, R. D., Heidemann, M., Hintermair, C. & Eick, D. Molecular evolution of the RNA polymerase II CTD. *Trends Genet.* **24**, 289–296 (2008).

34. Yang, C. & Stiller, J. W. Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain. *Proc. Natl. Acad. Sci.* **111**, 5920–5925 (2014).

35. Jeronimo, C., Bataille, A. R. & Robert, F. The Writers, Readers, and Functions of the RNA Polymerase II C-Terminal Domain Code. *Chem. Rev.* **113**, 8491–8522 (2013).

36. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

37. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017).

38. Material, S. O. *et al.* Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science (80-.).* **1729**, 1729–1732 (2009).

39. Uversky, V. N. Supramolecular Fuzziness of Intracellular Liquid Droplets: Liquid–Liquid Phase Transitions, Membrane-Less Organelles, and Intrinsic Disorder. *Molecules* **24**, 3265 (2019).

40. Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).

41. Guillen-Chable, F., Bayona, A., Rodríguez-Zapata, L. C. & Castano, E. Phase Separation of Intrinsically Disordered Nucleolar Proteins Relate to Localization and Function. *Int. J. Mol. Sci.* **22**, 13095 (2021).

42. Yang, P. *et al.* G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325-345.e28 (2020).

43. Nosella, M. L. & Forman-Kay, J. D. Phosphorylation-dependent regulation of messenger RNA transcription, processing and translation within biomolecular condensates. *Curr. Opin. Cell Biol.* **69**, 30–40 (2021).

44. Murthy, A. C. & Fawzi, N. L. The (un)structural biology of biomolecular liquid-liquid phase separation using NMR spectroscopy. *J. Biol. Chem.* **295**, 2375–2384 (2020).

45. Murthy, A. C. *et al.* Molecular interactions underlying liquid−liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **26**, 637–648 (2019).

46. Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).

47. Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e16 (2018).

48. Boyko, S., Qi, X., Chen, T.-H., Surewicz, K. & Surewicz, W. K. Liquid–liquid phase separation of tau protein: The crucial role of electrostatic interactions. *J. Biol. Chem.* **294**, 11054–11059 (2019).

49. Franzmann, T. M. & Alberti, S. Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *J. Biol. Chem.* **294**, 7128–7136 (2019).

50. Guillén-Boixet, J. *et al.* RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181**, 346-361.e17 (2020).

51. Alshareedah, I., Moosa, M. M., Raju, M., Potoyan, D. A. & Banerjee, P. R. Phase transition of RNA−protein complexes into ordered hollow condensates. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 15650–15658 (2020).

52. Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).

53. Sabari, B. R., Dall'Agnese, A. & Young, R. A. Biomolecular Condensates in the Nucleus. *Trends Biochem. Sci.* **45**, 961–977 (2020).

54. Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *J. Biol. Chem.* **294**, 7115–7127 (2019).

55. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science (80-.).* **360**, 918–921 (2018).

56. Schuster, B. S. *et al.* Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* **9**, 2985 (2018).

57. Onuchic, P. L., Milin, A. N., Alshareedah, I., Deniz, A. A. & Banerjee, P. R. Divalent cations can control a switch-like behavior in heterotypic and homotypic RNA coacervates. *Sci. Rep.* **9**, 12161 (2019).

58. Boeynaems, S. *et al.* Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci.* **116**, 7889–7898 (2019).

59. Riback, J. A. *et al.* Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**, 209–214 (2020).

60. Nott, T. J. *et al.* Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **57**, 936–947 (2015).

61. Bratek-Skicki, A., Pancsa, R., Meszaros, B., Van Lindt, J. & Tompa, P. A guide to regulation of the formation of biomolecular condensates. *FEBS J.* **287**, 1924–1935 (2020).

62. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).

63. Hastings, R. L. & Boeynaems, S. Designer Condensates: A Toolkit for the Biomolecular Architect. *J. Mol. Biol.* **433**, 166837 (2021).

64. Feng, Z., Chen, X., Wu, X. & Zhang, M. Formation of biological condensates via phase separation: Characteristics, analytical methods, and physiological implications. *J. Biol. Chem.* **294**, 14823–14835 (2019).

65. Dao, T. P. *et al.* Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Mol. Cell* **69**, 965-978.e6 (2018).

66. Cho, W. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (80-.).* **361**, 412–415 (2018).

67. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).

68. Buckley, M. S. & Lis, J. T. Imaging RNA Polymerase II transcription sites in living cells. *Curr. Opin. Genet. Dev.* **25**, 126–130 (2014).

69. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science (80-.).* **361**, eaar3958 (2018).

70. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e16 (2018).

71. Zamudio, A. V. *et al.* Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol. Cell* **76**, 753-766.e6 (2019).

72. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science (80-.).* **361**, eaar2555 (2018).

73. Chen, G. *et al.* Taf14 recognizes a common motif in transcriptional machineries and facilitates their clustering by phase separation. *Nat. Commun.* **11**, 4206 (2020).

74. Lu, H. *et al.* Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318–323 (2018).

75. Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).

76. Henninger, J. E. *et al.* RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225.e24 (2021).

77. Gurumurthy, A., Shen, Y., Gunn, E. M. & Bungert, J. Phase Separation and Transcription Regulation: Are Super-Enhancers and Locus Control Regions Primary Sites of Transcription Complex Assembly? *BioEssays* **41**, 1–14 (2019).

78. Strickfaden, H. *et al.* Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells. *Cell* **183**, 1772-1784.e13 (2020).

79. A & Weber. Evidence for and against Liquid-Liquid Phase Separation in the Nucleus. *Non-Coding RNA* **5**, 50 (2019).

80. McSwiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev.* **33**, 1619–1634 (2019).

81. Karamanof, L. Intrinsically Disordered Regions in Plasmodium falciparum Transcription Factors : Prevalence and Comparative Analysis of Liquid-Liquid Phase Separation Properties. (University of Cape Town, 2019).

82. Okonechnikov, K., Golosova, O. & Fursov, M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).

83. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).

84. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta - Proteins Proteomics* **1804**, 996–1010 (2010).

85. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2**, 924–932 (2007).

86. Su, X. Z., Wu, Y., Sifri, C. D. & Wellems, T. E. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Res.* **24**, 1574–1575 (1996).

87. Lawrence, A.-M. & Besir, H. Staining of Proteins in Gels with Coomassie G-250 without Organic Solvent and Acetic Acid. *J. Vis. Exp.* 2–4 (2009) doi:10.3791/1350.

88. Alshareedah, I., Kaur, T. & Banerjee, P. R. Methods for characterizing the material properties of biomolecular condensates. in *Methods in Enzymology* vol. 646 143–183 (Elsevier Inc., 2021).

89. van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).

90. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

91. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

92. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

93. Malecová, B., Gross, P., Boyer-Guittaut, M., Yavuz, S. & Oelgeschläger, T. The initiator core promoter element antagonizes repression of TATA-directed transcription by negative cofactor NC2. *J. Biol. Chem.* **282**, 24767–24776 (2007).

94. Hirai, H., Tani, T. & Kikyo, N. Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *Int. J. Dev. Biol.* **54**, 1589–1596 (2010).

95. Ceballos, A. V., McDonald, C. J. & Elbaum-Garfinkle, S. Methods and Strategies to Quantify Phase Separation of Disordered Proteins. *Methods Enzymol.* **611**, 31–50 (2018).

96. Mitrea, D. M. *et al.* Methods for Physical Characterization of Phase-Separated Bodies and Membrane-less Organelles. *J. Mol. Biol.* **430**, 4773–4805 (2018).

97. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

98. Klein, I. A. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science (80-.).* **368**, 1386–1392 (2020).

99. Basu, S. *et al.* Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* **181**, 1062-1079.e30 (2020).

100. Ghosh, A., Mazarakos, K. & Zhou, H.-X. Three archetypical classes of macromolecular regulators of protein liquid–liquid phase separation. *Proc. Natl. Acad. Sci.* **116**, 19474–19483 (2019).

101. Kroschwald, S., Maharana, S. & Simon, A. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters* 1–7 (2017) doi:10.19185/matters.201702000010.

102. Erdel, F. *et al.* Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation. *Mol. Cell* **78**, 236-249.e7 (2020).

103. Claeys Bouuaert, C. *et al.* DNA-driven condensation assembles the meiotic DNA break machinery. *Nature* **592**, 144–149 (2021).

104. Wysocka, J. & Herr, W. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends Biochem. Sci.* **28**, 294–304 (2003).

105. Milbradt, A. G. *et al.* Structure of the VP16 transactivator target in the Mediator. *Nat. Struct. Mol. Biol.* **18**, 410–415 (2011).

106. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

107. Kishore, S. P., Perkins, S. L., Templeton, T. J. & Deitsch, K. W. An Unusual Recent Expansion of the C-Terminal Domain of RNA Polymerase II in Primate Malaria Parasites Features a Motif Otherwise Found Only in Mammalian Polymerases. *J. Mol. Evol.* **68**, 706–714 (2009).

108. Li, W.-B. *et al.* An enlarged largest subuniit of Plasmodium falciparum RNA polymerase II defines conserved and variable RNA polymerase domains. *Nucleic Acids Res.* **17**, 9621–9636 (1984).

109. Rana, U., Brangwynne, C. P. & Panagiotopoulos, A. Z. Phase separation vs aggregation behavior for model disordered proteins. *J. Chem. Phys.* **155**, 1–13 (2021).

110. Shrinivas, K. *et al.* Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell* **75**, 549-561.e7 (2019).

111. Wei, M.-T. *et al.* Nucleated transcriptional condensates amplify gene expression. *Nat. Cell Biol.* **22**, 1187–1196 (2020).

112. Michowicz, J. Characterising the role of Plasmodium falciparum TATA-box binding-like Protein low complexity region 1 (Pf TLP-LCR1) in liquid-liquid phase separation. (University of Cape Town, 2019).

113. Shen, Y. *et al.* Biomolecular condensates undergo a generic shear-mediated liquid-to-solid transition. *Nat. Nanotechnol.* **15**, 841–847 (2020).

114. Quail, T. *et al.* Force generation by protein–DNA co-condensation. *Nat. Phys.* **17**, 1007–1012 (2021).

115. Suzuki, M. The heptad repeat in the largest subunit of RNA polymerase II binds by intercalating into DNA. *Nature* **344**, 562–565 (1990).

116. Rhine, K., Vidaurre, V. & Myong, S. RNA Droplets. *Annu. Rev. Biophys.* **49**, 247–265 (2020).

117. Wright, R. H. G., Le Dily, F. & Beato, M. ATP, Mg 2+ , Nuclear Phase Separation, and Genome Accessibility. *Trends Biochem. Sci.* **44**, 565–574 (2019).

118. Wiedner, H. J. & Giudice, J. It's not just a phase: function and characteristics of RNA-binding proteins in phase separation. *Nat. Struct. Mol. Biol.* **28**, 465–473 (2021).

119. Klein, I. A. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science (80-.).* **368**, 1386–1392 (2020).