Dissertation zur Erlangung des Doktorgrades der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Integrative structural and functional studies of Mediator and RNA polymerase II transcription initiation assemblies

Clemens Plaschka

aus

Wien, Österreich

2015

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Patrick Cramer betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

Göttingen, den 5.06.2015

……

Clemens Plaschka

Dissertation eingereicht am	18.03.2015
1. Gutachter	Prof. Dr. Patrick Cramer
2. Gutachter	PD Dr. Dietmar Martin
Mündliche Prüfung am	27.04.2015

Summary

Gene transcription is a fundamental process of the living cell. Eukaryotic transcription of messenger RNA requires the regulated recruitment of the conserved transcribing enzyme RNA polymerase (Pol) II to the gene promoter. The most heavily regulated step is transcription initiation that involves the ordered assembly of Pol II, the general transcription factors (TF) -IIA, -IIB, -IID, -IIF, -IIE, -IIH and the co-activator Mediator complex. Mediator communicates between transcription regulators and Pol II, and is associated with human disease. Mediator from the yeast *Saccharomyces cerevisiae (Sc)* has a molecular mass of 1.4 megadaltons and contains 25-subunits that constitute a head, middle, tail and kinase module. The core of Mediator contains the head and middle modules that are essential for viability in *Sc*, and directly contact Pol II. Mediator co-operates with TFIIH, to assist assembly and stabilization of the transcription initiation complex and stimulate TFIIH kinase activity. Because of the large size and complexity of Mediator and the initiation machinery, the underlying mechanism remains poorly understood. In this work we studied the structure and function of Mediator head and middle modules, the structure of the reconstituted Pol II-core Mediator transcription initiation complex, and reveal mechanisms of transcription regulation.

We report the crystal structure of the 6-subunit *Schizosaccharomyces pombe* Mediator head module at 3.4 Å resolution. The structure resembles the head of a crocodile and reveals eight elements that are part of three domains named neck, fixed jaw and movable jaw. The neck contains a spine, shoulder, arm and finger. The arm and essential shoulder elements contact the remainder of Mediator and Pol II. The head module jaws and central joint, important for transcription, also interact with Mediator and Pol II. The *Sp* head module structure is conserved and revises a 4.3 Å model of the *Sc* head module, explains known mutations, and provides an atomic model for one half of core Mediator. We further propose a model of the Mediator middle module based on protein crosslinking and mass spectrometry.

To determine how Mediator regulates initiation, we prepared recombinant *Sc* core Mediator by co-expression of its 15 subunits in bacteria. Core Mediator is active in transcription assays and bound an *in vitro* reconstituted core initially transcribing complex (cITC) that contains Pol II, the general factors TFIIB, TBP, TFIIF, and promoter DNA. We determined the cryo-electron microscopy structure of the initially

transcribing core initiation complex at 7.8 Å resolution. The structure reveals the arrangement of DNA, TBP, TFIIB, and TFIIF on the Pol II surface, the path of the complete DNA template strand and three TFIIF elements. The ‘charged helix’ and ‘arm’ of TFIIF subunit Tfg1, reach into the Pol II cleft and may stabilize open DNA. The linker region of TFIIF subunit Tfg2 extends between Pol II protrusion and TFIIB, and may stabilize TFIIB. The structure agrees with its human counterpart, and suggests a conserved architecture of the core initiation complex. Finally, we determined the cryo-electron microscopy architecture of the cITC-core Mediator complex to 9.7 Å resolution. Core Mediator binds Pol II at the Rpb4/Rbp7 stalk close to the carboxy-terminal domain (CTD). The Mediator head module contacts the Pol II dock and TFIIB ribbon and stabilizes the initiation complex. The Mediator middle module ‘plank’ domain touches the Pol II foot and may control polymerase conformation allosterically. The Med14 subunit bridges head and middle modules with a ‘beam’, and connects to the tail module that binds transcription activators located on upstream DNA. The ‘arm’ and ‘hook’ domains of core Mediator form part of a ‘cradle’ that may position CTD and the TFIIH kinase to stimulate Pol II phosphorylation. Taken together, our results provide a structural framework to unravel the role of Mediator in transcription initiation and determine mechanisms of gene regulation.

Acknowledgments

I thank my parents, my brother, and Carrie for their wonderful and constant support.

My great thanks go to Patrick Cramer, who gave me the opportunity to carry out this work in an excellent scientific environment, and constantly provided incredible advice and support throughout the years. I am particularly grateful for the great creative freedom we all enjoy in your lab. I learned a great deal from you, and thank you for always having an open ear and being an amazing mentor.

I wish to thank all current and past members of the Cramer group, who contribute to the excellent, fun, and collaborative atmosphere in the lab, insightful lab seminars, and laid important groundwork for the results presented in this thesis. In particular I wish to thank Laurent Larivière, Carrie Bernecky, and Larissa Wenzeck for great discussions and support. I thank Laurent Larivière, Carrie Bernecky, Elizabeth Villa, Jürgen Plitzko, Jenya Petrotchenko, and Franz Herzog for introducing me to the wonders of X-ray crystallography, electron microscopy, and protein crosslinking. You were wonderful teachers. I thank all my mentors, collaborators, and colleagues from the labs of Patrick Cramer, Christoph Borchers, Franz Herzog, Wolfgang Baumeister, Friedrich Förster, Roland Beckmann, Mario Halic, and Karl-Peter Hopfner. I feel very fortunate to have worked with so many excellent scientists during my PhD and thank all of you for shaping me as a scientist.

Publications

L. Larivière*, C. Plaschka*, M. Seizl, L. Wenzeck, F. Kurth, P. Cramer. Structure of the Mediator head module. *Nature* (2012) 492, 448–451.
*These authors contributed equally.

L. Larivière, C. Plaschka, M. Seizl, E. V. Petrotchenko, L. Wenzeck, C. H. Borchers, P. Cramer. Model of the Mediator middle module based on protein cross-linking. *Nucleic Acids Research* (2013) 41, 9266–9273.

C. Plaschka, L. Larivière, L. Wenzeck, M. Seizl, M. Hemann, D. Tegunov, E. V. Petrotchenko, C. H. Borchers, W. Baumeister, F. Herzog, E. Villa, P. Cramer. Architecture of the RNA polymerase II–Mediator core initiation complex. *Nature* (2015) 518, 376-380.

Contents

Erklärung ii

Summary iii

Acknowledgments v

Publications vi

Contents vii

1 Introduction 1

1.1 Transcription 1

1.1.1 Pol II transcription cycle and regulation 2

1.1.2 Pol II transcription initiation and general transcription factors 3

1.1.3 Structural studies of transcription initiation 5

1.2 The co-activator complex Mediator 6

1.2.1 Early Mediator studies 6

1.2.2 Mediator function in transcription 6

1.2.3 Structural studies of Mediator 9

1.2.4 Studies of core Mediator and Pol II-Mediator complexes 11

1.3 Structural biology methods 11

1.4 Aims and scope 12

2 Materials and Methods 14

2.1 Materials 14

2.1.1 Bacterial strains 14

2.1.2 Yeast strains 14

2.1.3 Plasmids and oligonucleotides 15

2.1.4 Antibodies 17

2.1.5 Growth media and additives 17

2.1.6 General buffers and solutions 17

2.2 Methods 18

2.2.1 Molecular Cloning 18

2.2.2 General methods 20
2.2.3 Specific methods for section 3.1 with focus on crystallography 22
2.2.4 Specific methods for section 3.2 with focus on crosslinking 28
2.2.5 Specific methods for section 3.3 with focus on EM 30
3 Results and Discussion 43
3.1 Structure of the Mediator head module 43
3.1.1 Structures of Med6 and Med17C–Med11C–Med22C 43
3.1.2 Expression, crystallization and structure solution of the *Sp* Mediator head module 45
3.1.3 Structure of *Sp* Mediator head module 47
3.1.4 Revised *Sc* Mediator head module backbone model 50
3.1.5 Structural comparisons reveal head module flexibility 50
3.1.6 Head module integrity and interactions 52
3.2 Model of the Mediator middle module based on protein cross-linking 54
3.2.1 Model of the Med4/Med9 dimer 56
3.2.2 Model of the Med4/Med9-Med7/Med21 tetramer 56
3.2.3 Positioning of Med10 and Med31 56
3.2.4 Elongated architecture of the middle module 57
3.3 Architecture of the RNA polymerase II–Mediator core initiation complex 58
3.3.1 Core Mediator and electron microscopy 60
3.3.2 cITC structure and DNA stabilization 63
3.3.3 cITC–cMed structure 64
3.3.4 Mediator topology 66
3.3.5 Initiation complex stabilization 69
3.3.6 Pol II activation and phosphorylation 72
4. Conclusions and Outlook 76
Appendix 82
List of abbreviations 88

List of figures 89

List of tables 90

References 102

1 Introduction

1.1 Transcription

Gene transcription is the first step to interpret a genome's content. During this process a DNA-dependent RNA polymerase synthesizes a RNA molecule that often serves as a template for protein translation. The polymerase enzyme is a multi-subunit machine featured in the three kingdoms of life[1]. Eukaryotes evolved three nuclear polymerases named RNA polymerase (Pol) I, II, and III that each produce a different class of RNA[2]. Pol I and III synthesize ribosomal 5.8S, 18S and 28S ribosomal RNA (rRNA) and non-translated RNAs such as tRNA and 5S rRNA, respectively. Pol II produces all protein-coding messenger RNA (mRNA) and other non-translated RNA species. In plants two additional polymerases, Pol IV and Pol V, synthesize RNA involved in gene silencing[3]. Mitochondrial and chloroplasts genomes are transcribed by dedicated single-subunit polymerases that most likely evolved from related phage enzymes[4].

Although the different multi-subunit polymerases produce diverse RNA, their core demonstrates great structural conservation. The largest subunits of the bacterial polymerase, β and β', show great homology to their archaeal and eukaryotic counterparts. Eukaryotic Pol I, II and III show a large degree of conservation in the polymerase subunits, associated transcription factors and the mechanism of transcription initiation[2] (Table 1). All three nuclear polymerases are subject to regulation. A particularly complex network of accessory proteins has developed to enable precise regulation of Pol II that generates messenger RNA that serves as the template for protein synthesis.

Table 1 | *Sc* RNA polymerase subunits and initiation factors. Adapted from[2].

Pol I	Pol II	Pol III
Polymerase Core		
A190	Rpb1	C160
A135	Rpb2	C128
AC40	Rpb3	AC40
Rpb5	Rpb5	Rpb5
Rpb6	Rpb6	Rpb6
Rpb8	Rpb8	Rpb8
A12.2	Rpb9	C11
Rpb10	Rpb10	Rpb10
AC19	Rpb11	AC19
Rpb12	Rpb12	Rpb12
Polymerase Stalk		
A14	Rpb4	C17
A43	Rpb7	C25
General Transcription Factors and Their Counterparts		
	Tfg1	C37
	Tfg2	C53
	Tfa1	C82
	Tfa2	C34
		C31
TBP	TBP	TBP
TAFs		
Rrn7	TFIIB	Brf1
Specific Factors		
		B
	Rrn6	
	Rrn11	
SAGA		
TFIIH		
Mediator	Rrn3	
	UAF	
		TFIIIC
		SNAPc

1.1.1 Pol II transcription cycle and regulation

Transcription can be divided into five stages, pre-initiation, initiation, elongation, termination and recycling[5] (Fig. 1). During pre-initiation Pol II and associated factors (Table 1) are recruited to the promoter and form the pre-initiation complex (PIC)[6], the closed DNA is melted and the DNA template strand inserted into the active site[5]. Initiation begins with synthesis of the first RNA nucleotides. When the RNA reaches a length of 13-14 nucleotides Pol II transitions into the elongation phase that accompanies the exchange of initiation for elongation factors. When Pol II reaches the end of the gene, termination occurs that requires the association of termination

factors. The nascent RNA and Pol II are released, and Pol II is recycled and can carry out further rounds of transcription.

Each step of the transcription cycle can be regulated by accessory proteins that interact with genomic DNA, nascent RNA, the Pol II core, and the Carboxy-terminal domain (CTD) of Pol II subunit Rpb1[5]. The CTD is specific to Pol II and consists of an imperfect heptad repeat with the amino-acid sequence $Y_1S_2P_3T_4S_5P_6S_7$. The hydroxyl groups of Y_1, S_2, T_4, S_5, and S_7 are post-translationally modified during the transcription cycle[7]. These modifications are reversible and dynamic and give rise to a complex 'CTD code' that together with nascent RNA regulates the association of transcription factors[8,9]. Pol II can further be regulated by nucleic acids of particular secondary structure[10-12].

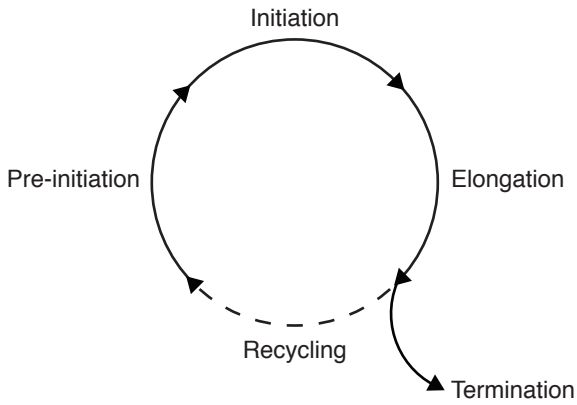


Figure 1 | Schematic of the transcription cycle.

1.1.2 Pol II transcription initiation and general transcription factors

Transcription initiation at eukaryotic protein-coding genes requires Pol II and the general transcription factors (GTF) TFIIA, -B, -D, -E, -F, and -H (Tables 1 and 2, Fig. 2). In the canonical view of initiation[6,13], promoter DNA first assembles with the TFIID subunit TATA-binding protein (TBP), TFIIB, TFIIA and the Pol II–TFIIF complex. TFIID enables promoter recognition and its subunit TBP binds the TATA-box element that resides at a fixed distance of 25-30 nucleotides upstream of the transcription start site in most eukaryotes[13]. TBP bends promoter DNA up to 90° and enables TFIIB to bind in a unidirectional manner, defined by flanking B-recognition elements[13]. TFIIA may associate to stabilize the ternary complex on DNA. Subsequently, TFIIB aids to recruit Pol II-TFIIF, reaches into the Pol II active site, and helps position the transcription start and assists initial RNA synthesis[14-16]. The resulting core initiation complex (excluding TFIIA) is minimally required for *in vitro*

transcription[17] and subsequently binds TFIIE and TFIIH. TFIIE interacts with Pol II, GTFs and single-stranded DNA to stabilize the open complex. TFIIH subunit Ssl2 unwinds promoter DNA using an ATP-dependent translocase[18]. The TFIIH kinase subunit Kin28 phosphorylates the Pol II CTD on S_5 and to lesser extent S_7 to pass a poorly understood initiation checkpoint[8,9]. The assembled initiation complex may undergo initial RNA synthesis that leads to the initially transcribing complex (ITC). After synthesis of the first 13-14 nucleotides the nascent RNA is predicted to clash with the TFIIB ribbon domain that occupies the RNA exit tunnel. This steric clash causes release of TFIIB, and later the release of general factors and formation of the elongation complex. After promoter clearance, TFIIA, TFIID, TFIIE and TFIIH may remain behind at the promoter as a 'scaffold complex' that facilitates additional rounds of initiation[19]. The order of assembly may differ *in vivo*[20].

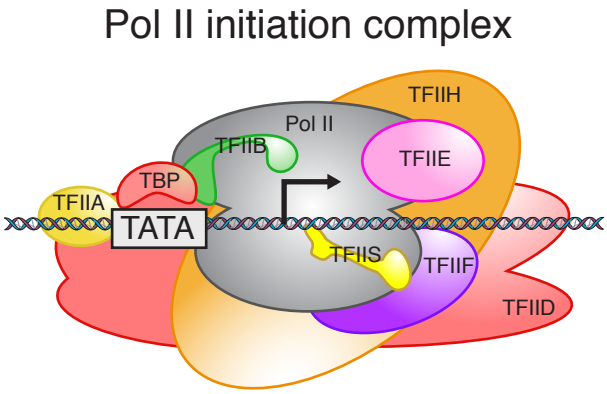


Figure 2 | Schematic of the Pol II transcription initiation complex. Pol II and General Transcription Factors are indicated. TBP binds the TATA-box element. TFIID and TFIIB determine promoter directionality. Adapted from[21].

The general factor TFIIS assists in both initiation and elongation and reaches into the active site[22]. During initiation TFIIS facilitates PIC formation through unknown interactions[23,24] and functionally co-operates with the Mediator co-activator complex to stimulate abortive transcription and productive elongation through the +1 nucleosome, located downstream of the transcription start site[25]. Importantly TFIIS stimulates elongation[26], and relieves unproductive Pol II states caused by backtracking or pausing through enzymatic cleavage of the RNA[22,27].

Table 2 | General Transcription Factors of *Sc* Pol II. Adapted from[28].

Factor	Subunits	Function
TFIIA	2	stabilizes the TFIID-DNA complex
TFIIB	1	stabilizes the TFIID-DNA complex, imposes transcription directionally, helps recruit Pol II-TFIIF, assists in start site selection and initial transcription
TFIID	14	consists of TBP and TBP-associated factors that aid in promoter recognition, interaction with transcription activators, co-activator complexes, and the +1 nucleosome
TFIIE	2	helps recruit TFIIH, stimulates TFIIH kinase and ATPase activities and stabilizes open DNA by interacting with single-stranded DNA
TFIIF	3	assists transcription start site selection, initial transcription and stabilizes the PIC
TFIIH	10	facilitates promoter opening through ATP-dependent translocase, phosphorylates the CTD to assist promoter clearance, carries out DNA repair
TFIIS	1	assists in PIC assembly and initial transcription, reaches into the Pol II active site to reduce pausing and backtracking during elongation

1.1.3 Structural studies of transcription initiation

Crystal structures of TBP in complex with DNA[29], DNA-TFIIB cyclin domains[30] and DNA-TFIIA[31] revealed a minimal upstream DNA binding complex. Crystallographic analysis of *Sc* Pol II revealed the complete 12-subunit enzyme[32]. The position of the initiation factor TFIIB on Pol II was first investigated using 2D electron crystallography[33]. Subsequent biochemical crosslinking analysis from the Hahn group revealed that the TFIIB-ribbon bound close to the Pol II dock domain[34] and was confirmed by crystallographic analysis of the binary Pol II-TFIIB complex[35]. A crystal of Pol II-TFIIB further revealed the position of the TFIIB N-terminal cyclin[14-16] and two TFIIB elements that reach into the Pol II active site and stabilize the DNA template strand[16]. The central dimerization domain of TFIIF[36] was placed on Pol II using protein crosslinking analysis[37,38]. EM analysis by the Nogales group of human initiation complex intermediates revealed the cITC at low resolution and confirmed the general location of DNA, TBP, TFIIB and the TFIIF dimerization domain[39]. Further, TFIIF elements were observed that may reach into the Pol II cleft and contact TFIIB[39]. A recent model of the *Sc* cITC agrees with the human minimal PIC architecture[40]. The general factors TFIIE and TFIIH were reconstituted with the human PIC and visualized by EM. This revealed that TFIIE interacts extensively with the Pol II clamp and stalk domains and that TFIIH resides on downstream DNA[39].

The flexible TFIIH kinase module may reach towards the CTD substrate[39]. Two studies of the yeast *Sc* minimal PIC-Mediator head complex[41] and complete *Sc* PIC[42] are inconsistent with current biochemical and structural findings[43]. Taken together, available evidence suggests that the transcription initiation complex is structurally and functionally conserved among eukaryotes[43] (Fig. 2).

1.2 The co-activator complex Mediator

1.2.1 Early Mediator studies

The *in vitro* reconstitution of transcription shed light on the role of the GTFs in basal transcription. However, the link between basal and activated transcription remained elusive[44-46]. The Young group initially identified suppressors of RNA polymerase B (SRB) that could rescue a cold-sensitive phenotype of Pol II CTD truncation[47-49]. It came as a great surprise when all SRB proteins and additional transcription regulators identified by earlier genetic screens[50] were purified as a single multi-protein assembly in complex with Pol II from yeast cell extract[51]. This complex was named Mediator (Table 3), and the Pol II-Mediator complex was termed the Pol II holoenzyme. Improved preparations of *Sc* Mediator were obtained by pulldown using the classical transcription activators VP16, Gal4 and the Gal4-VP16 fusion protein, and gave first evidence of a direct activator-Mediator interaction[45,51]. Although early studies indicated that Mediator is required for activated transcription only, it has been demonstrated that Mediator facilitates basal transcription when GTFs are present in limiting amounts[52]. Mediator has since been established as a global[53,54] and conserved[55] transcription regulator. A comparative genomics analysis has revealed an evolutionarily conserved set of 17 Mediator subunits[55]. Several species-specific Mediator proteins have been identified that may serve additional regulatory functions[56,57].

1.2.2 Mediator function in transcription

Mediator acts as a regulatory hub that integrates positive and negative transcription signals[58]. It has a central role in gene transcription and defects in Mediator are linked to human disease[59]. Mediator is recruited to enhancer and promoter regions by transcription activators[60,61]. Several activators can bind multiple Mediator subunits[62], such as p53 that can associate with Med17 of the head module or Med1 of the middle

module[63], leading to differential transcriptional output. The great majority of activators target tail module subunits[60] and may permit a co-operative recruitment of Mediator to chromatin DNA. Activator binding may also affect Mediator conformation as suggested from EM studies of human Mediator in complex with VP16, the Sterol response element binding protein, Thyroid hormone receptor, Vitamin D receptor or p53[63-65]. Alternative Mediator conformations may facilitate Pol II recruitment to the promoter by providing a preferred interaction interface. The requirement for an activator transactivation domain can be overcome *in vivo* by artificial recruitment of Mediator[66]. The yeast Adr1-DNA-binding domain was fused to Mediator tail subunits Med3 or Med15 and thereby bypassed a need for the Adr1 transactivation domain[66]. These findings suggest that the primary role of transcription activators is to recruit Mediator to enhancer/promoter DNA.

Promoter-bound Mediator may subsequently recruit GTFs, Pol II, and other transcription regulators. Pol II purified from yeast whole cell extract revealed a Pol II-TFIIF and Pol II-Mediator fraction[67] and suggests a key role of Mediator is its ability to bind the polymerase enzyme and the CTD[68]. TFIIH assists promoter DNA melting and phosphorylates the CTD on serine 5 and 7 *in vivo*[8]. The TFIIH kinase is greatly stimulated in presence of Mediator[45,68]. A yeast two-hybrid screen from the Werner lab first showed a direct interaction between Mediator head module subunit Med11 and TFIIH subunit Rad3[20]. Chromatin-immunoprecipitation (ChIP) experiments suggest that Mediator, TFIIH, and TFIIE form a complex at several yeast promoters prior to Pol II-TFIIF association[20]. This pre-formed complex may resemble a re-initiation competent scaffold that contains several GTFs and is stabilized by an activator[19]. Mediator facilitates recruitment of TFIIB[52], and interacts with TFIID[69] and the TFIID subunit TBP[70,71]. Mediator co-operates with TFIIS during initiation[23,24] and assists transcription through the +1 nucleosome[25]. Altogether these results suggest a role of Mediator in the assembly and/or stabilization of the PIC.

The reversibly associated Mediator kinase module regulates activated and repressive transcription. The kinase module subunit cdk8 was shown to phosphorylate other Mediator subunits, a histone tail, transcription factors and the TFIIH complex and can thereby regulate protein function[72-74]. The species-specific Mediator subunit Med26 is rarely observed in Mediator complexes that contain the kinase module[75]. This near mutually exclusive association has been suggested as evidence of a molecular switch. Structural and functional *in vitro* studies of yeast and human

Mediator complexes suggest an inhibitory function of the kinase module, by occluding the Pol II and/or CTD binding site on Mediator[76-78]. Regulation by the kinase module may include an activating function since it co-localizes with Mediator at active promoters[61,79], interacts with the human elongation factor P-TEFb[80], and can phosphorylate the CTD[72].

Table 3 | List of *Sc* Mediator modules and subunits.

Module	Subunit
Head	**Med6**
	Med8
	Med11
	Med17
	Med18
	Med20
	Med22
Head/Middle	Med19
Middle	Med1
	Med4
	Med7
	Med9
	Med10
	Med21
	Med31
Head/Middle/Tail	**Med14**
Tail	Med2/Med29
	Med3/Med27
	Med5/Med24
	Med15
	Med16
Kinase	Med12
	Med13
	CDK8
	CycC

Conserved *Sc* Mediator subunits are written in bold face.
Subunits essential for yeast viability are underlined. Adapted from[21].

Mediator performs several post-recruitment functions in addition to its established role in PIC formation. Gdown1 is a human Pol II-associated protein that acts inhibitory on transcription initiation[81-83]. This repression can be relieved by Mediator and facilitates exchange for TFIIF to enable productive initiation complex formation[84]. DSIF, the human homologue of yeast Spt4/5, enhances transcription elongation processivity and confers Mediator-dependence in an *in vitro* transcription system with purified factors[85]. In higher eukaryotes up to seven additional Mediator subunits have been identified using mass spectrometry[56,57] that may enable greater

regulatory control during transcription. Med23 was shown to affect alternative splice site selection through interaction with the hnRNP L particle[86]. Med25 may serve as an additional anchor for transcription activators[87,88]. Med26 may assist in the initiation-elongation transition by facilitating exchange of TFIID for the super-elongation complex that contains P-TEFb[89]. Pulldowns of human Mediator coupled to mass spectrometry have identified a plethora of factors that may also interact with Mediator, such as the co-activator SAGA, RNA processing factors, histone modifiers, DNA repair components, and other chromatin-associated factors[75].

1.2.3 Structural studies of Mediator

Yeast *Sc* Mediator is a 25-subunit of 1.4 MDa molecular mass. Mediator consists of biochemically defined head, middle, tail and kinase modules (Table 3). A detailed structural and functional understanding of Mediator is however lacking, due to large conformational flexibility, complex heterogeneity, and difficulty in preparation of large Mediator quantities. Single proteins and subassemblies have been studied at high-resolution using X-ray crystallography and Nuclear Magnetic Resonance (NMR) providing atomic structures for 20 % of Mediator[21] (Fig. 3).

The Mediator complex has been investigated at low resolution using negative stain, cryo-negative stain and cryo-electron microscopy (EM)[64,78,90-92] leading to a canonical view of its conserved modular architecture that was recently revised[93,94]. These studies suggest that Mediator has great intrinsic flexibility that may be important for its function as a transcription regulator.

The Mediator head module is the most highly conserved subassembly of Mediator and is composed of seven subunits, Med6, Med8, Med11, Med17, Med18, Med20, Med22[55]. The head module measures 223 kDa from *Sc*, and 197 kDa from *Schizosaccharomyces pombe* (*Sp*) respectively. Yeast two-hybrid screens and co-immunoprecipitation demonstrated inter-subunit contacts in the Mediator head module[95,96]. Recombinant co-expression of the head module in insect cells[97] enabled negative stain EM analysis of the head module[71]. This revealed three domains named neck, fixed jaw and movable jaw. Further it was shown that the movable jaw contains Med18/Med20. Bacterial co-expression of the Med8 C-terminus/Med18/Med20 heterodimer enabled crystal structure solution of the movable jaw from *Sc*[70] and *Sp*[98]. The Takagi and Kornberg groups solved the architecture of the complete *Sc* Mediator head module by crystallography to 4.3 Å resolution[99]. The structure revealed that the

fixed jaw contains the C-terminal regions of Med11/Med22/Med17, and that the neck contained Med6 and N-terminal regions of Med8/Med11/Med22/Med17. A separate study from the Cramer lab also identified the composition of the fixed jaw and revealed the N-terminal Med11/Med22 4-helix bundle at atomic resolution[100]. An atomic model of the complete head module was however lacking.

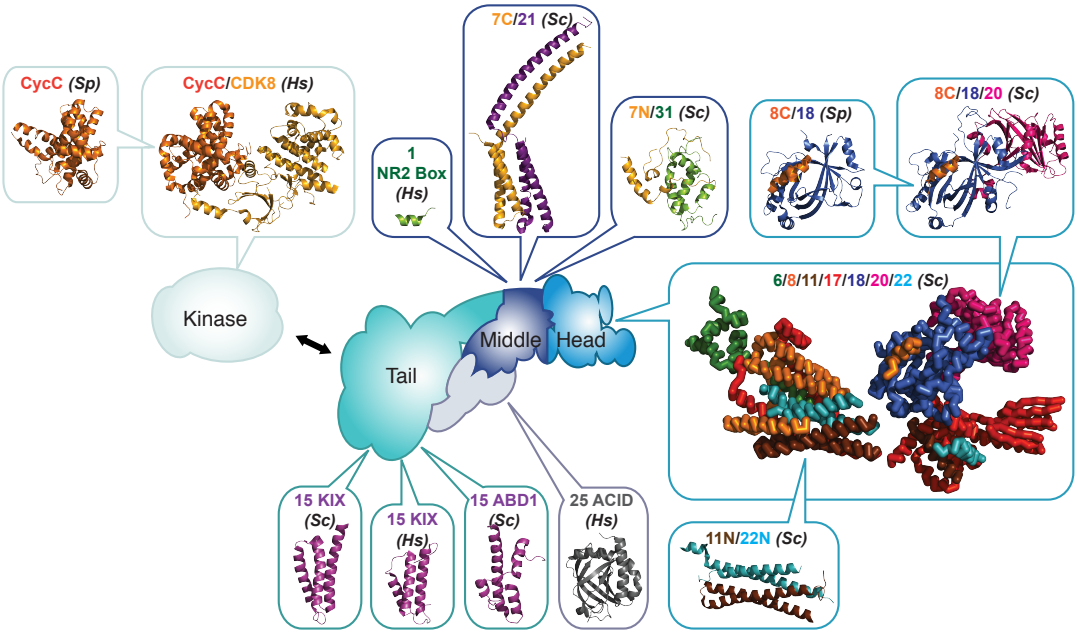


Figure 3 | Known structures of Mediator subunits and subcomplexes. Mediator modules are colored (Head/light blue, Middle/violet, Tail/cyan, Kinase/grey, Species-specific subunits/metallic). The low-resolution head module model is shown as ribbon. Adapted from[21].

The Mediator middle module comprises seven subunits, Med1, Med4, Med7, Med9, Med10, Med21 and Med31. The Med7C/Med21[101] and Med7N/Med31[102] subassemblies of the Mediator middle module were solved by crystallography. Additional subunit interactions have been identified using biochemistry and mass spectrometry tools[95,96,103]. A detailed model of the Middle module is nevertheless lacking, most likely due to a flexible character, revealed by native mass spectrometry and small-angle X-ray scattering[103].

The Mediator tail and kinase modules are less well understood. The tail module binds transcription regulators at enhancer and promoter regions and in *Sc* consist of Med2, Med3, Med5, Med15 and Med16. It is the evolutionary most diverse module and has acquired additional subunits in some species that may allow greater regulatory control. Structural work in yeast and human systems has focused on short transcription activator binding regions of the Med15[104,105] and human-specific

Med25[87,88] subunits. The reversibly associated Mediator kinase module contains Cdk8, CycC, Med12, and Med13 and acts as both a positive and negative regulator of transcription[76,77,79,80]. The kinase Cdk8[101] and the Cdk8/CycC[106] dimer have been studied at high resolution.

1.2.4 Studies of core Mediator and Pol II-Mediator complexes

Biochemical experiments suggest that Mediator has a modular architecture and that modules may be functionally distinct. Mediator head and middle modules act together and are referred to as the core of Mediator (cMed)[67], whereas Mediator tail and kinase modules may act independently[107,108]. cMed was initially identified through fractionation of yeast whole cell extract and mass spectrometry[67], and has a human counterpart named positive cofactor 2[109,110] that was recently reconstituted *in vitro*[111]. cMed further contains all subunits required for viability in yeast[21]. Purified Mediator head module bound to a minimal PIC[97] that contained Pol II, TFIIB, TBP, TFIIF and DNA. However, the head module shows very little activity *in vitro*[97] and the resulting head-minimal PIC was limited to a low resolution cryo-EM envelope[41]. A tail-less Mediator could be prepared by deletion of the Med14 C-terminal region (residues 1-745) or of the Med16 subunit[112]. The ΔMed16 Mediator forms a complex with Pol II, indicating that the tail module is not required for Pol II binding[91]. Together these results indicate that Mediator head and middle modules form the core of Mediator and bind Pol II.

The arrangement of the modules within Mediator was revised based on recent EM analysis[93,94]. EM also visualized endogenous Mediator in complex with Pol II at low resolution from yeast[78,90,113] and human proteins[114]. These studies led to inconsistent locations of Mediator on Pol II, probably owing to Mediator heterogeneity and a lack of general factors.

1.3 Structural biology methods

The combination of three structural biology methods was required to reveal details of Mediator and transcription initiation described in this thesis, protein X-ray crystallography, EM, and protein crosslinking coupled to mass spectrometry.

Protein X-ray crystallography enables the study of macromolecular structures at or near atomic resolution and has revealed many details of biology[115]. Structure determination requires an ordered three-dimensional array of the target molecule in a

crystal, subsequent irradiation using very bright x-rays, and the recording of resulting diffraction patterns. This pattern is related to the ordered lattice of atoms and electrons that interact with the incoming x-rays. This enables determination of the electron density and interpretation of the underlying atomic structure.

EM enables imaging of macromolecules[116]. Negative stain EM is based on the displacement of a heavy atom solution by the target molecule. The dense shell of the molecule can be visualized at low resolution in an electron microscope. Cryo-EM directly visualizes the macromolecule and may therefore reveal details near atomic resolution. To this purpose the sample is applied to a grid and frozen in a thin layer of vitreous ice. The ideally randomly oriented particles are then imaged under low dose in an electron microscope. The single particle images are computationally aligned and their relative orientations determined. Backprojection of the two-dimensional images is then used to calculate an average three-dimensional reconstruction. A benefit of this method is the ability to separate heterogeneity to reveal either conformational or compositional differences. Recent advances in imaging and data processing make EM a powerful alternative to study large assemblies at high resolution[116].

Protein crosslinking coupled to mass spectrometry can define multi-protein architecture or pseudo-atomic models, when individual structures are known[117]. Free amine groups, lysine residues or the protein amino (N)-terminus, are covalently linked by chemical crosslinking. Protease digestion and mass spectrometric analysis of the resultant peptides allows the detection of crosslink sites within (intra) and between (inter) proteins. This provides low-resolution restraints of ~30 Å and is often used in a hybrid structural biology approach with EM, crystallography and modeling.

1.4 Aims and scope

Mediator bridges between transcription activators and the Pol II initiation machinery[60]. Though the role of Mediator in gene activation is likely ubiquitous[53,54], its mechanism remains poorly understood. Structural and functional work has been hampered because the Mediator complex is not abundant in the cell, flexible, modular, and post-translationally modified. How does Mediator regulate transcription initiation? What is the structure of Mediator? How can Mediator interface with transcription regulators, general transcription factors and Pol II at the same time?

Here we aimed to determine the high-resolution structure of the Mediator head module using X-ray crystallography. This work built on previous efforts to obtain

recombinant Mediator complexes by co-expression in bacteria[21]. The head module structure revealed a complex multi-protein assembly, explained genetic and biochemical data, and identified at least one head module element that is functional in transcription. We studied the architecture of the middle module by protein crosslinking and mass spectrometry. Subsequently we aimed to prepare the essential and conserved[111] 15-subunit core Mediator from recombinant material. We determined its architecture in complex with the cITC using EM to 9.7 Å resolution. We further obtained a reconstruction of the cITC to 7.8 Å resolution. This enabled fitting of known atomic models into the cITC and cITC-core Mediator densities. We extended our analysis to the middle module using crosslinking analysis. Complementary functional assays demonstrate regions of cMed that are important for transcription *in vitro* and *in vivo*. Finally, we reveal the architecture of cMed, how Mediator interacts with the initiation complex, how Mediator stimulates TFIIH phosphorylation of the Pol II CTD, and provide new evidence that Mediator exercises global control of transcription.

2 Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

Table 4 | Bacterial strains used in this study.

Strain	Description	Source
XL1- blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE4 relA1 lac(F' proAB lacIqZΔM15Tn10(Tetr))	Stratagene
BL21-(DE3)RIL	B F- ompT hsdS(rB- mB-) dcm+ Tetr gal λ(DE3) endA The (argU ileY leuW Camr), extra copies of argU ileY and leuW tRNA genes on a COlE1-compatible plasmid with chloramphenicol resistance marker, protease deficiency, chromosomal T7-polymerase gene	Stratagene
Rosetta B834 (DE3)	F- ompT hsdSB(rB- mB-) dcm+ metB, methionine auxotroph	Novagen

2.1.2 Yeast strains

Table 5 | Yeast strains used in this study.

Strain	Genotype	Source
BY4741	MATa; his3D1; leu2D0; met15D0;	Euroscarf
Med17/med17Δ	MATa; his3D1; leu2D0; met15D0; ura3D0 srb4::KanMX	Euroscarf
MED17 shuffle - SRB4	MATa; his3D1; leu2D0; met15D0; ura3D0 srb4::KanMX pRS315-SRB4	this work
MED17 shuffle - srb4-ts	MATa; his3D1; leu2D0; met15D0; ura3D0 srb4::KanMX pRS315-srb4-ts	this work
MED17 shuffle - med17Δ321–369	MATa; his3D1; leu2D0; met15D0; ura3D0 srb4::KanMX pRS315-*med17Δ321–369*	this work
MED6/med6Δ	MATa; his3D1; leu2D0; met15D0; ura3D0 MED6::KanMX	Euroscarf
MED6 shuffle - MED6	MATa; his3D1; leu2D0; met15D0; ura3D0 MED6::KanMX pRS315-MED6	this work
MED6 shuffle - med6$_{1–225}$	MATa; his3D1; leu2D0; met15D0; ura3D0 MED6::KanMX pRS315-*med6$_{1–225}$*	this work
MED6 shuffle - med6$_{163-295}$	MATa; his3D1; leu2D0; met15D0; ura3D0 MED6::KanMX pRS315-*med6$_{163–295}$*	this work
SHY349	MATa; ade2D; his3D; leu2D0; met15D; trp1D; ura3D; SRB5-Flag3::kanMx	Steven Hahn
MED7-TAP	BY4743; YOL135C::kanMX4/YOL135C	Euroscarf
Y40343	W303; MATa; tor1-1; fpr1::NAT RPL13A-FKBP12::TRP1	Euroscarf
Med18-FRB-KanMX6	W303; MATa; ade2-1; trp1-1; can1-100; leu2-3,112; his3-11,15; ura3; GAL; psi+; tor1-1 fpr1::NAT RPL13A-2xFKB12::TRP1 SRB5-FRB::kanMX6	Euroscarf

2.1.3 Plasmids and oligonucleotides

Table 6 | Plasmids used in this study.

Vector	Species	Insert	Type	Res.	Source
LL322	*Sp*	Med17-Med22-Med11	pCDFDuet	Sm	LL
LL481	*Sp*	Med20-Med18-Med8	pET21	Amp	LL
LL365	*Sp*	Med17-Med22-Med11-Med6(C-6xHis)-Med20-Med18-Med8	pCDFDuet	Sm	LL
CP001	*Sp*	Med6 (1-180), N-6xHis	pET28b	Kan	this work
CP002	*Sp*	Med6 (118-148), N-6xHis	pET28b	Kan	this work
CP003	*Sp*	Med6 (118-180), N-6xHis	pET28b	Kan	this work
CP004	*Sp*	Med6 (113-148), N-6xHis	pET28b	Kan	this work
CP005	*Sp*	Med6 (113-180), N-6xHis	pET28b	Kan	this work
CP006	*Sp*	Med8 (9-158)	pET21d	Amp	this work
CP007	*Sp*	TBP, N-6xHis	pET28b	Kan	this work
CP008	*Sp*	TBP (52-242), N-6xHis	pET28b	Kan	this work
CP009	*Sp*	Med17 (77-545)-Med22-Med11	pCDFDuet	Sm	this work
CP010	*Sp*	Rpb4-Rpb7, C-6xHis	pET24b	Kan	this work
CP011	*Sp*	Med8-Med18	pET21d	Amp	this work
CP012	*Sp*	Med17 (88-545)	pCDFDuet	Sm	this work
CP013	*Sp*	Med17 (92-545)	pCDFDuet	Sm	this work
CP014	*Sp*	Med17 (97-545)	pCDFDuet	Sm	this work
CP015	*Sp*	Med17 (88-545)-Med22-Med11	pCDFDuet	Sm	this work
CP016	*Sp*	Med17 (92-545)-Med22-Med11	pCDFDuet	Sm	this work
CP017	*Sp*	Med17 (97-545)-Med22-Med11	pCDFDuet	Sm	this work
CP018	*Sp*	Med17ΔC-211	pCDFDuet	Sm	this work
CP019	*Sp*	Med17ΔN-77/ΔC-211	pCDFDuet	Sm	this work
CP020	*Sp*	Med8 (158-223)	pET21b	Amp	this work
CP021	*Sp*	Med17ΔN97ΔC223	pCDFDuet	Sm	this work
CP022	*Sp*	Med17ΔN97Δloop1-3	pCDFDuet	Sm	this work
CP023	*Sp*	Med17ΔC223-Med11-Med22 (1-89)	pCDFDuet	Sm	this work
CP024	*Sp*	Med17ΔC223-Med11(1-91)-Med22 (1-89)	pCDFDuet	Sm	this work
CP025	*Sp*	Med17ΔN97Δloop1-3-Med11-Med22	pCDFDuet	Sm	this work
CP026	*Sp*	Med17ΔN77ΔC223-Med11-Med22	pCDFDuet	Sm	this work
CP027	*Sp*	Med17ΔN77ΔC211-Med11-Med22	pCDFDuet	Sm	this work
CP028	*Sp*	Med18	pET21b	Amp	this work
CP029	*Sp*	Med18, N-6xHis	pET28b	Kan	this work
LL394	*Sc*	Med17-Med22-Med11-Med6-Med8-Med20-Med18	pETDuet	Amp	LL
LL395	*Sc*	Med19-Med14 (1-745), N-10xHis	pET28b	Kan	LL
LL451	*Sc*	Med9-Med4-Med31-Med10-Med7-Med21	pCDFDuet	Sm	LL
LL489	*Sc*	Med17-Med22-Med11-Med6-Med8 (1-189)	pETDuet	Amp	LL
LL490	*Sc*	Med17-Med22-Med1-Med8-Med18-Med20	pETDuet	Amp	LL
LL534	*Sc*	Med19-Med14 (1-745), N-10xHis	pETDuet	Amp	LL
LL535	*Sc*	Med17-Med22-Med11-Med6-Med8-Med20-Med18	pCOLADuet	Kan	LL
CP030	*Sc*	Med19-Med14 (1-745) N-10xHis-SUMO	pETDuet	Kan	this work
CP031	*Sc*	Med19-Med14 (1-745) N-10xHis-Gal4DBD(1-147)	pETDuet	Kan	this work
CP032	*Sc*	Med9-Med4-Med31-Med10	pCDFDuet	Sm	this work
CP033	*Sc*	Med9-Med4-Med7-Med21-Med31	pCDFDuet	Sm	this work
CP034	*Sc*	Med9-Med4-Med10	pCDFDuet	Sm	this work

CP035	*Sc*	Med17, C-Gal4DBD (1-147)	pETDuet	Amp	this work
CP036	*Sc*	Med17 (90-687)-Med22-Med11-Med6-Med8-Med18-Med20	pETDuet	Amp	this work
CP037	*Sc*	Med17 (110-687)-Med22-Med11-Med6-Med8-Med18-Med20	pETDuet	Amp	this work
CP038	*Sc*	Med17 (150-687)-Med22-Med11-Med6-Med8-Med18-Med20	pETDuet	Amp	this work
CP039	*Sc*	Med17 (190-687)-Med22-Med11-Med6-Med8-Med18-Med20	pETDuet	Amp	this work
CP040	*Sc*	Med17 (227-687)-Med22-Med11-Med6-Med8-Med18-Med20	pETDuet	Amp	this work
CP041	*Sc*	Med17 (320-687)-Med22-Med11-Med6-Med8-Med18-Med20	pETDuet	Amp	this work
CP042	*Sc*	Med17 (417-687)-Med22-Med11-Med6-Med8-Med18-Med20	pETDuet	Amp	this work
CP043	*Sc*	Med19-Med14 (40-745), N-10xHis	pET28b	Kan	this work
CP044	*Sc*	Med19-Med14 (77-745), N-10xHis	pET28b	Kan	this work
CP045	*Sc*	Med19-Med14 (100-745), N-10xHis	pET28b	Kan	this work
CP046	*Sc*	Med19-Med14 (130-745), N-10xHis	pET28b	Kan	this work
CP047	*Sc*	Med19-Med14 (164-745), N-10xHis	pET28b	Kan	this work
CP048	*Sc*	Med19-Med14 (204-745), N-10xHis	pET28b	Kan	this work
CP049	*Sc*	Med17-Med22-Med11-Med6-Med8 (29-222)-Med18-Med20	pETDuet	Amp	this work
CP050	*Sc*	Med17-Med22-Med11-Med6-Med8 (59-222)-Med18-Med20	pETDuet	Amp	this work
CP051	*Sc*	Med17-Med22-Med11-Med6-Med8 (87-222)-Med18-Med20	pETDuet	Amp	this work
CP052	*Sc*	Med17, C-GAL4DBD(1-147)-Med22-Med11-Med6-Med8-Med20-Med18	pETDuet	Amp	this work
CP053	*Sc*	Med17, C-GAL4DBD(1-147)-Med22-Med11-Med6-Med8-Med20-Med18	pETDuet	Amp	this work
CP054	*Sc*	Med17 (1-320)-Med22-Med11-Med6-Med8-Med20-Med18	pETDuet	Amp	this work
CP055	*Sc*	Med17 (1-417)-Med22-Med11-Med6-Med8-Med20-Med18	pETDuet	Amp	this work

Resistance, Res.; *Sc*, *Saccharomyces cerevisiae*; *Sp, Schizozaccharomyces pombe*; Kan, Kanamycin; Amp, Ampicillin; Sm, Streptomycin; cloned by Laurent Lariviére, LL. A detailed list of DNA oligonucleotides used for cloning can be obtained from the Cramer group.

Table 7 | Oligonucleotides used for EM.

Type	Sequence (5'-3')
Template DNA	CGA GAA CAG TAG CAC GCT GTG TAT ATA ATA GTG TGT TGT ACA TAG CGG AGG TCG GTG GGG CAC AAC TGC GCT
Non-template DNA	AGC GCA GTT GTG CTA TGA TAT TTT TAT GTA TGT ACA ACA CAC TAT TAT ATA CAC AGC GTG CTA CTG TTC TCG
RNA	AUA UCA

Table 8 | DNA used for transcription assays.

Type	Details	Source
Reverse transcription primer	Sequence: 5'-Cy5-TTCACCAGTGAGACGGGCAACAGCCAAGCTC-3'	Steven Hahn
HIS4 yeast promoter	Plasmid with 428 to +24 relative to start codon, pSH515	Steven Hahn

2.1.4 Antibodies

Table 9 | Antibodies used in this study.

Antibody	Dilution	Host	Source
anti-Rpb3	1:2,000	rat	NeoClone (1Y26)
anti-Ser5-P	1:20	mouse	D. Eick (Helmoltz-Zentrum-München, 3E10)
anti-rat IgG HRP	1:3,000	goat	Sigma (A9037)
anti-mouse IgG HRP	1:3,000	goat	Bio-Rad (170-6516)
anti-TFIIB	1:4,000	rabbit	Abcam (ab63909)
anti-TBP	1:400	rabbit	Santa Cruz (sc-33736)
anti-Med17	1:10,000	rabbit	Steven Hahn (Fred Hutchinson Cancer Research Center)
anti-Flag-tag	1:1,000	mouse	Sigma (F1804)
anti-Med2	1:1,000	goat	Santa Cruz (sc-28058)
anti-rabbit IgG HRP	1:3,000	goat	Santa Cruz (sc-2004)
anti-goat IgG HRP	1:3,000	donkey	Santa Cruz (sc-2020)

2.1.5 Growth media and additives

Table 10 | Growth media for *E. coli* and *S. cerevisiae* cultures.

Media	Organism	Description
LB	*E. coli*	1 % (w/v) tryptone, 0.5 % (w/v) yeast extract. 0.5 % (w/v) NaCl
SeMet	*E. coli*	22 g/L SeMet base, 5g/L nutrient mix, 40 µg/mL SeMet (Molecular Dimensions)
YPD	*S. cerevisiae*	2 % (w/v) peptone, 2 % (w/v) glucose, 1 % (w/v) yeast extract

Table 11 | Additives for *E. coli* cultures.

Additive	Application	Stock solution	Applied concentration
Ampicillin	Antibiotic	100 mg/mL in H2O	0.1 mg/L
Kanamycin	Antibiotic	20 mg/mL in H2O	0.02 mg/L
Streptomycin	Antibiotic	30 mg/mL in H2O	0.03 mg/L
Chloramphenicol	Antibiotic	30 mg/mL in Ethanol	0.03 mg/L
IPTG	Protein expression	1 M in H2O	0.5 mM

isopropyl-b-D-thiogalactoside, IPTG.

2.1.6 General buffers and solutions

Table 12 | List of general buffers and solutions.

Name	Description	Application
Electrophoresis buffer	10x MOPS NuPAGE buffer (Life Technologies)	SDS-PAGE
5x SDS loading dye	250 mM Tris pH 6.8, 50 % Glycerol, 0.03 % bromophenol blie, 7.5 % SDS, 100 mM DTT	SDS-PAGE
SDS-PAGE stain	Instantblue (Expedion)	SDS-PAGE staining
Transfer buffer	NuPAGE TransferBuffer (Life Technologies)	Western blot
PBS buffer	137 mM NaCL, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.76 mM KH_2PO_4, 0.1 % Tween 20, pH 7.4	Western blot
100x PI	0.028 mg/mL Leupeptin, 0.137 mg/mL Pepstatin A, 0.017 mg/mL PMSF, 0.33 mg/mL Benzamidine in Ethanol	Protease Inhibitor

PMSF	100 mM phenylmethanesulfonylfluoride (Sigma-Aldrich) stock solution, in 100 % DMSO	Protease Inhibitor
ABESF	100 mM 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (Sigma-Aldrich) stock solution, in H2O	Protease Inhibitor
10x TAE	50 mM EDTA pH 8, 2.5 M Tris-acetate	Agarose gel
10x TBE	900 mM Tris, 900 mM boric acid, 20 mM EDTA, pH 8.0	Transcription assay
2x ssRNA loading dye	20 % 10x TBE, 8 M RNase free urea, 0.03 % bromopehnol blue, 0.03 % xylene caynol FF	Transcription assay
5x transcription buffer	500 mM K Acetate, 100 mM Hepes pH 7.6, 5mM EDTA pH 8, 25mM Mg Acetate	Transcription assay
5x primer annealing buffer	25mM Tris pH 8.3, 375mM KCl, 5mM EDTA pH 8.0	Transcription assay
5x synthesis buffer	25 mM Tris pH 8.3, 375 mM KCl, 22.5 mM $MgCl_2$, 75 mM DTT	Transcription assay
TBE-Urea-PA Gel solution	8% acrylamide NF(40%/2% solution), 1x TBE buffer, 7 M Urea, filtered and degassed	Transcription assay

2.2 Methods

2.2.1 Molecular Cloning

Cloning strategy

Expression constructs were designed based on previous multi-protein cloning strategies[70,101]. A T7 promoter controlled by the Lac Repressor was placed before each open reading frame (ORF) or sum of ORFs, when the total length did not exceed ~2,000 basepairs. Sequential ORFs were separated by 20-50 basepairs. Each coding region was preceded by a ribosomal binding sequence six nucleotides from the start codon as described[101]. Known binding partners were cloned in sequence, generally under control of the same T7 promoter. For co-expression with multiple plasmids, antibiotic resistance, the origin of replication, and plasmid copy number were taken into account.

Polymerase Chain Reaction (PCR)

PCR primers were designed to contain a GC-rich 5'-overhang, an appropriate restriction site, and 18 or more nucleotides complementary to the sequence of interest[101]. PCR was performed with the Phusion High-Fidelity DNA Polymerase (Finnzymes), in a 50 µL reaction volume. DNA plasmids (Table 6), PCR products, or yeast genomic DNA were used as the PCR template at 1-50 ng or 200 ng respectively.

0.5 pmol/µL PCR primers were used per reaction. Reactions were generally carried out in a Biometra T3000 Thermocycler with 25-30 thermocycles. Primer annealing temperature and synthesis time varied according to DNA template and primer. PCR products were visualized by agarose gel electrophoresis and Sybr Safe stain (Invitrogen), and purified using the QIAquick PCR purification kit (Qiagen).

Restriction digest, dephosphorylation, and ligation

Vectors and purified PCR products were cleaved using the appropriate restriction enzymes and reaction buffer (NEB) overnight at 37 °C. Treated vectors were dephosphorylated with FastaAP (NEB) and purified using the QIAquick PCR purification kit (Qiagen). When a complete ORF was replaced by restriction digestion, the cleaved vector was purified by agarose gel electrophoresis and the QIAquick Gel Extraction Kit (Qiagen). Ligation was carried out with a 5-fold molar excess of insert relative to vector using the Quick ligation kit (NEB).

Transformation and sequencing

For plasmid transformation in *E. coli*, a 50 µL aliquot of chemically competent cells (Table 4) was thawed on ice and incubated with 1 µL plasmid for 10 minutes. Cells were transformed by heatshock at 42 °C for 45 seconds and immediately transferred on ice to recover for two minutes. Subsequently 200 µL LB was added and cells were incubated at 37 °C for 1 hour in a thermomixer. Cells were then plated on LB agar plates that contained the appropriate antibiotic, and were incubated at 37 °C overnight. For sequencing, single colonies were used to inoculate 5 mL LB cultures. Plasmids were purified using the Mini Miniprep purification kit (Qiagen) and sequenced (GATC). For protein co-expression using three plasmids, multiple rounds of transformation were performed and chemical competence re-established between transformations. Typically two plasmids, carrying Kanamycin and Ampicillin resistance, were transformed first. To restore chemical competence, cells were incubated in 2 mL ice-cold 100 mM $CaCl_2$ and 10 % Glcyerol for 30 minutes. Cells were then harvested by centrifugation, resuspended in 100 µL 100 mM $CaCl_2$ and 10 % Glcyerol and used for subsequent transformation.

2.2.2 General methods

Protein expression and purification

Proteins were expressed in *E. coli* and purified as indicated in the respective sections 2.2.3, 2.2.4, and 2.2.5. Generally, proteins were express in 1-8 l LB culture volumes, from a 50-150 mL pre-culture. Cells were grown to the target OD_{600} and induced with IPTG for protein expression at 37 °C for 3-5 hours or at 18 °C overnight. Recombinant material was purified using tailored strategies that usually include affinity purification, ion exchange and size exclusion chromatography (see 2.2.3, 2.2.4, and 2.2.5).

Protein concentrating, determination of concentration, and storage

Proteins were concentrated using AMICON Ultra spin concentrators (Millipore) using a molecular weight cutoff three-fold less than predicted for the target or as determined to recover most protein. Final protein concentration was determined by absorbance at 280 nm using a NanoDrop spectrophotometer. Absorption coefficients were determined using ProtParam (http://expasy.org/tools/protparam.html). Purified proteins were flash-frozen in liquid nitrogen at stored at -80 °C.

SDS-PAGE

Proteins were visualized by SDS-PAGE analysis. The sample was incubated with 5x SDS-PAGE loading dye and heated to 95 °C for 3-5 min and loaded on a pre-cast 4-12% NuPage gradient gel (Invitrogen). Gels were run using 1x MOPS buffer (Life Technologies) at 200 V for 45 minutes. For core Mediator, gels were run at 150 V for 15 minutes and 200V for 35 minutes to better separate protein bands. Gels were stained with InstantBlue (Expedeon).

Western blot

Proteins were applied to SDS-PAGE and transferred to a nitrocellulose membrane (whatman protran BA83 membrane, GE Healthcare) soaked in 1x transfer buffer containing 20 % Ethanol. The transfer was done for 1.5 hours on ice at 30 V using the XCell II Blot Module according to manufacturer instructions (Life Technologies). After transfer the membrane was blocked with 5 % milk in 1x PBS for 30 minutes. 10 mL of primary Antibody (Table 9) diluted in 5 % milk in 1x PBS were added and

incubated at 4 °C overnight. The membrane was washed four times in 30 minutes with 1x PBS and 10 mL secondary Antibody (Table 9), diluted in 5 % milk and 1x PBS, was applied and incubated for 30 minutes. The membrane was washed four times in 30 minutes with 1x PBS, developed using the Pierce enhanced chemiluminescence (ECL) western blotting substrate (Thermo Scientific), and imaged with an Advanced Fluorescence Imager (Intas).

Protein identification

Protein identification was carried out by mass spectrometry by the protein analysis core facility (Adolf-Butenandt-Institute, LMU) and Monika Raabe from the Urlaub group (Max Planck Institute for Biophysical Chemistry, Göttingen).

Scaffold preparation

Oligonucleotides that comprise the DNA-RNA scaffold for EM (Table 7) were dissolved separately in 1x TE at a concentration of 400 µM. Dissolved DNA and RNA were mixed in equimolar amounts (100 µM final concentration) and annealed in a Biometra T3000 Thermocycler that heated to 95 °C and cooled in 1 °C increments every 30 seconds until 10 °C were reached. Annealed scaffold was stored at -20°C.

Initial protein crystallization

Initial protein crystallization screens were performed at the MPI crystallization facility (Max Planck Institute of Biochemistry, Martinsried). On average eight 96-well sitting drop sparse matrix screens were used to identify initial crystallization conditions, from commercial or in house source including Complex (QIAGEN), Complex 2 (QIAGEN), Ammonium Sulphate (QIAGEN), Crystal platform Magic 1 (in house), Crystal platform Magic 2 (in house), Anions (QIAGEN), Cations (QIAGEN), Classics (QIAGEN), Cryos (QIAGEN), PEGs (QIAGEN), and Index (Hampton). Screens were performed at room temperature or 4 °C, with 200 nL drop size (100 nL protein and 100 nL reservoir solution). Additive screens were performed using the Additive Screen HT (Hampton) with a customized the reservoir solution.

Initial crystallization hits were optimized manually in 24- or 15-well hanging drop or 24-well sitting drop plates, or batch crystallization format. Generally, 1 µL reservoir buffer was added to 1 µL protein over a 600 µL reservoir solution at 20°C or

4°C. Total volume was varied from 2-8 µL. The ratio of protein to reservoir solution was varied between 1:1, 1:2, and 2:1. Seed stocks for micro-seeding were prepared with Seed Beads (Hampton) and appropriate seed stabilizing solution. For seeding, 2 µL reservoir buffer was added to 1.5 µL protein, sealed and equilibrated for 30 min, before addition of 0.5 µL seeds. Seed stocks were frozen at -80°C and used multiple times. Crystal feeding was achieved by addition of purified Mediator head module (at 3.5 mg/mL) to drops containing small head module crystals. To influence vapor diffusion rates, silicone and paraffin oils (Hampton) were added to the reservoir solution in different volumes (100-800 µL) and ratios (1:8-8:1).

2.2.3 Specific methods for section 3.1 with focus on crystallography

Vectors and sequences

Vectors used in this section are shown in Fig. 5 and Table 6. Vectors containing a single T7 promoter are based on pET21, pET24 or pET28 vectors (Novagen). Vectors containing two T7 promoters are based on Duet vectors (Novagen). The head co-expression vector that contains three T7 promoters was constructed by concatenating two Duet vectors. Open reading frames (ORFs) were cloned sequentially into vectors. For polycistronic constructs, ribosomal-binding sites were introduced as described[101]. Cells were grown in LB medium at 37 °C to an OD_{600} of 0.5 at 600 nm. Expression was induced with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) for 16 h at 18 °C. Sequence alignments were performed with MUSCLE[118] and manually adjusted based on the presented structures. The graphical representation in Supplementary Fig. 1 was generated with ESPript[119].

Sp Med6 crystal structure determination

For expression of *Sp* Med6 (residues 9–180), cells were transformed with the vector depicted in Fig. 5a and the protein was purified as described[120] except that a HiLoad 16/600 Superdex 200 pg (GE Healthcare) was used for size exclusion chromatography. Crystals were grown at 20 °C in hanging drops over reservoirs containing 100 mM HEPES, pH 7.5, and 400 mM sodium citrate. For crystallization of selenomethionine-labelled[121] protein, sodium citrate was replaced by 16% (v/v) tacsimate. Crystals were collected, cryo-protected by gradually adding glycerol to a final concentration of 30% (v/v), and flash-frozen in liquid nitrogen. Native and

selenomethione single-wavelength anomalous dispersion (SAD) diffraction data were collected at 100 K on a PILATUS 6M detector at the X06SA beamline at Swiss Light Source (SLS), Villigen, Switzerland. Data were processed with XDS and XSCALE[122]. Phenix.autosol[123] was used to detect selenium sites, phase the structure, perform density modification, and for initial model building. The model was manually adjusted using COOT[124] and refined with Phenix.refine[123], including high-resolution data[125] because the resulting final model had a better R_{free} value and stereochemistry. The refined structure has an R_{free} value of 21.7% and shows very good stereochemistry (Supplementary Table 1). Ninety-six per cent of the residues fall in favoured regions of the Ramachandran plot and none of them are in disallowed regions[126].

Sc Med17C-Med11C-Med22C crystal structure determination

Sc Med17C–Med11C–Med22C was expressed in cells co-transformed with the two vectors depicted in Fig. 5b and was purified as *Sp* Med6. For the Med11C–Med22C fusion construct, a linker with sequence GAGSGAGSG was inserted between the C terminus of Med11 and residue 96 of Med22. This covalent linker was essential for complex stability. Crystals were grown at 20 °C in hanging drops over reservoirs containing 100 mM MES, pH 6.0, and 4 M ammonium acetate. Crystals were collected, transferred to a solution containing 100 mM MES, pH 6.0 and 5 M ammonium acetate, and flash-frozen in liquid nitrogen. Selenomethionine labelling and diffraction data collection were as above for *Sp* Med6. Data were processed with XDS and XSCALE[122]. SHELXC/D/E[127] was used to detect selenium sites, phase the structure, and perform density modification. The resulting electron density map allowed for building of most of the model with COOT[124]. The model was refined with Phenix.refine[123], including high-resolution data[125] because the resulting final model had a better R_{free} value and stereochemistry. The refined structure has a R_{free} value of 23.3% and shows very good stereochemistry (Supplementary Table 1). Ninety-six of the residues fall in favoured regions of the Ramachandran plot and none of them is in disallowed regions[126].

Preparation of recombinant *Sc* head module

Sc head module was expressed in cells transformed with the single vector depicted in Fig. 5c. Cells were lysed by sonication in buffer A (50 mM HEPES-potassium hydroxide, pH 7.5, 400 mM potassium chloride, 10% (v/v) glycerol, 5 mM dithiothreitol (DTT)) containing protease inhibitors (Table 12). After centrifugation, the supernatant was loaded onto a 1 ml HisTrap column (GE Healthcare) equilibrated in buffer B (50 mM HEPES-potassium hydroxide, pH 7.5, 500 mM potassium acetate, 10% (v/v) glycerol, 50 mM imidazole, 5 mM DTT). The complex was eluted with a linear gradient from 50 mM to 300 mM imidazole in buffer B. The *Sc* head module was further purified by anion exchange chromatography with a 1 ml HiTrap Q HP column (GE Healthcare). The column was equilibrated in buffer C (50 mM HEPES-potassium hydroxide, pH 7.5, 150 mM potassium acetate, 10% (v/v) glycerol, 2 mM DTT), and proteins were eluted with a linear gradient from 150 mM to 1.25 M potassium acetate in buffer C. Fractions containing head module were applied to a HiLoad 16/600 Superdex 200 pg (GE healthcare) size exclusion column equilibrated in buffer D (20 mM HEPES-potassium hydroxide, pH 7.5, 150 mM potassium acetate, 10% (v/v) glycerol, 2 mM DTT). The *Sc* head module was concentrated to 5 mg ml^{-1}, flash-frozen, and stored at −80 °C. The identity of the polypeptides was confirmed by mass spectrometry. Static light scattering showed that the complex contains one copy of each subunit (not shown). The *Sc* head module variant lacking the Med17 part of the joint ($Med17_{\Delta 321–369}$) was expressed and purified the same way.

Yeast strains and functional assays

Plasmids pRS316-*SRB4* and pRS316-*MED6* were generated by cloning the respective ORF plus 500 base pairs (bp) upstream and 300 bp downstream sequence into pRS316 (ATCC; *URA3* marker). Plasmids pRS315-SRB4, pRS315-$srb4^{ts}$, pRS315-$med17_{\Delta 321-369}$, pRS315-MED6, pRS315-$med6_{1-225}$ and pRS315-$med6_{163-295}$ were generated by cloning the respective wild-type or mutant ORF plus 500 bp upstream and 300 bp downstream sequence into pRS315 (ATCC; *LEU2* marker). The heterozygous *MED17/med17*Δ and *MED6/med6*Δ *Sc* yeast strains were obtained from Euroscarf and transformed with pRS316-SRB4 and pRS316-MED6, respectively. Diploids were sporulated, tetrads were dissected and analysed, and a shuffle strain was selected. To assess functionality of mutants, pRS315 constructs were transformed

into the respective shuffle strain. Equal amounts of freshly grown yeast cells in SC (−Ura −Leu) medium were resuspended in water and tenfold dilutions were spotted on 5- fluoroorotic acid (5-FOA) and SC (−Ura −Leu) plates. Isogenic *SRB4* and *srb4*^{ts} strains used for nuclear extract preparation were obtained by transforming the *MED17* shuffle strain with plasmids pRS315-SRB4 or pRS315-srb4^{ts} and streaking transformants twice on 5-FOA plates. Nuclear extracts were prepared from 3 l of yeast cultures as described[19,100]. *In vitro* transcription and analysis by primer extension were performed as described[120]. Primer extension was done a 5′-Cy5-labelled oligonucleotide (Table 8) for all promoters. For activated transcription we added 200 ng of recombinant full-length Gcn4 that was purified as described[120]. *HIS4* and *TMT1* template plasmids were as described[120]. The *ACT1* template plasmid was generated by inserting the *ACT1* promoter sequence (425 bp upstream to 24 bp downstream of the start codon) in pBluescript KS+ with HindIII and BamHI.

Preparation of recombinant *Sp* head module

The complete *Sp* head module was expressed in cells co-transformed with the three plasmids depicted in Fig. 5d. The crystallized *Sp* head module was expressed in cells co-transformed with the three plasmids shown in Fig. 5e. Cells were lysed by sonication in buffer E (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% (v/v) glycerol, 5 mM DTT) containing protease inhibitors (Table 12). After centrifugation, the supernatant was precipitated with 35% (v/v) saturated ammonium sulphate, and pellets were resuspended in buffer E and loaded on a 2 ml Ni-NTA agarose beads column (QIAGEN) equilibrated in buffer E. The flow-through was reloaded on another 2 ml Ni-NTA column and both columns were washed with buffer E containing increasing concentration of imidazole (0, 10 and 20 mM). The complex was eluted from both columns with buffer E containing 300 mM imidazole. The His-tag was cleaved overnight with 0.7 U thrombin (Sigma-Aldrich) per mg of protein while dialysing against buffer F (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM DTT). The complex was further purified by anion exchange chromatography using a HiTrap Q HP 1 ml column (GE Healthcare) equilibrated in buffer G (20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10% (v/v) glycerol, 2 mM DTT) and eluted with a linear gradient from 50 mM to 500 mM NaCl in buffer G over 70 column volumes. Fractions containing the complex were collected and applied to a HiLoad 16/600 Superdex 200 pg (GE Healthcare) size exclusion column equilibrated in buffer F. The

Sp head module was concentrated to 6 mg ml^{-1}, flash-frozen and stored at −80 °C. The identity of the polypeptides was confirmed by mass spectrometry.

Crystal structure determination of the *Sp* head module

Crystals of the native complex were grown at 3.5 mg ml^{-1} *Sp* head module at 20 °C in hanging drops over reservoirs containing 50 mM MES, pH 6.0, and 1 M ammonium sulphate. Several treatments were required to obtain high-quality crystals, described in 2.2.2 (Initial protein crystallization). Crystals were collected and transferred in a stepwise manner to the final cryo-solution (50 mM MES, pH 6.0, 100% saturated lithium sulphate) and flash-frozen in liquid nitrogen. For structure solution native crystals were derivatized with Ta_6Br_{12} (refs.[128-130]) (Proteros biostructures) and Yb-DTPA-BMA[131] (NatX-ray). Ta_6Br_{12} was added directly to the crystallization drop at 2 mM for 1 h. Yb-DTPA-BMA was added to the final cryo-solution at 100 mM for 10 min and back-soaked 10 s before freezing. The *Sp* head module was further labelled with selenomethionine as described[121]. Crystals of the labelled protein were grown at 2 mg ml^{-1} *Sp* head module at 20 °C in hanging drops over a reservoir composed of 50 mM MES, pH 6.0, 0.95 M ammonium sulphate. Crystals were measured at 100 K at SLS, Villigen, Switzerland. Diffraction data were collected on a PILATUS 6M detector at the X06SA beamline (native crystals and Ta_6Br_{12} and Yb-DTPA-BMA derivatives) or on a PILATUS 2M detector at the X06DA beamline (selenomethionine-labelled crystals). Data were processed with XDS and XSCALE[122]. The HySS submodule in Phenix[123] identified four heavy-atom sites in the Ta_6Br_{12} derivative and a single heavy-atom site in the Yb-DTPA-BMA derivative. These sites were used as input for MIRAS phasing with autoSHARP[132]. Density modification was performed with Resolve[133] and yielded an interpretable electron density map. In parallel, an anomalous difference Fourier map was calculated with data from selenomethionine-labelled crystals and with experimental MIRAS phases. The experimental map, in combination with the selenomethionine sequence markers, allowed the unambiguous manual fitting of the amino-terminal part of Med6 (residues 10–117) and of the Med8C–Med18 subcomplex (PDB code 3C0T). Guided by knowledge of the folds and topologies of the Med11N–Med22N and Med17C–Med11C–Med22C subcomplexes, we built a polyalanine model corresponding to all remaining secondary structure elements with COOT[124]. MIRAS phases were then combined with phases from this initial model and density was modified using the

AutoBuild routine in Phenix[123]. This led to an improved electron density map that allowed for an extension of the initial model. This process was iterated until no further improvement of the map was observed. Then, 23 selenium sites were identified and used as input for SAD phasing with the AutoSol routine in Phenix[123]. The new phases were combined with phases from the partial model. After density modification, an electron density map of excellent quality was obtained, in which many side chains and most of the linkers between secondary structure elements were visible. After rebuilding, the model was refined with autoBUSTER[134]. The resulting $2F_o - F_c$ electron density map allowed further improvement of the model. After several rounds of rebuilding and refinement with autoBUSTER and Phenix.refine[123] using individual B-factor and TLS refinement, we obtained the final model, which shows good stereochemistry and a R_{free} value of 25.8% (Table 13). Including high-resolution data[125] resulted in a refined structure with better R_{free} values and stereochemistry. More than 95% and 99% of the residues fall in favoured and allowed regions of the Ramachandran plot, respectively[126].

Modeling of the *Sc* head module

A hybrid model was built using our structure of *Sc* Med17C–Med11C–Med22C and the structure of *Sc* Med8C–Med18–Med20 (ref.[70]) as models for the fixed and moveable jaws, respectively. For the neck, our *Sp* head module was used, replacing Med11N and Med22 helix α1 with the corresponding *Sc* elements in the *Sc* Med11N–Med22N structure[100]. Other amino acids of the neck were replaced with their *Sc* counterparts, based on sequence alignments. Hydrophobic character of the amino acid situated in the core in the resulting neck model was systematically checked. The obtained models for the two jaws, the spine and the shoulder were separately fitted onto the corresponding elements in the published architectural model of the *Sc* head module[99] (PDB code 3RJ1), using secondary structure matching in COOT[124]. This resulted in a model that explained most of the published electron density. Regions of the model that lacked convincing density were removed, as well as side-chain atoms. The resulting model was adjusted by rigid-body refinement with Phenix.refine[123], using the six structural elements present in the model (partial shoulder, partial arm, spine, tooth, nose and moveable jaw) as separate rigid-body groups. For figure

preparation and structure interpretation, *Sp* Med20 was modelled with MODELLER[135].

2.2.4 Specific methods for section 3.2 with focus on crosslinking

Preparation of *Sc* 6-subunit middle module

Bacterial co-expression of the *Sc* Mediator middle module was performed using a single plasmid based on a pCDFDuet-1 vector (Novagen), shown schematically in Fig. 13a. ORFs were cloned sequentially and additional ribosomal binding sites were introduced as described[101]. Med31 harbors a deca-histidine tag at its N-terminus. The exact sequence of the construct is available on request. The middle module was expressed in *E. coli* BL21 CodonPlus(DE3)RIL cells (Table 4). Cells were grown in LB at 37 °C to an optical density of 0.5 at 600 nm. Expression was induced with 0.5 mM IPTG for 16 h at 18 °C. Cells were lysed by sonication in buffer H (50 mM Tris pH 8.0, 150 mM sodium chloride, 5 mM DTT) containing protease inhibitors (Table 12). After centrifugation, the supernatant was loaded onto a 2 ml Ni-NTA agarose beads column (QIAGEN) equilibrated in buffer H. The column was washed with buffer H containing increasing concentration of imidazole (0, 20, 50 mM). The complex was eluted with buffer H containing 300 mM imidazole. The middle module was further purified by anion exchange chromatography with a 1 ml HiTrap Q HP column (GE Healthcare). The column was equilibrated in buffer I (50 mM Tris pH 8.0, 50 mM sodium chloride, 2 mM DTT), and proteins were eluted with a linear gradient from 50 mM to 1 M sodium chloride in buffer I. Fractions containing middle module were applied to a HiLoad 16/600 Superdex 200 pg (GE healthcare) size exclusion column equilibrated in buffer J (20 mM HEPES-KOH pH 7.5, 150 mM potassium acetate, 10% (v/v) glycerol, 2 mM DTT). The protein complex was concentrated to 3 mg/ml, flash-frozen, and stored at -80 °C.

Chemical protein crosslinking

The pure middle module was cross-linked using isotopically coded cyanurbiotindipropionyl succinimide[136] (CBDPS, Creative Molecules Inc.). The middle module was diluted to 0.5 mg/ml with buffer K (1x PBS, 2 mM DTT). CBDPS was dissolved in DMSO to 10 mM. To determine the optimal ratio of CBDPS to middle module, we mixed 3 µg of middle module with CBDPS at a concentration of 0.05–1.5 mM, and incubated for 30 min at 30 °C. The reaction was

stopped by addition of 0.5 M NH_4HCO_3 to a final concentration of 40 mM and incubation for 10 min at room temperature, and analyzed by SDS-PAGE (Fig. 13b). The optimum concentration of CBDPS was considered to result in a higher molecular weight band. We used a final concentration of 0.7 mM CBDPS.

The cross-linked sample was dialyzed twice in dialysis buttons (Hampton Research) against 20 ml buffer K. Trypsin and/or GluC were respectively added in a 1:10 or 1:1 ratio of protease to middle module and incubated overnight at 37 °C. Proteases were then inhibited by addition of 10 mM AEBSF and 20 mM PMSF, and incubation for 10 min at room temperature. Affinity enrichment was performed with monomeric avidin beads (ThermoFisher) equilibrated with 0.1 M ammonium acetate. The amount of bead slurry was adjusted to a ratio of 1:10 of total CBDPS to bead capacity (1.2 mg/ml). The sample was loaded five times. The column was washed with 300 µl ammonium acetate at concentrations of 0.1, 0.5, and 0.1 M, followed by three 300 µl H_2O washes. The pH was adjusted to 2-3 by addition of 0.1% TFA. Peptides were eluted with buffer containing 0.1% TFA and 50% acetonitrile. The sample was concentrated to 10 µl by lyophilization.

Mass spectrometry

Mass spectrometric analysis was carried out with a nano-HPLC system (Easy-nLC II, ThermoFisher) coupled to the ESI-source of an LTQ Orbitrap Velos mass spectrometer (ThermoFisher Scientific). Samples were injected onto a 100 µm ID, 360 µm OD IntegraFrit trap column (New Objective Inc.) packed with Magic C_{18}AQ (5 µm particle size, 100 Å pore size, Bruker-Michrom) and desalted by washing for 15 min at a flow rate of 300 nl/min with 0.1% (v/v) formic acid. Peptides were subsequently injected on a 75 µm ID, 360 µm OD IntegraFrit analytical column packed with Magic C_{18}AQ (5 µm particle size, 100 Å pore size), equilibrated with 95% solvent A (2% (v/v) acetonitrile, 98% (v/v) water, 0.1% (v/v) formic acid), 5% solvent B (90% (v/v) acetonitrile, 10% (v/v) water, 0.1% (v/v) formic acid). Peptides were separated at a flow rate of 300 nl/min using a 70 minutes gradient (0–60 min: 4–40% solvent B, 60–62 min: 40–80% solvent B, 62–70 min: 80% solvent B).

MS data were acquired with Xcalibur (ver. 2.1.0.1140) with Mass Tags and Dynamic Exclusion precursor selection methods enabled in global data dependent settings. For CBDPS-H8/D8 mass difference between light and heavy isotopic forms of 8.05824 Da was used in Mass Tags setting. Mass Tags and inclusion list runs used

a Top 3 method. MS scans (m/z range from 400 to 2000) and MS/MS scans were acquired in the Orbitrap mass analyzer at 60000 and 30000 resolution, respectively. Fragment ions for MS/MS acquisition were produced by collision-induced dissociation (CID) at normalized collision energy of 35% for 10 ms at activation q=0.25. Data analysis was performed with DXMSMS Match of ICC-CLASS[136]. Two additional cross-link pairs (Med7(K35)–Med7(K103) and Med4(K36)–Med9(K117)) were obtained from a preliminary 7-subunit middle module preparation and were included in the final data set.

Structure prediction and modelling

Protein sequences of Med4, Med9 and Med10 from fungal species[137] and higher eukaryotes were aligned with MUSCLE[118]. The generated alignments were used for structure prediction with HHPred[138]. Med4, Med9 and Med10 were predicted to be partly homologous to Med7, Med21 and Med8, respectively.

For middle module modeling, the *Sp* Mediator head structure[139] (PDB code 4H63) was used as a template. *Sc* Med11N/Med22N (PDB code 3R84) was superimposed on *Sp* Med11/Med22, either superimposing *Sc* Med22N on *Sp* Med11, or superimposing *Sc* Med22N on *Sp* Med22. Med7C/Med21 (PDB code 1YKH)[101] was then superimposed on *Sc* Med11N/Med22N for both positions. Protein superimposition was performed using secondary structure matching in COOT[124]. A second Med7C/Med21 dimer was positioned in the same relative orientation as in the crystal, letting the open ends of the C-terminal coiled-coils interacting with each other[101]. The two resulting models, formed of Med8 and two copies of Med7C/Med21, were used as templates for final middle module modeling using MODELLER[135]. For this purpose, Med8 and the second copy of Med7C/21 served as templates for their structurally homologous regions in Med10 and Med4/9, respectively.

2.2.5 Specific methods for section 3.3 with focus on EM

Vectors and sequences

Vectors for co-expression of *Saccharomyces cerevisiae* (*Sc*) core Mediator (cMed) subunits in *Escherichia coli* are shown in Fig. 17a. Open reading frames (ORFs) of Med19 and Med14 (1–745; ref.[112]) with an additional N-terminal 10×histidine tag were cloned sequentially into a pET Duet vector (Novagen, Table 6). The histidine

tags of previously cloned genes were removed[139,140]. Ribosomal binding sites were introduced as described[101]. Sequences are available upon request. All proteins were expressed in *E. coli* BL21(DE3)RIL (Stratagene). Mediator head and middle modules were expressed and purified as described in sections 2.2.3 and 2.2.4 (refs.[139,140]).

Preparation of core Mediator

For co-expression of the 15-subunit cMed, *E. coli* BL21(DE3)RIL cells were transformed with three plasmids (Fig. 17a) and grown in LB medium at 37 °C to an optical density of 0.7 at 600 nm. Expression was induced with 0.5 mM IPTG for 24 h at 18 °C. Cells were harvested and resuspended in buffer M (50 mM HEPES-KOH pH 7.5, 400 mM KCl, 10% glycerol, 10 mM imidazole, 5 mM DTT) containing protease inhibitors (Table 12). After sonication and centrifugation, the supernatant was loaded on a HisTrap HP 5 ml column (GE Healthcare) equilibrated in buffer N (25 mM HEPES-KOH pH 7.5, 400 mM KCl, 10% glycerol, 25 mM imidazole, 5 mM DTT). The column was washed with seven column volumes (CVs) of buffer N. The complex was eluted with a linear gradient from 25 mM to 300 mM imidazole in buffer N over 10 CVs. Fractions containing the complex were diluted 1:3 in buffer O (25 mM HEPES-KOH pH 7.5, 100 mM KCl, 10% glycerol, 1 mM EDTA, 5 mM DTT) and applied to a HiTrap Q HP 5 ml column (GE Healthcare) equilibrated in buffer O. The complex was eluted with a gradient from 100 mM to 800 mM KCl in buffer O over 25 CVs. Fractions containing the complex were concentrated and applied to gel filtration using a Superose 6 10/600 GL (GE Healthcare) equilibrated in buffer P (25 mM HEPES-KOH pH 7.5, 400 mM KCl, 3 mM DTT). Purified cMed was concentrated to 3 mg ml^{-1}, flash-frozen and stored at −80 °C. Up to 2 mg of pure cMed could be obtained from 8 l cell culture (Fig. 16a). The identity of the protein subunits was confirmed by mass spectrometry.

Promoter-dependent *in vitro* transcription

Nuclear extract (NE) from 3 l yeast culture of strain SHY349 was prepared as described[19,100]. SHY349 NE was immuno-depleted of endogenous Mediator a s described[19], with minor modifications. 150 µl NE were dialysed for 90 min against buffer Q (20 mM HEPES-KOH pH 7.6, 75 mM ammonium sulphate, 10 mM $MgSO_4$, 20% glycerol, 1 mM EGTA pH 8.0, 0.5 mM DTT) containing protease inhibitors

(Table 12). 75 µl anti-Flag M2 agarose beads (Sigma) were washed twice with buffer E, and incubated in NE for 1 h at 4 °C on a turning wheel. After centrifugation, the supernatant was depleted with the same amount of beads used initially. Beads were removed by centrifugation. The doubly depleted NE was flash-frozen and stored at −80 °C. Tandem affinity purification (TAP) of endogenous Mediator was performed using a C-terminal TAP tag on Med7 as described[100]. Purified TAP-Mediator was flash-frozen and stored at −80 °C. Activator-dependent and independent promoter-dependent *in vitro* transcription and primer extension was performed as described[100]. Purified TAP-Mediator (~0.25 pmol), 7-subunit head module (1.25 pmol), 6-subunit middle module (1.25 pmol) or cMed (2.5 pmol) were added as indicated in Fig. 16b.

Preparation of cITC–cMed complex

S. cerevisiae 12-subunit Pol II[141], TBP (residues 61–240)[29], TFIIB[16], and TFIIF (*Saccharomyces mikatae* Tfg1, *S. cerevisiae* Tfg2)[37] were prepared as described. The nucleic acid scaffold previously used to generate the core Pol II initially transcribing complex (cITC)[40], was used for assembly of the cITC–cMed complex (Table 7). Pol II (150 µg at 3 mg ml^{-1}) was incubated with a fourfold molar excess of TFIIF for 5 min at 25 °C. A twofold molar excess of nucleic acid scaffold and a fourfold molar excess of TBP and TFIIB were added to buffer R (25 mM HEPES-KOH pH 7.5, 180 mM potassium acetate, 5% glycerol, 5 mM DTT) and incubated with pre-formed Pol II–TFIIF complex for 10 min at 25 °C. cMed was added to the cITC in a 1.2-fold molar excess over Pol II and incubated for 50 min at 25 °C. The sample was cooled in 5 min intervals from 25 °C to 20 °C, 15 °C, 10 °C and 4 °C. The cITC–cMed complex was purified by gel filtration using a Superose 6 10/600 GL equilibrated in buffer R. Fractions containing the complex were concentrated to ~0.6 mg ml^{-1} and extra nucleic acid scaffold was added in equimolar amount.

Binary interaction assays

To test the binary interaction of cMed with glutathione-*S*-transferase (GST)–CTD, GST and GST–CTD were expressed separately in *E. coli* BL21(DE3)RIL cells and induced at an optical density of 0.5 at 600 nm with 0.5 mM IPTG. Cells were lysed by sonication in buffer S (20 mM HEPES-KOH pH 7.5, 100 mM KCl, 10% glycerol, 100 µM EDTA, 0.1% NP40, 1 mM PMSF, 5 mM DTT). After centrifugation, the

supernatant was applied to 1 ml glutathione Sepharose 4B resin (GE Healthcare), equilibrated in buffer S. The resin was washed twice with 10 ml buffer S. 10 µg purified cMed were incubated for 3 h on ice with 30 µl immobilized GST or GST–CTD resin. The resin was washed four times with 600 µl buffer G and bound proteins were analysed by SDS–PAGE.

To test the binary Pol II–cMed interaction by protein pull-down, 3 µg purified Pol II was biotinylated on the Rpb3 subunit as described[142] and immobilized on 20 µl Dynabeads M280 Streptavidin resin (Life Technologies), equilibrated in buffer R. 1.5 µg cMed were incubated with immobilized Pol II or control beads for 1 h at 4 °C. Beads were washed four times and bound proteins were analysed by SDS–PAGE. To further assess the stability of Pol II–cMed by gel filtration, 80 µg Pol II were incubated with 1.5-fold less cMed as done for the cITC–cMed complex (see Preparation of cITC–cMed complex). Pol II–cMed complex was applied to size exclusion chromatography on a Superpose 6 10/600 GL equilibrated in buffer R, and peak fractions were analysed by SDS–PAGE.

To test the interaction of Pol II with cMed variants comprising mutations in interface A, B and C, 2 µg recombinant cMed variant were incubated with 3 µg biotinylated Pol II immobilized on 15 µl Dynabeads M280 Streptavidin resin (Life Technologies) equilibrated in buffer R. Beads were washed four times and bound proteins were analysed by SDS–PAGE. cMed variants were purified according to the protocol for complete cMed, yet gave much lower yield.

Protein crosslinking and mass spectrometry

80 µg of purified cITC–cMed were crosslinked[117] with 1–4 mM di-succinimidyl-suberate (DSS-d0/d12, Creative Molecules Inc.) or 1–4 mM bis-sulfo-succinimidyl-suberate (BS3-d0/d12, Creative Molecules Inc.) as described[143]. Crosslinked samples were digested with trypsin or AspN. Crosslinked peptides were enriched and analysed on a liquid chromatography system coupled to the electrospray ionization (ESI) source of an Orbitrap Elite mass spectrometer (Thermo Scientific)[143]. Crosslinks were identified by xQuest[144] as described[143]. 35 µg of purified cMed were crosslinked with 0.1–0.15 mM cyanurbiotin-dipropionyl succinimide (CBDPS-d0/d8, Creative Molecules Inc.) as described[136]. Crosslinked cMed was digested with trypsin and/or GluC, peptides were enriched and analysed on a liquid chromatography system coupled to the ESI source of an LTQ Orbitrap Velos mass spectrometer (Thermo

Scientific) as described[140]. Crosslinks were identified by DXMSMS Match of ICC-CLASS[136].

Electron microscopy

Purified cITC–cMed complex was crosslinked with 2 mM BS3 (Sigma) for 30 min at 30 °C, and the reaction quenched with 50 mM ammonium bicarbonate. The crosslinked sample was purified by gel filtration using a Superose 6 10/600 GL equilibrated in buffer R. Fractions containing the complex were concentrated to ~0.15 mg ml^{-1}. Negatively stained samples of the cITC–cMed were prepared on continuous carbon coated grids (Quantifoil). Grids were glow-discharged 30 s before deposition of 4 µl sample (~0.15 mg ml^{-1}) and incubated for 1 min. Grids were blotted between sequential transfers to two 40 µl drops distilled water, stained for 1 min in a 40 µl drop 2% (w/v) uranyl acetate solution, and blotted until dry. To obtain an unbiased initial model, a 3D reconstruction was generated from particles selected from negative-stain tomography data. Single axis tilt series were recorded using Serial EM[145] on a FEI Tecnai F20 microscope operated at 200 keV. Images were acquired from −54° to 54° with an angular increment of 3°. Images were recorded on a 4k×4k Gatan Ultrascan CCD camera with a defocus of –2 µm and at a nominal magnification of 68,000× (2.21 Å pixel^{-1}). The cumulative dose per tomogram did not exceed 100 e^{-} Å^{-2}. To refine the negative-stain tomography reconstruction, a further 103 micrographs of the untitled sample were acquired with a range of defocus values (from –0.5 µm to –1.5 µm) and used for single-particle analysis.

Cryo samples of the cITC–cMed were prepared on lacey carbon copper grids (Quantifoil). Grids were glow-discharged for 20 s before deposition of 4 µl sample (~0.15 mg ml^{-1}) and incubated for 30 s. Grids were washed twice with 4 µl distilled water, blotted, and vitrified by plunging into liquid ethane with a manual plunger. Data was acquired using the TOM toolbox[146] on a FEI Titan Krios operated in EFTEM mode at 300 keV. 2,972 movies were collected using a Gatan K2 Summit direct detector with a range of defocus values (from –1 µm to –2.5 µm) at a nominal magnification of 37,000× (1.35 Å pixel^{-1}). The camera was operated in 'super-resolution' mode (0.675 Å pixel^{-1}) with exposure times of 0.2–0.3 s per frame, a dose rate of ~8 e^{-} pixel^{-1} s^{-1}, and a target total dose of 25–30 e^{-} Å^{-2}. Movies were binned once in Fourier space, and partitioned into 2,048^2 quadrants that were aligned and

averaged using a CUDA implementation of a previously described algorithm[147]. The averaged $2{,}048^2$ images were used for image processing.

Image processing

Negative stain tomography data was processed using the TOM toolbox[146]. The tilt series was contrast-transfer function (CTF) corrected as described[148]. Due to the absence of colloidal markers, images were aligned using feature tracking before weighted back projection. 175 subvolumes were selected manually using EMAN2[149] and extracted with a 160^3 voxel box size from the reconstructed tomogram. Reference-free alignment was performed in PyTom[150]. The obtained volume was used as the template for a 6D correlation search of the same tomogram[151], yielding 675 subvolumes. Subvolumes were aligned and averaged in PyTom to obtain a 3D reconstruction with an estimated resolution of 28 Å (Fourier shell correlation, FSC = 0.5) (Supplementary Fig. 4a).

For negative-stain single-particle analysis, 21,365 particles were selected semi-automatically using e2boxer.py from EMAN2[149]. All 3D reconstructions were performed in RELION[152]. The unbiased negative-stain tomography reconstruction was low-pass filtered to 60 Å and used as an initial model for 3D reconstruction. The selected particles were extracted with a 160^2 pixel box and pre-processed to normalize images and to remove pixel values more than 5 standard deviations from the mean value. Particle images were sorted by unsupervised 3D classification in RELION[152] into four classes with the regularization parameter T set to 1.5, an initial angular sampling interval of 7.5°, an offset search range of 5 pixels, and offset search steps of 1 pixel (Supplementary Fig. 3a). This yielded one class of 8,815 particles that showed density for all components of the cITC–cMed. This class was refined using the 3D auto-refine procedure in RELION[152] with default parameters, to an estimated resolution of 25.2 Å with the 'gold-standard' FSC = 0.143 (Supplementary Fig. 4c-e). For validation of the reconstruction quality, reference-free 2D class averages were calculated from all particles included in the final reconstruction using RELION[152]. The obtained averages were compared with SPIDER-generated[153] forward projections of the final reconstruction (Supplementary Fig. 4b).

For cryo-EM single-particle analysis, 89,769 particles were selected semi-automatically using e2boxer.py from EMAN2[149]. CTF parameters were estimated using CTFFIND[154]. CTF correction and 3D reconstruction were performed in

RELION[152]. The selected particles were extracted with a 280^2 pixel box and pre-processed to normalize images and remove pixel values more than 5 standard deviations from the mean value. Sorting of particle images by unsupervised 3D classification led to three classes, Pol II–DNA/RNA (14,777 particles), cITC (4,439 particles), and cITC–cMed (3,267 particles) (Supplementary Fig. 3b). Each class was refined only against the respective particles within this class using RELION[152]. All reference maps were filtered to 60 Å before refinement using the 3D auto-refine procedure in RELION[152]. Pol II–DNA/RNA and cITC classes were refined using a soft spherical reference mask (300 Å diameter), an initial angular sampling interval of 7.5°, an offset search range of 3 pixels, and offset search steps of 1 pixel. Pol II–DNA/RNA and cITC 3D reconstruction were refined to an estimated resolution of 6.6 Å and 7.8 Å, respectively, at FSC = 0.143 (Supplementary Fig. 4h, m). For visualization and structural modelling of the EM densities, temperature factors of –240 $Å^2$ (Pol II–DNA/RNA) and –340 $Å^2$ (cITC) were applied. The cITC–cMed class was refined using the 3D auto-refine procedure in RELION as Pol II–DNA/RNA and cITC, except with a soft SPIDER-generated[153] reference mask in the shape of the cITC–cMed (maximum diameter of 360 Å) and an initial offset search range of 5 pixels. The cITC–cMed 3D reconstruction was refined to an estimated resolution of 9.7 Å at FSC = 0.143 (Supplementary Fig. 4r). For visualization and structural modelling of the cITC–cMed EM density, a temperature factor of –340 $Å^2$ was applied.

For validation of the reconstruction quality, reference-free 2D class averages were calculated from all particles included in the final reconstructions. The obtained averages were compared with SPIDER-generated[153] forward projections of the final reconstructions (Supplementary Fig. 4f, k, p). To confirm the presence of only a single cITC–cMed conformation, we repeated 3D classification with rejected particles from class 1, 2, 5, 7, 8 of round 1 and cITC–cMed particles from class 4 of round 1 (see 3D classification of cryo-EM data) with the Pol II–DNA/RNA reconstruction filtered to 60 Å resolution as initial model (Supplementary Fig. 3c, d). For these 3D classifications, an initial angular sampling of 7.5°, an offset search range of 5 pixels, and offset search steps of 1 pixel were employed.

3D classification of cryo-EM data

Unsupervised 3D classification of the cryo-EM data was performed using RELION[152] in a pseudo-hierarchical manner that consisted of seven rounds of classification

(Supplementary Fig. 3b). The obtained classes were iteratively reclassified and/or combined, based on the structures of Pol II[155], human minimal pre-initiation complex[39] and cITC–cMed. The regularization parameter T was set to 4 for all rounds of classification. The negative-stain single-particle reconstruction of the cITC–cMed complex was low-pass filtered to 60 Å as the initial model for round 1 of 3D classification. The cryo-EM reconstruction of the human minimal pre-initiation complex (EMD-2305) was low-pass filtered to 60 Å as the initial model for round 2b of 3D classification. All classification was performed with soft reference masks as detailed for 3D refinement of Pol II–DNA/RNA, cITC and cITC–cMed (see Image processing).

Round 1 served to sort all 89,769 particles into eight classes to discard faulty particle images and identify a population of the complete cITC–cMed. The following initial sampling parameters were used: angular sampling of 7.5°, an offset search range of 15 pixels, and offset search steps of 3 pixels. Round 1 converged after 55 iterations and gave rise to Class 3 (~13,400 particles, 14.8% of data) that showed strong density for the cITC and partial density for cMed. This class was submitted to a second round of classification (round 2a) into four classes with an initial angular sampling of 7.5°, an offset search range of 5 pixels, and offset search steps of 1 pixel. All following rounds of classification were performed using the same initial sampling parameters used for round 2a. Class 3 of round 2a (3,267 particles, 3.7%) revealed the cITC–cMed complex with equal intensity of cITC and cMed, and was consequently refined as detailed (see Image processing). To explore any remaining heterogeneity, class 3 of round 2a was sorted into four classes (round 3a). This yielded class 1 and 2, lacking density for either upstream DNA–TFIIF–TFIIB–TBP or upstream DNA–TFIIB–TBP and the mobile plank of cMed (Supplementary Fig. 3b), and class 3 and 4 that varied appreciably in the mobile plank domain (Fig. 24d). Due to the small number of particle images, classes from round 3a were not refined.

Class 3 and 6 of round 1 showed density for Pol II and were consequently combined (~50,000 particles, 55.5%) and classified into eight classes (round 2b). This yielded class 2 (~14,000 particles, 15.5%) that presented a cITC with weak density for general transcription factors and upstream DNA. Class 2 of round 2b was therefore further classified into three classes (round 3b) that resulted in the cITC containing class 2 (~3,700 particles, 4.1%). To obtain a larger set of homogenous cITC particles, class 1 of round 2a and class 2 of round 3b were combined (~7,300

particles, 8.1%) and sorted into four classes (round 4b). This led to class 2 of round 4b (4,439 particles, 4.9%) that revealed the complete cITC. The complete cITC was subsequently refined as detailed (see Image processing).

Class 1 and 8 of round 2b and class 3 of round 3b displayed density only for Pol II–DNA/RNA and were consequently combined (~25,000 particles, 27.7%) and further sorted into four classes (round 5). The resulting class 3 (14,777 particles, 16.4%) showed density for the Pol II–DNA/RNA complex with weak density for the Rpb4–Rpb7 subcomplex and was subsequently refined as detailed (see Image processing). Class 1 of round 5 (~3,200 particles, 3.5%) presented density for 12-subunit Pol II–DNA/RNA, but was not refined due to the small number of particles and poor orientational distribution (Supplementary Fig. 3b). Class 4 of round 2a and class 1 of round 3b displayed density for the binary Pol II–TFIIF–DNA/RNA complex and were combined (~12,900 particles, 15%, round 6). This complex presented with identical features found in the cITC and was not analysed further (Supplementary Fig. 3b).

Local resolution, filtering and variance estimation

Local resolution maps (LRMs) were obtained using a method[156] implemented to run on a Graphics Processing Unit. A sliding window of 40^3 voxels was centred around each voxel, extending the original half-maps through mirroring at the borders. The FSC at 0.143 was then calculated within the window and assigned to the central voxel. Given the low particle counts involved in the reconstructions and the resultant non-uniform sampling, five differently randomized pairs of half-maps generated in RELION[152] were processed independently and the results averaged to obtain a more robust estimate. Nevertheless, the maximal value for the local resolution was capped at the global FSC = 0.143 value to prevent enhancing of noise. Therefore, even though some regions, notably the Pol II density, clearly exhibited higher resolution than the nominal, LRMs were used only to limit the resolution locally, for a conservative interpretation of our data.

To perform local filtering, maps were downscaled for the Nyquist frequency to match the highest frequency value within the respective LRM. Look-up maps were then created by low-pass filtering the original map to each integer frequency value present in the LRM. For each voxel, a value linearly interpolated between the two

look-up maps closest to its non-integer frequency was copied to the output. Finally, the locally filtered maps were rescaled to their original size.

Variance maps were generated for each cryo-EM reconstruction as an additional metric for assessing reconstruction quality and structural variability. These were obtained by implementing a described bootstrap approach[157], using the direct Fourier inversion method[158] for the reconstructions. As RELION's data processing pipeline could not be used for this task, the aligned micrographs were CTF-corrected by means of automatic defocus determination and phase flipping in the TOM toolbox[146]. Single particle views were then extracted using RELION's position and rotation estimates. Differently sampled 12.5% fractions of the entire particle set were used to create 4,000 reconstructions. The variance between these reconstructions was calculated for each voxel position and normalized by the respective intensity value (Supplementary Fig. 4j, o, t).

Structural modelling

To generate unbiased models, known structures or homology models were sequentially rigid-body fitted using an automated global 6D correlation search in Situs[159] (Figs 19a and 21a). As the majority of fitted models account for only a small fraction of the density, a Laplacian filter was applied for fitting (except for fitting the 10-subunit Pol II structure into the Pol II–DNA/RNA and cITC maps, and for fitting the cITC model lacking Rpb4–Rpb7 into the cITC–cMed map). After fitting of each component, difference maps were generated using UCSF Chimera[160] to reduce the search space in subsequent searches.

For the cITC model, we automatically fitted crystal structures of Pol II [16,161], TFIIB residues 22–213[16], TBP[29], and homology models of TFIIF[40], and the TFIIB C-terminal cyclin domain[40] (Fig. 19a). Pol II, DNA, and the Tfg1 charged helix were slightly adjusted manually using COOT[124] to accurately reflect the density. The crystal structure of the initially transcribing complex[27] (PDB code 4A3D) was used for fitting of Pol II–DNA/RNA and initial fitting of cITC, as it presents the most complete model of Pol II except for the protrusion domain. For the Pol II–DNA/RNA model Rpb4–Rpb7 was excluded, due to the weak density in this region. For fitting of the 10-subunit Pol II–TFIIB model, we combined structures of initially transcribing Pol II (PDB code 4A3D, excluding Rpb4–Rpb7) and initially transcribing Pol II–

TFIIB complex (PDB code 4BBS, chain M, TFIIB residues 22–118). After visual inspection of the cITC density, the nucleic acids and Pol II domains clamp core, wall and protrusion in the initially transcribing Pol II structure (PDB code 4A3D) were replaced with the corresponding regions from the initially transcribing Pol II–TFIIB complex structure (PDB code 4BBS). For fitting of Rpb4–Rpb7, chains D and G from the initially transcribing Pol II structure (PDB code 4A3D) were used.

For fitting the structure of the TFIIB N-terminal cyclin domain–TBP–DNA complex, the atomic coordinates of the equivalent *Homo sapiens* complex (PDB code 1VOL) were used as a template. TBP and the TFIIB N-terminal cyclin of 1VOL were replaced with *S. cerevisiae* TBP (PDB code 1YTB, chain A) and the *S. cerevisiae* TFIIB N-terminal cyclin (PDB code 4BBS, chain M and residues 122–213), respectively. The model of the TFIIF dimerization domain was generated as described[40], containing *S. mikatae* Tfg1 residues 92–143 and 343–412, and *S. cerevisiae* Tfg2 residues 54–138 and 208–227. The homology model of the TFIIB C-terminal cyclin (residues 233–342) was based on the *H. sapiens* TFIIB C-terminal cyclin (PDB code 1VOL). The Tfg1 charged helix was placed manually using COOT[124] according to the chemical environment. Further density in shape of duplex DNA was observed upstream and downstream of TATA-box DNA, and downstream of the active site. This density was modelled with canonical duplex B-DNA, generated in COOT[124]. NAMD[162] was used to optimize DNA geometry using the CHARMM force field[163]. The DNA sequence of all nucleic acid models was edited in COOT[124] to match the sequence of the employed nucleic acid scaffold.

The cITC–cMed model was generated by sequential fitting of cITC lacking Rpb4–Rpb7, and the three regions of an improved *S. cerevisiae* Mediator head module model (Fig. 21a). The construction of an improved head module was required because the published structures are at 4.3 Å resolution and contain out-of-register errors and are lacking several regions[99,164]. To account for the known flexibility within the head module[139], its regions neck (including joint), movable jaw, and fixed jaw were fitted independently. To model the fixed jaw tooth and the nose domains, we used the structure of *S. cerevisiae* Med17C–Med11C–Med22C (PDB code 4H62[139]). For the movable jaw, the structure of *S. cerevisiae* Med8C–Med18–Med20 (PDB code 2HZS[70]) was completed with Med18 residues 69-93 from the structure of *S. cerevisiae* Med18–Med20 (PDB code 2HZM, chain D) to account for additional residues visible in the electron density. The shoulder domain was modelled with

MODELLER[135] based on its *S. pombe* counterpart (PDB code 4H61[139]). For the arm, finger, spine and joint domains of the head module, our *S. pombe* head module (PDB code 4H63[139]) was used as a model. Med11 and Med22 helices α1 and α2 were replaced with the corresponding helices from the *S. cerevisiae* Med11N–Med22N structure (PDB code 3R84). Other *S. pombe* residues were replaced with their *S. cerevisiae* counterparts, based on sequence alignments. The obtained models for all head module domains except the finger were individually fitted onto the corresponding elements in the published structure of the *S. cerevisiae* head module (PDB code 4GWP), using secondary structure matching in COOT[124]. A $2F_o$-F_c electron density map corresponding to the published *S. cerevisiae* head module structure was calculated. The map was used to fit our model of the finger element and to adjust the rest of our model when required. The linker between helices α4 and α5 in Med8 and the *Sc* specific helix between strand β1 and helix α3 were built manually. The geometry of the obtained head module model was regularized in COOT[124] and with PHENIX[123]. All figures and movies were generated using UCSF Chimera[160] and PyMol.

Immobilized template assay

The immobilized template assay was performed essentially as described[100] using nuclear extract from *srb4-ts*[54] or wild-type (BY4741) *S. cerevisiae* strains and a linear *HIS4* promoter[19,100]. The *srb4-ts* extract was complemented with recombinant head module (4.5 pmol) as indicated in Fig. 25b. Samples were applied to SDS–PAGE, transferred onto a polyvinylidene difluoride membrane (Millipore) and probed by antibodies anti-Rpb3 (1Y26, NeoClone, dilution 1:1000), anti-TFIIB (ab63909, Abcam, dilution 1:4000), anti-TBP (sc-33736, Santa Cruz, dilution 1:400), anti-Med17 (1:10000 dilution), kindly provided by Steven Hahn (Fred Hutchinson Cancer Research Center), anti-Flag-tag (F1804, Sigma, 1:1000 dilution), and anti-Med2 (sc-28058, Santa Cruz, 1:1000 dilution). Secondary antibodies anti-rat IgG HRP (A9037, Sigma, dilution 1:3000), anti-mouse IgG HRP (170-6516, Bio-Rad, dilution 1:3000), anti-rabbit IgG HRP (sc-2004, Santa Cruz, dilution 1:3000) and anti-goat IgG HRP (sc-2020, Santa Cruz, dilution 1:3000) were used. Antibody detection was achieved with Pierce enhanced chemiluminescence (ECL) western blotting substrate (Thermo Scientific) and Amersham Hyperfilm ECL (GE Healthcare).

Comparative Dynamic Transcriptome Analysis (cDTA)

cDTA enables global analysis of newly synthesized RNA[165] that reveals defects in transcription with much higher sensitivity than conventional steady-state methods. cDTA was carried out as described[165] using the *S. cerevisiae* heterozygous *Med17/med17Δ* strain (Euroscarf) transfected with plasmids pRS315-SRB4[139] or pRS315-srb4-ts[139], and Y40343-wildtype (Euroscarf) or Med18-FRB-KanMX6 (Euroscarf) strains. Heat shock of *SRB4* and *srb4-ts* strains was applied for 18 or 60 min at 37 °C before RNA labelling as described[165]. To deplete the Med18 subunit from the nucleus, anchor-away experiments were performed by rapamycin treatment (1 μg ml^{-1} in 200 ml YPD) for 18 or 60 min at 30 °C before RNA labelling[166]. Data analysis was as described[165].

CTD phosphorylation assay

Endogenous TFIIH was purified as described[17]. Purified Pol II (2.5 pmol), 7-subunit head module (1.25 pmol), 6-subunit middle module (1.25 pmol), cMed (2.5 pmol) were added as indicated in Fig. 27f and incubated on ice for 10 min before addition of TFIIH (~0.02 pmol) and 300 μM ATP. Reactions were incubated for 25 min at 25 °C, applied to SDS–PAGE, transferred to a nitrocellulose membrane (GE Healthcare) and probed with primary antibodies anti-Ser5-P (3E10, dilution 1:20), provided by D. Eick (Helmholtz-Zentrum München), and anti-Rpb3 (1Y26, NeoClone, dilution 1:2,000). Secondary antibodies anti-rat IgG HRP (A9037, Sigma, dilution 1:3,000) and anti-mouse IgG HRP (170-6516, Bio-Rad, dilution 1:3,000) were used. Antibody detection was achieved with Pierce enhanced chemiluminescence (ECL) western blotting substrate (Thermo Scientific) and an Advanced Fluorescence Imager (Intas).

3 Results and Discussion

3.1 Structure of the Mediator head module

The work presented in this chapter was published.

L. Larivière*, C. Plaschka*, M. Seizl, L. Wenzeck, F. Kurth, P. Cramer. Structure of the Mediator head module. *Nature* (2012) 492, 448–451.
*These authors contributed equally.

3.1.1 Structures of Med6 and Med17C–Med11C–Med22C

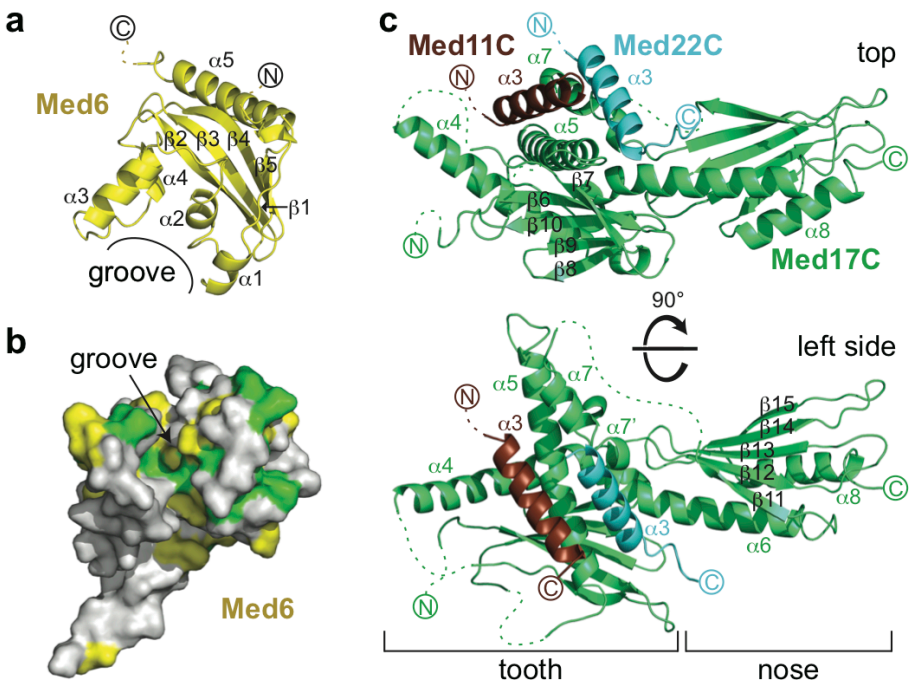


Figure 4 | Structures of Med6 and Med17C–Med11C–Med22C. a, Ribbon model of the *Sp* Med6 crystal structure. **b,** Conserved surface groove in Med6. Residues that are invariant or conserved among seven yeast species are in green or yellow, respectively (Supplementary Fig. 1). **c,** Ribbon model of the *Sc* subcomplex Med17C–Med11C–Med22C crystal structure. 'Tooth' and 'nose' domains are indicated. Dashed lines indicate disordered regions.

Extending structural analysis of Mediator[21], we determined the crystal structure of a bacterially expressed Med6 variant (*Sp* residues 9–180) at 2.7 Å resolution (Fig. 4a, Supplementary Fig. 1, and Supplementary Table 1). The structure revealed a core domain with a five-stranded antiparallel β-sheet, two pairs of α-helices flanking a conserved groove, and a flexible C-terminal helix α5 (Fig. 4a, b). Bacterial expression also provided the head subcomplex Med17C–Med11C–Med22C, which consists of C-terminal regions in Med17, Med11 and Med22 (ref.[100]) and constitutes the fixed jaw[99] (see 2.2.3 and Fig. 4b). The crystal structure at 3.0 Å resolution (Supplementary Table

1) revealed two subdomains we call ‘tooth’ and ‘nose’ (Fig. 4c). The tooth contains Med11C, Med22C and Med17C residues 383–541 and 594–611. It forms a five-stranded β-sheet that is flanked by five helices, three from Med17C and one each from Med11C and Med22C. The nose comprises Med17C residues 542–580 and 612–687 and forms another five-stranded β-sheet with two flanking helices.

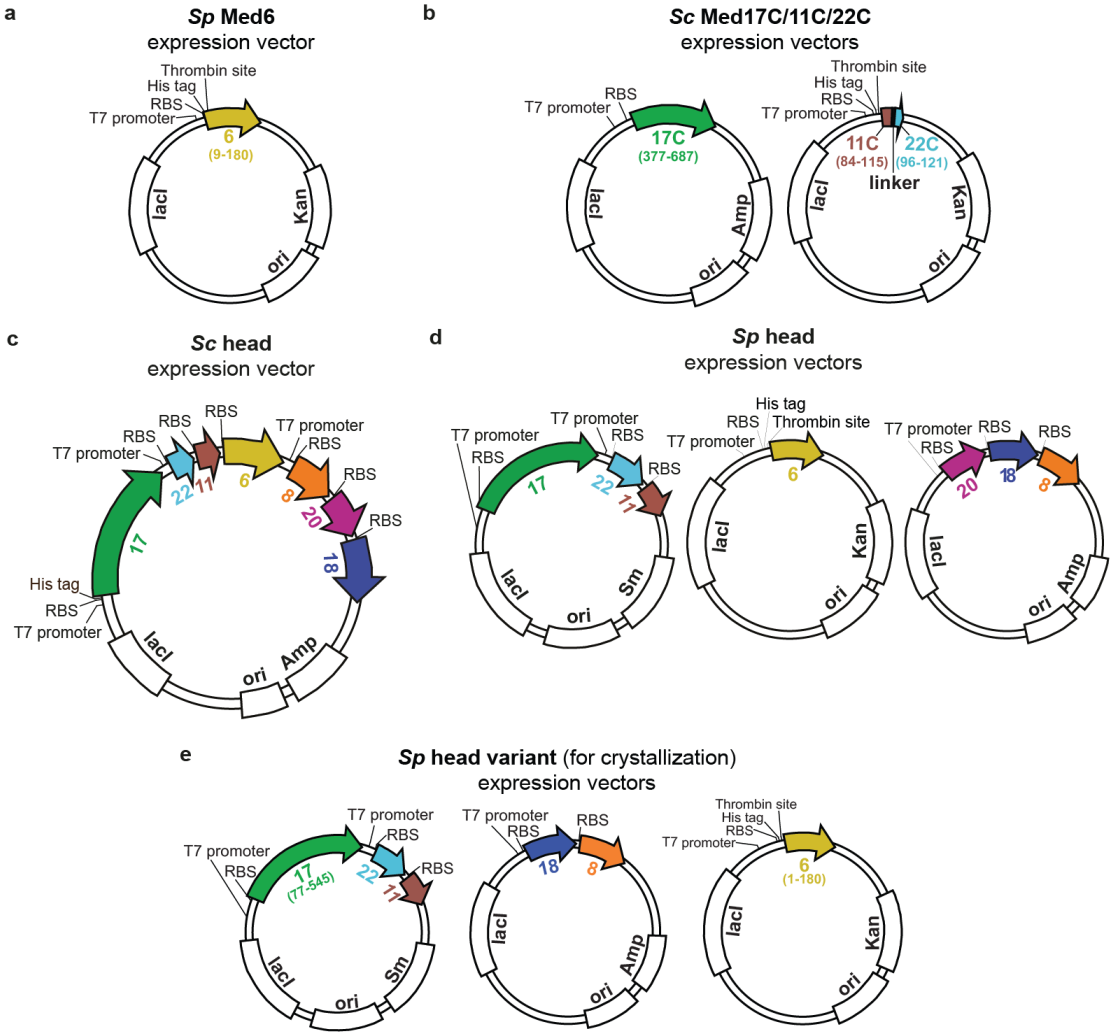


Figure 5 | Bacterial expression vectors. Schematic representation of the plasmids used in this study for the expression of the **a,** crystallization construct of *Sp* Med6. The variant lacks only the non-essential terminal tails. **b,** *Sc* Med17C/Med11C/Med22C subcomplex. Two plasmids were used for co-expression, one expressing Med17C, the other expressing a Med11C-Med22C fusion polypeptide (see Supplementary Methods). **c,** full length *Sc* head module. Co-expression was driven from a single plasmid with three T7 promoters, one for Med17 expression, one for tricistronic expression of Med22, Med11, and Med6, and one for tricistronic expression of Med8, Med18, and Med20. **d,** full length *Sp* head module. Three plasmids were used for co-expression, each harboring a single T7 promoter, one expressing Med6, one expressing Med8, Med18, and Med20, and one expressing Med17, Med22, and Med11. **e,** *Sp* head variant used for crystallization. RBS, ribosome binding site; His tag, 6xHistidine tag; *ori*, origin of replication; *lacI*, gene encoding Lac repressor. *Sm*, *Amp* and *Kan* refer to streptomycin, ampicillin and kanamycin resistance genes, respectively.

3.1.2 Expression, crystallization and structure solution of the *Sp* Mediator head module

We prepared the entire *Sc* head module by co-expressing its seven subunits in *Escherichia coli* (see 2.2.3 and Fig. 5c). The purified module (Fig. 6a) supported activator- and promoter-dependent transcription from three different promoters *in vitro* (Fig. 6b). Crystals of the *Sc* head module were obtained in 2007, but persistently diffracted to low resolution. We therefore prepared the head module from *Sp* (Fig. 5d). The purified module (Fig. 6a) did not crystallize, but crystals were obtained for a variant lacking three non-essential parts, Med6 residues 181–216, Med17 residues 1–76 and Med20 (Fig. 5e and 6a).

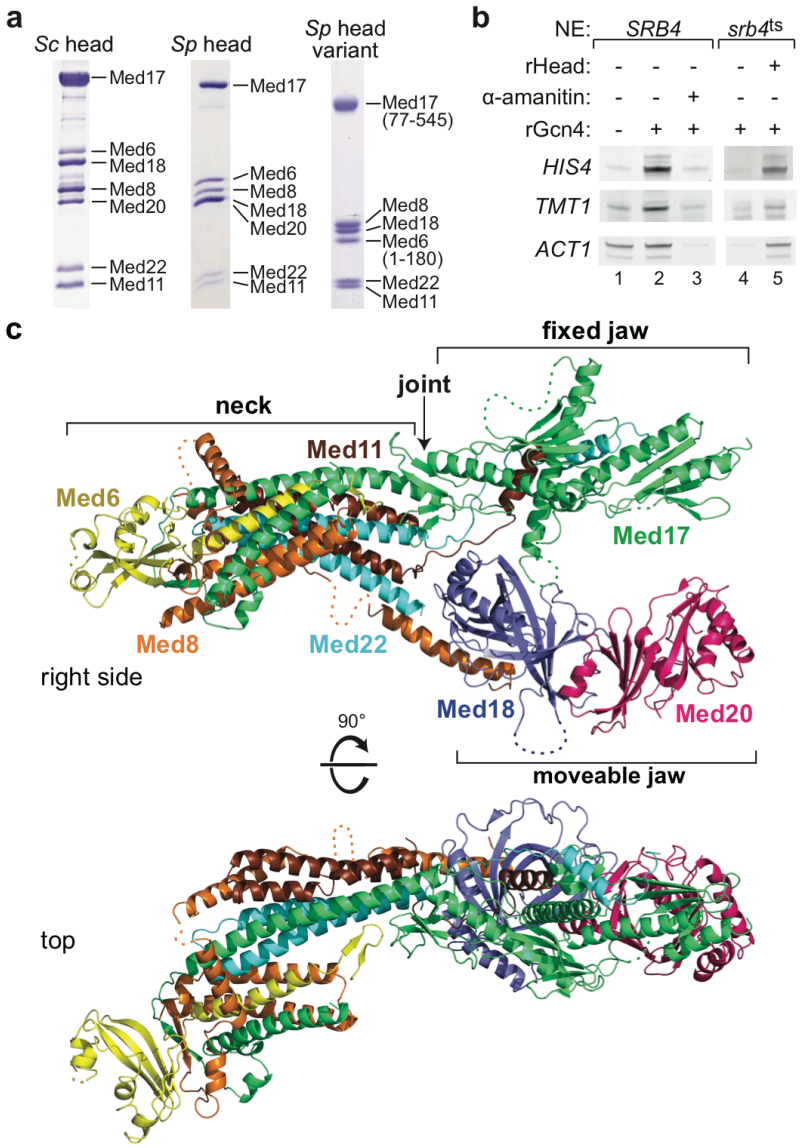


Figure 6 | Structure of Sp Mediator head module. a, SDS–polyacrylamide gel electrophoresis (PAGE) analysis of purified recombinant head modules of *Sc* (left), full-length Sp (middle) and the crystallized *Sp* variant (right). **b,** Recombinant *Sc* head module supports

in vitro transcription from three different promoter types. Activity of nuclear extracts from *SRB4* and *srb4-138* (*srb4^{ts}*) strains was assessed on promoters *HIS4*, *TMT1* and *ACT1*. Pol II-dependency was confirmed by a-amanitin sensitivity (0.04 mM, lane 3). Addition of 5 pmol recombinant *Sc* head module (lane 5) to inactive *srb4^{ts}* extracts (lane 4) restored activity. **c,** Two views of a ribbon representation of *Sp* Mediator head module structure. Subunits are in different colours. Subunit Med20 is not present in the crystals but was unambiguously modelled and positioned by superimposing the previous *Sc* Med18–Med20 heterodimer structure[70].

We determined the *Sp* head module structure *de novo* by multiple isomorphous replacement with anomalous scattering (see 2.2.3 and Fig. 8a). Modelling required selenomethionine sequence markers (Fig. 7b) and the four available structures of subcomplexes (Fig. 4, ref[70,100]). Diffraction data to 3.4 Å resolution (Table 13) resulted in excellent electron density (Fig. 7c, d) and a refined structure that has a free *R* factor (R_{free}) of 25.8% (Table 13) and contains 87% of the residues in the variant.

Table 13 | Data collection and refinement statistics for *Sp* head module crystals.

	Native	Ta_6Br_{14}	Yb-DTPA-BMA	SeMet
Data collection				
Space group	$P3_221$	$P3_221$	$P3_221$	$P3_221$
Cell dimensions				
a, *b*, *c* (Å)	145.6, 145.6, 241.6	145.4, 145.4, 233.2	146.0, 146.0, 232.6	146.0, 146.0, 238.6
Wavelength (Å)	0.99977	1.25537	1.38633	0.97940
Resolution (Å)†	3.4 (3.49-3.40)*	4.7 (4.82-4.70)	4.0 (4.1-4.0)	3.8 (3.9-3.8)
$CC_{1/2}$‡	99.9 (24.0)	99.9 (13.8)	99.7 (23.0)	99.6 (22.1)
Completeness (%)	100.0 (99.9)	99.9 (99.9)	99.7 (97.2)	99.9 (99.2)
Redundancy	10.0 (9.2)	4.6 (4.7)	5.2 (5.2)	5.2 (4.7)
R_{merge} (%)	12.5 (412.4)	9.1 (474.1)	20.4 (860.3)	18.1 (246.5)
$I/\sigma I$	15.2 (0.6)	8.4 (0.4)	7.6 (0.3)	8.9 (0.6)
Refinement				
Resolution (Å)	126–3.4			
No. reflections	41,454			
R_{work}/R_{free}	23.2/25.8			
No. atoms				
Protein	9254			
Ligand				
B-factors				
Protein	166			
Ligand				
R.m.s deviations				
Bond lengths (Å)	0.009			
Bond angles (°)	1.13			

*Highest resolution shell is shown in parenthesis.

†Resolution limits are provided using the $CC_{1/2} > 10\%$ criterion[167]. Using the traditional criterion of $I/\sigma I > 2.0$, the resolution limit of the native crystal is 3.7 Å.

‡$CC_{1/2}$ = percentage of correlation between intensities from random half-datasets[167].

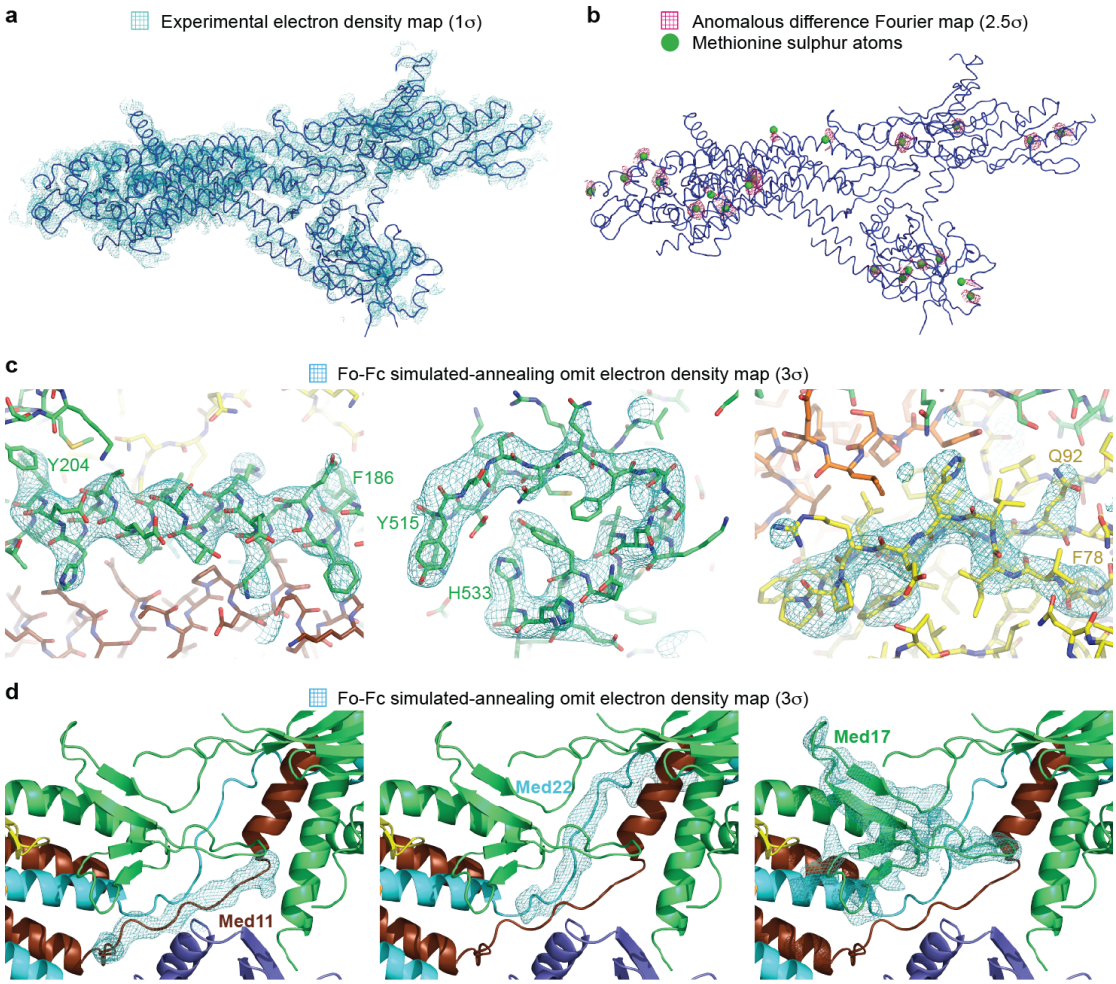


Figure 7 | Electron density maps for the Sp Mediator head module. a, Initial unbiased MIRAS electron density map (cyan, contour level 1σ) with the final structure superimposed (blue coils, right side view). **b,** Anomalous difference Fourier map for selenomethionine-containing head module (magenta, contour level 2.5σ) with an outline of the structure superimposed (blue ribbon, right side view). Sulphur atoms of methionine residues are shown as green spheres. **c,** mFo–DFc simulated annealing omit maps (cyan, contour level 3s) with the final model superimposed, shown as sticks with carbon atoms coloured by subunit (see Fig. 6). Shown are three regions of the head module, Med17 helix α3 (left), a part of the nose (middle), and a part of Med6 (right). Med17 residues 186 to 204, Med17 residues 515 to 533, and Med6 residues 78 to 92 were omitted in the map calculation, respectively. **d,** mFo–DFc simulated annealing omit maps (cyan, contour level 3σ) for Med11 and Med22 linker regions (left and middle, respectively) and the Med17 b-sheet from the 'joint' (right). The final model is superimposed, shown as ribbons and coloured according to subunits (see Fig. 6). Med17 helix a4 was omitted for clarity.

3.1.3 Structure of *Sp* Mediator head module

The *Sp* head module structure reveals an asymmetric multiprotein assembly that resembles the head of a crocodile with one limb (Figs 6c and 8a). The structure reveals all folds and previously unobserved regions, and is partitioned into eight elements (Fig. 8a, b). The neck submodule is formed by Med6, Med8 and parts of Med11, Med17 and Med22, and consists of 'spine', 'shoulder', 'arm' and 'finger'

elements (Figs 8a, b and 9). The spine comprises seven helices, six forming a long bundle (Med8 α5, Med17 α3, Med11N–Med22N (ref.[100]), and a seventh perpendicular helix (Med8 α4). The arm consists of a short four-helix bundle (Med6 α5, Med8 α1/2, Med17 α1), two pairs of helices (Med17 η1/α2, Med8 α2′/α3), and a two-stranded sheet (Med8 β1, Med17 β1). The arm binds the shoulder, which consists of the Med6 core (Fig. 4a). The finger is an exposed C-terminal β-hairpin in Med6 (Fig. 8a, d). The fixed and moveable jaws consist of subcomplexes Med17C–Med11C–Med22C (Figs 4c and 8a) and Med8C–Med18–Med20 (ref.[70]), respectively.

Both jaws are connected to the spine. The fixed jaw is connected via the 'joint' that contains a four-stranded β-sheet formed by Med17 residues 215–259, and short linkers in Med11 and Med22. The joint may change its structure as it contains many conserved glycine and serine residues. Deletion of the Med17 part of the joint did not cause lethality in yeast, but essentially abolished activated transcription *in vitro* (Fig. 10a, b). The moveable jaw binds the spine via Med8C, a helix extending from the Med18–Med20 heterodimer[98] (Fig. 8a, d), and also contacts the joint and tooth.

Conservation of the module surface is highest at the shoulder and arm (Fig. 8c), probably because this region binds the middle module[96,101]. Deletion of the shoulder caused synthetic lethality (Fig. 10).

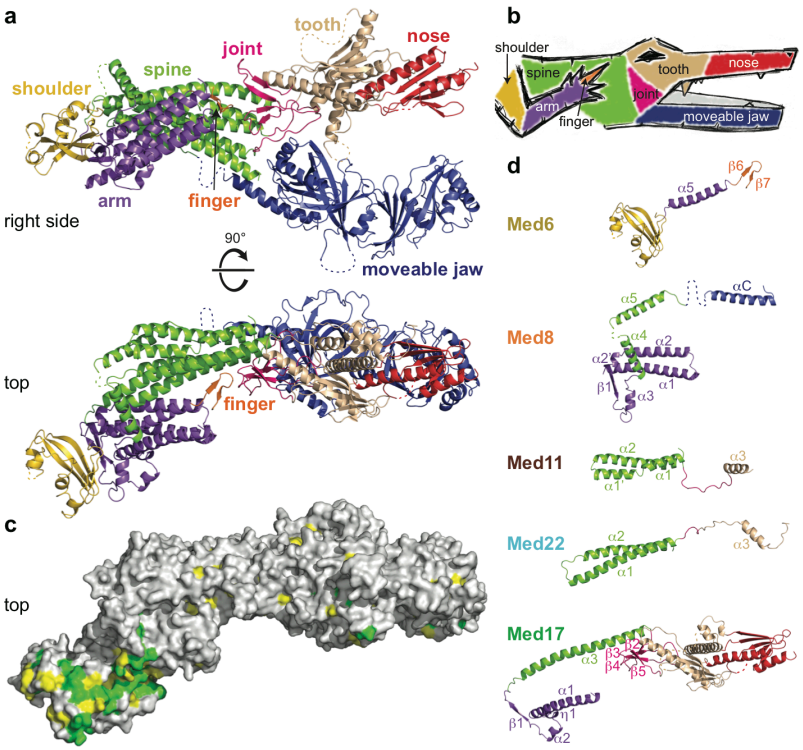


Figure 8 | Structural elements and surface conservation. a, Eight structural elements in the Mediator head module. Ribbon model of the Sp head module structure is coloured accordingly (see Fig. 9). **b,** The structure resembles the head of a crocodile with one limb. The eight structural elements received corresponding names, and previously introduced names for the three submodules neck, fixed jaw and moveable jaw are preserved. **c,** Surface conservation of the Sp Mediator head module. Residues that are invariant or conserved among seven yeast species (Supplementary Fig. 1) are in green or yellow, respectively. **d,** Ribbon representation of the five non-globular head subunits (top view) coloured according to structural elements as in a (see Fig. 9).

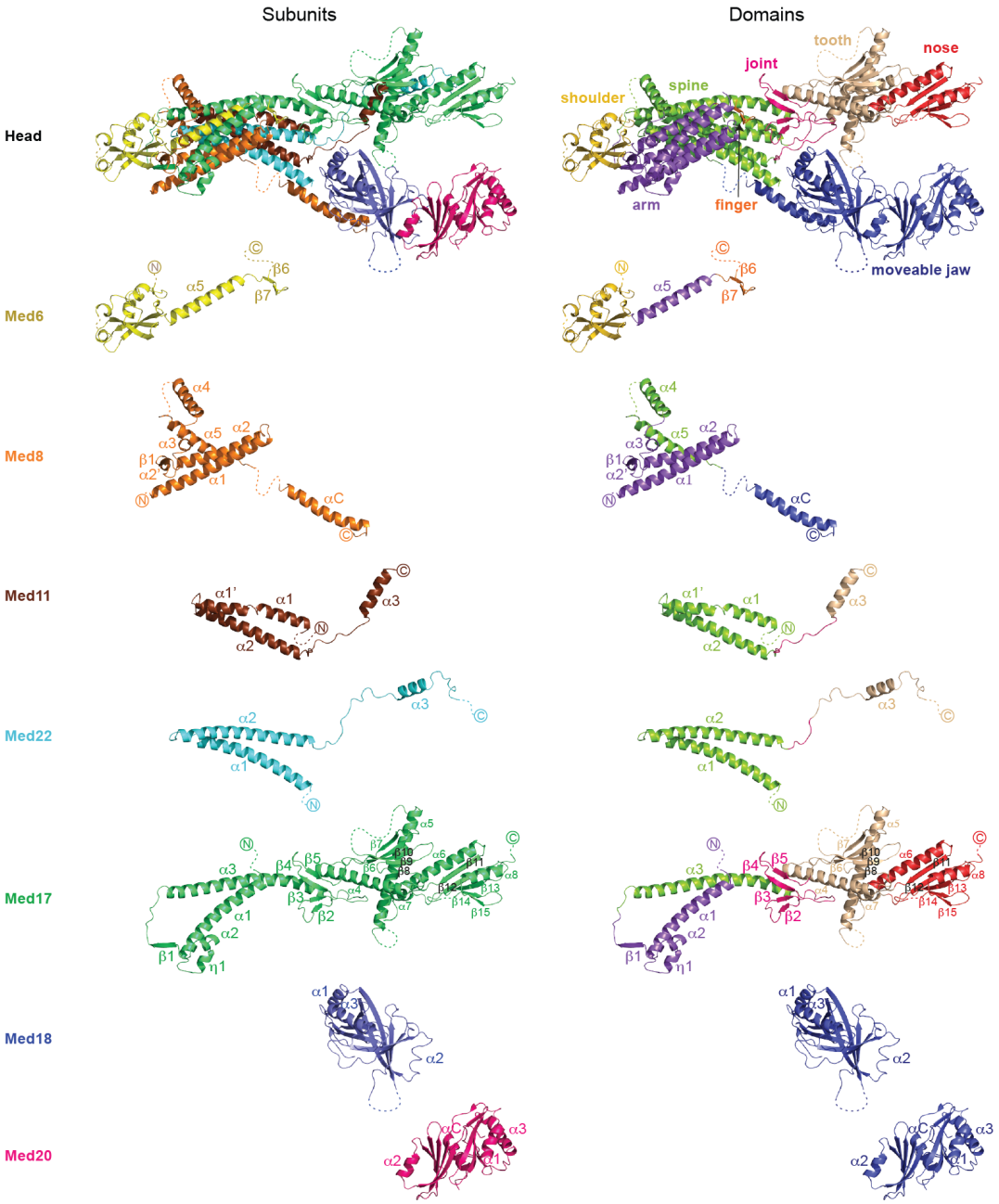


Figure 9 | Subunit structure within the Sp Mediator head module. The head module and its individual subunits are represented as ribbons (right side view), coloured by subunit (left) or according to structural elements (right).

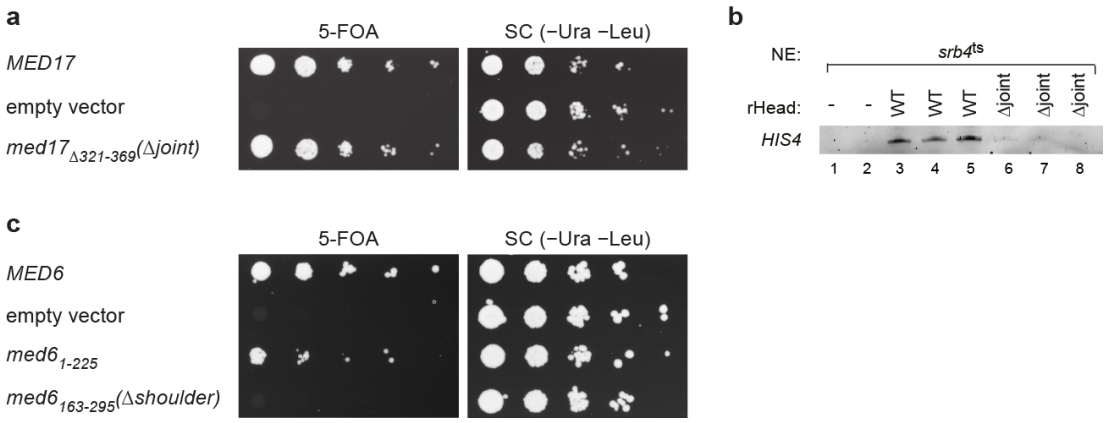


Figure 10 | Structural element deletions affect Mediator head function in vivo and in vitro. a, Yeast complementation assays were performed with wild-type MED17, an empty vector, or MED17 lacking the joint (Δjoint, Δ321-369) (see 2.2.3). **b,** In vitro transcription was performed with extract from the srb4-138 (srb4ts) strain in the absence of recombinant head module (rHead; lanes 1-2), in the presence of 5 pmol wild-type rHead (lanes 3-5), or 5 pmol rHead Δjoint (lanes 6-8) as described (see 2.2.3). **c,** Yeast complementation assays were performed with wild-type MED6, an empty vector, a MED6 variant corresponding to the construct used for the Sp head module crystallization (residues 1- 225), or MED6 lacking the shoulder (Δshoulder, residues 163-295) as described (see 2.2.3).

3.1.4 Revised *Sc* Mediator head module backbone model

We prepared a homology model for the *Sc* head module based on our structures of the *Sp* module and the *Sc* subcomplexes Med17C–Med11C–Med22C, Med11N–Med22N (ref.[100]) and Med8C–Med18–Med20 (ref.[70]). After minor adjustments, the model explained the published electron density and selenium positions[99]. Compared with the published model[99], four α-helices were swapped, a β-sheet was added in the tooth, and the amino acid register was adjusted for 46% of the residues outside the moveable jaw. Thus the structures of the *Sp* and *Sc* head modules are well conserved although the structured regions show only 15% sequence identity (Supplementary Fig. 1). *Sc*-specific differences are found in six helices; three are shorter (Med8 αC, Med17 α5, Med22 α1), one is longer (Med11 α1) and two are rotated (Med17 α7, Med22 α3). Because of poor electron density[99], the revised *Sc* model lacks the joint, part of the arm, the sheet in the shoulder, the finger and five protein linkers.

3.1.5 Structural comparisons reveal head module flexibility

Structural comparisons reveal flexibility within the head module. The position of the shoulder changes in different *Sp* module crystals (Fig. 11a), owing to a flexible connection to Med6 helix α5. The finger is mobile in the free Med6 structure. Flexibility of the moveable jaw is indicated by high B-factors (Fig. 11b) and electron

microscopic results[71]. Movement of the nose with respect to the tooth is revealed by comparison of the *Sc* model with the free fixed jaw structure (Fig. 11c). Comparison of the *Sp* structure with the *Sc* model indicates that the neck can rotate with respect to the jaws (Fig. 11e). The rotation may be accommodated by the joint, which uses conserved hydrophobic residues to anchor the tooth to the spine. Thus the head module contains at least four mobile peripheral elements, the shoulder, finger, moveable jaw and nose.

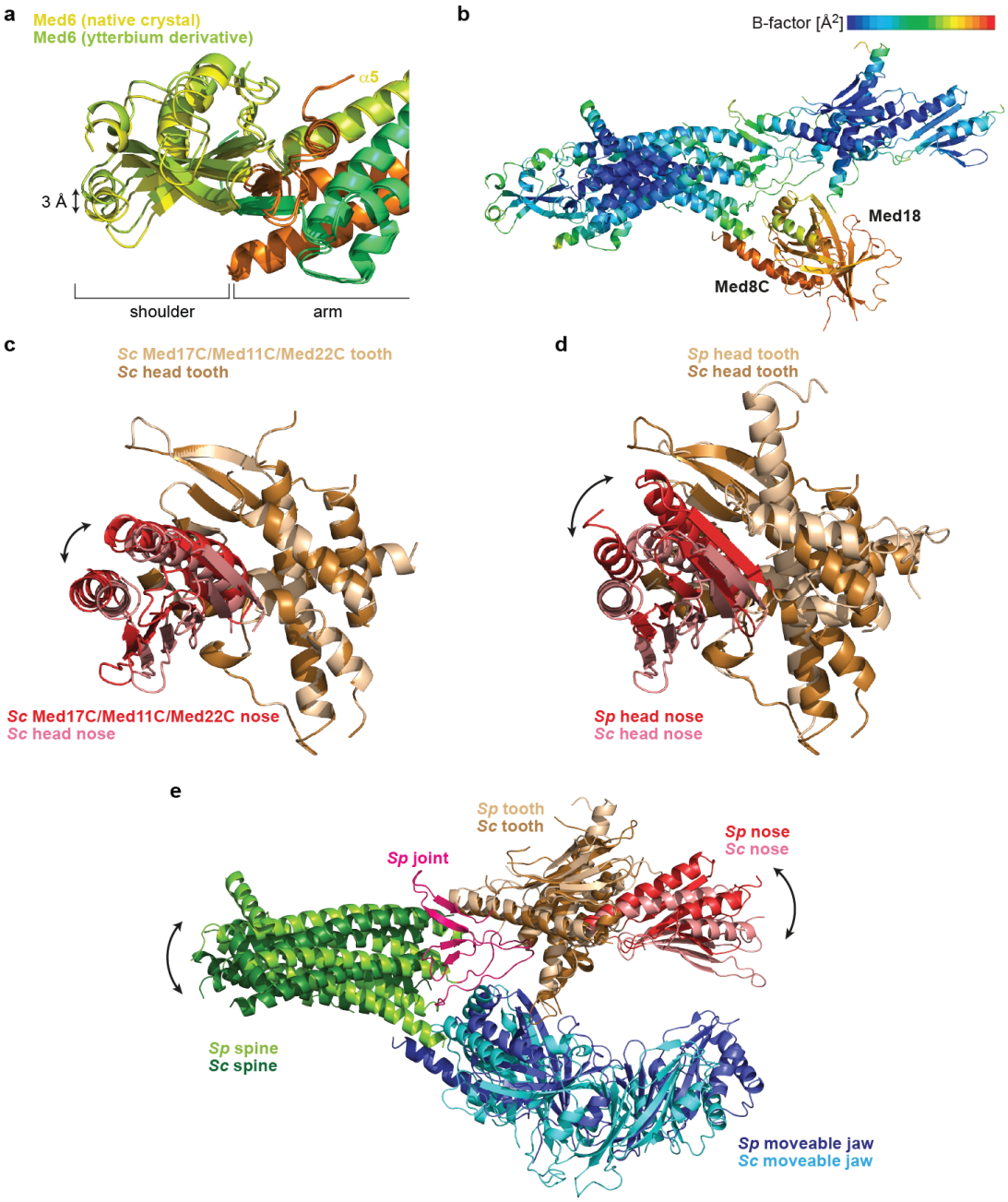


Figure 11 | Structural comparisons reveal head module flexibility. a, The shoulder adopts different orientations relative to the arm in different *Sp* head module crystals. **b,** B-factor distribution within the refined *Sp* head module structure reveals a high degree of flexibility of the moveable jaw. **c,** Comparison of the orientation of the nose relative to the

tooth between the new *Sc* head module backbone model and the free *Sc* Med17C/Med11C/Med22C structure. Tooth elements are superimposed. **d,** Comparison of the orientation of the nose relative to the tooth between the new *Sc* head module backbone model and the *Sp* head module structure. Tooth elements are superimposed. **e,** Superposition of *Sc* and *Sp* tooth elements reveals a different orientation of the spine relative to the jaws. Arm, shoulder and finger elements were omitted for clarity.

3.1.6 Head module integrity and interactions

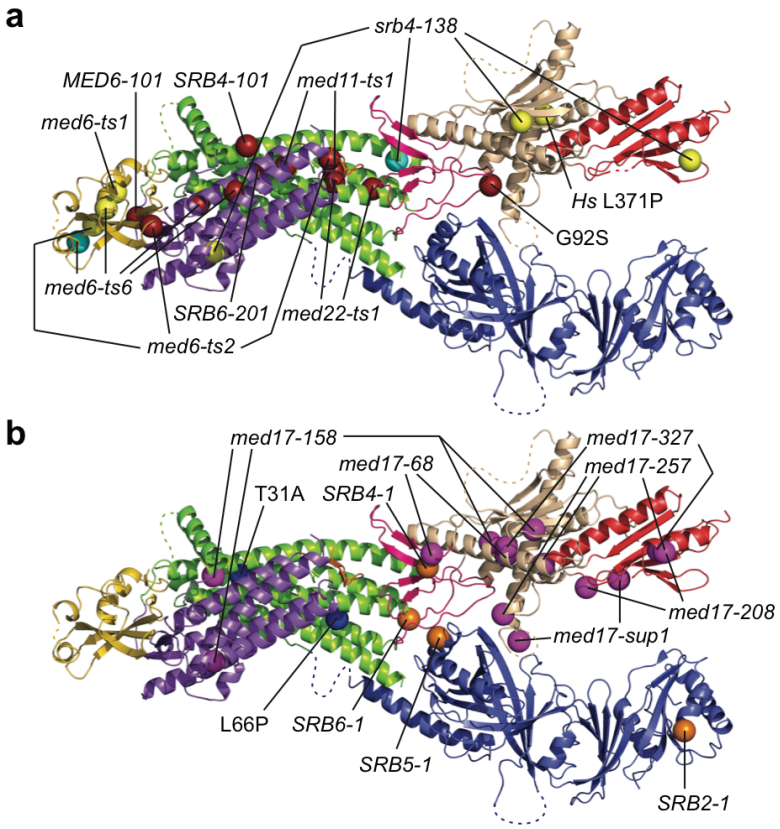


Figure 12 | Head module integrity and interactions. a, Location of sites of *Sc* mutations that may influence module stability. Mutations were mapped onto the *Sp* head module structure using alignments shown in Supplementary Fig. 1. Cα atoms of mutated residues are shown as spheres. Mutations that may affect intrasubunit stability, intersubunit stability or with uncertain affects are in yellow, dark red or cyan, respectively. **b,** Location of sites of *Sc* mutations that may influence interactions with the Pol II core, CTD and TFIIH. Mutations were mapped on the *Sp* head module using alignments shown in Supplementary Fig. 1, with Cα atoms represented as spheres. Mutations that suppress CTD truncation phenotypes, interact genetically with Rpb3 mutations or lead to a loss of interaction with TFIIH are shown as orange, magenta or blue spheres, respectively. A Med11 mutation in the spine, Thr31Ala, leads to reduced interaction with TFIIH[20].

To determine the phenotypes of known mutations, we mapped 47 sites of 55 *Sc* head module mutations onto the *Sp* structure (Fig. 12, Supplementary Fig. 1). Most temperature-sensitive mutations map to protein cores or domain interfaces and apparently influence module stability (Fig. 12a). These include mutant *srb4-138* that contains six Med17 mutations required for temperature sensitivity[54], including mutations Ser226Pro and Phe649Ser that apparently destabilize the arm and nose,

respectively. The human MED17 mutation Leu371Pro (residues 504 and 389 in *Sc* and *Sp*, respectively) is associated with infantile cerebral and cerebellar atrophy[168] and perturbs MED17 helix α5 in the tooth. The temperature-sensitive phenotype for Med11 mutations Glu17Lys/Leu24Lys (ref.[100]), Leu66Pro (refs[20,100]), and Glu92Ser (refs[20,100]) is also explained by fold destabilizations. The mutation *med6-ts1* and mutations in the *med6-ts6* allele[169] may destabilize the shoulder. Mutant *med6-ts2* contains six point mutations[169], of which Gln49Leu changes the conserved surface implicated in binding other Mediator parts, Phe125Tyr changes the shoulder–arm interface, and Phe194Leu may destabilize the finger–spine interface.

The structure also explains how mutant phenotypes are suppressed by secondary mutations. The *SRB4-101* mutation Glu286Lys rescues the *med6-ts2* phenotype[169], apparently by creating a new salt bridge between Med17 helix α3 and Med6 helix α5. The *MED6-101* mutation[170] may suppress the *srb4-138* phenotype because mutation Asp152Tyr between the shoulder and arm compensates for decreased stability of the neck. The *SRB6-201* (ref.[170]) mutation Asn59His may also suppress the *srb4-138* phenotype by stabilizing the neck.

Other genetic data implicate the fixed jaw in interactions with Pol II (Fig. 12b, Supplementary Fig. 1 and Supplementary Table 2). The head module interacts with Pol II near subunit Rpb3, and this contact is required for regulated transcription[171,172]. Mutations causing co-lethality with the Rpb3 mutation *rpb3-2* map to the fixed jaw (Fig. 12b). Mutants *med17-208* and *med17-257* both contain the mutation Glu669Asp that is located on the nose surface and may abolish the head–polymerase interaction that is weakened by Rpb3 mutation. Mutant *med17-sup1* rescues the phenotype of Rpb3 mutation Ala159Gly (ref.[171]) and contains two mutations in the fixed jaw (Fig. 12b). Consistent with an interaction between the head module and the Pol II region around Rpb3, a mutation in the adjacent polymerase dock domain has a cold-sensitive phenotype that is rescued by *SRB* mutations[48] in the joint and moveable jaw (*rpb1-14*). Mutants *med17-68*, *med17-158* and *med17-327* all contain mutations that are predicted to destabilize the fixed jaw.

Mediator also interacts with the carboxy-terminal domain (CTD)[173,174] that extends from the polymerase core near the dock domain. CTD truncation causes a cold-sensitive phenotype that is rescued by mutations[44,47,48] that map mainly to the

joint (Fig. 12). The mutation *SRB4-1* maps to the joint–tooth interface, whereas the mutation *SRB6-1* may stabilize the spine and adjacent joint. The mutation *SRB5-1* in Med18 helix α1 may strengthen joint interaction with the moveable jaw, whereas mutation *SRB2-1* alters the moveable jaw surface. These results suggest that the joint and moveable jaw are involved in CTD binding and/or that they are required for structural changes in Mediator that enable CTD binding. However, other Mediator regions probably contribute to interactions with the CTD because the head module is insufficient for CTD binding *in vitro*[97]. The flexibility and extended shape of the head module, which is 170 Å long, 95 Å high and 65 Å wide, may allow for several interactions not only with Pol II but also with transcription factors TFIIB[52,95], TFIIH[20,100] and the TATA box-binding protein[70,71].

3.2 Model of the Mediator middle module based on protein cross-linking

The work presented in this chapter was published.

L. Larivière, C. Plaschka, M. Seizl, E. V. Petrotchenko, L. Wenzeck, C. H. Borchers, P. Cramer. Model of the Mediator middle module based on protein cross-linking. *Nucleic Acids Research* (2013) 41, 9266–9273.

The essential core of Mediator consists of two conserved modules, the head and middle modules. Whereas the structure of the head module is known, the structure of the middle module is lacking. We therefore determined a model of the middle module based on protein cross-linking. Recombinant middle module was obtained by co-expression of its six subunits, Med4, Med7, Med9, Med10, Med21, and Med31 from a single plasmid (Fig 13a).

50 µg of middle module were incubated with isotope-labeled CBDPS[175]. CBDPS reacts with primary amines in lysine side chains and protein N-termini, and harbors a biotin moiety. Cross-linking efficiency was monitored by SDS-PAGE (Fig 13b). After protease digestion, cross-linked peptides were enriched by affinity chromatography using avidin. Peptides and their fragments were detected with high-resolution LC-MS (see 2.2.4). Measurements and subsequent analysis resulted in 55 mass spectra that matched cross-linked peptides[140]. An example is shown in Fig. 13c. These spectra correspond to 40 unique linkage pairs, among which 19 and 21 were intra- and inter-subunit cross-link pairs, respectively (Fig 14).

Eight cross-link pairs could be mapped on the two known crystal structures of middle module subcomplexes, seven on the Med7C/Med21 subcomplex, and one on the Med7N/Med31 subcomplex (Figs 14 and 15). The mapped cross-link pairs fell within the distance cut-off between Cα atoms of 26 Å.

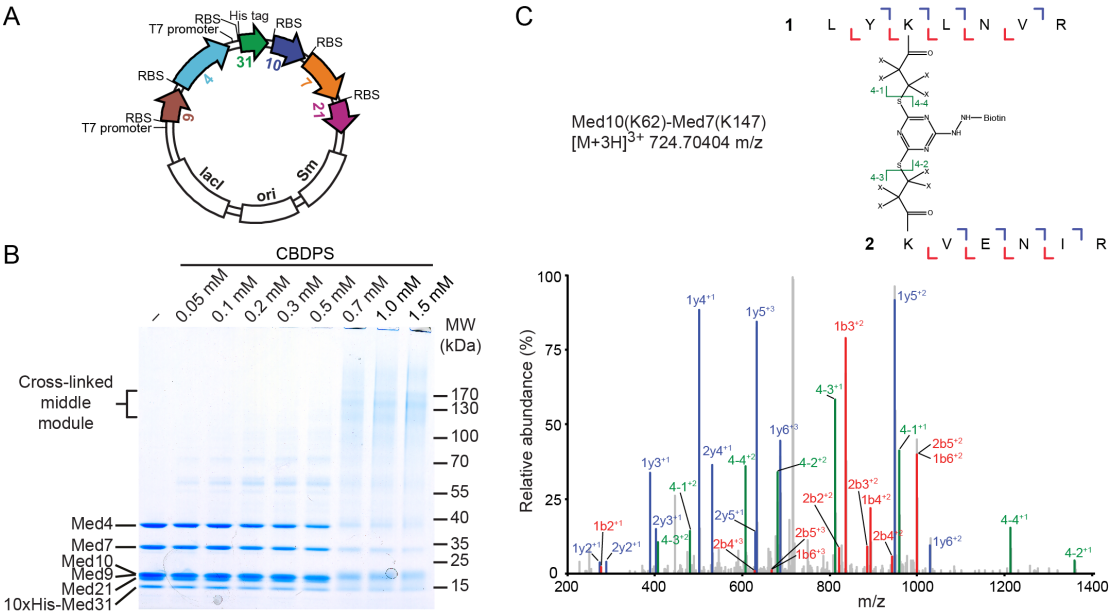


Figure 13 | Preparation and CX-MS analysis of the Mediator middle module. a, Schematic representation of the plasmid used for Mediator middle module recombinant expression. Coding sequences are colored according to a code used throughout (Med4, cyan; Med7, orange; Med9, brown; Med10, slate; Med21, magenta; Med31, green). Co-expression was driven from a single plasmid with two T7 promoters, one for bicistronic expression of Med9 and Med4, and one for tetracistronic expression of Med31, Med10, Med7 and Med21. His tag, deca-histidine tag; *ori*, origin of replication; *lacI*, gene encoding Lac repressor; RBS, ribosome binding site; *Sm*, streptomycin resistance gene. **b,** SDS-PAGE analysis of the middle module cross-linked with different concentrations of CBDPS. **c,** Fragmentation spectrum of a cross-linked peptide.

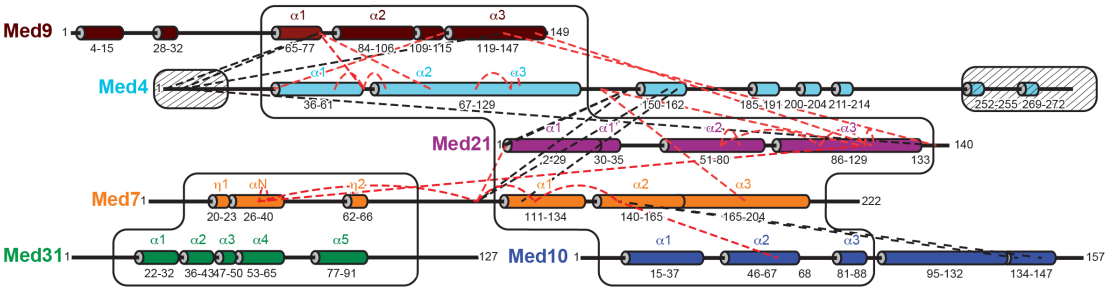


Figure 14 | Map of lysine-lysine cross-link pairs of the Mediator middle module. The primary structure of the six subunits is drawn schematically to scale. α-helices from crystal structures or from predictions are indicated as cylinders and colored as in Fig. 13. α-helices for which crystal structures are available or which could be modeled are labeled. White boxes enclose regions that could be modeled. Hatched boxes indicate regions that are dispensable for module assembly. Black and green dashed lines indicate cross-links that could be mapped onto the middle module model or not, respectively.

3.2.1 Model of the Med4/Med9 dimer

The structure of the Med4/Med9 heterodimer is unknown. Secondary structure prediction suggests that Med4/Med9 shares a common 4-helix bundle fold with head module subunits Med11/Med22[100] and middle module subunits Med7/Med21[101]. Based on this we generated a homology model of the Med4/Med9 heterodimer (Fig. 15) that explained all seven cross-links observed within the Med4/Med9 heterodimer (3 intra-subunit cross-links, 4 inter-subunit cross-links, Figs 14 and 15). The model was further supported by the resulting location of hydrophobic residues, which form the interface between the two subunits.

3.2.2 Model of the Med4/Med9-Med7/Med21 tetramer

To position the Med4/Med9 model relative to the Med7C/Med21 heterodimer crystal structure, we assumed that the two heterodimers interact as observed for symmetry-related Med7C/Med21 heterodimers in the Med7C/Med21 crystal structure[101]. In both crystal forms of Med7C/Med21, heterodimers pack against each other via the conserved ends of their protruding coiled-coils[101]. The resulting Med4/Med9–Med7C/Med21 model is strongly supported by four cross-links observed between the two heterodimers (Figs 14 and 15). The model is further supported by a clustering of hydrophobic residues in the heterodimer-heterodimer interface.

3.2.3 Positioning of Med10 and Med31

Tertiary structure prediction of the Med10 subunit using HHpred[138] suggests that Med10 is homologous to the N-terminal region of the head module subunit Med8, for which the structure is available[99,139,164]. We modeled the corresponding Med10 region, which forms three helices, α1, α2 and α3, corresponding to helices α1, α2 and α4 in Med8[139]. In the head module structure, these three helices interact with Med11N/Med22N. In particular, Med8 helix α2 interacts with Med22 helix α2.

Finally, two cross-links guided an approximate positioning of the external Med7N/Med31 subcomplex structure onto the obtained model. One cross-link was mapped between Med7N and a flexible linker in Med7C, and another one between Med7N and helix α3 in Med21. These restraints were not sufficient to ascribe a unique position to Med7N/Med31, but strongly suggested that this subcomplex is located near the Med7C/Med21 coiled-coil (Fig. 15). These results are consistent with

our previous notion that the Med7N/Med31 forms a surface subcomplex on Mediator[102].

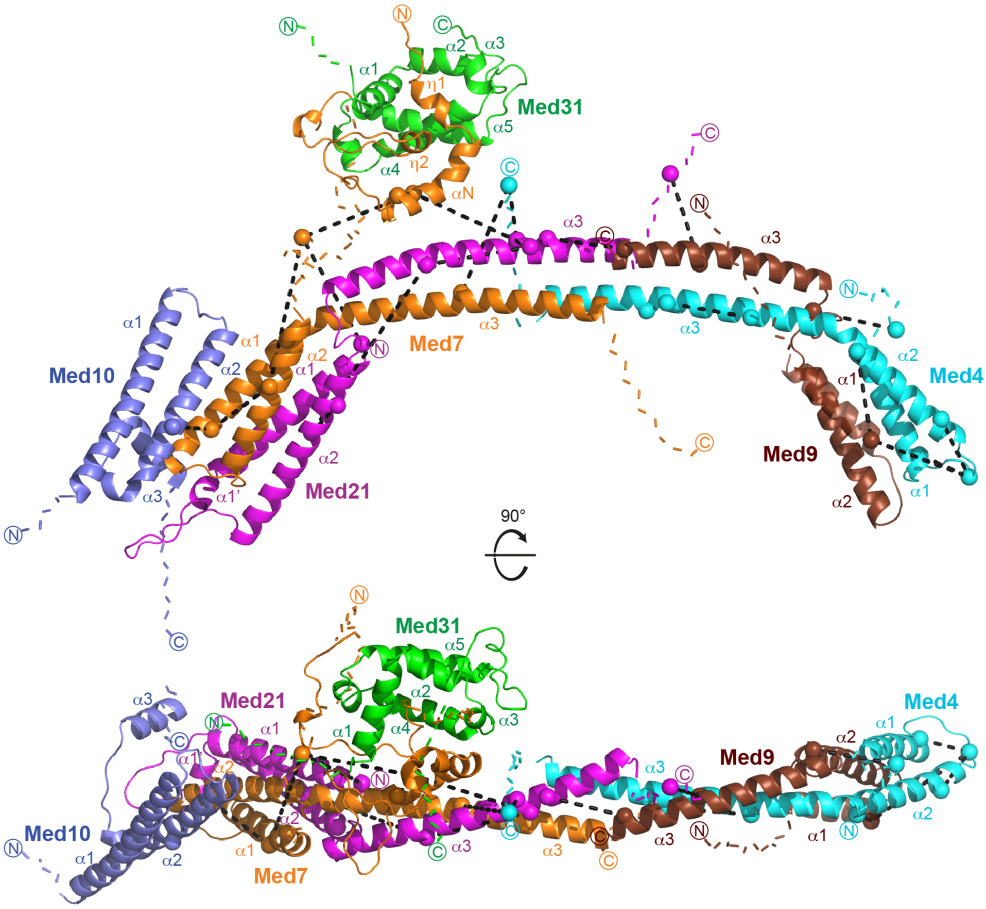


Figure 15 | Model of the Mediator middle module. Subunits are shown as ribbon and colored as in Fig. 13. Med9 helix α1 is represented as a semi-transparent ribbon because its exact position remains unclear. The Cα atoms of cross-linked residues are shown as spheres. Colored dashed lines indicate the first ten residues extending from both termini of the model. Black dashed lines indicate inter-subunit or intra-subunit cross-links. The two views are related by a 90° rotation around a horizontal axis.

3.2.4 Elongated architecture of the middle module

Our model is in agreement with reported data for the middle module. The length of the model is compatible with the previously determined hydrodynamic radius of the module[103]. The elongated shape of the model is reminiscent of the recently published electron microscopic density of the endogenous middle module[78]. All subunit interactions observed in the model have been reported using biochemical methods[78,96,101,103]. There are only two described subunit interactions that the model does not account for, between Med10 and subunits Med4 and Med31[78,103] but these interactions may be made by protein regions that could not be modeled and did not crosslink (Fig. 14).

The most notable feature of our model is its length of around 180 Å, which is consistent with the idea that Mediator envelopes a large surface area of Pol II. The model further emphasizes the flexible nature of the Mediator middle module, which explains why crystallization attempts were unsuccessful. The Med7C/Med21 subcomplex contains a flexible hinge between its four-helix bundle domain and its coiled-coil protrusion[101]. The Med4/Med9 heterodimer adopts a similar fold and likely harbors a similar hinge. This suggests that the terminal helix bundle domains in the 6-subunit middle module adopt their relative orientation and distance upon binding of Med1 and formation of higher-order complexes.

3.3 Architecture of the RNA polymerase II–Mediator core initiation complex

The work presented in this chapter was published.

C. Plaschka, L. Larivière, L. Wenzeck, M. Seizl, M. Hemann, D. Tegunov, E. V. Petrotchenko, C. H. Borchers, W. Baumeister, F. Herzog, E. Villa, P. Cramer. Architecture of the RNA polymerase II–Mediator core initiation complex. *Nature* (2015) 518, 376-380.

To reveal how Mediator regulates transcription initiation we prepared an active, homogeneous, recombinant 15-subunit core Mediator (cMed) that contains the head and middle modules. cMed corresponds to the endogenous core Mediator purified from yeast[67], and has a human counterpart that was reported while our manuscript was in revision[111]. cMed binds the core ITC (cITC), which lacks TFIIE and TFIIH and is stabilized by a short six-nucleotide RNA[16,40]. We resolved the architecture of the cITC–cMed complex by cryo-EM and protein crosslinking. Our results reveal the location of Mediator on Pol II and suggest mechanisms of transcription regulation.

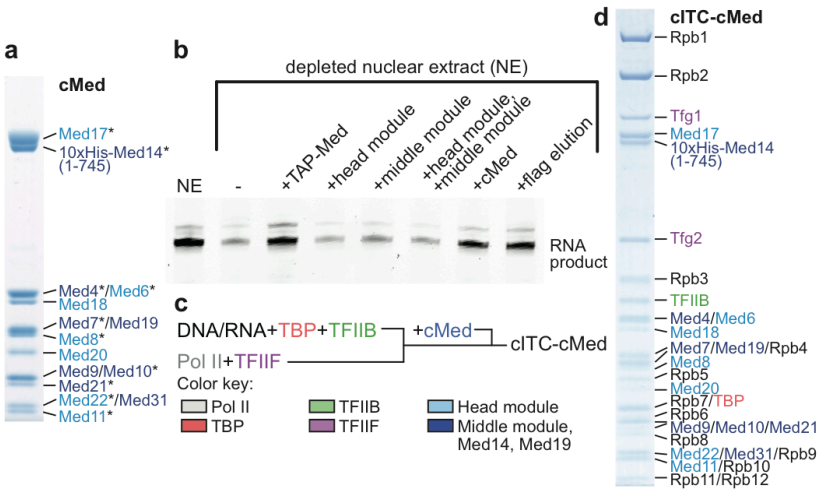


Figure 16 | Reconstitution of Pol II-Mediator complex cITC–cMed. a, SDS–PAGE analysis of recombinant 15-subunit cMed. Head and middle module subunits are labelled in blue and violet, respectively. Asterisks mark essential subunits. **b,** Recombinant cMed is functional in promoter-dependent transcription, whereas head and middle modules are not (see 2.2.5). As a positive control, Mediator-depleted nuclear extract was complemented with a Mediator fraction, either TAP–Mediator or Flag elution. The RNA product was visualized by primer extension ins triplicate experiments. **c,** Assembly of the cITC–cMed complex and protein colour key used throughout. **d,** SDS–PAGE analysis of the cITC–cMed complex after size exclusion chromatography.

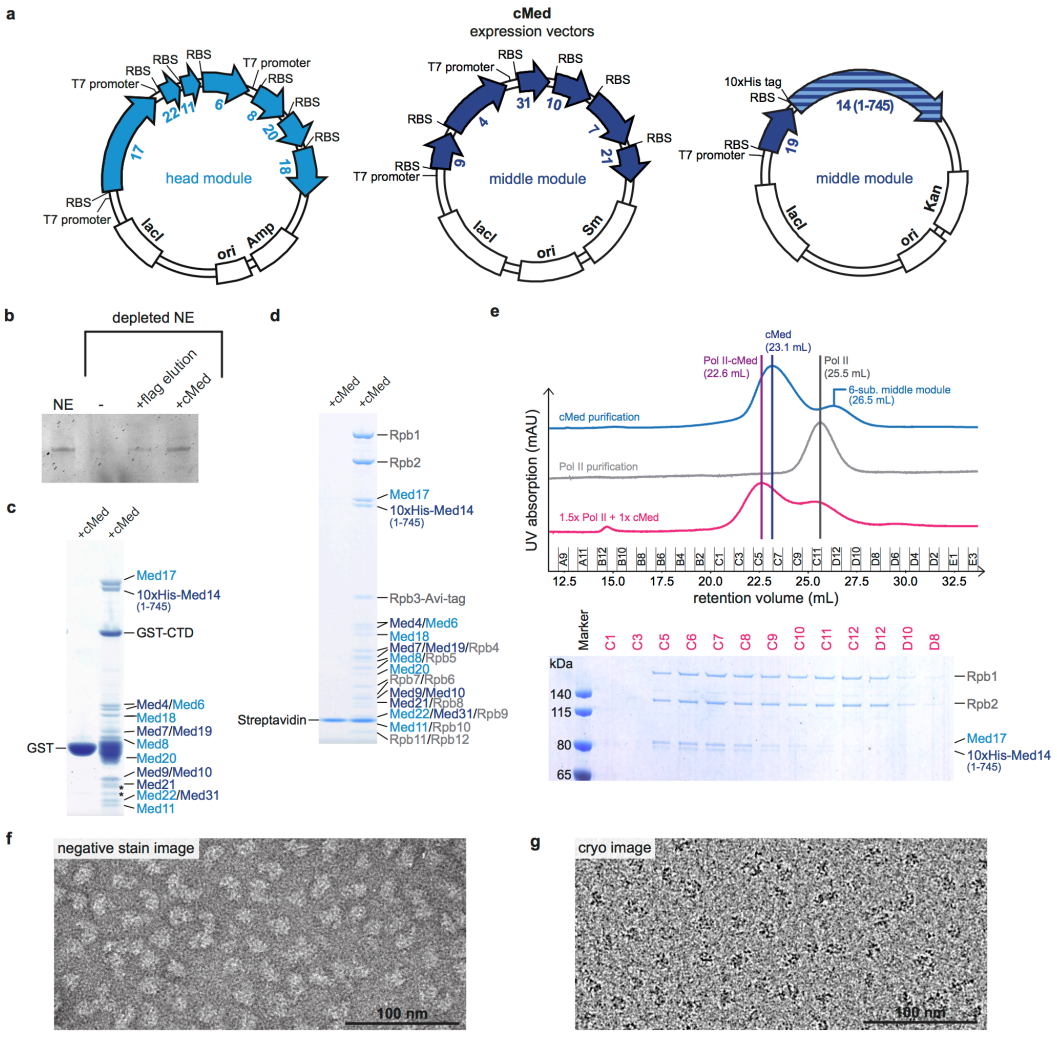


Figure 17 | cMed reconstitution, activities, and cITC binding. a, Schematic view of vectors used for co-expression of cMed. RBS, ribosome binding site; 10×His tag, 10×histidine tag; *ori*, origin of replication; *lacI*, gene encoding Lac repressor. *Sm*, *Amp* and *Kan* refer to streptomycin, ampicillin and kanamycin resistance genes, respectively. **b,** Recombinant cMed is active in activator-independent promoter-dependent transcription53. Compare with Fig. 16b. **c,** Recombinant cMed binds purified GST–CTD fusion protein, but not GST alone that was immobilized on glutathione resin. An asterisk marks two contaminant bands. **d,** Recombinant cMed binds Pol II in a pull-down experiment. Pol II was immobilized on streptavidin resin via a biotin tag on the Rpb3 subunit. **e,** Recombinant cMed binds purified Pol II during size exclusion chromatography. Chromatograms of cMed (blue), Pol II (grey) and Pol II–cMed complex (magenta) are shown (top). Peak fractions of the Pol II–cMed were analysed by SDS–PAGE (bottom). **f,** Negative-stain EM image of cITC–cMed complex. The scale bar is 100 nm. **g,** Cryo-EM image of cITC–cMed complex.

3.3.1 Core Mediator and electron microscopy

We previously prepared Mediator head and middle modules by co-expression of their subunits in bacteria[139,140]. The head module comprised subunits Med6, Med8, Med11, Med17, Med18, Med20, and Med22. The middle module contained Med4, Med7, Med9, Med10, Med21, and Med31. Recombinant head and middle modules did not bind each other, but their co-expression with subunits Med14 (residues 1–745) and Med19 enabled purification of a stable cMed complex (see 2.2.5, Figs 16 and 17a). Med14 enabled association of the two modules, consistent with a scaffold function[176]. cMed comprised 15 conserved[55] subunits (Fig. 16a), including all subunits essential for yeast viability[21]. cMed was functional in basal and activated promoter-dependent transcription, whereas the head and middle modules were not (Figs 16b and 17b).

Pure cMed bound to Pol II and the CTD (Fig. 17 c-e), and to the cITC (Fig. 16c, d). The cITC–cMed complex was stable during size exclusion chromatography and contained all 31 polypeptides (Fig. 16d, Supplementary Table 3). The sample comprised particles of the expected size when observed by EM in negative stain, and was used for cryo-EM data collection with a direct detection device (Fig. 17f, g; see 2.2.5). An unbiased reference structure was generated by tomographic reconstruction (see 2.2.5). Unsupervised 3D classification of particle images revealed that the cITC–cMed complex co-existed with free cITC and Pol II–DNA/RNA complex (see 2.2.5). We obtained high-resolution reconstructions for cITC–cMed, cITC, and Pol II–DNA/RNA complexes at resolutions of 9.7 Å, 7.8 Å, and 6.6 Å, respectively (Figs 18a, 19, 20, see 2.2.5).

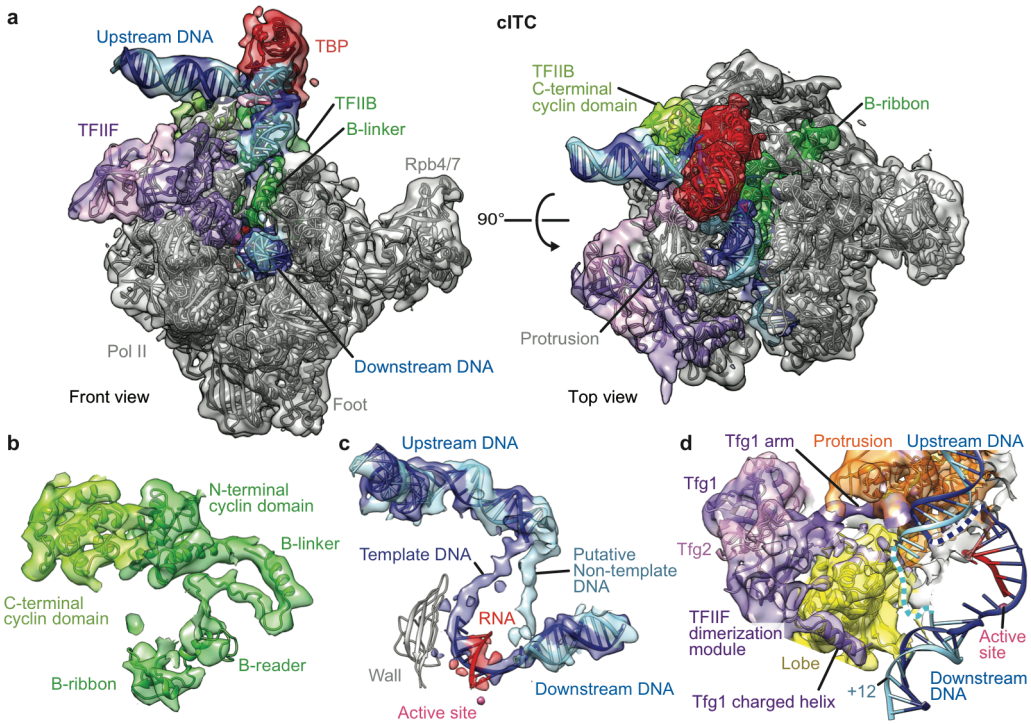


Figure 18 | EM structure of Pol II initiation complex cITC. a, EM reconstruction (4,439 particle images, 0.143 FSC = 7.8 Å resolution) with fitted structures as ribbon models. The views are previously defined front and top views of Pol II[155]. The density was filtered by local resolution. **b,** TFIIB X-ray structure[14] (ribbon) explains the EM density. View is from the side[155]. **c,** Course of the DNA template strand through the active centre and putative non-template strand density in the transcription bubble. View is as in b. **d,** TFIIF Tfg1 elements. The 'charged helix' protrudes towards downstream DNA, whereas the 'arm' tracks along the protrusion towards the upstream DNA bubble.

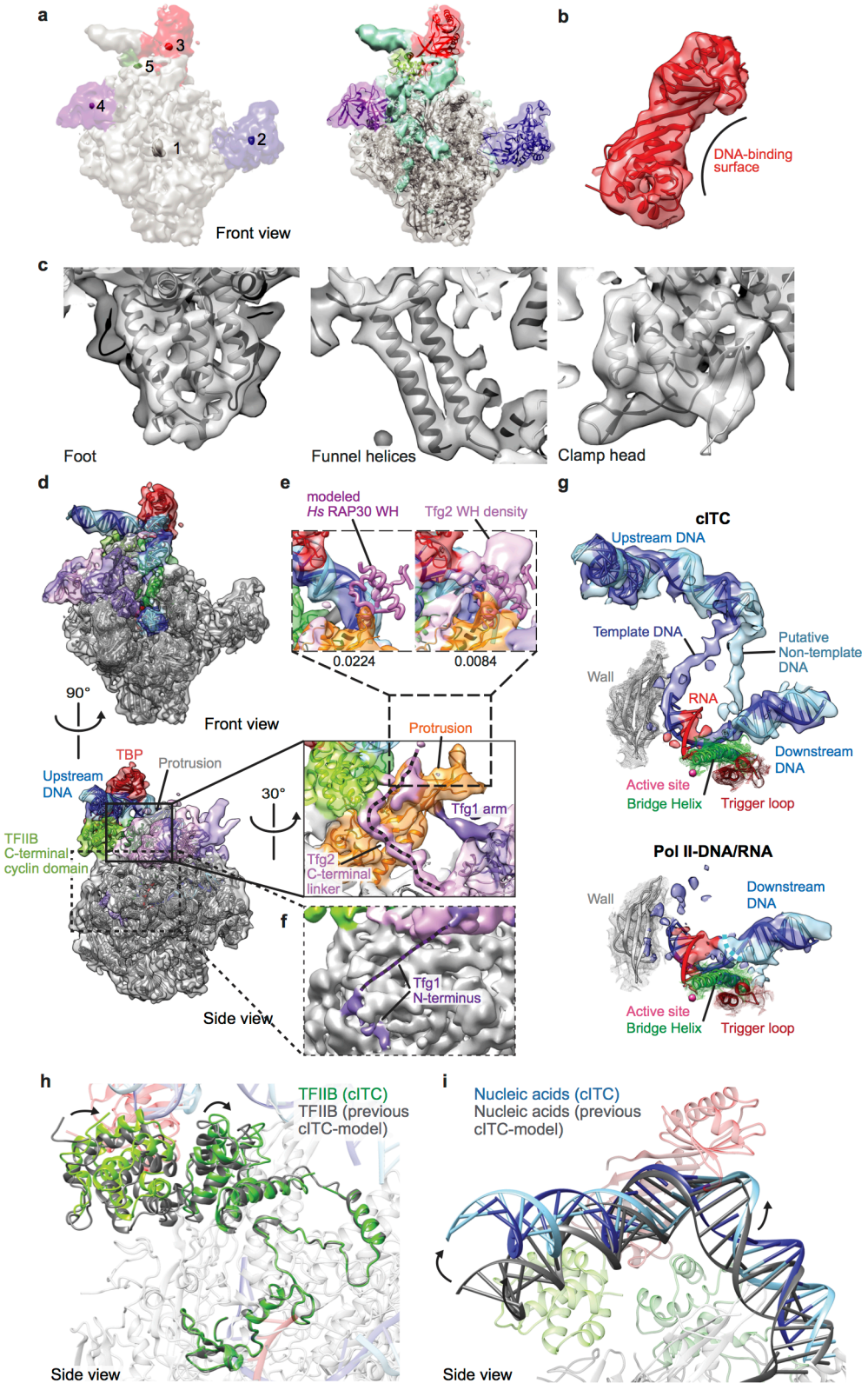


Figure 19 | Structural modelling of cITC into the cryo-EM density. a, Automatic fitting of structures into cITC cryo-EM density. The order of structure fitting and the corresponding translation correlation peaks are indicated (left). After fitting of all structures (right), the remaining density (cyan) was attributed mainly to DNA and TFIIF. **b,** Fit of TBP to the cITC EM density. **c,** Detailed views of Pol II domains foot, funnel, and clamp head in the cITC EM density. **d,** Two views of the cITC EM reconstruction corresponding to previously defined front and side views of Pol II[155]. The final cITC model is coloured (DNA template/non-template,

blue/cyan; RNA, red; Pol II, silver; TBP, red; TFIIB, green; TFIIF Tfg1/Tfg2, violet/magenta). Tfg1 and Tfg2 contain non-conserved insertions in the TFIIF dimerization module. **e,** The cITC map reveals density for the Tfg2 C-terminal linker and winged helix (WH) domain. The view corresponds to the side view in c, but is rotated around a vertical axis by 30°. The mobile Tfg2 WH domain is visible at lower density threshold. The homologous human RAP30 WH domain (Homo sapiens, Hs) was modelled on the basis of its position in the human initiation complex[39] and resides at a similar location. **f,** The cITC map reveals density that may correspond to the Tfg1 N-terminus, as suggested by protein–protein crosslinking[38,40]. **g,** Comparison of EM densities for promoter DNA in cITC (top) and in the Pol II–DNA/RNA complex (bottom). The Pol II elements active site (magenta), bridge helix (green), trigger loop (dark red), wall (grey), and template DNA (dark blue), non-template DNA (light blue) and RNA (red) are depicted. **h,** Minor repositioning of TFIIB cyclin domains compared to our previous model9. **i,** Minor rotation of the TBP–DNA–TFIIB complex on the Pol II wall in the cITC structure compared to the previous open complex model[40].

3.3.2 cITC structure and DNA stabilization

The cITC reconstruction enabled automated fitting of crystal structures of Pol II, TFIIB, TBP, and a TFIIF homology model (Figs 18a, b and 19a–c; see 2.2.5). The remaining EM density belonged to DNA and parts lacking from the TFIIF model (Figs 18c, d and 19d–g). The resulting cITC structure was consistent with a model derived biochemically[177], and confirmed and extended our previous model[14,16,40]. The TFIIB–TBP–DNA complex resided on the Pol II wall, but was slightly rotated towards the Pol II protrusion compared to an earlier model[40] (Fig. 19h, i). The previously unobserved[14] TFIIB C-terminal cyclin domain bound upstream DNA and the Pol II protrusion and subunit Rpb12 (Fig. 18).

The cITC reconstruction revealed the path of the DNA template through the Pol II cleft (Figs 18c and 19g). The template strand descended from the wall to the active site through the template tunnel[14] lined by TFIIB and the Pol II fork loop 1 and rudder. At the location expected for the non-template strand[178] we observed tubular densities (Fig. 18c). The TFIIF dimerization module resided on the Pol II lobe as in the Pol II–TFIIF complex[38,177]. The ‘charged helix’ of TFIIF subunit Tfg1 extended from the dimerization module to DNA at position +12 downstream of the transcription start site (Fig. 18d) and may stabilize loaded DNA[39]. The Tfg1 ‘arm’ extended along the Pol II protrusion towards the transcription bubble (Fig. 18d). The linker region of TFIIF subunit Tfg2 passed between the protrusion and TFIIB, and connected to the C-terminal winged helix (WH) domain located at the upstream bubble (Fig. 19 d, e). The cITC structure is highly conserved[39], and indicated that the Tfg2 linker and C-terminal winged helix domain stabilize TFIIB and open DNA[17].

3.3.3 clTC–cMed structure

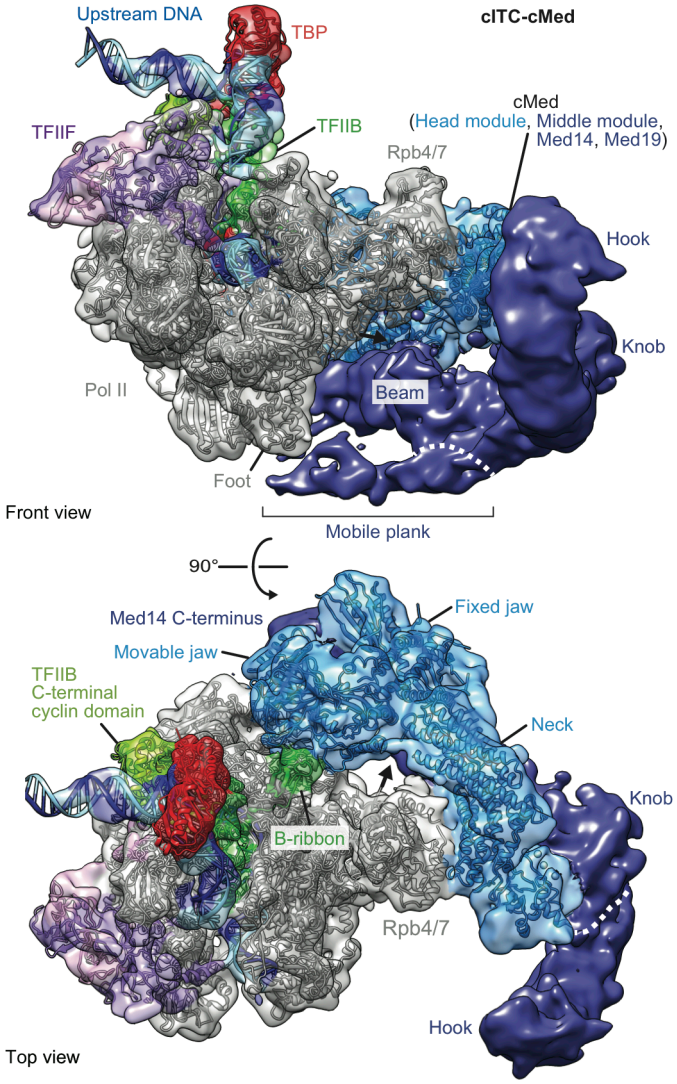


Figure 20 | EM structure of the clTC–cMed complex. EM reconstruction (3,267 particle images, 0.143 FSC = 9.7 Å resolution) viewed as in Fig. 18a with fitted structures (ribbon models). The density was filtered by local resolution. Mediator head and middle modules are coloured in blue and violet, respectively. We observe a single location of cMed on Pol II. Head module regions neck, fixed jaw and movable jaw are indicated. A black arrow marks the beginning of the linker to the Pol II CTD.

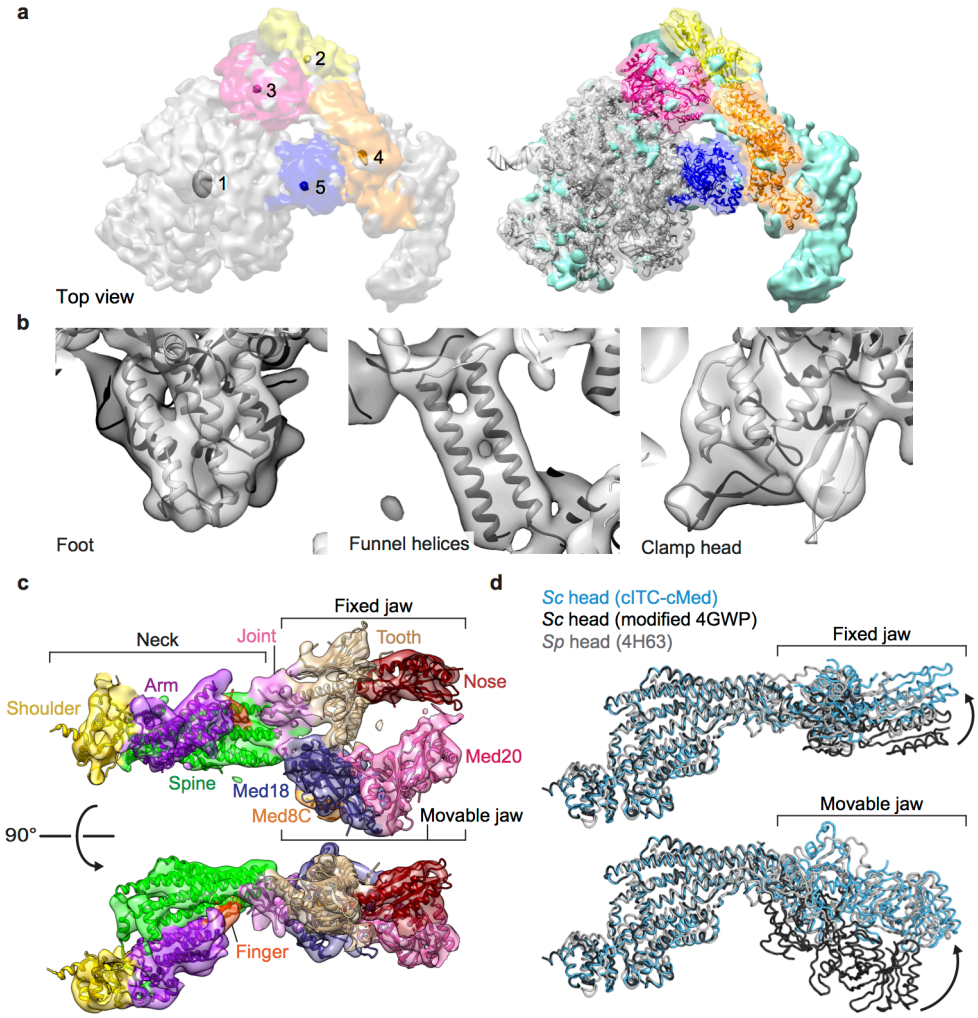


Figure 21 | Structural modelling of cITC–cMed into the cryo-EM density. a, Automatic fitting of atomic models into the cITC–cMed density. The order of model fitting and the corresponding translation correlation peaks are indicated (left). After fitting of all models (right), the remaining density (cyan) was attributed to the middle module and to some minor additional protein regions in the cITC and head module. **b,** Detailed views of Pol II domains foot, funnel, and clamp head in the cITC–cMed EM density. **c,** Fit of improved Mediator head crystallographic model to the corresponding cryo-EM density for cITC–cMed. The different head domains are depicted in different colours. Views are from the previously defined right side and top views of the head module[139]. **d,** The head module undergoes minor changes in conformation upon formation of the cITC–cMed complex. The EM fit with a modified model of the *S. cerevisiae* (*Sc*) Mediator head is compared to the crystal structures of head modules from *S. pombe* (*Sp*) (PDB code 4H63) and *Sc* (modified based on PDB code 4GWP). Previously defined top view of the head module[139].

The cITC model was placed unambiguously into the EM reconstruction of the cITC–cMed complex (Figs 20 and 21a, b). The remaining density belonged to cMed and was 240 Å long, 120 Å high, and 110 Å wide. We observed a single location of cMed on Pol II that differed from previously reported locations (Fig. 20, see 2.2.5). About half of the cMed density corresponded to the head module, and revealed the module's neck, fixed jaw, and movable jaw[139,164]. We fitted structures for the two jaws[70,139] and an improved model for the *S. cerevisiae* neck[164] based on the *Schizosaccharomyces*

pombe head module structure[139] (see 2.2.5, Fig. 21c). The relative position of the neck and jaws was similar in the free module from *S. pombe*[139] but differed in the *S. cerevisiae* structure[99,164], indicating mobility (Fig. 21d).

The remaining cMed density belonged to the middle module, Med14, and Med19. The density agreed with an extended shape of the middle module[140] and was divided into four regions that we call hook, knob, beam, and plank domains (Fig. 20). The middle module was connected to the shoulder and arm of the head module as proposed[139]. Additional contacts to the joint and fixed jaw domains in the head module revealed an intimate head–middle module interaction, in agreement with recent data[93,94].

3.3.4 Mediator topology

Since the middle module is structurally unresolved, we elucidated its subunit topology with protein crosslinking. For the cITC–cMed complex we obtained 706 high-confidence lysine–lysine crosslinks that provided 243 distance restraints (Figs 22a and 23a-d). Our structure explained 196 crosslinks within the cITC, 26 crosslinks within the head module, 26 crosslinks between the head module and the remainder of cMed, and 21 crosslinks between cITC and cMed. Crosslinking analysis of free cMed provided 52 additional crosslinks (Fig. 23a-d).

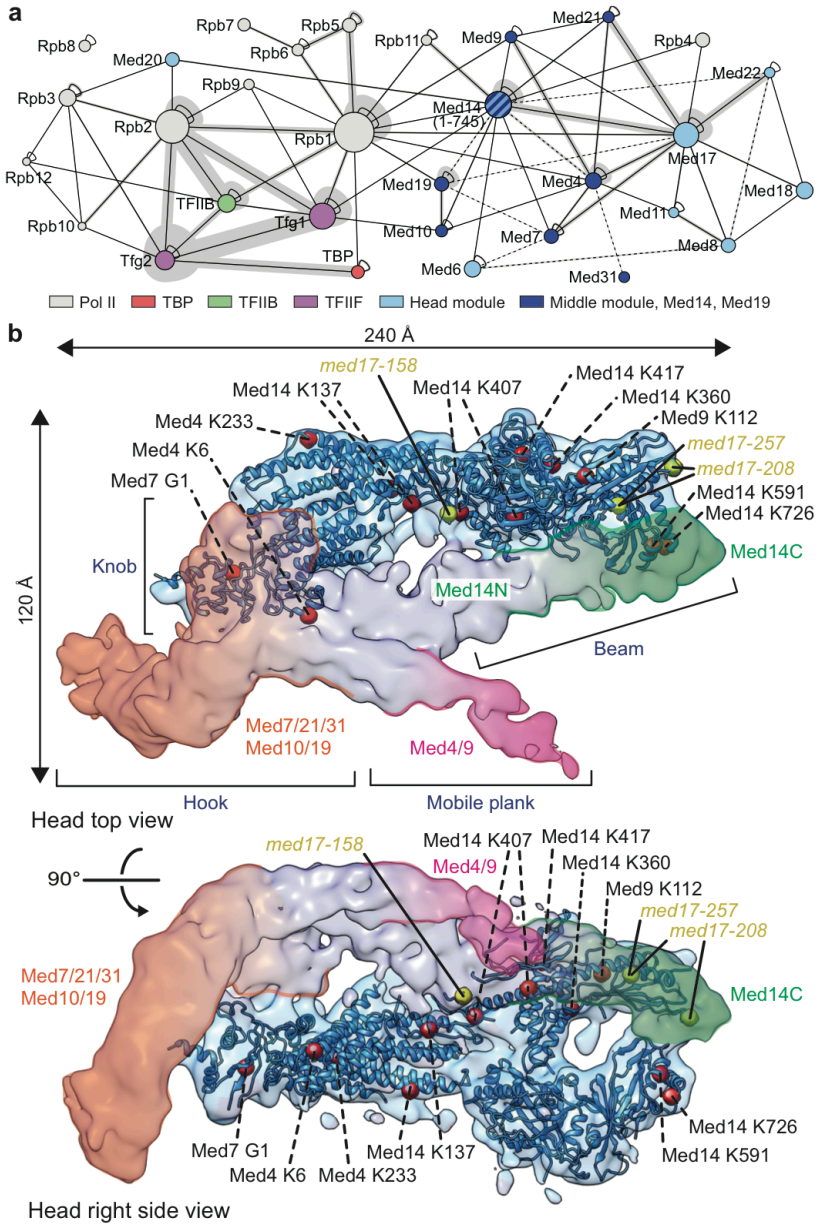


Figure 22 | Protein crosslinking and cMed architecture. a, Overview of protein–protein crosslinks visualized with xiNET. Solid and dashed lines indicate crosslinks for cITC–cMed and free cMed, respectively. Circle area correlates with protein mass. Line thickness correlates to the number of crosslinks. 13 crosslinks in flexible regions that exceeded the 27 ± 3 Å distance restraint between Ca atoms were excluded. **b,** Architecture of cMed and location of middle module subunits. Views are as defined for the head module[139]. Crosslinking residues were mapped onto the head module (red spheres) and labelled with crosslinking residues in the middle module. Mutants *med17-158*, *med17-208*, and *med17-257* (yellow spheres) that cause synthetic lethality with a Pol II mutation[171] map to the head–middle module interface. Putative locations of the Med14 N-terminal (residues 1–299) and C-terminal (residues 300–745) regions are indicated.

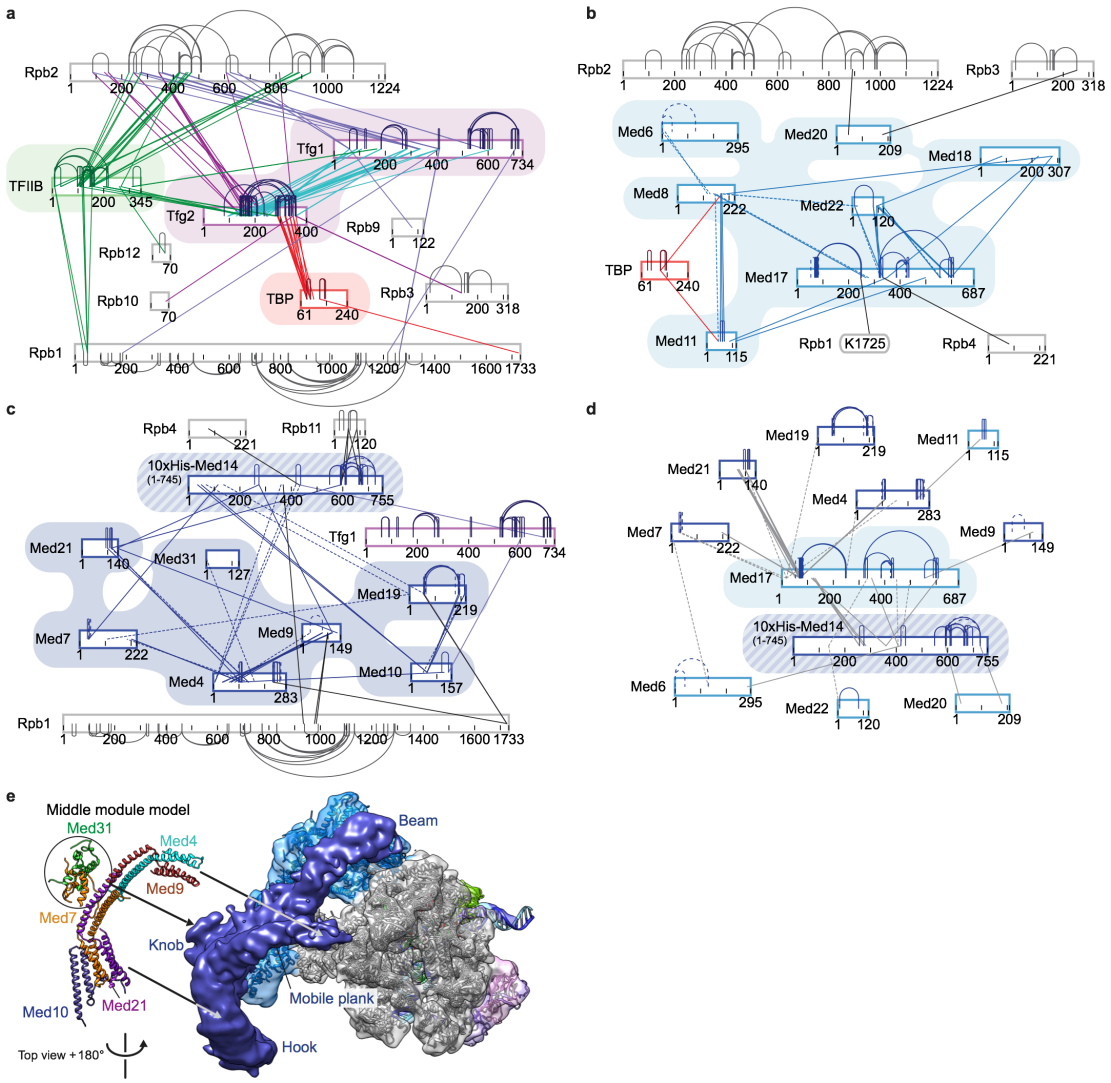


Figure 23 | Protein–protein crosslinking. a, Crosslinks between Pol II and TBP, TFIIB, and TFIIF confirm cITC architecture. Crosslinks were visualized using xiNET (Rappsilber laboratory). **b,** Crosslinks within the cMed head module and between the head module and Pol II. Solid lines indicate crosslinks derived from cITC–cMed data; dashed lines indicate additional crosslinks obtained only for free cMed. **c,** Crosslinks within the cMed middle module and between the middle module and Pol II. Crosslinks within the middle module agree well with the proposed middle module architecture[103,140]. **d,** Crosslinks between head and middle modules elucidate cMed architecture. e, Possible location of the previous model of the Mediator middle module (ribbons, left[140]) within the observed cMed density in the cITC–cMed reconstruction based on protein crosslinking. The cITC–cMed complex is viewed from the 'bottom', which corresponds to the top view rotated by 180° around a vertical axis.

The data indicated that the middle module hook and knob contained subunits Med7, Med10, Med19, Med21, and Med31, and that the globular Med31 domain[103] formed part of the knob (Figs 22b and 23e). The Med7–Med21 heterodimer resided near the arm and shoulder of the head module. The beam domain contained Med14. The Med14 N-terminal region (residues 1–299) resided between the middle module and the joint and spine domains of the head module. The Med14 C-terminal region (residues 300–745) crosslinked to the head module jaws. The plank contained the

Med4–Med9 heterodimer. This resulted in a revised cMed topology that is not consistent with the canonical Mediator architecture, but with the recent alternative Mediator topology[93,94] and the architecture of the human cMed counterpart[111].

3.3.5 Initiation complex stabilization

In the cITC–cMed structure, cMed binds around the Pol II Rpb4–Rpb7 stalk that is required for initiation[179] and forms three interfaces with the cITC (interfaces A to C, Fig. 24). In interface A, the movable jaw heterodimer Med18–Med20 binds the TFIIB B-ribbon domain, the Pol II dock, and the Rpb3–Rpb11 heterodimer (Fig. 24a, b). This contact can explain Mediator-facilitated recruitment of TFIIB[52] and a requirement of the B-ribbon for transcription[180], and is consistent with *SRB* mutations in the joint and movable jaw of the head module that rescue the phenotype of the *rpb1-14* mutant that alters the dock[48]. The location of Med20 between the Rpb3 zinc loop and Rpb3–Rpb11 dimerization domain explains why mutation of the Rpb3 zinc loop impairs Mediator binding[171] and is consistent with two Mediator-Pol II crosslinks (Figs 23b, 24b and 25a). The mutant *SRB2-1* comprises a point mutation in Med20 located in interface A (Fig. 25b), suggesting that weakening this interface suppresses a lethal CTD truncation[47]. The C-terminal region of Med14 may stabilize interface A because it contacts Med20 and crosslinks to Rpb11 (Figs 23c, d, 25a).

Interface B is formed between the conserved arm domain in the head module and the Rpb4–Rpb7 stalk. The interaction involves the arm domain helices α1 and α2 in Med8 and α2 in Med17[139] and α-helices H*, H3, and H6 in Rpb4[32] (Fig. 24c). The EM density suggests ordering of a mobile insertion between Rpb4 helices H1 and H* and explains a crosslink between H1 and the joint of the head module (Fig. 24c). Interface C is formed between the plank domain in the middle module and the Pol II foot and explains a crosslink from Med9 in the plank to the foot (Fig. 24d). Interface C is transient because the plank is detached from the foot in a subgroup of EM particles (Fig. 24d). The interfaces are hydrophilic and contain conserved residues, suggesting a conserved Pol II–Mediator interaction.

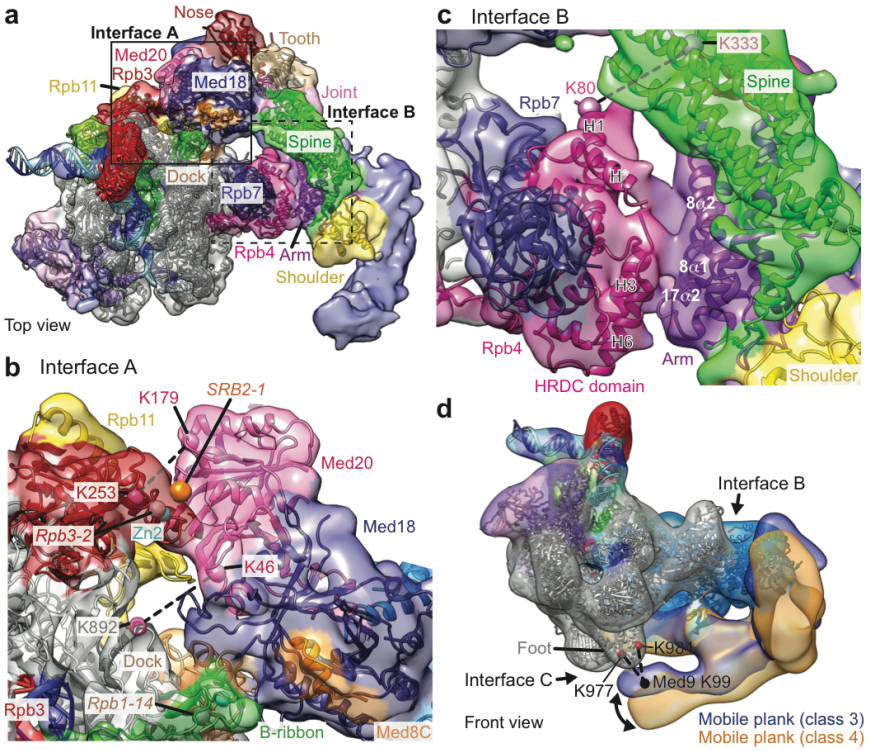


Figure 24 | cITC–cMed interfaces. a, View of cITC–cMed complex from the top. Previously defined domains of Pol II[155] and Mediator head module[139] are in different colours. **b,** Interface A. Spheres depict Cα atoms of residues mutated in yeast strains *SRB2-1*, *Rpb3-2*, and *Rpb1-14*. Crosslinked residues in Pol II and Med20 are indicated. **c,** Interface B. Secondary structure elements are labelled according to the previous nomenclature[32,139]. A crosslink from Rpb4 to the head module joint is indicated. **d,** Transient interface C. Classification of cITC–cMed particles reveals movement of the plank domain in the middle module. Unsharpened maps at 22 Å resolution for classes 3 and 4 are in violet and gold, respectively (Supplementary Fig. 3). Crosslinks from the Pol II foot to Med9 K99 (cMed) are indicated. View as in Fig. 20 (top).

Functional data are consistent with the observed cITC–cMed interfaces. *In vitro*, promoter recruitment of Pol II and TFIIB is lost in extracts from a *srb4-ts* strain, in which the head module dissociates from Mediator[54,181]. Addition of recombinant head module rescued this defect (Fig. 25b), consistent with previous results[97] and with our observation that most cITC contacts are formed by the head module. Variants of cMed with amino-terminal truncations of Med8 perturb interface B and were strongly impaired in Pol II binding (Fig. 25c). In contrast, cMed variants perturbing either interface A (lacking the movable jaw) or interface C (lacking Med4–Med9) still bound Pol II, showing that interface B is critical for cMed binding to Pol II. Consistent with these findings, perturbation of interface A *in vivo* by nuclear depletion of Med18 led to a mild decrease in messenger RNA synthesis rates[165], whereas perturbation of interface B by deletion of Rpb4 or dissociation of the head

module (interfaces A and B) resulted in a strong, global reduction of mRNA synthesis (Supplementary Fig. 2 and ref.[182]).

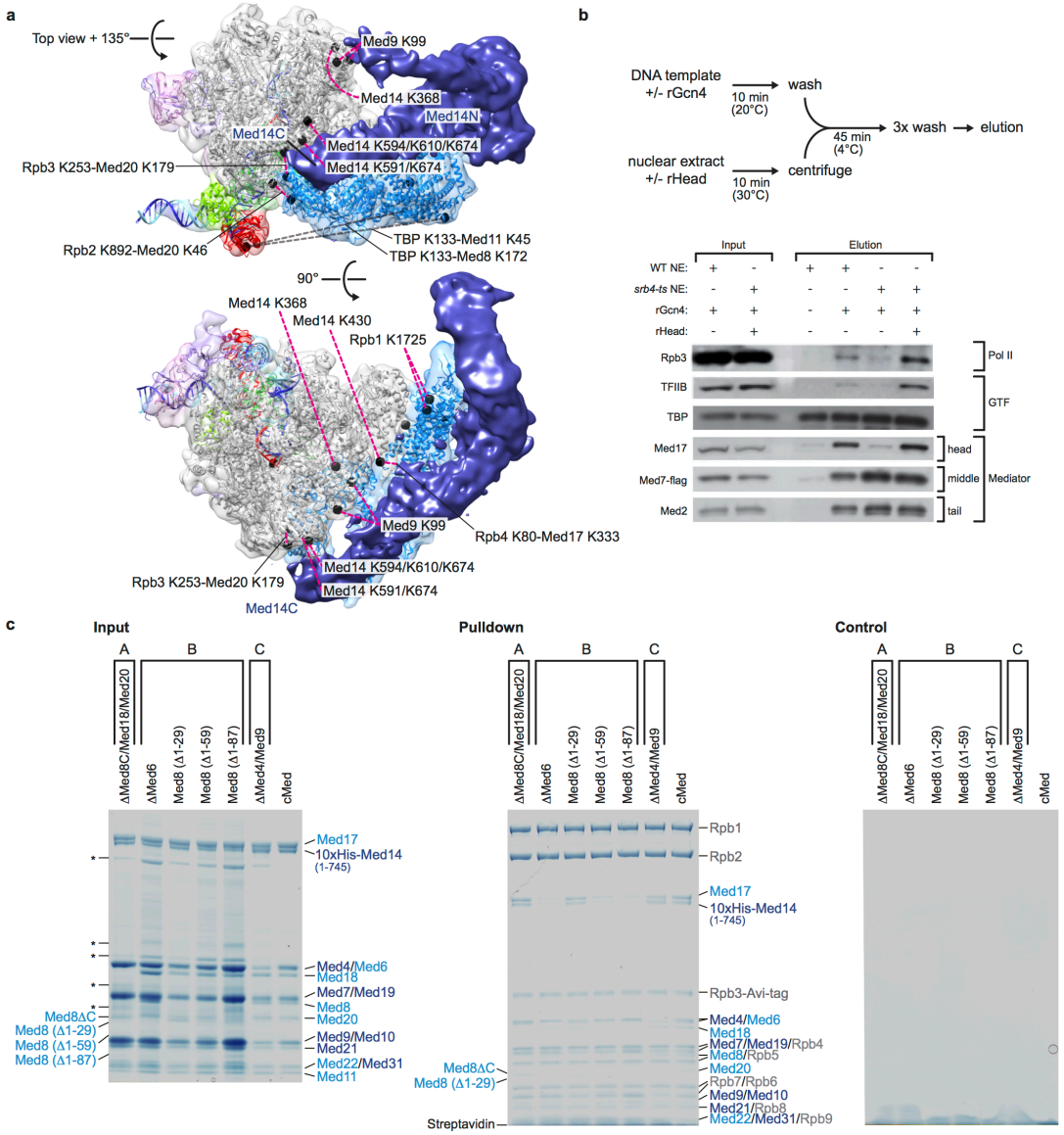


Figure 25 | Analysis of the cITC–cMed interface. a, Crosslinks between cITC and cMed were mapped on available models of cITC and Mediator head module. Two views of cITC–cMed are shown, a top view rotated by 135° around the horizontal axis and an additional 90° around the same axis. Mapped crosslinks exceeding the 30 Å distance restraint were coloured in grey. Crosslinks between lysine residues (black spheres) are labelled or connect (dashed magenta lines) to proximal cMed density, where plausible. Med14 N- and C-terminal regions are indicated. **b,** Immobilized template assay using nuclear extract of *srb4-ts* and wild-type yeast strains demonstrates that recruitment of Pol II and TFIIB depends on the Mediator head module. Schematic protocol for the immobilized template (ITA) assay[100] (top). The loss of Pol II, TFIIB and Mediator head module from promoter DNA by heat inactivation of the *srb4-ts* nuclear extract is rescued by addition of recombinant head (rHead) module[100] (bottom). **c,** Pull-down assays with recombinant cMed variants carrying mutations in interfaces A, B, and C reveal that interface B is essential for Pol II binding, whereas interfaces A and C are not. For definition of interfaces see Fig. 24. The assay measures retention of cMed variants on Pol II coupled to beads (see Fig. 17d). Interface A was perturbed by deletion of the movable jaw, Med8C(residues 1–189)–Med18–Med20. Interface B was perturbed by removal of parts of the arm domain, either by N-terminal truncation of Med8 (Δ1–29, deleting only the flexible N-terminal tail; Δ1–59, deleting the tail and helix α1; Δ1–87,

deleting the tail and helices α1 and α2), or by removal of Med6, which contributes a helix to the arm domain (ΔMed6). Interface C was perturbed by deletion of Med4–Med9. Shown are SDS–PAGE analyses (Coomassie staining). On the left, 2 µg of purified input cMed variants were analysed and the integrity of the complexes confirmed. The identity of the cMed variant-specific bands was confirmed by mass spectrometry. Some minor contaminant bands are marked with asterisks. On the middle gel, the bead elutions after the pull-down assays are shown. Wild-type cMed bound to Pol II-coupled beads, providing a positive control (rightmost lane). cMed variants were retained by Pol II-coupled beads to various degrees. Whereas cMed variants with perturbations in interfaces A and C bound to Pol II, several variants with perturbations in interface B were impaired for Pol II binding. On the right, the bead controls are shown. None of the analysed cMed variants bound to beads only, providing a negative control.

3.3.6 Pol II activation and phosphorylation

Transcription activators located on upstream DNA can recruit Mediator by binding its tail module. Consistent with this, the tail module is predicted to reside near upstream DNA as shown by superposition of the EM density of free Mediator[94] onto our structure (Fig. 27a). The superposition further argues that Mediator does not undergo major structural changes upon Pol II binding. Transcription activation may however include conformational control of Pol II. We propose that cMed stabilizes an active arrangement[183] of the two major polymerase modules[155] 'core' and 'shelf', because it bridges between these modules via interfaces A and C, respectively. Mediator may also communicate with TFIIS allosterically (Fig. 27b), because TFIIS reorients the shelf module[22] that contains the foot, and cooperates with Mediator during initial transcription[25]. Mediator may also facilitate DNA opening by positioning the clamp and the Rpb4–Rpb7 stalk (Fig. 27c).

We considered how Mediator may stimulate CTD phosphorylation at serine-5 residues by TFIIH[68]. Comparison of our structure with the EM reconstruction of the human initiation complex that includes TFIIE and the core of TFIIH[39] suggested that TFIIE reaches the head module arm and shoulder, and that the TFIIH core may bind the middle module (Figs 26a and 27c, d). This may explain how Mediator stabilizes TFIIE and TFIIH in the initiation complex[20]. The TFIIH kinase subcomplex, which was modelled onto the human complex[39], projects into a 'cradle' formed between Pol II and the Mediator arm and hook domains (Figs 26a and 27b, c). A part of the CTD was apparently ordered inside the cradle on the surface of the arm domain at a location similar to that of a CTD peptide observed in head module crystals[164] (Figs 26b and 27e). Thus Mediator may stimulate CTD phosphorylation by orienting the mobile TFIIH kinase and CTD within the cradle, consistent with a previous

suggestion[68]. Indeed, cMed could stimulate CTD phosphorylation by TFIIH *in vitro* (Fig. 27e), although less strongly than endogenous Mediator (not shown).

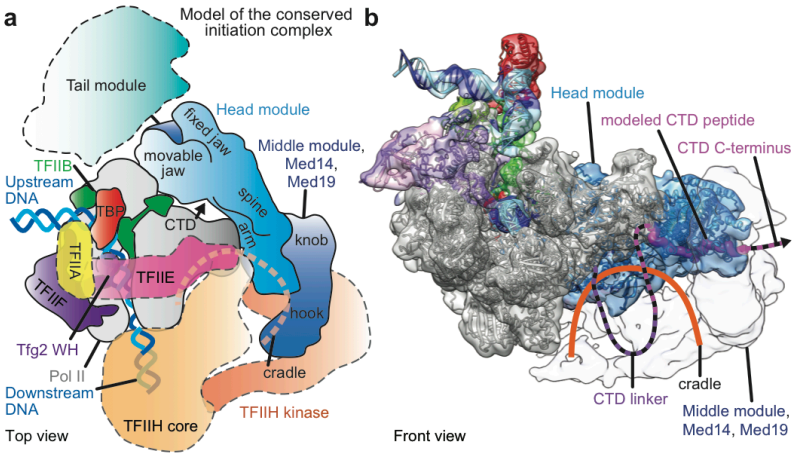


Figure 26 | Initiation complex model and CTD cradle. a, Schematic of the Pol II–Mediator initiation complex viewed from the top[155]. Locations of TFIIA, TFIIE, and TFIIH were inferred from the EM reconstruction of the human initiation complex[39] (Fig. 27d). The location of the Mediator tail module was inferred from the revised Mediator architecture[94] (Fig. 27a). **b,** View from the front into the 'cradle' of the cITC–cMed complex. A modelled CTD peptide (magenta) was positioned according to the head module–CTD structure[164] (PDB code 4GWQ). The middle module is transparent.

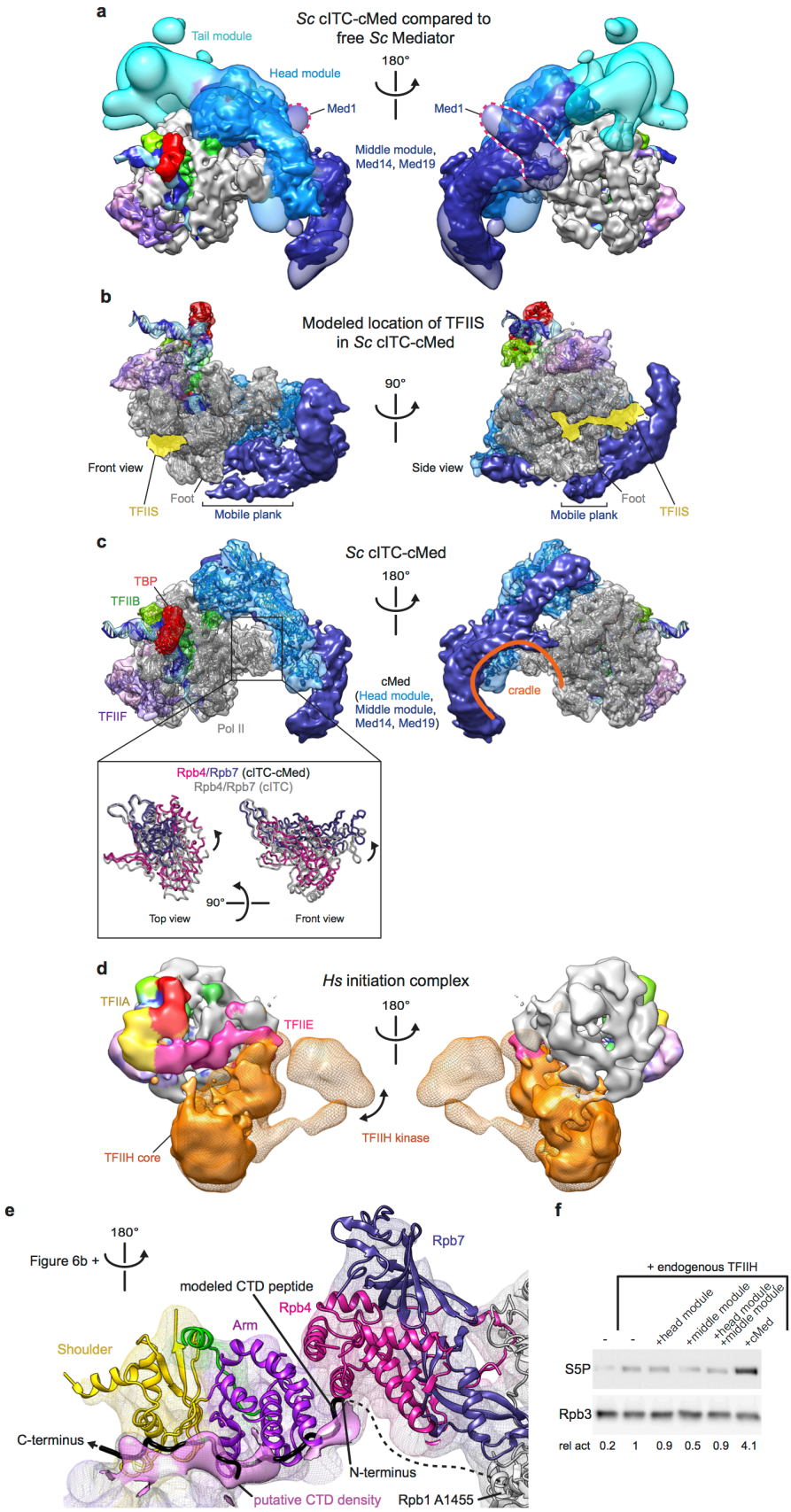


Figure 27 | Inferred locations of the Mediator tail module, TFIIS, the general factors TFIIA, TFIIE, and TFIIH, and the CTD, and comparison with the human initiation complex. a, Superposition of revised free yeast Mediator EM reconstruction[94] (transparent lower-resolution surface, coloured in shades of blue) onto clTC–cMed (coloured as in Fig. 20). This reveals a high degree of similarity in the cMed region, and suggests locations of the

tail module and subunit Med1 that are not present in cMed. **b,** Modelled location of TFIIS (yellow) within the cITC–cMed complex based on the Pol II–TFIIS complex (PDB code 1Y1V)[22]. **c,** Movement of the Rpb4–Rpb7 stalk upon cMed binding to the cITC and location of the 'cradle'. Two views of the *S. cerevisiae* (*Sc*) cITC–cMed complex related by a 180 degree rotation around a vertical axis. The left view corresponds to the previously defined top view of Pol II[155]. The box contains a zoom-in view of the Rpb4–Rpb7 complex revealing its movement upon cMed binding. The right view reveals the location of the 'cradle' on the 'bottom' of Pol II. The density is transparent, with the final cITC-head module model underneath and coloured (DNA template/non-template, blue/cyan; RNA, red; Pol II, silver; TBP, red; TFIIB, green; TFIIF Tfg1/Tfg2, violet/magenta; head module, blue; middle module, violet). **d,** Human (*Homo sapiens*, *Hs*) initiation complex EM reconstruction[39] viewed as in a. The complex reveals equivalent locations for TFIIB, TBP, TFIIF, and DNA, and additionally contains TFIIA (yellow), TFIIE (magenta), and the core subcomplex of TFIIH (solid orange surface) and a modelled TFIIH kinase subcomplex (orange mesh). **e,** Putative density for the CTD (magenta) near a modelled peptide (black) positioned according to the head module-CTD complex co-crystal structure (PDB code 4GWQ). The putative density superposes moderately well with the modelled CTD peptide and suggests the CTD may adopt a different conformation in cITC–cMed with respect to the binary head module–CTD complex. This density cannot be assigned to the CTD with certainty because it may also stem from unresolved protein regions in Mediator. Coloured as in Fig. 24a. The region C-terminal of the CTD remained flexible and crosslinked to distant proteins (not shown, see Fig. 23). **f,** Recombinant cMed stimulates TFIIH kinase activity, whereas the free head and middle modules do not. The relative activity (rel act) of Pol II phosphorylation at CTD residue serine-5 (S5P) was determined with respect to background (Rpb3).

4. Conclusions and Outlook

This thesis covers structural and functional studies of the Mediator complex and its role in Pol II transcription regulation. We determined the 3.4 Å crystal structure of the 6-subunit *Sp* Mediator head module that constitutes one half of the conserved and essential core Mediator (Fig. 6). New structural and mobile elements were identified that confirm and extend previous mutational and structural analysis of Mediator sub-complexes[21] and revised models of the *Sc* head module at 4.2-4.3 Å resolution[99,164]. Protein crosslinking of the recombinant 6-subunit Mediator middle module led to a first architectural model of this second half of core Mediator (Fig. 15). Finally, we prepared a recombinant 15-subunit core Mediator, which contained the complete 7-subunit head module, the 6-subunit middle module, and subunits Med19 and Med14. Med14 bridges the two modules, whereas Med19 increased yield. Core Mediator was active in basal and activated transcription, and bound to a Pol II core initially transcribing complex (cITC) that further contained the general factors TBP, TFIIB, TFIIF, and promoter DNA.

Cryo-EM of the cITC elucidated the evolutionarily conserved core initiation complex and revealed functions of general transcription factors (Fig. 18). We determined the complete path of the DNA template strand and TFIIF elements that reach into the Pol II cleft, likely to stabilize DNA. Cryo-EM and protein crosslinking of the cITC-core Mediator complex agreed with a revised Mediator architecture[93,94] and revealed a single position of cMed on Pol II that is different from previous studies[21,94] (Fig. 20). The cITC-Mediator core architecture suggests how Mediator interacts with transcription activators bound to upstream DNA, stabilizes the initiation complex, stimulates Pol II activation and CTD phosphorylation. In particular, the location of the tail module was inferred by crosslinking (Figs 23 and 24a) and EM of free *Sc* Mediator[93,94] that suggests close spatial proximity of the Mediator tail module to the average transcription factor binding site 250 basepairs[184] upstream of the transcription start site. Our results reveal that only minor adjustments in Mediator structure are required to bind Pol II (Fig. 27a), in contrast to earlier findings[21,94]. This indicates that the interaction of Mediator with the initiation complex relies on co-operative binding of Pol II and general factors[60] rather than Mediator unfolding[185].

Mediator control may carry an allosteric component. Mediator and TFIIS cooperate during transcription initation[25], and may communicate through the Pol II foot and the Mediator plank domain that contains Med1, Med4, and Med9. Core Mediator did not contain Med1, and may be added in an ongoing effort to prepare a complete recombinant Mediator complex. Mediator may influence Pol II conformation since it bridges the Pol II 'shelf' and 'core' modules and affects the Pol II stalk position (Fig. 27c). Thereby Mediator may assist in stabilization of the Pol II clamp, DNA loading, and recruitment of additional factors. These regulatory strategies may extend to other transcription systems. For example the Pol I-specific factor Rrn3 binds Pol I in the region corresponding to interfaces A and B (Fig. 24a) and may stabilize the TFIIB-like factor Rrn7[186] together with Rrn6[187]. This work further explains mutational and genetic data on the head module, and their impact on initiation complex formation (Figs 12, 24, and 25). We provide evidence that Mediator is a global (Supplementary Fig. 2) and basal (Fig. 6a) regulator of transcription, in agreement with previous studies[53,54]. In summary, these results contribute to a structural and functional framework of the conserved mechanism of transcription initiation in eukaryotes (Table 1).

Many questions regarding transcription initiation remain. What is the structure and function of the remaining Mediator tail and kinase module subunits? Do transcription activators influence Mediator tail module structure or inter-module contacts? How do they affect transcriptional output? Little structural information is available for tail and kinase modules and their subcomplexes (see 1.2.3), yet they may be important pivots of Mediator control. Transcription activators may affect yeast[94] and human[188] Mediator conformation, yet a detailed understanding is lacking. The kinase module may inhibit initiation by competing with the Pol II CTD to bind Mediator[78]. Additionally, the kinase module may occlude the Pol II binding site on Mediator through a steric or allosteric mechanism. However, the kinase module may also have an activating function[80]. For example, the cdk8 kinase can activate transcription factors, such as the oncogenic beta-catenin[189]. The kinase module further influences Mediator recruitment, for example through associated activating non-coding RNA in human cells[190]. It will be important to extend recombinant expression of core Mediator established here to tail and kinase modules to enable structural and functional analysis.

How are Pol II, Mediator and GTFs released during Pol II promoter escape? Mediator may remain on Pol II after promoter escape, because modelling suggests it neither interferes with RNA exit or with binding of the elongation factors TFIIS[22] and Spt4/5[191]. Pulldown experiments suggest that Pol II, *in vitro* phosphorylated at residue Ser5 of the CTD, does not inhibit the binary Pol II-Mediator core interaction (Plaschka et al., unpublished), in contrast to pulldowns using heterogeneous endogenous proteins[68]. Mediator association with the initiation complex is transient *in vivo* and was shown to depend on kinase activity of the general factor TFIIH[184,192]. Therefore, promoter escape may depend on the combined force of a transcribing Pol II[193], nascent RNA and CTD phosphorylation. The nascent transcript assists TFIIB release and facilitates formation of the elongation complex. CTD phosphorylation enables the recruitment of elongation factors that disrupt the Pol II-Mediator interaction and facilitate transcription elongation.

What are the structures of the general factors TFIID, -IIE and -IIH? How do these co-operate among themselves and with Mediator to form the complete transcription initiation complex? Does Mediator assist in DNA melting and loading into the Pol II active site? *Sc* TFIID has a molecular mass of 1.2 MDa and consists of 14 subunits[43]. Future work may reveal the structure of this flexible[194] complex, how the observed co-operativity with Mediator is achieved[69], and how the +1 nucleosome regulates the initiation complex[195]. The 10-subunit TFIIH will remain of great interest due to its role in promoter opening, CTD phosphorylation, and DNA repair. We anticipate important contacts between the core Mediator 'cradle' and TFIIH to stimulate CTD phosphorylation and possibly DNA melting. TFIIE is a core GTF that together with TFIIB and TBP is conserved from archea to eukaryotes[1], yet its underlying mechanism remains unknown. TFIIE makes contact with the Pol II clamp domain[18] and may cooperate with Mediator since the TFIIE Zn-ribbon domain binds to the Pol II stalk domain, which is repositioned by Mediator (Fig. 27c).

How does Mediator influence chromatin architecture? The cohesin complex regulates genome architecture and showed significant interaction with human Mediator by pulldown coupled to mass spectrometry[75]. This cohesin-Mediator contact was confirmed by co-IP and ChIP experiments in a mouse cell line and revealed that this interaction is crucial to maintain enhancer-promoter chromatin architecture[61]. Future studies on Mediator may therefore look into the many additional regulators

that interact with Mediator and tune transcription output, for example factors involved in RNA processing, chromatin maintenance and modification, and DNA repair[75].

Answers to these many questions will require assembly of the complete transcription initiation complex. This goal will benefit from strategies established here. However, to fully unravel the molecular details of transcription requires the continued development of strategies to reconstitute transient macromolecular complexes, to integrate structural biology and single-molecule methods, and to apply functional *in vitro* and *in vivo* techniques. The final challenge will be to integrate all available data into a single model that can explain and predict how the cell controls gene transcription.

Structural biology will continue to play a central role in deciphering the cell's inner workings. For more than half a century macromolecular X-ray crystallography has delivered key insights into biology[115]. Automation from crystallization to data processing, new third generation beamlines, the free-electron laser and electron diffraction experiments are likely to push the limits of crystallography even further. Mass spectrometry has made great contributions to integrated structural biology with the advent of protein crosslinking, hydrogen-deuterium exchange, surface modification and protein-RNA crosslinking[196]. Importantly, we should aim to integrate available mass spectrometry software into a single package that serves the community analogous to the collaborative computational project No. 4 (CCP4). Further, we should continue development of a mass spectrometry data bank[197] and encourage scientists to share published data with the community.

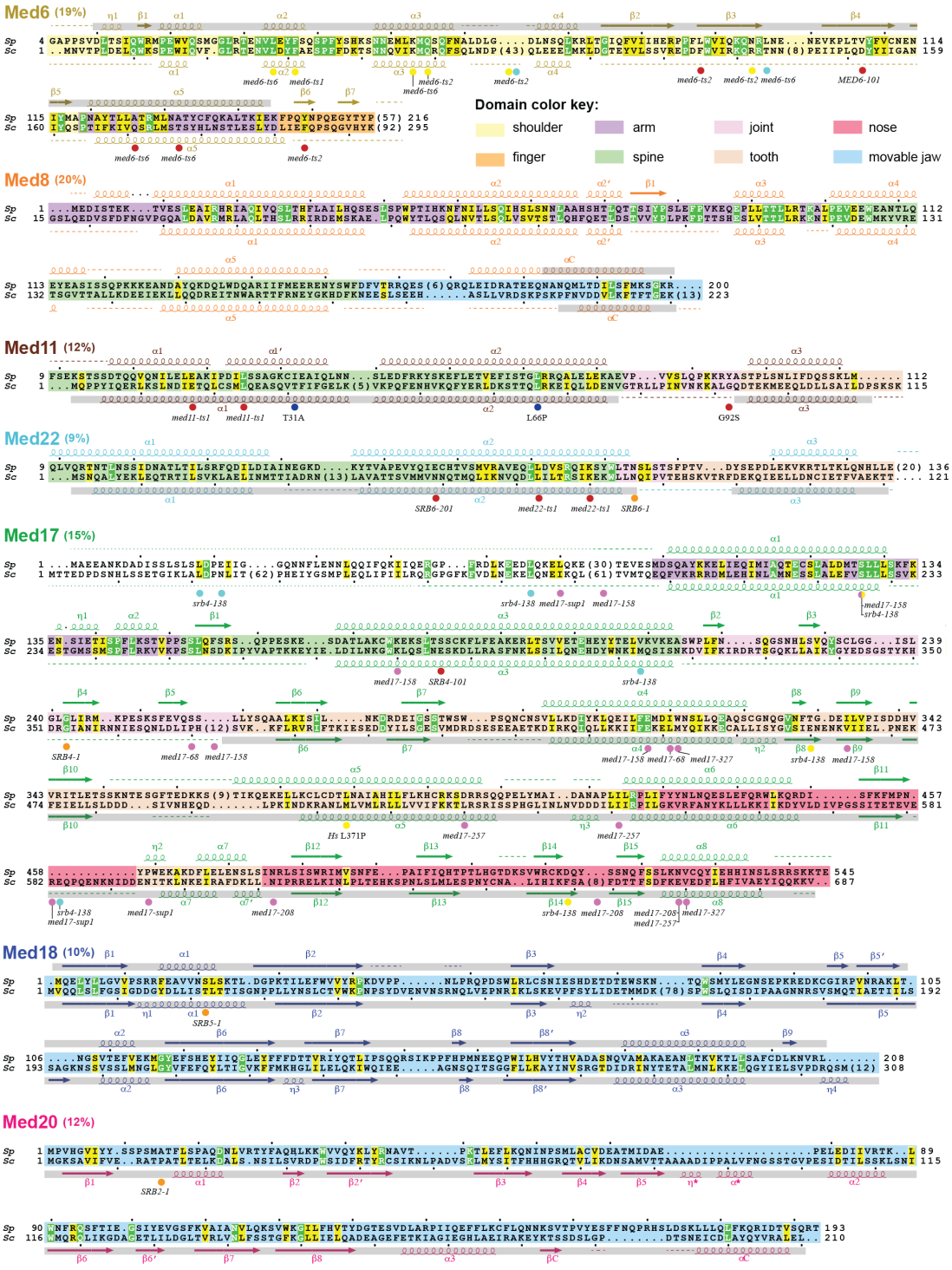
We may expect the greatest progress in EM. Improved imaging hardware and processing software have enabled a 'resolution revolution'[116] in EM. Even asymmetric macromolecules of a molecular weight less than 0.5 MDa can now yield sub-nanometer and even near-atomic resolution[198]. These new developments will make single particle analysis a routine method of structure determination. In parallel, cryo-electron tomography permits the study of protein complexes in their native cellular environment. A recent breakthrough demonstrates the promise of this technique[199]. By the combined use of a new type of phase plate and direct detection device, the Baumeister group was able to quantitatively study conformational states of the 26S proteasome in neuronal cells[199]. This technique may reveal nuclear chromatin

structure, enhancer-promoter contacts, and gene looping, in short, the *in situ* structure of a gene.

It will require a combined national and international effort, to make EM facilities easily accessible and to facilitate software and hardware developments. It will be paramount to improve software that automates operation of the EM microscope and imaging hardware, to increase the throughput of biological samples. Great progress has been made, through for example open-source projects such as TOM^2 (ref.[146]) or LEGINON[200] that control the microscope and enable automated data acquisition. Automatic alignment of several orders of astigmatism has been achieved on a C_S-corrected microscope[201]. Image analysis will continue to improve and benefit from a communal effort, such as the recently established CCPEM project[202], closely modelled on CCP4.

Combining these orthogonal structural and system-wide data will require continued development of integrated modelling platforms[203]. Most importantly, we must continue to support a coordinated and communal research infrastructure and enable scientific discovery.

Appendix



Supplementary Figure 1 | Conserved primary and secondary structure of head module subunits. Amino acid sequence alignments for Mediator head subunits Med6, Med8, Med11, Med22, Med17, Med18, and Med20 from *Schizosaccharomyces pombe* (*Sp*, top), and *Saccharomyces cerevisiae* (*Sc*, bottom). Sequence identity between *Sp* and *Sc* subunits is indicated in brackets next to the protein name. Secondary structure elements are indicated above and below the sequences as observed in the *Sp* head module structure and the *Sc* backbone model, respectively (spirals, α-helices and 310/η-helices; arrows, β-strands; dashed lines, disordered regions; dots, not present in protein variant). Regions for which high-

resolution structures are available are highlighted in grey over the secondary structure elements. Residues that are invariant or conserved among the yeasts *Sp*, *Sc*, *Candida glabrata*, *Candida albicans*, *Ashbya gossypii*, *Kluyveromyces lactis*, and *Debaryomyces hansenii* are highlighted in green or yellow, respectively. Subunit names are given in a colour code that is used throughout (Med6, yellow; Med8, orange; Med11, brown; Med22, cyan; Med17, green; Med18, blue; Med20, magenta). Portions of the sequences forming the eight structural elements of the head module are highlighted in eight different colours (domain colour key in Fig. 3). Spheres indicate known mutation sites (*med6-ts1*[169], *med6-ts2*[169], *med6-ts6*[169], *MED6-101*[170], *med11-ts1*[100], G92S[20], *srb4-138*[49,97], *SRB4-101*[169], *med17-68*[171], *med17-158*[171], *med17-208*[171], *med17-257*[171], *med17-327*[171], *med22-ts1*[100], *Hs* L371P[168], T31A[20], L66P[20], *SRB4-1*[48], *med17-sup1*[171], *SRB5-1*[48], *SRB2-1*[47], *SRB6-1*[48], *SRB6-201*[170]) and are coloured according to their predicted effect (intrasubunit stability, yellow; intersubunit stability, red; unknown effect, cyan; suppressor of CTD truncation, orange; mutations affecting TFIIH interaction, dark blue; mutations affecting Pol II core interaction, magenta).

Supplementary Table 1 | Data collection and refinement statistics for Med6 and Med17C/Med11C/Med22C crystals.

	Med6 (native)	Med6 (SeMet)	Med17C/Med11C/Med22C (SeMet)
Data collection			
Space group	$P2_13$	$P2_13$	R32
Cell dimensions			
a, *b*, *c* (Å)	126.0, 126.0, 126.0	125.0, 125.0, 125.0	261.2, 261.2, 47.7
Wavelength (Å)	0.91890	0.97960	0.97972
Resolution (Å)[†]	2.7 (2.77-2.70)*	3.4 (3.49-3.40)	3.0 (3.08-3.00)
$CC_{1/2}$[‡]	100.0 (32.4)	–	99.9 (17.1)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	99.9 (100.0)
Redundancy	20.0 (19.3)	10.4 (10.4)	5.1 (5.2)
R_{merge} (%)	5.2 (1010.2)	16.6 (145.1)	6.1 (692.2)
$I/\sigma I$	41.4 (0.6)	11.8 (2.4)	16.5 (0.6)
Refinement			
Resolution (Å)	89–2.7		65–3.0
No. reflections	18562		12487
R_{work}/R_{free}	18.1/21.7		20.4/22.8
No. atoms			
Protein	2254		2547
Ligand			12
B-factors			
Protein	136		151
Ligand			190
R.m.s deviations			
Bond lengths (Å)	0.006		0.007
Bond angles (°)	0.967		1.221

*Highest resolution shell is shown in parenthesis.

[†]Resolution limits are provided using the $CC_{1/2} > 10\%$ criterion[125]. Using the traditional criterion of $I/\sigma I > 2.0$, resolution limits are 2.9 Å and 3.2 Å for Med6 and Med17C/11C/22C crystals, respectively.

[‡]$CC_{1/2}$ = percentage of correlation between intensities from random half-datasets[125]

Supplementary Table 2 | Mediator head module mutations, their location, and predicted effect.

Yeast strain	Site(s) in *Sc* (*Sp*)	Location	Predicted effect	Ref.
med6-ts1	F31S (F37)	Med6, shoulder	Fold destabilization	[169]
med6-ts2	Q49L (Q55)	Med6, shoulder	Fold destabilization	[169]
	I68L (α3-α4)	Med6, shoulder	Unknown	
	L94P (α3-α4)	Med6, shoulder	Unknown	
	F125Y (L88)	Med6, shoulder	Intersubunit destabilization	

	R132G (N95)	Med6, shoulder	Fold destabilization	
	F194L (Y149)	Med6, finger	Intersubunit destabilization	
med6-ts6	L28P (L34)	Med6, shoulder	Fold destabilization	[169]
	K47T (K53)	Med6, shoulder	Fold destabilization	
	T134A (L97)	Med6, shoulder	Unknown	
	Q171R (A126)	Med6, shoulder	Intersubunit destabilization	
	T177M (A132)	Med6, arm	Intersubunit destabilization	
	M273L (β7-C-term)	Med6 (n/a)	Unknown	
	I275V (β7-C-term)	Med6 (n/a)	Unknown	
MED6-101	D152Y (V107)	Med6, shoulder	Intersubunit stabilization	[170]
med11-ts1	E17K (E28)	Med11, spine	Intersubunit destabilization	[100]
	L24K (L35)	Med11, spine	Intersubunit destabilization	
T31A*	T31A (I42)	Med11, spine	Intersubunit destabilization	[20]
L66P*	L66P (L72)	Med11, spine	Fold destabilization	[20]
G92S*	G92S (Y95)	Med11, joint	Intersubunit destabilization	[20]
srb4-138	L21S (L18)	Med17 (n/a)	Not required for *ts*	[49,97]
	N24I (E21)	Med17 (n/a)	Unknown	
	L124M (L56)	Med17 (n/a)	Not required for *ts*	
	S226P (S127)	Med17, arm	Fold destabilization	
	M313I (V209)	Med17, spine	Unknown	
	E460G (T329)	Med17, tooth	Fold destabilization	
	E583G (β11-α7)	Med17, nose	Unknown	
	F649S (D513)	Med17, nose	Fold destabilization	
SRB4-1	G353C (G242)	Med17, joint	Intersubunit stabilization	[48]
SRB4-101	E286K (S182)	Med17, spine	Intersubunit stabilization	[169]
med17-68	P370S (S257)	Med17, joint	Fold destabilization	[171]
	L441P (I310)	Med17, tooth	Fold destabilization	
med17-158	F159Y (N-term-α1)	Med17 (n/a)	Unknown	[171]
	S226T (S127)	Med17, arm	Fold destabilization	
	K280M (K176)	Med17, spine	Intersubunit destabilization	
	K377N (β5-β6)	Med17, joint	Unknown	
	E438G (E307)	Med17, tooth	Fold destabilization	
	V465E (I333)	Med17, tooth	Fold destabilization	
med17-208	N611H (N476)	Med17, nose	Unknown	[171]
	A655T (β14-β15)	Med17, nose	Unknown	
	E669D (E525)	Med17, nose	Unknown	
med17-257	L520S (D405)	Med17, tooth	Intersubunit destabilization	[171]
	I541K (I424)	Med17, nose	Fold destabilization	
	E669G (E525)	Med17, nose	Unknown	
med17-327	M442V (W311)	Med17, tooth	Fold destabilization	[171]
	V670E (V526)	Med17, nose	Fold destabilization	
med17-sup1	I128V (L60)	Med17 (n/a)	Unknown	[171]
	R582G (β11-α7)	Med17, nose	Unknown	
	N595D (P459)	Med17, tooth	Unkonwn	
SRB5-1	T22I (S21)	Med18, movable jaw	Intersubunit stabilization	[48]
SRB2-1	P14H (A15)	Med20, movable jaw	Unknown	[47]
SRB6-1	N86K (N84)	Med22, spine	Intersubunit stabilization	[48]
med22-ts1	L73E (L71)	Med22, spine	Intersubunit destabilization	[100]
	K80E (K78)	Med22, spine	Intersubunit destabilization	

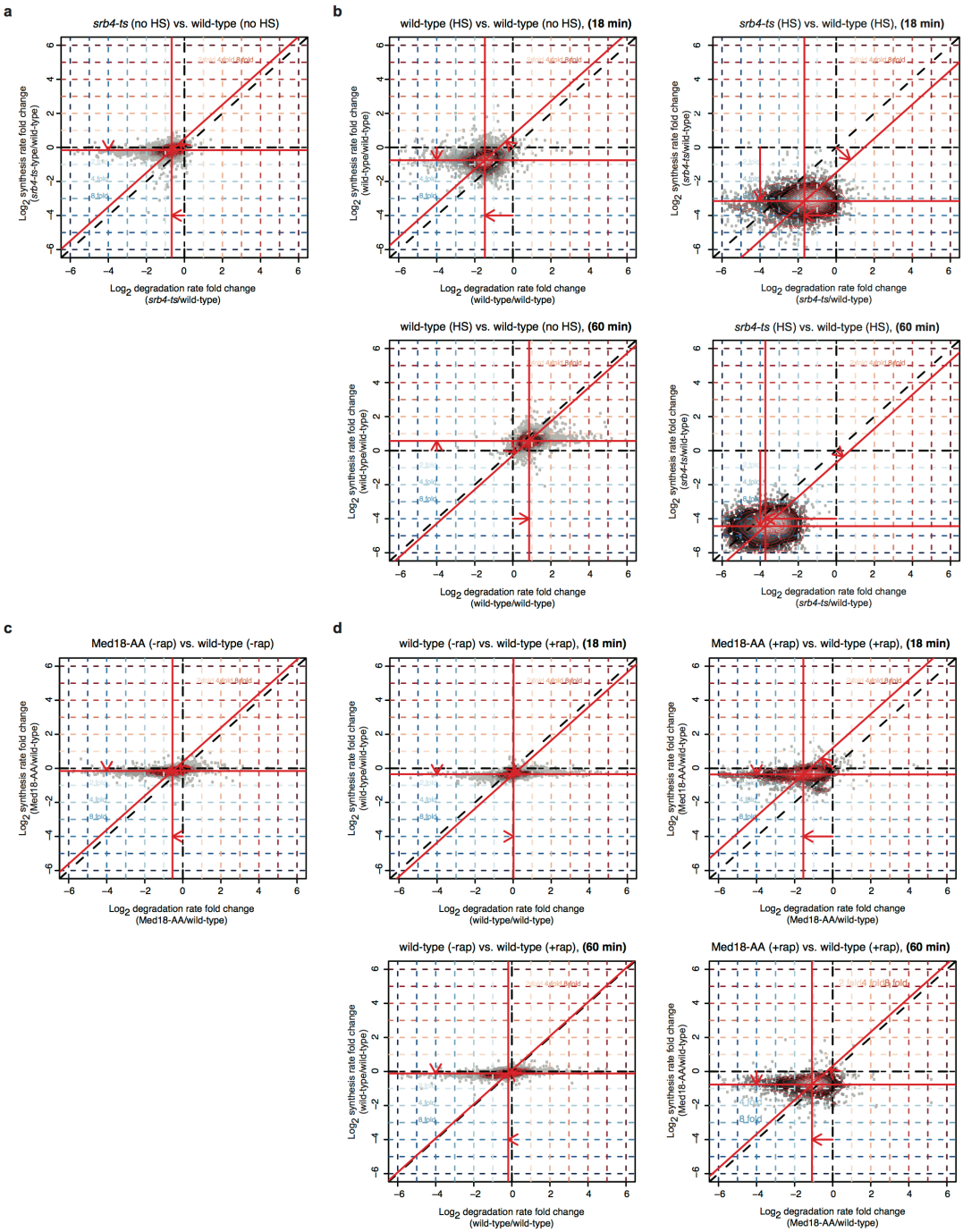
SRB6-201	N59H (E57)	Med22, spine	Intersubunit stabilization	[47]
Human	**Site in *Hs* (*Sc*, *Sp*)**	**Location**	**Predicted effect**	**Ref.**
Hs L371P	L371P (M504, L389)	Med17, tooth	Fold destabilization	[168]

Sc, *Saccharomyces cerevisiae*; *Sp*, *Schizosaccharomyces pombe*; *Hs*, *Homo sapiens*; Ref., Reference.
Residues with no equivalent in *Sp* were annotated by the N- and C-terminal secondary structure elements closest to that residue ('secondary structure'-'secondary structure'). When mutations fell within three amino acids of the built *Sc* or *Sp* models, they were included in Fig. 12a or 12b, respectively for illustrative purposes.
* Med11 mutations were numbered according to the corrected annotation[100]. T31A, L66P and G92S refer to T47A, L82P and G108S, respectively in the original publication[20].

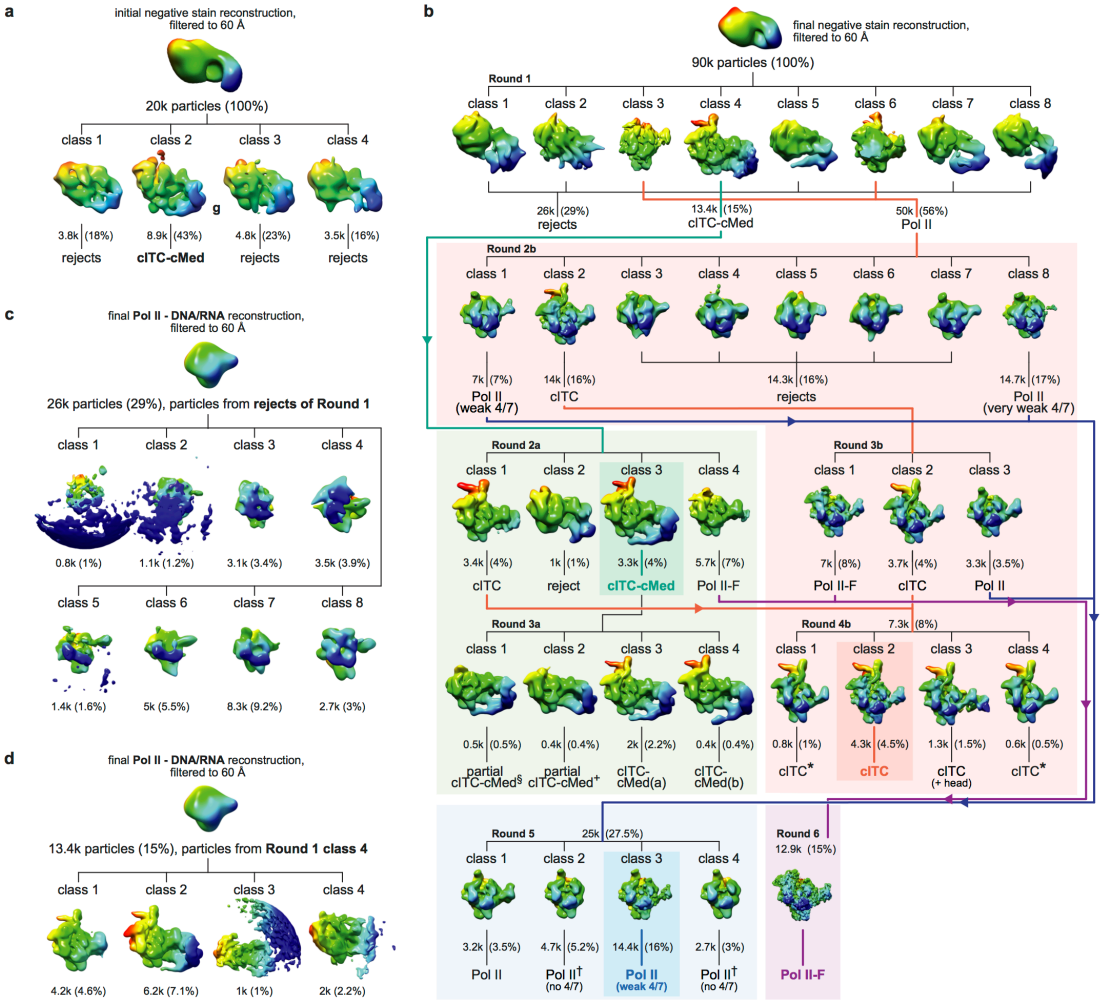
Supplementary Table 3 | Components of the cITC-cMed complex.

	Protein subunits	Length (aa)	Molecular weight (kDa)
RNA polymerase II	Rpb1	1733	191.6
	Rpb2	1224	138.8
	Rpb3‡	318	35.3
	Rpb4	221	25.4
	Rpb5	215	25.1
	Rpb6	155	17.9
	Rpb7	171	19.1
	Rpb8	146	16.5
	Rpb9	122	14.3
	Rpb10	70	8.3
	Rpb11	120	13.6
	Rpb12	70	7.7
General initiation factors	TFIIB‡	345	38.2
	TBP‡	61-240	20.2
	Tfg1*‡	734	82.3
	Tfg2	400	46.6
Mediator head module	Med6	295	32.8
	Med8	222	25.3
	Med11	115	13.3
	Med17	687	78.5
	Med18	307	34.3
	Med20	210	22.9
	Med22	121	13.7
Mediator middle module	Med4	284	32.2
	Med7	222	25.6
	Med9	149	17.4
	Med10	157	17.9
	Med21	140	16.1
	Med31	127	14.7
Mediator (other subunits)	Med14†	745	84.6
	Med19	220	24.9
Total	31 subunits	10,220	1,155.1
Nucleic acid strands		Length (nt)	Molecular weight (kDa)
Template	DNA	72	22.2
Non-template	DNA	72	22.5
Initial transcript	RNA	6	2.1

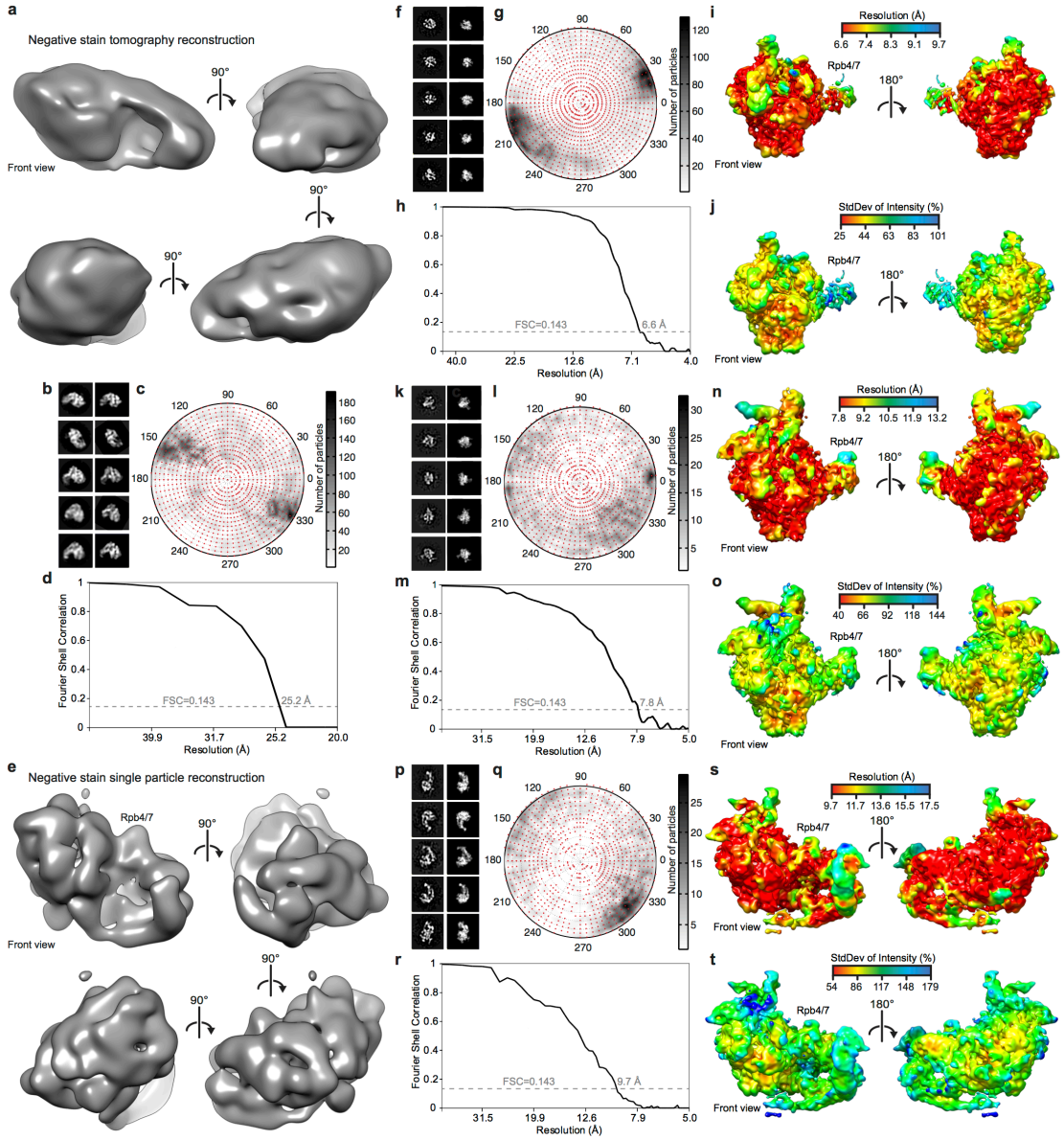
aa, amino acids; nt, nucleotides; kDa, kilo Dalton.
*Tfg1 was from *S. mikatae*[37].
†Med14 construct contains an additional N-terminal 10xHis tag.
‡Constructs contain an N- or C-terminal 6xHis tag as described[40].



Supplementary Figure 2 | Global requirement of the Mediator head module for transcription. a, Fold changes in RNA degradation (log folds, x-axis) and synthesis (log folds, y-axis) rates observed in strains *srb4-ts* versus wild-type in the absence of heat shock. Each point corresponds to one mRNA and the density of points is reflected in their brightness. Red contour lines define regions of equal intensity. The centre of the distribution results from the median synthesis and degradation rates, whose relative contributions are indicated by shifts of the red lines parallel to synthesis or degradation rate axis, respectively. **b,** Global shutdown of RNA synthesis upon heat shock (HS) of the *srb4-ts* mutant. Fold changes in degradation (log fold, x-axis) and synthesis (log folds, y-axis) rates of *srb4-ts* and wild-type strains are indicated, after 18 or 60 min of HS treatment, respectively. **c,** As for a, but using a Med18-anchor-away (AA) strain in absence of rapamycin (rap) treatment. **d,** Global down regulation of RNA synthesis upon anchor-away of the Med18 subunit, after 18 or 60 min of rapamycin (rap) treatment, as in b.



Supplementary Figure 3 | 3D classification of negative-stain and cryo-EM data. a, 3D classification of the negative-stain EM data set into four classes. The percentage of data in each class is given in parenthesis. To help visualize structural differences, all 3D reconstructions were radially coloured in UCSF Chimera. **b,** Pseudo-hierarchical 3D classification of the cryo-EM data set. The percentage of the data in each class is given in parentheses. Rejects refer to EM reconstructions that did not reflect the known structures of Pol II, cITC or cITC–cMed. §, class of partial cITC–cMed particles lacking upstream DNA–TFIIF–TFIIB–TBP; +, class of partial cITC–cMed particles lacking upstream DNA–TFIIB–TBP and the mobile plank of cMed; *, class of cITC particles that do not average well with the main cITC class; †, class of Pol II–DNA/RNA particles lacking Rpb4–Rpb7. cITC–cMed(a) and cITC–cMed(b) classes from Round 3a correspond respectively to class 3 and class 4 superimposed in Fig. 24d. The Pol II–TFIIF class (Round 6) presented with density for the TFIIF dimerization module and the Tfg1 'charged helix', but weak to no density for Tfg1 'arm' and Tfg2 'linker' regions due to the absence of upstream DNA stabilizing factors TFIIB and TBP. Classes were visualized as in a. **c,** 3D classification of particles from rejects of round 1 using the Pol II–DNA/RNA reconstruction as initial model. Particles were sorted into eight groups, resulting in poor 3D reconstructions. Classes were visualized as in a. **d,** 3D classification of particles from class 4 of round 1 using the Pol II–DNA/RNA reconstruction as initial model. Particles were sorted into four groups, resulting in EM reconstructions of cITC and cITC–cMed. These results suggest high data quality, and further the presence of a single detectable cITC–cMed conformation in the cryo-EM data, even in absence of a cITC–cMed reference. Classes were visualized as in a.



Supplementary Figure 4 | Negative stain and cryo-EM reconstructions of Pol II–DNA/RNA, cITC, and cITC–cMed complexes. a, Four views of the negative-stain tomography reconstruction of the cITC–cMed related by 90° rotation, starting from the previously defined front view of Pol II[155]. **b,** Comparison of five reference-free 2D class averages calculated from all particles used in the final negative-stain single-particle reconstruction with corresponding forward projections of the reconstruction. **c,** Orientational distribution plot of all particles in the final negative-stain single-particle reconstruction. The estimated angular accuracy is 3.2°. **d,** Fourier shell correlation of the final negative-stain single-particle cITC–cMed reconstruction (0.143 FSC = 25.2 Å resolution bin). **e,** Four views of the negative-stain single-particle reconstruction of the cITC–cMed related by 90° rotation, starting from the previously defined front view of Pol II[155]. **f,** Comparison of five reference-free 2D class averages calculated from all particles used in the final Pol II–DNA/RNA cryo-EM single-particle reconstruction with corresponding forward projections of the reconstruction. **g,** Orientational distribution plot of all particles in the final cryo-EM Pol II–DNA/RNA single-particle reconstruction. The estimated angular accuracy is 3.2°. **h,** Fourier shell correlation of the final Pol II–DNA/RNA cryo-EM single-particle reconstruction (FSC = 0.143). **i,** Two views of the Pol II–DNA/RNA cryo-EM map are shown from the previously defined front view of Pol II[155] and rotated by 180°, and are coloured by local resolution. **j,** Two views of the Pol II–DNA/RNA cryo-EM map are shown from the previously defined front view of Pol II[155] and rotated by 180°, and are coloured by variance (the standard deviation, StdDev, of the normalized intensity value). **k–o,** As f–j but for the cITC reconstruction. **p–t,** As f–j but for the cITC–cMed reconstruction.

List of abbreviations

Å	Angstrom
ATP	Adenosine triphosphate
Amp	Ampicillin
BS3	Bis-sulfo-succinimidyl-suberate
C-terminus	Carboxy-terminus
CBDPS	Cyanurbiotindipropionyl succinimide
cDTA	comparative Dynamic Transcriptome Analysis
cITC	core Initially Transcribing Complex
cMed	core Mediator
CV	Column volume
CTD	Carboxy-terminal domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSS	Di-succinimidyl-suberate
DTT	1,4-dithio-D,L-threitol
E. coli	Escherichia coli
EM	Electron microscopy
EMDB	Electron microscopy data bank
FSC	Fourier shell correlation
GTF	General transcription factor
His-tag	Histidine-tag
Hs	*Homo sapiens*
IPTG	Isopropyl-b-d-thiogalactoside
Kan	Kanamycin
MDa	Megadalton
MIRAS	Multiple isomorphous replacement and anomalous diffraction
mRNA	messenger ribonuecleic acid
N-terminus	Amino-terminus
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data bank
PIC	Pre-initiation complex
Pol II	RNA Polymerase II
RNA	Ribonuecleic acid
SAD	Single-wavelength anomalous diffraction
Sc	*Saccharomyces cerevisiae*
SDS	Sodium dedocyl sulfate
Sm	Streptomycin
Sp	*Schizosaccharomyces pombe*
TAF	TBP-associated factor
TAP	Tandem-affinity purification
TBP	Tata-box binding protein
TFII	Transcription factor II

List of figures

Figure 1 | Schematic of the transcription cycle. 3
Figure 2 | Schematic of the Pol II transcription initiation complex. 4
Figure 3 | Known structures of Mediator subunits and subcomplexes. 10
Figure 4 | Structures of Med6 and Med17C–Med11C–Med22C. 43
Figure 5 | Bacterial expression vectors. 44
Figure 6 | Structure of Sp Mediator head module. 45
Figure 7 | Electron density maps for the Sp Mediator head module. 47
Figure 8 | Structural elements and surface conservation. 48
Figure 9 | Subunit structure within the Sp Mediator head module. 49
Figure 10 | Structural element deletions affect Mediator head function in vivo and in vitro. 50
Figure 11 | Structural comparisons reveal head module flexibility. 51
Figure 12 | Head module integrity and interactions. 52
Figure 13 | Preparation and CX-MS analysis of the Mediator middle module. 55
Figure 14 | Map of lysine-lysine cross-link pairs of the Mediator middle module. 55
Figure 15 | Model of the Mediator middle module. 57
Figure 16 | Reconstitution of Pol II-Mediator complex cITC–cMed. 59
Figure 17 | cMed reconstitution, activities, and cITC binding. 59
Figure 18 | EM structure of Pol II initiation complex cITC. 61
Figure 19 | Structural modelling of cITC into the cryo-EM density. 62
Figure 20 | EM structure of the cITC–cMed complex. 64
Figure 21 | Structural modelling of cITC–cMed into the cryo-EM density. 65
Figure 22 | Protein crosslinking and cMed architecture. 67
Figure 23 | Protein–protein crosslinking. 68
Figure 24 | cITC–cMed interfaces. 70
Figure 25 | Analysis of the cITC–cMed interface. 71
Figure 26 | Initiation complex model and CTD cradle. 73
Figure 27 | Inferred locations of the Mediator tail module, TFIIS, the general factors TFIIA, TFIIE, and TFIIH, and the CTD, and comparison with the human initiation complex. 74
Supplementary Figure 1 | Conserved primary and secondary structure of head module subunits. 77
Supplementary Figure 2 | Global requirement of Mediator head module for transcription. 81
Supplementary Figure 3 | 3D classification of negative-stain and cryo-EM data. 82
Supplementary Figure 4 | Negative stain and cryo-EM reconstructions of Pol II– DNA/RNA, cITC, and cITC–cMed complexes. 83

List of tables

Table 1 | Sc RNA polymerase subunits and initiation factors. 2
Table 2 | General Transcription Factors of *Sc* Pol II. 5
Table 3 | List of Sc Mediator modules and subunits. 8
Table 4 | Bacterial strains used in this study. 14
Table 5 | Yeast strains used in this study. 14
Table 6 | Plasmids used in this study. 15
Table 7 | Oligonucleotides used for EM. 16
Table 8 | DNA used for transcription assays. 16
Table 9 | Antibodies used in this study. 17
Table 10 | Growth media for *E. coli* and *S. cerevisiae* cultures. 17
Table 11 | Additives for *E. coli* cultures. 17
Table 12 | List of general buffers and solutions. 17
Table 13 | Data collection and refinement statistics for Sp head module crystals. 46
Supplementary Table 1 | Data collection and refinement statistics for Med6 and Med17C/Med11C/Med22C crystals. 78
Supplementary Table 2 | Mediator head module mutations, their location, and predicted effect. 78
Supplementary Table 3 | Components of the cITC-cMed complex. 80

References

1 Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nature reviews. Microbiology* **9**, 85-98, doi:10.1038/nrmicro2507 (2011).

2 Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Molecular cell* **45**, 439-446, doi:10.1016/j.molcel.2012.01.023 (2012).

3 Matzke, M., Kanno, T., Daxinger, L., Huettel, B. & Matzke, A. J. M. RNA-mediated chromatin-based silencing in plants. *Current opinion in cell biology* **21**, 367-376 (2009).

4 Cheetham, G. M. T. & Steitz, T. A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305-2309, doi:10.1126/science.286.5448.2305 (1999).

5 Svejstrup, J. The RNA polymerase II transcription cycle: cycling through chromatin. *Biochimica et biophysica acta* **1677**, 64-73, doi:10.1016/j.bbaexp.2003.10.012 (2004).

6 Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549-561 (1989).

7 Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical reviews* **113**, doi:10.1021/cr400071f (2013).

8 Mayer, A. *et al.* Uniform transitions of the general RNA polymerase II transcription complex. *Nature structural & molecular biology* **17**, 1272-1278, doi:10.1038/nsmb.1903 (2010).

9 Kim, H. *et al.* Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nature structural & molecular biology* **10**, 1279-1286, doi:10.1038/nsmb.1913 (2010).

10 Aguilera, A. & García-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Molecular cell* **46**, 115-124, doi:10.1016/j.molcel.2012.04.009 (2012).

11 Lehmann, E., Brueckner, F. & Cramer, P. Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* **450**, 445-449, doi:10.1038/nature06290 (2007).

12 Allen, T. A., Kaenel, V. S. & Goodrich, J. A. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nature structural & molecular biology*, doi:10.1038/nsmb813 (2004).

13 Hahn, S. & Young, E. T. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705-736, doi:10.1534/genetics.111.127019 (2011).

14 Kostrewa, D. *et al.* RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323-330, doi:10.1038/nature08548 (2009).

15 Liu, X., Bushnell, D., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206-209 (2010).

16 Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437-440, doi:10.1038/nature11715 (2012).

17 Fishburn, J. & Hahn, S. Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. *Molecular and cellular biology* **32**, 12-25, doi:10.1128/MCB.06242-11 (2012).

18 Grünberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature structural & molecular biology* **19**, 788-796, doi:10.1038/nsmb.2334 (2012).

19 Yudkovsky, N., Ranish, J. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225-229, doi:10.1038/35041603 (2000).

20 Esnault, C. *et al.* Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Molecular cell* **31**, 337-346, doi:10.1016/j.molcel.2008.06.021 (2008).

21 Larivière, L., Seizl, M. & Cramer, P. A structural perspective on Mediator function. *Current opinion in cell biology* **24**, 305-313, doi:10.1016/j.ceb.2012.01.007 (2012).

22 Kettenberger, H., Armache, K.-J. & Cramer, P. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347-357 (2003).

23 Kim, B. *et al.* The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16068-16073, doi:10.1073/pnas.0704573104 (2007).

24 Guglielmi, B., Soutourina, J., Esnault, C. & Werner, M. TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16062-16067, doi:10.1073/pnas.0704534104 (2007).

25 Nock, A., Ascano, J., Barrero, M. & Malik, S. Mediator-Regulated Transcription through the +1 Nucleosome. *Molecular cell* **48**, 837-848, doi:10.1016/j.molcel.2012.10.009 (2012).

26 Ishibashi, T. *et al.* Transcription factors IIS and IIF enhance transcription efficiency by differentially modifying RNA polymerase pausing dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 3419-3424, doi:10.1073/pnas.1401611111 (2014).

27 Cheung, A. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249-253, doi:10.1038/nature09785 (2011).

28 Sikorski, T. W. & Buratowski, S. The basal initiation machinery: beyond the general transcription factors. *Current opinion in cell biology* **21**, doi:10.1016/j.ceb.2009.03.006 (2009).

29 Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520 (1993).

30 Nikolov, D. B. *et al.* Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119-128, doi:10.1038/377119a0 (1995).

31 Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272**, 830-836 (1996).

32 Armache, K. J., Mitterweger, S., Meinhart, A. & Cramer, P. Structures of Complete RNA Polymerase II and Its Subcomplex, Rpb4/7. *Journal of Biological Chemistry* **280**, 7131-7134, doi:10.1074/jbc.M413038200 (2004).

33 Leuther, K., Bushnell, D. & Kornberg, R. Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes: implications for start site selection and initiation complex formation. *Cell* **85**, 773-779, doi:10.1016/S0092-8674(00)81242-8 (1996).

34 Chen, H. T. & Hahn, S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* **119**, 169-180 (2004).

35 Bushnell, D. A., Westover, K. D., Davis, R. E. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* **303**, 983-988, doi:10.1126/science.1090838 (2004).

36 Gaiser, F., Tan, S. & Richmond, T. J. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. *Journal of molecular biology* **302**, 1119-1127 (2000).

37 Chen, H.-T., Warfield, L. & Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nature Structural & Molecular Biology* **14**, 696-703, doi:10.1038/nsmb1272 (2007).

38 Chen, Z. *et al.* Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO journal* **29**, 717-726, doi:10.1038/emboj.2009.401 (2010).

39 He, Y., Fang, J., Taatjes, D. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481-486, doi:10.1038/nature11991 (2013).

40 Mühlbacher, W. *et al.* Conserved architecture of the core RNA polymerase II initation complex. *Nature Communications*, in press (2014).

41 Cai, G. *et al.* Interaction of the mediator head module with RNA polymerase II. *Structure (London, England : 1993)* **20**, 899-910, doi:10.1016/j.str.2012.02.023 (2012).

42 Murakami, K. *et al.* Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342**, 1238724, doi:10.1126/science.1238724 (2013).

43 Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, doi:10.1038/nrm3952 (2015).

44 Koleske, A. J., Buratowski, S., Nonet, M. & Young, R. A. A novel transcription factor reveals a functional link between the RNA polymerase II CTD and TFIID. *Cell* **69**, 883-894 (1992).

45 Kim, Y., Björklund, S., Li, Y., Sayre, M. & Kornberg, R. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608, doi:10.1016/0092-8674(94)90221-6 (1994).

46 Koleske, A. J. & Young, R. A. An RNA polymerase II holoenzyme responsive. *Nature* **368**, 466-496 (1994).

47 Nonet, M. L. & Young, R. A. Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of Saccharomyces cerevisiae RNA polymerase II. *Genetics* **123**, 715-724 (1989).

48 Thompson, C., Koleske, A., Chao, D. & Young, R. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**, 1361-1375, doi:10.1016/0092-8674(93)90362-T (1993).

49 Thompson, C. & Young, R. General requirement for RNA polymerase II holoenzymes in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 4587-4590 (1995).

50 Myers, L. C. & Kornberg, R. D. Mediator of transcriptional regulation. *Annual review of biochemistry* **69**, 729-749, doi:10.1146/annurev.biochem.69.1.729 (2000).

51 Hengartner, C. *et al.* Association of an activator with an RNA polymerase II holoenzyme. *Genes & development* **9**, 897-910, doi:10.1101/gad.9.8.897 (1995).

52 Baek, H., Kang, Y. & Roeder, R. Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. *The Journal of biological chemistry* **281**, 15172-15181, doi:10.1074/jbc.M601983200 (2006).

53 Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717-728 (1998).

54 Takagi, Y. & Kornberg, R. D. Mediator as a general transcription factor. *The Journal of biological chemistry* **281**, 80-89, doi:10.1074/jbc.M508253200 (2006).

55 Bourbon, H.-M. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic acids research* **36**, 3993-4008, doi:10.1093/nar/gkn349 (2008).

56 Sato, S. *et al.* Identification of mammalian Mediator subunits with similarities to yeast Mediator subunits Srb5, Srb6, Med11, and Rox3. *The Journal of biological chemistry* **278**, 15123-15127, doi:10.1074/jbc.C300054200 (2003).

57 Bäckström, S., Elfving, N., Nilsson, R., Wingsle, G. & Björklund, S. Purification of a plant mediator from Arabidopsis thaliana identifies PFT1 as the Med25 subunit. *Molecular cell* **26**, 717-729, doi:10.1016/j.molcel.2007.05.007 (2007).

58 Conaway, R. & Conaway, J. Origins and activity of the Mediator complex. *Seminars in cell & developmental biology* **22**, 729-734, doi:10.1016/j.semcdb.2011.07.021 (2011).

59 Spaeth, J. M., Kim, N. H. & Boyer, T. G. Mediator and human disease. *Seminars in cell & developmental biology* **22**, 776-787 (2011).

60 Malik, S. & Roeder, R. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature reviews. Genetics* **11**, 761-772, doi:10.1038/nrg2901 (2010).

61 Kagey, M. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435, doi:10.1038/nature09380 (2010).

62 Taatjes, D. The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends in biochemical sciences* **35**, 315-322, doi:10.1016/j.tibs.2010.02.004 (2010).

63 Meyer, K., Lin, S.-C., Bernecky, C., Gao, Y. & Taatjes, D. p53 activates transcription by directing structural shifts in Mediator. *Nature structural & molecular biology* **17**, 753-760, doi:10.1038/nsmb.1816 (2010).

64 Taatjes, D. J., Näär, A. M., Andel, F., Nogales, E. & Tjian, R. Structure, function, and activator-induced conformations of the CRSP coactivator. *Science* **295**, 1058-1062, doi:10.1126/science.1065249 (2002).

65 Taatjes, D., Schneider-Poetsch, T. & Tjian, R. Distinct conformational states of nuclear receptor-bound CRSP-Med complexes. *Nature structural & molecular biology* **11**, 664-671, doi:10.1038/nsmb789 (2004).

66 Young, E. *et al.* Artificial recruitment of mediator by the DNA-binding domain of Adr1 overcomes glucose repression of ADH2 expression. *Molecular and cellular biology* **28**, 2509-2516, doi:10.1128/MCB.00658-07 (2008).

67 Liu, Y., Ranish, J., Aebersold, R. & Hahn, S. Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes. *The Journal of biological chemistry* **276**, 7169-7175, doi:10.1074/jbc.M009586200 (2001).

68 Søgaard, T. M. M. & Svejstrup, J. Q. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *The Journal of biological chemistry* **282**, 14113-14120, doi:10.1074/jbc.M701345200 (2007).

69 Johnson, K., Wang, J., Smallwood, A., Arayata, C. & Carey, M. TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes & development* **16**, 1852-1863, doi:10.1101/gad.995702 (2002).

70 Larivière, L. *et al.* Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nature structural & molecular biology* **13**, 895-901, doi:10.1038/nsmb1143 (2006).

71 Cai, G. *et al.* Mediator head module structure and functional interactions. *Nature structural & molecular biology* **17**, 273-279, doi:10.1038/nsmb.1757 (2010).

72 Akoulitchev, S., Chuikov, S. & Reinberg, D. TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature* **407**, 102-106, doi:10.1038/35024111 (2000).

73 Fryer, C., White, J. & Jones, K. Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. *Molecular cell* **16**, 509-520, doi:10.1016/j.molcel.2004.10.014 (2004).

74 Meyer, K. *et al.* Cooperative activity of cdk8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3. *The EMBO journal* **27**, 1447-1457, doi:10.1038/emboj.2008.78 (2008).

75 Ebmeier, C. & Taatjes, D. Activator-Mediator binding regulates Mediator-cofactor interactions. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11283-11288, doi:10.1073/pnas.0914215107 (2010).

76 Elmlund, H. *et al.* The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15788-15793, doi:10.1073/pnas.0607483103 (2006).

77 Knuesel, M., Meyer, K., Bernecky, C. & Taatjes, D. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes & development* **23**, 439-451, doi:10.1101/gad.1767009 (2009).

78 Tsai, K.-L. *et al.* A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nature structural & molecular biology* **20**, 611-619, doi:10.1038/nsmb.2549 (2013).

79 Andrau, J.-C. *et al.* Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Molecular cell* **22**, 179-192, doi:10.1016/j.molcel.2006.03.023 (2006).

80 Donner, A. J., Ebmeier, C. C., Taatjes, D. J. & Espinosa, J. CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nature structural & molecular biology* **17**, 194-201, doi:10.1038/nsmb.1752 (2010).

81 Hu, X. *et al.* A Mediator-responsive form of metazoan RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9506-9511, doi:10.1073/pnas.0603702103 (2006).

82 Cheng, B. *et al.* Functional association of Gdown1 with RNA polymerase II poised on human genes. *Molecular cell* **45**, 38-50, doi:10.1016/j.molcel.2011.10.022 (2012).

83 Jishage, M. *et al.* Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. *Molecular cell* **45**, 51-63, doi:10.1016/j.molcel.2011.12.014 (2012).

84 Davis, M. A. M., Guo, J., Price, D. H. & Luse, D. S. Functional interactions of the RNA polymerase II-interacting proteins Gdown1 and TFIIF. *Journal of Biological Chemistry* **289**, 11143-11152 (2014).

85 Malik, S., Barrero, M. & Jones, T. Identification of a regulator of transcription elongation as an accessory factor for the human Mediator coactivator. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 6182-6187, doi:10.1073/pnas.0608717104 (2007).

86 Huang, Y. *et al.* Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Molecular cell* **45**, 459-469, doi:10.1016/j.molcel.2011.12.022 (2012).

87 Milbradt, A. G. *et al.* Structure of the VP16 transactivator target in the Mediator. *Nature structural & molecular biology* **18**, 410-415, doi:10.1038/nsmb.1999 (2011).

88 Vojnic, E. *et al.* Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nature structural & molecular biology* **18**, 404-409, doi:10.1038/nsmb.1997 (2011).

89 Takahashi, H. *et al.* Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* **146**, 92-104, doi:10.1016/j.cell.2011.06.005 (2011).

90 Asturias, F. J., Jiang, Y. W., Myers, L. C., Gustafson, C. M. & Kornberg, R. D. Conserved Structures of Mediator and RNA Polymerase II Holoenzyme. *Science* **283**, 985-987, doi:10.1126/science.283.5404.985 (1999).

91 Dotson, M. *et al.* Structural organization of yeast and mammalian mediator complexes. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 14307-14310, doi:10.1073/pnas.260489497 (2000).

92 Cai, G., Imasaki, T., Takagi, Y. & Asturias, F. J. Mediator structural conservation and implications for the regulation mechanism. *Structure* **17**, 559-567, doi:10.1016/j.str.2009.01.016 (2009).

93 Wang, X. *et al.* Redefining the modular organization of the core Mediator complex. *Cell research*, doi:10.1038/cr.2014.64 (2014).

94 Tsai, K.-L. *et al.* Subunit Architecture and Functional Modular Rearrangements of the Transcriptional Mediator Complex. *Cell* **157**, 1-15 (2014).

95 Kang, J. *et al.* The structural and functional organization of the yeast mediator complex. *The Journal of biological chemistry* **276**, 42003-42010, doi:10.1074/jbc.M105961200 (2001).

96 Guglielmi, B. *et al.* A high resolution protein interaction map of the yeast Mediator complex. *Nucleic acids research* **32**, 5379-5391, doi:10.1093/nar/gkh878 (2004).

97 Takagi, Y. *et al.* Head module control of mediator interactions. *Molecular cell* **23**, 355-364, doi:10.1016/j.molcel.2006.06.007 (2006).

98 Larivière, L. *et al.* Structure-system correlation identifies a gene regulatory Mediator submodule. *Genes & development* **22**, 872-877, doi:10.1101/gad.465108 (2008).

99 Imasaki, T. *et al.* Architecture of the Mediator head module. *Nature* **475**, 240-243, doi:10.1038/nature10162 (2011).

100 Seizl, M., Larivière, L., Pfaffeneder, T., Wenzeck, L. & Cramer, P. Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization. *Nucleic acids research* **39**, 6291-6304, doi:10.1093/nar/gkr229 (2011).

101 Baumli, S., Hoeppner, S. & Cramer, P. A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer. *The Journal of biological chemistry* **280**, 18171-18178, doi:10.1074/jbc.M413466200 (2005).

102 Koschubs, T. *et al.* Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *The EMBO journal* **28**, 69-80, doi:10.1038/emboj.2008.254 (2009).

103 Koschubs, T. *et al.* Preparation and topology of the Mediator middle module. *Nucleic acids research* **38**, 3186-3195, doi:10.1093/nar/gkq029 (2010).

104 Thakur, J. K., Arthanari, H., Yang, F., Pan, S. J. & Fan, X. A nuclear receptor-like pathway regulating multidrug resistance in fungi. *Nature* **452**, 604-609, doi:10.1038/nature06836 (2008).

105 Brzovic, P. S. *et al.* The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Molecular cell* **44**, 942-953, doi:10.1016/j.molcel.2011.11.008 (2011).

106 Schneider, E. V. *et al.* The structure of CDK8/CycC implicates specificity in the CDK/cyclin family and reveals interaction with a deep pocket binder. *Journal of molecular biology* **412**, 251-266 (2011).

107 Ansari, S., He, Q. & Morse, R. Mediator complex association with constitutively transcribed genes in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 16734-16739, doi:10.1073/pnas.0905103106 (2009).

108 Gonzalez, D. *et al.* Suppression of Mediator is regulated by Cdk8-dependent Grr1 turnover of the Med3 coactivator. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 2500-2505, doi:10.1073/pnas.1307525111 (2014).

109 Malik, S., Gu, W., Wu, W., Qin, J. & Roeder, R. G. The USA-derived transcriptional coactivator PC2 is a submodule of TRAP/SMCC and acts synergistically with other PCs. *Molecular cell* **5**, 753-760 (2000).

110 Malik, S., Baek, H., Wu, W. & Roeder, R. G. Structural and functional characterization of PC2 and RNA polymerase II-associated subpopulations of metazoan Mediator. *Molecular and cellular biology* **25**, 2117-2129, doi:10.1128/MCB.25.6.2117-2129.2005 (2005).

111 Cevher, M. *et al.* Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nature structural & molecular biology* **21**, 1028–1034, doi:10.1038/nsmb.2914 (2014).

112 Li, Y. *et al.* Yeast global transcriptional regulators Sin4 and Rgr1 are components of mediator complex/RNA polymerase II holoenzyme. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 10864-10868, doi:10.1073/pnas.92.24.10864 (1995).

113 Davis, J., Takagi, Y., Kornberg, R. & Asturias, F. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Molecular cell* **10**, 409-415, doi:10.1016/S1097-2765(02)00598-1 (2002).

114 Bernecky, C., Grob, P., Ebmeier, C., Nogales, E. & Taatjes, D. Molecular architecture of the human Mediator-RNA polymerase II-TFIIF assembly. *PLoS biology* **9**, doi:10.1371/journal.pbio.1000603 (2011).

115 Cramer, P. A tale of chromatin and transcription in 100 structures. *Cell* **159**, 985-994 (2014).

116 Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443-1444 (2014).

117 Leitner, A. *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Molecular & cellular proteomics* **9**, 1634-1649, doi:10.1074/mcp.R000001-MCP201 (2010).

118 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

119 Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308 (1999).

120 Seizl, M. *et al.* A conserved GA element in TATA-less RNA polymerase II promoters. *PloS one* **6**, doi:10.1371/journal.pone.0027595 (2011).

121 Budisa, N. *et al.* High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *European journal of biochemistry / FEBS* **230**, 788-796, doi:10.1111/j.1432-1033.1995.0788h.x (1995).

122 Kabsch, W. XDS. *Acta crystallographica. Section D, Biological crystallography* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).

123 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

124 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

125 Karplus, P. & Diederichs, K. Linking crystallographic model and data quality. *Science (New York, N.Y.)* **336**, 1030-1033, doi:10.1126/science.1218231 (2012).

126 Chen, V. B., Arendall, W. B. & Headd, J. J. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **D66**, 12-21, doi:10.1107/S0907444909042073 (2009).

127 Sheldrick, G. M. A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography* **A64**, 112-122, doi:10.1107/S0108767307043930 (2007).

128 Schneider, G. & Lindqvist, Y. Ta6Br14 is a useful cluster compound for isomorphous replacement in protein crystallography. *Acta Cryst.* **D50**, 186-191 (1994).

129 Cramer, P. *et al.* Architecture of RNA polymerase II and implications for the transcription mechanism. *Science (New York, N.Y.)* **288**, 640-649 (2000).

130 Knäblein, J. *et al.* Ta6Br(2+)12, a tool for phase determination of large biological assemblies by X-ray crystallography. *Journal of molecular biology* **270**, 1-7 (1997).

131 Girard, E., Stelter, M., Vicat, J. & Kahn, R. A new class of lanthanide complexes to obtain high-phasing-power heavy-atom derivatives for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* **59**, 1914-1922 (2003).

132 Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods in molecular biology* **364**, 215-230, doi:10.1385/1-59745-266-1:215 (2007).

133 Terwilliger, T. C. Maximum-likelihood density modification. *Acta crystallographica. Section D, Biological crystallography* **56**, 965-972, doi:10.1107/S0907444900005072 (2000).

134 Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta crystallographica. Section D, Biological crystallography* **D60**, 2210-2221, doi:10.1107/S0907444904016427 (2004).

135 Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics* **Chapter 5**, 5.6.1-5.6.30, doi:10.1002/0471250953.bi0506s15 (2006).

136 Petrotchenko, E. & Borchers, C. ICC-CLASS: isotopically-coded cleavable crosslinking analysis software suite. *BMC bioinformatics* **11**, 64, doi:10.1186/1471-2105-11-64 (2010).

137 Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54-61, doi:10.1038/nature06107 (2007).

138 Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* **33**, W244-W248 (2005).

139 Larivière, L. *et al.* Structure of the Mediator head module. *Nature* **492**, 448-451, doi:10.1038/nature11670 (2012).

140 Larivière, L. *et al.* Model of the Mediator middle module based on protein cross-linking. *Nucleic acids research* **41**, 9266-9273, doi:10.1093/nar/gkt704 (2013).

141 Sydow, J. F. *et al.* Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Molecular cell* **34**, 710-721, doi:10.1016/j.molcel.2009.06.002 (2009).

142 Kireeva, M., Lubkowska, L., Komissarova, N. & Kashlev, M. Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from Saccharomyces cerevisiae. *Methods Enzymol.* **370**, 138-155 (2003).

143 Herzog, F. *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337**, 1348-1352, doi:10.1126/science.1221483 (2012).

144 Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nature protocols* **9**, 120-137, doi:10.1038/nprot.2013.168 (2014).

145 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology* **152**, 36-51 (2005).

146 Korinek, A., Beck, F., Baumeister, W., Nickell, S. & Plitzko, J. M. Computer controlled cryo-electron microscopy--TOM² a software package for high-throughput applications. *Journal of structural biology* **175**, 394-405, doi:10.1016/j.jsb.2011.06.003 (2011).

147 Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods* **10**, 584-590, doi:10.1038/nmeth.2472 (2013).

148 Eibauer, M. *et al.* Unraveling the structure of membrane proteins in situ by transfer function corrected cryo-electron tomography. *Journal of structural biology* **180**, 488-496, doi:10.1016/j.jsb.2012.09.008 (2012).

149 Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157**, 38-46, doi:10.1016/j.jsb.2006.05.009 (2007).

150 Chen, Y., Pfeffer, S., Hrabe, T., Schuller, J. & Förster, F. Fast and accurate reference-free alignment of subtomograms. *Journal of structural biology* **182**, 235-245 (2013).

151 Hrabe, T. *et al.* PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *Journal of structural biology* **178**, 177-188, doi:10.1016/j.jsb.2011.12.003 (2012).

152 Scheres, S. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* **180**, 519-530, doi:10.1016/j.jsb.2012.09.006 (2012).

153 Shaikh, T. *et al.* SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature protocols* **3**, 1941-1974, doi:10.1038/nprot.2008.156 (2008).

154 Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of structural biology* **142**, 334-347 (2003).

155 Cramer, P., Bushnell, D. & Kornberg, R. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863-1876, doi:10.1126/science.1059493 (2001).

156 Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *Journal of structural biology* **184**, 226-236 (2013).

157 Penczek, P. A., Yang, C., Frank, J. & Spahn, C. M. T. Estimation of variance in single-particle reconstruction using the bootstrap technique. *Journal of structural biology*, 168-183 (2006).

158 Penczek, P. A., Renka, R. & Schomberg, H. Gridding-based direct Fourier inversion of the three-dimensional ray transform. *JOSA A* **21**, 499-509, doi:10.1364/JOSAA.21.000499 (2004).

159 Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophysical reviews* **2**, 21-27, doi:10.1007/s12551-009-0026-3 (2010).

160 Pettersen, E. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

161 Cheung, A. C. M., Sainsbury, S. & Cramer, P. Structural basis of initial RNA polymerase II transcription. *The EMBO journal* **30**, 4755-4763, doi:10.1038/emboj.2011.396 (2011).

162 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781-1802, doi:10.1002/jcc.20289 (2005).

163 Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry* **4**, 187-217, doi:10.1002/jcc.540040211 (1983).

164 Robinson, P. J., Bushnell, D., Trnka, M., Burlingame, A. & Kornberg, R. Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17931-17935, doi:10.1073/pnas.1215241109 (2012).

165 Sun, M. *et al.* Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome research* **22**, 1350-1359, doi:10.1101/gr.130161.111 (2012).

166 Sun, M. *et al.* Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Molecular cell* **52**, 52-62, doi:10.1016/j.molcel.2013.09.010 (2013).

167 Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033, doi:10.1126/science.1218231 (2012).

168 Kaufmann, R. *et al.* Infantile cerebral and cerebellar atrophy is associated with a mutation in the MED17 subunit of the transcription preinitiation mediator complex. *American journal of human genetics* **87**, 667-670, doi:10.1016/j.ajhg.2010.09.016 (2010).

169 Lee, Y. & Kim, Y. Requirement for a functional interaction between mediator components Med6 and Srb4 in RNA polymerase II transcription. *Molecular and cellular biology* **18**, 5364-5370 (1998).

170 Lee, T. *et al.* Interplay of positive and negative regulators in transcription initiation by RNA polymerase II holoenzyme. *Molecular and cellular biology* **18**, 4455-4462 (1998).

171 Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C. & Werner, M. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* **331**, 1451-1454, doi:10.1126/science.1200188 (2011).

172 Tan, Q., Linask, K., Ebright, R. & Woychik, N. Activation mutants in yeast RNA polymerase II subunit RPB3 provide evidence for a structurally conserved surface required for activation in eukaryotes and bacteria. *Genes & development* **14**, 339-348 (2000).

173 Lawrence, C. M. *et al.* The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes & Development* **12**, 45-54, doi:10.1101/gad.12.1.45 (1998).

174 Näär, A. M., Taatjes, D. J., Zhai, W., Nogales, E. & Tjian, R. Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. *Genes & development* **16**, 1339-1344, doi:10.1101/gad.987602 (2002).

175 Petrotchenko, E., Serpa, J. & Borchers, C. An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics. *Molecular & cellular proteomics : MCP* **10**, doi:10.1074/mcp.M110.001420 (2011).

176 Lee, Y., Park, J., Min, S., Han, S. & Kim, Y. An activator binding module of yeast RNA polymerase II holoenzyme. *Molecular and cellular biology* **19**, 2967-2976 (1999).

177 Eichner, J., Chen, H.-T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *The EMBO journal* **29**, 706-716, doi:10.1038/emboj.2009.386 (2010).

178 Zhang, Y. *et al.* Structural Basis of Transcription Initiation. *Science* **338**, 1076-1080, doi:10.1126/science.1227786 (2012).

179 Edwards, A. M., Kane, C. M., Young, R. A. & Kornberg, R. D. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *The Journal of biological chemistry* **266**, 71-75 (1991).

180 Ranish, J. A., Yudkovsky, N. & Hahn, S. Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes & Development* **13**, 49-63, doi:10.1101/gad.13.1.49 (1999).

181 Linder, T., Zhu, X., Baraznenok, V. & Gustafsson, C. The classical srb4-138 mutant allele causes dissociation of yeast Mediator. *Biochemical and Biophysical Research Communications* **349**, 948-953, doi:10.1016/j.bbrc.2006.08.099 (2006).

182 Schulz, D., Pirkl, N., Lehmann, E. & Cramer, P. Rpb4 Subunit Functions Mainly in mRNA Synthesis by RNA Polymerase II. *The Journal of biological chemistry* **289**, 17446-17452, doi:10.1074/jbc.M114.568014 (2014).

183 Engel, C., Sainsbury, S., Cheung, A., Kostrewa, D. & Cramer, P. RNA polymerase I structure and transcription regulation. *Nature* **502**, 650-655, doi:10.1038/nature12712 (2013).

184 Célia, J. & François, R. Kin28 regulates the transient association of Mediator with core promoters. *Nature structural & molecular biology* **21**, 449-455, doi:10.1038/nsmb.2810 (2014).

185 Kornberg, R. D. Mediator and the mechanism of transcriptional activation. *Trends in biochemical sciences* (2005).

186 Blattner, C. *et al.* Molecular basis of Rrn3-regulated RNA polymerase I initation and cell growth. *Genes & Development* **25**, 2093-2105 (2011).

187 Peyroche, G. *et al.* The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *The EMBO journal* **19**, 5473-5482, doi:10.1093/emboj/19.20.5473 (2000).

188 Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology* **AOP**, 1-12, doi:10.1038/nrm3951 (2015).

189 Firestein, R. *et al.* CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* **455**, 547-551, doi:10.1038/nature07179 (2008).

190 Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497-501, doi:10.1038/nature11884 (2013).

191 Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. & Cramer, P. Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity. *The EMBO Journal* **30**, 1302–1310, doi:10.1038/emboj.2011.64 (2011).

192 Koon, W., Yi, J. & Kevin, S. TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Molecular cell* **54**, 1-12, doi:10.1016/j.molcel.2014.03.024 (2014).

193 Galburt, E. A. *et al.* Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**, 820-823, doi:10.1038/nature05701 (2007).

194 Cianfrocco, M. A. *et al.* Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* **152**, 120-131, doi:10.1016/j.cell.2012.12.005 (2013).

195 Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950-953, doi:10.1126/science.1229386 (2013).

196 Serpa, J. *et al.* Using multiple structural proteomics approaches for the characterization of prion proteins. *Journal of proteomics*, doi:10.1016/j.jprot.2012.10.008 (2012).

197 Vizcaíno, J. A., Deutsch, E. W., Wang, R. & Csordas, A. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature biotechnology* **32**, 223-226, doi:10.1038/nbt.2839 (2014).

198 Bai, X.-c. C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends in biochemical sciences* **40**, 49-57, doi:10.1016/j.tibs.2014.10.005 (2015).

199 Asano, S. *et al.* Proteasomes. A molecular census of 26S proteasomes in intact neurons. *Science* **347**, 439-442, doi:10.1126/science.1261197 (2015).

200 Suloway, C. *et al.* Automated molecular microscopy: The new Leginon system. *Journal of Structural Biology* **151**, 41-60, doi:10.1016/j.jsb.2005.03.010 (2005).

201 Fischer, N. *et al.* Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature*, doi:10.1038/nature14275 (2015).

202 Wood, C. *et al.* Collaborative Computational Project for Electron cryo-Microscopy. *Acta crystallographica. Section D, Biological crystallography* **71**, 123-126, doi:10.1107/S1399004714018070 (2014).

203 Russel, D. *et al.* Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS biology* **10**, e1001244, doi:10.1371/journal.pbio.1001244 (2011).