Published in final edited form as:

Nat Struct Mol Biol. 2014 September ; 21(9): 810–816. doi:10.1038/nsmb.2873.

Architecture of the *S. cerevisiae* RNA polymerase I Core Factor complex

Bruce A. Knutson[1,3], **Jie Luo**[2], **Jeffrey Ranish**[2], and **Steven Hahn**[1,*]

[1]Division of Basic Sciences Fred Hutchinson Cancer Research Center Seattle, WA 98109, USA

[2]Institute for Systems Biology Seattle, WA 98109, USA

Abstract

Core Factor (CF) is a conserved RNA polymerase (Pol) I general transcription factor and is comprised of Rrn6, Rrn11, and the TFIIB-related subunit Rrn7. CF binds TBP, Pol I, and the regulatory factors Rrn3 and UAF. We used chemical crosslinking-mass spectrometry (CXMS) to determine the molecular architecture of CF and its interactions with TBP. The CF subunits assemble through an interconnected network of interactions between five structural domains that are conserved in orthologous subunits of the human Pol I factor SL1. The crosslinking-derived model was validated through a series of genetic and biochemical assays. Our combined results show the architecture of CF and the functions of the CF subunits in assembly of the complex. We extend these findings to model how CF assembles into the Pol I preinitiation complex, providing new insight into the roles of CF, TBP and Rrn3.

Introduction

Each eukaryotic RNA polymerase (Pol I-III) requires their own unique set of general transcription factors (GTFs) to recruit Pol to promoter DNA, respond to regulatory signals, and accurately initiate transcription. For Pol I, four main GTFs orchestrate the assembly of the Pol I preinitiation complex (PIC): the UAS-binding upstream activation factor (UAF), the promoter-binding complex termed Core Factor (CF), TATA binding protein (TBP) and the regulatory GTF Rrn3 [1,2]. Yeast CF is comprised of three subunits: Rrn6, Rrn11, and the TFIIB-related subunit Rrn7 [1-5]. The human CF ortholog, termed Selectivity Factor 1 (SL1), contains five subunits: TAF1A (TAFI48), TAF1B (TAFI63), TAF1C (TAFI110) orthologous to yeast Rrn11, Rrn7, and Rrn6 [1,2,6], metazoan-specific TAF1D (TAFI41) [7], and the shared SAGA and TFIID subunit TAFII12 [8].

*Corresponding author, Phone 206-667-5261, Fax 206-667-6497, shahn@fhcrc.org.

[3]Current address (as of August 2014): Department of Biochemistry and Molecular Biology SUNY Upstate Medical University Syracuse, NY 13210, USA

Author contributions

B.A.K, J.L., J.R, and S.H. designed the experiments. B.A.K. performed all structural modeling, genetic assays, and biochemical experiments. B.A.K and J.L. performed the crosslinking experiments, and J.L. performed the database searches for the CXMS analysis. B.A.K. and S.H. prepared the manuscript. All authors discussed and interpreted the data and approved the manuscript.

Accession Codes

4c2m, 3tj1, 3nfi, 4bbr, 2oit, 3hym, 1ytb

The Pol I PIC is formed by a network of protein-DNA and protein-protein interactions between Pol I, the Pol I GTFs, and the rDNA promoter. Yeast rDNA promoters consist of two *cis*-regulatory elements: the upstream activating sequence (UAS) and the core element (CE) [1,2]. UAF and CF target these promoter elements during the initial steps in PIC assembly [1,2]. Pol I is found in two populations: an initiation incompetent Pol I dimer, and an initiation competent form bound by Rrn3 that acts as a bridge between Pol I and the GTFs [9-11]. Rrn3 folds into an elongated HEAT repeat superhelix that binds to the Pol I A43 subunit in the stalk domain and wraps itself along the side of Pol I passing near the Rpa190 dock domain and the Rpb3 paralog AC40 [12]. Rrn3 also binds DNA, thereby adding to the important protein-DNA interactions necessary for Pol I PIC formation [13]. During a typical yeast rDNA transcription cycle, UAF, TBP, and CF interact with each other and promoter DNA and recruit the Pol I-Rrn3 complex [1,2,14]. Human rDNA promoters are also bipartite, consisting of an Upstream Control element (UCE) and a Core element (CE) that are targeted by Upstream Binding factor and SL1, respectively [1,2,6].

Pol I has an evolutionarily conserved structural and functional composition with its Pol II and III counterparts [15,16]. All three eukaryotic Pols contain a core set of five shared subunits, Rpb5, 6, 8, 10, and 12. Pols I and III share two additional subunits, AC40 and AC19, that are paralogs of Pol II subunits Rpb3 and Rpb11[17,18]. The remaining Pol I-specific subunits are paralogs to both Pol II and III subunits or Pol II GTFs [17-19]. Pol I contains several unique structural features that include a DNA-mimicking or extender loop that lies within the cleft, a connector loop that physically links Pol I dimers, and a flexible bridge helix [15,16,20]. Each of these unique elements likely plays a role in regulating Pol I transcription.

From previous work, there is limited information about Pol I PIC architecture and the molecular functions of the Pol I GTFs. For example, the structure of CF, the mechanism of CF interaction with Pol I and TBP, and the function of CF during transcription initiation are all unknown. We also know little about the architecture of the human CF ortholog SL1, a promising therapeutic target for cancer treatment using molecules that inhibit its DNA binding function [21-25]. To overcome these challenges and to learn more about the conserved functional and structural framework of CF, we employed an integrated approach to determine the molecular architecture of CF, and map structural and architectural domains essential for function. Based on this work, we derive a model for the role of CF and its orthologous factor SL1 in Pol I PIC formation and transcription initiation.

Results

Predicted structural domains within Core Factor subunits

Rrn7 is predicted to have domain architecture similar to TFIIB and Brf1 [1,26,27] with the Rrn7 N-terminal half containing two structured regions: a zinc ribbon domain (ZR) and tandem cyclin fold repeats. An unstructured region analogous to the TFIIB B-reader (BR) and B-linker (BL) regions connects these structured domains **(Fig. 1, Supplementary Fig. 1a)**. The Rrn7 C-terminal domain (CTD) contains a large Pol I-specific region that is similar to the Brf1-CTD in terms of size and secondary structure but unrelated in sequence [1,26]. For Rrn6 and Rrn11, we detected high probability matches with protein repeat motifs that often

function as protein-protein interaction interfaces. Rrn11 contains a predicted stretch of 7 tetratricopeptide repeats (TPR) consisting of short antiparallel alpha helices that typically stack into a superhelical structure [28,29] **(Fig. 1, Supplementary Fig 1b)**. The N- and C-terminal portions of Rrn11 lack similarity with known protein structures, and the N-terminal domain contains a large segment with high degree of predicted disorder **(Supplementary Fig. 1b)**. For Rrn6, the N-terminal half contains a large region homologous to WD40 β-propeller proteins [30,31], while the C-terminal half contains a region that matches several helical bundle proteins **(Fig 1c, Supplementary Fig. 1c)**. The N- and C-terminal ends of Rrn6 lack similarity to known structural domains. As shown below, these domains all act as major structural features important for CF assembly.

Chemical Crosslinking and Mass spectrometry of CF

We coexpressed three yeast CF subunits from a single vector in bacteria and purified recombinant CF for use in crosslinking analysis. **(Fig. 2)**. Rrn7 and Rrn6, were tagged with His_6 for initial purification by Ni-sepharose followed by cation-exchange and size exclusion chromatography, resulting in a pure stoichiometric complex **(Fig. 2b).** The purified recombinant CF restored Pol I transcription activity of an Δ*rrn7* yeast extract that is deficient in CF function **(Fig. 2c).** Adding back recombinant GST-Rrn7 alone could not restore transcription activity (**Supplementary Fig. 2**).

We next used a combined chemical crosslinking and mass-spectrometry (CXMS) approach to map protein-protein interactions within CF, where the complex was crosslinked with the homo-bifunctional reagent BS3, digested with trypsin, and the resulting peptides were analyzed by mass spectrometry. We analyzed the MS data with crosslink database search algorithms pLink[32] and Nexus to identify crosslinked peptides and assemble a linkage map of all the crosslinked peptides within and between the CF protein subunits. The BS3 crosslinker reacts with accessible primary amine groups found in lysine side chains and at the N-terminus that are within a theoretical maximum Cα-Cα distance of 30 Å [30,31,33-36]. Each CF subunit contains 35 or more lysine residues that are well distributed throughout the polypeptides with the only exceptions being the Rrn11-CTD, and Rrn7-ZR and -BR domains that are devoid of lysine residues **(Supplementary Fig. 1)**, making CF a good substrate for the CXMS approach.

Incubation of CF with increasing concentrations of BS3 crosslinker showed that all three subunits crosslink with 1:1:1 stoichiometry, resulting in a complex that migrates near the predicted mass sum of the entire complex (~200 KDa) **(Fig. 2d)**. MS analysis detected a total of 71 intramolecular and 39 intermolecular crosslinks within and between the CF subunits, respectively **(Fig. 3, Supplementary Table 1, 2)**. Among them, 43 were identified by both pLink and Nexus; 33 were identified by pLink only and 34 were identified by Nexus only. We detected extensive crosslinking between each of the CF subunits, suggesting an interconnected interaction network where each CF subunit contacts the other two subunits. This conclusion agrees with coexpression studies of different CF subunit combinations in bacteria where all pairwise combinations formed stable heterodimers that can be isolated by His_6 affinity chromatography **(Fig. 2e)**. Together, these studies indicate that CF assembles through an extensive network of protein–protein interactions.

Intramolecular crosslinking within CF subunits

We used the intramolecular crosslinking results within the predicted CF domains to validate the crosslinking approach and the molecular modeling of these domains. The predicted Rrn11 TPR domain contains 17 intramolecular crosslinks, and 14 of these crosslinks are within the structure model. All of the Cα-Cα distances between crosslinked lysine pairs are within or very close to the theoretical maximum crosslinking distance for BS3 **(Fig. 4a,e)**. The predicted Rrn6-WD40 domain contains 13 intramolecular crosslinks of which 12 could be measured in the model and each of these crosslinks are within the theoretical BS3 crosslinking distance **(Fig. 4b,e)**. We also detected 2 intramolecular crosslinks within the first repeat of the Rrn7 cyclin repeat domain that are both within allowable BS3 distance constraint **(Fig. 4c,e)**.

Because of limited sequence similarity to HB domains of known structure, it was difficult to obtain a reasonable sequence only-based homology model. Instead, we used Robetta[37] to generate models for the Rrn6-HB domain and then used the 10 intramolecular crosslinks as restraints to select the model with the lowest cumulative measured Cα-Cα distance where all crosslinks were within the theoretical BS3 crosslinking distance **(Fig. 4d)**. Our combined results support the models for the structured CF domains and validate the crosslinking approach.

We also detected extensive intramolecular crosslinking between regions lacking a predicted 3D structure **(Fig. 3, Supplementary Table 1)**. For example, we identified 5 crosslinks within the Rrn11-NTD and 10 between the Rrn11-NTD and -TPR domains as well as 3 crosslinks between the Rrn6-WD40 and HB domains that suggest their close proximity. Finally, we identified 9 crosslinks within the Rrn7-CTD, and one crosslink between the Rrn7-CTD and the BL domain.

Intermolecular crosslinking between CF subunits and domains

Many of the intermolecular crosslinks are between the predicted CF structured domains **(Fig. 3, Supplementary Table 2)**. We detected 10 interdomain crosslinks between the Rrn11-TPR and Rrn6-WD40 domains, suggesting that these two structural domains likely interact. TPR and WD40 protein-protein interactions are relatively common and have been solved in atomic detail [38,39], giving precedence for this interaction within the CF complex. We also detected three crosslinks between the Rrn11-TPR and Rrn6-HB domain. Given that the Rrn6-WD40 and HB domains are in close proximity as measured by their interdomain crosslinking **(Fig. 3, Supplementary Table 1)**, this positions the Rrn11-TPR domain so that it can interact with both of these Rrn6 domains.

We also detected an extensive crosslinking network for the Rrn7-CTD **(Fig. 3, Supplementary Table 2)**. The last five Rrn7-CTD lysine residues crosslink to the Rrn11-NTD and -TPR domains, and the Rrn6-WD40 and -HB domains. This positions the Rrn7-CTD at the center of all these potential domain contacts. A total of 5 crosslinks were observed between the Rrn7 cyclin repeat domain and the Rrn6-HB domain, suggesting an interaction between these two domains.

CF domains important for growth and complex integrity

BS3 crosslinking yields information on proximity of two amino acids within a protein or protein complex, but not necessarily information on protein-protein interaction. To test the functional relevance of the crosslinking results, we created a series of CF subunit variants to determine their impact on cell growth and CF complex integrity. Rrn7 deletions of the entire cyclin fold domain, smaller deletions of the cyclin repeats, and CTD domain truncations were all lethal and unable to associate with Rrn6 and Rrn11 **(Supplementary Fig. 3, summarized in Fig 5)**. This suggests that the cyclin folds and CTD are important protein-protein interaction surfaces for Rrn6 and Rrn11.

There is a strong correlation between the lethal growth phenotypes and the inability of Rrn7 to associate with Rrn6 and Rrn11 **(Fig. 5)**. Exceptions to this are the lethal N-terminal domain deletions (ZR, BR, BL) that still bind Rrn6 and Rrn11. The Rrn7-BL domain is predicted to be flexible and is dispensable for CF integrity, so the crosslinking with the Rrn6-WD40 and -HB domains is likely not functionally relevant. These findings are consistent with previous studies showing that the Taf1B and Brf1 N-termini are dispensable for SL1 and TFIIIB complex integrity and promoter recruitment but are necessary for transcription activity during a post-polymerase recruitment step [1,26,27,40,41].

The Rrn7 CTD contains seven tandem helical segments based on secondary structure predictions [26]. Rrn11 crosslinks near Rrn7 helical segment H3 and deletion of this region is lethal, and specifically disrupts the association with Rrn11 but not Rrn6 **(Supplementary Fig. 3)**. This phenotype is in excellent agreement with the CF crosslinking results where the region near helical segment H3 crosslinks to only Rrn11. Deletion of Rrn7 helical segments H6 and H7 were also lethal and both are required for association with Rrn6 and Rrn11, again in agreement with the crosslinking results **(Supplementary Fig. 3, Fig. 3)**. Finally, targeted deletions of helical segments H1, H2, H4, and H5 gave rise to viable yeast where deletions H2 and H5 showed a slow growth phenotype. All these deletions are competent for formation of CF **(Supplementary Fig. 3b)**.

Rrn11 deletion growth and CF integrity phenotypes also correlate well with the CF crosslinking network **(Fig. 5)**. For instance, deletions of either the N-terminal portion of the Rrn11-NTD that exhibits extensive crosslinking, TPR 1-3, or TPR 4-7 were all lethal, while deletion of the CTD domain gave wild-type growth **(Supplementary Fig. 4a)**. All of the individual TPR deletions were inviable with the exception of TPR 7 at the end of the tandem TPR stretch. Crosslinking was observed in TPR 1, 2, 6, and 7, which agrees well with the deletion growth phenotypes showing that TPRs 1-6 are essential. CF complex integrity analysis of the Rrn11 deletion variants is also consistent with the CF crosslinking network. Individual deletions of TPR repeats 1-5 exhibit reduced association with Rrn6 and Rrn7 **(Supplementary Fig. 4b)**, indicating a defect in CF complex assembly. An exception is the Rrn11-TPR repeat 6 deletion, and the Rrn11-NTD deletions that still form an intact CF complex but are lethal, indicating the importance of these regions for transcriptional functions outside CF complex assembly such as interaction with other Pol I initiation and/or regulatory factors.

Deletions of the Rrn6-NTD, -WD40, and -HB domains resulted in lethal growth phenotypes, and disrupted CF complex assembly **(Fig. 5, Supplementary Figs. 5)**. Smaller deletions within the WD40 domain indicated that all but WD40 repeat 2 are necessary for CF complex integrity **(Supplementary Fig. 5b)**. Rrn6-WD40 domain deletions are specific to Rrn11 but not Rrn7, since these deletions variants remain associated with Rrn7. Although, we detected no crosslinking within or with the Rrn6-NTD, it is essential for growth and the two most N-terminal Rrn6-NTD deletions were defective in Rrn11 association **(Supplementary Fig. 5c,d)**. This phenotype is similar to the Rrn6-WD40 domain deletions and may indicate that the Rrn6-Rrn11 interaction interface extends beyond the WD40 domain and into the NTD domain. Overall, our findings agree with the extensive Rrn11-TPR and Rrn6-WD40 crosslinking network and indicate that these two domains function as a major protein interface for CF assembly.

Smaller Rrn6-HB deletions resulted in range of growth phenotypes. Deletion of helical patch H2 grew at wild-type levels and formed an intact CF complex, while deletion of patches H3 and H4 resulted in slower growth phenotypes and impacted CF integrity evident by reduced levels of both Rrn7 and Rrn11 **(Supplementary Fig. 5a,b)**. Deletion of helical patches H1 and H5 were lethal and exhibited severe reductions in their association with Rrn7 and Rrn11. These findings suggest that the Rrn6-HB domain is essential for CF assembly and is consistent with the Rrn6-HB crosslinking network with Rrn7 and Rrn11. Finally, the Rrn6-CTD is neither required for growth nor CF complex integrity, consistent with the absence of crosslinking **(Supplementary Fig. 5a,b)**. In summary, there is a strict correlation between CF complex integrity and cell growth, and nearly all of the crosslinking overlaps the essential CF domains and regionsrequired for CF complex assembly **(Fig.5).**

Recruitment of CF variants to the rDNA promoter

We identified five regions of CF dispensable for complex integrity but essential for cell growth, indicating their potential importance for transcriptional activity outside of CF assembly. An important role for CF is to facilitate Pol I PIC formation. We used chromatin IP to determine whether any of these CF mutants retain the ability to bind the rDNA core promoter. As expected, CF mutants defective in CF complex integrity are similarly defective in rDNA recruitment, indicating that an intact CF complex is required for promoter recruitment **(Supplementary Fig. 6)**. Lethal CF mutants that still form an intact complex are also defective in rDNA promoter recruitment, indicating their importance in Pol I PIC formation, possibly through roles in interacting with other Pol I factors and/or promoter DNA. One exception is the modest 50% reduction in Rrn7-ZRΔ recruitment that mirrors its similarly reduced protein levels. This is analogous to the Rrn7 ortholog TAF1B and paralog Brf1 ZRΔ mutants that are still recruited to their respective promoters but are defective for post-recruitment functions[27,40,41].

A model for the structure of CF

We used a combination of crosslinking, bioinformatics, and protein modeling to derive a topological model of protein-protein interactions required for assembly of CF **(Fig. 6, Supplementary Fig. 7)**. The model shows the predicted protein-protein contacts between each of the CF domains and the interconnectedness of CF assembly. We used interdomain

crosslinks to orientate the CF domain models so that all of the crosslinks satisfy the BS3 crosslinker distance constraints **(Fig. 6b,c)**. The Rrn6-WD40 and Rrn11-TPR contacts predicted by the crosslinking are the most intimate in that the TPR cradles the WD40 domain where individual TPR repeats are in close contact with the WD40 repeats. The Rrn6-HB lies at the junction of all three essential CF domains where it simultaneously contacts the Rrn7 N-terminal cyclin repeat and CTD in addition to the Rrn6-WD40, and Rrn11-TPR domains. This arrangement explains why targeted deletions within the HB domain are lethal and disrupt CF complex integrity, and helps explain the loss of Rrn11 interaction due to the Rrn7-CR deletion. Rrn6 and Rrn11 make extensive contacts and disrupting the Rrn6-HB and Rrn7-CR domain interaction would likely result in loss of both proteins. The model predicts that CF forms a globular structure with flexible termini that emanate from a central core composed of interdomain contacts between the Rrn6-WD40, Rrn6-HB, Rrn11-TPR, and Rrn7-CR and -CTD domains.

Mapping interactions between the Rrn7 CTD and other CF domains, suggests a key function for the Rrn7-CTD in CF stability **(Supplementary Figs. 7a,b)**. Strikingly, the crosslinks in Rrn7-CTD helical segments H6 and H7 are clustered within the central region at the predicted junction between each of the CF domains and suggest that the C-terminal end of the Rrn7-CTD provides an additional surface to link the CF domains together. In addition, Rrn7-CTD lysines found in helical segments H1 and in the linker between H2 and H3 crosslink with two lysines within the first three Rrn11-TPR repeats. These interactions likely act as a separate interaction interface between the Rrn7-CTD and Rrn11-TPR domain as these regions are also required for complex integrity. Although there is no sequence similarity between the Pol-specific domains of Rrn7 and Brf1, the function of the Rrn7 CTD appears analogous to the Brf1-CTD, which is essential for coordinating interactions with other TFIIIB subunits [42-44].

CF-TBP crosslinking

Little is known about the interactions between TBP and the Pol I GTFs. Previous protein interaction assays suggest that each CF subunit can independently interact with TBP, but that the strongest TBP interactions are with Rrn6 [45,46]. To examine the interactions between TBP and CF, we crosslinked CF-TBP complexes to identify regions of these factors in close proximity. We detected 23 intermolecular crosslinks between TBP and CF subunits that clustered into several essential domains within the CF subunits **(Supplementary Table 3, Fig. 7)**. We also detected 23 intramolecular crosslinks within TBP **(Supplementary Table 3, Fig. 7)**. We mapped seven of these intramolecular crosslinks to the TBP structure and all fall within the BS3 distance constraints **(Supplementary Fig. 8b,c)**. Crosslinks between the TBP N-terminus and CF subunits are less informative since several reports suggest that the TBP N-terminus is not required for rDNA transcription nor CF and SL1 interaction[47].

We detected 10 crosslinks between the TBP core domain and CF domains with predicted structures (**Fig. 7a, Supplementary Fig. 8a**). These crosslinks suggest that several CF domains are in close proximity to TBP that include the Rrn6-HB and -WD40 domains, Rrn11-TPR and -NTD domains, and Rrn7-CR and -CTD domains. Highlighting the lysine residues within the CF model that crosslink to TBP show that they all reside on a continuous

surface that accommodates TBP and DNA **(Supplementary Fig. 8d,e,f)**. Together, our results agree with previous reports showing that all three CF subunits and corresponding SL1 subunits make specific and unique contacts with TBP.

Discussion

Here we describe the molecular architecture of the essential and conserved Pol I CF complex. We discovered that CF assembles through an interconnected network of protein-protein interactions through essential structural domains found in each of the CF subunits. The CF CXMS studies are in excellent agreement with a systematic series of genetic and biochemical results. Our combined results lead to a topological model of the CF complex that explains the role of many structural domains within CF and provides insight into assembly of the Pol I PIC. The coupling of crosslinking, biological assays, and modeling further highlight the usefulness of integrating multiple approaches to understand the architecture and function of macromolecular protein complexes that are often difficult to decipher by a single experimental approach.

Model for architecture of the Pol I PIC

The macromolecular architectures of the Pol II and III PICs have been deduced at varying degrees of resolution and may give clues to the architecture of the Pol I PIC. A common theme between the Pol II and III PIC architectures is the position of the TFIIB and Brf1 cyclin fold repeats and their position with respect to Pol, TBP and promoter DNA. Using the Pol II and III PIC models as a guide[12,48-53], we created a Pol I PIC model containing the Pol I and Rrn3 structures with TBP, DNA, and our Rrn7 homology model. A striking feature of this Pol I PIC model is an empty canyon between Rrn7 and Rrn3 where Pol I forms the floor of the canyon, and the Rrn7 cyclin folds, DNA, and TBP form one wall, and Rrn3 forms the opposing wall **(Fig. 8a)**. Rrn7 lysines within the cyclin domain that crosslink to the Rrn6-HB domain are accessible and face the empty canyon pointed toward Rrn3, and the Rrn6-HB can be positioned within the canyon and satisfy the BS3 crosslinking distance constraints **(Fig. 8b)**. This model explains how Rrn6 mediates simultaneous interactions with Rrn7, Rrn3, and TBP.

In our Pol I PIC model, the Rrn6-HB domain lies above the Rrn3 HEAT repeats along the side of Pol I. Previous studies have shown that Rrn6 directly interacts with Rrn3, and that the Rrn6 C-terminus is sufficient for this interaction and is essential for growth[11]. The tested Rrn6 C-terminal fragment overlaps extensively with the essential H5 region of the Rrn6-HB that we demonstrate is also important for cell growth, CF complex integrity, and rDNA promoter recruitment, while the remainder of the Rrn6 C-terminal region is non-essential. These results suggest that Rrn6-H5 mediates interaction with Rrn3. This agrees well with our positioning of Rrn6-H5 close to Rrn3, and explains the direct interaction between Rrn3 and Rrn6. The Rrn6-HB domain is also positioned close to TBP in the model, which can explain the Pol I-specific interaction and crosslinking between Rrn6 and TBP[1,4,5,45]. Together, the Pol I PIC modeling suggests that the Rrn6-HB acts a protein-protein interaction hub that mediates interactions within the CF complex and between Rrn3 and TBP.

The Pol I PIC modeling also has implications for an additional role of Rrn3 in that it helps coordinate interactions between CF and Pol I by anchoring Rrn6 and possibly the entire CF complex into the transcription competent configuration within the PIC. This mechanism helps explain why the absence of Rrn3 in the Pol I PIC results in inactive complexes[14,54]. Furthermore, placement of the CF model into the Pol I PIC also positions portions of the Rrn11-TPR and Rrn6-WD40 near the PIC periphery **(Fig. 8c),** where they may interact with other upstream Pol I regulatory factors such as TBP, UAF, and possibly the rDNA promoter. This mechanism can explain the defects in the Rrn6 and Rrn11 deletions that result in lethal growth phenotypes, but retain CF complex integrity. The Pol I PIC model will help direct future studies employing biochemical and structural approaches to describe these regulatory interactions in greater detail.

Conservation between yeast CF and human SL1 complexes

SL1 subunits TAF1A, B, and C are predicted to contain identical structural features as the orthologous CF subunits[1], suggesting that the SL1 and CF complexes have similar architecture. However, we note there are several intriguing differences between CF and SL1 subunits that will be important to address in future studies. For instance, Rrn7 contains a metazoan-specific insertion within the first cyclin fold repeat that is rich in phosphorylatable residues [1,26]. Given that the Rrn7 cyclin fold domain is essential for CF complex integrity, the insertion may play a regulatory role that impinges upon SL1 assembly or provide additional sites for interaction with other Pol I factors. The N- and C-termini of TAF1A and TAF1C are also poorly conserved with their CF counterparts, so these domains may also have species-specific functions in Pol I transcription. Further, it is unclear how the SL1-specific subunits TAF1D and TAFII12 interact with TAF1A, B, and C.

Pol I transcription is upregulated in cancer cells but the precise molecular mechanism is not clear and targeting the DNA binding activity of SL1 has promising anti cancer activity [21-25]. Our study provides the first detailed and systematic analysis of the yeast SL1 ortholog that will help guide future studies into the structural and mechanistic roles of CF and SL1 and how they can be targeted to treat cancer and disease.

Online Methods

CF expression and purification

CF was expressed and purified using a similar strategy as described in [56] with several modifications detailed below. Briefly, recombinant CF protein was expressed in BL21(DE3)RIL cells (Stratagene) by auto-induction in Zy5052 media (25 mM Na_2HP0_4, 25 mM KH2PO4, 50 mM NH_4Cl, 5 mM $NaSO_4$, 2 mM $MgSO_4$, 0.5 % glycerol, 0.05% glucose, 0.2% α-lactose, 0.01% tryptone, 0.005% yeast extract) for 16 hrs at 37°C [57]. Cells were harvested by centrifugation and cell pellets were washed once with Extraction buffer (50 mM HEPES pH 8.0, 500 mM KCl, 20 mM Imidazole, 5 mM $MgCl_2$, 0.1 mM EDTA, 20% Glycerol, 1 mM TCEP) supplemented with protease inhibitors. Cell pellets were resuspended in 5 ml extraction buffer per 1 gram of cells, and lysed with 1 mg/ml lysozyme (Sigma) for 30 min on ice followed by sonication. Lysates were cleared by centrifugation, and cleared lysate was added to Ni-sepharose affinity medium (GE Healthcare) and

incubated for 4 hrs at 4°C. CF-bound beads were washed twice with extraction buffer containing 1 M KCl and twice with extraction buffer with 0.2 M KCl. CF was eluted with elution buffer (50 mM HEPES pH 8.0, 200 mM KCl, 200 mM Imidazole, 5 mM $MgCl_2$, 0.1 mM EDTA, 20% Glycerol, 1 mM TCEP). Peak elutions were pooled and further purified over a Heparin column (GE Healthcare) using a linear gradient of buffer A (50 mM HEPES pH 8.0, 200mM KCl, 5 mM MgCl2, 0.1 mM EDTA, 5% Glycerol, 1 mM TCEP) to buffer B (50 mM HEPES pH 8.0, 1 M KCl, 5 mM MgCl2, 0.1 mM EDTA, 5% Glycerol, 1 mM TCEP) over 20 column volumes. CF eluted at approximately 500 mM KCl and was subsequently filtered and concentrated on a 100 KDa cut off centricon filter (Millipore) and further purified by size exclusion SE-200 (GE Healthcare) chromatography in buffer A. Peak elutions of CF were pooled and was stored at −80°C.

Pol I transcription competent extracts

Pol I transcription competent extracts were prepared as previously described with minor modifications[58]. Briefly, 1 liter of wild-type or *rrn7*Δ strains were grown to an OD 1.0 in YP media containing galactose as a carbon source. Cells were harvested by centrifugation and washed twice with cold extraction buffer (100 mM HEPES KOH pH 7.5, 250 mM KCl, 5 mM EGTA, 1 mM EDTA, 2.5 mM DTT) supplemented with protease inhibitors. Cell pellets were weighed, resuspended in 1.5 ml extraction buffer per gram of cells, and loaded into a syringe. The cell paste was then extruded onto a weigh boat floating over liquid nitrogen to freeze the cells, and then stored at −80°C. Frozen cells were broken by manual grinding with a mortar and pestle. The resulting cell powder was thawed and another 1.5 volumes of extraction buffer per gram of cells was added to the ground cells. Broken cells were cleared by centrifugation for 3 hours at 100,000 × g, and the resulting supernatant was dialyzed for a total of 6 hrs in dialysis buffer (20 mM HEPES-KOH pH 7.9, 50 mM KCl, 5 mM EGTA, 0.05 mM EDTA, 2.5 mM DTT, 20% glycerol) supplemented with protease inhibitors, and then stored in aliquots at −80°C. Extracts contained approximately 20 mg/ml protein.

In vitro transcription and primer extension

Transcription assays were as previously described with minor modifications [59]. Briefly, assays were assembled in 50 μl reactions containing 15 μl extract, 1 μl RNase Out (Invitrogen), 100 μM each NTP, 200 μg plasmid template DNA, 10 mg/ml α-amanitin, 12.5 μL 4X transcription buffer (80 mM HEPES-KOH pH 7.9, 200 mM KCl, 40 mM MgCl2, 20 mM EGTA, 0.2 mM EDTA, 10 mM DTT, 40% glycerol). Transcription reactions were incubated at room temperature for 30 min, and stopped by addition of 180 μl stop solution (100 mM NaOAc, 10 mM EDTA, 0.5% SDS) supplemented with 17 μg/ml tRNA (Sigma). RNA was extracted with phenol-chloroform (2:1) once, ethanol (EtOH) precipitated, and dried. RNA pellets were suspended in 10 μl primer annealing mix (5 mM Tris pH 8.3, 75 mM KCl, 1 mM EDTA, and 65 μM fluorescently (700IR) labeled LacI primer). Primer extension reactions were incubated for 45 min at 55°C followed by addition of 20 μl cDNA synthesis mix (25 mM Tris pH 8.3, 75 mM KCl, 4.5 mM MgCl2, 15 mM DTT, 150 μM deoxynucleoside triphosphates (dNTPs), 100 units MMLV reverse transcriptase (Invitrogen)). Reactions were incubated for 30 min at 37°C and were stopped by EtOH precipitation. Pellets were washed with 80% EtOH, dried, resuspended in 3 ul RNase A (40

μg/ml), and incubated for 3 min at room temperature followed by addition of 3 μl formamide loading dye containing bromophenyl blue. Resuspended pellets were heated for 1 min at 90°C, cooled on ice, and run on a denaturing 7% urea acrylamide gel. Gels were visualized by an Odyssey scanner (700 IR, Li-Cor).

CF sequence analysis and structural modeling

PSIpred v3.0 (http://bioinf.cs.ucl.ac.uk/psipred/) was used to predict CF subunit secondary structure [60]. DISOpred2 was used to predict CF subunit disordered regions [61]. Confidence values for the predictions were plotted as area graphs for each residue position. Structure similarity searches were performed by HHpred (http://toolkit.tuebingen.mpg.de/hhpred) using HMM PDB under default settings and thresholds [62]. Rrn7-NTD, Rrn6-WD40, and Rrn11-TPR domains were searched by HHpred to identify and align homologous proteins of known structure. PIR-alignments of the query and target sequences were generated by HHPred using the “create model” option in the HHPred results page and manually converted to FASTA format. Templates were selected based on the highest probability score and highest protein identity. These alignments were used to generate structural homology models by Modeller 9v10 using the UCSF chimera interface [63]. Ten models of each domain were generated and best scoring model was chosen. Sequence and templates used for structural modeling: Rrn7 residues 1-375 aligned with yeast TFIIB residues 1-345 (Pdb, 4bbr)[64], Rrn6-WD40 residues 163-559 aligned with human Nuclear pore subunit Nup214 residues 1-411 (Pdb, 2oit)[65], and Rrn11-TPR was aligned in two overlapping segments 129-320 and 208-401 were aligned to Anaphase-promoting complex subunit CDC26 TPR subdomain residues 229-431 (Pdb, 3hym)[66]. De novo protein structure of the Rrn6-HB domain residues 571-769 was generated by Robetta using default settings and thresholds [67]. The models are reasonable given their low protein identity scores and received good to fair scores on number of model evaluation algorithms that include: ProQ[68], QMEAN[69], SelectPro[70], and Verify3D[71]. Model quality results are listed in **Supplementary Table 5**. The rigid body protein docking program Patchdock[72] was used to guide fitting of the CF domains together and assembly of CF into the Pol I PIC. Default thresholds were employed with the exception that the crosslinked lysine Cα distance constraints were included as an added modeling constraint for the docking. The resulting models were then manually optimized to further refine the models and represent a single unique solution.

BS3 crosslinking and Mass spectrometry sample preparation

50 μg of CF was crosslinked with BS3 (Bis[sulfosuccinimidyl] suberate, Thermo Scientific) at a final concentration of 2 mM and 5 mM for 2 hrs at room temperature and then stored overnight at 4°C in 200 μl reactions. For CF-TBP crosslinking, 50 μg of CF was incubated with equal molar concentration of TBP for 2 hours at 4°C and then passed through an S200 MicroSpin column (GE Healthcare) to remove unbound TBP protein. The CF-TBP complex was then crosslinked with BS3 at a final concentration of 2 mM. An equal volume of Trifluoroethanol was added to the BS3-crosslinked protein complex and the sample was incubated at 60°C for 30 min to denature the proteins. The proteins were then reduced by addition of 5 mM TCEP for 30 min at 37°C. The denatured and reduced sample was alkylated with iodoacetamide at a 10 mM final concentration for 30 min in the dark at room temperature. The sample was diluted 10 fold with 20 mM Triethanolamine and the proteins

were digested with 1 μg of trypsin overnight at 37°C. The peptides were further purified on C18 spin columns (Nest Group), dried and resuspended in 5% Acetonitrile/0.1% TFA solution and either analyzed by MS or further fractionated by microcapillary SCX HPLC. For HPLC, peptides were loaded onto the column equilibrated in 5% Acetonitrile/0.1% TFA and washed with 20% Acetonitrile/0.1% Formic Acid followed by a gradient of 10% Acetonitrile/0.1 % Formic Acid containing 10-100 mM Ammonium Acetate. Remaining column bound peptides were eluted with 10% Acetonitrile/0.1% Formic acid containing 1 M Ammonium Acetate. Most crosslinked peptides were eluted between 50-100 mM Ammonium Acetate, and the peptide fractions were dried down.

Mass spectrometry analysis and crosslinked peptide identification

BS3-crosslinked peptides were analyzed on a Thermo Scientific Orbitrap Elite with HCD fragmentation and serial MS events that included one FTMS1 event at 30,000 resolution followed by 10 FTMS2 events at 15,000 resolution. Other instrument settings included: MS mass range greater than 1800; use m/z value as masses enabled; Charge state rejection: +1, +2, and unassigned charges; Monoisotopic precursor selection enabled; Dynamic exclusion enabled: repeat count 1, exclusion list size 500, exclusion duration 30s; HCD normalized collision energy 35%, isolation width 3Da, minimum signal count 5000; FTMS MSn AGC target 50,000. The RAW files were converted to mzXML files and analyzed by two different crosslink database searching algorithms: pLink [73] and in-house designed Nexus (details to be published elsewhere; J.L., and J.R.).

A protein database containing the forward and reversed sequences was used for the Nexus analysis with the following parameter settings: (a) up to three miscleavages; (b) static modification on Cysteines (+57.0215 Da); (c) differential oxidation modification on Methionines (+15.9949 Da); (d) differential modification on the peptide N-terminal Glutamic acid residues (-18.0106 Da) or N-terminal Glutamine residues (-17.0265 Da); (e) differential mono-BS3 modification on Lysine residue (+156.0806 Da). All possible tryptic peptide pairs within 20 ppm of the precursor mass are used for crosslinked peptide searches. For each candidate peptide pair, the theoretical b- and y- ion series for each peptide is compared to the acquired spectrum with a modification mass of either the partner peptide and the BS3 linker at the crosslinking site, or a crosslinker-derived Lysine immonium ion plus the BS3 linker (modification mass of 221.1416 Da) at the crosslinking site. The score is calculated as the sum of the weighted dot product of the weighted candidate ion series and the normalized intensity of each fragment ion within 60 ppm of the theoretical fragment ion (similar to the normalization procedure used by the Sequest algorithm[74]. The weighted ion matrix considers both the coexistence and connectivity of fragment ions: Each candidate ion is weighted 1 at a given charge state; if the preceding ion of the same charge state is present, the score is increased by 0.5; if not, the score is decreased by 0.5; the same rules apply to the subsequent ion in the ion series. If the weight of an ion equals zero, then a minimum weight of 0.5 is given. The highest candidate score is kept for each spectrum. The false-positive rate (FDR) is calculated as the number of identifications containing one decoy sequence (U) minus the number of identifications containing two decoy sequences (F) divided by the number of identifications containing no decoy sequences (T): FDR=(U-F)/T. This is similar to the FDR calculation used by pLink [73].

After performing the pLink and the Nexus analysis against a protein database containing the sequences of Core factors subunits and applying a 5% FDR cutoff, the search results are combined and each spectrum is manually evaluated for the quality of the match to each peptide using the COMET/Lorikeet Spectrum Viewer (TPP). The crosslinked peptides were considered confidently identified if at least 4 consecutive b or y ions for each peptide were observed and the majority of the observed ions are accounted for.

Yeast growth assays

Yeast strains (*MAT* α ade2-1 ade3::hisG ura3-1 his3-11 trp1-1 leu2-112 can1) contained chromosomal deletions of *RRN7*, *RRN6*, or *RRN11* genes and the plasmid pNOY103 that expresses the 35S ribosomal RNA transcript under control of the *GAL7* promoter [73,75,76]. These yeast strains were transformed with either wild-type or mutant CF subunit gene derivatives **(Supplementary Table 4)**. Cells were grown in galactose complete media lacking leucine and then streaked on glucose containing plates to monitor glucose sensitive phenotypes. Plates were incubated for 3 days at 30°C, and then assessed for growth relative to wild-type (+++).

Yeast CF immunoprecipitations

HA-tagged Rrn7, Myc-tagged Rrn6, or Myc-tagged Rrn11 were purified from 200 ml mid-log-phase cultures grown in minimal glucose containing media. For HA-tagged Rrn7 immunoprecipitation, either wild-type HA-tagged Rrn7, or deletion variant plasmid constructs were transformed into wild-type yeast strains containing chromosomally Flag-tagged Rrn6 or Rrn11. For Myc-tagged Rrn6 and Rrn11 immunoprecipations, either Rrn6 or Rrn11 wild-type and deletion variant plasmid constructs were transformed into wild-type yeast strains containing chromosomally tagged HA tagged Rrn7 and either flagged tagged Rrn11 or Rrn6, respectively. Cells were collected by centrifugation and lysed by bead beating in lysis buffer (100 mM Tris-HCl pH 7.9, 250 mM AmSO4, 1 mM EDTA, 10% glycerol) supplemented with 0.5 mM DTT, 0.5 mM PMSF, and protease inhibitors, and lysates were cleared by centrifugation. Approximately 4 mg of the whole-cell extract diluted with two volumes of dilution buffer (25 mM HEPES pH 7.5, 50 mM NaCl, 1 mM EDTA) was incubated overnight at 4°C with either 20 μl of Anti-HA affinity agarose (Sigma, E6779) or 20 μl of Anti-c-Myc affinity agarose (Sigma, A7470) for HA and Myc immunoprecipitations, respectively. Beads were washed 3 times in 1 ml 1X Tris-buffered saline (TBS) (50 mM Tris pH 7.5, 150 mM NaCl) containing 0.1% Tween, 2 times with 1× TBS, and then the proteins were eluted with 1X NuPAGE LDS sample buffer (Invitrogen) at 70°C for 10 min. Samples were resolved on 4-12% polyacrylamide gradient gels (Invitrogen) in 1× MOPs buffer (50 mM MOPS, 50 mM Tris base pH 7.7, 3.5 mM SDS, 1 mM EDTA) for 50 min at 200 V. Proteins were transferred to PVDF and probed with mouse monoclonal antibodies against Flag (Sigma, F1805), HA (Covance, MMS-101R), and Myc (Covance, MMS-150R).

Chromatin Immunoprecipitation

ChIP assays were conducted as previously described with a few modifications [77-79]. Briefly, yeast strains expressing various CF subunit variants were grown in GC-media lacking

leucine. Cells were crosslinked with 1% formaldehyde (Sigma) for 5 min at room temperature. After crosslinking, cells were harvested and washed twice with cold 1× TBS buffer. Cells were resuspended in ChIP buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% NP-40, 1.0% Triton X-100) supplemented with 0.5 mM DTT, 0.5 mM PMSF, and protease inhibitors. Cells were lysed with 0.5 mm zirconia beads (Biospec) using a Mini-Beadbeater-96 (Biospec) at 3 min intervals until greater than 95% cell breakage was achieved. Lysates were sonicated 4 × 10 min on high setting with the Bioruptor UCD-200 (Diagencode). Sonicated lysates were cleared by centrifugation and protein concentrations were determined by Bradford assay. Next, 500 μg of sheared chromatin was incubated overnight at 4°C with either 10 μl Anti-HA (Sigma, E6779) or Anti-Myc (Sigma, A7470) affinity agarose. Beads were washed three times in ChIP buffer containing 1M NaCl, once with standard ChIP buffer, and once with water. Next, 100 μl of 10% Chelex-100 (Bio-Rad) was added to beads and boiled for 10 min. The suspension was centrifuged and supernatant containing the eluted DNA was collected. The beads were washed with 60 μl water, centrifuged, and the supernatants were combined. Eluted DNA was used directly in quantitative PCR. To determine the immunoprecipitated ChIP signal, values were calculated as the ratio of the percent precipitated at a specific locus to the percent precipitated at the *POL1* reference locus. All values are expressed relative to wild -type set at 1.0. Experiments were performed in biological duplicate. Primers sets used for qPCR were previously described [79] and are listed as follows: Pro-rDNAF, 5'-TCGGCGAGAAATACGTAGTTAAG-3'; Pro-rDNA-R, 5'-CCTCACACTTGTACTCCATGAC-3'; ETS-rDNA-F, 5'-AATAGCCGGTCGCAAGACT-3'; ETS-rDNA-R, 5'-TCACGGAATGGTACGTTTGA-3'; 25S-rDNA-F, 5'-AGGATGCTGGCATAATGGTT-3'; 25S-rDNA-R, 5'-CACCCAAACACTCGCATAGA-3'; POL1-F, 5'-TTTCTGCTGAGGTGTCTTATAGAATTCA-3'; POL1-R, 5'-CGTTTGGGCCCATGCAT-3'.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank members of the S. Hahn lab (FHCRC), E. Young lab (Univ. Washington), and J. Ranish Lab (ISB) for the comments throughout the course of the work. We also thank the P. Gafkin and the FHCRC proteomics shared resources for assistance with mass spectrometry. We also thank the P. Cramer lab (Univ. Munich) for sharing an early release of their yeast Pol I crystal structure. This work was supported by grant NIH-NIGMS 2RO1GM053451 to S.H and NIH-NIGMS 2P50 GM076547 and NIH-NCI R21CA175849 to J.R.

References

1. Knutson BA, Hahn S. TFIIB-related factors in RNA polymerase I transcription. Biochim Biophys Acta. 2013; 1829:265–273. [PubMed: 22960599]
2. Schneider DA. RNA polymerase I activity is regulated at multiple steps in the transcription cycle: recent insights into factors that influence transcription elongation. Gene. 2012; 493:176–184. [PubMed: 21893173]
3. Bedwell GJ, Appling FD, Anderson SJ, Schneider DA. Efficient transcription by RNA polymerase I using recombinant core factor. Gene. 2012; 492:94–99. [PubMed: 22093875]

4. Lin CW, et al. A novel 66-kilodalton protein complexes with Rrn6, Rrn7, and TATA-binding protein to promote polymerase I transcription initiation in Saccharomyces cerevisiae. Mol Cell Biol. 1996; 16:6436–6443. [PubMed: 8887672]
5. Lalo D, Steffan JS, Dodd JA, Nomura M. RRN11 encodes the third subunit of the complex containing Rrn6p and Rrn7p that is essential for the initiation of rDNA transcription by yeast RNA polymerase I. J Biol Chem. 1996; 271:21062–21067. [PubMed: 8702872]
6. Russell J, Zomerdijk JC. The RNA polymerase I transcription machinery. Biochem Soc Symp. 2006:203–216. [PubMed: 16626300]
7. Gorski JJ, et al. A novel TBP-associated factor of SL1 functions in RNA polymerase I transcription. EMBO J. 2007; 26:1560–1568. [PubMed: 17318177]
8. Denissov S, et al. Identification of novel functional TBP-binding sites and general factor repertoires. EMBO J. 2007; 26:944–954. [PubMed: 17268553]
9. Milkereit P, Schultz P, Tschochner H. Resolution of RNA polymerase I into dimers and monomers and their function in transcription. Biol Chem. 1997; 378:1433–1443. [PubMed: 9461342]
10. Milkereit P, Tschochner H. A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. EMBO J. 1998; 17:3692–3703. [PubMed: 9649439]
11. Peyroche G, et al. The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. EMBO J. 2000; 19:5473–5482. [PubMed: 11032814]
12. Blattner C, et al. Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. Genes Dev. 2011; 25:2093–2105. [PubMed: 21940764]
13. Stepanchick A, et al. DNA binding by the ribosomal DNA transcription factor rrn3 is essential for ribosomal DNA transcription. J Biol Chem. 2013; 288:9135–9144. [PubMed: 23393135]
14. Aprikian P, Moorefield B, Reeder RH. New model for the yeast RNA polymerase I transcription cycle. Mol Cell Biol. 2001; 21:4847–4855. [PubMed: 11438642]
15. Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P. RNA polymerase I structure and transcription regulation. Nature. 2013; 502:650–655. [PubMed: 24153182]
16. Fernandez-Tornero C, et al. Crystal structure of the 14-subunit RNA polymerase I. Nature. 2013; 502:644–649. [PubMed: 24153184]
17. Vannini A. A structural perspective on RNA polymerase I and RNA polymerase III transcription machineries. Biochim Biophys Acta. 2013; 1829:258–264. [PubMed: 23031840]
18. Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell. 2012; 45:439–446. [PubMed: 22365827]
19. Geiger SR, et al. RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. Mol Cell. 2010; 39:583–594. [PubMed: 20797630]
20. Zomerdijk J. Structural biology: Pivotal findings for a transcription machine. Nature. 2013; 502:629–630. [PubMed: 24153180]
21. Bywater MJ, Pearson RB, McArthur GA, Hannan RD. Dysregulation of the basal RNA polymerase transcription apparatus in cancer. Nat Rev Cancer. 2013; 13:299–314. [PubMed: 23612459]
22. Bywater MJ, et al. Inhibition of RNA polymerase I as a therapeutic strategy to promote cancer-specific activation of p53. Cancer Cell. 2012; 22:51–65. [PubMed: 22789538]
23. Drygin D, et al. Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. Cancer Res. 2011; 71:1418–1430. [PubMed: 21159662]
24. Drygin D, Rice WG, Grummt I. The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. Annu Rev Pharmacol Toxicol. 2010; 50:131–156. [PubMed: 20055700]
25. Hannan KM, Sanij E, Rothblum LI, Hannan RD, Pearson RB. Dysregulation of RNA polymerase I transcription during disease. Biochim Biophys Acta. 2013; 1829:342–360. [PubMed: 23153826]
26. Knutson BA, Hahn S. Yeast Rrn7 and human TAF1B are TFIIB-related RNA polymerase I general transcription factors. Science. 2011; 333:1637–1640. [PubMed: 21921198]

27. Naidu S, Friedrich JK, Russell J, Zomerdijk JC. TAF1B is a TFIIB-like component of the basal transcription machinery for RNA polymerase I. Science. 2011; 333:1640–1642. [PubMed: 21921199]
28. Allan RK, Ratajczak T. Versatile TPR domains accommodate different modes of target protein recognition and function. Cell Stress Chaperones. 2011; 16:353–367. [PubMed: 21153002]
29. D'Andrea LD, Regan L. TPR proteins: the versatile helix. Trends Biochem Sci. 2003; 28:655–662. [PubMed: 14659697]
30. Stirnimann CU, Petsalaki E, Russell RB, Muller CW. WD40 proteins propel cellular networks. Trends Biochem Sci. 2010; 35:565–574. [PubMed: 20451393]
31. Xu C, Min J. Structure and function of WD40 domain proteins. Protein Cell. 2011; 2:202–214. [PubMed: 21468892]
32. Yang B, et al. Identification of cross-linked peptides from complex samples. Nat Methods. 2012; 9:904–906. [PubMed: 22772728]
33. Chen ZA, et al. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J. 2010; 29:717–726. [PubMed: 20094031]
34. Kalkhof S, Sinz A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. Anal Bioanal Chem. 2008; 392:305–312. [PubMed: 18724398]
35. Madler S, Bich C, Touboul D, Zenobi R. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. J Mass Spectrom. 2009; 44:694–706. [PubMed: 19132714]
36. Muller MQ, Sinz A. Chemical cross-linking and high-resolution mass spectrometry to study protein-drug interactions. Methods Mol Biol. 2012; 803:205–218. [PubMed: 22065227]
37. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic acids research. 2004; 32:526–531.
38. Chao WC, Kulkarni K, Zhang Z, Kong EH, Barford D. Structure of the mitotic checkpoint complex. Nature. 2012; 484:208–213. [PubMed: 22437499]
39. ter Haar E, Musacchio A, Harrison SC, Kirchhausen T. Atomic structure of clathrin: a beta propeller terminal domain joins an alpha zigzag linker. Cell. 1998; 95:563–573. [PubMed: 9827808]
40. Colbert T, Hahn S. A yeast TFIIB-related factor involved in RNA polymerase III transcription. Genes Dev. 1992; 6:1940–1949. [PubMed: 1398071]
41. Hahn S, Roberts S. The zinc ribbon domains of the general transcription factors TFIIB and Brf: conserved functional surfaces but different roles in transcription initiation. Genes Dev. 2000; 14:719–730. [PubMed: 10733531]
42. Kassavetis GA, Geiduschek EP. Transcription factor TFIIIB and transcription by RNA polymerase III. Biochem Soc Trans. 2006; 34:1082–1087. [PubMed: 17073756]
43. Khoo SK, Wu CC, Lin YC, Lee JC, Chen HT. Mapping the protein interaction network for the TFIIB-related factor Brf1 in the RNA polymerase III pre-initiation complex. Mol Cell Biol. 2013; 34:551–559. [PubMed: 24277937]
44. Schramm L, Hernandez N. Recruitment of RNA polymerase III to its target promoters. Genes Dev. 2002; 16:2593–2620. [PubMed: 12381659]
45. Steffan JS, Keys DA, Dodd JA, Nomura M. The role of TBP in rDNA transcription by RNA polymerase I in Saccharomyces cerevisiae: TBP is required for upstream activation factor-dependent recruitment of core factor. Genes Dev. 1996; 10:2551–2563. [PubMed: 8895657]
46. Steffan JS, Keys DA, Vu L, Nomura M. Interaction of TATA-binding protein with upstream activation factor is required for activated transcription of ribosomal DNA by RNA polymerase I in Saccharomyces cerevisiae in vivo. Mol Cell Biol. 1998; 18:3752–3761. [PubMed: 9632758]
47. Rudloff U, Eberhard D, Grummt I. The conserved core domain of the human TATA binding protein is sufficient to assemble the multisubunit RNA polymerase I-specific transcription factor SL1. Proceedings of the National Academy of Sciences of the United States of America. 1994; 91:8229–8233. [PubMed: 8058785]
48. Grunberg S, Hahn S. Structural insights into transcription initiation by RNA polymerase II. Trends Biochem Sci. 2013; 38:603–611. [PubMed: 24120742]

49. Grunberg S, Warfield L, Hahn S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. Nature structural & molecular biology. 2012; 19:788–796.
50. He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. Nature. 2013; 495:481–486. [PubMed: 23446344]
51. Murakami K, et al. Architecture of an RNA polymerase II transcription pre-initiation complex. Science. 2013; 342:1238724. [PubMed: 24072820]
52. Wu CC, et al. RNA polymerase III subunit architecture and implications for open promoter complex formation. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:19232–19237. [PubMed: 23132938]
53. Wu CC, Lin YC, Chen HT. The TFIIF-like Rpc37/53 dimer lies at the center of a protein network to connect TFIIIC, Bdp1, and the RNA polymerase III active center. Mol Cell Biol. 2011; 31:2715–2728. [PubMed: 21536656]
54. Cavanaugh AH, Evans A, Rothblum LI. Mammalian Rrn3 is required for the formation of a transcription competent preinitiation complex containing RNA polymerase I. Gene Expr. 2008; 14:131–147. [PubMed: 18590050]
55. Kim Y, Geiger JH, Hahn S, Sigler PB. Crystal structure of a yeast TBP/TATA-box complex. Nature. 1993; 365:512–520. [PubMed: 8413604]
56. Bedwell GJ, Appling FD, Anderson SJ, Schneider DA. Efficient transcription by RNA polymerase I using recombinant core factor. Gene. 2012; 492:94–99. [PubMed: 22093875]
57. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005; 41:207–234. [PubMed: 15915565]
58. Schultz MC, Choe SY, Reeder RH. Specific initiation by RNA polymerase I in a whole-cell extract from yeast. Proceedings of the National Academy of Sciences of the United States of America. 1991; 88:1004–1008. [PubMed: 1992452]
59. Fishburn J, Hahn S. Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. Mol Cell Biol. 2012; 32:12–25. [PubMed: 22025674]
60. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000; 16:404–405. [PubMed: 10869041]
61. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. Bioinformatics. 2004; 20:2138–2139. [PubMed: 15044227]
62. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic acids research. 2005; 33:244–248. [PubMed: 15647507]
63. Yang Z, et al. UCSF Chimera, MODELLER, and IMP: an integrated modeling system. J Struct Biol. 2012; 179:269–278. [PubMed: 21963794]
64. Sainsbury S, Niesser J, Cramer P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature. 2013; 493:437–440. [PubMed: 23151482]
65. Napetschnig J, Blobel G, Hoelz A. Crystal structure of the N-terminal domain of the human protooncogene Nup214/CAN. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:1783–1788. [PubMed: 17264208]
66. Wang J, Dye BT, Rajashankar KR, Kurinov I, Schulman BA. Insights into anaphase promoting complex TPR subdomain assembly from a CDC26-APC6 structure. Nature structural & molecular biology. 2009; 16:987–989.
67. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic acids research. 2004; 32:526–531.
68. Wallner B, Elofsson A. Can correct protein models be identified? Protein Sci. 2003; 12:1073–1086. [PubMed: 12717029]
69. Benkert P, Kunzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic acids research. 2009; 37:510–514.
70. Randall A, Baldi P. SELECTpro: effective protein model selection using a structure-based energy function resistant to BLUNDERs. BMC Struct Biol. 2008; 8:52. [PubMed: 19055744]
71. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science. 1991; 253:164–170. [PubMed: 1853201]

72. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic acids research. 2005; 33:363–367.

73. Yang B, et al. Identification of cross-linked peptides from complex samples. Nat Methods. 2012; 9:904–906. [PubMed: 22772728]

74. Eng JK, Fischer B, Grossmann J, Maccoss MJ. A fast SEQUEST cross correlation algorithm. J Proteome Res. 2008; 7:4598–4602. [PubMed: 18774840]

75. Nogi Y, Vu L, Nomura M. An approach for isolation of mutants defective in 35S ribosomal RNA synthesis in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America. 1991; 88:7026–7030. [PubMed: 1871118]

76. Nogi Y, Yano R, Nomura M. Synthesis of large rRNAs by RNA polymerase II in mutants of Saccharomyces cerevisiae defective in RNA polymerase I. Proceedings of the National Academy of Sciences of the United States of America. 1991; 88:3962–3966. [PubMed: 2023944]

77. Nelson J, Denisenko O, Bomsztyk K. The fast chromatin immunoprecipitation method. Methods Mol Biol. 2009; 567:45–57. [PubMed: 19588084]

78. Nelson JD, Denisenko O, Bomsztyk K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. Nat Protoc. 2006; 1:179–185. [PubMed: 17406230]

79. Zhang Y, et al. The SWI/SNF chromatin remodeling complex influences transcription by RNA polymerase I in Saccharomyces cerevisiae. PLoS One. 2013; 8:e56793. [PubMed: 23437238]

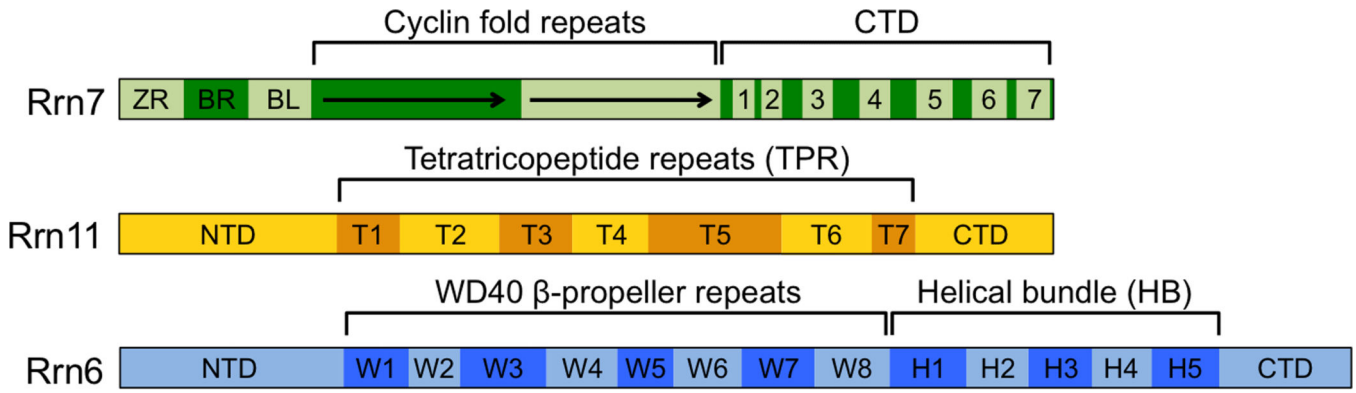

Figure 1. Predicted domain organization of Core Factor subunits

Domain maps of each CF subunit are shown for Rrn7 (Green), Rrn11 (Yellow-orange), and Rrn6 (Blue). ZR, zinc ribbon; BR, B-reader; BL, B-linker; CR, cyclin repeat. T, Tetratricopeptide repeat (TPR); W, WD40; NTD, N-terminal domain; CTD, C-terminal domain.

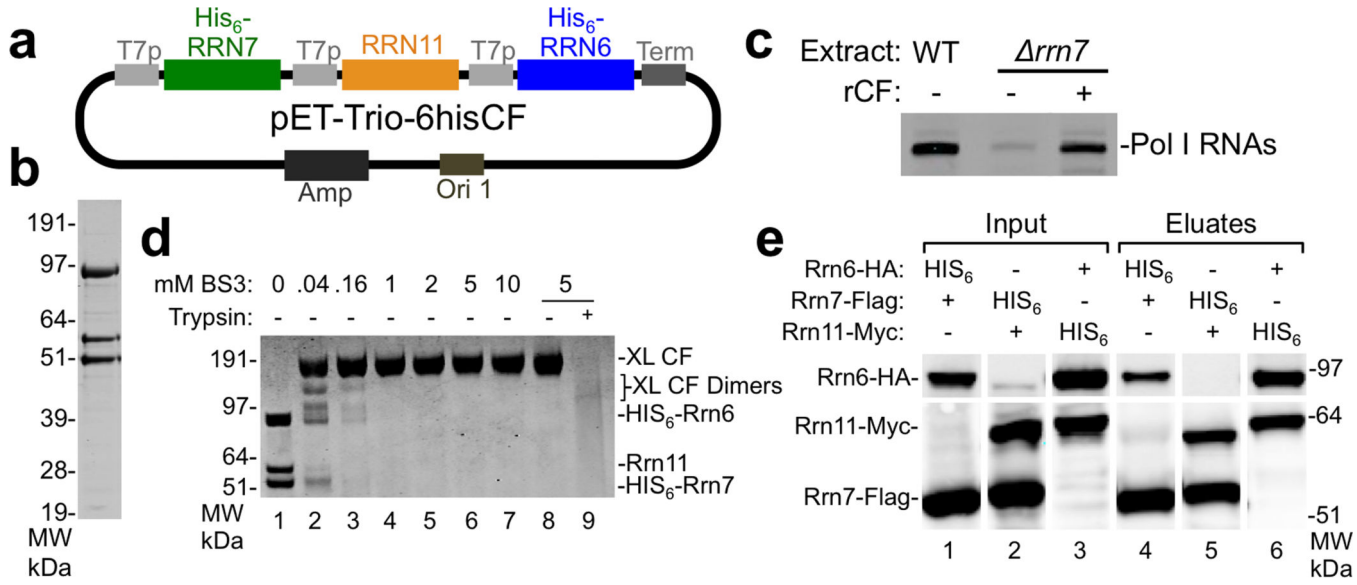

Figure 2. Expression, purification, and chemical crosslinking of active recombinant Core Factor

(a) Diagram of the single T7- vector to express recombinant Core Factor (rCF) heterotrimer. **(b)** Coomassie-stained gel of purified rCF. **(c)** In vitro transcription assays with a Pol I reporter plasmid and either wild-type (WT) or Δ*rrn7* yeast extracts in the presence or absence of rCF. Pol I transcripts were detected by primer extension. **(d)** Chemical crosslinking of rCF. The indicated concentrations of BS3 were incubated with rCF and products were visualized by SDS-PAGE and coomassie-staining. **(e)** Coexpression and association of pairwise CF subunit combinations. Shown are the inputs of coexpressed proteins and eluates from Ni-agarose pulldowns analyzed by SDS-PAGE and Western blots of the tagged CF subunits.

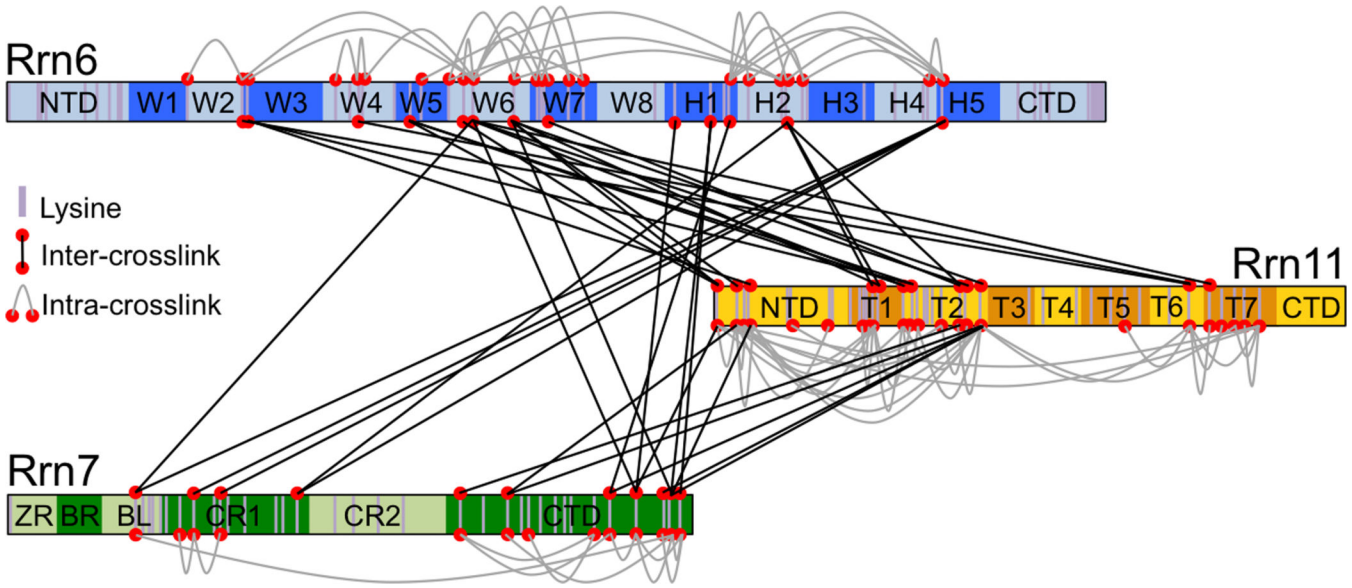

Figure 3. Core Factor crosslinking map

Linkage map of crosslinked CF lysine residues. Schematics of each CF subunit showing the predicted domain organization. Purple bars denote lysine positions and the N-terminal amine, while red spheres connected by dashed black lines indicate intra- and inter-molecular crosslinked lysine pairs.

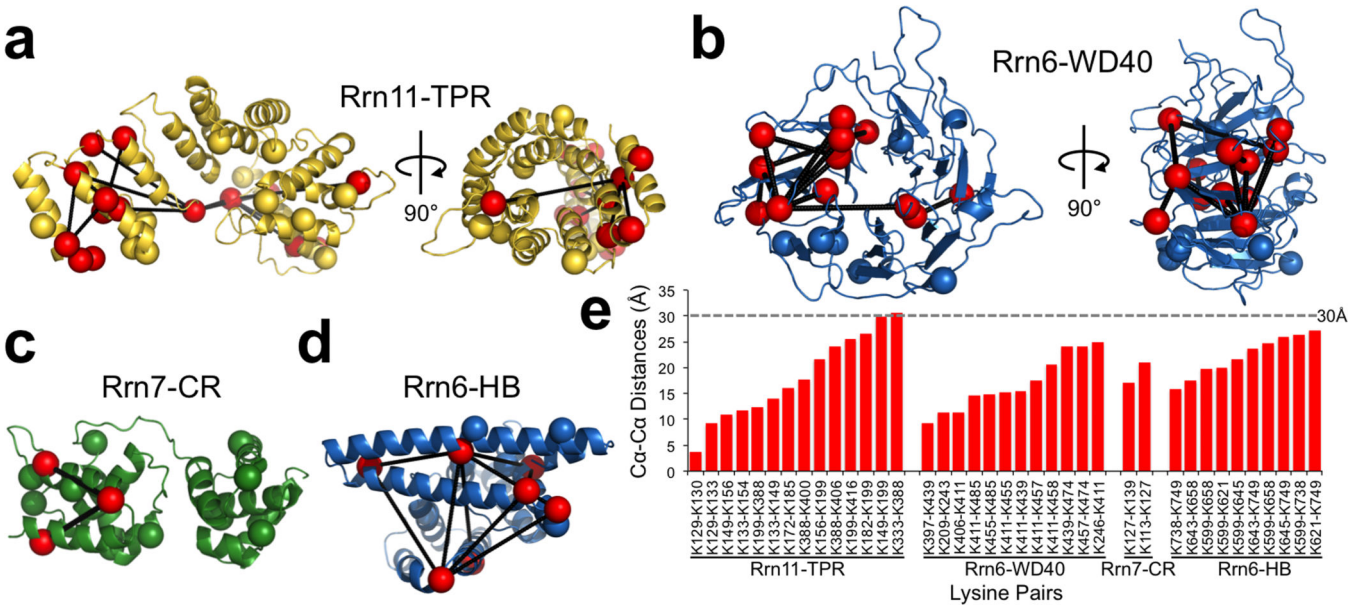

Figure 4. Intramolecular crosslinking within predicted structured domains of Core Factor

(a-d) Crosslinks within the structure models for: **(a)** Rrn11-TPR, **(b)** Rrn6-WD40, **(c)** Rrn7-CR, and **(d)** Rrn6-HB. Noncrosslinked lysine Cα atoms are depicted as spheres. Cα atoms of crosslinked lysine pairs are depicted as red spheres connected by black lines. **(e)** Calculated Cα-Cα distances between crosslinked lysine pairs within CF subunit domains. Dashed grey line denotes the theoretical maximum crosslinking distance for BS3 of 30 Å.

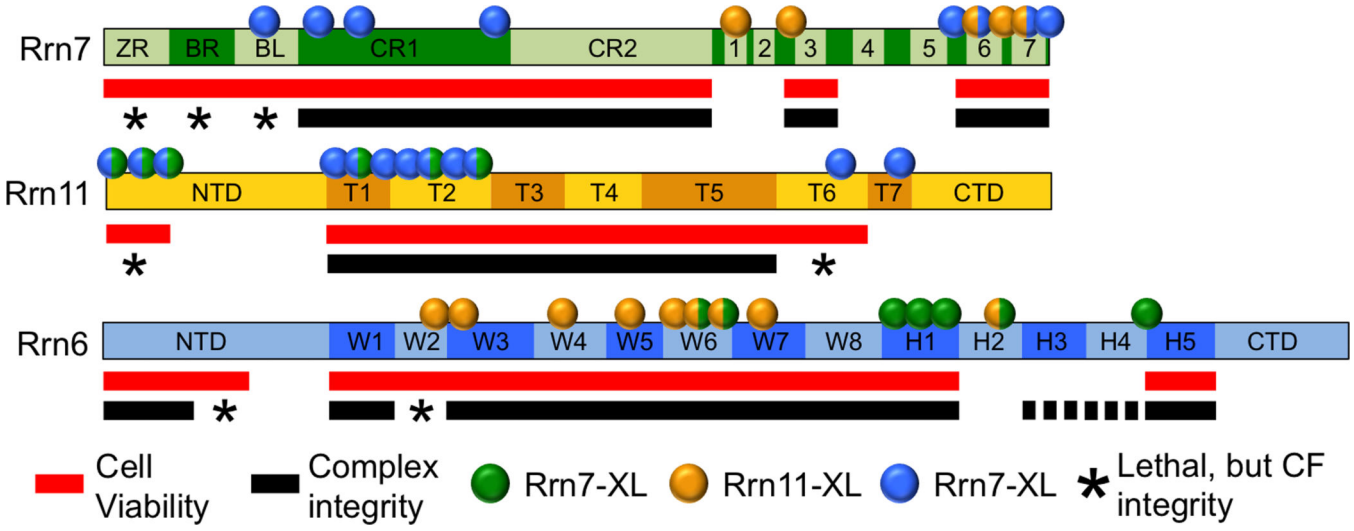

Figure 5. Summary of intermolecular crosslinking, genetic growth phenotypes, and biochemical integrity phenotypes

Domain maps of each CF subunit are shown. Spheres along the top of the maps indicate the approximate sites of BS3 crosslinking and are colored by crosslinking partner: green, Rrn7; orange, Rrn11; blue, Rrn6. Red and black bars below the maps indicate the domains and regions important for cell viability and/or CF complex integrity, respectively. Asterisks under the CF domain maps denote domains required for cell growth but dispensable for CF complex integrity.

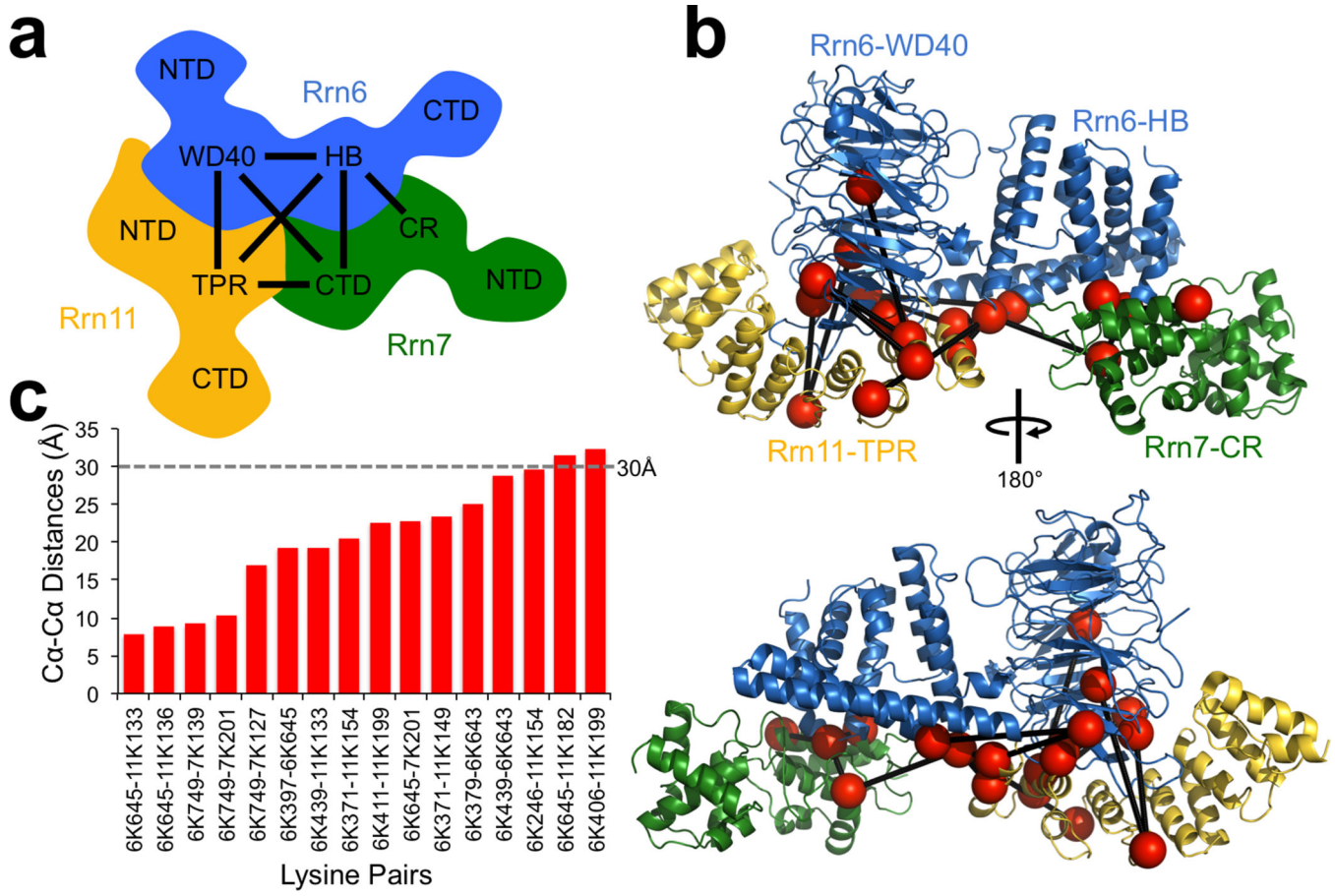

Figure 6. Molecular model of Core Factor

(a) Cartoon representation of the essential domain contacts within CF. Solid black lines between the CF subunit domains indicate a domain contact. **(b)** Model of CF architecture showing interactions between structured domains. Each CF subunit is arranged and positioned to satisfy their domain contacts based on crosslinking, genetic, and biochemical data. **(c)** Calculated Cα-Cα distances between crosslinked lysine pairs within CF subunit domains in CF model. Dashed grey line denotes theoretical maximum crosslinking distance for BS3 of 30 Å.

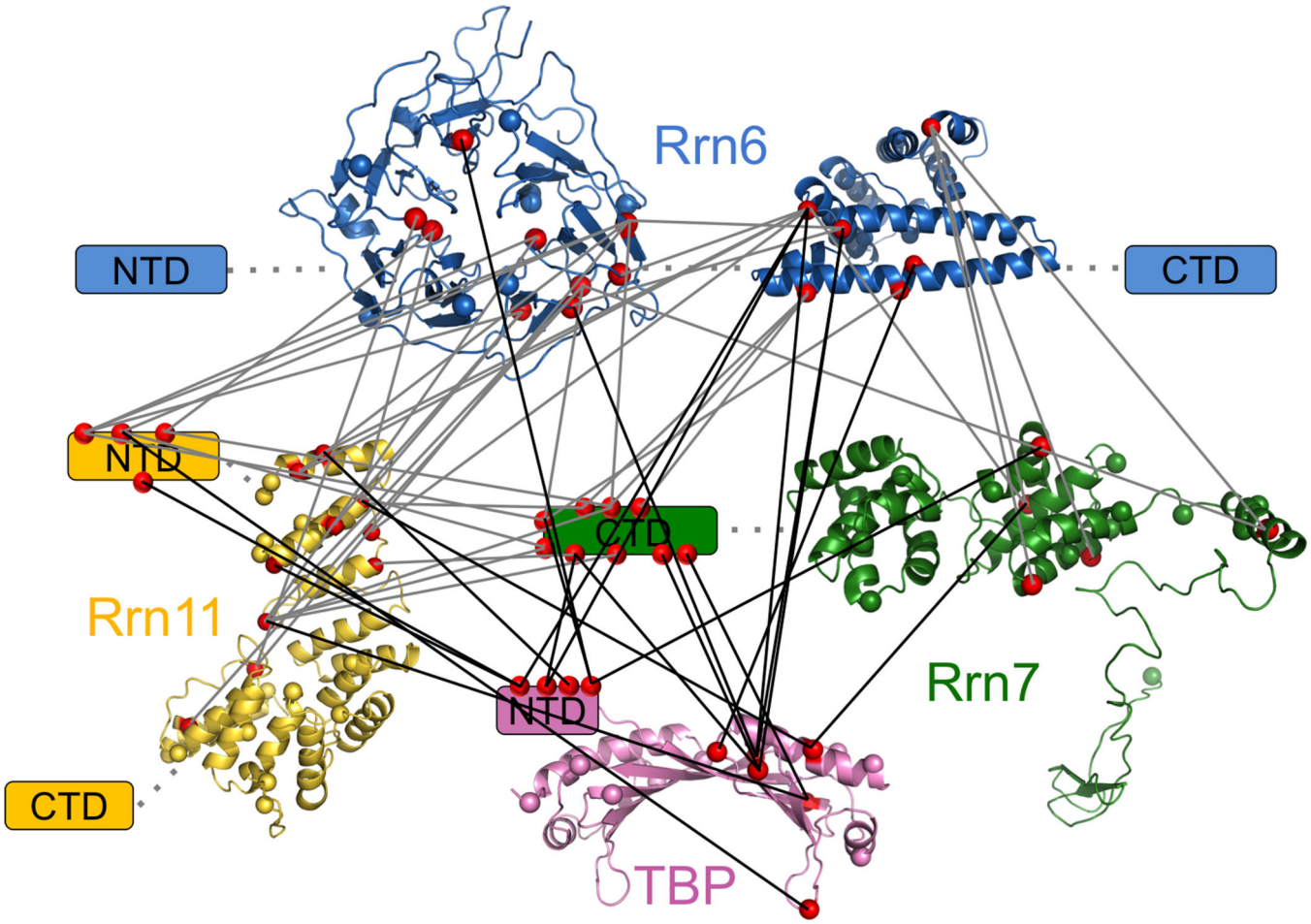

Figure 7. CF-TBP crosslinks

Topological linkage map showing all the observed intermolecular and interdomain crosslinks between the CF subunit domains and TBP. Non-crosslinked lysine Cα atoms are depicted as spheres in the same color as the domain. Crosslinked lysine pairs are shown as red spheres connected by grey or black lines. Gray lines denote crosslinks between CF domains and black lines denotes crosslinks between CF and TBP.

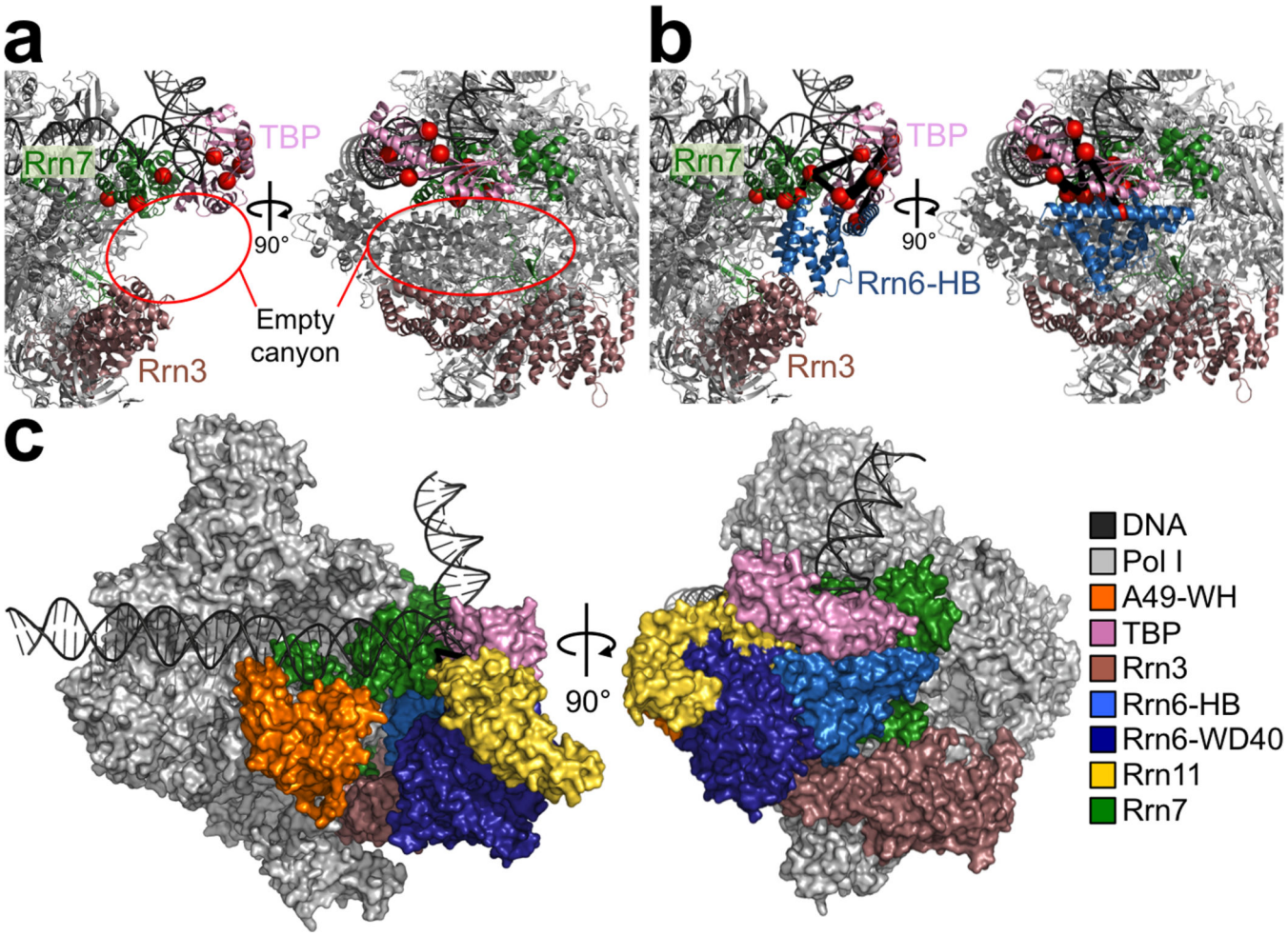

Figure 8. Model for the Pol I PIC

(a-c) The Pol I PIC model is based on current Pol II/III PIC models. The following structures were used to create the model: Pol I (Pdb, 4C2M)[15], Rrn3 (Pdb, 3TJ1)[12], A49-WH (Pdb, 3NFI)[19], TBP (Pdb, 1YTB)[55]. **(a)** The Pol I PIC model creates an empty canyon (circled in red) between the Rrn7-CR, TBP, and Rrn3. Cα atoms of lysines within the Rrn7-CR and TBP that crosslink to the Rrn6-HB are shown as red colored spheres **(b)** Positioning of the Rrn6-HB into the empty canyon guided by crosslinking restraints between the Rrn6-HB and Rrn7-CR domains and TBP. Crosslinked lysine pairs are shown as red spheres connected by black lines, and the Cα-Cα distances between each crosslinked lysine pair are below 23 Å. **(c)** Surface representation of the Pol I PIC model including the entire CF domain model.