

NAR Breakthrough Article

Reconstitution of the CstF complex unveils a regulatory role for CstF-50 in recognition of 3'-end processing signals

Wen Yang[†], Peter L. Hsu[†], Fan Yang, Jae-Eun Song and Gabriele Varani*

Department of Chemistry, University of Washington, Seattle, WA 98195-1700, USA

Received September 25, 2017; Revised November 02, 2017; Editorial Decision November 08, 2017; Accepted November 13, 2017

ABSTRACT

Cleavage stimulation factor (CstF) is a highly conserved protein complex composed of three subunits that recognizes G/U-rich sequences downstream of the polyadenylation signal of eukaryotic mRNAs. While CstF has been identified over 25 years ago, the architecture and contribution of each subunit to RNA recognition have not been fully understood. In this study, we provide a structural basis for the recruitment of CstF-50 to CstF via interaction with CstF-77 and establish that the hexameric assembly of CstF creates a high affinity platform to target various G/U-rich sequences. We further demonstrate that CstF-77 boosts the affinity of the CstF-64 RRM to the RNA targets and CstF-50 fine tunes the ability of the complex to recognize G/U sequences of certain lengths and content.

INTRODUCTION

Cleavage and polyadenylation (3'-end processing) at the 3'-end of genes is a crucial step in the maturation of mRNAs prior to packaging and export to the cytoplasm and for many non-coding RNAs as well. This co-transcriptional process is well conserved between yeast and humans and represents a critical stage of quality control, since improperly processed transcripts are subject to nuclear retention and degradation. While 3'-end processing is a relatively simple reaction, consisting of an endonucleolytic cleavage and subsequent polyadenylation at the newly formed free 3'-OH, several dozen proteins, assembled into various sub-

complexes, are required to execute and regulate this reaction (1,2).

Early fractionation experiments identified three complexes responsible for the recognition of sequence elements within the polyadenylation signals (3). The Cleavage and Polyadenylation Specificity Factor (CPSF), comprised of six different subunits, is responsible for the recognition of the nearly universal metazoan polyadenylation signal, AAUAAA, via WDR33 and CPSF30. It also coordinates both cleavage and polyadenylation reactions (4,5). Cleavage Factor I (CFI_m) recognizes a strong consensus sequence (UGUA, USE) upstream of the polyA site (6–8). Finally, the Cleavage Stimulation Factor (CstF) recognizes the downstream sequence element (DSE), which contains a conserved G/U-rich sequence, and plays a key role in polyA site selection (9). Unlike the polyA site and the USE, however, the DSE has no clear consensus sequence (10,11), making it somewhat counterintuitive to understand how it can be specifically recognized by CstF.

CstF is composed of three subunits of molecular weights 77, 64 and 50 kDa. The 77 kDa subunit is highly conserved, with homologs found in all eukaryotes. Crystal structures of CstF-77 show an N-terminal helical HAT domain [Half-a-TPR (tetratricopeptide repeat)] that dimerizes strongly and interacts directly with CPSF160, linking the two complexes to establish a connection between the DSE and polyA sites (12,13). In addition, CstF-77 acts as the scaffold around which the other two components of CstF assemble (14–16).

CstF-64 is the primary RNA binding component of the complex, with an RNA recognition motif (RRM) domain within the N-terminal ~100 residues. Both solution and crystallographic studies have shown that CstF-64 and its yeast homolog bind RNAs representative of the G/U-rich DSE with low μM affinity, irrespective of the detailed se-

*To whom correspondence should be addressed. Tel: +1 206 543 7113; Fax: +1 206 685 8665; Email: varani@chem.washington.edu

[†]These authors contributed equally to this work as first authors.

Present addresses:

Peter L. Hsu, Department of Pharmacology, University of Washington, Seattle, WA 98195, USA.

Jae-Eun Song, Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA.

quence composition, but discriminates against adenine and cytosine nucleotides (17,18). The hinge domain of CstF-64 binds to the CstF-77 C-terminal tail in a highly intertwined fashion, suggesting the two proteins have co-evolved to be stoichiometrically related to one another (15). The C-terminus of CstF-64 contains a small, but conserved domain that recruits other components of the 3'-end processing and termination complex (19). Like CstF-77, CstF-64 is also highly conserved, with identified homologs in all eukaryotes.

The smallest subunit, CstF-50 has been identified only in multicellular eukaryotes and has no known yeast homolog. It contains a predicted seven-bladed WD40 domain that interacts with CstF-77 (14), though how this occurs is unknown. In addition, its N-terminus contains a small domain that mediates homodimerization (20). However, CstF-50 has not been studied as much as the two other components and its precise biochemical role in the 3'-end processing reactions is unclear.

Since two of the three subunits of CstF form homodimers, it has been suggested that CstF forms a dimer of heterotrimers (12,20). However, no confirmation of the hexamerization of the CstF complex has been reported and only limited biochemical studies on the holocomplex have been reported. Here we present the *in vitro* reconstitution of the CstF complex from recombinant sources and demonstrate that the assembly is hexameric, containing two copies of each subunit. Using structural and biochemical techniques, we demonstrate the structural basis of CstF-50 recruitment to the CstF complex and dissect the contribution of each subunit to RNA recognition.

MATERIALS AND METHODS

Protein preparation

The CstF-64 RRM domain and all human CstF-77 constructs were cloned into a modified pET-28a (Novagen) vector with a N-terminal Protein G B1 domain (GB1) to facilitate expression and increase solubility. The genes encoding full-length human CstF-50 and the WD40 domain of CstF-50 were inserted into a modified pFastBac (Life Technologies) vector to incorporate an N-terminal GST tag to facilitate purification, while the N-terminal dimerization domain was cloned into a pGEX backbone for bacterial expression. Sequences coding for human CstF-77 Δ N and CstF-64RH were cloned into a modified pRSF-Duet-1 (Novagen) vector with a 6His-GB1 tag at the N-terminus of CstF-77 Δ N. Identity of all clones were verified by DNA sequencing.

All bacterial expression vectors were transformed into BL21 (DE3) competent cells. Transformants were then grown in LB media at 37°C to an OD₆₀₀ reading of 0.6. The recombinant proteins were overexpressed in *Escherichia coli* at 18°C overnight upon induction by addition of 0.5mM IPTG. Standard methods were used for production and expansion of baculovirus Sf9 cells expressing CstF-50. Viruses at P3 stage of expansion were used to infect High Five (Invitrogen) monolayer insect cells for 72 h. The cells were harvested by centrifugation and resuspended in buffer A (20 mM Tris pH 8.0, 150 mM NaCl) and lysed by sonication at 4°C. The crude extracts were centrifuged at 27 000g for 60 min to remove the cell debris.

The proteins with a 6His-GB1 tag were purified by nickel affinity chromatography. Unspecific bound proteins were extensively washed with buffer B (buffer A + 20 mM imidazole) and the target proteins were eluted from the column with buffer C (buffer A + 250 mM imidazole). The GST tagged proteins were loaded onto a glutathione sepharose (GE Healthcare) column then washed with buffer A, and finally eluted with buffer D (buffer A + 10 mM reduced glutathione). TEV protease was added to the purified material to remove the 6His-GB1 tag and GST tag, and the mixture was then dialyzed against buffer E (20 mM Tris pH 8.0, 1 mM DTT).

Following dialysis, proteins were further purified on a Q HP column (GE Healthcare) using a linear gradient of 0–1M NaCl in buffer E. Fractions containing pure material were concentrated and loaded onto a Superdex 200 10/300 GL (GE Healthcare) column equilibrated in storage buffer (20 mM Tris pH8.0, 150 mM NaCl, 1 mM DTT). The purified proteins were flash frozen for future use.

GST pull-down assays

All purified CstF-77 constructs and the CstF-77 Δ N/CstF-64RH subcomplex were incubated with GST-CstF-50 beads for one hour at 4°C in storage buffer and washed extensively with storage buffer following incubation. The samples were eluted from beads with buffer D and applied to SDS-PAGE gel for analysis. The same protocol was scaled up to produce milligram amounts of the CstF complex for assays used in this study. Samples from the scale up were subject to ion exchange, and gel filtration as described under 'Protein preparation'.

Crystallization, data collection, structure determination and refinement

For crystallization, the WD40 domain of CstF-50 was concentrated to ~15mg/ml in storage buffer then mixed in 1:1.2 ratio with the CstF-77 peptide obtained from Genscript. Crystals were obtained using the hanging-drop vapor diffusion method at 4°C by mixing one volume of protein sample with an equal volume of precipitant (0.1 M MES pH 6.0–6.5, 1.6–1.9 M ammonium sulfate, 2–5% PEG400, 0.2 M NaCl). Crystals appeared in ~1 day and matured to final size after a week. Crystals were cryoprotected by addition of 20% (v/v) glycerol to the crystallization solutions, then flash frozen in liquid nitrogen and harvested for further data collection.

All X-ray diffraction data were collected at the Advanced Light Source at the Lawrence Berkeley National Laboratory at beam lines BL8.2.1. All diffraction data were indexed, integrated, and scaled with the HKL2000 package (21). Complete data collection statistics are summarized in Table 1.

For structure determination, phases were solved by molecular replacement using Phaser (22) in the CCP4 suite (23) with a model generated by the Phenix MR_Rosetta protocol (24,25) using an HHPred (26) derived alignment. The initial models were then rebuilt using COOT (27) and refined by Refmac5 (28). A detailed summary of the refinement statistics are provided in Table 1. Graphics figures

Table 1. Data collection and refinement statistics

	CstF-50-WD40/CstF-77 (581–600)
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	43.9 74.5 95.1
α , β , γ (°)	90 90 90
Wavelength (Å)	1.000
Resolution (Å)	50–2.3 (2.34–2.30)
R_{sym} or R_{merge}	0.116 (0.584)
$I/\sigma I$	29.3 (3.2)
Completeness (%)	98.9 (93.2)
Redundancy	5.9 (3.8)
Refinement	
Resolution (Å)	47–2.3
No. reflections	13548
$R_{\text{work}} / R_{\text{free}}$	21.2/26.0
No. atoms	2810
Protein	2708
Water	102
B-factors	
Protein	48
Water	55
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.076
Ramachandran plot (%)	
Favored	97
Allowed	3
Outliers	0

were rendered in PyMOL and conservation analysis was performed by Chimera.

Size exclusion chromatography multi-angle light scattering

The SEC-MALS system consisted of a P900 HPLC pump (GE Healthcare), a UV-2077 detector (Jasco), a Tri Star Mini Dawn light scattering instrument (Wyatt), and an Opti Lab T-Rex refractive index instrument (Wyatt). Approximately 200 μg of purified CstF complex was injected into a Superdex 200 10/300 GL column and eluted at 0.5 ml/min in buffer containing 20 mM Tris pH 8.0, 250 mM NaCl, 0.05% NaN_3 . The specific refractive index of the CstF complex was assumed to be 0.186 ml/g. Data collection and analysis were performed with Astra 6 software (Wyatt). Total molecular mass of the complex was determined with Astra 6 software using protein analysis. Both peak overlap and peak broadening were corrected with Astra 6 software. The SEC-MALS system was pre-calibrated with BSA.

Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) was performed at 293 K with a MicroCal ITC200. All protein samples were dialyzed against PBS buffer. All of the RNA ligands used in this study were purchased from Integrated DNA Technologies and the CstF-77 peptide was synthesized by GenScript. RNAs and peptide were dissolved in RNase-free water and desalted by a PD MiniTrap G-10 or Illustra NAP-10 (GE), finally dissolved in PBS buffer after overnight lyophilization. The concentrations of proteins and RNAs were calculated by absorbance spectroscopy with extinction coefficients at 280 and 260 nm, respectively, and the peptide

concentration was estimated by weighting lyophilized powder. The typical titration was carried out with proteins in the cell as titrate (5 μM) and the ligands in the syringe as titrant (100 μM), except 300 μM UUUUU-containing RNA (GUGUGACCCUUUUU) was titrated into 15 μM CstF-77 ΔN /CstF-64RH subcomplex due to weak binding. Data were analyzed by Origin 7 and all measurements were performed at least in duplicate.

RESULTS

CstF-50 binds to a conserved patch of CstF-77 between the HAT domain and CstF-64 binding site

The interaction between CstF-77 and CstF-64 has been well established by structural studies of yeast homologs (Rna14/Rna15) (15,16), but it is unclear how CstF-50 is integrated into the CstF complex. Preliminary studies suggested that CstF-50 binds to a region C-terminal to CstF-77's HAT domain and N-terminal to CstF-64's binding site (14). To investigate this interaction, we performed GST pull-down assays to confirm binding of CstF-77 (residues 241–610) using purified full length GST-CstF-50 fusion protein as bait. As shown in Figure 1A, CstF-77 was successfully pulled down by GST-CstF-50. To establish whether dimerization of CstF-50 is important for CstF-77 binding, we repeated the GST pull-down assay by using GST-tagged CstF-50 N-terminal dimerization domain and C-terminal WD40 domain as baits. As shown in Figure 1A, the WD40 of CstF-50 but not the dimerization domain binds to CstF-77.

To further define the binding site of CstF-50 within CstF-77, we made a series of C-terminal truncations of CstF-77

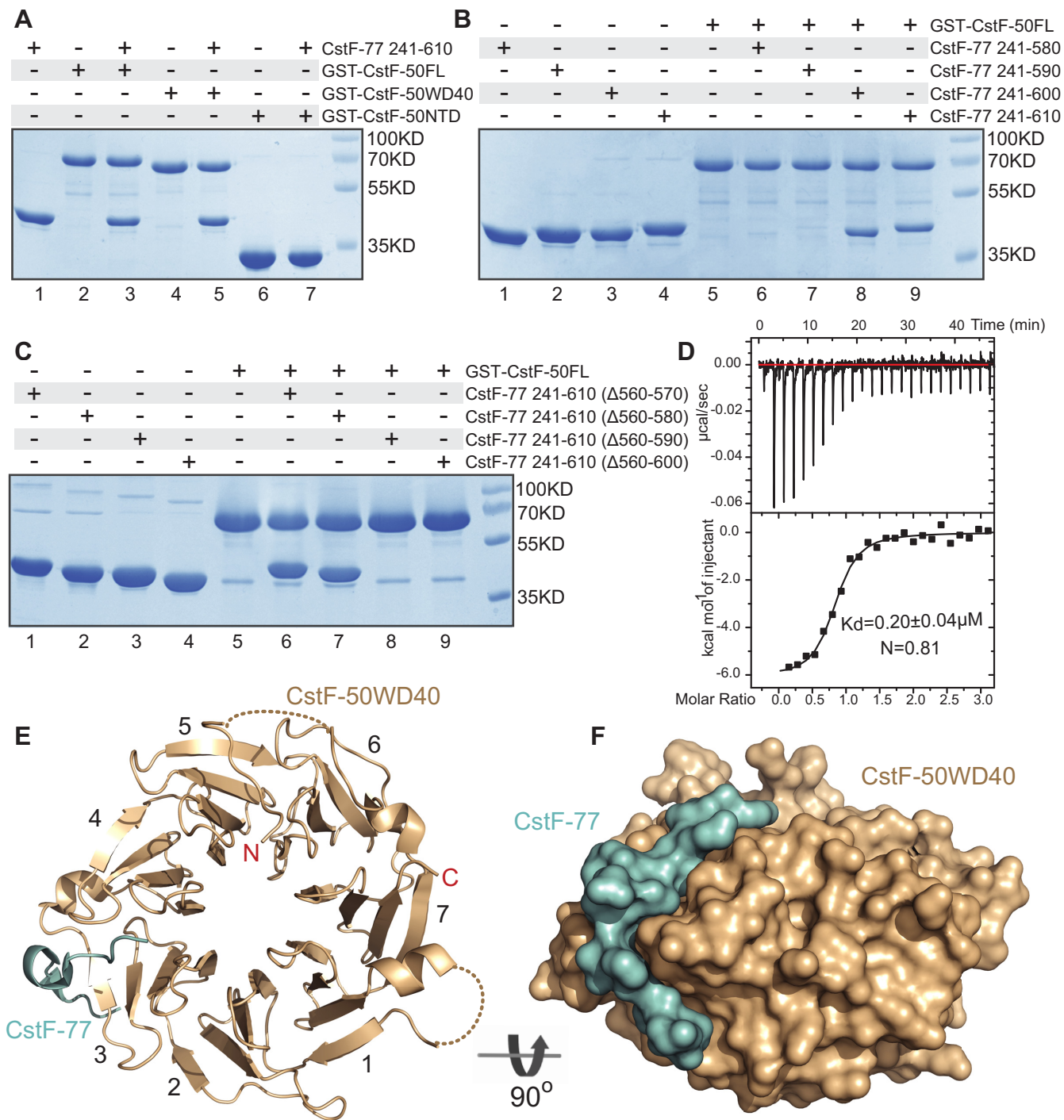


Figure 1. The WD40 domain of CstF-50 binds to a conserved segment of CstF-77. (A) Mapping the domains of CstF-50 required for interaction with CstF-77 by pull-down assays. GST-tagged CstF-50FL, CstF-50-WD40 or CstF-50-NTD were pre-bound to glutathione sepharose beads (lanes 2, 4, 6) and the pull-down assays were carried out using excess pre-purified CstF-77 (241–610) (lane 1). The results clearly show that CstF-77 can be pulled down by both full-length CstF-50 (lane 3) and the WD40 domain (lane 5) but not by the NTD (lane 7). (B) Mapping the C-terminal boundary of CstF-77 required for association with CstF-50. GST-tagged CstF-50FL protein was pre-bound to glutathione sepharose beads (lane 5) and the pull-down assays were performed using different C-terminal extensions of CstF-77 (lanes 1–4). Only the proteins extended at least to residue 600 bind to CstF-50 strongly. (C) Identifying the N-terminal boundary of CstF-77 that recognizes CstF-50. A series of internal deletions of CstF-77 based on residues 241–610 were purified (lanes 1–4) and the pull-down assays were performed using GST-tagged CstF-50FL (lane 5). CstF-77 retains the ability to bind CstF-50 after removal of residues 560–570 or 560–580 (lanes 6–7), while binding is lost upon deletion of 560–590 or 560–600 (lanes 8–9). (D) Representative ITC curve for the binding of the WD40 domain of CstF-50 and the CstF-77 peptide (581–600). Raw injection heats are shown in the upper panel and the corresponding integrated heat changes are shown in the bottom panels versus the molar ratio of CstF-77 peptide to the WD40 domain of CstF-50. (E) Ribbon diagram of the WD40 domain of CstF-50 in complex with the CstF-77 peptide. The two proteins are colored in wheat and cyan. The canonical seven blades in the WD40 domain of CstF-50 were numbered and both N- and C-termini were labeled. The dashed lines represent two flexible loops that were not modeled due to poor electron density. (F) Surface representation of the complex structure (orthogonal view to E) with domains colored with the same scheme as in Figure 1E.

based on the above construct (residues 241–610) and successful purification of all relevant proteins (Figure 1B). Using GST-CstF-50 as bait, we tested the ability of these CstF-77 constructs to bind to CstF-50. As shown in Figure 1B, the CstF-77 HAT-C domain (residues 241–550) and the region consisting of amino acids 551–590 lack the ability to bind CstF-50. However, proteins truncated at residue 600 and beyond can bind to CstF-50 without significant difference in binding between, e.g. constructs 241–600 and 241–610 (Figure 1B). These results show that the C-terminal boundary of the CstF-50 binding site is located around residue 600. To identify the minimal binding site recognized by CstF-50, we made a series of internal deletions of CstF-77 based on residues 241–610 and repeated the above GST pull-down assay. As shown in Figure 1C, CstF-77 retains the ability to bind CstF-50 after removal of residues 560–580, while binding is lost upon deletion of 560–590. Altogether, these results identify residues 581–600 as the minimal binding site that recognizes CstF-50. Consistent with this conclusion, comparison of sequences spanning metazoan species also showed a highly conserved patch between residues 581–600 (Supplementary Figure S1).

We further performed ITC to measure the binding between the WD40 domain of CstF-50 and a synthetic peptide spanning residues 581–600 of CstF-77. As predicted, we observed a 1:1 complex with a K_d of ~200 nM (Figure 1D). Altogether, we conclude that the WD40 domain of CstF-50 binds to CstF-77 via a conserved patch (residues 581–600) between the HAT domain and the CstF-64's binding site.

Structural basis for recognition of CstF-77 by CstF-50

In order to gain structural insight into the CstF-50/CstF-77 interaction, we determined the co-crystal structure of the WD40 domain of human CstF-50 in complex with the CstF-77 peptide (residues 581–600) to 2.3 Å resolution (Table 1). CstF-50 has a relatively canonical seven-bladed β -propeller domain (Figure 1E) and the observed CstF-77 peptide (residues 581–594) fits well into a side cleft that is formed by blades 3 and 4, and extends from the bottom face to the top face of the β -propeller of CstF-50 (Figure 1E–F and Supplementary Figure S2).

To prevent ambiguities, we refer to residues in the CstF-77 peptide using the three-letter code, while residues on CstF-50 will be referred to using single letter abbreviations. The first two amino acids, Tyr581 and Pro582, in the CstF-77 peptide form non-specific interactions with P227 and A274 in CstF-50 through hydrophobic contacts (Figure 2A). To further anchor the N-terminus of the peptide on the bottom face of CstF-50, the backbone carbonyl oxygen of Pro582 forms a hydrogen bond with the side chain (N ϵ 2 atom) of H226 (Figure 2A).

Pro584 is the first residue that contributes specific contacts between the peptide and CstF-50, predominantly through intermolecular hydrophobic interactions. A hydrophobic pocket formed by F231, L233, L243, Y277 and V292 in CstF-50 can accommodate Pro584 in that position (Figure 2B). Although no specific interaction was found between the side chain of Asp585 and CstF-50, the main chain amide nitrogen of Asp585 in the peptide forms a hydrogen bond with the backbone carbonyl oxygen of V292 in

CstF-50 to further stabilize the peptide conformation (Figure 2B). Residues 586–588 form a short α helix to facilitate the redirection of the peptide from the bottom to the top face of CstF-50. Several intramolecular hydrogen bonds stabilize this turn (Figure 2C). No specific interactions were observed between this turn and CstF-50, but the side chain of Thr586 contacts F252 of CstF-50 through hydrophobic interactions (Figure 2B and C).

The second residue that is specifically recognized by CstF-50 is Met589, whose side chain protrudes into a hydrophobic pocket next to the Pro584 binding site, formed by L243, F252 and N294 on CstF-50 (Figure 2C). To reinforce the interaction and lock this second determinant residue in position, three residues immediately following Met589 form several backbone interactions with CstF-50 (Figure 2D). No specific recognition was identified between these three residues and CstF-50, except that the side chain of Phe592 contacts V253 through hydrophobic interactions (Figure 2D).

In order to validate the significance of the molecular interactions observed in the structure, we carried out a mutational analysis on CstF-77 (residues 241–610). Using GST tagged CstF-50-WD40 domain as bait, we tested the ability of these CstF-77 mutants to bind to CstF-50 (Figure 2E). Each of the single mutants Pro584Ala, Thr586Ala, Met589Ala and Phe592Ala significantly decreased the binding affinity to CstF-50, while Tyr581Ala, Lys583Ala, Ile590Ala and Pro591Ala mostly retained the ability to interact with CstF-50. Of note, a new CstF-77 construct containing residues 241–592 also retained the full binding ability to CstF-50, suggesting residues Gln593 and Pro594 of CstF-77 do not participate in recognition of CstF-50.

These data, combined with our structural investigation, demonstrate that Pro584, Thr586, Met589 and Phe592 predominantly contribute to recognition of CstF-50 and function as the structural determinants for the specific recognition of CstF-77 by CstF-50, whereas the peripheral residues in the peptide predominantly interact with CstF-50 through backbone hydrogen bonds to stabilize and lock the peptide conformation in place.

In vitro reconstitution of the CstF complex demonstrates that CstF is a dimer of heterotrimers

To further study the architecture and biochemical activity of the entire CstF complex, we sought to assemble it using recombinant proteins (Figure 3A). While an early reconstitution of CstF from baculovirus was reported (29), we attempted to setup an efficient method to reconstitute a stoichiometric complex. Full-length CstF-77 can be readily purified when co-expressed with the CstF-64's RRM and Hinge domains (denoted CstF-64RH) in bacteria. Using a GST-pull-down strategy, we immobilized full-length GST-CstF-50 and used it to capture a preformed CstF-77/CstF-64RH heterodimeric subcomplex. With this approach, we were able to robustly form a stoichiometric 3-subunit complex, which was amenable to scale-up to prepare large quantities of protein. However, we observed that CstF-77 underwent extensive N-terminal degradation (data not shown), as was also seen in previous studies (30). Removal of the HAT-

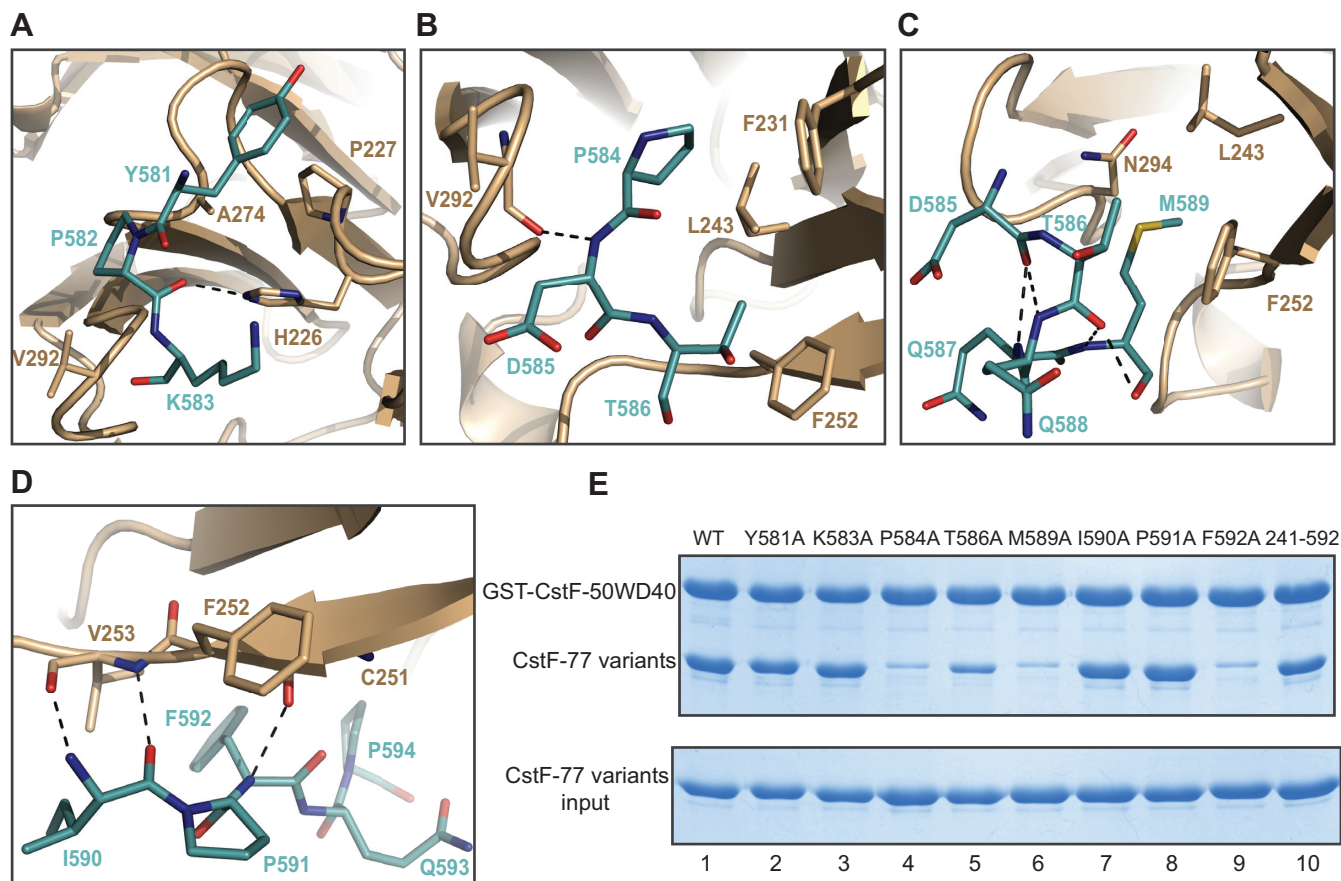


Figure 2. Structural and mutagenesis analyses of the CstF-50/CstF-77 interface. (A–D) Close-up views of interactions at the CstF-50/CstF-77 interface, colored in the same scheme as in Figure 1E. Dash lines indicate intermolecular hydrogen bonds. (E) The effect of structure-based CstF-77 mutations assayed by GST pull-down methods. The CstF-77 variants P584A, T586A, M589A and F592A have significantly decreased binding affinity for CstF-50, while other mutants largely retain the ability to interact with CstF-50.

N domain from CstF-77 (residues 1–240, denoted CstF-77 Δ N) resulted in highly homogenous preparations, yet did not impair complex assembly (Figure 3B).

To validate our pull-down results, the eluted CstF complex was loaded onto a gel filtration column, and all three proteins co-eluted in a single monodisperse peak. We also observed a supershift of free CstF-50 to an earlier elution volume, indicating stable integration into the CstF complex (Figure 3C). Curiously, given that CstF-77 Δ N/CstF-64RH subcomplex has a similar mass to CstF-50, comparison of elution volumes of the CstF-77 Δ N/CstF-64RH subcomplex with that of the full CstF complex only shows a small shift on gel filtration (Figure 3C), suggesting that CstF-50 must be incorporated in a fashion that minimally increases the overall hydrodynamic size of the complex, perhaps due to a non-globular shape. To validate the assembly and the molecular weight of the complex, we used size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS), which allows for accurate molar mass determinations of macromolecules based on light scattering of column fractions. The results demonstrated that CstF migrated as a single species with an approximate molecular weight of 250 kDa on a Superdex 200 column (Figure 3C). These results are consistent with a hexameric complex containing two copies each of CstF-77 Δ N (~50 kDa), CstF-

64RH (~25 kDa) and CstF-50 (~50 kDa), and conclusively demonstrate that CstF is indeed a dimer of heterotrimers, with dimerization reinforced by both CstF-77 and CstF-50.

CstF binds various G/U-rich target RNAs with high affinity and specificity

Early SELEX experiments and bioinformatics studies showed that CstF complex isolated from native sources binds to G/U-rich RNAs regardless of the particular sequence (9,31,32), while discriminating against A/C-rich sequences. However, to date, the RNA binding activities by a recombinant CstF complex have yet to be studied, because such a study would require properly expressed and reconstituted complex, which we now have available.

First, we sought to define the length of G/U-rich RNAs that can be optimally recognized and bound by the CstF complex. To address this point, we synthesized a series of G/U-rich RNAs (GU6 to GU14) as binding targets, and tested protein-RNA interactions quantitatively by isothermal titration calorimetry (ITC). As shown in Figure 4A, B and Supplementary Figure S3A–D, CstF binds to G/U-rich RNAs with two different stoichiometries. Each CstF binds two G/U RNA molecules independently in the case of GU6 and GU8, but exhibits equimolar binding with the longer

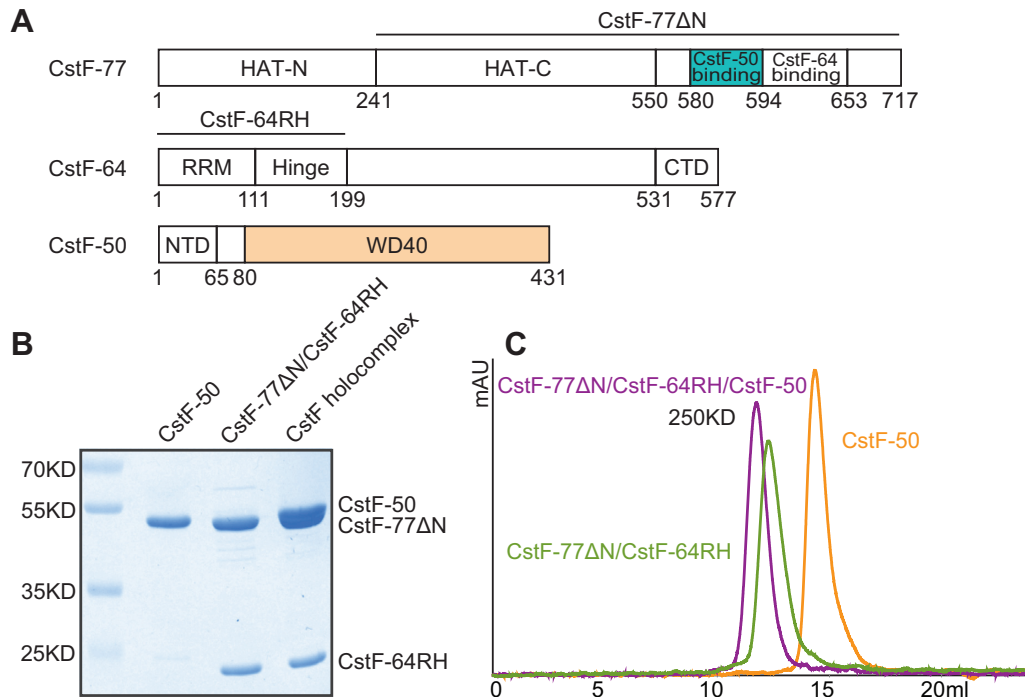


Figure 3. *In vitro* reconstitution of the hexameric CstF complex. (A) Domain organization of CstF-77, CstF-64 and CstF-50. All numbering and domain breakdowns are based on the human proteins. The fragments used for assembling the CstF complex are underlined and the regions of crystallized CstF-50 and CstF-77 are colored in wheat and cyan, as in the previous figures. (B) SDS-PAGE gel of the purified CstF-50, CstF-77 Δ N/CstF-64RH subcomplex, and of the full complex used in gel filtration runs and RNA binding assays. All of the affinity tags have been removed. (C) Size exclusion chromatography elution profiles of the isolated CstF-50 (brown), CstF-77 Δ N/CstF-64RH subcomplex (green), and of their complex (purple).

RNA substrates (GU10-GU14). Control experiments using a CstF complex assembled in the absence of the CstF-64's RRM domain show no binding to G/U-rich RNA (data not shown), confirming that primary RNA interactions occur via the RRM even in the context of the complete complex. Additionally, control experiments with an A/C sequence of comparable length (AC14) show no detectable binding to the complex either, indicating that CstF retains its specificity to G/U-rich sequences (Supplementary Figure S3I). Considering that each CstF complex contains two copies of CstF-64's RRM domain, the different stoichiometry suggests that 10 G/U-rich nucleotides is the minimum length of RNA that enables the two RRMs present in each CstF complex to bind RNA simultaneously, under the structural constraints of complex assembly.

Independently of the binding stoichiometry, longer RNA substrates consistently exhibited stronger binding to CstF, e.g. GU8 binds to CstF \sim 2.5-fold tighter than GU6, and GU14 has a low nanomolar K_d (120 nM) compared with GU10's more moderate binding ($K_d \sim$ 500 nM) (Figure 4A, B and Supplementary Figure S3A–D). The stronger affinity to the longer RNAs are likely a result of multiple frames of binding along a longer RNA, allowing for CstF to 'slide' and provide greater opportunities for the RNA to interact.

Previous work identified DSE's as likely composed of two variably spaced G/U- and U-rich elements (10,11,33), and the GU10 RNA used in above experiments could be conceptually separated into two segments, in which GUGUG acts as the G/U-rich element while UGUGU is the U-rich element. Since no spacer sequence is left in GU10 RNA, we

investigated if and how spacing between the two elements plays a role in CstF recognition. Thus, we prepared a series of GU10 RNAs where G/U- and U-rich sequences are separated by a varying number of adenosines inserted in the middle as spacers (Figure 4A), and repeated the ITC experiments against the CstF complex. Since CstF-64 does not recognize adenosine, these bidentate RNAs will reflect direct recognition of DSE sequences. As shown in Figure 4A and Supplementary Figure S3E–H, the binding stoichiometry observed for all RNAs is 1:1 and the CstF complex displays no obvious preference with regards to spacer lengths, with K_d 's consistently at \sim 500 nM, up to a longer spacer sequence of 19 adenosines. Given the relatively low uridine content of the U-rich element (UGUGU) used above, we optimized the RNA by replacing UGUGU with a UU-UUU sequence that reportedly significantly increases the efficiency of 3'-end processing (34), inserted a 4 nts A/C spacer between the two elements, and repeated the ITC experiment. As shown in Figure 4A and D, the CstF complex exhibits a binding affinity of \sim 900 nM, \sim 2-fold weaker than against GU10-derived RNAs ($K_d \sim$ 500 nM). Together, these data demonstrate that CstF is able to simultaneously bind two various downstream elements separated by spacers of different length.

Dissecting the contribution of each subunit in CstF to RNA recognition

A previous study has shown that the intact CstF complex binds G/U-rich RNAs much more strongly than the

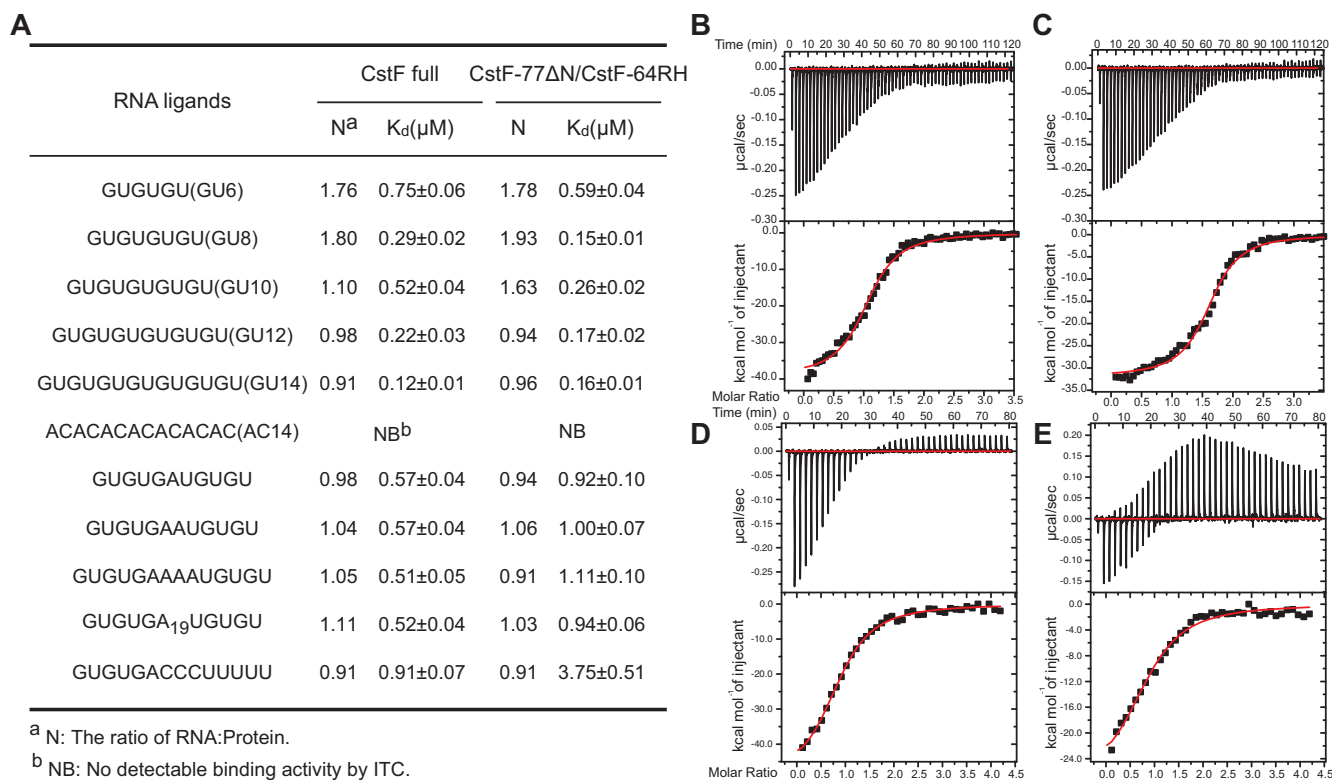


Figure 4. RNA binding of fully assembled CstF complex and CstF-77/CstF-64 subcomplex. (A) The binding stoichiometry and affinities of full CstF complex and CstF-77 Δ N/CstF-64RH subcomplex to G/U-rich RNAs. AC14 was used as negative control for specificity. Representative ITC curves for the binding of full CstF complex to (B) GU10 and (D) UUUUU-containing RNA (GUGUGACCCUUUUU); binding of CstF-77 Δ N/CstF-64RH subcomplex to (C) GU10 and (E) UUUUU-containing RNA. Raw injection heats are shown in the upper panels and the corresponding integrated heat changes are shown in the bottom panels versus the molar ratio of RNA to protein.

isolated CstF-64 RRM domain (9), suggesting the complex assembly creates a high affinity platform that optimally targets G/U-rich sequences. To dissect the contribution of each subunit to RNA recognition, we performed ITC experiments using the GU14 sequence against CstF the complex, CstF-77 Δ N/CstF-64RH subcomplex and the CstF-64 RRM alone. Using the isolated CstF-64 RRM domain, we measured a $K_d \sim 2.5 \mu\text{M}$ (Supplementary Figure S3I; small differences from previous observations may arise from the exact RNA lengths and sequences used in the experiments (17,18)). Both the CstF complex and CstF-77 Δ N/CstF-64RH subcomplex showed superior binding to GU14, with a measured K_d of approximately 120 and 160 nM, respectively, ~ 20 -fold stronger binding relative to the single RRM, suggesting CstF-77 is the key factor that boosts the affinity of the CstF-64 RRM.

The data presented above show CstF-77/CstF-64 subcomplex assembly plays an important role in boosting the affinity against G/U-rich RNAs. In agreement with this observation, the lack of an obvious yeast homolog of CstF-50 suggests it might not be strictly required for basal cleavage and polyadenylation. We therefore wondered about the functional role for metazoan's CstF-50 in the fully assembled complex, and specifically whether it affects recognition of G/U-rich RNAs. ITC experiments performed using CstF-77 Δ N/CstF-64RH subcomplex against G/U-rich RNAs of varying length (GU6-GU14; Figure 4A,

C and Supplementary Figure S3A–D) showed that CstF-77 Δ N/CstF-64RH binds G/U-rich RNAs with generally comparable affinities to those observed for the entire complex, suggesting CstF-50 plays a small role in the recognition of these RNAs. However, the CstF-77 Δ N/CstF-64RH subcomplex binds to GU10 with an unexpected protein:RNA molar ratio of 1:1.63, close to a 1:2 binding mode, while 1:1 binding was observed for the intact complex (Figure 4A–C).

We also repeated the ITC experiments with the CstF-77 Δ N/CstF-64RH subcomplex against the bidentate RNAs with different length spacers. As shown in Figure 4A and Supplementary Figure S3E–H, the subcomplex also displays no obvious preference for spacer length with a comparable K_d of $\sim 1 \mu\text{M}$ to the fully assembled complex, suggesting CstF-50 only affects subtle improvements in RNA recognition. However, CstF-77 Δ N/CstF-64RH subcomplex binds the UUUUU-containing DSE much more weakly, with a K_d of $\sim 3.7 \mu\text{M}$ (Figure 4A and D–E), a binding constant that resembles that of a single CstF-64 RRM domain, suggesting that CstF-50 is likely needed to boost affinities towards a bidentate G/U-, U-rich that have been described in the literature as being part of a strong 3'-end signal (34).

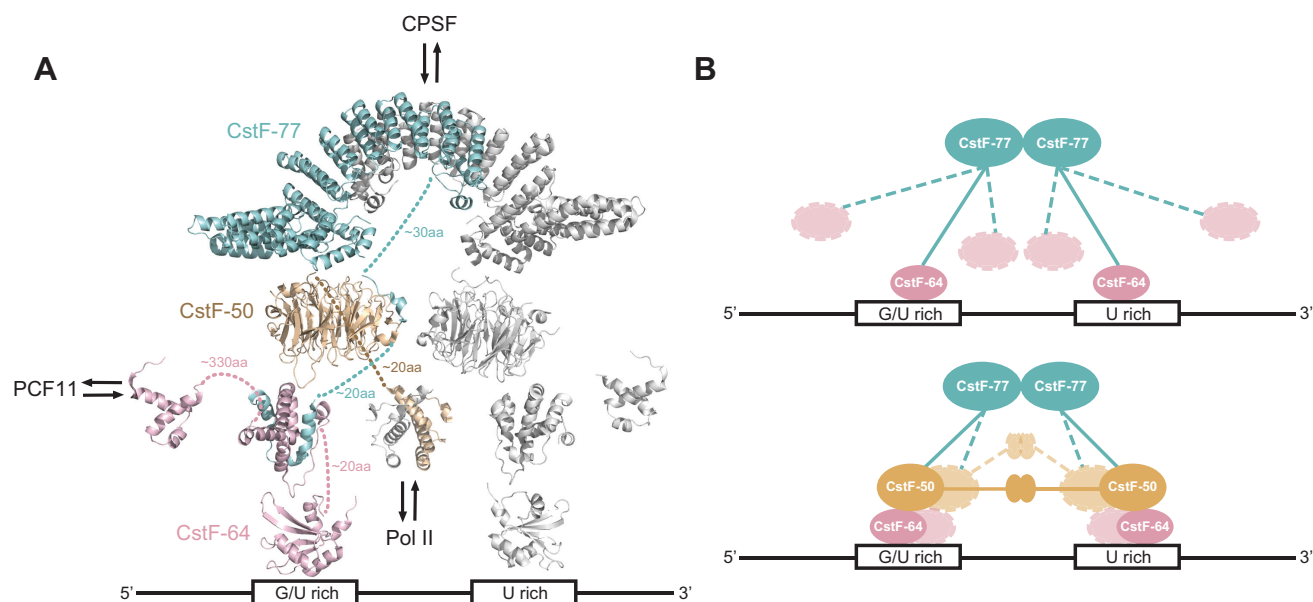


Figure 5. A model for DSE recognition by CstF complex and the contribution of each subunit to RNA recognition. (A) A molecular graphic model showing all available structures of CstF complex components (PDBs: 2OOE (CstF-77HAT), 2L9B (Rna14/Rna15), 1P1T (CstF-64RRM), 2J8P (CstF-64CTD), 2XZ2 (CstF-50NTD) and 6P3X (CstF-50WD40)). The three proteins are colored in cyan, pink and wheat. The symmetry copy is shown in grey for clarity. The dashed lines represent flexible loop regions unresolved in the structures and the numbers of missing residues are also labeled. Potential cross-talk with other components of the 3'-end processing apparatus are shown with black arrows. (B) Top: CstF-77 boosts the affinity of the CstF-64 RRM for RNA due to its dimeric assembly that incorporates two RNA binding motifs in a single complex in close spatial proximity. However, without CstF-50, the CstF-77/CstF-64 subcomplex binds to DSE signals less favorably due to the highly flexible linker region between the HAT domain and CstF-64 hinge binding site on CstF-77, which results in the random movement of the two RRM of CstF-64 (represented with dashed lines). Bottom: The dimerized CstF-50 acts as a clamp to restrict the flexibility of the linker region by binding to a conserved patch of CstF-77 between the HAT domain and CstF-64 binding site, limiting the movement of the two RRM and allowing more efficient engagement of G/U- and U-rich motifs. The dashed lines indicates that the full complex has modest flexibility allowed by the linker between the dimerization and WD40 domains of CstF-50, which allows recognition of DSE signals of different length.

DISCUSSION

Biochemical and structural analysis of the polyadenylation machinery has been hampered by the ability to obtain sufficient amounts of highly pure recombinant proteins and their reconstitution into functional complexes. Earlier studies used complexes purified from native sources such as transformed cell lines (3,35–39), but these may contain contaminating proteins, or could be isolated as non-stoichiometric complexes. Using these sources as starting material also becomes difficult to design mutants to probe activity. In this work, we describe an efficient method to reconstitute a functional RNA-binding core of the cleavage stimulatory factor from purified recombinant proteins, and demonstrate that CstF assembles as a hexameric complex that binds G/U-rich sequences with high affinity and selectivity. In contrast, the RNA-binding component of the complex, the RRM of CstF-64, only binds RNA with moderate low μM activity, even to long G/U-rich sequence (GU14) suggesting that it is the assembly of functional complexes, rather than cooperative interactions between weakly binding proteins, that drive recognition of processing sites. The molecular basis for the super-activity of CstF-64 when assembled into CstF likely is that dimeric assembly incorporates two RNA binding motifs in a single complex in close spatial proximity.

A general consensus sequence for DSEs has yet to be established because these elements are extremely variable, and

this also makes it difficult to establish quantitative properties of the CstF-DSE interaction. Our studies indicate that ten G/U bases is the minimal length that can tolerate binding of two CstF-64 molecules to a single strand of RNA, while shorter G/U-rich elements can be occupied by only one of the two CstF-64's RRM domains.

In addition to G/U-rich RNAs, we also observed that the CstF complex is able to bind to two independent G/U- and U-rich elements separated by variable A/C spacing, ranging from 0 to 19, with no obvious preference for spacer length. To date, there is still some debate as to the existence of a two part DSE, which would include distinct G/U- and U-rich signal (33,40,41). Our findings strongly support the idea that the CstF complex may serve as the platform to read G/U- and U-rich signals simultaneously using two copies of CstF-64. Computational statistical analysis of DSE sequences observed the typical separation between the G/U-rich and U-rich elements is approximately 15nts (11), which is consistent with our results as well. The observed length of 4–7 nts for a typical G/U-rich sequence (10,11) would not allow simultaneously binding of two CstF-64's (10 nts is the minimal length, as we clarified above), suggesting that one CstF-64 in a CstF holocomplex can occupy the G/U-rich motif, while the second must bind to the U-rich element.

Our crystal structure and biochemical analysis of the CstF-50/CstF-77 complex reveals a unique binding surface on the circumference of the β -propeller (Supplementary Figure S4) and identifies critical residues that con-

tribute to the recognition of CstF-50 by CstF-77. Sequence analysis of CstF-50 among metazoans shows that the peptide binding cleft is highly conserved, in agreement with its function in CstF-77 recognition (Supplementary Figure S5). Of note, CstF complexes in metazoans have three protein components, while the yeast equivalent has only two (Rna14/Rna15), with no identified homolog of CstF-50. Consistent with this, sequence analysis of CstF-77 between metazoans and yeasts shows high conservation in the HAT domain and the CstF-64 binding site, but no conservation is observed for residues 581–594, suggesting the highly conserved patch that recognizes CstF-50 does not exist in fungi (Supplementary Figure S6).

The conservation of the interface and absence of it in yeast strongly suggests that the CstF-50/CstF-77 interaction motif is important in metazoans, yet similar RNA binding characteristics exhibited by the heterodimeric CstF-77 Δ N/CstF-64RH subcomplex and the absence of CstF-50 in yeast made us curious about the function of CstF-50 implied by the co-evolution of its interface with CstF-77. The WD40 domain of CstF-50 interacts with BARD1 (42) and the N-terminal dimerization domain is weakly involved in pol II CTD binding (43), but the exact function of CstF-50 remains a mystery. In this study, we showed that the CstF holocomplex efficiently and indiscriminately binds to different RNAs that contain both G/U- and U-rich elements, but loss of CstF-50 disfavors CstF binding to DSE, particularly when uridine content is high. Considering that the U-rich element consists of four uridines out of five bases (10,31), our findings likely implies a potential role of CstF-50 in efficient selection of DSEs, but further experimentation will be necessary to elucidate the mechanistic details of its contribution.

We speculate, based on current structural understanding, that dimerized CstF-50 acts as a clamp to restrict the flexibility of the linker region between the HAT domain and CstF-64 hinge binding site on CstF-77 (Figure 5A and B) and therefore restricts the two RRM domains of CstF-64 to a single DSE motif. This arrangement would force the CstF complex to recognize DSE signals in *cis* on a single strand of mRNA by maximally coupling binding of both CstF-64 RRMs, as opposed to a less discriminate binding mode without CstF-50 which would result in unfavorable binding. In agreement with this suggestion, partial RNAi knockdown of CstF-50 in *C. elegans* resulted in the upshift of CstF-64 binding to more proximal, and likely weaker, 3'-end processing sites on several genes (44).

CONCLUSIONS

Reconstitution of CstF with recombinant purified proteins has allowed us to interrogate the hexameric complex's RNA binding ability, dissect the contribution of each subunit to RNA recognition and provide structural insight into the recruitment of CstF-50 by CstF-77. Unfortunately, even this complex retains conformational flexibility which prevents successful crystallization or single particle EM analysis. We demonstrate that CstF-77 boosts the affinity of the CstF-64 RRM to the RNA targets and uncover a potential role for metazoan CstF-50 in the recognition of 3'-end processing signals by restricting flexibility in the CstF complex rather

than through direct contacts with the RNA. Our experiments provide a rationale for *in vivo* observations of the role of CstF-50 in recognition of processing sites, and provide a platform for future experiments to further investigate the role of CstF in RNA 3'-end processing.

AVAILABILITY

Atomic coordinates and structure factors of CstF-50-WD40 domain in complex with the peptide of CstF-77 have been deposited in the Protein Data Bank with the accession code 6B3X.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the beamline staff of the Advanced Light Source at the University of California at Berkeley for help with the X-ray data collection. The ITC experiments were conducted by John Sumida at the Molecular Analysis Facility, a National Nanotechnology Coordinated Infrastructure site at the University of Washington. We also thank other members of the Varani lab for discussion and help in the laboratory.

FUNDING

National Institutes of Health [GM064440 to G.V.]. Funding for open access charge: National Institutes of Health [GM064440 to G.V.].

Conflict of interest statement. None declared.

REFERENCES

- Mandel, C.R., Bai, Y. and Tong, L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cell. Mol. Life Sci.*, **65**, 1099–1122.
- Xiang, K., Tong, L. and Manley, J.L. (2014) Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. *Mol. Cell. Biol.*, **34**, 1894–1910.
- Takagaki, Y., Ryner, L.C. and Manley, J.L. (1989) Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev.*, **3**, 1711–1724.
- Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A.R., Keller, W., Zavolan, M. and Wahle, E. (2014) Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev.*, **28**, 2381–2393.
- Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Iii, J.R.Y., Ule, J., Manley, J.L. and Shi, Y. (2014) CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev.*, **28**, 2370–2380.
- Yang, Q., Gilmartin, G.M. and Doublé, S. (2010) Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 10062–10067.
- Yang, Q., Coseno, M., Gilmartin, G.M. and Doublé, S. (2011) Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. *Structure*, **19**, 368–377.
- Li, H., Tong, S., Li, X., Shi, H., Ying, Z., Gao, Y., Ge, H., Niu, L. and Teng, M. (2011) Structural basis of pre-mRNA recognition by the human cleavage factor Im complex. *Cell Res.*, **21**, 1039–1051.
- Takagaki, Y. and Manley, J.L. (1997) RNA recognition by the human polyadenylation factor CstF. *Mol. Cell. Biol.*, **17**, 3907–3914.

10. Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L. and Hovorun, D.M. (2003) Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res.*, **31**, 1375–1386.
11. Salisbury, J., Hutchison, K.W. and Graber, J.H. (2006) A multispecies comparison of the metazoan 3'-processing downstream elements and the CstF-64 RNA recognition motif. *BMC Genomics*, **7**, 55.
12. Bai, Y., Auperin, T.C., Chou, C.Y., Chang, G.G., Manley, J.L. and Tong, L. (2007) Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. *Mol. Cell*, **25**, 863–875.
13. Legrand, P., Pinaud, N., Minvielle-Sébastien, L. and Fribourg, S. (2007) The structure of the CstF-77 homodimer provides insights into CstF assembly. *Nucleic Acids Res.*, **35**, 4515–4522.
14. Takagaki, Y. and Manley, J.L. (2000) Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol. Cell Biol.*, **20**, 1515–1525.
15. Moreno-Morcillo, M., Minvielle-Sébastien, L., Fribourg, S. and MacKereth, C.D. (2011) Locked tether formation by cooperative folding of Rna14p monkeytail and Rna15p hinge domains in the yeast CF IA complex. *Structure*, **19**, 534–545.
16. Paulson, A.R. and Tong, L. (2012) Crystal structure of the Rna14-Rna15 complex. *RNA*, **18**, 1154–1162.
17. Pérez Cañadillas, J.M. and Varani, G. (2003) Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.*, **22**, 2821–2830.
18. Pancevac, C., Goldstone, D.C., Ramos, A. and Taylor, I.A. (2010) Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors. *Nucleic Acids Res.*, **38**, 3119–3132.
19. Qu, X., Perez-Canadillas, J.-M., Agrawal, S., De Baecke, J., Cheng, H., Varani, G. and Moore, C. (2007) The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing. *J. Biol. Chem.*, **282**, 2101–2115.
20. Moreno-Morcillo, M., Minvielle-Sébastien, L., Mackereth, C. and Fribourg, S. (2011) Hexameric architecture of CstF supported by CstF-50 homodimerization domain structure. *RNA*, **17**, 412–418.
21. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
22. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
23. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**, 235–242.
24. DiMaio, F., Terwilliger, T.C., Read, R.J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H.L. *et al.* (2011) Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature*, **473**, 540–543.
25. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 213–221.
26. Söding, J., Biegert, A. and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, 244–248.
27. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 486–501.
28. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**, 355–367.
29. Takagaki, Y. and Manley, J.L. (1994) A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein. *Nature*, **372**, 471–474.
30. Bai, Y., Auperin, T.C. and Tong, L. (2007) The use of in situ proteolysis in the crystallization of murine CstF-77. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **63**, 135–138.
31. MacDonald, C.C., Wilusz, J. and Shenk, T. (1994) The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol. Cell Biol.*, **14**, 6647–6654.
32. Beyer, K., Dandekar, T. and Keller, W. (1997) RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA. *J. Biol. Chem.*, **272**, 26769–26779.
33. McDevitt, M.A., Hart, R.P., Wong, W.W. and Nevins, J.R. (1986) Sequences capable of restoring poly(A) site function define two distinct downstream elements. *EMBO J.*, **5**, 2907–2913.
34. Chou, Z.F., Chen, F. and Wilusz, J. (1994) Sequence and position requirements for uridylylate-rich downstream elements of polyadenylation signals. *Nucleic Acids Res.*, **22**, 2525–2531.
35. Bienroth, S., Wahle, E., Suter-crazzolaras, C. and Keller, W. (1991) Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. *J. Biol. Chem.*, **266**, 19768–19776.
36. Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J. and Shenk, T. (1990) A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes Dev.*, **4**, 2112–2120.
37. Rügsegger, U., Beyer, K. and Keller, W. (1996) Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J. Biol. Chem.*, **271**, 6107–6113.
38. Vries, H., Rügsegger, U., Hübner, W., Friedlein, A., Langen, H. and Keller, W. (2000) Human pre-mRNA cleavage factor IIm contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.*, **19**, 5895–5904.
39. Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., Frank, J. and Manley, J.L. (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell*, **33**, 365–376.
40. Gil, A. and Proudfoot, N.J. (1987) Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β -globin mRNA 3' end formation. *Cell*, **49**, 399–406.
41. Legendre, M. and Gautheret, D. (2003) Sequence determinants in human polyadenylation site selection. *BMC Genomics*, **4**, 7.
42. Kleiman, F.E. and Manley, J.L. (1999) Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. *Science*, **285**, 1576–1579.
43. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M. and Bentley, D.L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, **385**, 357–361.
44. Garrido-Lecca, A., Saldi, T. and Blumenthal, T. (2016) Localization of RNAPII and 3' end formation factor CstF subunits on *C. elegans* genes and operons. *Transcription*, **7**, 96–110.