

# Crystal structure of the Rna14–Rna15 complex

ASHLEY R. PAULSON and LIANG TONG<sup>1</sup>

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

## ABSTRACT

A large protein machinery is required for 3'-end processing of mRNA precursors in eukaryotes. Cleavage factor IA (CF IA), a complex in the 3'-end processing machinery in yeast, contains four subunits, Rna14, Rna15, Clp1, and Pcf11. Rna14 has a HAT (half a TPR) domain at the N terminus and a region at the C terminus that mediates interactions with Rna15. Rna15 contains a RNA recognition module (RRM) at the N terminus, followed by a hinge region. These two proteins are homologs of CstF-77 and CstF-64 in the cleavage stimulation factor (CstF) of the mammalian 3'-end processing machinery. We report the first crystal structure of Rna14 in complex with the hinge region of Rna15, and the structures of the HAT domain of Rna14 alone in two different crystal forms. The complex of the C-terminal region of Rna14 with the hinge region of Rna15 does not have strong interactions with the HAT domain of Rna14, and this complex is likely to function independently of the HAT domain. Like CstF-77, the HAT domain of Rna14 is also a tightly associated dimer with a highly elongated shape. However, there are large variations in the organization of this dimer among the Rna14 structures, and there are also significant structural differences to CstF-77. These observations suggest that the HAT domain and especially its dimer may have some inherent conformational variability.

**Keywords:** pre-mRNA processing; HAT domain; HEAT repeat; CstF-77; CstF-64; protein complex

## INTRODUCTION

3'-end processing, including cleavage and polyadenylation, is required for the maturation of most eukaryotic messenger RNA precursors (pre-mRNAs) (Colgan and Manley 1997; Zhao et al. 1999; Mandel et al. 2008). Defects in 3'-end processing affect mRNA export, stability, translation, and transcription by RNA polymerase II.

A large number of protein factors are required for 3'-end processing, with more than 20 proteins identified for this machinery in yeast. These proteins form several subcomplexes, such as cleavage factor IA (CF IA), cleavage factor II (CF II), and polyadenylation factor I (PF I). CF IA has four subunits, Rna14, Rna15, Clp1, and Pcf11, and this factor is also required for gene looping (Medler et al. 2011). Rna15 contains a RNA recognition module (RRM) at the N terminus (Pancevac et al. 2010), followed by a hinge region and a C-terminal domain (CTD) (Fig. 1A). The CTD of Rna15 interacts with Pcf11 (Qu et al. 2007), while the hinge region interacts with the C-terminal 40 residues of Rna14 (Moreno-Morcillo et al. 2011). The N-terminal region of Rna14, like its mammalian counterpart CstF-77 (Bai et al. 2007a), contains a HAT domain, formed by 12 HAT (half

a TPR) motifs (Fig. 1A). The HAT motif is related to the tetratricopeptide repeat (TPR) (Preker and Keller 1998). Proteins containing TPRs are generally involved in protein–protein interactions.

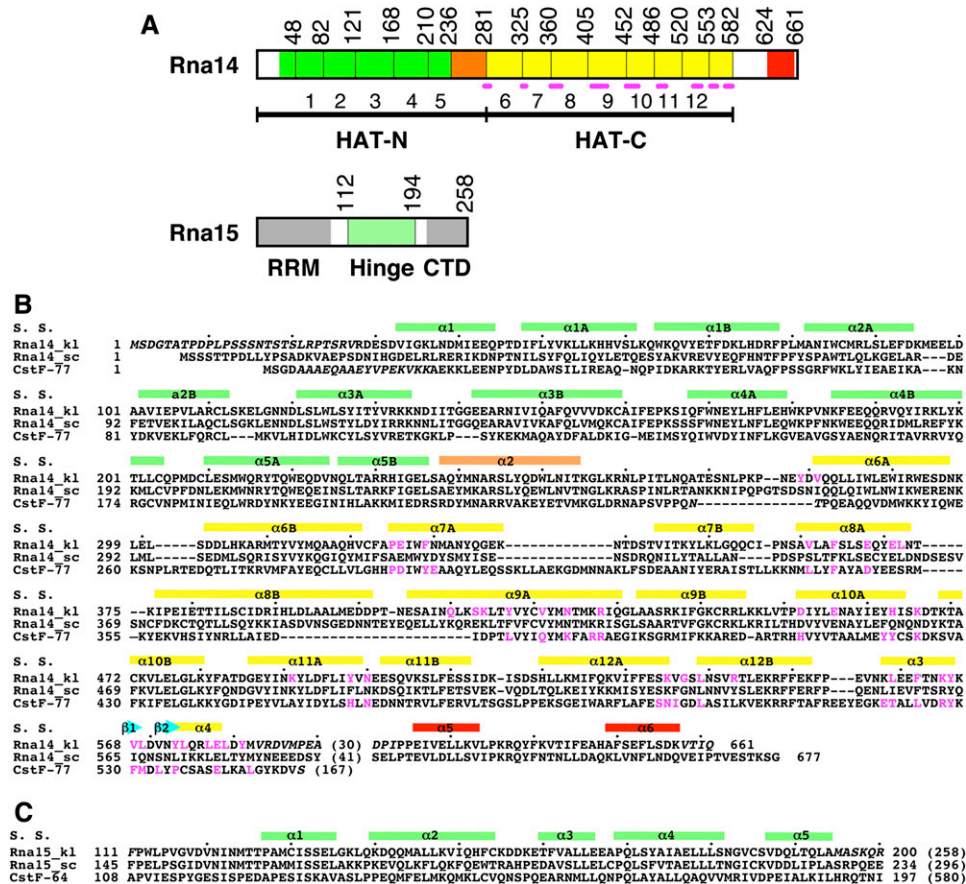
The crystal structure of the HAT domain of murine CstF-77 shows that it consists of two subdomains, a HAT-N domain with five HAT motifs and a HAT-C domain with seven HAT motifs (Bai et al. 2007a). Each HAT motif forms a pair of anti-parallel helices, and the repeats are assembled into a curved structure, with the first helix of each repeat (the  $\alpha$ A helix) on one face and the second helix ( $\alpha$ B helix) on the other face. The HAT domain is a stable dimer, mediated by the HAT-C domain, and this dimeric association has significant implications for the architecture of the mammalian 3'-end processing machinery (Bai et al. 2007a). Structural studies on the CstF-77 homolog from a protozoan parasite confirm the observations with the mammalian CstF-77 (Legrand et al. 2007).

Rna14 is essential for pre-mRNA 3'-end processing in yeast (Minvielle-Sebastia et al. 1994) and is tightly associated with Rna15 (Minvielle-Sebastia et al. 1994; Noble et al. 2004; Moreno-Morcillo et al. 2011). It shares 29% amino acid sequence identity with CstF-77, and many of the residues in the dimer interface of the HAT domain are conserved (Fig. 1B; Bai et al. 2007a). Rna14 has been shown to mediate self-association of the Rna14–Rna15 heterodimer (Noble et al. 2004). The hinge region of Rna15 is weakly conserved with that of CstF-64 (Fig. 1C). To begin

<sup>1</sup>Corresponding author.

E-mail [ltong@columbia.edu](mailto:ltong@columbia.edu).

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.032524.112>.



**FIGURE 1.** Sequence alignments of Rna14 and Rna15. (A) Schematic drawing of the domain organization of *K. lactis* Rna14 and Rna15. Motifs in the HAT-N domain of Rna14 are shown in green and labeled; those in the HAT-C domain, in yellow. The linker between the two domains is shown in orange. The dots in magenta indicate residues in the dimer interface of the HAT domain. The C-terminal region that interacts with Rna15 is shown in red. The hinge region of Rna15 is shown in light green; the RNA recognition module (RRM) and the C-terminal domain (CTD), in gray. (B) Structure-based sequence alignment of *K. lactis* Rna14 (Rna14\_kl), yeast Rna14 (Rna14\_sc), and murine CstF-77. 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 crystal structures are shown in italics. (C) Sequence alignment of the hinge region of *K. lactis* Rna15 (Rna15\_kl), yeast Rna15 (Rna15\_sc), and murine CstF-64. The total number of residues for each protein is indicated in the parenthesis.

to understand the CF IA complex at the molecular level, we have produced samples of *Kluyveromyces lactis* Rna14 (lacking only the first 17 amino acid residues) in complex with *K. lactis* Rna15 (hinge region and CTD) and obtained crystals from these samples, in three different crystal forms. We report the first structure of the Rna14–Rna15 complex and the structures of the Rna14 HAT domain alone based on these crystals.

## RESULTS AND DISCUSSION

### Overall structure of Rna14 HAT domain dimer

The recombinant *K. lactis* Rna14 used in these studies lacks only 17 residues from the N terminus of the full-length protein. In addition, the crystals for both crystal forms 1 and 2 were obtained using a sample of Rna14 in complex with the hinge region and CTD of Rna15. Based on the

crystallographic analysis, however, the C-terminal region of Rna14 (residues 583–661) was not observed. This region contains the two helices that interact with Rna15 (Fig. 1A; Moreno-Morcillo et al. 2011), and therefore, Rna15 was not observed in these crystals either. Our further studies indicated that the C-terminal region of Rna14 was likely removed by proteolysis during crystallization (possibly due to a contaminated tryptone solution; data not shown), similar to our experiences earlier of in situ proteolysis by a contaminating protease in the crystallization of yeast CPSF-100 homolog Ydh1 (Mandel et al. 2006a,b), murine CstF-77 HAT-C domain (Bai et al. 2007b), and the *E. coli* enzyme DXS (Xiang et al. 2007). This suggests that the linker between the HAT domain and the region at the C terminus of Rna14 that interacts with Rna15 is accessible to proteases and may be flexible. The flexibility of this linker has also been reported for the *S. cerevisiae* Rna14–Rna15 complex (Gordon et al. 2011).

The crystal structure of *K. lactis* Rna14 in crystal form 1 was determined at 2.8 Å resolution by the selenomethionyl single-wavelength anomalous diffraction (SAD) method (Hendrickson 1991). The structure was refined against a data set at 2.3 Å resolution collected on a native crystal. The structure of Rna14 in crystal form 2 was solved by the molecular replacement method and refined at 3.0 Å resolution. 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 (97%) are in the most favored and additional allowed regions of the Ramachandran plot for both structures.

The structures in the two crystal forms show that residues 28–582 form the HAT domain of *K. lactis* Rna14, with 12 pairs of anti-parallel helices (12 HAT motifs) (Fig. 2A). As first observed in the structure of murine CstF-77 (Bai et al. 2007a), the HAT domain consists of two subdomains. The HAT-N domain, with five pairs of helices (HAT motifs 1–5), covers residues 28–236 (Fig. 1A). There is an additional helix ( $\alpha 1$ ) at the N terminus of this domain. The HAT-C domain has seven pairs of helices (HAT motifs 6–12) for residues 281–552, and residues 553–582 form a C-terminal extension that contains two small helices ( $\alpha 3$  and  $\alpha 4$ ) and two small anti-parallel  $\beta$ -strands ( $\beta 1$  and  $\beta 2$ ). Residues 237–280, the linker between the two domains, form a helix ( $\alpha 2$ ) that interacts with motif 6 in the HAT-C domain and long loop that lies over motifs 1 and 2 in the HAT-N domain.

A dimer of Rna14 can be generated by the crystallographic twofold axis in crystal form 1 (Fig. 2A), with the two monomers arranged anti-parallel to each other. As in CstF-77, the dimer is formed by juxtaposing the concave faces of the HAT-C domains of the two monomers, while the HAT-N domains are located away from the dimer interface (Fig. 2A). The overall shape of the dimer is highly elongated. Viewed down the twofold axis, the dimer

appears to be rectangular, with a width of 50 Å and a height of 150 Å. From the side, the dimer is shaped like a bow (Fig. 2B).

Approximately 2500 Å<sup>2</sup> of the surface area of each monomer is buried at the interface of the dimer, indicating that the dimer is likely to be stable. The protein is also a dimer in solution, based on gel-filtration experiments, in agreement with observations on the CstF-77 HAT domain (Bai et al. 2007a). Residues from HAT motifs 7–12 and, especially, the C-terminal extension contribute to the dimer interface (Fig. 1A). They are generally conserved among Rna14 homologs in fungal species (Fig. 1B), consistent with the fact that *S. cerevisiae* Rna14 has been reported to dimerize as well (Noble et al. 2004). The overall architecture of the Rna14 HAT domain dimer is similar to that of CstF-77 (Bai et al. 2007a), although there are also important differences (see below).

Our structure of the Rna14 HAT domain dimer is supported by published mutagenesis data on *S. cerevisiae* Rna14 (Gordon et al. 2011), which showed that the R562E/Y563S double mutation disrupted its dimerization. These two residues are equivalent to Lys565 and Tyr566 of *K. lactis* Rna14 and are located in the center of the HAT domain dimer interface. The mutation reduced both the affinity of the Rna14–Rna15 complex for binding RNA and the activity of the 3'-end processing machinery for cleavage and polyadenylation of pre-mRNA (Gordon et al. 2011), demonstrating the functional importance of the dimerization of Rna14.

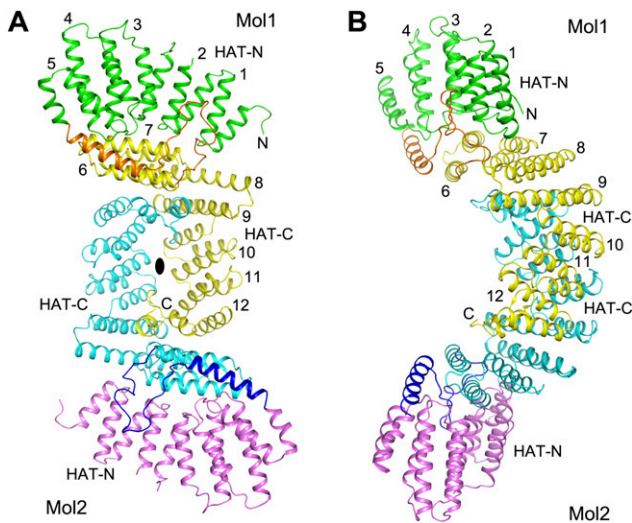
### Overall structure of the Rna14–Rna15 complex

Crystal form 3 was obtained under the same crystallization condition that produced crystal forms 1 and 2, but using a fresh solution of tryptone to exclude contaminating proteases. These crystals were small and showed very weak X-ray diffraction. After extensive efforts at optimization (which

**TABLE 1.** Summary of crystallographic information

| Crystal form                      | 1                                   | 2                                             | 3                   |
|-----------------------------------|-------------------------------------|-----------------------------------------------|---------------------|
| Protein                           | Rna14 HAT domain                    | Rna14 HAT domain                              | Rna14–Rna15 complex |
| Space group                       | C2                                  | P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> | P3 <sub>2</sub>     |
| Unit cell parameters (Å, °)       | 117.4, 59.8, 123.5 $\beta$ = 101.6° | 114.9, 241.7, 49.0                            | 162.0, 162.0, 177.5 |
| Resolution range (Å) <sup>a</sup> | 50–2.3 (2.38–2.3)                   | 50–3.0 (3.11–3.0)                             | 50–3.3 (3.42–3.3)   |
| No. of observations               | 155,841                             | 96,539                                        | 124,861             |
| <i>R</i> <sub>merge</sub> (%)     | 5.3 (24.7)                          | 10.6 (37.8)                                   | 5.5 (44.2)          |
| Redundancy                        | 4.3 (4.2)                           | 3.6 (2.7)                                     | 1.7 (1.5)           |
| <i>I</i> / $\sigma$ <i>I</i>      | 21.7 (4.2)                          | 8.7 (2.4)                                     | 10.9 (1.2)          |
| No. of reflections                | 33,535                              | 24,884                                        | 67,757              |
| Completeness (%)                  | 89 (68)                             | 88 (61)                                       | 86 (60)             |
| <i>R</i> factor (%)               | 21.0 (29.3)                         | 23.3 (41.2)                                   | 22.6 (39.3)         |
| Free <i>R</i> factor (%)          | 27.0 (33.8)                         | 29.3 (49.0)                                   | 29.8 (42.5)         |
| rms deviation in bond lengths (Å) | 0.015                               | 0.014                                         | 0.008               |
| rms deviation in bond angles (°)  | 1.7                                 | 1.7                                           | 1.3                 |
| Protein Data Bank accession code  | 4E6H                                | 4E85                                          | 4EBA                |

<sup>a</sup>The numbers in parentheses are for the highest resolution shell.



**FIGURE 2.** Structure of the HAT domain dimer of *K. lactis* Rna14. (A) Schematic drawing of the structure of the HAT domain dimer, viewed down the twofold axis (black oval). The HAT-C domain of one monomer is shown in yellow; the other, in cyan. The HAT-N domain of one monomer is shown in green; the other, in magenta. The linker between the HAT-N and HAT-C domains is shown in orange and blue, respectively. (B) Structure of the HAT domain dimer, after 90° rotation around the vertical axis from panel A. All structure figures were produced with PyMOL ([www.pymol.org](http://www.pymol.org)).

led to the inclusion of 2% benzamidine and 1% dioxane as additives) and screening through a large number of crystals, we were able to collect an X-ray diffraction data set to 3.3 Å resolution (Table 1). The crystal belongs to space group  $P3_2$ , with three Rna14 dimers in the asymmetric unit, related by a noncrystallographic threefold symmetry axis along the  $c$ -axis of the unit cell (Fig. 3A,B). The three HAT domain dimers have different conformations in the crystal (see below).

The structure of this new crystal form was solved by the molecular replacement method, using the structure of the Rna14 HAT domain dimer as the model. After crystallographic refinement, new electron density was observed that corresponded to the C-terminal region of Rna14 in complex with the hinge region of Rna15. The solution structure of this complex for *S. cerevisiae* Rna14–Rna15 (Moreno-Morcillo et al. 2011) was manually docked into the electron density and adjusted to fit the density and the sequence of *K. lactis* Rna14–Rna15. No electron density was observed for the CTD of Rna15, which was probably disordered in this crystal.

The overall structure of the complex between the C-terminal region of Rna14 and the hinge region of Rna15 (to be referred to as the complex here, for simplicity, with the HAT domain being treated as a separate entity) is similar to that observed in solution by NMR (Fig. 3C; Moreno-Morcillo et al. 2011). The rms distance is 1.5 Å for 105 equivalent Cα atoms located within 3 Å of each other (the first model of the NMR structure ensemble was used for this comparison).

The complex consists of a seven-helical bundle, with the two helices in the C-terminal region of Rna14 wrapping around the central three helices in the hinge region of Rna15 (Fig. 3C). There are recognizable differences in the positions of some of the helices and loops. The differences for the helices are likely due in part to the fact that the X-ray structure is for the *K. lactis* complex while the NMR structure is for the *S. cerevisiae* complex, as the helices are well defined by the NMR structure ensemble. The differences in the loops may also be due to inherent flexibility. Especially, the loop connecting the two helices of Rna14 shows large variations in the NMR structure ensemble and has large differences to the X-ray structure (Fig. 3C).

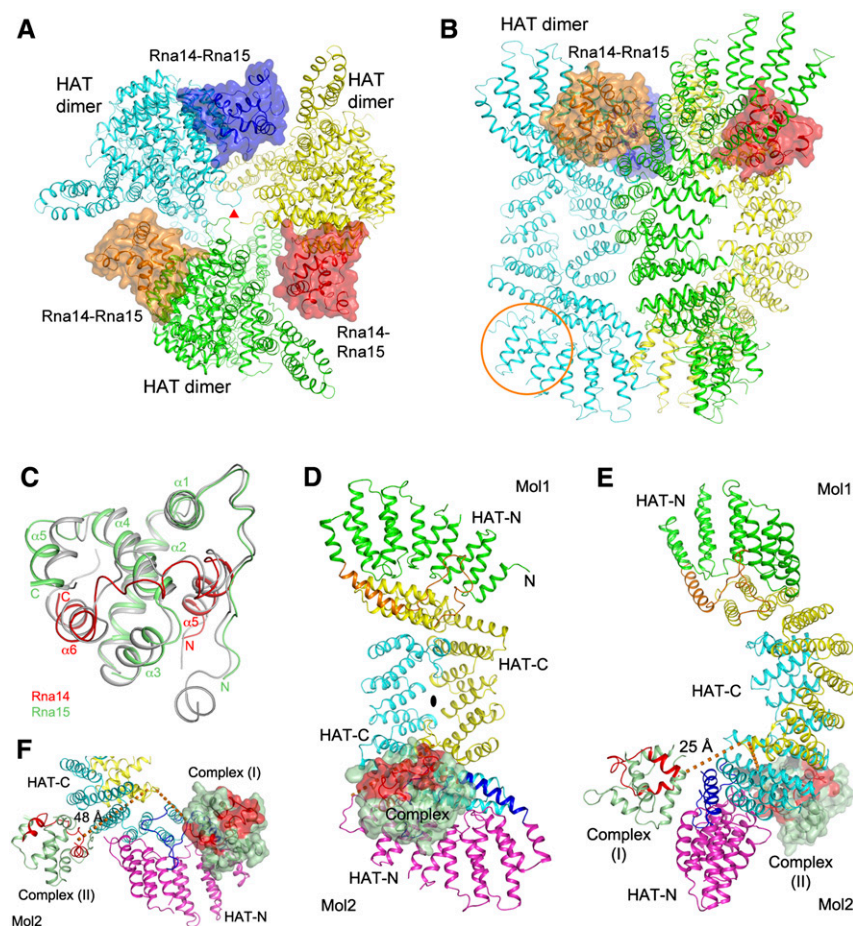
The linker (residues 583–623) between the HAT domain and the C-terminal region of Rna14 is disordered in the crystal. The molecules are arranged in the crystal such that each Rna14–Rna15 complex is located between the HAT domains of two neighboring Rna14 dimers, related by the noncrystallographic threefold symmetry axis (Fig. 3A). As a consequence, there are two possible interpretations for the position of the Rna14–Rna15 complex relative to the HAT domain. In the first arrangement, the complex is located on the concave face of the HAT domain dimer, on top of the HAT-N domain of the other monomer (Fig. 3D). The complex is on the same face as the C terminus of the HAT domain, and the distance between the last residue (582) of the HAT domain and the first residue (624) of the C-terminal region is 28 Å (Fig. 3E), which can be readily bridged by the 40 residues of the linker. However, there is only minor contact between the complex and the HAT-N domain here (Fig. 3E), with a surface area burial of only 120 Å<sup>2</sup>.

In the second arrangement, the complex is located on the convex face of the HAT domain dimer, burying 500 Å<sup>2</sup> surface area against the HAT-C domain of the other monomer of the dimer (Fig. 3F). However, this places the complex on the opposite face from the C terminus of the HAT domain. The distance between the two entities is 45 Å, and therefore, it is still possible for them to be connected by the linker. The interactions between the complex and the HAT-C domain are mostly ionic in nature, but the residues in the interface are not conserved among Rna14 and Rna15 homologs.

The first arrangement may be the more likely organization of the complex relative to the HAT domain, although the complex is stabilized by contacts with a neighboring HAT domain in the crystal. Overall, the structural observations suggest that there may be no strong interactions between the complex and the HAT domain. The position of the complex relative to the HAT domain is likely to be flexible, which is consistent with the flexibility of the linker between them. The HAT domain dimer and the Rna14–Rna15 complex therefore may function independently of each other in pre-mRNA 3'-end processing.

Only one Rna15 molecule was observed for each Rna14 dimer (Fig. 3A,D). The expected position of the complex in the second Rna14 molecule is blocked by crystal packing





**FIGURE 3.** Structure of *K. lactis* Rna14–Rna15 complex. (A) Three Rna14 HAT domain dimers (in cyan, green, and yellow) related by a noncrystallographic threefold axis (red triangle). The complex between the C-terminal region of Rna14 and the hinge region of Rna15 (in orange, red, and blue) are located between the dimers. (B) Side view of the trimeric association of the HAT domain dimers. The complex is only found between the HAT domains at the *top*, while the equivalent position for the complex between the *bottom* HAT domains (indicated with the orange circle) is blocked by crystal packing. (C) Overlay of the crystal structure of the C-terminal region of *K. lactis* Rna14 (in red) in complex with the hinge region of Rna15 (in light green) with the solution structure of the *S. cerevisiae* complex (in gray) (Moreno-Morcillo et al. 2011). The first model of the NMR structure is used for this comparison. (D) Structure of the *K. lactis* Rna14 dimer in complex with Rna15 hinge domain (light green). The C-terminal region of Rna14 is shown in red. A semi-transparent molecular surface for the complex is also shown. (E) Two possible interpretations (I and II) for the locations of the complex relative to the HAT domain dimer. The connection between the HAT domain and the C-terminal region is indicated with the dashed lines (in orange). Residues in this linker are disordered in the structure. Panel D corresponds to the first arrangement. (F) A close-up view of the second arrangement for the complex relative to the HAT domain dimer. The position of the first arrangement is shown as a molecular surface. The view is related to that of panel E by a  $\sim 180^\circ$  rotation around the vertical axis.

(Fig. 3B), which may have rendered the second Rna15 molecule (and the associated C-terminal region of Rna14) disordered. At the same time, we cannot exclude the possibility that the Rna14:Rna15 stoichiometry in the crystal may actually be 2:1. Only Rna15 carried the His-tag in the coexpression experiment, and the expression level of this protein was lower than that of Rna14 (which was the reason why the His-tag was placed on this protein). A single Rna15 molecule can however pull down two Rna14 molecules as the

latter is a dimer. Therefore, Rna15 may be present in a substoichiometric amount relative to Rna14 in the purified protein (we see some evidence of this in the SDS gels), and crystal form 3 may have selected only those with 2:1 stoichiometry of Rna14:Rna15. It has recently been reported that reconstituted *S. cerevisiae* CF IA has the stoichiometry 2:2:1:1 for Rna14:Rna15:Clp1:Pcf11 (Gordon et al. 2011).

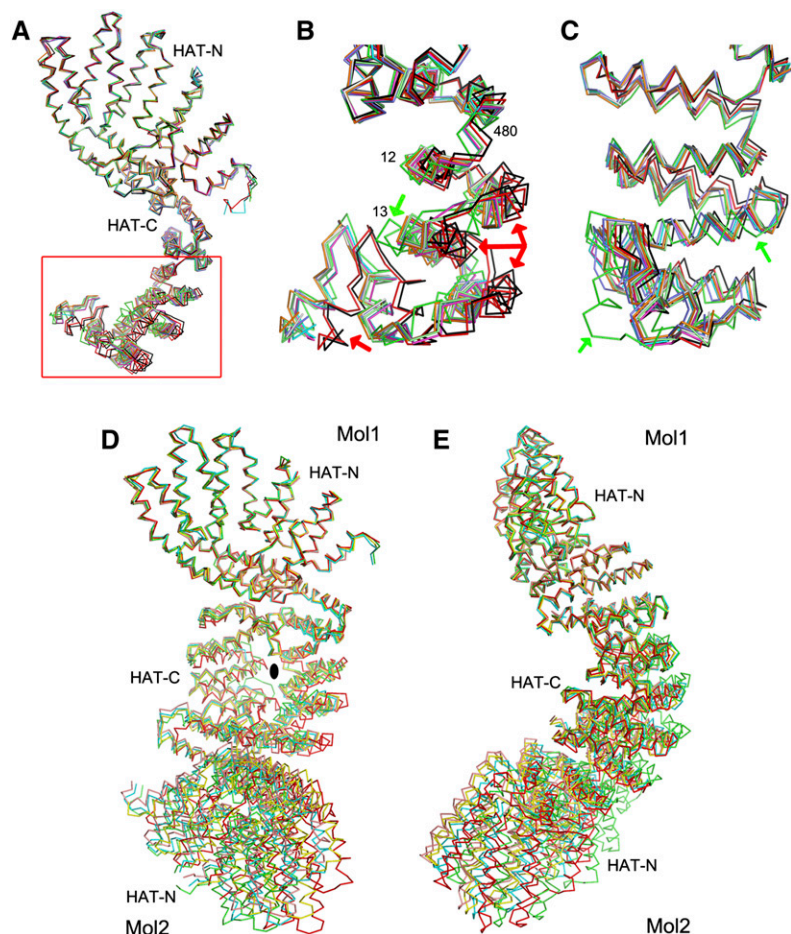
Under the same conditions that produced crystal form 3, we observed another crystal form (crystal form 4). These crystals have roughly the same morphology as crystal form 3 and are in space group *R*3, with unit cell parameters of  $a=b=163.0$  Å and  $c=181.6$  Å, nearly the same as those for crystal form 3. These crystals are therefore related to crystal form 3, with the interchange of the crystallographic and noncrystallographic threefold symmetry axis, and have only one Rna14 dimer in the asymmetric unit. We were able to collect an X-ray diffraction data set at 3.4 Å resolution on these crystals. Electron density was observed for the Rna14–Rna15 complex from the crystallographic analysis, at the same position at that in crystal form 3. However, the electron density for the complex was weak, and this crystal form will not be discussed further here.

### Conformational variability of Rna14 HAT domain dimer

Crystal form 1 has a monomer of the HAT domain in the asymmetric unit, while crystal forms 2 and 3 have one and three HAT domain dimers, respectively. This gives nine independent observations of the HAT domain monomer and five observations of the HAT domain dimer, in three different crystal packing environments, allowing us to

analyze the conformational variability of the monomers and dimers of the *K. lactis* HAT domain.

The overall structures of the nine monomers are similar to each other, with rms distances of 0.2–0.4 Å between pairs of monomers in the same crystal and 0.5–0.9 Å between pairs of monomers in different crystals (Fig. 4A). However, conformational differences are observed for the C-terminal segment (residues 480–582, HAT motifs 11 and 12 and the C-terminal extension, excluded from the rms distance cal-



**FIGURE 4.** Conformational variability of the *K. lactis* HAT domain dimer. (A) Overlay of the nine HAT domain monomers in the three different crystal forms. Residues 28–480 are superposed for this overlay. The red box highlights the structural differences in the C-terminal segment (residues 481–582) of the monomers. (B) Close-up view of the structural differences in the C-terminal region. Red arrows point to large differences for two of the monomers (red and black) of a HAT domain dimer in crystal form 3. Green arrows point to large differences for the monomer in crystal form 1 (green). (C) Panel B viewed after a 90° rotation around the vertical axis. (D) Overlay of the five HAT domain dimers in the three different crystal forms. Residues 28–480 in the first monomer are superposed for this overlay. Large differences in the positions of the second monomer, especially the HAT-N domain, are clearly visible. (E) Panel D viewed after a 90° rotation around the vertical axis.

culations above) (Fig. 4A), especially for the monomer in crystal form 1 and for two monomers of a dimer in crystal form 3 (Fig. 4B,C). On the other hand, the position of the HAT-N domain relative to the HAT-C domain appears to be conserved among the nine monomers (Fig. 4A). The weak NCS restraints that were used during structure refinement are unlikely to have greatly affected the structures of the monomers, as clear differences for the C-terminal segment (residues 480–582) are observed among the nine monomers (Fig. 4B,C).

Residues that show structural differences among the monomers are located in the center of the dimer interface (Fig. 2A). Therefore, the observed differences in the monomers are coupled to significant reorganizations of the HAT

domain dimers (Fig. 4D,E). With one monomer of the dimer in overlay, large variations in the position of the HAT-N domain of the other monomer are observed, and changes in the position of the HAT-C domain of the other monomer are also apparent (Fig. 4D). Monomers that show larger differences in the C-terminal segment produce dimers that show larger differences as well. Overall, structural comparisons among these monomers and dimers indicate that there is significant conformational variability in the HAT domain, especially its dimer. At the same time, these different dimers are stabilized at least in part by crystal packing forces, suggesting that there may not be a large difference in the stability of these dimers.

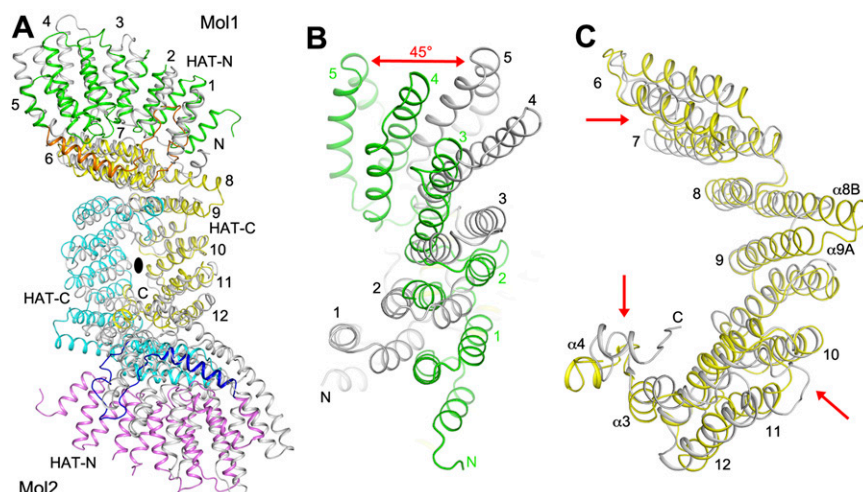
#### Large structural differences to the HAT domain of CstF-77

While the overall architecture of the monomer and dimer of the Rna14 HAT domain is similar to that of CstF-77 (Bai et al. 2007a), there are large differences between these two structures (Fig. 5A). First of all, the position of the HAT-N domain relative to the HAT-C domain in the Rna14 monomer is different to that in CstF-77. With the HAT-C domains of the two structures in overlay, the orientations of the HAT-N domains differ by a rotation of 45° (Fig. 5B) around an axis that is mostly perpendicular to the twofold axis of the dimer. As a consequence, the position of the HAT-N domain is significantly different in the two structures.

Within the HAT-C domain, only HAT motifs 8–10 show good overlay between the two structures, although helices  $\alpha$ 8B and  $\alpha$ 9A in Rna14 are much longer than their equivalents in CstF-77 (Fig. 5C). The other HAT motifs (Minvielle-Sebastia et al. 1994; Noble et al. 2004; Qu et al. 2007; Moreno-Morcillo et al. 2011) and, especially, the C-terminal extension show large differences to CstF-77. In addition, helix  $\alpha$ 4 in this extension is formed by different residues in the primary sequence of the two proteins, and the axes of this helix differ by  $\sim 90^\circ$  in the two structures (Fig. 5C). As a consequence, residues involved in the dimer interface are different in this region of the two proteins (Fig. 1B).

There is also a significant difference in the organization of the HAT domain dimer between Rna14 and CstF-77





**FIGURE 5.** Comparison with the structure of the HAT domain of CstF-77. (A) Schematic drawing of the overlay of the HAT domain dimer structure of *K. lactis* Rna14 in crystal form 1 (in color) with that of CstF-77 (in gray). The HAT-C domains of monomer 1 are superimposed for this overlay. (B) Large differences in the orientation and position of the HAT-N domains between the structures of Rna14 (green) and CstF-77 (gray), after the HAT-C domains are superimposed. The difference in orientation between the two structures is indicated with the red arrow. (C) Large structural differences between the HAT-C domains of Rna14 (yellow) and CstF-77 (gray), indicated with the red arrows.

(Fig. 5A). With the HAT-C domains of one monomer of the two structures in overlay, the orientations of the HAT-C domains of the other monomer differ by a 21° rotation, while those of the HAT-N domains differ by a 51° rotation.

The rms distance among equivalent C $\alpha$  atoms in the HAT-C domain of the two structures is 1.7 Å, while that for the HAT-N domain is 1.5 Å. The linker between the HAT-N and HAT-C domains in Rna14 is much longer than that in CstF-77. These large structural differences between Rna14 and CstF-77 also explain why attempts at solving the structure of Rna14 by the molecular replacement method, even with only the HAT-N or HAT-C domain, were not successful.

In summary, the crystal structures confirm that the HAT domain of Rna14 is also a stable dimer with a highly elongated shape, indicating that CF IA, like CstF, may also be dimeric in the yeast 3'-end processing machinery. A recent report suggests that Rna14 and Rna15 are dimeric in CF IA, while Clp1 and Pcf11 are monomeric (Gordon et al. 2011). The C-terminal region of Rna14, in complex with the hinge region of Rna15, appears to function independently of the HAT domain. Observations from the three different crystal forms of Rna14 and the comparison to CstF-77 suggest that the HAT domain may have some inherent degrees of variability. It remains to be seen whether this variability, especially in the HAT domain dimer, has any effect on the organization of CF IA and whether it could have a functional role in pre-mRNA 3'-end processing.

## MATERIALS AND METHODS

### Protein expression and purification

For expression of *K. lactis* Rna14 alone, residues 18–661 were subcloned into the pET28a vector (Novagen) and overexpressed in *Escherichia coli* Rosetta2 cells at 16°C by the addition of 0.5 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) at a OD<sub>600</sub> of 0.7. The expression construct introduced a hexa-histidine 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), 200 mM NaCl, and 150 mM imidazole and was further purified by gel filtration chromatography in a running buffer of 20 mM Tris (pH 8.5), 200 mM NaCl, and 10 mM DTT. The protein was concentrated to 16 mg/mL and stored at –80°C in the presence of 5% (v/v) glycerol.

For coexpression of *K. lactis* Rna14 and Rna15, residues 105–258 (hinge region and the CTD) of Rna15 were cloned into the pET28a vector, while residues 18–661 of Rna14 were cloned into the pCDFDuet

vector (without His tag). The plasmids were cotransformed into *E. coli* Rosetta2 cells, and protein expression and purification followed the same protocols as those for Rna14 alone.

The selenomethionyl protein was produced in B834(DE3) cells (Novagen) that have been cotransformed with a plasmid for rare codon tRNAs, 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 were obtained at 20°C by the sitting-drop vapor diffusion method, using protein samples of Rna14 alone or the Rna14–Rna15 complex. The reservoir solution contained 50 mM Hepes (pH 7.0), 12% (w/v) PEG3350, and 1% (w/v) tryptone. The crystals were cryo-protected with the reservoir solution supplemented to 25% (w/v) PEG3350 and 15% (v/v) glycerol and flash-frozen in liquid nitrogen for data collection at 100 K.

Crystal form 1 was observed from the initial screening. During the optimization of these crystals, two additional crystal forms were obtained. After the structures of crystal forms 1 and 2 were solved, it was realized that the C-terminal segment of Rna14 was removed by proteolysis (see below), possibly because of a contaminated tryptone solution. By use of a fresh tryptone solution, crystal form 3 could be obtained reproducibly, and the inclusion of 2% benzamidine and 1% dioxane as additives was essential in making these crystals suitable for data collection.

### Data collection and processing

X-ray diffraction data were collected on an ADSC charge-coupled device at the X29A beamline of National Synchrotron Light Source (NSLS). The diffraction images were processed and scaled

with the HKL package (Otwinowski and Minor 1997). A selenomethionyl multi-wavelength anomalous diffraction (MAD) data set was collected to 2.8 Å resolution, and a native data set was collected to 2.3 Å resolution for crystal form 1. The crystals belong to space group C2, with cell dimensions of  $a=117.4$  Å,  $b=59.8$  Å,  $c=123.5$  Å, and  $\beta=101.6^\circ$  for the native crystal. There is one molecule of Rna14 in the crystallographic asymmetric unit. The data processing statistics are summarized in Table 1.

A native data set to 3.0 Å resolution was collected on crystal form 2. The crystal belongs to space group  $P2_12_12_1$ , with cell dimensions of  $a=114.9$  Å,  $b=241.7$  Å and  $c=49.0$  Å. There are two molecules of Rna14 in the crystallographic asymmetric unit.

For crystal form 3, a large number of crystals were screened, and the best data set extended to 3.3 Å resolution. The crystal belongs to space group  $P3_2$ , with cell dimensions of  $a=b=162.0$  Å and  $c=177.5$  Å. There are three dimers of the Rna14–Rna15 complex in the crystallographic asymmetric unit.

### Structure determination and refinement

The Se sites were located from the data set at the high-energy remote wavelength (0.9715 Å) with the program BNP (Weeks et al. 2003). The data set collected at the peak wavelength, on a different part of the crystal, was of lower quality due to the presence of some smeared reflections. Reflection phases were calculated at 2.8 Å resolution with the data set at the remote wavelength, using the program SOLVE (Terwilliger 2003), and transferred to a native data set at 2.3 Å resolution. Subsequent phase improvement with the program RESOLVE extended to phase information to 2.3 Å resolution, allowing easier tracing and automatic model building. The complete atomic model was fit into the electron density with the programs O (Jones et al. 1991) and Coot (Emsley and Cowtan 2004). The structure refinement was carried out with the programs CNS (Brunger et al. 1998) and Refmac (Murshudov et al. 1997). The statistics on the structure refinement are summarized in Table 1.

For the second crystal form, the structure was solved by the molecular replacement method with the program COMO (Jogl et al. 2001), using the refined structure of the first crystal form as the model. Weak noncrystallographic symmetry (NCS) restraints were used during the structure refinement.

For the Rna14–Rna15 complex in crystal form 3, the structure was solved by the molecular replacement method with the program COMO (Jogl et al. 2001), using the refined structure of the HAT domain as the model. After one round of refinement, structural differences in the C-terminal segment of the HAT domain of one dimer were observed, and the atomic model was manually rebuilt. Electron density was also observed for the Rna14–Rna15 complex, and the solution structure of the *Saccharomyces cerevisiae* complex (Moreno-Morcillo et al. 2011) was manually docked into the density. For the NCS restraints during refinement, the HAT domain dimer with structural differences in the C-terminal segment was restrained in a separate group from the other two dimers. The Rna14–Rna15 complex part was restrained as a third group. The strength of the restraints was weak, but they led to a reduction in the free  $R$  factor, confirming their validity in the refinement.

### ACKNOWLEDGMENTS

We thank Stuart Myers and Neil Whalen for setting up the X29A beamline at the NSLS. This research is supported in part by grants

from the NIH (GM077175) to L.T. A.R.P. was also supported by an NIH training program in Cellular and Molecular Foundations of Biomedical Science (GM008798).

Received January 19, 2012; accepted February 26, 2012.

### REFERENCES

- Bai Y, Auperin TC, Chou C-Y, Chang G-G, Manley JL, Tong L. 2007a. Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. *Mol Cell* **25**: 863–875.
- Bai Y, Auperin TC, Tong L. 2007b. The use of in situ proteolysis in the crystallization of murine CstF-77. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**: 135–138.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, et al. 1998. Crystallography and NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**: 905–921.
- Colgan DF, Manley JL. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11**: 2755–2766.
- Emsley P, Cowtan KD. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126–2132.
- Gordon JMB, Shikov S, Kuehner JN, Liriano M, Lee E, Stafford W, Poulsen MB, Harrison C, Moore C, Bohm A. 2011. Reconstitution of CF IA from overexpressed subunits reveals stoichiometry and provides insights into molecular topology. *Biochemistry* **50**: 10203–10214.
- Hendrickson WA. 1991. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**: 51–58.
- Hendrickson WA, Horton JR, LeMaster DM. 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.
- Jogl G, Tao X, Xu Y, Tong L. 2001. COMO: A program for combined molecular replacement. *Acta Crystallogr D Biol Crystallogr* **57**: 1127–1134.
- Jones TA, Zou JY, Cowan SW, 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.
- Legrand P, Pinaud N, Minvielle-Sebastia L, Fribourg S. 2007. The structure of CstF-77 homodimer provides insights into CstF assembly. *Nucleic Acids Res* **35**: 4515–4522.
- Mandel CR, Gebauer D, Zhang H, Tong L. 2006a. A serendipitous discovery that *in situ* proteolysis is required for the crystallization of yeast CPSF-100 (Ydh1p). *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**: 1041–1045.
- Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. 2006b. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**: 953–956.
- Mandel CR, Bai Y, Tong L. 2008. Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* **65**: 1099–1122.
- Medler S, Al Husini N, Raghunayakula S, Mukundan B, Aldea A, Ansari A. 2011. Evidence for a complex of transcription factor IIB (TFIIB) with poly(A) polymerase and cleavage factor I subunits required for gene looping. *J Biol Chem* **286**: 33709–33718.
- Minvielle-Sebastia L, Preker PJ, Keller W. 1994. RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. *Science* **266**: 1702–1705.
- Moreno-Morcillo M, Minvielle-Sebastia L, Fribourg S, Mackereth CD. 2011. Locked tether formation by cooperative folding of Rna14p monkeytail and Rna15p hinge domains in the yeast CFIA complex. *Structure* **19**: 534–545.
- Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240–255.



- Noble CG, Walker PA, Calder LJ, Taylor IA. 2004. Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA. *Nucleic Acids Res* **32**: 3364–3375.
- Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307–326.
- Pancevac C, Goldstone DC, Ramos A, Taylor IA. 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.
- Preker PJ, Keller W. 1998. The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem Sci* **23**: 15–16.
- Qu X, Perez-Canadillas JM, Agrawal S, de Baecke J, Cheng H, Varani G, Moore CL. 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* **283**: 2101–2115.
- Terwilliger TC. 2003. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* **374**: 22–37.
- Weeks CM, Adams PD, Berendzen J, Brunger AT, Dodson EJ, Grosse-Kunstleve RW, Schneider TR, Sheldrick GM, Terwilliger TC, Turkenburg MG, et al. 2003. Automatic solution of heavy-atom substructures. *Methods Enzymol* **374**: 37–83.
- Xiang S, Usunow G, Lange G, Busch M, Tong L. 2007. Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. *J Biol Chem* **282**: 2676–2682.
- Zhao J, Hyman L, Moore CL. 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.