

Crystal Structure of Murine CstF-77: Dimeric Association and Implications for Polyadenylation of mRNA Precursors

Yun Bai,¹ Thierry C. Auperin,¹ Chi-Yuan Chou,² Gu-Gang Chang,² James L. Manley,¹ and Liang Tong^{1,*}

¹Department of Biological Sciences, Columbia University, New York, NY 10027, USA

²Faculty of Life Sciences, National Yang-Ming University, Taipei 112, Taiwan

*Correspondence: ltong@columbia.edu

DOI 10.1016/j.molcel.2007.01.034

SUMMARY

Cleavage stimulation factor (CstF) is a heterotrimeric protein complex essential for polyadenylation of mRNA precursors. The 77 kDa subunit, CstF-77, is known to mediate interactions with the other two subunits of CstF as well as with other components of the polyadenylation machinery. We report here the crystal structure of the HAT (*half a TPR*) domain of murine CstF-77, as well as its C-terminal subdomain. Structural and biochemical studies show that the HAT domain consists of two subdomains, HAT-N and HAT-C domains, with drastically different orientations of their helical motifs. The structures reveal a highly elongated dimer, spanning 165 Å, with the dimerization mediated by the HAT-C domain. Light-scattering studies, yeast two-hybrid assays, and analytical ultracentrifugation measurements confirm this self-association. The mode of dimerization and the relative arrangement of the HAT-N and HAT-C domains are unique to CstF-77. Our data support a role for CstF dimerization in pre-mRNA 3' end processing.

INTRODUCTION

The synthesis of mRNA requires transcription, 5' capping, splicing, and 3' polyadenylation. Polyadenylation occurs in the nuclei of all eukaryotic cells and can be divided into two distinct steps—endonucleolytic cleavage in the 3' noncoding sequence of the pre-mRNA followed by synthesis of poly(A) tail. Polyadenylation has multiple links with other steps in gene expression, including transcription (Calvo and Manley, 2003; Hirose and Manley, 2000; Maniatis and Reed, 2002; Zorio and Bentley, 2004), mRNA export from nucleus to cytoplasm (Vinciguerra and Stutz, 2004), mRNA stability, and translational efficiency (Wilusz et al., 2001). The length of the poly(A) tail of mature mRNAs in the cytoplasm can also be regulated

by polyadenylation in early development (Barnard et al., 2004).

Multiple protein factors are involved in the 3' end cleavage and polyadenylation process (Colgan and Manley, 1997; Zhao et al., 1999). In mammalian cells, the core machinery includes four subunits of cleavage and polyadenylation specificity factor (CPSF) of molecular weights 160, 100, 73, and 30 kDa, three subunits of cleavage stimulation factor (CstF) of 77, 64, and 50 kDa, cleavage factor I (CFI), cleavage factor II (CFII), poly(A) polymerase (PAP), and poly(A)-binding protein 2 (PABP2). Recent studies suggest that Fip1 is another subunit of CPSF (Kaufmann et al., 2004), and symplekin appears to play a scaffolding role (Takagaki and Manley, 2000). CPSF is required for both steps of 3' end polyadenylation, and CPSF-73 is the endoribonuclease that catalyzes the cleavage of the pre-mRNA (Mandel et al., 2006; Ryan et al., 2004).

CstF is required for the cleavage reaction but is dispensable for the synthesis of the poly(A) tail. The N-terminal region of CstF-64 contains an RNA-recognition motif (RRM) that binds a conserved G/U-rich motif downstream of the cleavage site in the pre-mRNA (Beyer et al., 1997; Takagaki and Manley, 1997). CstF-77 mediates the interactions among the subunits of the CstF complex as well as with the 160 kDa subunit of CPSF (Murthy and Manley, 1995; Takagaki and Manley, 1994, 2000). It may also participate in mRNA masking in the cytoplasm of *Xenopus* oocytes (Rouget et al., 2006). CstF-77 comprises an N-terminal HAT (*half a TPR*) domain (residues 1–550) with 12 HAT repeats (Preker and Keller, 1998) (Figure 1A) and a proline-rich segment from residue 560 to residue 630 (with 24% proline). The interactions between CstF-77 and the other two subunits of CstF are mediated by residues in the proline-rich segment (Takagaki and Manley, 2000).

The function of the HAT domain of CstF-77 is not known, although it might mediate the interactions with CPSF-160 (Takagaki and Manley, 2000). The HAT repeat is a variant of the tetratricopeptide repeat (TPR) (Preker and Keller, 1998). The TPR is a 34 residue module that is often found in tandem arrays (Goebel and Yanagida, 1991) and mediates protein-protein interactions (Lamb et al., 1995). It consists of a helix-loop-helix structural segment, with the two helices oriented in an antiparallel fashion (Das et al., 1998; Gatto et al., 2000; Lapouge et al.,

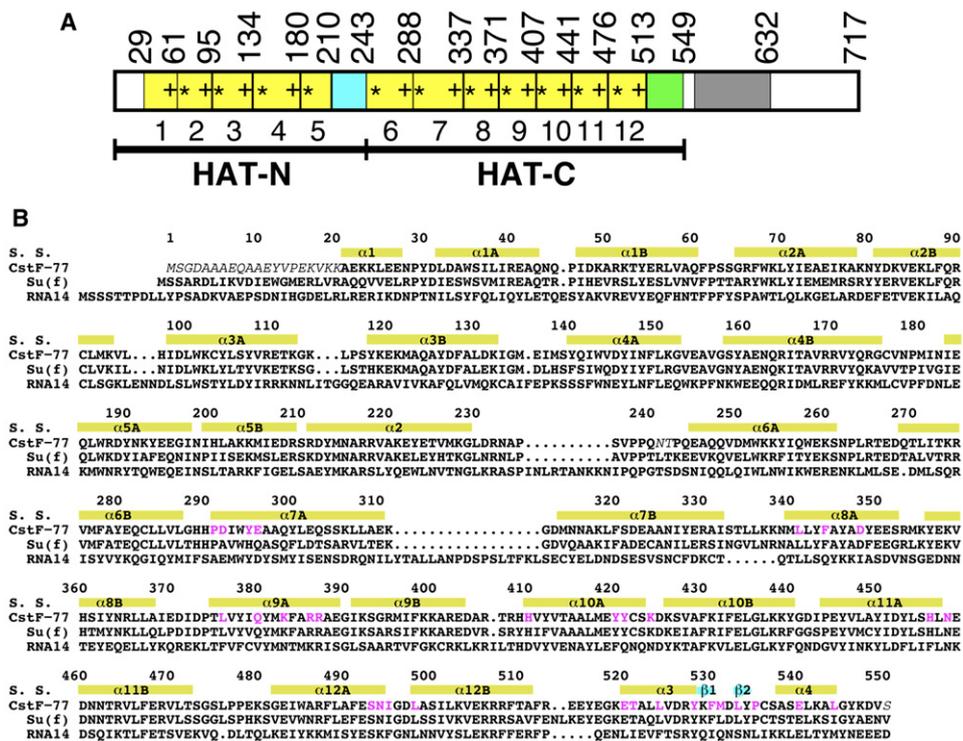


Figure 1. Sequence Alignment of CstF-77

(A) Schematic drawing of the domain organization of murine CstF-77. The HAT motifs are shown in yellow and labeled. The segment near the end of the HAT-N domain is shown in cyan, and that near the end of the HAT-C domain is shown in green. The proline-rich segment is shown in gray. The * and + symbols indicate the conserved tryptophan and tyrosine residues in each HAT repeat, respectively.

(B) Sequence alignment of murine CstF-77, *Drosophila* Suppressor of forked (Su[f]), and yeast Rna14. The secondary structure elements (S.S.) are labeled. Residues shown in magenta are in the dimer interface of the HAT domain. Residues missing in the current model of CstF-77 are shown in italic.

2000; Scheufler et al., 2000). There are two highly conserved aromatic residues in the HAT repeats, one in each helix (Preker and Keller, 1998). However, the pattern of conservation in the HAT repeats is distinct from the TPRs. Besides CstF-77, HAT repeats have been identified in Prp6, a component of the U4/U6.U5 snRNP (Liu et al., 2006).

CstF-77 is highly conserved among eukaryotes (Figure 1B) (Mitchelson et al., 1993; Takagaki and Manley, 1994). Its ortholog in *Drosophila*, known as Suppressor of forked (Su[f]), is required for pre-mRNA cleavage (Benoit et al., 2002; Mitchelson et al., 1993). Mutations in this gene can affect the utilization of alternative polyadenylation sites in unrelated genes that contain retroviral-like transposon insertions and can thereby modulate (suppress or enhance) their mutant phenotypes (Benoit et al., 2002; Mitchelson et al., 1993). The ortholog of CstF-77 in yeast, Rna14, is essential for pre-mRNA 3' end processing in that organism (Minvielle-Sebastia et al., 1994). In addition, Rna14 is tightly associated with Rna15, the yeast ortholog of CstF-64 (Minvielle-Sebastia et al., 1994; Noble et al., 2004).

Despite the identification of this large collection of protein factors that are required for pre-mRNA 3' end pro-

cessing, and their importance in cellular metabolism, very little structural information is currently available for them. This is especially true of CstF, as only the RRM and the extreme C terminus of CstF-64 have been studied in detail at the molecular level (Canadillas and Varani, 2003; Deka et al., 2005; Qu et al., 2007). We describe here structural information on CstF-77 and provide new insights into how it functions in the polyadenylation complex.

RESULTS AND DISCUSSION

Structure Determination

We first determined the crystal structure of residues 242–549 of murine CstF-77 at 2.8 Å resolution by the selenomethionyl single-wavelength anomalous diffraction (SAD) method (Hendrickson, 1991). The existence of this domain was revealed by limited proteolysis and mass spectrometry analyses on protein samples containing the entire HAT domain, the details of which are presented elsewhere (Bai et al., 2007). The bacterial expression construct actually contained residues 200–600 of CstF-77, and in situ proteolysis, with the enzyme subtilisin, was essential for the crystallization (Bai et al., 2007).

To determine the structure of the entire HAT domain of CstF-77, we coexpressed the N-terminal segment (residues 21–240) with the C-terminal segment (residues 241–550). Strong interactions between the two segments were observed, as nickel affinity chromatography copurified both proteins even though only the C-terminal segment carried a histidine tag. The coexpression strategy was probably essential for the crystallization of the HAT domain, because protein samples containing the entire HAT domain were partly aggregated in solution (Bai et al., 2007). The crystal structure was determined by the selenomethionyl SAD method (Hendrickson, 1991).

The refined structures have excellent agreement with the crystallographic data and the expected bond lengths, bond angles, and other geometric parameters (Table 1). The majority of the residues (88%) are in the most favored region of the Ramachandran plot, and none of the residues are in the disallowed region.

Overall Structure of the HAT-C Domain

Amino acid sequence analysis suggests that the CstF-77 HAT domain covers residues 1–550 (Preker and Keller, 1998). However, our limited proteolysis and mass spectrometry data showed that the C-terminal part of this segment may form a stable domain on its own (Bai et al., 2007). We refer to this region as the HAT-C domain and have confirmed its existence by the determination of its structure.

The crystal structure of the HAT-C domain (residues 242–549) contains seven pairs of antiparallel α helices for residues 242–513, forming HAT motifs 6–12 (Figure 2A). To facilitate discussion of these helices, we devised a systematic nomenclature such that, for example, the two helices in HAT motif 6 are named $\alpha 6A$ and $\alpha 6B$ (Figure 1B). Notably, the seven HAT motifs are arranged in a curved fashion, and the two helices in motif 12 are almost at a right angle to those in motif 6 (Figure 2A).

In contrast to predictions from the sequence analysis, residues 514–549 at the end of the HAT-C domain do not form an HAT motif. Instead, this segment contains two short helices ($\alpha 3$ and $\alpha 4$) and a small β hairpin ($\beta 1$ and $\beta 2$) (Figure 2A), with two residues in each strand (Figure 1B). As described below, this segment mediates dimerization of the HAT-C domain.

A Dimer of the HAT-C Domain

Our crystallographic analysis revealed a dimer of the HAT-C domain (Figure 2A), formed by the two molecules in the crystallographic asymmetric unit. A total of 2500 Å² of the surface area of each monomer is buried at the dimer interface, consistent with our biophysical characterizations of this dimer (see below). Significantly, residues in the interface of this dimer are mostly conserved in CstF-77 from other organisms (Figure 1B), suggesting that the dimer may be a common feature among these proteins. In fact, yeast Rna14 has been observed to dimerize (Noble et al., 2004), possibly mediated by the HAT domain. In addition, analysis of the shape complementarity of the dimer

Table 1. Summary of Crystallographic Information

Protein	HAT-C Domain (Residues 242–549)	HAT Domain (Residues 21–240, 243–549)
Space group	<i>P6₃22</i>	<i>R32</i>
Maximum resolution (Å)	2.8	3.0
Number of observations	587,105	57,260
R_{merge} (%) ^a	6.9 (41.3)	7.1 (19.9)
$I/\sigma I$	33.5 (6.3)	12.1 (3.1)
Resolution range used for refinement	30–2.8	30–3.0
Number of reflections ^b	52,247	15,047
Completeness (%)	97 (92)	95 (89)
R factor (%) ^c	22.4 (34.8)	21.4 (31.1)
Free R factor (%)	26.2 (41.3)	29.4 (37.7)
Rmsd in bond lengths (Å)	0.007	0.008
Rmsd in bond angles (°)	1.2	1.3
PDB accession code	2OND	2OOE

^a $R_{\text{merge}} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$. The numbers in parentheses are for the highest resolution shell.

^b The number for the HAT-C domain includes both Friedel pairs.

^c $R = \sum_h |F_h^o - F_h^c| / \sum_h F_h^o$.

interface, with the program Sc (Lawrence and Colman, 1993), gave a shape correlation score of 0.69, which is in the range of scores for oligomeric protein interfaces. The two monomers of the dimer have essentially the same conformation, with rms distance of 0.6 Å for their equivalent C α atoms.

The two monomers are arranged in an antiparallel fashion in the dimer. HAT motif 12 and the C-terminal segment (residues 514–549) of one monomer are in contact with HAT motifs 7–10 of the other monomer (Figure 2A). When looking down the two-fold axis, the dimer has an elongated shape, with a height of 90 Å and a width of 45 Å (Figure 2A). From the side, the dimer is shaped like a crescent (Figure 2B).

The C-terminal segment (residues 514–549) has a crucial role in dimerization (Figure 2C). The two helices and the β hairpin in this segment follow the curved contour of HAT motifs 7–10 of the other monomer. One face of this segment contains many hydrophobic side chains (Figure 2C), and this hydrophobic face is buried in the dimer interface. In addition, residues in and near the linker between $\alpha 12A$ and $\alpha 12B$ are located in the dimer interface (Figure 2C). His456, near the C terminus of helix $\alpha 11A$, is located next to the two-fold axis of the dimer and

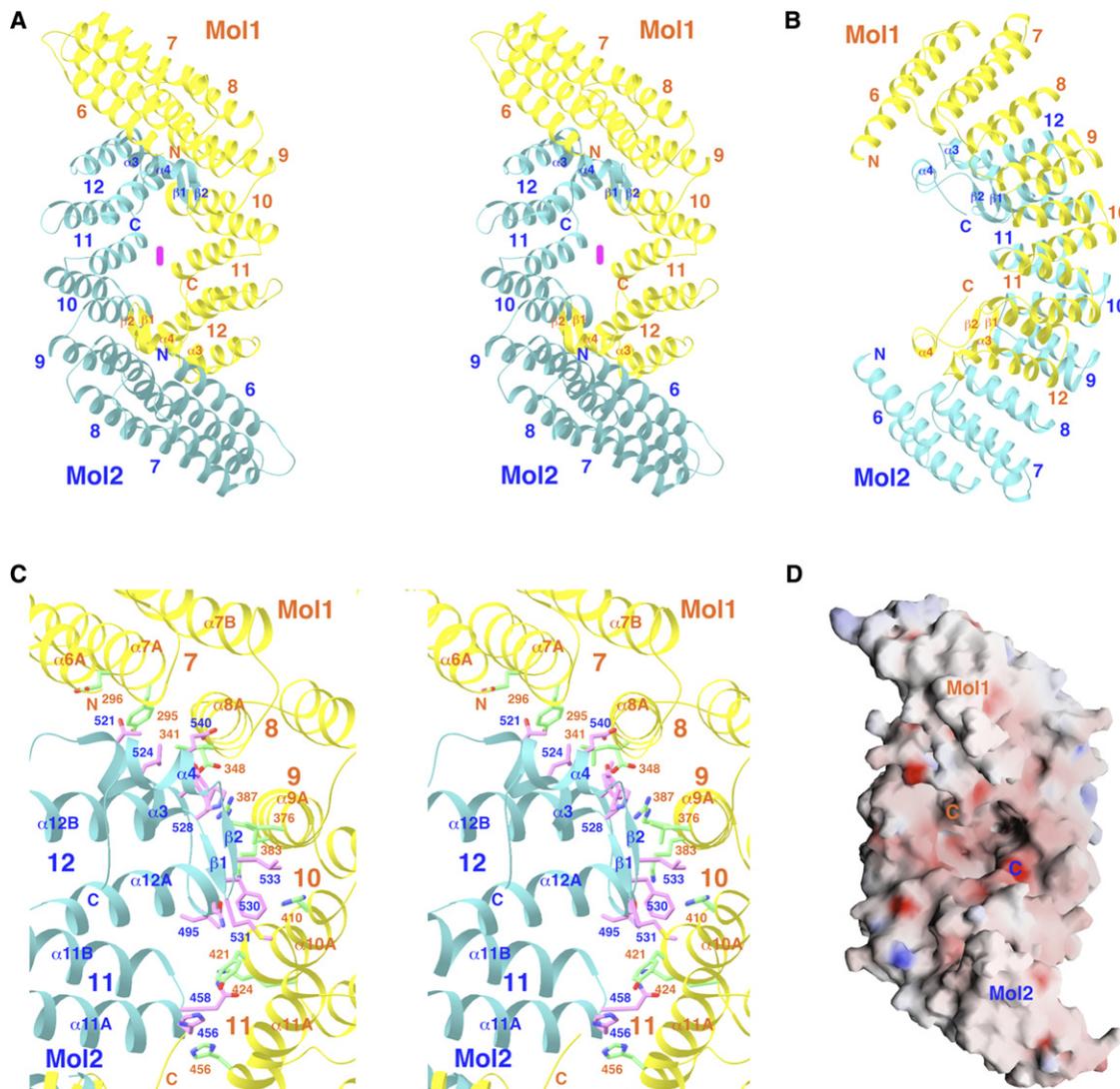


Figure 2. Structure of the HAT-C Domain of Murine CstF-77

(A) Stereo drawing of the structure of the HAT-C domain dimer. One monomer is shown in yellow, and the other is shown in cyan. The HAT motifs are labeled, and the magenta oval indicates the two-fold axis of the dimer.

(B) Structure of the HAT-C domain dimer, after 90° rotation around the vertical axis from (A).

(C) Stereo drawing showing detailed interactions in the dimer interface of the HAT-C domain. Side chains for residues contributing >40 Å² of surface area to the dimer interface are shown in green and magenta for the two monomers, respectively.

(D) Molecular surface of the HAT-C domain dimer, colored based on electrostatic potential. The C termini of the two monomers are labeled, located at the rim of the pocket. (A)–(C) are produced with Ribbons (Carson, 1987), and (D) is produced with Grasp (Nicholls et al., 1991).

interacts with its symmetry mate in the other monomer (Figure 2C).

A prominent pocket exists at the center of the HAT-C dimer (Figure 2D), formed by HAT motifs 10–12 of each monomer (Figure 2A). The opening of the pocket has a rectangular shape, with dimensions of 20 Å × 10 Å, and the depth of the pocket is about 10 Å. It is lined with both hydrophobic and hydrophilic (charged) residues, and the C-terminal residue (amino acid 549) of the monomer is located at its rim (Figure 2D). This pocket could be used for binding the additional C-terminal residues of

CstF-77 or a peptide segment from another subunit of the 3' end processing machinery. We soaked crystals of the HAT-C domain with millimolar concentrations of various amino acids but did not observe any binding based on subsequent crystallographic analyses. Further studies are needed to characterize the biological function(s) of this pocket.

The Structure of the Entire HAT Domain

The crystal structure of the entire HAT domain (residues 21–240, 243–549) of murine CstF-77 was determined at

3.0 Å resolution (Table 1). The N-terminal segment of the HAT domain, which we will refer to as the HAT-N domain (Figure 1A), contains another five HAT motifs (Figure 3A). Surprisingly, these motifs do not follow the curved pattern established by the motifs in the HAT-C domain. Instead, they make an acute angle with the HAT-C domain, representing a change of nearly 130° in the direction of motif 5 compared to motif 6 (Figure 3B). This demonstrates that the entire HAT domain is made of two subdomains and explains our earlier data indicating the presence of the HAT-C domain (Bai et al., 2007). The motifs in the HAT-N domain are also arranged in a curved fashion, although this is not as apparent when compared to the HAT-C domain, as the HAT-N domain only has five repeats.

Interactions between the HAT-N and HAT-C domains are mediated predominantly by residues 211–240 at the end of the HAT-N domain. These residues do not form an HAT motif but rather function as a linker between the two domains (Figure 1A). About 1100 Å² of the surface area of each domain is buried at this interface. Residues 211–229 form a long helix ($\alpha 2$) that is docked onto the exposed face of motif 6 in the HAT-C domain (Figure 3A). In addition, the loop following this helix (residues 230–240) interacts with motifs 6 and 7 in the HAT-C domain. In comparison, there are few direct contacts among the HAT motifs of the two domains in this interface. Our structure is based on the coexpression of an N-terminal (residues 21–240) and a C-terminal (241–550) segment of the HAT domain. However, the C terminus of the HAT-N domain and the N terminus of the HAT-C domain are located close together in the structure (Figure 3B and see Figure S1 in the Supplemental Data available with this article online), suggesting that the conformation of the entire HAT domain in the native protein should be similar to what is described here.

As expected from the structure of the HAT-C domain on its own, the entire HAT domain also forms a dimer, mediated by the HAT-C domain (Figure 3A). The structure of the HAT-C dimer in the entire HAT domain is essentially the same as that of the HAT-C dimer alone, with rms distance of 0.87 Å among their equivalent C α atoms. The HAT-N domain is positioned far from the dimer interface and makes no contributions to it (Figure 3B). In fact, the HAT dimer is much more elongated than the HAT-C dimer. The longest dimension of the HAT dimer is about 165 Å, while the width of the dimer remains at 45 Å (Figure 3A). With the addition of the HAT-N domain, the entire HAT dimer is shaped like a bow (Figure 3B).

Most of the hydrophobic residues of the HAT domain are buried in the core of the monomer or the interface of the dimer. The exposed surface of the HAT domain dimer is highly hydrophilic in nature, with predominantly charged residues (Figure S2). In fact, CstF-77 is enriched in charged and hydrophilic residues (Takagaki and Manley, 1994). The calculated isoelectric point of the HAT domain is 8.8, suggesting that it should carry a net positive charge at physiological pH.

Prior to HAT motif 1, we observed a small helix ($\alpha 1$) for residues 21–27 (Figure 3A). As our expression construct

started at residue 21, it is possible that additional residues from the N-terminal segment of CstF-77 belong to this helix (Figure 1B). This helix helps to close the exposed face of motif 1.

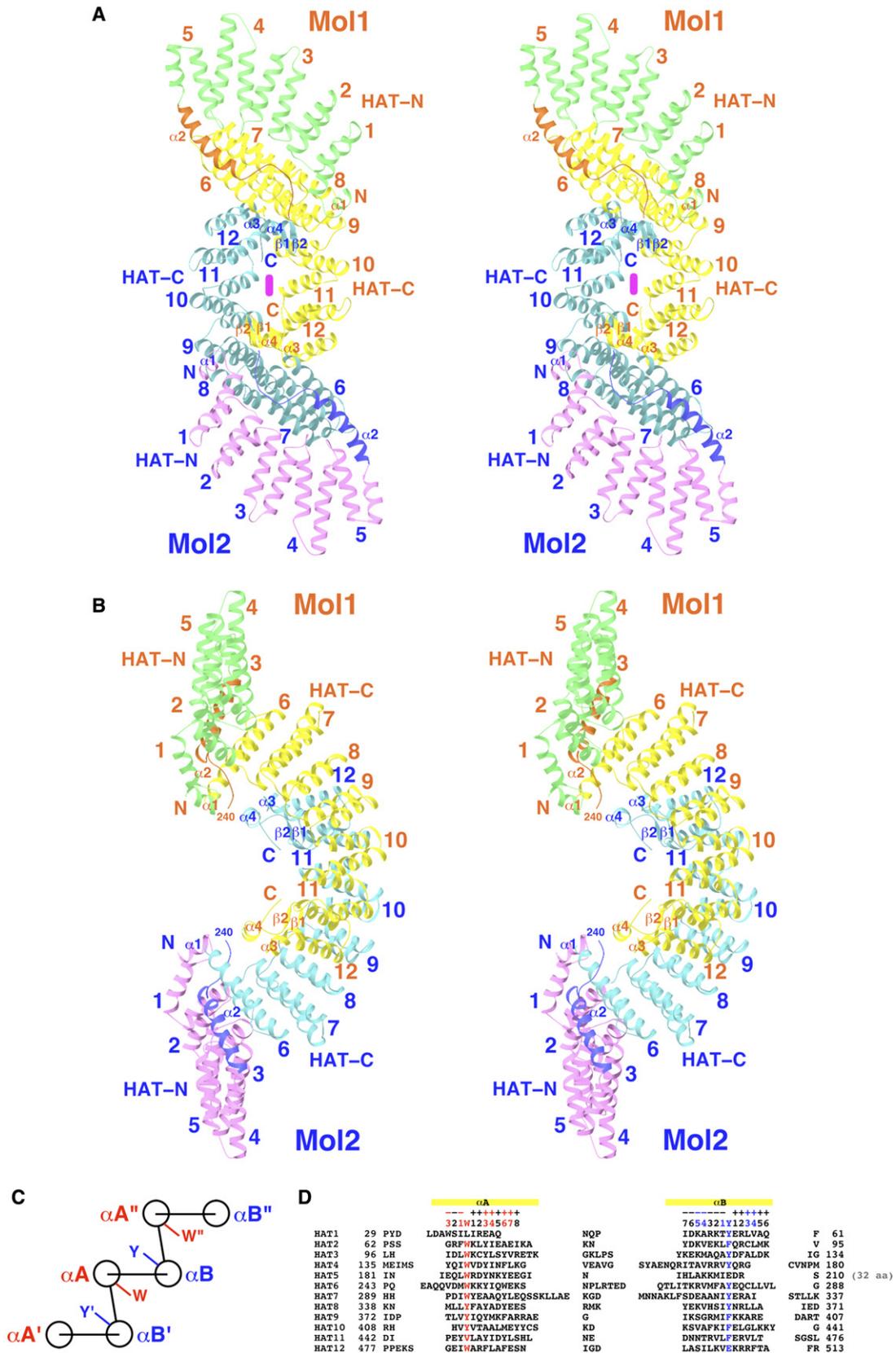
The HAT Motif

Previous sequence analysis suggests that each HAT motif has two helices, the first one containing a highly conserved tyrosine residue and the second a tryptophan residue (Preker and Keller, 1998). However, our structure shows that the tyrosine residue is in the αB helix (the second helix), while the tryptophan residue is in the αA helix (the first helix) (Figure 1B). In fact, the HAT-N and HAT-C domains are contiguous structures, and the HAT motifs are arranged such that the αA helix of one motif is in contact with the αB helix of that motif as well as the αB helix of the previous motif ($\alpha B'$) (Figure 3C). Similarly, the αB helix is in contact with the αA helix of the same motif as well as the αA helix of the following motif ($\alpha A''$) (Figure 3C). Therefore, it is somewhat arbitrary which pair of consecutive helices (αA and αB , or $\alpha B'$ and αA) is considered to belong to the same HAT motif. Our current assignment is probably more appropriate because the HAT-C domain would contain seven complete HAT motifs. With the alternative assignment, the HAT-C domain would have only six HAT motifs, with an unpaired helix at both ends.

With our assignment, each HAT motif contains two helices, the first (αA) with a highly conserved tryptophan (W) residue near its beginning and the second (αB) with a tyrosine (Y) near its middle (Figures 1A and 3D). The two helices in each repeat are oriented the same relative to each other among most of the HAT motifs (Figure S3). While most of the HAT motifs have between 34 and 37 residues, a few of them (HAT4, HAT6, and HAT7) are significantly larger, with 46–49 residues (Figure 3D). For ease of discussion, the residues in the helices are numbered relative to the conserved tryptophan and tyrosine residues (Figure 3D).

There are two types of interactions in the HAT domain, one for those between the αA and αB helices of the same motif, and the other for those between the αA and $\alpha B'$ helices of neighboring motifs. For interactions within the HAT motif, one face of the tryptophan residue in helix αA is placed next to the side chain of the Y + 3 residue (Figure S3), which is generally small and hydrophobic in nature (Figure 3D). This is a close contact; as the only case where the Y + 3 residue is β branched (valine, in motif 11), the tryptophan residue is replaced by a valine as otherwise there would be steric clashes between them (Figure 3D). One face of the tyrosine residue in helix αB is positioned against the side chain of the W + 4 residue (Figure S3), generally also small and hydrophobic (Figure 3D). The W + 3 residue is frequently a tyrosine (Figure 3D) and interacts with the Y – 1 residue (Figure S3). There are additional interactions for residues near the ends of these helices (Figure S3).

For interactions between neighboring HAT motifs, the other face of the tryptophan residue interacts with the



$Y' + 4$ residue (Figure S3), which most frequently is a leucine residue (Figure 3D). The side chain of the $W + 3$ residue (most frequently tyrosine) interacts with the amide bond linking residues Y' and $Y' + 1$, and the Y' residue interacts with the amide bond linking residues $W + 2$ and $W + 3$ (Figure S3). The $W - 1$ residue has hydrophobic interactions with the $Y' + 3$ residue. Overall, these interactions within and between HAT motifs bury the conserved tryptophan and tyrosine residues and produce the hydrophobic core of the HAT domain.

Unique Mode of Dimerization and Organization of Helical Repeats

The HAT domain bears structural resemblance to the TPR domains (Das et al., 1998; Fukuhara et al., 2005; Gatto et al., 2000; Jinek et al., 2004; Lapouge et al., 2000; Pekkala et al., 2004; Rice and Brunger, 1999; Scheufler et al., 2000) as well as other helical repeat domains, such as that in farnesyltransferase (Park et al., 1997), the MalT ATP-dependent transcriptional activator (Steegborn et al., 2001), the cysteine-rich proteins of *Helicobacter pylori* (Luthy et al., 2004), the Pumilio homology domain (Wang et al., 2001), the Armadillo repeats, and the HEAT repeats. However, in contrast to the dimeric association observed for the HAT domain, most of these other helical repeats are monomeric. Moreover, the relative positions of the HAT-N and HAT-C domains are unique to our structure and have not been observed in other structures containing helical repeats.

The HAT and TPR domains share only 6%–12% amino acid sequence identity, as calculated with the program Dali (Holm and Sander, 1993). As in the TPR domain, the helices in the HAT domain are arranged in a right-handed superhelix, and the curvature of the HAT domain is similar to that of TPR domains (Figure S4). However, the orientation of the individual repeats in these domains can be very different (Figure S4). This is also reflected by the exceedingly large rms distances among equivalent $C\alpha$ atoms of these structures, which range between 5 and 10 Å as calculated by Dali.

The TPR structures support the hypothesis that the pocket in the center of the HAT-C dimer (Figure 2D) binds a peptide segment. For example, in the structure of human PEX5 in complex with the peroxisomal targeting signal-1 (PTS1) peptide (Gatto et al., 2000), the three N-terminal TPRs are juxtaposed against the three C-terminal TPRs in the same monomer (Figure S4). This overall arrangement is reminiscent of the HAT motifs in the dimer of the HAT-C domain (Figure 2A), although the relative orientations of the two TPR halves in PEX5 are different from the HAT-C

dimer. There is a tunnel in the center of the TPR structure of PEX5, and the PTS1 peptide is bound here, along the helices in the TPRs (Figure S4). In the structures of the adaptor protein Hop in complex with peptides from Hsp70 and Hsp90, the peptides are also bound along the helices in the TPRs (Scheufler et al., 2000).

Biochemical and Biophysical Evidence for HAT Domain Self-Association

Our structural studies revealed a dimeric association of the HAT and HAT-C domains. Because this finding has significant implications for the function of CstF in 3' processing, we wished to obtain additional evidence for the existence of this dimer. We therefore characterized the oligomeric state of these proteins in solution. Protein samples (at 10 mg/ml concentration) were run through a gel filtration column, and the eluates were examined by static light scattering. Both the HAT-C domain (as an MBP fusion protein) and the HAT domain were found to be dimeric (data not shown). No monomeric species were observed in the eluates at the concentrations used for these experiments.

We next carried out yeast two-hybrid assays to obtain further evidence for the dimerization of the HAT domain (Figure 4A). Various segments of the HAT domain were cloned into the DNA-binding domain (bait) and activation domain (prey) vectors. Interactions between them were examined qualitatively using the *HIS3* reporter (Figure 4B), as well as quantitatively with *LacZ* as the reporter (Figure 4C). Both assays produced qualitatively similar results and show that residues 241–555, covering the entire HAT-C domain, have strong self-association (Figure 4B). Residues 241–513, lacking the C-terminal segment of the HAT-C domain, displayed weak self-association, while residues 241–480, lacking HAT motif 12 and the C-terminal segment, did not show any association with either residues 241–513 or 241–555 (Figure 4B). The self-association of the 241–480 segment could not be tested, as the bait vector produced very high background signals (data not shown). These observations support our structural data showing that motif 12 and the C-terminal segment of the HAT-C domain are crucial for dimerization (Figure 2A).

The yeast two-hybrid assay also demonstrated a strong interaction between the HAT-N (residues 1–240) and HAT-C (241–555) domains (Figure 4B). In contrast, residues 1–240 did not show any self-association in the assay (Figure 4B). These data are fully consistent with our structural observations.

Figure 3. Structure of the Entire HAT Domain of Murine CstF-77

(A) Stereo drawing of the structure of the HAT domain dimer. The HAT-C domain of one monomer is shown in yellow, and the other is shown in cyan. The HAT-N domain of one monomer is shown in green, and the other is shown in magenta. The linker between the HAT-N and HAT-C domains (residues 211–240) is shown in orange and blue, respectively.

(B) Structure of the HAT domain dimer, after 90° rotation around the vertical axis from (A).

(C) Schematic drawing of the arrangement of the HAT motifs in CstF-77.

(D) Alignment of the sequences of the αA and αB helices of the 12 HAT motifs in murine CstF-77. The numbering schemes for the residues in the two helices are indicated at the top. (A) and (B) are produced with Ribbons (Carson, 1987).

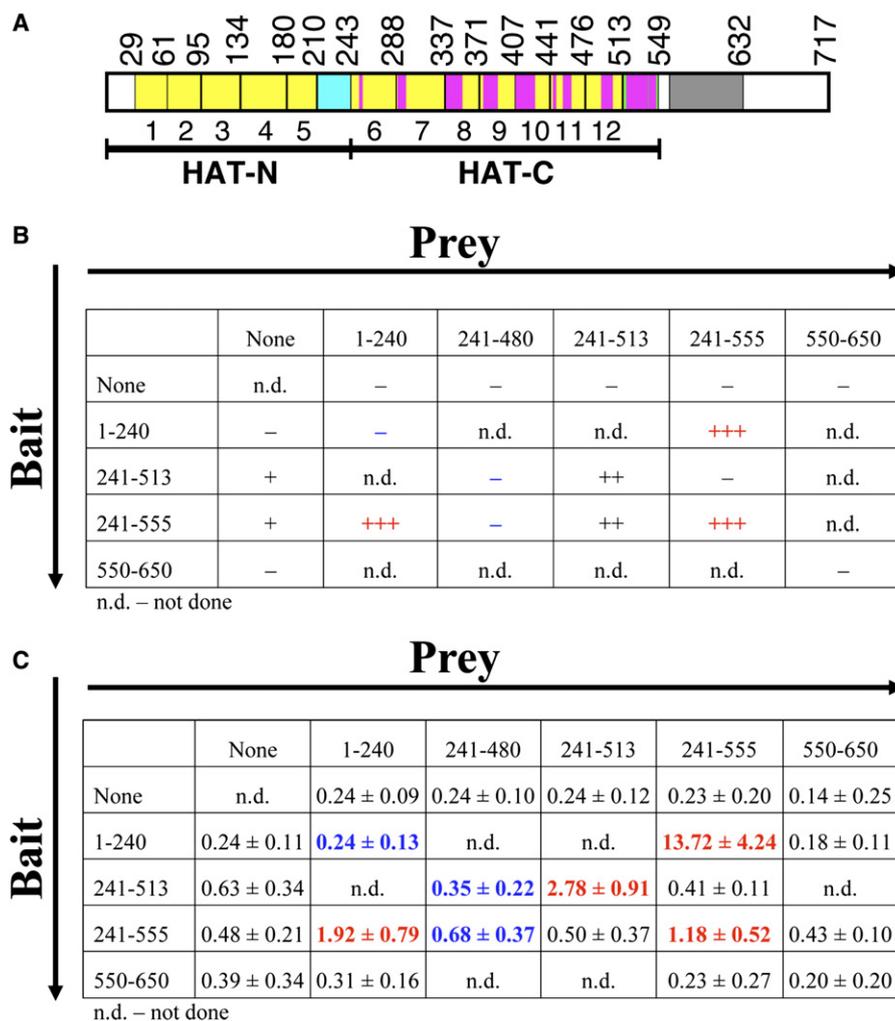


Figure 4. HAT Domain Yeast Two-Hybrid Assays

(A) Schematic drawing of the HAT motifs in CstF-77. Residues that contribute to the dimer interface are indicated with the vertical bars in magenta. (B) Yeast two-hybrid assay for various segments of CstF-77, with the *HIS3* reporter. Strong interactions between bait and prey are indicated with three plus signs in red. A representative experiment from three different repeats is shown.

(C) Yeast two-hybrid assay for various segments of CstF-77, with the *LacZ* reporter. Strong interactions between bait and prey are indicated in red, and expected weak interactions are shown in blue. The average and standard deviation from five different repeats are shown.

To obtain quantitative data on the stability of the HAT domain dimers, we determined their K_d values by sedimentation velocity (SV, Figure 5A) and sedimentation equilibrium (SE, Figure 5B) analytical ultracentrifugation (AUC) experiments. The SV data were obtained at three different protein concentrations (Figures S5 and S6). The HAT-C domain, the entire HAT domain, and the HAT domain with two different portions of the proline-rich segment were examined in these experiments. K_d values for the dimers were obtained from global fitting to the SV and SE AUC data (Figure 5C).

The AUC data confirm the dimerization of the entire HAT domain, with a K_d of 0.25 μ M (Figure 5C). The HAT-C domain appears to have the highest affinity for self-association, with K_d of 0.028 μ M. With the inclusion of the proline-rich segment, the K_d values are somewhat higher, about

2 μ M. The dissociation rate constants of the dimer for the four protein samples are similar (Figure 5C). The K_d value measured for murine CstF-77 here is consistent with that for the dimer of full-length yeast Rna14 (Noble et al., 2004). Therefore, the AUC data suggest that full-length CstF-77 and Rna14 have moderate affinity for dimerization.

Characterization of the Interactions between CstF-77 and Other Proteins

The interactions between CstF-77 and the other two subunits of CstF (CstF-64 and CstF-50) are mediated by the proline-rich segment, based on far western experiments (Takagaki and Manley, 2000). We have confirmed the interactions between the proline-rich segment of CstF-77 and the hinge region of CstF-64, which covers about

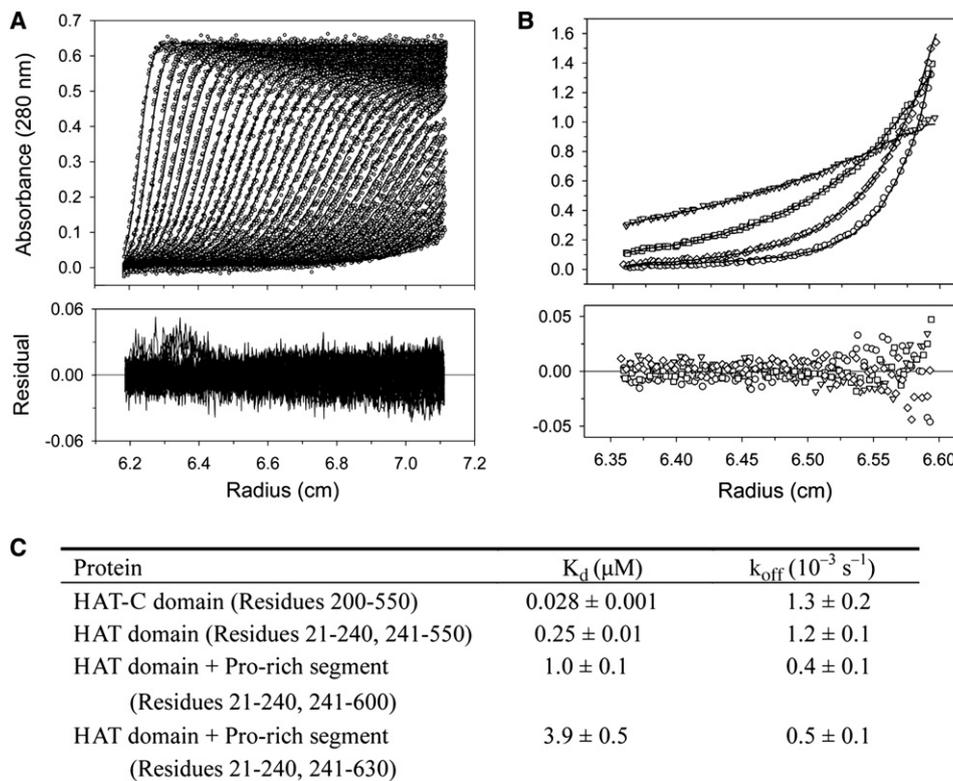


Figure 5. AUC Data on the HAT Domain

(A) Sedimentation velocity AUC data on the HAT domain (residues 21–240 and 241–550), at 7.5 μM concentration. The open circles are the observed data, and the solid lines are the global fit to the data.

(B) Sedimentation equilibrium AUC data on the HAT domain (residues 21–240 and 241–550), at 7.5 μM concentration. The raw data at rotor speeds of 8,000 (triangles), 12,000 (squares), 16,000 (diamonds), and 20,000 (circles) rpm are shown.

(C) Observed K_d and k_{off} values for the protein samples of the HAT-C domain, HAT domain, and HAT domain with proline-rich segments.

100 residues immediately following the RRM. The two proteins can be copurified by nickel affinity chromatography following coexpression in *E. coli* (Figure 6A), even though only the proline-rich segment carried a histidine tag. The stoichiometry of the two proteins in the complex appears to be 1:1, based on Coomassie staining of the protein gel (Figure 6A). We further characterized the stability of this heterodimer by AUC. The data suggest that the dimer is very stable, with K_d of about 1 nM (Y.B., C.-Y.C., G.-G.C., and L.T., unpublished data). Therefore, the proline-rich segment of CstF-77 and the hinge domain of CstF-64 may mediate the constitutive association between the two proteins. Our data are supported by observations with the yeast Rna14:Rna15 complex (Noble et al., 2004). Electron microscopy images showed that the Rna14:Rna15 heterodimer is very stable, but it can dimerize to produce (Rna14:Rna15)₂ heterotetramers.

We also characterized the interactions between CstF-77 and CPSF-160 by yeast two-hybrid assays. It has been suggested that sequences within the HAT domain are important for these interactions (Takagaki and Manley, 2000). CPSF-160 contains 16 copies of a β propeller sequence motif, and motifs IV–VIII (residues 501–750) may

be involved in binding the AAUAAA motif upstream of the cleavage site (Dichtl et al., 2002). Due to the size of this protein, we divided it into three fragments in our experiments. Our yeast two-hybrid data showed strong interactions between the HAT-C domain of CstF-77 and residues 400–973 of CPSF-160 (Figure 6B). The 400–973 segment of CPSF-160 also contains the domain for binding the AAUAAA motif (Dichtl et al., 2002). While further studies are needed to define the precise boundaries of the interacting domains between CstF-77 and CPSF-160, it is not unusual for one domain in these proteins to have multiple functions. For example, the proline-rich segment of CstF-77 interacts with both CstF-64 and CstF-50, and the hinge region of CstF-64 interacts with both CstF-77 and symplekin (Takagaki and Manley, 2000).

A Dimeric Model for the CstF Complex

Our structural, light-scattering, yeast two-hybrid, and AUC studies indicate that CstF-77 has the potential to dimerize, and its ortholog in yeast, Rna14, has also been observed to dimerize (Noble et al., 2004). Moreover, distinct lethal *Drosophila su(f)* alleles can partially complement each other, and based on this it has been suggested

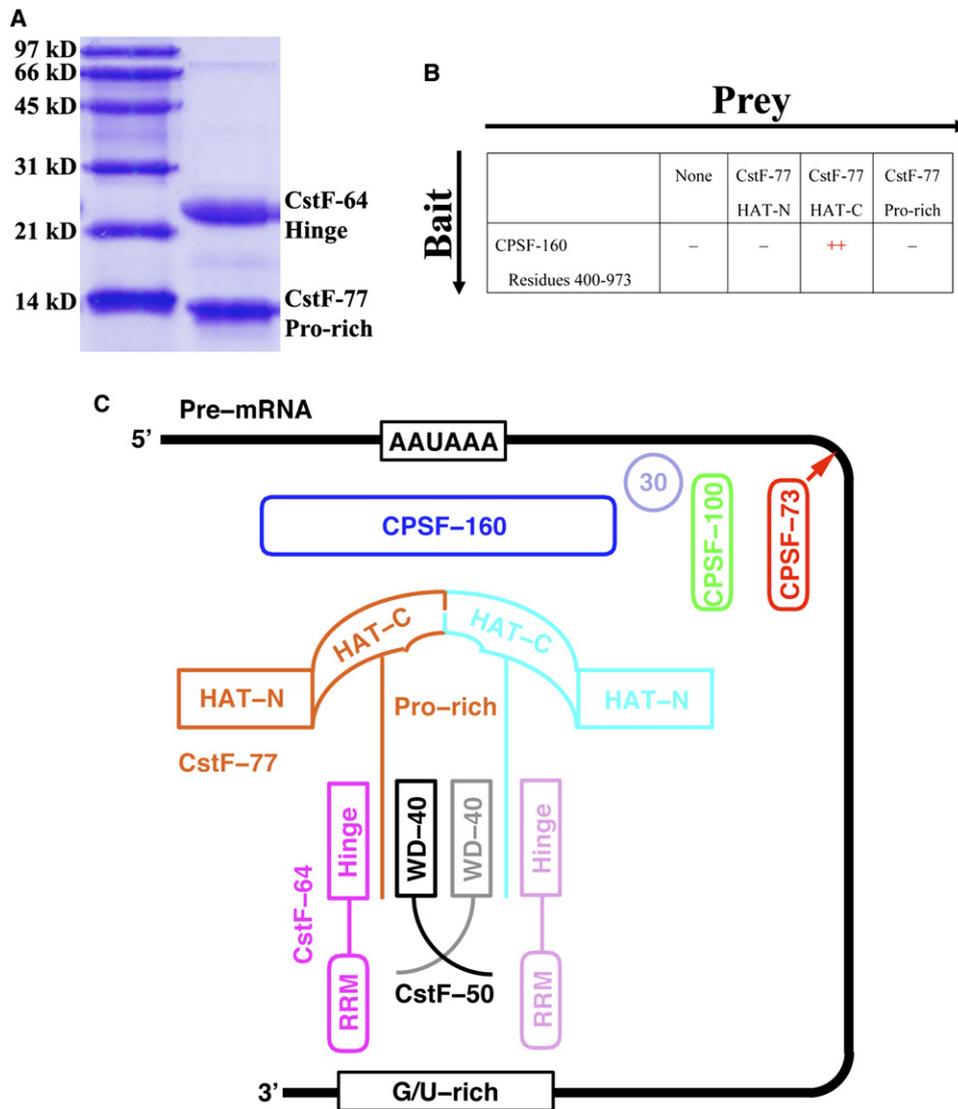


Figure 6. Model for the CstF Complex

(A) Biochemical evidence for direct interactions between the hinge domain of CstF-64 and the proline-rich segment of CstF-77. SDS gel of the two domains after coexpression and purification is shown. The two proteins appear to have 1:1 stoichiometry in the complex.

(B) Yeast two-hybrid assay for interactions between various segments of CstF-77 and CPSF-160, with the *HIS3* reporter.

(C) CstF-77, in its dimeric form, is the central component that holds the CstF complex together. The proline-rich segment interacts with the hinge region in CstF-64 and the WD-40 repeats in CstF-50 (Takagaki and Manley, 2000). The RRM in CstF-64 binds the G/U-rich motif in the pre-mRNA, while the N-terminal region of CstF-50 may mediate its self-association (Takagaki and Manley, 2000). In addition, CstF-77 interacts with CPSF-160 and brings CPSF-73, the endonuclease (Mandel et al., 2006; Ryan et al., 2004), close to the cleavage site of the pre-mRNA.

that Su(f) may self-associate during its function in pre-mRNA processing (Benoit et al., 2002; Simonelig et al., 1996). Taken together, these data support the hypothesis that CstF-77, and thereby the entire CstF complex, functions as a dimer at some point during pre-mRNA 3' end processing. The observed mode of dimerization is also supported by far western experiments showing that residues 450–550 are crucial for self-association of CstF-77 (Takagaki and Manley, 2000). At the same time, the AUC data show that the CstF-77 dimer only has moderate

stability, with K_d in the micromolar range. It is possible that this dimerization is a transient, although important, event in pre-mRNA 3' end processing.

Based on our studies, a dimeric model for the CstF complex would have the HAT-C dimer of CstF-77 as the central platform upon which the HAT-N domain and the proline-rich segment are assembled (Figure 6C). The proline-rich segment is crucial for recruiting CstF-64 (Figure 6A) and CstF-50 (Takagaki and Manley, 2000), while the HAT domain is not involved in interacting with

these two subunits. Two characterized *su(f)* mutations are located in the proline-rich segment (Benoit et al., 2002; Simonelig et al., 1996). The formation of a wild-type:mutant heterodimer, mediated by the HAT-C domain, could still enable the recruitment of CstF-50 and CstF-64, explaining the complementation among these mutants.

In contrast to the HAT domain, the proline-rich segment does not dimerize on its own (Takagaki and Manley, 2000), as also shown by our yeast two-hybrid data (Figure 4B). However, far western experiments showed that CstF-50 may dimerize through its N-terminal region (Takagaki and Manley, 2000), which could facilitate the dimeric organization of this region of the CstF complex (Figure 6C).

An important insight of our model is the presence of two RRM in the CstF complex (Figure 6C). Previous studies have demonstrated that CstF-64 and Rna14 recognize G/U-rich motifs that contain more than ten nucleotides (Beyer et al., 1997; Canadillas and Varani, 2003; Takagaki and Manley, 1997). However, a single RRM appears to bind only 4 to 5 nucleotides (Canadillas and Varani, 2003; Wang and Tanaka Hall, 2001). Therefore, the presence of two RRMs in the CstF complex may be required to recognize the entire G/U-rich motif in the pre-mRNA substrate. Observations with the Rna14:Rna15 complex also suggest that dimerization may enhance the RNA-binding ability of the complex, possibly due to the presence of two RRMs (Noble et al., 2004).

The molecular weight of the entire CstF complex has been reported to be 200 kDa, based on glycerol density gradient centrifugation measurements (Takagaki et al., 1989). This is consistent with a monomeric association of the three subunits. However, the moderate stability of the dimer suggests that it may have dissociated during the centrifugation or the purification steps.

Our yeast two-hybrid data show that the HAT-C domain may also be important for the interactions with CPSF-160 (Figure 6C). Because CPSF-160 contains many copies of the β propeller motif, it is possible that a dimeric CstF complex can interact with several of these repeats in a single CPSF-160 molecule. Further studies are needed to clarify the relative stoichiometry between the CstF and CPSF complexes in the 3' end processing machinery. For example, the possibility that the interaction with CPSF-160 is mediated by the dimer interface of the HAT-C domain, such that CstF-77 is in fact monomeric in this machinery, cannot be ruled out. By this scenario, to explain the *su(f)* complementation data (Benoit et al., 2002; Simonelig et al., 1996), we would suggest that CstF-77 dimerization plays an essential but transient role at another step in the polyadenylation reaction.

In summary, we have produced structures of the HAT domain of CstF-77 as well as its C-terminal subdomain (HAT-C domain). The helical motifs in the HAT-N and HAT-C domains are arranged in a right-handed superhelix. A dimeric association of the HAT domain was revealed by the structure, primarily mediated by the HAT-C domain. Solution light scattering, yeast two-hybrid assays, and AUC have confirmed this self-association and suggest

that CstF functions as a dimer during a crucial stage in polyadenylation. The unique, highly elongated HAT domain structure provides a central scaffold for the recruitment of other protein factors involved in pre-mRNA 3' end processing.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

Residues 200–600 of murine CstF-77 (the HAT-C domain) were subcloned into the pET28a vector (Novagen) and overexpressed in *E. coli* at 20°C by the addition of 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG) at OD₆₀₀ of 0.8. The expression construct introduced a hexahistidine tag at the N terminus of the protein. The soluble protein was eluted from a nickel-agarose affinity column using a buffer containing 20 mM Tris (pH 8.5), 250 mM NaCl, and 150 mM imidazole and further purified by gel filtration chromatography in a running buffer of 20 mM Tris (pH 8.5) and 200 mM NaCl. The protein was concentrated to 6 mg/ml in a buffer containing 20 mM Tris (pH 8.5), 200 mM NaCl, and 5% (v/v) glycerol and stored at –80°C.

Residues 21–240 and 241–550 of murine CstF-77 were subcloned into the pET28a vector (Novagen) for coexpression, with a ribosomal-binding site upstream of each DNA sequence. A histidine tag is placed at the N terminus of residues 241–550. The proteins were overexpressed in *E. coli* Rosetta2 (DE3) cells, and purified by nickel-agarose affinity chromatography followed by gel filtration chromatography. The protein was concentrated to 10 mg/ml in a buffer containing 20 mM Tris (pH 8.5), 200 mM NaCl, 5% (v/v) glycerol, and 5 mM DTT.

The selenomethionyl protein was produced in B834(DE3) cells (Novagen), grown in defined LeMaster media supplemented with selenomethionine (Hendrickson et al., 1990), and purified following the same protocol as that for the native protein.

Protein Crystallization

Crystals of the HAT-C domain (residues 200–600) of murine CstF-77 were obtained at 20°C by the sitting-drop vapor diffusion method. The reservoir solution contained 200 mM sodium tartrate and 20% (w/v) PEG3350. The drop also contained 10 mM EDTA and 5000:1 weight ratio subtilisin. The crystals were cryoprotected with 55% (w/v) D-glucose (in water) and then flash frozen in liquid nitrogen for data collection at 100 K. Details on the crystallization of the HAT-C domain are presented elsewhere (Bai et al., 2007).

Crystals of coexpressed HAT-N and HAT-C domains were obtained at 4°C by the sitting-drop vapor diffusion method. The reservoir solution contained 50 mM Bis-Tris (pH 6.0), 50 mM ammonium sulfate, and 30% (v/v) pentaerythritol ethoxylate (15/4 EO/OH). The crystals were protected by a cryosolution with 50 mM Bis-tris (pH 6.0), 50 mM ammonium sulfate, and 40% (v/v) pentaerythritol ethoxylate and flash frozen in liquid nitrogen.

Data Collection and Processing

X-ray diffraction data were collected on an ADSC charge-coupled device at the X4A beamline of National Synchrotron Light Source (NSLS). The diffraction images were processed and scaled with the HKL package (Otwinowski and Minor, 1997). A selenomethionyl SAD data set to 2.8 Å resolution was collected for the HAT-C domain. The crystal belongs to space group *P*6₂22, with cell dimensions of *a* = *b* = 155.4 Å and *c* = 161.6 Å. There are two molecules of the HAT-C domain in the crystallographic asymmetric unit.

A selenomethionyl SAD data set to 3.2 Å resolution and a native data set to 3.0 Å resolution were collected for the entire HAT domain. The crystal belongs to space group *R*32, with cell dimensions of *a* = *b* = 105.7 Å and *c* = 358.2 Å. There is one molecule of the HAT domain in the crystallographic asymmetric unit. The data processing statistics are summarized in Table 1.

Structure Determination and Refinement

The Se sites were located from the SAD data with the program SnB (Weeks et al., 2003). Reflection phases were calculated and improved with the program SOLVE/RESOLVE (Terwilliger, 2003), which also automatically located a portion of the residues in the crystal. The atomic model was fit into the electron density with the program O (Jones et al., 1991). For the HAT-C domain, the second molecule in the asymmetric unit was located by the molecular replacement method, with the program COMO (Jogl et al., 2001). The structure refinement was carried out with the program CNS (Brunger et al., 1998). The statistics on the structure refinement are summarized in Table 1.

Yeast Two-Hybrid Assays

Yeast two-hybrid experiments were conducted using the MATCH-MAKER GAL4 two-hybrid system (Clontech). Various segments of murine CstF-77 were subcloned into vector pGAD424 as preys and vector pGBT9 as baits. Bait and prey plasmids were then cotransformed into *S. cerevisiae* HF7c strain using the LiAc yeast transformation procedure.

The *HIS3* gene was used as a reporter in one binding assay. In this assay, transformants were streaked on *His*⁻ plates. Interaction between bait and prey will induce the expression of *HIS3* gene, enabling transformants to grow on such plates. To reduce background noise in this assay, 5–30 mM 3-amino-1,2,4-triazole was included in the plates.

The *LacZ* gene was used as a reporter for the β -galactosidase assay in liquid phase with ONPG as substrate. In this assay, clones were grown, harvested, and lysed by freeze thaw. The β -galactosidase activity was determined at 30°C, and the product absorbance at 420 nm was normalized for cell density (divided by absorbance at 600 nm).

Analytical Ultracentrifugation

The detailed protocols of the AUC experiments are described in the Supplemental Data. Briefly, AUC experiments were performed using a Beckman model XL-A analytical ultracentrifuge (Fullerton, CA). Three different concentrations were used for each protein sample in the SV experiment, at 20°C with a rotor speed of 42,000 rpm. Absorbance of the sample at 280 nm was monitored in a continuous mode. Four different rotor speeds (8,000, 12,000, 16,000, and 20,000 rpm) were used for the SE experiments. Absorbance at 280 nm was measured with time interval of 8–10 min until the sample reached equilibrium. The K_D and K_{off} values were obtained from global fitting to the SV and SE AUC data, using the program SEDPHAT (Schuck, 2003).

Supplemental Data

Supplemental Data include six figures, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at <http://www.molecule.org/cgi/content/full/25/6/863/DC1/>.

ACKNOWLEDGMENTS

We thank Randy Abramowitz and John Schwanof for setting up the X4A beamline at the NSLS, and Emanuel Rosonina for help with the yeast two-hybrid assays. This research is supported in part by grants from the NIH.

Received: August 28, 2006
Revised: December 20, 2006
Accepted: January 22, 2007
Published: March 22, 2007

REFERENCES

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.

Barnard, D.C., Ryan, K., Manley, J.L., and Richter, J.D. (2004). Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell* 119, 641–651.

Benoit, B., Juge, F., Iral, F., Audibert, A., and Simonelig, M. (2002). Chimeric human CstF-77/Drosophila Suppressor of forked proteins rescue suppressor of forked mutant lethality and mRNA 3' end processing in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 99, 10593–10598.

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.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54, 905–921.

Calvo, O., and Manley, J.L. (2003). Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* 17, 1321–1327.

Canadillas, J.M.P., and Varani, G. (2003). Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.* 22, 2821–2830.

Carson, M. (1987). Ribbon models of macromolecules. *J. Mol. Graph.* 5, 103–106.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* 11, 2755–2766.

Das, A.K., Cohen, P.W., and Barford, D. (1998). The structure of the tetrapeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* 17, 1192–1199.

Deka, P., Rajan, P.K., Canadillas, J.M.P., and Varani, G. (2005). Protein and RNA dynamics play key roles in determining the specific recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *J. Mol. Biol.* 347, 719–733.

Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S., and Keller, W. (2002). Yhh1p/Ctf1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J.* 21, 4125–4135.

Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. (2005). SMG7 is a 14-3-3-like adaptor in the nonsense-mediated decay pathway. *Mol. Cell* 17, 537–547.

Gatto, G.J., Jr., Geisbrecht, B.V., Gould, S.J., and Berg, J.M. (2000). Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nat. Struct. Biol.* 7, 1091–1095.

Goebel, M., and Yanagida, M. (1991). The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem. Sci.* 16, 173–177.

Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 254, 51–58.

Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* 9, 1665–1672.

Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev.* 14, 1415–1429.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123–138.

Jinek, M., Rehwinkel, J., Lazarus, B.D., Izaurralde, E., Hanover, J.A., and Conti, E. (2004). The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α . *Nat. Struct. Mol. Biol.* 11, 1001–1007.

Jogl, G., Tao, X., Xu, Y., and Tong, L. (2001). COMO: A program for combined molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* 57, 1127–1134.

- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
- Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. (2004). Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J.* **23**, 616–626.
- Lamb, J.R., Tugendreich, S., and Hieter, P. (1995). Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* **20**, 257–259.
- Lapouge, K., Smith, S.J.M., Walker, P.A., Gamblin, S.J., Smerdon, S.J., and Rittinger, K. (2000). Structure of the TPR domain of p67phox in complex with Rac.GTP. *Mol. Cell* **6**, 899–907.
- Lawrence, M.C., and Colman, P.M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.
- Liu, S., Rauhut, R., Vornlocher, H.-P., and Luhrmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA* **12**, 1418–1430.
- Luthy, L., Grutter, M.G., and Mittl, P.R.E. (2004). The crystal structure of helicobacter cysteine-rich protein C at 2.0 Å resolution: similar peptide-binding sites in TPR and SEL1-like repeat proteins. *J. Mol. Biol.* **340**, 829–841.
- Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953–956.
- Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506.
- Minvielle-Sebastia, L., Preker, P.J., and Keller, W. (1994). RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. *Science* **266**, 1702–1705.
- Mitchelson, A., Simonelig, M., Williams, C., and O'Hare, K. (1993). Homology with *Saccharomyces cerevisiae* RNA14 suggests that phenotypic suppression in *Drosophila melanogaster* by suppressor of forked occurs at the level of RNA stability. *Genes Dev.* **7**, 241–249.
- Murthy, K.G.K., and Manley, J.L. (1995). The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev.* **9**, 2672–2683.
- Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296.
- Noble, C.G., Walker, P.A., Calder, L.J., and Taylor, I.A. (2004). Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA. *Nucleic Acids Res.* **32**, 3364–3375.
- Otwiniowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- Park, H.-W., Boduluri, S.R., Moomaw, J.F., Casey, P.J., and Beese, L.S. (1997). Crystal structure of a protein farnesyltransferase at 2.25 angstrom resolution. *Science* **275**, 1800–1804.
- Pekkala, M., Hieta, R., Bergmann, U., Kivirikko, K.I., Wierenga, R.K., and Myllyharju, J. (2004). The peptide-substrate-binding domain of collagen prolyl 4-hydroxylase is a tetratricopeptide repeat domain with functional aromatic residues. *J. Biol. Chem.* **279**, 52255–52261.
- Preker, P.J., and Keller, W. (1998). The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem. Sci.* **23**, 15–16.
- 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. Published online November 20, 2006. 10.1074/jbc.M609981200.
- Rice, L.M., and Brunger, A.T. (1999). Crystal structure of the vesicular transport protein Sec17: implications for SNAP function in SNARE complex disassembly. *Mol. Cell* **4**, 85–95.
- Rouget, C., Papin, C., and Mandart, E. (2006). Cytoplasmic CstF-77 protein belongs to a masking complex with cytoplasmic polyadenylation element-binding protein in *Xenopus* oocytes. *J. Biol. Chem.* **281**, 28687–28698. Published online August 1, 2006. 10.1074/jbc.M601116200.
- Ryan, K., Calvo, O., and Manley, J.L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* **10**, 565–573.
- Scheuffler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* **101**, 199–210.
- Schuck, P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104–124.
- Simonelig, M., Elliott, K., Mitchelson, A., and O'Hare, K. (1996). Inter-allelic complementation at the suppressor of forked locus of *Drosophila* reveals complementation between suppressor of forked proteins mutated in different regions. *Genetics* **142**, 1225–1235.
- Steebhorn, C., Danot, O., Huber, R., and Clausen, T. (2001). Crystal structure of transcription factor MalT domain III: a novel helix repeat fold implicated in regulated oligomerization. *Structure* **9**, 1051–1060.
- 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.
- Takagaki, Y., and Manley, J.L. (1997). RNA recognition by the human polyadenylation factor CstF. *Mol. Cell. Biol.* **17**, 3907–3914.
- 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.
- 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.
- Terwilliger, T.C. (2003). SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374**, 22–37.
- Vinciguerra, P., and Stutz, F. (2004). mRNA export: an assembly line from genes to nuclear pores. *Curr. Opin. Cell Biol.* **16**, 285–292.
- Wang, X., and Tanaka Hall, T.M. (2001). Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.* **2001**, 141–145.
- Wang, X., Zamore, P.D., and Hall, T.M.T. (2001). Crystal structure of a Pumilio homology domain. *Mol. Cell* **7**, 855–865.
- Weeks, C.M., Adams, P.D., Berendzen, J., Brunger, A.T., Dodson, E.J., Grosse-Kunstleve, R.W., Schneider, T.R., Sheldrick, G.M., Terwilliger, T.C., Turkenburg, M.G., and Uson, I. (2003). Automatic solution of heavy-atom substructures. *Methods Enzymol.* **374**, 37–83.
- Wilusz, C.J., Wormington, M., and Peltz, S.W. (2001). The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell Biol.* **2**, 237–246.
- Zhao, J., Hyman, L., and Moore, C.L. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63**, 405–445.
- Zorio, D.A.R., and Bentley, D. (2004). The link between mRNA processing and transcription: communication works both ways. *Exp. Cell Res.* **296**, 91–97.