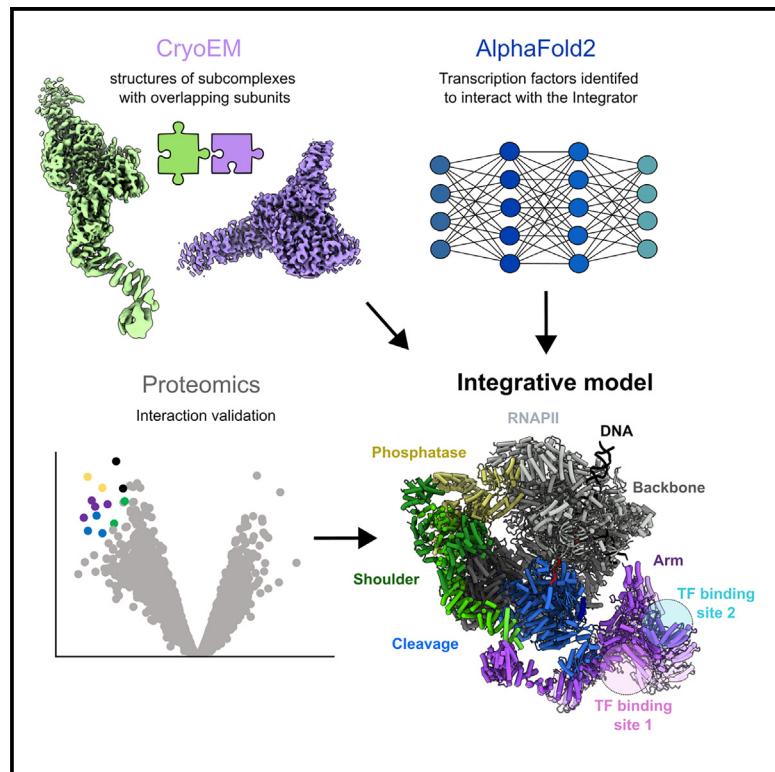


Structural basis of the Integrator complex assembly and association with transcription factors

Graphical abstract



Authors

Michał Razew, Angélique Fraudeau,
Moritz M. Pfleiderer, Romain Linares,
Wojciech P. Galej

Correspondence

wgalej@embl.fr

In brief

Razew et al. report two cryo-EM structures that complete the model of the fully assembled Integrator complex. Employing an AlphaFold2-based protein-protein interaction screen, the authors unravel interactions between Integrator and transcription factors, suggesting an additional layer of regulation of transcription attenuation.

Highlights

- Cryo-EM reveals structures of two complexes: INTS10/13/14/15 and INTS5/8/10/15
- New structures allow modeling of the fully assembled Integrator complex
- AlphaFold2 interaction screen identifies Integrator-binding transcription factors
- ZNF655 interacts with INTS13 within fully assembled Integrator



Short article

Structural basis of the Integrator complex assembly and association with transcription factors

Michał Razew,¹ Angélique Fraudeau,¹ Moritz M. Pfleiderer,^{1,2} Romain Linares,¹ and Wojciech P. Galej^{1,3,*}

¹European Molecular Biology Laboratory, EMBL Grenoble, 71 Avenue des Martyrs, 38042 Grenoble, France

²Present address: Department of Biochemistry, University of Zurich, Zurich, Switzerland

³Lead contact

*Correspondence: wgalej@embl.fr

<https://doi.org/10.1016/j.molcel.2024.05.009>

SUMMARY

Integrator is a multi-subunit protein complex responsible for premature transcription termination of coding and non-coding RNAs. This is achieved via two enzymatic activities, RNA endonuclease and protein phosphatase, acting on the promoter-proximally paused RNA polymerase II (RNAPII). Yet, it remains unclear how Integrator assembly and recruitment are regulated and what the functions of many of its core subunits are. Here, we report the structures of two human Integrator sub-complexes: INTS10/13/14/15 and INTS5/8/10/15, and an integrative model of the fully assembled Integrator bound to the RNAPII paused elongating complex (PEC). An *in silico* protein-protein interaction screen of over 1,500 human transcription factors (TFs) identified ZNF655 as a direct interacting partner of INTS13 within the fully assembled Integrator. We propose a model wherein INTS13 acts as a platform for the recruitment of TFs that could modulate the stability of the Integrator's association at specific loci and regulate transcription attenuation of the target genes.

INTRODUCTION

In metazoans, one of the critical layers of transcription regulation involves promoter-proximal pausing of the RNA polymerase II (RNAPII), controlled by its association with negative regulators of transcription-DRB sensitivity inducing factor (DSIF) and negative elongation factor (NELF).¹ Recruitment of the positive transcription elongation factor b (P-TEFb) complex and its kinase activity triggers the release of the paused polymerase into the gene body and allows for productive transcription elongation.² However, not all stalled RNAPII are destined to produce mature transcripts, and the paused polymerase can undergo a premature transcription termination through the recruitment of the multi-subunit Integrator complex.^{3–10} Integrator was discovered initially as the 3' end processing endonuclease of non-polyadenylated small nuclear RNA (snRNA) transcripts.³ More recent studies pointed to its emerging role in several aspects of the transcription of protein-coding genes and other non-coding RNAs.^{11–18} Integrator's association with the protein phosphatase PP2A-C counteracts cyclin-dependent kinase 9 (CDK9) of the P-TEFb complex and provides compelling evidence that a fine balance between these two activities plays a key role in determining the outcome of gene expression.^{19–22} Consequently, perturbations of the Integrator function are known to result in severe genetic disorders.^{23–25}

The Integrator complex assembles from fifteen canonical subunits (INTS1–15), including recently uncovered INTS15,^{26–30} and together with PP2A-C heterodimer forms a 1.5-MDa Integrator-

PP2A complex that harbors endonuclease (INTS11) and phosphatase (PP2A-C) enzymatic activities.^{31–34} Numerous other factors have been reported to interact with the Integrator, most notably transcription regulators (e.g., polymerase associated factor 1 complex; PAF1C), transcription factors (TFs) (e.g., ZNF592, ZNF687, and ZFP609) and chromatin reading/modifying enzymes (e.g., ZMYND8).^{35–37}

Biochemical and structural studies have revealed several stable sub-complexes of the Integrator, including the “cleavage module” (INTS4/9/11) containing β-CASP/MBL (CPSF-Artemis-SNM1-PSO2/Metallo-β-lactamase) catalytic endonuclease INTS11,^{38–40} as well as the INTS10/13/14 module^{41,42} and its extended variant INTS10/13/14/15,^{27,29} referred to as the “arm module” hereafter. Cryo-electron microscopy (cryo-EM) structures of higher-order Integrator-PP2A complexes revealed its overall architecture and intimate interactions between different subunits, as well as the structural basis of its recruitment to the RNAPII paused elongating complex (PEC)⁴³ and the mechanism of INTS11 endonuclease activation.^{20,44,45}

Despite the tremendous progress that has been made in structural studies of the Integrator, many of its subunits, including the entire arm module (INTS10/13/14/15), have never been visualized and their role in the context-dependent function of the Integrator is not well understood. It is also not clear how TFs and chromatin factors associate with the Integrator and what their role is in Integrator recruitment to target genes.

Here, we report structures of two Integrator sub-complexes and a structural model of a fully assembled 17-subunit Integrator



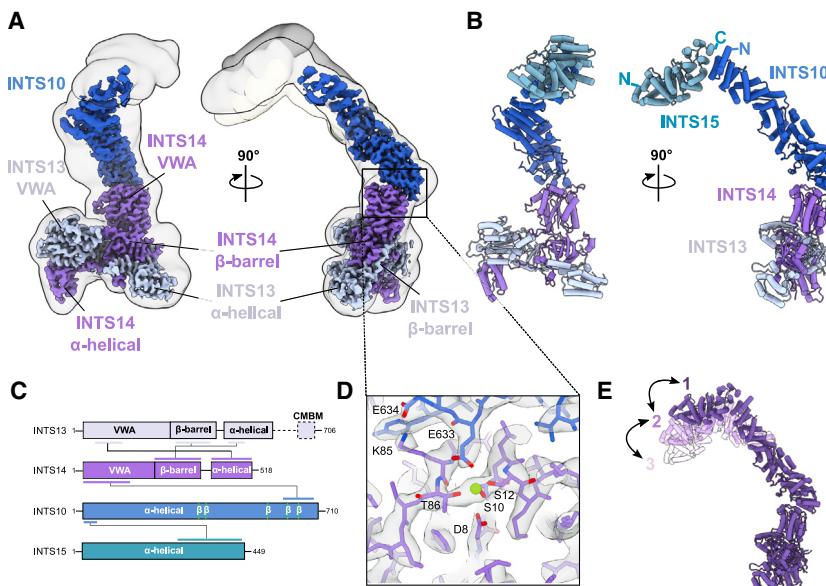


Figure 1. Cryo-EM structure of the INTS10/13/14/15 complex (arm module)

(A) High-resolution cryo-EM map colored by the subunit identity fitted in a low-contour-level map showing peripheral regions.

(B) Atomic model of the Integrator arm module.

(C) Schematic of the domain organization of each subunit of the Integrator arm module. Colored lines indicate interacting regions. The dashed line in INTS13 indicates regions that are missing in our density, including the cleavage-module-binding motif (CMBM).

(D) Zoom-in on the INTS10-INTS14 interface showing amino acid residues fitting into the cryo-EM density. The metal ion within the metal-ion-dependent adhesion site (MIDAS) of the INTS14 VWA domain is highlighted as a green sphere.

(E) Atomic models representing different states of the Integrator arm module captured by three-dimensional (3D) classification of the peripheral regions shown in (A). State 2 corresponds to the atomic model shown in (B).

bound to PEC. This, together with an AlphaFold2 (AF2)-based high-throughput *in silico* protein-protein interaction screen, revealed the structural basis for the Integrator interaction with a zinc-finger TF ZNF655. Our case study may serve as a more general model for understanding the mechanism of Integrator association with a broader spectrum of adaptor proteins.

RESULTS

Structure of the Integrator arm module

Integrator arm-module-containing subunits INTS10/13/14/15 have been previously characterized biochemically,^{27,29} but experimental data supporting their overall architecture remained missing. We expressed the 270-kDa INTS10/13/14/15 quaternary complex recombinantly in insect cells (Figures S1A and S1B) and determined its structure using single-particle cryo-EM to 3.3 Å resolution (Figures 1A, 1B, and S2; Table 1). The overall architecture of the arm module is elongated and hook-shaped, with INTS15 and INTS13/14 separated from each other by nearly 200 Å and connected exclusively by an extended and mostly α-helical INTS10 (Figure 1B). INTS15 interacts with INTS10 in a head-to-tail arrangement, and a weaker density in this region suggests potential conformational dynamics (Figure 1A). Focused classification and refinement of this region resolved three distinct states of complex accounting for the mobility of INTS15 and the N-terminal region of INTS10 (Figures 1A and 1E). The INTS14 Von Willebrand factor A domain (INTS14^{VWA}) coordinates a metal ion at the interface with INTS10, which is also bound by E633 of INTS10 (Figure 1D). This interaction resembles the previously reported Mg²⁺-mediated intercellular adhesion molecule 3 (ICAM-3) and integrin α_Lβ₂ interface.⁴⁶ Binary interactions formed by the components of the arm module are in good agreement with the previously proposed integrative model²⁹ and the crystal structure of INTS13/14⁴² (Figure S3).

Interaction of the arm module with the core of the Integrator complex

Although we obtained insights into the structure of the isolated Integrator arm module, it remained unclear how it interacts with other Integrator components within the fully assembled complex. Two contacts between the arm module and the Integrator's core have been previously reported:

(1) cleavage-module-binding motif (CMBM) located at the C terminus of INTS13 interaction with INTS4/9/11^{42,47} and (2) INTS5-INTS15 interaction predicted using the AF2 interaction screen.²⁹ The molecular basis for both of these interactions remains unknown.

To directly address this knowledge gap, we reconstituted a quaternary complex of INTS5/8/10/15 and determined its cryo-EM structure at the overall 3.2 Å resolution (Figures 2 and S1D-S1G). High anisotropy of the reconstruction does not allow for accurate side-chain modeling, therefore the map was only used after low-pass filtering for fitting secondary structure elements (Figures S4 and S5). The best-defined region of the map is centered around INTS5/8 (Figure 2A), which remains unchanged when compared with previously reported structures of the Integrator-PP2A²⁰ and Integrator-PEC complexes.^{44,45} The map quality gradually decreases toward the region encompassing INTS15 and INTS10 (Figure 2A). Although the limited resolution did not allow for the *de novo* building of the atomic models of INTS10 and INTS15, the structures of both proteins could be fitted into the map based on the arm module reconstruction (Figure 2B). INTS5/8/10/15 complex has an elongated architecture spanning nearly 300 Å in the longest dimension. INTS15 plays a key role in bridging INTS5/8 and INTS10, which otherwise make no contact with one another.

To validate the INTS5-INTS15 interface observed in our reconstruction (Figure 2E), we designed point mutation variants of INTS15 (INTS15^{M120AV124A} and INTS15^{L69AVL72A}) that are expected to abolish this interaction and we analyzed their effects

Table 1. Cryo-EM data collection and refinement statistics

| | INTS10/13/14 core (EMDB-19851) (PDB: 9EOC) | INTS10/13/14/15 complete (EMDB-19871) (PDB: 9EP1) | INTS5/8/15 core (EMDB-19872) (PDB: 9EP4) | INTS5/8/10/15 complete (EMDB-19853) (PDB: 9EOF) |
|--|--|--|--|--|
| Data collection and processing | | | | |
| Magnification | 105,000 | – | 105,000 | 130,000 |
| Voltage (kV) | 300 | – | 300 | 200 |
| Camera | Gatan Quantum-K3 | – | Gatan Quantum-K3 | Falcon 4i |
| Electron exposure (e ⁻ /Å ²) | 40.0 | – | 40.0 | 40.0 |
| Exposure rate (e ⁻ /pixel ² s) | 20.5 | – | 14.6 | 10.5 |
| Number of frames | 40 | – | 40 | 40 |
| Defocus range (μm) | –0.9 to –2.2 | – | –0.8 to –2.5 | –0.8 to –2.5 |
| Pixel size (Å) | 0.82 | – | 0.82 | 0.86 |
| Symmetry imposed | C1 | – | C1 | C1 |
| Movies collected (no.) | 9,253 | – | 3,682 | 8,425 |
| Initial particle images (no.) | 1,282,824 | – | 734,626 | 1,396,521 |
| Final particle images (no.) | 258,952 | 54,007 | 95,992 | 65,387 |
| Map resolution Å | | | | |
| FSC _{0.143} (unmasked/masked) | 4.1/3.3 | 4.0/3.9 | 4.5/3.2 | 10.0/7.7 |
| Map resolution range (Å) | 2.8–6.0 | 2.8–15.0 | 2.8–7.0 | 6.9–15.0 |
| Refinement | | | | |
| Initial model used (PDB code) | 6SN1 | 6SN1 | 7PKS | 7PKS |
| Resolution cutoff (Å) | 3.3 | 3.9 | 3.2 | – |
| Map sharpening B factor (Å ²) | –112 | –88 | –65 | –524 |
| Model composition | | | | |
| Non-hydrogen atoms | 11,654 | 14,339 | 15,399 | 19,577 |
| Protein residues | 1,563 | 1,923 | 2,064 | 2,613 |
| Ligands | 1 | – | – | – |
| B factors (Å²) | | | | |
| Protein | 28.3 | 93.8 | 79.8 | not refined |
| Ligand | 14.8 | – | – | – |
| RMSD | | | | |
| Bond lengths (Å) | 0.002 | 0.003 | 0.003 | 0.004 |
| Bond angles (°) | 0.547 | 0.610 | 0.601 | 0.698 |
| Validation | | | | |
| MolProbity score | 2.17 | 2.20 | 1.92 | 2.01 |
| Clashscore | 8.6 | 9.7 | 12.0 | 13.0 |
| Poor rotamers (%) | 2.9 | 2.8 | 1.6 | 0.3 |
| Ramachandran plot | | | | |
| Favored (%) | 92.5 | 95.0 | 96.2 | 96.6 |
| Allowed (%) | 7.4 | 4.9 | 3.7 | 3.3 |
| Disallowed (%) | 0.1 | 0.1 | 0.1 | 0.1 |

in vitro and *in vivo*. First, we transiently expressed INTS15 and its variants in HEK293T cells, purified them using hemagglutinin (HA)-agarose resin, and subjected them to tandem mass tag (TMT) differential quantitative proteomics workflow (Figure 2D). When compared with the INTS15^{WT} (wild type), both the INTS15^{M120A/V124A} and INTS15^{L69A/L72A} variants have lost their ability to co-purify numerous Integrator subunits (Figures 2D and

S6A). These include INTS5, a direct interacting partner of INTS15 that, together with INTS8, was among the highest-ranking hits. Similarly, enrichments of the phosphatase (PP2A-INTS6) and backbone modules (INTS1/2/7/12) were decreased in the INTS15^{M120A/V124A} and INTS15^{L69A/L72A} mutants, which is to be expected as both of these modules are known to interact with INTS5/8.²⁰ Notably, no significant changes to the abundance of

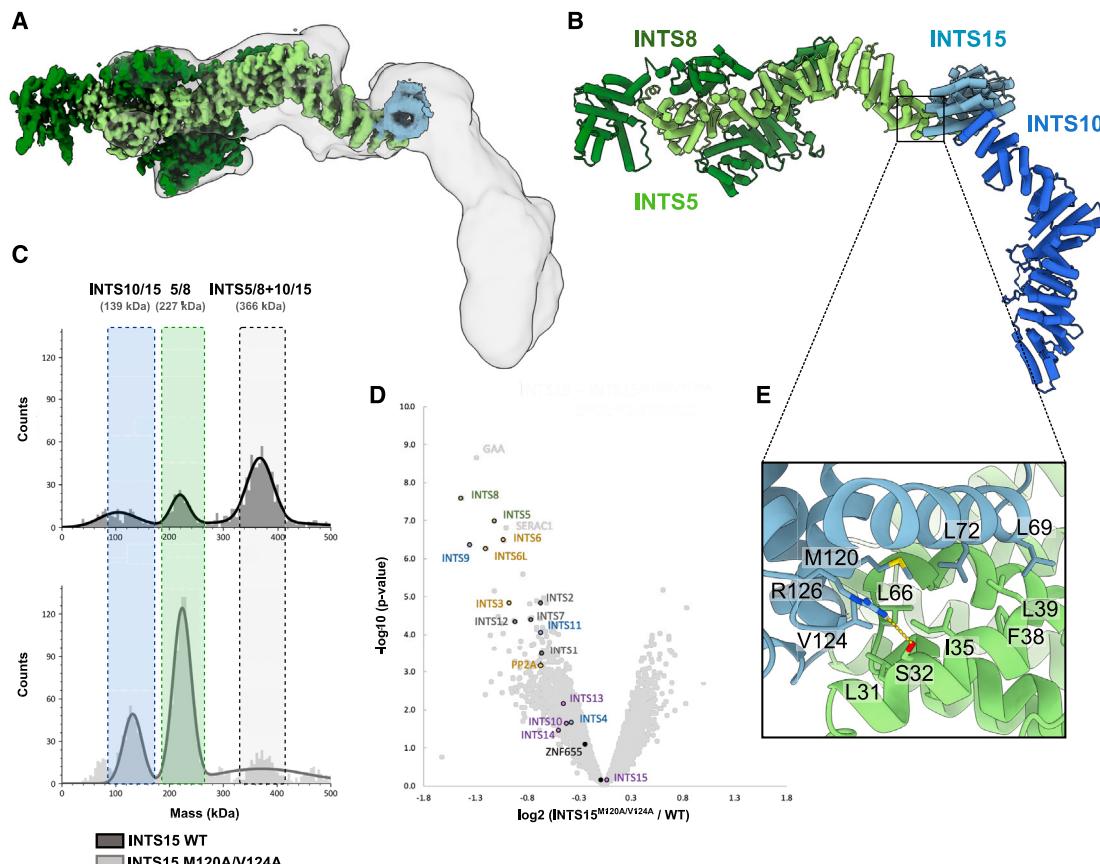


Figure 2. Cryo-EM structure of the INTS5/8/10/15 complex

- (A) High-resolution cryo-EM map colored by the subunit identity fitted in a low-contour-level map showing peripheral regions.
- (B) Integrative atomic model of the INTS5/8/10/15 complex based on experimental cryo-EM density and AF2 model.
- (C) Mass photometry analysis of the INTS5/8/10/15 complex formation from recombinant proteins containing either a wild-type INTS15 or its M102A/V124A variant predicted to abolish complex formation.
- (D) Differential quantitative mass spectrometry analysis of the proteins co-purifying with 3xHA-INTS15 (WT or M102A/V124A) transiently expressed in HEK293T cells and purified on anti-HA-agarose resin.
- (E) Zoom-in at the interface of INTS5-INTS15 modeled with AlphaFold2 and fitted into low-resolution cryo-EM reconstruction.

the INTS15 and other components of the arm module were observed, as those interactions were not targeted by the mutations discussed.

Next, we purified recombinant INTS5/8 and INTS10/15 (WT and M102A/V124A) and analyzed their association *in vitro* using mass photometry.⁴⁸ When INTS15^{WT} was used, we observed an efficient formation of a large complex with the measured molecular weight corresponding very well to the theoretical mass of INTS5/8/10/15 (367 vs. 366 kDa), while the INTS15^{M102A/V124A} mutant severely hampered complex formation (Figure 2C; Table S1).

Taken together, these data establish the molecular basis of the INTS5-INTS15 interaction and provide new insights into the mode of the arm module recruitment to the core of the Integrator.

Model of the fully assembled Integrator complex bound to RNAPII

The Integrator bound to paused elongating RNAPII forms a compact structure with multiple contacts between RNAPII and Integrator subunits, including INTS1, INTS6, INTS7, and

INTS11.^{44,45} Negative regulators of transcription DSIF and NELF play critical roles in the Integrator recruitment and activation of its endonuclease INTS11. The structure of the 4-subunit arm module has never been visualized in the context of the fully assembled Integrator. To address this issue, we first superimposed the two structures presented here via overlapping subunits, INTS10/15, to create a model of the INTS5/8/10/13/14/15 complex (Figure 3A). This structure was then oriented with respect to the Integrator-PEC complex by superposition via INTS5, a subunit well resolved in current and previous studies.^{44,45} Re-analysis of the previously deposited maps of the Integrator-PP2A complexes (EMD-33741 and EMD-30473)^{20,45} (Figures 3B and S7) shows low-resolution unassigned density, which is in good agreement with the location of the arm module obtained by the superposition and rigid body fitting (Figures 3C and S7). This density has not been interpreted previously, most likely due to the limited resolution of this region of the map and missing high-resolution information for the arm module presented here. The positioning of the arm module is also compatible with the previously reported interaction between

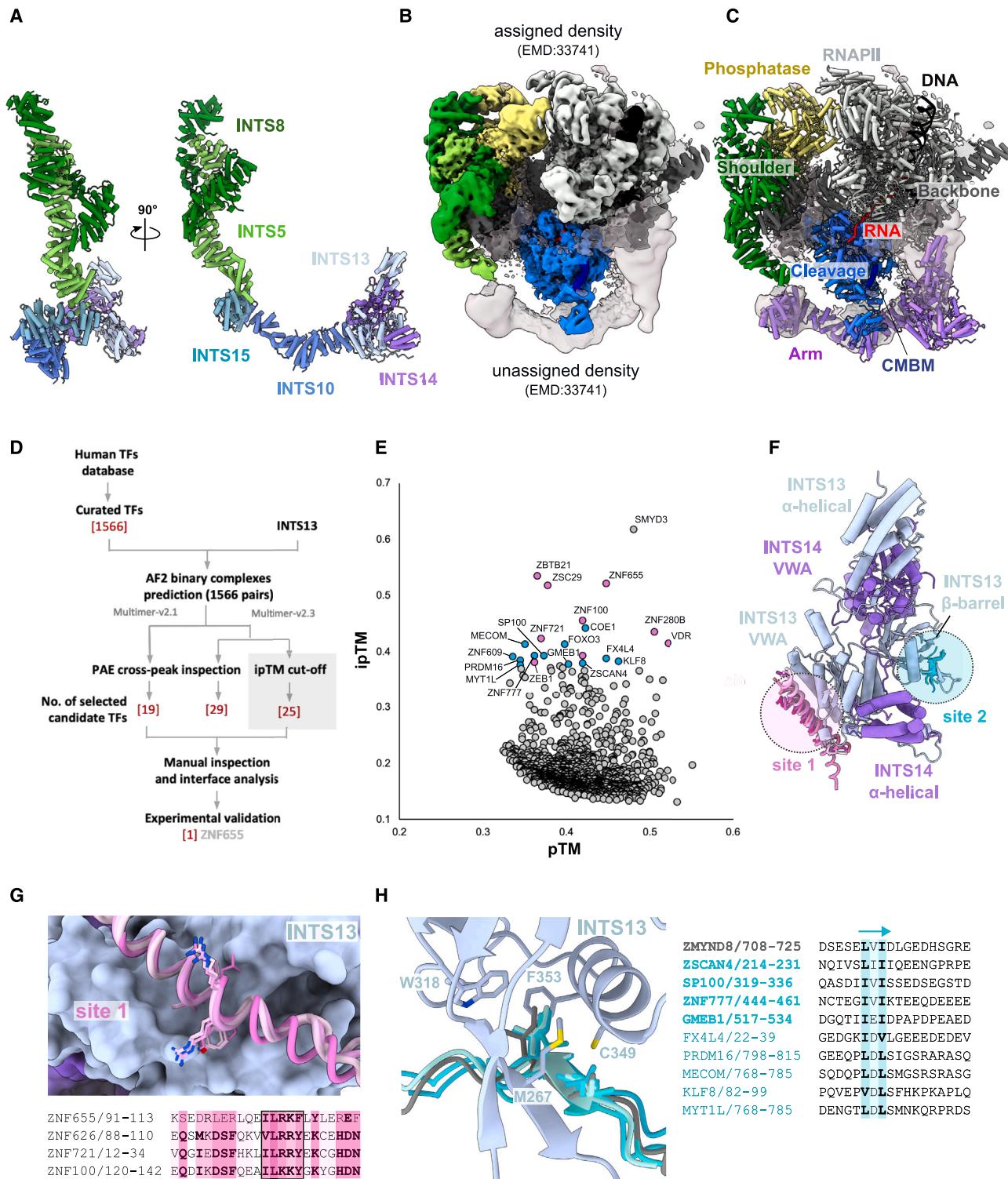


Figure 3. Model of the fully assembled Integrator complex bound to paused elongating RNAPII and transcription factors identified in an AlphaFold2 screen

(A) A composite model of the Integrator shoulder-arm module (INTS5/8/10/13/14/15) based on two experimental cryo-EM structures aligned using shared subunits (INTS10/15).

(legend continued on next page)

INTS13^{CMBM} and the cleavage module.^{42,47} Interestingly, the EMD-33741 map contains additional density near the C-terminal domain 1 (CTD1) of the INTS9/11 heterodimer, which is in excellent agreement with the INTS13^{CMBM} interaction site predicted by AF2 (Figures S8A–S8D).

Our new data, together with previous studies, allow modeling of the arm module in association with the Integrator-PEC complex. One striking feature of this model is the location of the arm module in the proximity of the DNA upstream of the transcription pause site, which would, in principle, allow their direct or indirect interaction (Figure S8E). We hypothesized that, based on its location, INTS13/14 might act as a platform to recruit additional proteins, such as TFs, which might be involved in mediating Integrator's interactions with DNA. Therefore, we set off to investigate the structural basis of the Integrator association with selected TFs.

Integrator association with TFs

Previous studies have shown that Integrator interacts with various TFs—including the so-called Z3 complex composed of ZNF592, ZNF687, and ZMYND8^{35,49}—as well as ZFP609^{35,36} and ZNF655.²⁹ However, the structural basis of these interactions remains elusive.

In order to identify a more comprehensive set of Integrator-binding TFs, we used AF2^{50,51} as a tool to screen protein-protein interactions. We used a self-curated version of the database of all human TFs⁵² and performed high-throughput structure prediction of over 1,500 TFs paired with INTS13 (Figure 3D). Based on intermolecular Predicted Aligned Error (PAE) cross-peaks from the two independent runs using different versions of the AF2 model and predicted interface template modeling scores (ipTM), we selected a subset of 25 INTS13-TFs pairs for a more detailed inspection (Figures 3D–3H). Interestingly, the majority of the analyzed TFs clustered around two distinct binding sites: on INTS13 VWA domain (site 1) and INTS13 β-barrel domain (site 2) (Figures 3E and 3F; Table S2). Site 1 is occupied by an α helix, which for at least 4 TFs has similar residues involved in the interaction with INTS13 (Figure 3G). Site 2 is occupied by a short β strand with conserved hydrophobic residues on both ends, forming a parallel or anti-parallel intermolecular β sheet with INTS13. This interface was indeed recently shown to interact with several TFs, including Integrator-associated chromatin reader ZMYND8^{53,54} (Figure 3H). One of the top hits in site 1, ZNF655, was previously co-purified with INTS13/14²⁹; therefore, we focused on the structural and biochemical characterization of the ZNF655-INTS13 interaction in more detail.

ZNF655 interaction with the fully assembled Integrator complex

ZNF655 is a 491AA protein belonging to the Kruppel-like family of TFs containing six cysteine2-histidine2-type (C2H2) zinc-finger domains at the C terminus⁵⁵ (Figure 4A). ZNF655 is known to be upregulated in human pancreatic cancers.⁵⁶ To gain further insights into ZNF655-Integrator interaction, we first reversed the AF2 screening procedure to look for binary interactions between ZNF655 and all known Integrator subunits. Based on intermolecular PAE cross-peaks, we identified INTS13 as the sole hit and the most likely binding partner of ZNF655, consistent with the result of the initial screen (Figures S9A–S9C). The predicted interface is formed between the N-terminal helix of ZNF655 and a hydrophobic cleft in the VWA domain of INTS13 (Figures 4A and 4B). To validate this interaction, we designed a series of point mutations in ZNF655 at the INTS13 interface and performed a pull-down experiment from HEK293T cells transiently co-expressing 3xFLAG-ZNF655 (WT or mutated variants) and 3xHA-INTS13. Proteins associated with the HA-agarose beads were analyzed by western blot against the FLAG-tag to probe for the presence of ZNF655 (Figure 4C). INTS13 could efficiently enrich ZNF655^{WT}, and the interaction was diminished to much lower levels in the case of ZNF655 double mutants of interface residues: F106A/I107A and I102A/I103A (Figure 4C). ZNF655^{R152A/I153A} was used as a negative control (residues not located at the interface).

Although we established that ZNF655 interacts specifically with INTS13, a broader functional context of this association remains unclear—specifically, whether this binding concerns only the arm module or also the fully assembled Integrator. To address this, we compared the proteome of ZNF655^{WT} with the ZNF655^{F106A/I107A} mutant deficient in INTS13 interaction (Figure 4D). Disruption of the ZNF655-INTS13 interface leads to the loss of not only INTS13 and the arm module subunits but also, essentially, the entire Integrator complex (Figures 4D and S6B). This provides evidence that ZNF655 associates stably with the fully assembled Integrator complex.

The ZNF655 interaction with INTS13 established here implies that its zinc fingers would be loosely tethered to Integrator via an 82 AA disordered linker, and could potentially probe their surroundings within a radius of approximately 260 Å (an equivalent of ±80 bp from the transcription start site [TSS]) (Figures 4E and S9D). That would allow them to reach the DNA upstream or downstream of the transcription pause site, as well as nascent RNA. Therefore, ZNF655 zinc fingers could be, in principle,

- (B) Cryo-EM map of the Integrator-RNAPII structure (EMD-33741) contoured at 3.2σ, colored by regions assigned to RNAPII and Integrator modules. The unassigned density was contoured at 0.25σ.
- (C) The atomic model of the fully assembled Integrator-RNAPII with the same orientation and coloring as in (B). Cleavage-module-binding motif (CMBM) of INTS13 position was identified with the AlphaFold2 prediction. The Integrator shoulder-arm module from (A) was first superimposed on the Integrator-PEC structure via INTS5 and the arm module was rigid body fitted to the unassigned density map.
- (D) A workflow of the *in silico* protein-protein interaction screen used in this study.
- (E) Predicted template modeling score (pTM) and interface predicted template modeling score (ipTM) for INTS13-TFs pairs analyzed with AlphaFold multimer v2.3. Top hits corresponding to two different INTS13 binding sites are color-coded.
- (F) INTS13 binding sites 1 and 2 of selected TFs indicated in (E) (see also Table S2).
- (G) Top: AlphaFold2 predicted helices of four different TFs bound to site 1 of INTS13. Bottom: structure-aided sequence alignment of the four helices with shared, similar residues in bold. Residues are highlighted based on sequence identity.
- (H) Left: AlphaFold2 predicted interface of five TFs bound to INTS13 site 2. Right: structure-aided sequence alignment of TFs identified in the AlphaFold2 screen that binds to site 2 of INTS13. Conserved, hydrophobic residues involved in the interaction are highlighted in bold and colored based on sequence identity.

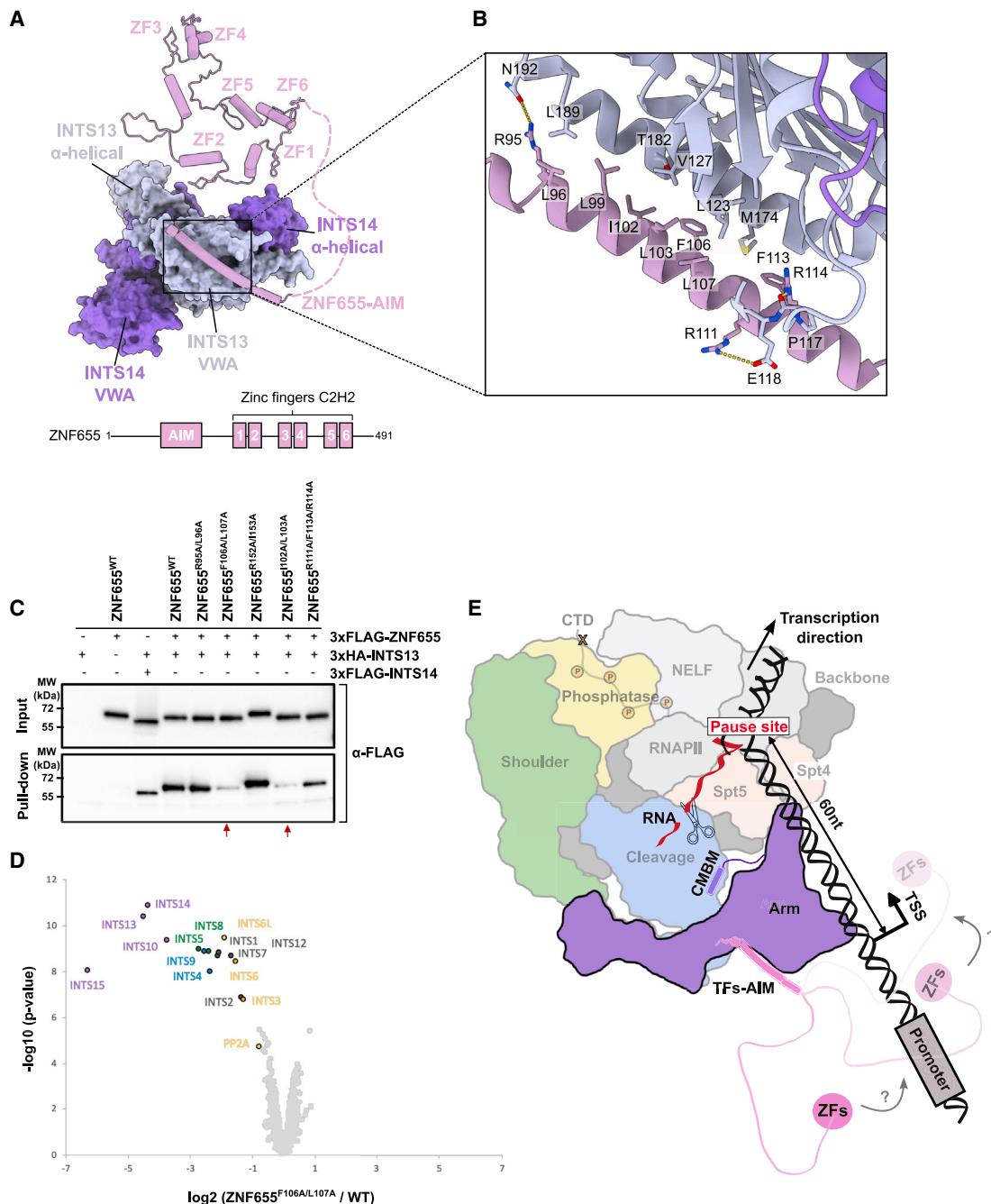


Figure 4. Transcription factor ZNF655 interacts with the Integrator complex via its arm module interacting motif

(A) Structural model of ZNF655 bound to INTS13/14 of the Integrator arm module based on the AlphaFold2 prediction and a schematic of ZNF655 domain organization. AIM, arm interacting motif; ZF1–6, zinc fingers domains 1–6.

(B) Zoom-in on the AlphaFold2 predicted ZNF655^{AIM}-INTS13^{VWA} interface, colored as in (A).

(C) ZNF655 mutagenesis and INTS13 pull-down experiment validating the interface described in (B). ZNF655 variants were used as prey and INTS13 as bait. INTS14 was used as a positive control and ZNF655^{R152A/R153A} as a negative control.

(D) Differential quantitative proteomics experiment of the 3xFLAG-ZNF655^{F106A/L107A} variant and 3xFLAG-ZNF655^{WT} transiently expressed in HEK293T cells and purified on anti-FLAG affinity resin.

(E) Schematic model of ZNF655 bound to Integrator-PEC complex. ZNF655 zinc fingers (ZFs) are connected to the arm interacting motif (AIM) via a long, disordered linker that enables it to probe the surroundings for DNA binding sites (see also Figure S9).

utilized to modulate the stability of Integrator's association with the paused RNAPII.

DISCUSSION

In this report, we present two structures of large Integrator sub-complexes, INTS5/8/10/15 and INTS10/13/14/15 (arm module), which, together with the previously deposited data, allowed us to create a complete model of the Integrator complex assembled in the context of the promoter-proximally paused RNAPII. Our model revealed the location of the previously unresolved sub-units INTS10/13/14/15, with respect to the core Integrator and RNAPII. INTS15 was identified as the pivotal factor linking the INTS13/14 heterodimer with the core of the Integrator via the extended α -helical structure of INTS10. INTS13/14 is positioned in the path of the upstream DNA extrapolated from the RNAPII structure.^{44,45} INTS13/14 has been shown to have weak and non-specific affinity for the nucleic acids,⁴² but the functional implications of this observation remained unclear. Our model suggests that these subunits could be indeed involved in direct or indirect interaction with DNA, and their positioning by the fully assembled Integrator could explain weak DNA binding properties when tested in isolation.⁴²

While this manuscript was under review, the structure of the arm-module-containing Integrator in three conformational states has been reported.⁵⁷ These new structures are in excellent agreement with our integrative model. In one of the reported conformations, Integrator subunits INTS13/14 occupy the same position as the upstream DNA and DSIF of the PEC complex; thus, the authors suggest that the arm module could be involved in opening of the DSIF clamp and facilitating displacement of nucleic acids from the RNAPII cleft. Such function would be fully compatible with our model and is further supported by the intrinsic flexibility of the arm module demonstrated in our work.

Based on INTS13/14 location, we hypothesized that it could be involved in recruiting additional DNA-binding proteins, including TFs. We used AF2⁵⁰ to exhaustively screen *in silico* all possible interactions between Integrator subunit INTS13 and all annotated human TFs.⁵² We previously showed, on a much smaller scale, that such a screening approach can uncover new and unexpected protein-protein interactions.²⁹ Here, our screen identified a small subset of TFs predicted to form high-confidence interfaces with INTS13. As a proof of concept, we performed a detailed characterization of one of them, ZNF655, whose interaction with INTS13 was confirmed experimentally. Three other factors identified in this screen (i.e., ZNF609, ZMYND8, and ZEB1) have been shown to interact with INTS13 in independent studies.^{35,53,54} These examples show the potential of a more general application of AF2-based interaction screening approaches that may aid future biochemical and structural studies.

Our data provide the first example of a structural basis for the association of Integrator with a zinc-finger TF.⁵⁸ The binding mode of ZNF655 to INTS13 is compatible with the fully assembled Integrator bound to PEC, and we show experimentally that ZNF655 co-purifies with virtually all Integrator subunits. At present, there is no clear biochemical evidence supporting ZNF655's association with DNA, chromatin, or RNA, as shown for some TFs containing RNA-binding motifs.⁵⁹ However,

ZNF655 was found in a screen for RNA-binding regions within the nuclear proteome of embryonic stem cells⁶⁰ and several studies reported its links to cell proliferation in various types of cancer,^{56,61,62} supporting its potential role as a transcription regulator. Future studies of ZNF655 DNA/RNA binding abilities and its chromatin localization are necessary to better understand its regulatory functions. Given that several other TFs are known to associate with the Integrator, the example of ZNF655 described here may represent a more general paradigm for Integrator interaction with other adaptor proteins. It remains to be seen how other TFs and/or chromatin-binding proteins are associated with the Integrator and whether their combinatorial or mutually exclusive association can play a functional role in Integrator targeting.

Integrator association with TFs points to the possibility of its targeting to specific loci through intrinsic sequence specificities of TFs.⁶³ Early work on snRNA processing suggested possible specific Integrator recruitment to the snRNA gene loci via its interaction with the so-called 3' box located in pre-snRNA transcripts.⁶⁴ However, more recent findings demonstrate that Integrator acts on diverse classes of RNAs, and no clear consensus sequences—either in the DNA around Integrator binding sites or in the RNA transcripts near its cleavage sites—could be identified.³⁴ Therefore, a universal mechanism wherein the Integrator would be directed/instructed in a sequence-specific manner seems unlikely.

Here, we envisage that the Integrator's interaction with TFs could be utilized for fine-tuning and modulation of the Integrator's binding stability and/or dwell time at the transcription pause sites rather than as a means of specific recruitment. Altering the stability of Integrator association with RNAPII at particular loci would shift the force balance in the PP2A-CDK9 axis and consequently alter the output of gene expression. Such context-specific regulation by TFs would be compatible with numerous reports on the Integrator-mediated rapid response to environmental stress and signal induction.^{18,65,66}

Limitations of the study

Although we show experimentally that ZNF655 associates with the fully assembled Integrator, our data cannot unambiguously prove that this interaction takes place in the context of the promoter proximally paused RNAPII. Additionally, it remains unclear whether ZNF655 is present together with the Integrator at all of its target sites and whether it can play an instructive role in Integrator targeting.

Our data suggest that ZNF655 zinc-finger domains could, in principle, interact with the template DNA or nascent transcript, but due to the poor solubility of the recombinant ZNF655, we could not directly test its DNA/RNA binding properties.

Our AF2 screen identified a number of TFs that could potentially associate with INTS13; however, only a few of those (i.e., ZNF655, ZNF609, ZMYND8, and ZEB1) have been verified experimentally in this and other studies. It remains to be seen whether the remaining predictions can also be substantiated by the experimental data.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
 - Protein expression and purification
 - Mass photometry analysis
 - AlphaFold2 interaction screen
 - Pull-down analysis and Western blotting
 - Negative stain electron microscopy
 - CryoEM sample preparation
 - CryoEM data collection and processing
 - Model building

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2024.05.009>.

ACKNOWLEDGMENTS

We would like to acknowledge Arnaud Krebs and Kasit Chatsirisupachai for advice and insightful discussions; Mandy Rettel, Per Haberkant, and Frank Stein from the EMBL Proteomic Core Facility for their support with the mass spectrometry data acquisition and analysis; Sarah Schneider and Joe Bartho for ensuring smooth running of the EMBL cryo-EM facilities; Magdalena Stoykova, Yonca Ozturk, and Tara Best for the technical support; Martin Pelosse from the EMBL Grenoble EEF platform for assistance with cell culture; and Caroline Mas for access to the biophysics platform. A. Peuch (IBS) and the EMBL Grenoble IT team for the support with high-performance computing. Sebastian Eustermann for critical comments on the manuscript. This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecole Universitaire de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). M.R. was supported by a fellowship from the EMBL Interdisciplinary Postdoctoral (EIPOD) program under Marie Skłodowska-Curie Actions COFUND (grant agreement no. 847543).

AUTHOR CONTRIBUTIONS

M.R. expressed and purified recombinant proteins, performed mass photometry experiments, prepared cryo-EM grids, collected and analyzed data, and built corresponding models. A.F. performed pull-down, western blotting, and sample preparation for quantitative proteomics. M.M.P. generated insect cell expression constructs and optimized recombinant expression and purification of the Integrator subunits. R.L. assisted with cryo-EM data collection. W.P.G. designed and performed the AF2 screen and supervised the project. All authors analyzed and interpreted data. M.R. and W.P.G. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 1, 2023

Revised: March 18, 2024

Accepted: May 9, 2024

Published: May 31, 2024

REFERENCES

1. Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* 13, 720–731. <https://doi.org/10.1038/nrg3293>.
2. Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* 23, 297–305. <https://doi.org/10.1016/j.molcel.2006.06.014>.
3. Baillat, D., Hakimi, M.-A., Näär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 123, 265–276. <https://doi.org/10.1016/j.cell.2005.08.019>.
4. Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* 56, 128–139. <https://doi.org/10.1016/j.molcel.2014.08.004>.
5. Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S., et al. (2014). Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* 5, 5531. <https://doi.org/10.1038/ncomms5531>.
6. Yamamoto, J., Hagiwara, Y., Chiba, K., Isobe, T., Narita, T., Handa, H., and Yamaguchi, Y. (2014). DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat. Commun.* 5, 4263. <https://doi.org/10.1038/ncomms5263>.
7. Elrod, N.D., Henriques, T., Huang, K.-L., Tatomer, D.C., Wilusz, J.E., Wagner, E.J., and Adelman, K. (2019). The integrator complex attenuates promoter-proximal transcription at protein-coding genes. *Mol. Cell* 76, 738–752.e7. <https://doi.org/10.1016/j.molcel.2019.10.034>.
8. Tatomer, D.C., Elrod, N.D., Liang, D., Xiao, M.-S., Jiang, J.Z., Jonathan, M., Huang, K.-L., Wagner, E.J., Cherry, S., and Wilusz, J.E. (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev.* 33, 1525–1538. <https://doi.org/10.1101/gad.330167.119>.
9. Lykke-Andersen, S., Žumer, K., Molska, E.Š., Rouvière, J.O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., and Jensen, T.H. (2021). Integrator is a genome-wide attenuator of non-productive transcription. *Mol. Cell* 81, 514–529.e6. <https://doi.org/10.1016/j.molcel.2020.12.014>.
10. Stein, C.B., Field, A.R., Mimoso, C.A., Zhao, C., Huang, K.-L., Wagner, E.J., and Adelman, K. (2022). Integrator endonuclease drives promoter-proximal termination at all RNA polymerase II-transcribed loci. *Mol. Cell* 82, 4232–4245.e11. <https://doi.org/10.1016/j.molcel.2022.10.004>.
11. Cazalla, D., Xie, M., and Steitz, J.A. (2011). A primate herpesvirus uses the integrator complex to generate viral microRNAs. *Mol. Cell* 43, 982–992. <https://doi.org/10.1016/j.molcel.2011.07.025>.
12. Lai, F., Gardini, A., Zhang, A., and Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399–403. <https://doi.org/10.1038/nature14906>.
13. Beckedorff, F., Blumenthal, E., daSilva, L.F., Aoi, Y., Cingaram, P.R., Yue, J., Zhang, A., Dokaneheifard, S., Valencia, M.G., Gaidosh, G., et al. (2020). The human integrator complex facilitates transcriptional elongation by endonucleolytic cleavage of nascent transcripts. *Cell Rep.* 32, 107917. <https://doi.org/10.1016/j.celrep.2020.107917>.
14. Dasilva, L.F., Blumenthal, E., Beckedorff, F., Cingaram, P.R., Gomes dos Santos, H., Edupuganti, R.R., Zhang, A., Dokaneheifard, S., Aoi, Y., Yue, J., et al. (2021). Integrator enforces the fidelity of transcriptional termination at protein-coding genes. *Sci. Adv.* 7, eabe3393. <https://doi.org/10.1126/sciadv.abe3393>.
15. Kirstein, N., Dokaneheifard, S., Cingaram, P.R., Valencia, M.G., Beckedorff, F., Gomes dos Santos, H., Blumenthal, E., Tayari, M.M., Gaidosh, G.S., and Shiekhattar, R. (2023). The Integrator complex regulates microRNA abundance through RISC loading. *Sci. Adv.* 9, ead0597. <https://doi.org/10.1126/sciadv.adf0597>.
16. Beltran, T., Pahita, E., Ghosh, S., Lenhard, B., and Sarkies, P. (2021). Integrator is recruited to promoter-proximally paused RNA Pol II to generate *Caenorhabditis elegans* piRNA precursors. *EMBO J.* 40, e105564. <https://doi.org/10.15252/embj.2020105564>.
17. Rubtsova, M.P., Vasilkova, D.P., Moshareva, M.A., Malyavko, A.N., Meerson, M.B., Zatsepin, T.S., Naraykina, Y.V., Beletsky, A.V., Ravin,

- N.V., and Dontsova, O.A. (2019). Integrator is a key component of human telomerase RNA biogenesis. *Sci. Rep.* 9, 1701. <https://doi.org/10.1038/s41598-018-38297-6>.
18. Rosa-Mercado, N.A., Zimmer, J.T., Apostolidi, M., Rinehart, J., Simon, M.D., and Steitz, J.A. (2021). Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. *Mol. Cell* 87, 502–513.e4. <https://doi.org/10.1016/j.molcel.2020.12.002>.
 19. Huang, K.-L., Jee, D., Stein, C.B., Elrod, N.D., Henriques, T., Mascibroda, L.G., Baillat, D., Russell, W.K., Adelman, K., and Wagner, E.J. (2020). Integrator recruits protein phosphatase 2A to prevent pause release and facilitate transcription termination. *Mol. Cell* 80, 345–358.e9. <https://doi.org/10.1016/j.molcel.2020.08.016>.
 20. Zheng, H., Qi, Y., Hu, S., Cao, X., Xu, C., Yin, Z., Chen, X., Li, Y., Liu, W., Li, J., et al. (2020). Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science* 370, eabb5872. <https://doi.org/10.1126/science.abb5872>.
 21. Vervoort, S.J., Welsh, S.A., Devlin, J.R., Barbieri, E., Knight, D.A., Offley, S., Bjelosevic, S., Costacurta, M., Todorovski, I., Kearney, C.J., et al. (2021). The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell* 184, 3143–3162.e32. <https://doi.org/10.1016/j.cell.2021.04.022>.
 22. Hu, S., Peng, L., Song, A., Ji, Y.-X., Cheng, J., Wang, M., and Chen, F.X. (2023). INTAC endonuclease and phosphatase modules differentially regulate transcription by RNA polymerase II. *Mol. Cell* 83, 1588–1604.e5. <https://doi.org/10.1016/j.molcel.2023.03.022>.
 23. Oegema, R., Baillat, D., Schot, R., van Unen, L.M., Brooks, A., Kia, S.K., Hoogeboom, A.J.M., Xia, Z., Li, W., Cesaroni, M., et al. (2017). Human mutations in integrator complex subunits link transcriptome integrity to brain development. *PLoS Genet.* 13, e1006809. <https://doi.org/10.1371/journal.pgen.1006809>.
 24. Tilley, F.C., Arrendel, C., Chhuon, C., Boisson, M., Cagnard, N., Parisot, M., Menara, G., Lefort, N., Guerrera, I.C., Bole-Feysot, C., et al. (2021). Disruption of pathways regulated by Integrator complex in Galloway-Mowat syndrome due to WDR73 mutations. *Sci. Rep.* 11, 5388. <https://doi.org/10.1038/s41598-021-84472-7>.
 25. Tepe, B., Macke, E.L., Niceta, M., Weisz Hubshman, M., Kanca, O., Schultz-Rogers, L., Zarate, Y.A., Schaefer, G.B., Granadillo De Luque, J.L., Wegner, D.J., et al. (2023). Bi-allelic variants in INTS11 are associated with a complex neurological disorder. *Am. J. Hum. Genet.* 110, 774–789. <https://doi.org/10.1016/j.ajhg.2023.03.012>.
 26. Drew, K., Wallingford, J.B., and Marcotte, E.M. (2021). hu.MAP 2.0: integration of over 15,000 proteomic experiments builds a global compendium of human multiprotein assemblies. *Mol. Syst. Biol.* 17, e10016. <https://doi.org/10.15252/msb.202010016>.
 27. Replogle, J.M., Saunders, R.A., Pogson, A.N., Hussmann, J.A., Lenail, A., Gunz, A., Mascibroda, L., Wagner, E.J., Adelman, K., Lithwick-Yanai, G., et al. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* 185, 2559–2575.e28. <https://doi.org/10.1016/j.cell.2022.05.013>.
 28. Pan, J., Kwon, J.J., Talamas, J.A., Borah, A.A., Vazquez, F., Boehm, J.S., Tsherniak, A., Zitnik, M., McFarland, J.M., and Hahn, W.C. (2022). Sparse dictionary learning recovers pleiotropy from human cell fitness screens. *Cell Syst.* 13, 286–303.e10. <https://doi.org/10.1016/j.cels.2021.12.005>.
 29. Offley, S.R., Pfleiderer, M.M., Zucco, A., Fraudeau, A., Welsh, S.A., Razew, M., Galej, W.P., and Gardini, A. (2023). A combinatorial approach to uncover an additional Integrator subunit. *Cell Rep.* 42, 112244. <https://doi.org/10.1016/j.celrep.2023.112244>.
 30. Azuma, N., Yokoi, T., Tanaka, T., Matsuzaka, E., Saida, Y., Nishina, S., Terao, M., Takada, S., Fukami, M., Okamura, K., et al. (2023). Integrator complex subunit 15 controls mRNA splicing and is critical for eye development. *Hum. Mol. Genet.* 32, 2032–2045. <https://doi.org/10.1093/hmg/ddad034>.
 31. Pfleiderer, M.M., and Galej, W.P. (2021). Emerging insights into the function and structure of the Integrator complex. *Transcription* 12, 251–265. <https://doi.org/10.1080/21541264.2022.2047583>.
 32. Sabath, K., and Jonas, S. (2022). Take a break: transcription regulation and RNA processing by the Integrator complex. *Curr. Opin. Struct. Biol.* 77, 102443. <https://doi.org/10.1016/j.sbi.2022.102443>.
 33. Welsh, S.A., and Gardini, A. (2023). Genomic regulation of transcription and RNA processing by the multitasking Integrator complex. *Nat. Rev. Mol. Cell Biol.* 24, 204–220. <https://doi.org/10.1038/s41580-022-00534-2>.
 34. Wagner, E.J., Tong, L., and Adelman, K. (2023). Integrator is a global promoter-proximal termination complex. *Mol. Cell* 83, 416–427. <https://doi.org/10.1016/j.molcel.2022.11.012>.
 35. Malovannaya, A., Lanz, R.B., Jung, S.Y., Bulynko, Y., Le, N.T., Chan, D.W., Ding, C., Shi, Y., Yucer, N., Krenciute, G., et al. (2011). Analysis of the human endogenous coregulator complexome. *Cell* 145, 787–799. <https://doi.org/10.1016/j.cell.2011.05.006>.
 36. van den Berg, D.L.C., Azzarelli, R., Oishi, K., Martynoga, B., Urbán, N., Dekkers, D.H.W., Demmers, J.A., and Guillemot, F. (2017). Nipbl interacts with Zfp609 and the integrator complex to regulate cortical neuron migration. *Neuron* 93, 348–361. <https://doi.org/10.1016/j.neuron.2016.11.047>.
 37. Wang, Z., Song, A., Xu, H., Hu, S., Tao, B., Peng, L., Wang, J., Li, J., Yu, J., Wang, L., et al. (2022). Coordinated regulation of RNA polymerase II pausing and elongation progression by PAF1. *Sci. Adv.* 8, eabm5504. <https://doi.org/10.1126/sciadv.abm5504>.
 38. Albrecht, T.R., Shevtsov, S.P., Wu, Y., Mascibroda, L.G., Peart, N.J., Huang, K.-L., Sawyer, I.A., Tong, L., Dundr, M., and Wagner, E.J. (2018). Integrator subunit 4 is a ‘Symplekin-like’ scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res.* 46, 4241–4255. <https://doi.org/10.1093/nar/gky100>.
 39. Pfleiderer, M.M., and Galej, W.P. (2021). Structure of the catalytic core of the Integrator complex. *Mol. Cell* 81, 1246–1259.e8. <https://doi.org/10.1016/j.molcel.2021.01.005>.
 40. Lin, M.-H., Jensen, M.K., Elrod, N.D., Huang, K.-L., Welle, K.A., Wagner, E.J., and Tong, L. (2022). Inositol hexakisphosphate is required for Integrator function. *Nat. Commun.* 13, 5742. <https://doi.org/10.1038/s41467-022-33506-3>.
 41. Barbieri, E., Trizzino, M., Welsh, S.A., Owens, T.A., Calabretta, B., Carroll, M., Sarma, K., and Gardini, A. (2018). Targeted enhancer activation by a subunit of the integrator complex. *Mol. Cell* 71, 103–116.e7. <https://doi.org/10.1016/j.molcel.2018.05.031>.
 42. Sabath, K., Stäubli, M.L., Marti, S., Leitner, A., Moes, M., and Jonas, S. (2020). INTS10–INTS13–INTS14 form a functional module of Integrator that binds nucleic acids and the cleavage module. *Nat. Commun.* 11, 3422. <https://doi.org/10.1038/s41467-020-17232-2>.
 43. Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Poi II–DSIF–NELF. *Nature* 560, 601–606. <https://doi.org/10.1038/s41586-018-0442-2>.
 44. Fianu, I., Chen, Y., Dienemann, C., Dybkov, O., Linden, A., Urlaub, H., and Cramer, P. (2021). Structural basis of Integrator-mediated transcription regulation. *Science* 374, 883–887. <https://doi.org/10.1126/science.abk0154>.
 45. Zheng, H., Jin, Q., Wang, X., Qi, Y., Liu, W., Ren, Y., Zhao, D., Xavier Chen, F., Cheng, J., Chen, X., et al. (2023). Structural basis of INTAC-regulated transcription. *Protein Cell* 14, 698–702. <https://doi.org/10.1093/procel/pwad010>.
 46. Song, G., Yang, Y., Liu, J.-H., Casasnovas, J.M., Shimaoka, M., Springer, T.A., and Wang, J.-H. (2005). An atomic resolution view of ICAM recognition in a complex between the binding domains of ICAM-3 and integrin alpha1beta2. *Proc. Natl. Acad. Sci. USA* 102, 3366–3371. <https://doi.org/10.1073/pnas.0500200102>.
 47. Mascibroda, L.G., Shboul, M., Elrod, N.D., Colleaux, L., Hamamy, H., Huang, K.-L., Peart, N., Singh, M.K., Lee, H., Merriman, B., et al. (2022). INTS13 variants causing a recessive developmental ciliopathy disrupt

- assembly of the Integrator complex. *Nat. Commun.* 13, 6054. <https://doi.org/10.1038/s41467-022-33547-8>.
48. Young, G., Hundt, N., Cole, D., Fineberg, A., Andrecka, J., Tyler, A., Olerinyova, A., Ansari, A., Marklund, E.G., Collier, M.P., et al. (2018). Quantitative mass imaging of single biological macromolecules. *Science* 360, 423–427. <https://doi.org/10.1126/science.aar5839>.
 49. Baillat, D., Russell, W.K., and Wagner, E.J. (2016). CRISPR-Cas9 mediated genetic engineering for the purification of the endogenous integrator complex from mammalian cells. *Protein Expr. Purif.* 128, 101–108. <https://doi.org/10.1016/j.pep.2016.08.011>.
 50. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
 51. Mirdita, M., Schütze, K., Moriawki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. <https://doi.org/10.1038/s41592-022-01488-1>.
 52. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The human transcription factors. *Cell* 175, 598–599. <https://doi.org/10.1016/j.cell.2018.09.045>.
 53. Sabath, K., Nabih, A., Arnold, C., Moussa, R., Domjan, D., Zaugg, J.B., and Jonas, S. (2024). Mechanistic basis of gene-specific transcription regulation by the Integrator complex. Preprint at bioRxiv. <https://doi.org/10.1101/2024.01.24.576984>.
 54. Fan, P., Shang, X.-Y., Song, A., Chen, S., Mao, R.-Y., Chen, J., Wang, Z., Zheng, H., Tao, B., Xu, W., et al. (2024). Catalytic-independent functions of INTAC in conferring sensitivity to BET inhibition. Preprint at bioRxiv. <https://doi.org/10.1101/2024.02.07.579305>.
 55. Houlard, M., Romero-Portillo, F., Germani, A., Depaux, A., Regnier-Ricard, F., Gisselbrecht, S., and Varin-Blank, N. (2005). Characterization of VIK-1: a new Vav-interacting Kruppel-like protein. *Oncogene* 24, 28–38. <https://doi.org/10.1038/sj.onc.1208043>.
 56. Shao, Z., Li, C., Wu, Q., Zhang, X., Dai, Y., Li, S., Liu, X., Zheng, X., Zhang, J., and Fan, H. (2022). ZNF655 accelerates progression of pancreatic cancer by promoting the binding of E2F1 and CDK1. *Oncogenesis* 11, 44. <https://doi.org/10.1038/s41389-022-00418-2>.
 57. Fianu, I., Ochmann, M., Walshe, J.L., Dybkov, O., Cruz, J.N., Urlaub, H., and Cramer, P. (2024). Structural basis of Integrator-dependent RNA polymerase II termination. *Nature* 629, 219–227. <https://doi.org/10.1038/s41586-024-07269-4>.
 58. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* 10, 252–263. <https://doi.org/10.1038/nrg2538>.
 59. Oksuz, O., Henninger, J.E., Warneford-Thomson, R., Zheng, M.M., Erb, H., Vancura, A., Overholt, K.J., Hawken, S.W., Banani, S.F., Lauman, R., et al. (2023). Transcription factors interact with RNA to regulate genes. *Mol. Cell* 83, 2449–2463.e13. <https://doi.org/10.1016/j.molcel.2023.06.012>.
 60. He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., and Bonasio, R. (2016). High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol. Cell* 64, 416–430. <https://doi.org/10.1016/j.molcel.2016.09.034>.
 61. Yang, C., Zheng, J., Liu, X., Xue, Y., He, Q., Dong, Y., Wang, D., Li, Z., Liu, L., Ma, J., et al. (2020). Role of ANKHD1/LINC00346/ZNF655 feedback loop in regulating the glioma angiogenesis via Staufen1-mediated mRNA decay. *Mol. Ther. Nucleic Acids* 20, 866–878. <https://doi.org/10.1016/j.omtn.2020.05.004>.
 62. Teng, Z., Yao, J., Zhu, L., Zhao, L., and Chen, G. (2021). ZNF655 is involved in development and progression of non-small-cell lung cancer. *Life Sci.* 280, 119727. <https://doi.org/10.1016/j.lfs.2021.119727>.
 63. Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* 152, 327–339. <https://doi.org/10.1016/j.cell.2012.12.009>.
 64. Hernandez, N., and Weiner, A.M. (1986). Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell* 47, 249–258. [https://doi.org/10.1016/0092-8674\(86\)90447-2](https://doi.org/10.1016/0092-8674(86)90447-2).
 65. Latos, P.A., Goncalves, A., Oxley, D., Mohammed, H., Turro, E., and Hemberger, M. (2015). Fgf and Esrrb integrate epigenetic and transcriptional networks that regulate self-renewal of trophoblast stem cells. *Nat. Commun.* 6, 7776. <https://doi.org/10.1038/ncomms8776>.
 66. Yue, J., Lai, F., Beckedorff, F., Zhang, A., Pastori, C., and Shiekhkhattar, R. (2017). Integrator orchestrates RAS/ERK1/2 signaling transcriptional programs. *Genes Dev.* 31, 1809–1820. <https://doi.org/10.1101/gad.301697.117>.
 67. Beniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem. Sci.* 37, 49–57. <https://doi.org/10.1016/j.tibs.2011.10.005>.
 68. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* 27, 14–25. <https://doi.org/10.1002/pro.3235>.
 69. Casañal, A., Lohkamp, B., and Emsley, P. (2020). Current developments in coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* 29, 1069–1078. <https://doi.org/10.1002/pro.3791>.
 70. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296. <https://doi.org/10.1038/nmeth.4169>.
 71. Weissmann, F., Petzold, G., VanderLinden, R., Huis In 't Veld, P.J., Brown, N.G., Lampert, F., Westermann, S., Stark, H., Schulman, B.A., and Peters, J.-M. (2016). biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proc. Natl. Acad. Sci. USA* 113, E2564–E2569. <https://doi.org/10.1073/pnas.1604935113>.
 72. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv. <https://doi.org/10.1101/2021.10.04.463034>.
 73. Mastronarde, D.N. (2003). SerialEM: A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* 9, 1182–1183. <https://doi.org/10.1017/S1431927603445911>.
 74. Weis, F., and Hagen, W.J.H. (2020). Combining high throughput and high quality for cryo-electron microscopy data collection. *Acta Crystallogr. D Struct. Biol.* 76, 724–728. <https://doi.org/10.1107/S2059798320008347>.
 75. Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J., and Berger, B. (2019). Topaz: A positive-unlabeled convolutional neural network CryoEM particle picker that can pick any size and shape particle. *Microsc. Microanal.* 25, 986–987. <https://doi.org/10.1017/S143192761900566X>.
 76. Scheres, S.H.W., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* 9, 853–854. <https://doi.org/10.1038/nmeth.2115>.
 77. Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J.M., Sorzano, C.O.S., and Vargas, J. (2021). DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* 4, 874. <https://doi.org/10.1038/s42003-021-02399-1>.
 78. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* 74, 531–544. <https://doi.org/10.1107/S2059798318006551>.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--|---------------------------------|-------------------------------|
| Antibodies | | |
| Mouse anti-FLAG (HRP conjugate) | Sigma | Cat#A8592; RRID:AB_439702 |
| Mouse anti-HA (HRP conjugate) | Santa Cruz | Cat#SC7392HRP; RRID:AB_627809 |
| Bacterial and virus strains | | |
| NEB10 | NEB | Cat# C3020K |
| DH10EMBacY | Bieniossek et al. ⁶⁷ | N/A |
| Chemicals, peptides, and recombinant proteins | | |
| L-Glutamine | GIBCO | Cat#A29168-01 |
| Penicillin-Streptomycin (10,000 U/mL) | GIBCO | Cat#15140-122 |
| DMEM | GIBCO | Cat#31966-021 |
| ExpressFive™ SFM | GIBCO | Cat#10486025 |
| SF-900™ II SFM | GIBCO | Cat#11497013 |
| Opti-MEM | GIBCO | Cat#31985-062 |
| Trypsin-EDTA | GIBCO | Cat#25200-056 |
| FBS | GIBCO | Cat#10270-106 |
| PEI 40k | Polysciences | Cat#24765-1 |
| Tween20 | Sigma | Cat#P1379-500ML |
| BS3 (bis(sulfosuccinimidyl)suberate) | ThermoScientific | A39266 |
| Triton X-100 | ICN Biomedicals | Cat#807426 |
| Pierce™ ECL Western Blotting Substrate | Thermo Fisher | Cat#32209 |
| Immobilon PVDF Membrane | Merck | Cat#PVH000010 |
| Anti-FLAG Resin | Sigma Aldrich | Cat#F2426-1ML |
| Anti-HA Resin | Sigma Aldrich | Cat#A2905-1ML |
| High Capacity Streptavidin Agarose Resin | Thermo Fisher | Cat#20361 |
| His60 Ni Superflow Resin | Takara | Cat#635660 |
| Amylose Resin | NEB | Cat#E8021S |
| Experimental models: Cell lines | | |
| HEK293T | ATCC | Cat#CRL-3216 |
| SF21 | GIBCO | Cat#11497013 |
| <i>Trichoplusia ni</i> High Five Cells | Invitrogen | Cat#B85502 |
| Recombinant DNA | | |
| pBIG1a-8xHis-INTS10/INTS13/ INTS14/SBP-INTS15 | this study | N/A |
| pBIG1a-8xHis-INTS5/SBP-INTS8 | this study | N/A |
| pBIG1c-SBP-INTS10/8xHis-INTS15 | this study | N/A |
| pcDNA3.1(+)FLAG-ZNF655 | this study | N/A |
| pcDNA3.1(+)FLAG-ZNF655 ^{R95A/L96A} | this study | N/A |
| pcDNA3.1(+)FLAG-ZNF655 ^{R95A/L96A} | this study | N/A |
| pcDNA3.1(+)FLAG-ZNF655 ^{F106A/L107A} | this study | N/A |
| pcDNA3.1(+)FLAG-ZNF655 ^{R152A/I153/A} | this study | N/A |
| pcDNA3.1(+)FLAG-ZNF655 ^{I102A/L103A} | this study | N/A |
| pcDNA3.1(+)FLAG-ZNF655 ^{R111A/F113A/R114A} | this study | N/A |
| pMG-3xFLAG-INTS14 | this study | N/A |
| pMG-3xHA-INTS13 | this study | N/A |

(Continued on next page)

Continued

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|------------------------------|---|
| pMG-3xHA-INTS15 | this study | N/A |
| pMG-3xHA-INTS15 ^{M120A/V124A} | this study | N/A |
| pMG-3xHA-INTS15 ^{L69A/L72A} | this study | N/A |
| Deposited data | | |
| cryoEM map of the INTS10/13/14 | this study | EMDB: EMD-19851 |
| cryoEM map of the INTS10/13/14/15 (state 2) | this study | EMDB: EMD-19871 |
| cryoEM map of the INTS10/13/14/15 (state 1) | this study | EMDB: EMD-50267 |
| cryoEM map of the INTS10/13/14/15 (state 3) | this study | EMDB: EMD-50268 |
| cryoEM map of the INTS5/8/15 | this study | EMDB: EMD-19872 |
| cryoEM map of the INTS5/8/10/15 | this study | EMDB: EMD-19853 |
| Atomic coordinates of the INTS10/13/14 | this study | PDB: 9EOC |
| Atomic coordinates of the INTS10/13/14/15 (state 2) | this study | PDB: 9EP1 |
| Atomic coordinates of the INTS10/13/14/15 (state 1) | this study | PDB: 9FA4 |
| Atomic coordinates of the INTS10/13/14/15 (state 3) | this study | PDB: 9FA7 |
| Atomic coordinates of the INTS5/8/15 | this study | PDB: 9EP4 |
| Atomic coordinates of the INTS5/8/10/15 | this study | PDB: 9EOF |
| AlphaFold2 interaction screen results (atomic coordinates, PAE and pLDDT plots) | this study | https://doi.org/10.17632/nkz5ypf3b2.1 |
| Software and algorithms | | |
| PyMol v2.3.4 | Schrodinger, Inc. | https://pymol.org/2/ |
| UCSF ChimeraX v1.7.1 | Goddard et al. ⁶⁸ | https://www.cgl.ucsf.edu/chimerax/ |
| Coot v0.9.8.7 | Casañal et al. ⁶⁹ | https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/ |
| cryoSPARC v4.4.1 | Punjani et al. ⁷⁰ | https://cryosparc.com/ |
| ColabFold (local) v1.3.0 and v1.5.2 | Mirdita et al. ⁵¹ | https://github.com/sokrypton/ColabFold |
| AlphaFold2, multimer v2.3 | Jumper et al. ⁵⁰ | N/A |
| Inkscape v0.92 | Inkscape project | https://inkscape.org/ |
| Refeyn Acquire and Discover MP | Refeyn | https://refeyn.com/ |
| Other | | |
| UltrAuFoil R1.2/1.3 300 mesh | Quantifoil | Cat#Q350AR13A |
| Quantifoil Au 300 mesh R1.2/1.3 | Quantifoil | Cat#N1-C14nAu30-01 |
| CF300-Cu | Electron Microscopy Sciences | Cat#CF300-CU-50 |
| Vitrobot Mark IV | ThermoScientific | N/A |
| ChemiDoc MP | Bio-Rad | N/A |
| Vibra Cell VCX750 | Sonics | https://www.sonics.com/liquid-processing/products/vibra-cell-processors/vcx-500-vcx-750/ |

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Wojciech P. Galej (wgalej@embl.fr).

Materials availability

Unique and stable reagents generated in this study are available upon request.

Data and code availability

- CryoEM maps obtained within this study have been deposited in the EMDB database with the following accession codes: EMDB: EMD-19851, EMD-19871, EMD-19872, EMD-19853, EMD-50267, EMD-50268. The atomic coordinates have been

deposited in PDBe with the following accession codes: PDB:9EOC, 9EP1, 9EP4, 9EOF, 9FA4, and 9FA7. AlphaFold2 structure prediction data including atomic coordinates, PEA and pLDDT plots is available through Mendeley Data: <https://doi.org/10.17632/hkz5ypf3b2.1>

- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

HEK293T cells (ATCC) were propagated in DMEM medium (GIBCO) supplemented with 10% FBS (Thermo Fisher) and Pen Strep (GIBCO). SF21 Insect cells were cultured in SF-900™ II SFM media (GIBCO). Hi5 cells were grown in Express Five™ SFM media (GIBCO) supplemented with L-Glutamine (GIBCO).

METHOD DETAILS

Protein expression and purification

Full-length ORFs of INTS10 (with an N-terminal 8xHis-tag), INTS13, INTS14 and INTS15 (with an N-terminal SBP-tag) were cloned into a single vector using the biGBac system.⁷¹ Using the same system, 8xHis-INTS5 and SBP-INTS8, as well as 8xHis-INTS15 and SBP-INTS10, were cloned into two separate vectors. Baculovirus was generated in SF21 cells, cultured in SF-900™ II SFM media.⁶⁷ For the high-yield protein expression, Hi5 cells were used, cultured in Express Five™ SFM media and infected at a density of 1×10^6 cells/ml with 1% volume of the baculovirus stock harvested from the SF21 culture. Cells were incubated at 27°C for 72 hours and harvested at 300 g for 10 minutes at 4°C (JLA 8.1000 rotor). The supernatant was discarded and the cell pellet was resuspended in 10 volumes of 20 mM HEPES KOH pH 7.8, 150 mM KCl, 30 mM imidazole, 10% (v/v) glycerol, flash frozen in liquid nitrogen and stored in -80°C prior to purification. Frozen pellets were thawed and sonicated in an ice bath for 3 minutes (10/30 sec ON/OFF cycles). The cell lysate was cleared by ultracentrifugation at 185,700 $\times g$ for 40 minutes (Beckman 45 Ti rotor) and filtered through a 0.8 µm syringe filter. For the two-step purification, the cleared lysate was first incubated with 10% (v/v) Ni-NTA resin (QIAGEN) at 4°C for 2 hours on a turning wheel. Next, the resin was transferred to a Poly-Prep Chromatography Column (Bio-Rad), washed extensively with 20 mM K-HEPES pH 7.8, 150 mM KCl, 30 mM Imidazole and eluted in 1 ml fractions with 20 mM K-HEPES pH 7.8, 150 mM KCl, 300 mM Imidazole. 5 ml of eluted fractions were pooled and incubated with 400 µl Streptavidin agarose resin (Thermo Scientific) and incubated at 4°C for 2 hours on a turning wheel. The resin was washed in 20 mM K-HEPES pH 7.8, 150 mM KCl and eluted in 200 µl fractions with 20 mM K-HEPES pH 7.8, 150 mM KCl, 25 mM desthiobiotin. For the single-step purification, the cleared lysate was transferred directly to 1 ml Streptavidin agarose resin and proceeded as described above. Sample quality was assessed by SDS-PAGE at each stage of the purification process.

Mass photometry analysis

The mass photometry experiment was performed using the Refeyn OneMP instrument. INTS5/8 and INTS10/15 were measured at 100 nM. Contrast-to-mass conversion was done by calibrating the instrument with a native protein ladder, following manufacturer's guidelines.

AlphaFold2 interaction screen

The database of the transcription factors⁵² was curated to include only proteins with sequences shorter than 1500AA. The resulting set of 1566 TFs was paired with INTS13, and binary structures were predicted with the ColabFold⁵¹ implementation of the AlphaFold2 algorithm.⁵⁰ Multimer v2.1 (ColabFold: alphafold2_multimer_v2) and Multimer v2.3 (ColabFold: alphafold2_multimer_v3) were used with 3 recycling steps and a single model for each protein.⁷² The results of the Multimer v2.1 screen were inspected manually and the outcome of interaction prediction was considered positive when the Predicted Aligned Error (PAE) plots showed inter-molecular cross-peaks with values better than 5 Å. For the Multimer v2.3, pTM vs ipTM plot was used to select candidates with the top 25 ipTM scores. These candidates were aligned against common reference (INTS13) and all interfaces were inspected manually.

Pull-down analysis and Western blotting

HEK293T cells were seeded into 6-well plates 24h before transfection at a density of 500 000 cells per well in 1.5 ml D-MEM medium with 10% FBS. 1 µg of total DNA (1:1 3xFLAG-ZNF655 and 3xHA-INTS13) in 50 µl of opti-MEM was mixed with 3 µg of PEI MAX 40K in 50 µl of opti-MEM and incubated at room temperature for 10 min. Transfection solutions were added drop-by-drop to each well. The cells were harvested by centrifugation 48h after transfection, dissolved in 400 µl of lysis buffer (150 mM KCl, 20 mM K-HEPES pH 7.8 and 0.1% Triton X-100) and sonicated for 10 s at 30% amplitude. Lysates were cleared by centrifugation in a table-top centrifuge at 20 000 g at 4°C for 30 min. The supernatant was incubated for 2 h with 5% (v/v) of HA-agarose to capture the bait protein. Affinity resin was washed 3 times with 10 resin volumes of Buffer3 (150 mM KCl and 20 mM K-HEPES pH 7.8) and subsequently resuspended in SDS sample buffer and heated up to 95°C for 5 min to release bound proteins. Input and elution fractions were analyzed by western blotting.

A PVDF membrane (Merck) was activated for 5 minutes in 100% EtOH and incubated for 5 minutes in the transfer buffer (1xTris-Glycine, 20%EtOH). A wet transfer was performed for 60-90min at 30V in an Invitrogen XCell II Blot Module. The membrane was blocked with 5% milk in PBS supplemented with 0.2% Tween 20 (PBST) for 1h at room temperature. Primary antibodies were added in the following dilutions: anti-HA 1:5000 ([HA-7] HRP ab49969, abcam); anti-FLAG – 1:5000 (AntiFLAG M2-Peroxidase (HRP) sigma A8592-.2MG). The membrane was washed 3 times for 5 min with 20 ml of PBST, and chemiluminescence was detected with an HRP substrate kit (Pierce ECL Western Blotting Substrate) in a Chemidoc imager (Bio-Rad).

For the differential quantitative proteomics experiments of ZNF655 and INTS15 variants, cells were grown in 15 cm dishes in triplicates and proteins were purified using EZview Red anti-FLAG M2 Affinity Gel (Sigma) or anti-HA resin (Sigma), respectively. Label-free TMT-10plex mass spectrometry experiments were performed as previously described.³⁹

Negative stain electron microscopy

Negative stain electron microscopy was used to check the quality of the samples prior to the preparation of cryoEM grids. The purified samples were crosslinked with 1mM BS3 (bis(sulfosuccinimidyl)suberate; Thermo Fisher Scientific) for 30 minutes on ice and quenched by the addition of Tris-HCl pH 8.0 to the final concentration of 10 mM. Negative stain grids, CF300-Cu (Electron Microscopy Sciences), were glow-discharged for 30 s at 25 mA at 0.3 bar (Pelco EasiGlow). 3 µl of the native or cross-linked sample was applied, incubated for 60 s, blotted away, washed once in a drop of 1.5% (w/v) uranyl acetate solution and incubated for 30 s in a second drop of uranyl acetate. After blotting away the liquid and drying, the grid was imaged on a Tecnai G2 Spirit BT microscope (Thermo Fisher Scientific) operated at 120 kV.

CryoEM sample preparation

EM grids were glow discharged for 40 s at 25 mA at 0.3 bar (Pelco EasiGlow), blotted and plunge-frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific) at 4°C, 100% humidity. For INTS10/13/14/15 complex (Arm module), 3 µl of BS3 cross-linked sample at 1.1 mg/ml concentration was applied to Quantifoil Au 300 mesh R1.2/1.3 grids (Quantifoil Micro Tools) before blotting for 2 s at -5 force. Purified INTS5/8 and INTS10/15 sub-complexes were mixed together in a 1:1 molar ratio, and 3 µl sample at 1.34 mg/ml concentration was applied to UltraAuFoil Au 300 mesh R1.2/1.3 grids (Quantifoil Micro Tools) before blotting for 2 s at 0 force.

CryoEM data collection and processing

INTS10/13/14/15 and INTS5/8/10/15 high-resolution data were collected on a Titan Krios TEM (Thermo Fisher Scientific) operated at 300 kV, equipped with a Gatan Quantum-K3 camera, using SerialEM⁷³ at EMBL Heidelberg.⁷⁴ 105 000 x magnification was used, corresponding to 0.82 Å/pixel. Movies were fractionated into 40 frames recorded with a total dose of 40 e⁻/Å². A total of 9 253 movies with defocus range from -0.9 to -2.2 µm for INTS10/13/14/15 and 3 682 movies with defocus range from -0.8 to -2.5 µm for INTS5/8/10/15 complex were acquired. Additionally, a total of 8 425 movies of INTS5/8/10/15 complex was collected using a Glacios microscope equipped with a Falcon 4i camera and SelectrisX energy filter, at 130 000 x magnification (0.86 Å/pixel) with a 40 e⁻/Å² total dose and fractionated into 40 frames. All image processing was performed using cryoSPARC⁷⁰ unless otherwise stated. Raw movies were motion-corrected using CryoSPARC Patch Motion Correction, followed by Patch CTF Estimation. Particles were picked using a cryoSPARC Blob picker. For the INTS10/13/14/15 complex, an initial subset of 1 550 micrographs was processed, and 360 000 particles were extracted with an original 600-pixel box size, binned to 256 pixels (1.9 Å/pixel). 2D classification was performed, and 7 different classes, corresponding to 40 000 particles, were selected as a template to improve particle picking using Topaz,⁷⁵ a neural network algorithm that was trained on a subset of 300 micrographs using the ResNet8 model architecture. Next, we used the trained model to pick particles from a total of 9 253 micrographs, which yielded 1 282 952 particles, extracted as described above. After 2D classification, 258 952 particles were selected for *ab initio* reconstruction, followed by homogenous and non-uniform refinement, yielding a 3.92 Å resolution map. Next, using local motion correction, the particles were unbinned to the original box size. After another round of non-uniform and global CTF refinement, the final resolution reached 3.34 Å, based on Gold Standard FSC criteria.⁷⁶ To improve the density map of the peripheral region particle subtraction was done with a mask covering the well-resolved regions of INTS10/13/14, followed by 3D classification which yielded 3 distinct classes of INTS15 and part of INTS10. Next, reconstitution to full-size particles and refinement yielded maps with improved densities of the peripheral regions where AF2 predicted model of INTS10/15 could be rigid-body fitted.

For the INTS5/8/10/15 complex, 734 626 particles were extracted, and box size was binned as described for the INTS10/13/14/15 complex. After 2D classification, 208 137 particles were selected for *Ab-initio* reconstruction, yielding 3 classes. One class that resembled the INTS5/8 model from the available INTAC structure (PDB ID: 7cun) was selected for homogenous refinement, resulting in a 3.94 Å resolution map. After local motion correction and unbining to the original box size, global CTF and non-uniform refinement were done, yielding a final 3.18 Å resolution map. However, the map is anisotropic and high-resolution features are focused around the core as indicated by local resolution estimation.

Several data sets for INTS5/8/10/15 were also collected on a Glacios microscope, in total 8 325 micrographs were merged and processed together. The initial set of 2 594 micrographs was processed, where we identified low-resolution 2D classes that contained features otherwise missing from the core INTS5/8/10/15 structure processed from the Krios dataset described above. Selected 2D classes were used as templates for particle picking using Topaz,⁷⁵ which was trained on a subset of 300 micrographs

using ResNet8 model architecture. A total of 638 977 particles were extracted and binned (2.0 Å/pixel). After 2D classification, 194 293 particles were selected for *ab initio* reconstruction. 3 classes were obtained and heterogeneously refined. 1 class (8.26 Å map) with features corresponding to the missing region of INTS5/8/10/15 map described before, was further refined using 435 135 particles that were Topaz-picked, extracted and 2D classified from all of the acquired 8 325 micrographs. Using 3D classification we generated 5 classes that were heterogeneously refined, which yielded 1 class that was enriched with particles containing low-resolution features that were missing. Non-uniform refinement resulted in a map of 5.94 Å resolution with improved particle distribution.

Model building

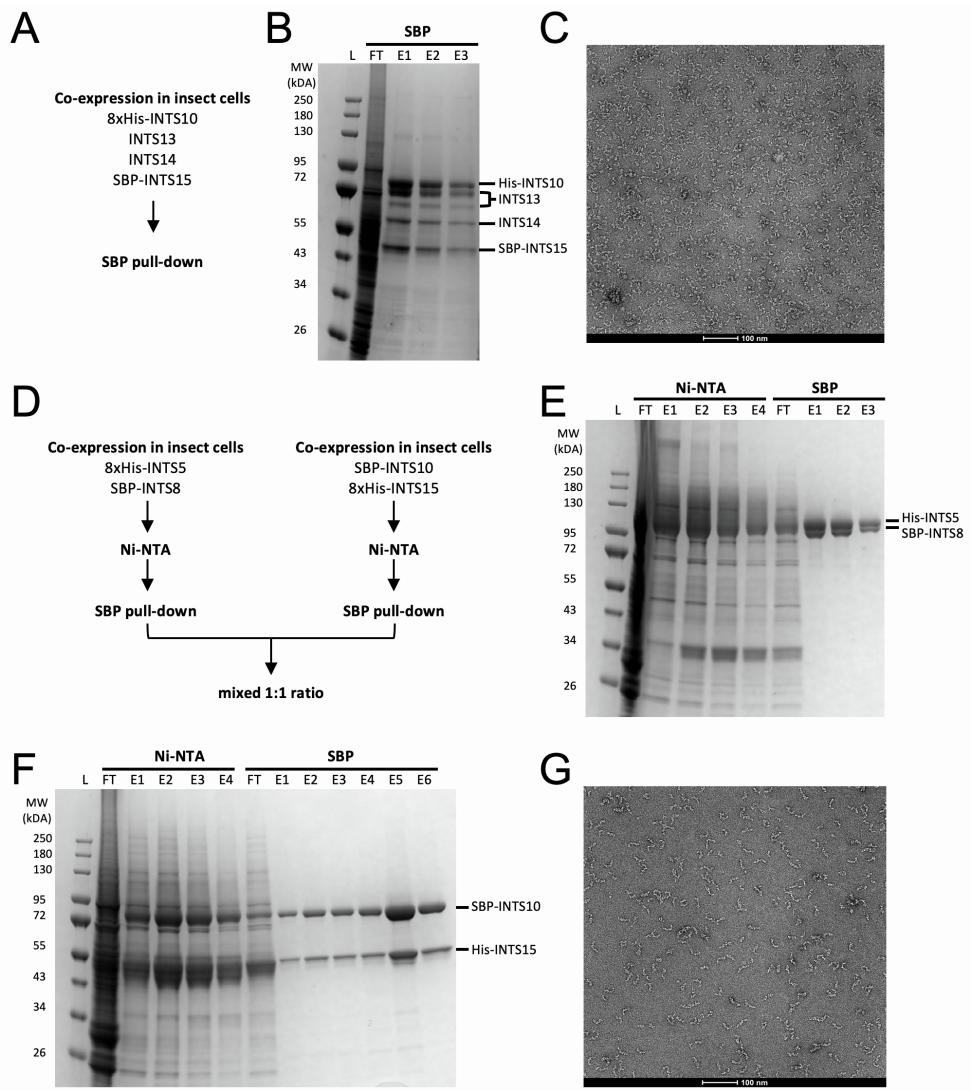
For the INTS10/13/14/15 complex INTS13/14 crystal structure (PDB ID: 6SN1) and INTS10 AlphaFold2 model (Uniprot ID: Q9NVR2) were rigid-body fitted into the map in ChimeraX⁶⁸ and used as templates for manual curation and rebuilding in Coot.⁶⁹ The deepEMhancer-generated map was used in model building of poorly resolved regions.⁷⁷ The final 3D reconstruction of INTS10/13/14/15 complex lacks high-resolution densities of the N-terminal part of INTS10 (residues 1–123) along with INTS15 (Figures 1A and 1B) thus these regions were modelled based on their AF2 predicted structures. The final model was refined using phenix.refine⁷⁸ and validated using the wwPDB OneDep Validation System.

For the modelling of INTS5/8/10/15, the available INTS5/8 experimental models (PDB: 7CUN, 7YCX, 7PKS) are very similar to one another and could be unambiguously fitted in our map. However, the N-terminal part of INTS5 is missing in these structures. Low-resolution tubular densities in our map were interpreted with AF2 and correspond to residues 28–253 of INTS5. This helical region forms an interface with INTS15. INTS10 and INTS15 models obtained from the Arm module were rigid-body fitted into the low-resolution region of the map (Figures 2A and 2B).

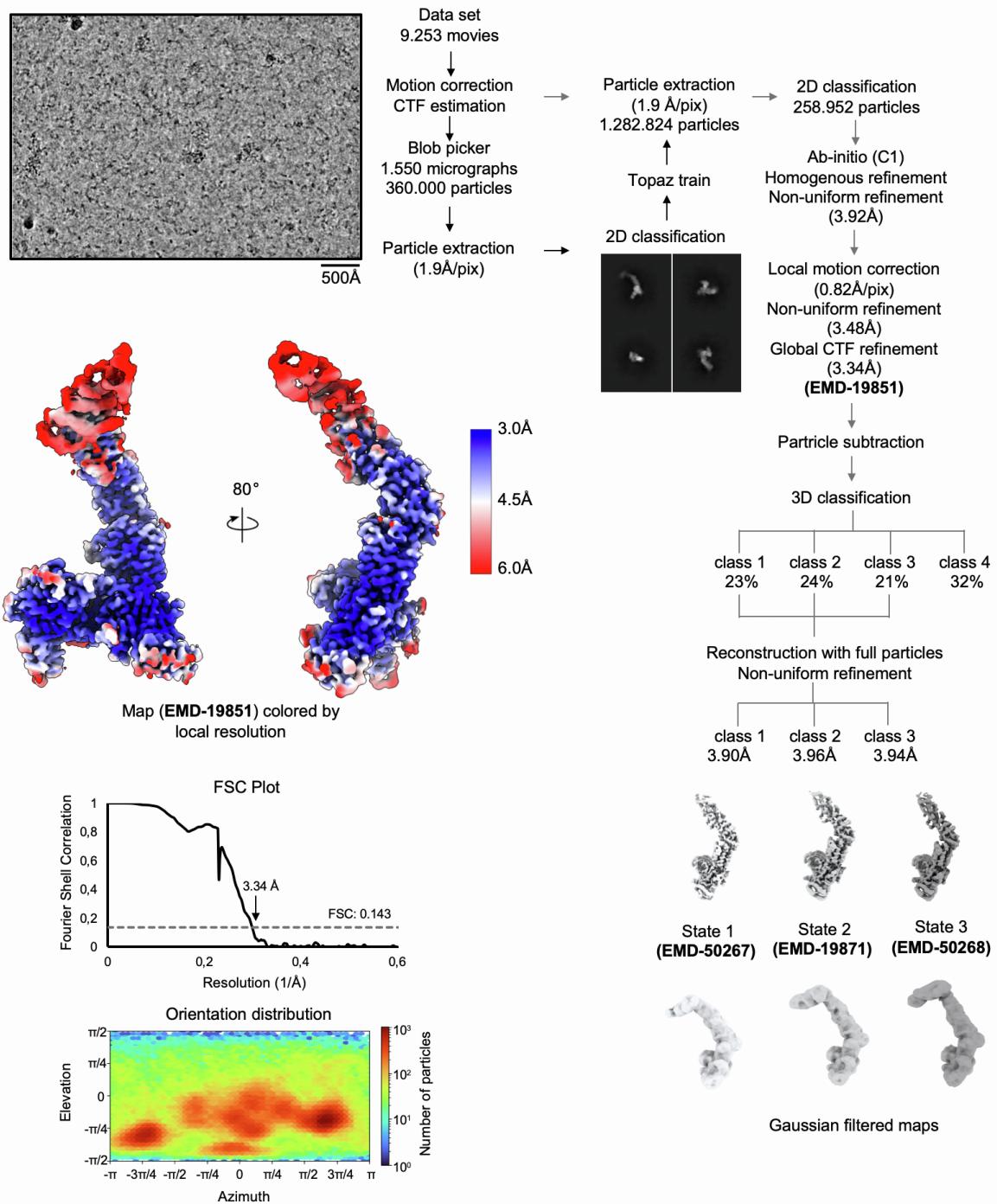
Supplemental information

**Structural basis of the Integrator complex assembly
and association with transcription factors**

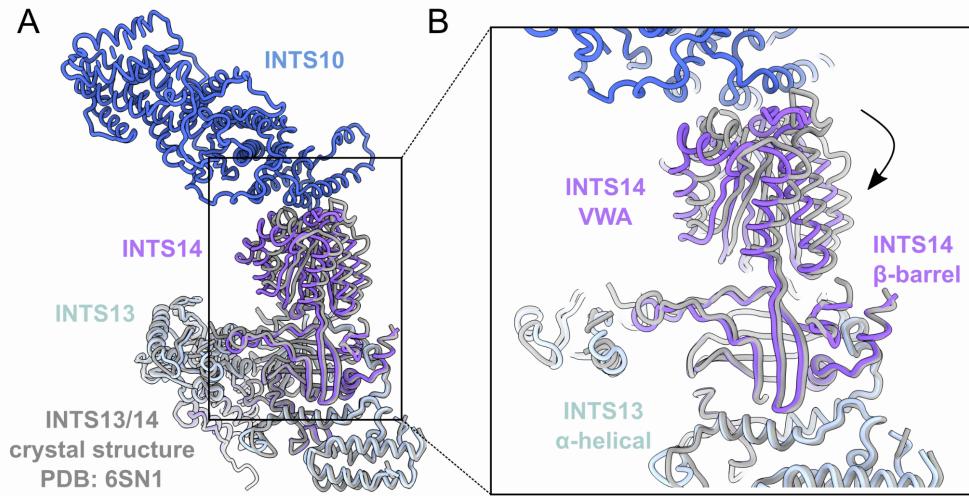
Michał Razew, Angelique Fraudeau, Moritz M. Pfleiderer, Romain Linares, and Wojciech P. Galej



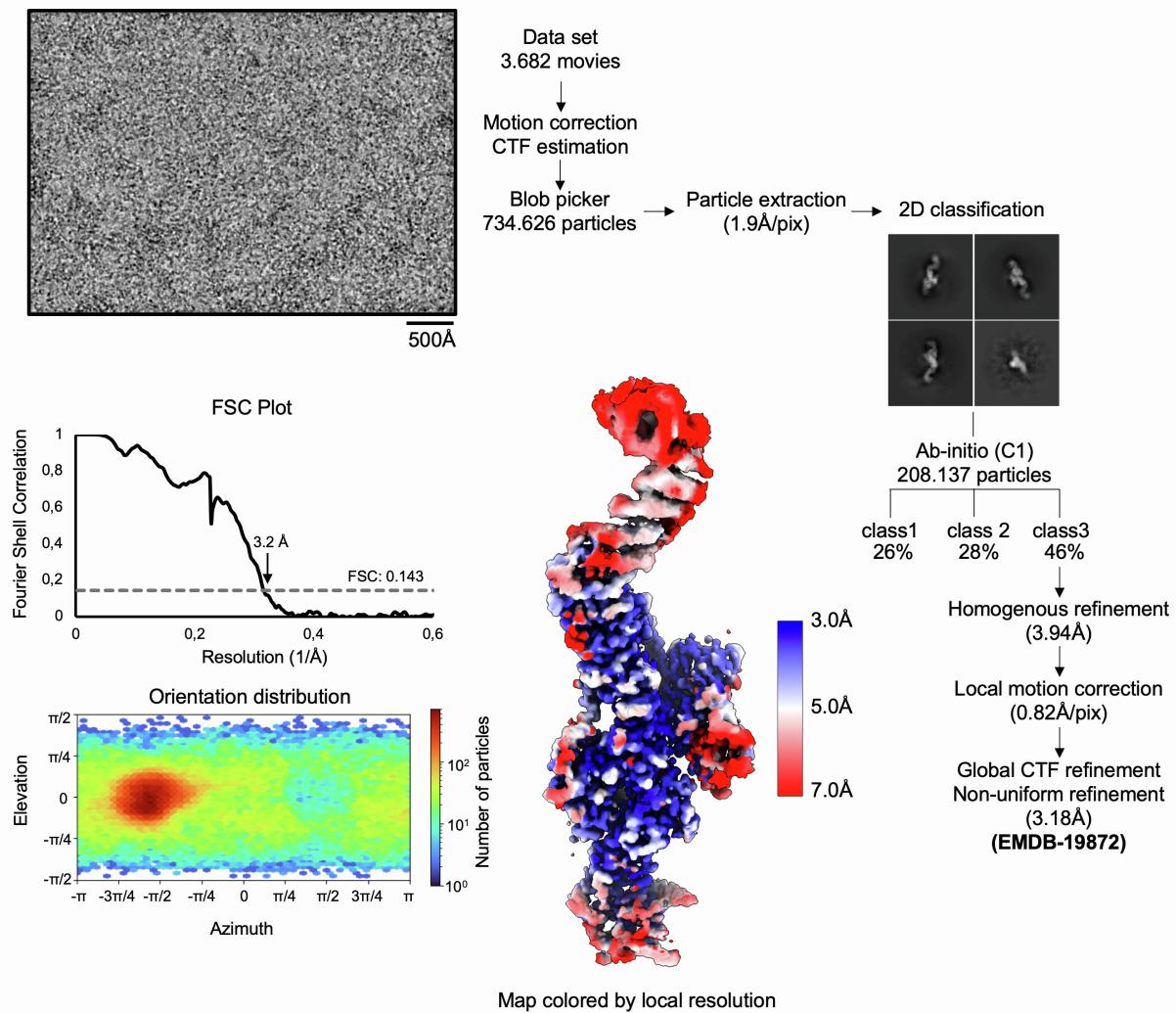
Supplementary Figure 1. INTS10/13/14/15 and INTS5/8/10/15 complex sample preparation. Related to Figures 1 and 2. **A)** Expression and purification strategy for INTS10/13/14/15 complex. **B)** SDS-PAGE image of the single-step purification of recombinant INTS10/13/14/15 complex on Streptavidin-agarose resin. **C)** Negative stain electron microscopy image of a BS3 cross-linked sample shown in B. **D)** Expression and purification strategy for INTS5/8/10/15 complex. **E)** SDS-PAGE image of the two-step purification of recombinant INTS5/8 sub-complex. **F)** SDS-PAGE image of the two-step purification of recombinant INTS10/15 sub-complex. **G)** Negative stain electron microscopy image of an INTS5/8 and INTS10/15 mixed in 1:1 ratio and BS3 cross-linked.



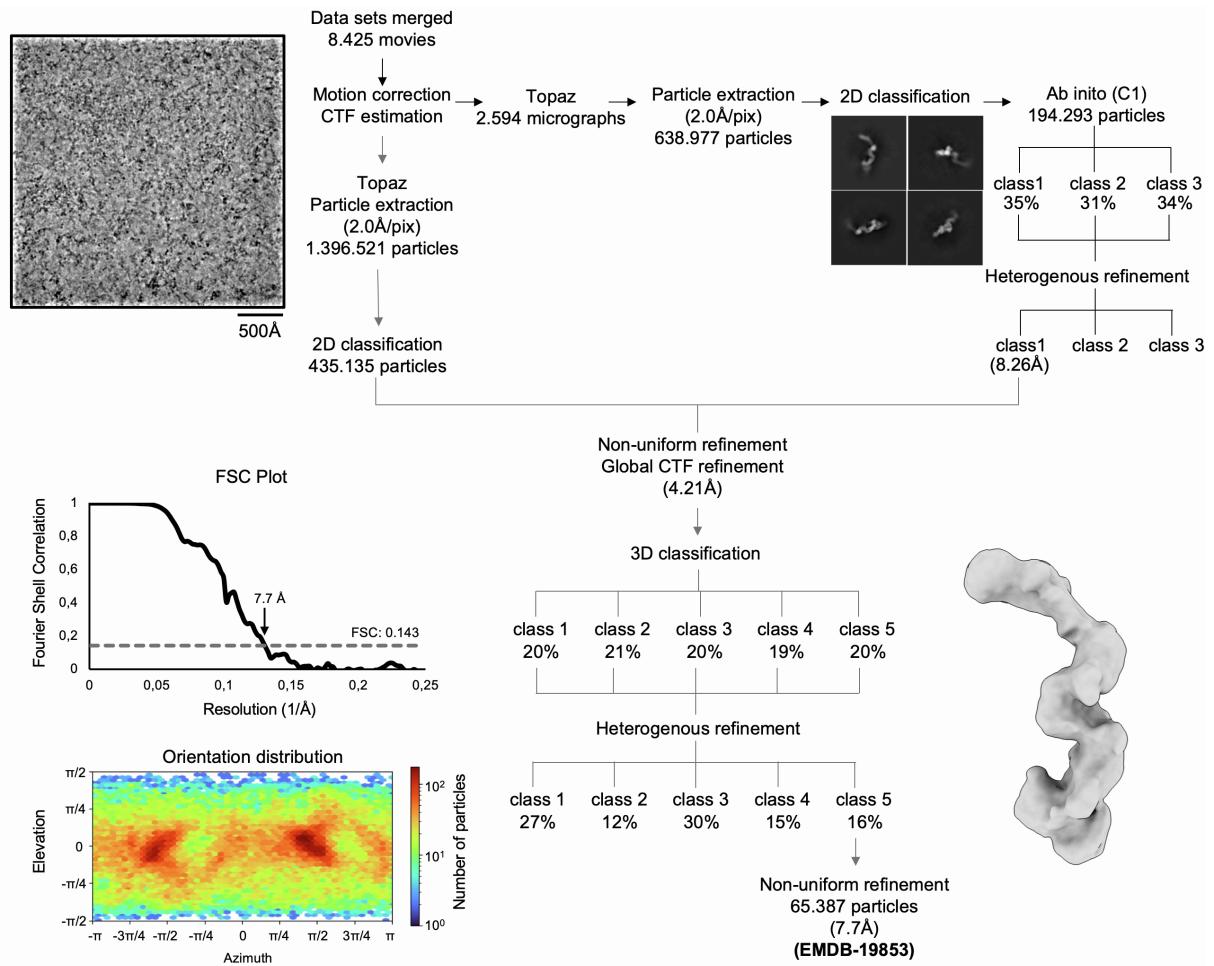
Supplementary Figure 2. Workflow of the cryoEM data processing of INTS10/13/14/15 complex.
Related to Figure 1. Data processing was described in detail in the Method section.



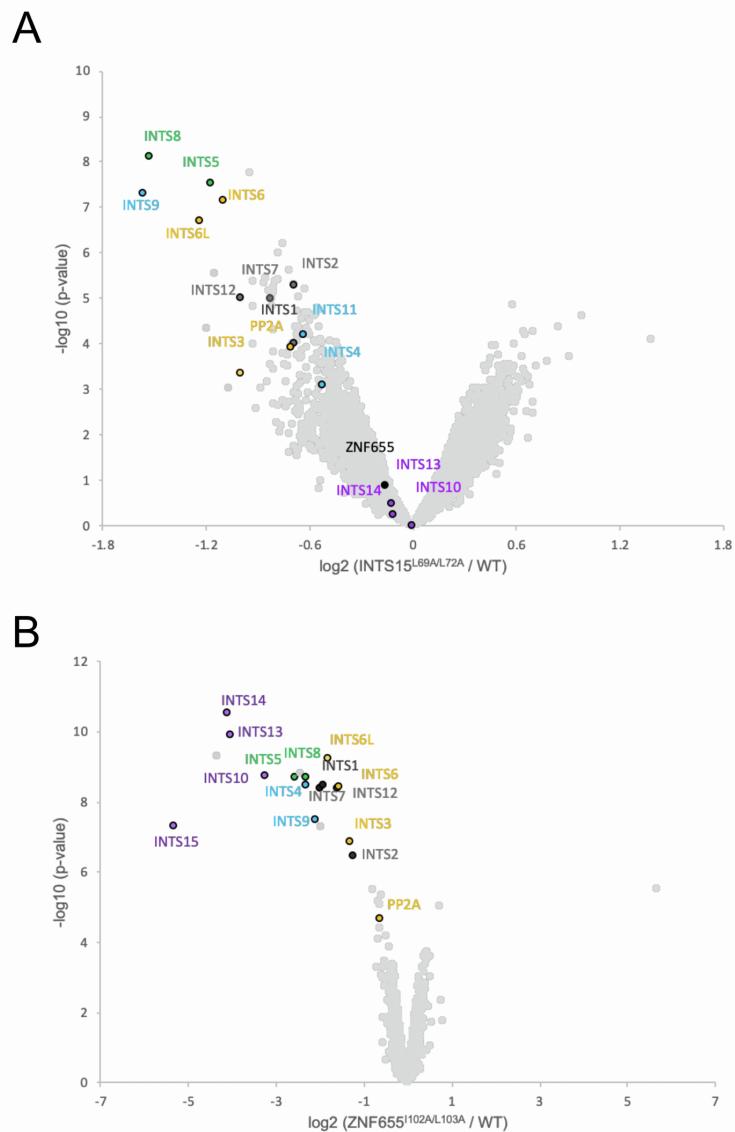
Supplementary Figure 3. Comparison of the cryoEM structure of the Integrator Arm module with the crystal structure of INTS13/14. Related to Figure 1. **A)** Superposition of INTS13 subunits of the cryoEM structure of the Arm module (coloured) and INTS13/14 crystal structure (grey). All domains of INTS13 and INTS14 in both structures are in good agreement with an overall RMSD (root mean square deviation) value 1.3 Å across 512 residues. In the cryoEM structure, the VWA domain of INTS14 is shifted by 4.2 Å (indicated with an arrow) due to its interaction with INTS10.



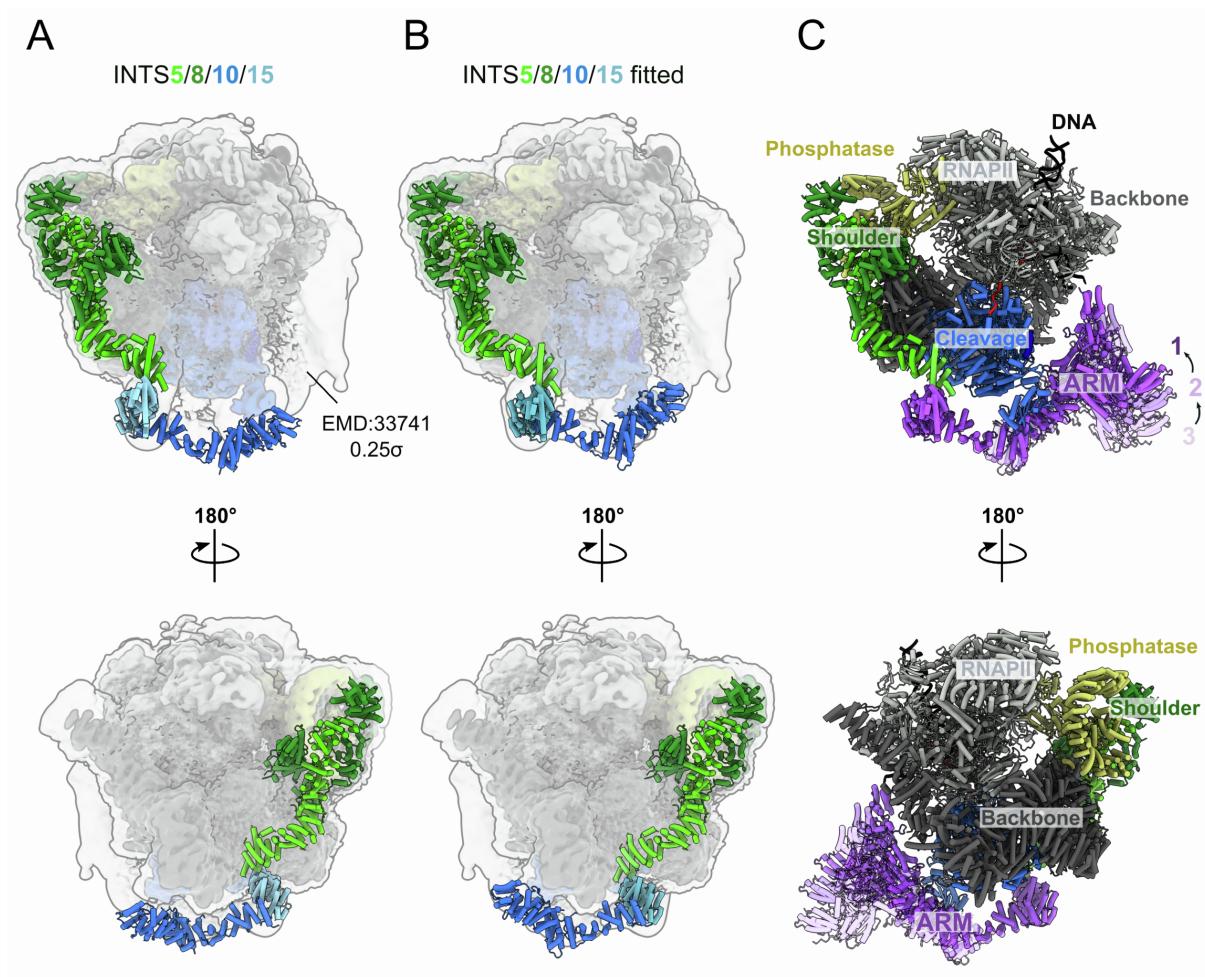
Supplementary Figure 4. Workflow of the high-resolution cryoEM data processing of INTS5/8/10/15 complex. Related to Figure 2. Data processing was described in detail in the Method section.



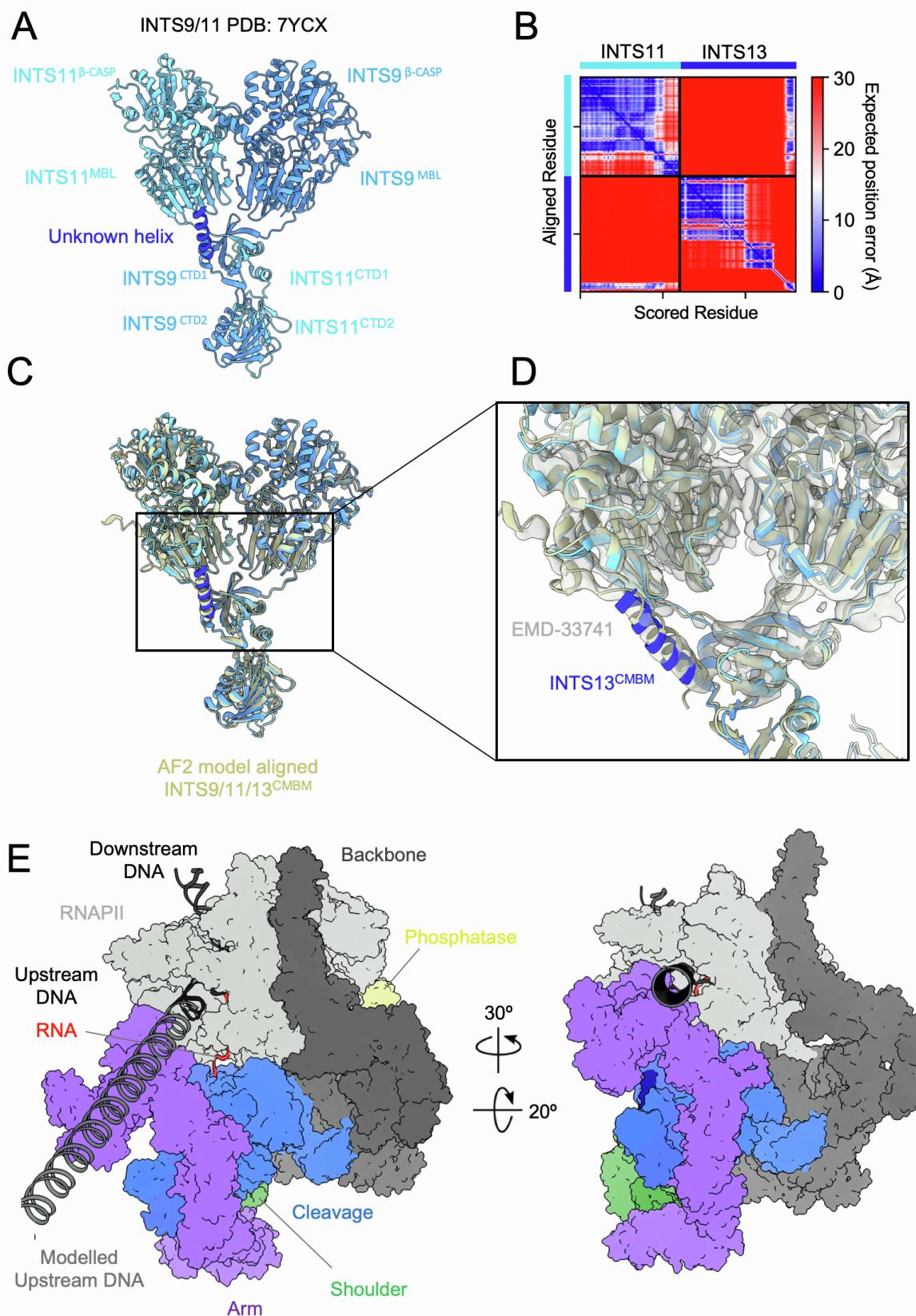
Supplementary Figure 5. Workflow of the low-resolution cryoEM data processing of INTS5/8/10/15 complex. Related to Figure 2. Data processing was described in detail in the Method section.



Supplementary Figure 6. Differential quantitative mass spectrometry analysis. Related to Figures 2 and 4. **A)** proteins co-purifying with 3xHA-INTS15 (WT or L69A/L72A) and **B)** 3xFLAG-ZNF655 (WT or I102A/I103A) transiently expressed in HEK293T cells and purified on anti-FLAG affinity resin.

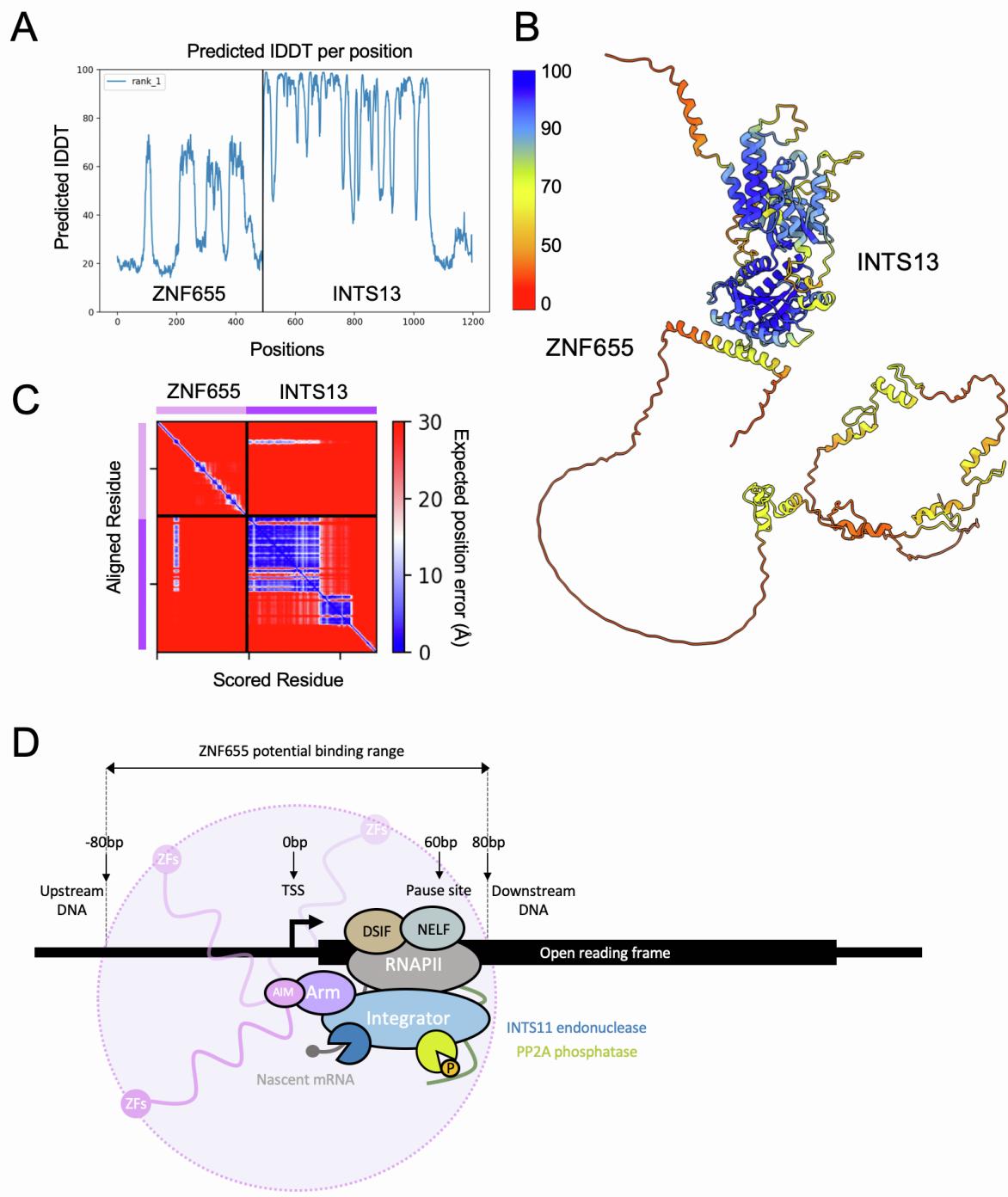


Supplementary Figure 7. Analysis of the low-resolution densities of EMD:33741 of the Integrator-RNAPII complex. Related to Figure 3. **A)** Integrator-RNAPII complex cryoEM density map contoured at 3.2σ fitted in a low-contour overall density map contoured at 0.25σ with visible unassigned regions. The structure of the Integrator INTS5/8/10/15 is fitted to the density map corresponding to the position of INTS5/8 subunits. **B)** N-terminal region of INTS5, INTS15 and INTS10 from A fitted to the unassigned density map show a conformation change upon binding to the Integrator-RNAPII complex. **C)** The atomic model of the fully assembled Integrator-RNAPII with the same orientation as in B, with the Arm module is represented in 3 different states aligned based on INTS15 position in B, with state 1 having the best fit to the experimental density map.



Supplementary Figure 8. INTS13 Cleavage Module Binding Motif interaction with INTS9/11 and modelling of the upstream DNA in the context of the full Integrator complex. Related to Figures 3 and 4. **A)** Structure of INTS9/11 of the Integrator cleavage module (PDB: 7YCX). The unassigned helix between INTS11^{MBL} and INTS9^{CTD1} domains is shown in dark blue. **B)** PAE (Predicted Alignment Error)

plot of the AlphaFold2 multimer prediction of the binary interaction between Integrator subunits INTS11 and INTS13. **C)** Superposition of INTS9/11/13^{CMBM} AlphaFold2 structure based on B) (coloured in light yellow) and INTS9/11 experimental structure from A) (PDB:7YCX) with an RMSD value of 1.03 Å across 633 residues. **D)** Zoom in on the superimposed structures shown in C). The previously unknown helix shown in dark blue aligns well with the AlphaFold2 predicted position of INTS13^{CMBM} within the INTS9/11 binding pocket. The cryoEM density map (EMD:33741) is contoured at 3.2 σ , shown in light grey. **E)** Surface representation of the integrative model of the full Integrator complex (including the Arm module from this study) bound to paused RNAPII. Additional upstream DNA was modelled (shown in grey) to visualise its compatibility with the Arm module position.



Supplementary Figure 9. ZNF655 binds to INTS13 via its N-terminal Arm Interacting Motif (AIM). Related to Figure 4. **A)** AlphaFold2 pLDDT (per-residue confidence score) plot for the predicted ZNF655-INTS13 structure. **B)** ZNF655-INTS13 AlphaFold2 predicted structure coloured based on the pLDDT score shown in A. **C)** PAE (Predicted Alignment Error) plot for ZNF655 and INTS13 (right) indicating a possible binding candidate, which was verified experimentally. **D)** Proposed model of ZNF655 bound to the Integrator-RNAPII complex via its Arm Interacting Motif (AIM). DRB Sensitivity Inducing Factor (DSIF) and Negative Elongation Factor (NELF) are bound to RNAPII in the paused state, 60 base pairs downstream of the transcription start site (TSS). RNAPII C-terminal domain is shown in green. The phosphorylated site (in orange) is positioned in close proximity to the PP2A phosphatase module of the Integrator (in yellow), while the nascent capped mRNA (shown in grey) is in the vicinity of INTS11 endonuclease (in blue). Both enzymatic activities lead to RNAPII pause-release and

transcription termination. We hypothesised that ZNF655 or, in general, any other transcription factor bound to the Integrator could modulate its association with the promoter-proximally paused RNAPII. ZNF655 zinc fingers are connected to the INTS13 interacting helix via a disordered region composed of 82 amino acids (AA). Typically, an extended linker of 10 AA has a length of approximately 32 Å; thus, the zinc fingers could, in principle, probe their surroundings within a radius of approximately 260 Å (or +/- 80bp) from the anchoring site at the INTS13-ZNF655 interface. This corresponds roughly to the limiting positions of 140 bp upstream and 20 bp downstream of the pause site.

Supplementary Table 1. Mass photometry data. Related to Figure 2.

| | sample | number of counts | % of total counts | S.D | calculated MW | theoretical MW |
|-------------------------------------|----------------|------------------|-------------------|------|---------------|----------------|
| INTS15^{WT} | INTS10/15 | 193 | 13 | 35 | 104 | 139 |
| | INTS5/8 | 179 | 12 | 15.9 | 219 | 227 |
| | INT5/8 + 10/15 | 810 | 54 | 65 | 367 | 366 |
| INTS15^{M120A/V124A} | INTS10/15 | 452 | 19 | 17 | 131 | 139 |
| | INTS5/8 | 1084 | 47 | 16.4 | 223 | 227 |
| | INT5/8 + 10/15 | 470 | 20 | 84 | 369 | 366 |

Supplementary Table 2. Potential TFs interacting with INTS13 of the Arm module. Related to Figure 3 and Supplementary Figure 9.

| ID | Protein name | pLDDT | pTM | ipTM | INTS13 binding site | Tissue specificity ¹ |
|---------|---|-------|-------|-------|---------------------|---|
| ZNF655 | Zinc finger protein 655 | 54,5 | 0,448 | 0,52 | | low tissue specificity |
| ZBTB21 | Zinc finger and BTB domain-containing protein 21 | 46,8 | 0,366 | 0,534 | | tissue enhanced (bone marrow) |
| ZSC29 | Zinc finger and SCAN domain-containing protein 29 | 52,7 | 0,378 | 0,518 | | low tissue specificity |
| ZNF100 | Zinc finger protein 100 | 56,7 | 0,42 | 0,455 | | low tissue specificity |
| ZNF721 | Zinc finger protein 721 | 55,3 | 0,37 | 0,423 | site 1 | low tissue specificity |
| VDR | Vitamin D receptor | 70,7 | 0,523 | 0,415 | | tissue enhanced (parathyroid gland) |
| ZNF626 | Zinc finger protein 626 | 57,2 | 0,42 | 0,392 | | low tissue specificity |
| ZEB1 | Zinc finger E-box binding homeobox 1 | 43 | 0,362 | 0,381 | | low tissue specificity |
| COE1 | EBF transcription factor 1 | 60,2 | 0,424 | 0,442 | | tissue enhanced (adipose tissue) |
| FOXO3 | Forkhead box protein O3 | 48,9 | 0,399 | 0,413 | | low tissue specificity |
| MECOM | MDS1 and EVI1 complex locus | 44,2 | 0,351 | 0,413 | | tissue enhanced (stomach) |
| ZNF777 | Zinc finger protein 777 | 48,1 | 0,374 | 0,392 | | low tissue specificity |
| SP100 | SP100 nuclear antigen | 50,3 | 0,362 | 0,392 | | tissue enhanced (bone marrow) |
| ZNF609 | Zinc finger protein 609 | 35,8 | 0,336 | 0,39 | site 2 | low tissue specificity |
| FX4L4 | Forkhead box protein D4-like 4 | 56,1 | 0,449 | 0,387 | | not detected |
| PRDM16 | Histone-lysine N-methyltransferase | 42,7 | 0,346 | 0,384 | | tissue enhanced (stomach) |
| KLF8 | KLF transcription factor 8 | 58,7 | 0,463 | 0,382 | | tissue enhanced (skin) |
| ZSCAN4 | Zinc finger and SCAN domain containing 4 | 56,2 | 0,42 | 0,379 | | tissue enhanced (stomach) |
| GMEB1 | Glucocorticoid modulatory element binding protein 1 | 53,7 | 0,403 | 0,378 | | low tissue specificity |
| MYT1L | Myelin transcription factor 1 like | 41,6 | 0,345 | 0,375 | | group enriched (brain, pituitary gland) |
| ZNF280B | Zinc finger protein 280B | 61 | 0,506 | 0,434 | site 1, 2 | retina, testis |
| ZBTB38 | Zinc finger and BTB domain containing 38 | 44,9 | 0,362 | 0,435 | VWA domain | low tissue specificity |
| ZHX3 | Zinc fingers and homeoboxes 3 | 50,9 | 0,38 | 0,413 | VWA domain | low tissue specificity |
| PHF1 | PHD finger protein 1 | 64,6 | 0,442 | 0,378 | clashes with INTS14 | low tissue specificity |
| SMYD3 | Histone methyltransferase SMYD3 | 77 | 0,481 | 0,617 | CTD | low tissue specificity |

¹tissue specificity based on Human Protein Atlas data [S1] (<https://www.proteinatlas.org/>)

Supplementary References

- S1. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419.