nature structural & molecular biology

Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex

Bo Liang[1,5], Song Xue[1,2,5], Rebecca M Terns[3], Michael P Terns[3,4] & Hong Li[1,2]

The most complex RNA pseudouridylases are H/ACA ribonucleoprotein particles, which use a guide RNA for substrate capture and four proteins (Cbf5, Nop10, Gar1 and L7Ae/NHP2) for substrate modification. Here we report the three-dimensional structure of a catalytically deficient archaeal enzyme complex (including the guide RNA and three of the four essential proteins) bound to a substrate RNA. Extensive interactions of Cbf5 with one guide-substrate helix and a guide RNA stem shape the forked guide–substrate RNA complex structure and position the substrate in proximity of the Cbf5 catalytic center. Our structural and complementary fluorescence analyses also indicate that precise placement of the target uridine at the active site requires a conformation of the guide–substrate RNA duplex that is brought about by the previously identified concurrent interaction of the guide RNA with L7Ae and a composite Cbf5-Nop10 surface, and further identify a residue that is critical in this process.

Numerous noncoding RNAs that participate in nearly all aspects of gene expression and regulation are being discovered at a rapid rate[1,2]. Owing to their base-pairing potential, many noncoding RNAs act as guides in biochemical processes catalyzed by their partner proteins[3]. Ribonucleoprotein complexes guided by noncoding RNAs constitute an unconventional but widespread class of enzymes. One important class of these RNA-guided enzymes is the H/ACA small nucleolar (sno) RNPs that function in ribosome biogenesis in the nucleolus, isomerizing nearly 100 selected uridines to pseudouridines in human ribosomal RNA (rRNA)[4–6]. Many of these modified sites are within functionally critical regions of the ribosome[7], and site-specific disruption of modifications impairs ribosome biogenesis or function[8].

Since the role of H/ACA RNAs in rRNA pseudouridylation was first identified[9,10], their known functional repertoire has expanded to include modification of small nuclear RNAs[11] as well as rRNA processing[12] and telomerase maturation[13]. H/ACA snoRNAs that guide rRNA modification are also found in the archaeal kingdom[14], revealing the ancient evolutionary origin of H/ACA RNAs. Pseudouridylation contributes to RNA stability[15–18], to increases in the number of available RNA functional groups[19] and ultimately to the cellular function of the substrate RNA[7,8,20–24].

The H/ACA RNAs responsible for isomerization of uridine function with partner proteins to form complex, multicomponent pseudouridylases (H/ACA RNPs)[25,26]. The H/ACA RNPs contain four core proteins that are known to participate directly in pseudouridylation[27–29]. These are Cbf5 (dyskerin in human, NAF57 in rodents), Gar1, Nop10 and L7Ae (NHP2 in eukaryotes). Cbf5 shares close sequence and structural homology with the bacterial pseudouridylase TruB and is the catalytic subunit for H/ACA RNP-mediated pseudouridylation[30]. Mutations in dyskerin cause the X-linked form of the rare genetic disease dyskeratosis congenita, which is characterized by abnormal skin pigmentation, bone marrow failure and an increased predisposition to cancer[31,32]. Archaeal H/ACA RNPs have been reconstituted from purified components and shown to possess pseudouridylation activity *in vitro*[28,29]. These studies showed that the ability of Cbf5 to catalyze efficient pseudouridylation is dependent upon each of the other three core H/ACA RNP components[28,29]. The specific roles of Gar1, Nop10 and L7Ae in pseudouridylation are not well understood.

In contrast to stand-alone pseudouridylases, which modify one or several specific uridine nucleotides[33], H/ACA RNP pseudouridylases are capable of modifying as many different sites as are specified by the H/ACA guide RNAs expressed in the organism. The pseudouridylation guide H/ACA RNAs range in size from 70 to 250 nucleotides and are comprised of one to three helix-internal loop-helix-tail units[2]. Each unit is sufficient for binding all four protein factors and for guiding isomerization of a target uridine *in vitro*[28,29,34]. The strictly conserved 3′ tail (containing the ACA trinucleotide) and the asymmetric internal loop (pseudouridine pocket) are required for efficient binding of Cbf5 (refs. 28,29,34,35). The mechanistic details of how the target uridine of the substrate RNA is faithfully presented to the active site of Cbf5 are largely unknown. It is clear that the extensive base-pairing that occurs between the single-stranded substrate RNA and complementary nucleotides of the pseudouridine pocket of the H/ACA guide RNA determines the specificity for substrate RNAs[9,10,35].

Recent structural studies of archaeal H/ACA RNPs have provided significant insight into the detailed architecture of the complex. Structures of isolated Cbf5–Nop10 and Cbf5–Nop10–Gar1 complexes have been determined[36–38], and a structure of the complex containing

[1]Institute of Molecular Biophysics and [2]Department of Chemistry and Biochemistry, 91 Chiefton Way, Florida State University, Tallahassee, Florida 32306, USA. [3]Departments of Biochemistry and Molecular Biology and [4]Department of Genetics, University of Georgia, Green Street, Athens, Georgia 30602, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to H.L. (hongli@sb.fsu.edu).

Received 14 June; accepted 15 October; published online 2 December 2007; doi:10.1038/nsmb1336

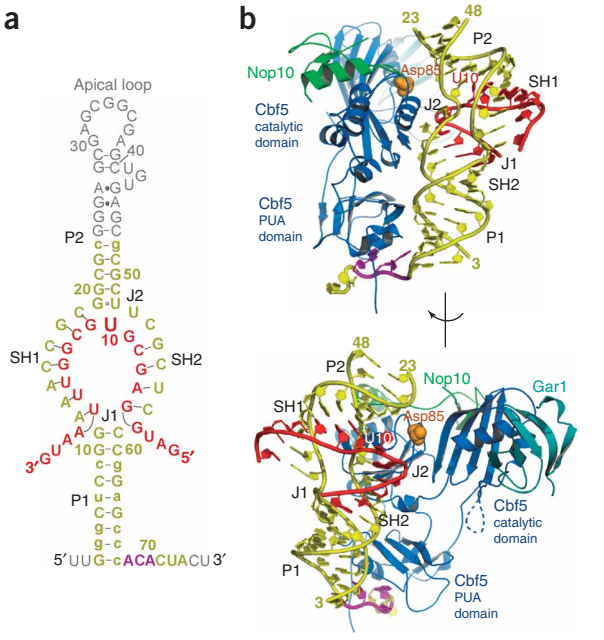


Figure 1 Overview of the substrate-loaded H/ACA RNP subcomplex structure. (**a**) Schematic of the composite Pf9 guide RNA (yellow) and the target RNA (red) used for crystallization. Gray nucleotides represent Pf9 sequences not included in the molecules for cocrystallization. Lower-case letters indicate nucleotides that were modified from the native sequence. (**b**) Two views of the overall structure of the protein–RNA complex. The same colors as above are used here for the guide and target RNA. Ribbon models of the three proteins are shown in blue (Cbf5), green (Nop10) and aquamarine (Gar1). The active-site residue Asp85 is shown in orange spheres.

all four proteins and a guide RNA is also available[39]. In addition, two structures of guide and target RNA complexes in solution were recently obtained[40,41]. These structures have all provided glimpses of the RNP enzyme. However, the basis for the accurate placement of target uridines in the active site of the enzyme, and for the requirement of all accessory proteins for function, is still not understood.

Toward the ultimate goal of understanding the molecular mechanism of H/ACA RNP function and the roles of the guide RNA and each protein in this process, we have determined the crystal structure of the *Pyrococcus furiosus* (*Pf*) H/ACA RNP bound to a wild-type substrate RNA for the Pf9 H/ACA guide RNA[28] in the absence of L7Ae at 2.87 Å. To inhibit any potential chemical reaction during crystallization, we used a form of the Cbf5 protein containing an active-site mutation (D85A). The RNP structure revealed interactions between the guide–target RNA complex and conserved residues of the Cbf5 catalytic domain, and showed that guide–target RNA base-pairing in the context of Cbf5–Nop10–Gar1 complex places the target uridine in the vicinity of the active site of Cbf5. Structural comparison with the previously determined RNP structure in the absence of the substrate RNA but in the presence of L7Ae[39] suggests a functional role of L7Ae in remodeling the guide RNA to further deliver the target uridine into the active site of Cbf5.

RESULTS

Overall structure of the RNP

The structure of the *P. furiosus* H/ACA RNP bound to a substrate RNA in the absence of L7Ae was determined by molecular replacement methods using the previously determined crystal structure of the *P. furiosus* Cbf5–Nop10–Gar1 complex (PDB accession code 2EY4) as the search model. Details of crystallization and the structural determination processes are described in Methods. The final structure was refined to an R_{free} of 30.1%, including all reflections, and to a satisfactory stereochemical quality. The refined structural model contains residues 11–340 for Cbf5 (full-length 1–343), 3–55 for Nop10 (full-length 1–60), 1–73 for Gar1 (full-length 1–97), nucleotides 4–25 and 48–72 of the guide RNA (this represents all but the last three nucleotides of the bimolecular model guide RNA) and nucleotides 5–18 of the 21-mer unmodified target RNA (**Fig. 1a**). The predicted catalytic residue Asp85 in Cbf5 was mutated to alanine to inhibit potential pseudouridylation reactions during crystallization. For clarity, in what follows all guide RNA nucleotides are prefixed with 'g' and all target nucleotides are prefixed with 't'.

In this structure, the guide RNA is stably bound with the target RNA in the absence of the apical loop and kink-turn (**Fig. 1a**). The extended guide–target RNA helical complex adheres to an inclined plane primarily formed by the Cbf5 protein (**Fig. 1b**), which positions the target uridine close to the active site of Cbf5. The Cbf5-induced incline seems to be essential for positioning the substrate nucleotide, arguing for the importance of proteins in this process. Moreover, in this RNP structure, which lacks L7Ae, the target uridine is located ~11 Å away from the active site of Cbf5 and is thus unavailable for modification. Previous *in vitro* reconstitution experiments had demonstrated a critical but unidentified role for L7Ae in pseudouridylation by the H/ACA RNP[28,29]. This RNP structure corroborates the biochemical data and furthermore indicates that L7Ae is also essential for target uridine placement within the active site of Cbf5.

Structural features of the guide–target RNA complex

As is also observed in the solution structures of the RNA–RNA complexes[40,41], the target RNA is bound to the guide RNA through a constrained three-way junction (**Fig. 1a,b**). The upper stem of the guide RNA (P2) and the two helices formed between the 5′ and 3′ halves of the pseudouridine pocket and the substrate RNA (SH1 and SH2 for substrate helices 1 and 2) constitute the three branches of the junction, whereas the lower stem of the guide RNA (P1) restricts the direction of the branches (**Fig. 1a,b**). The pseudouridine pocket of the guide RNA is an elongated opening with a distance from the apex to the base of ~40 Å (**Fig. 2**). The target RNA is bent into a V-shaped loop that fits snugly into the opening by base pairing to both sides of the pseudouridine pocket. The distance between the phosphate

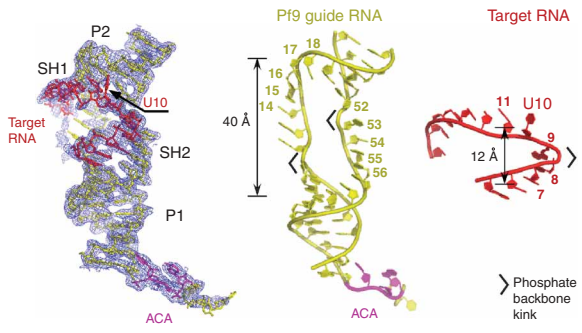


Figure 2 Structural features of the guide and target RNAs in the complex. Left, final electron density map computed as a composite omit (omitting 5% of the model each time), cross-validated and SigmaA-weighted $3F_o - 2F_c$. Right, guide (yellow) and target (red) RNA, respectively. The locations of the three phosphate backbone kinks are indicated with black angle brackets.

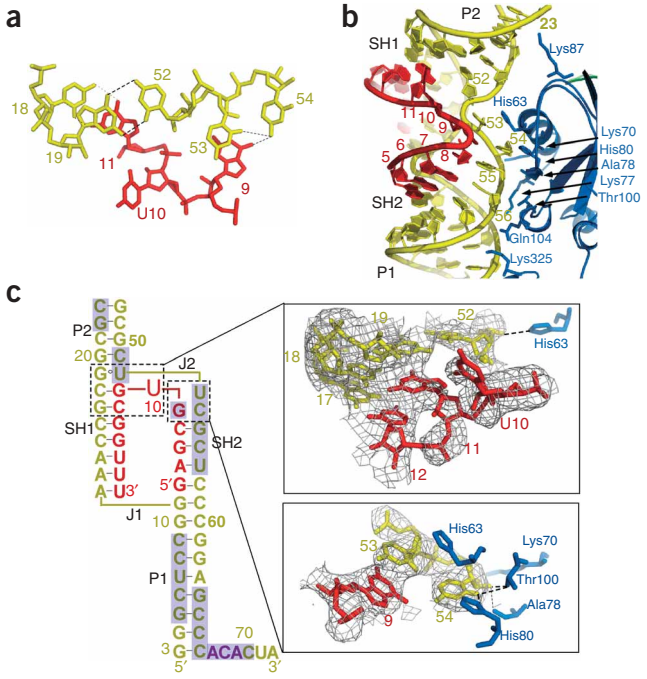


Figure 3 RNA-RNA and RNA-protein interactions in the H/ACA RNP structure. (**a**) Detailed view of the J2 junction (see also the notations for **c**). (**b**) The RNA-protein interface involving the pseudouridine pocket. Cbf5 residues within 3.4 Å of any RNA atoms are shown in stick models and are labeled.(**c**) Detailed RNA-protein contacts involving the J2 junction nucleotides. In the schematic of the guide–target RNA complex, nucleotides covered by gray boxes are protected by proteins. Note that the primary protein contacts occur in the SH2 stem and that His63 forms two interactions, stacking with gU53 and hydrogen bonding with the 2′-hydroxyl group of gU52.

backbones of the two sides of the V-loop ranges from ~6 Å to ~14 Å (**Fig. 2**). SH1 and SH2 are each coaxial with the upper and lower stem of the guide RNA, P2 and P1, respectively (**Fig. 1a**,**b**), similar to what is seen in the guide–target RNA complexes formed in the absence of proteins and determined by NMR[40,41]. However, the two pseudo-continuous helices (P1-SH2 and P2-SH1) are inclined at a ~120° angle at junction J2, in contrast to the nearly parallel helices in the guide–target RNA-only complexes[40,41]. This inclination, which appears to be a consequence of interaction with Cbf5, is necessary for the substrate RNA to enter the catalytic pocket, suggesting that Cbf5 has a functional role in the placement of the substrate RNA.

The guide–target RNA complex has an unprecedented RNA architecture wherein a single RNA strand docks to an asymmetric internal loop from one side. Unique to this architecture are the structural features observed at the helical junctions between P1 and SH1 (J1) and between SH2 and P2 (J2). J2, which has more complex features than J1, includes the target uridine and makes extensive contacts with Cbf5 residues (**Fig. 3a**). Both junctions involve rotation of downstream nucleotides away from the upstream helical procession. In J1, gA12 rotates away from SH2 helical stack to form an AU base pair with tU17 in the substrate RNA. This creates a sharp bend at the phosphate backbone between gA12 and gG11 (**Fig. 2**). Nucleotide rotations in J2 occur over nucleotides 52–54 in the guide as well as nucleotides 9–10 in the target strand. As a result of its extensive interactions with Cbf5 residues, the nucleobase of gC54 rotates toward the protein and away from tG9 (**Fig. 3b**,**c**). This eases the constraint on the phosphate backbone of tG9, allowing reversal of the target chain direction at this nucleotide; the phosphate backbone flips more than 90° away from the A-form position (**Fig. 3b**,**c**). The rotation at tG9 further disrupts the Watson-Crick edge-to-edge interaction between tG9 and gC54, leaving only one potential hydrogen bond between O6 of tG9 and N4 of gC54 (**Fig. 3b**,**c**). gU53 at the center of the J2 junction, which does not pair with a nucleotide, stabilizes the unpaired tG9 by stacking on its nucleobase (**Fig. 3c**). The sugar-phosphate of the target uridine, tU10, further facilitates the chain reversal, but its nucleobase is disordered in our structure. Finally, gU52 rotates nearly 90° from gU53 to pair with gG19 and initiate the P2 helix, completing the transition from SH2 to P2. The structure of J2 observed here is more open than that reported for the 3′ guide-target structure of the human U65 RNA[40] in solution but appears to be less open than that of the 5′ guide-target structure of the same RNA in solution[41]. Whether these differences are due to the different RNA secondary structures or to the roles of the proteins remains to be explored. However, given that each of these RNAs guides the modification of its respective substrate by the same enzyme, it is likely that a similar structure is formed in the active site of the enzyme. Therefore, H/ACA RNPs could assemble a diverse range of guide–target RNA complexes to that required for catalysis. As discussed in greater detail below, phosphate backbone bending at the two junctions may also allow parts of the phosphate backbone to act as hinges for a structural transition required to dock the target RNA into the active site of the enzyme.

Interactions between the guide–target RNA complex and Cbf5

In the previously determined structure of the archaeal H/ACA RNP lacking target RNA, the pseudouridine pocket of the guide RNA is either disordered or stabilized by crystal packing interactions[39]. The primary contacts between the guide RNA and Cbf5 occur mostly in the lower stem of the RNA. Almost no interactions are observed between the pseudouridine pocket and the catalytic subunit of the pseudouridylase[39]. Our structure indicates that, upon association of the target RNA, the pseudouridine pocket of the guide RNA becomes ordered and forms specific contacts with conserved residues of Cbf5.

The most extensive contacts between the guide–target RNA complex and Cbf5 occur in the SH2 stem. As indicated by changes in the solvent-accessible surface areas attributable to the bound RNA, the Cbf5 residues that contact the SH2 stem are His63, Val66, Ala67, Ala68, Lys70, Gly79, His80, Thr100, Arg101, Val103, Gln104 and Lys325. Sequence alignments show that His63 and His80 are strictly conserved; homologous amino acids are found in the other positions in phylogenetically diverse organisms (**Supplementary Fig. 1** online). Despite the extensive RNA-protein interface, few interactions involve nucleobases (**Fig. 3a**,**b**). This mode of interaction is consistent with the requirement that the pseudouridine pocket of the H/ACA RNA be available to interact specifically with the target RNA[35,42,43]. In contrast to the close interactions between SH2 and Cbf5, the SH1 stem formed between the 3′ half of the target RNA and the 5′ half of the pseudouridine pocket does not contact the proteins in our structure (**Fig. 3b**,**c**), which may contribute to the relatively high degree of disorder in this particular region of the RNA structure. In summary, the conservation of Cbf5 residues involved in stabilizing the pseudouridine pocket predicts the general importance of these residues in substrate binding and suggests that these interactions will be maintained in the fully assembled holoenzyme.

A mutation (S121G) in dyskerin (the human homolog of Cbf5) that affects an amino acid adjacent to the strictly conserved histidine that corresponds to *Pf* Cbf5 His80 (**Supplementary Fig. 1**) is associated with dyskeratosis congenita[31]. This is consistent with the previous

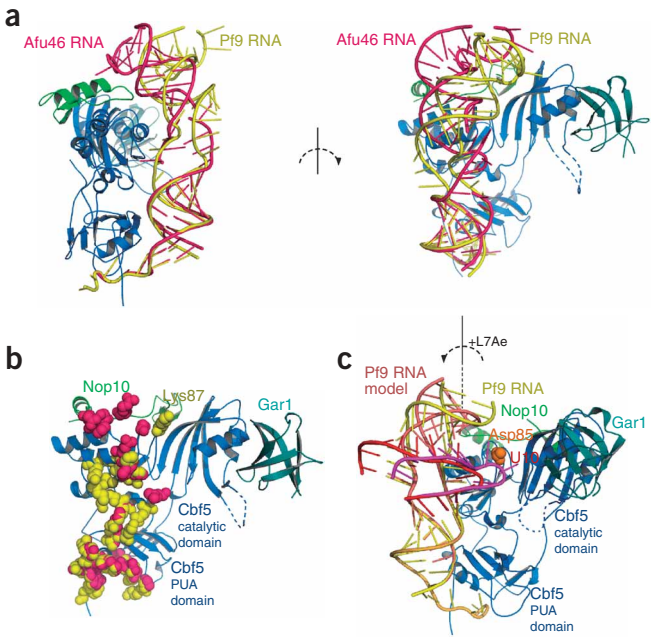


Figure 4 Guide RNA conformation and RNA-protein contacts in the presence and absence of L7Ae. (**a**) Comparison of the guide RNA structure in the presence of L7Ae and absence of target RNA (magenta, from ref. 39) and in the absence of L7Ae and presence of target RNA (yellow, from this study). Two orthogonal views are presented. (**b**) Guide RNA-protein contacts observed in the presence (magenta, from ref. 39) and absence (yellow, from this study) of L7Ae. (**c**) Modeled configurations of the guide and target RNAs in the presence of L7Ae. The upper stem of the guide RNA is shifted to the position observed in the presence of L7Ae (pink, Pf9 RNA model), which threads helix P2S forward and places the target uridine (U10, magenta) in the active site (Asp85, orange spheres). The target RNA position observed in our structure is shown in red.

suggestion that many dyskeratosis congenita–associated mutations affect amino acids in regions important for RNA binding[37,38]. Protein sequence alignment shows that the mutations underlying dyskeratosis congenita do not generally affect dyskerin amino acids that are predicted to be directly involved in RNA binding[39] (which might cause a more severe phenotype), but rather neighboring amino acids (**Supplementary Fig. 1**). Notably, a mutation from an individual with dyskeratosis congenia was also recently mapped to human Nop10 (ref. 44) in a region expected to be involved in nonspecific binding the upper stem of the guide RNA in the holoenzyme[39].

The conserved β7–β10 loop of Cbf5, which is predicted to interact with the guide–target RNA complex[37–39], is disordered and not engaged in RNA-protein interactions in our structure (**Fig. 4**, dashed line). This loop may close the active site upon correct placement of the target RNA and the triggering mechanism may rely upon residues within the active site of the enzyme. The disengagement of the analogous 'thumb loop' of TruB upon mutation of the catalytic aspartate to alanine[45] corroborates this possibility. Similarly, Gar1 may play a role in target RNA binding[38,39], though it was not observed to interact with target RNA in our structure. Gar1 may also move closer to the target RNA when the target uridine is engaged in the active site.

Role of L7Ae in target RNA placement in the active site

Comparison of our *Pf* RNP structure with that of the *Pf* RNP containing L7Ae (and lacking target RNA) reveals substantial differences in the upper stem (P2) of the bound guide RNA. Structural and biochemical evidence described below suggests that these differences are attributable to L7Ae. We superimposed the Cbf5 molecules of both structures and examined the resulting positions of the other molecules. The overall structures of Cbf5, Gar1 and Nop10 are highly similar in both RNP structures (overall r.m.s. deviation 0.94 Å for 440 Cα atoms). In addition, the structures of the P1 lower stem, including the 3′ ACA trinucleotides of the guide RNAs, do not deviate substantially between the two (r.m.s. deviation 1.03 Å for 21 phosphate atoms). The largest differences between the two structures are in the pseudouridine pocket and the upper stem of the guide RNAs. Whereas the difference in the pseudouridine pockets is very likely the result of target binding as described above, that observed between the upper stems is likely the result of L7Ae binding. In the absence of L7Ae, the upper stem is less inclined forward toward Nop10 and more bent sideways in the direction of Gar1 (**Fig. 4a**). As a result, the contacts established between the upper stem and the protein complex are different (**Fig. 4b**). For instance, the upper stem extensively contacts residues 34–38 of Nop10 in the presence but not the absence of L7Ae (**Fig. 4b**). Similarly, the close contact formed between the upper stem of the guide RNA and Lys87 of Cbf5 in the RNP lacking L7Ae is not found in the RNP containing L7Ae (**Fig. 4b**). Notably, the results of protein protection assays indicate that full-length Pf9 also contacts Cbf5 Lys87 in the absence of both L7Ae and target RNA in solution[46]. These findings indicate that the difference in the position of the upper stem of the guide RNA in our structure does not reflect any alteration in the model guide RNA sequence or the presence of target RNA, but rather results from the lack of L7Ae.

L7Ae seems to bend the upper stem of the guide RNA away from the active site of Cbf5 and toward Nop10 (**Fig. 4a**). To specifically address what effect the L7Ae-induced movement of the upper stem might have on the proximity of the target uridine to the active site of Cbf5, we superimposed the upper stems (P2) of the guide RNAs from the two structures[39] and allowed SH1 (which is not subject to extensive RNA-protein interaction in our structure) to follow while fixing the positions of P1 and SH2 (which are engaged in interaction with protein). Adjustments were restricted to the nucleotides in the two junction regions, J1 and J2, and were limited by steric constraints. The modeling produces an RNA configuration in which the target uridine is redirected toward the active site of Cbf5, to the position equivalent to that occupied by the target uridine in TruB-substrate RNA structures[47,48] (**Fig. 4c**). The model requires unpairing of tG11 from gC18, and possibly pairing of gC18 with gU53 to 'close' the more open pocket observed in our structure and to more closely resemble the predicted consensus guide-target RNA structure[35,42,43].

To investigate the proposed role of L7Ae in the target RNA placement, we devised a fluorescence assay that monitors the conformation of the nucleotide immediately 3′ of the target uridine (tG11). We replaced tG11 with the fluorescent analog of guanosine and adenosine, 2-aminopurine (2-AP), and the target uridine with 5-fluorouridine (5-FU). We believe that this target RNA mimics the rRNA substrate targeted by Pf9 RNA and that the 5-FU is converted by the enzyme to (5*S*,6*R*)-5-fluoro-6-hydroxypseudouridine (5-FhΨ), as has been observed in the case of the related TruB enzyme[46]. The fluorescence intensity of 2-AP directly reflects its stacking state and can, therefore, reveal changes in the conformational state of the RNA near the target uridine during the processes of substrate docking and the chemical reaction. The 2-AP–labeled target RNA was annealed with the guide RNA at a 1:1 molar ratio and then incubated with a ten-fold molar excess of Cbf5–Nop10–Gar1 trimeric protein complex to allow formation of an RNP similar to that observed in our crystal

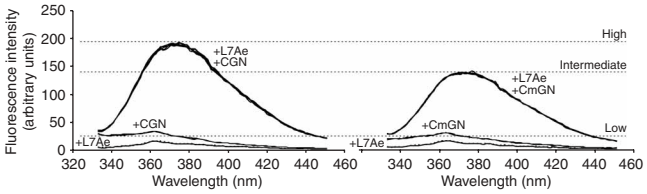


Figure 5 Effect of L7Ae on target RNA conformation. Fluorescence intensity traces of target RNA labeled with 2-AP and 2-FU and of unlabeled guide RNA assembled with a ten-fold molar excess of Cbf5–Nop10–Gar1 (+CGN), Cbf5(D85A)–Nop10–Gar1 mutant (+CmGN), L7Ae (+L7Ae), Cbf5–Nop10–Gar1 plus L7Ae (+CGN +L7Ae) or Cbf5(D85A)–Nop10–Gar1 plus L7Ae (+CmGN +L7Ae). The maximum fluorescence intensity reached as a result of L7Ae titration in the presence of Cbf5–Nop10–Gar1 trimer is designated as the 'high- fluorescence state', that reached as a result of L7Ae titration in the presence of the D85A mutant complex as the 'intermediate-fluorescence state' and that reached without L7Ae as the 'low-fluorescence state'.

structure. In the absence of the L7Ae protein, the assembled H/ACA RNPs (either wild-type Cbf5 or the D85A mutant) showed a comparably low 2-AP fluorescence intensity, which we refer to as the 'low-fluorescence state' (**Fig. 5**). L7Ae was subsequently titrated into the RNP solution, which resulted in substantial increases in the 2-AP fluorescence intensity (**Fig. 5**), indicative of remodeling of the target RNA in the region of the target uridine. A maximum fluorescence intensity change was observed with a ten-fold molar excess of L7Ae. The change in 2-AP fluorescence required the presence of the Cbf5–Nop10–Gar1 trimeric protein complex because L7Ae titration of the guide RNA and target RNA in the absence of other proteins resulted in no change in fluorescence intensities (**Fig. 5**). We interpreted the L7Ae-induced change as a result of the bound 2-AP unstacking during the placement of the target RNA into the catalytic site of Cbf5 followed by conversion of 5-FU to 5-FhΨ. Consistent with this interpretation, when L7Ae was titrated into a solution of the 2-AP RNA bound with the trimeric protein bearing the D85A mutation in Cbf5 (the complex used for crystallization), the maximum fluorescence intensity change was substantially less than that produced by the wild-type RNP (**Fig. 5**), suggesting that this mutation prevents the target RNA from completing the L7Ae-induced transformation and/or the conversion of 5-FU to 5-FhΨ. We designate the final state reached by the wild-type RNP as the 'high-fluorescence state' and that reached by the mutant RNP as the 'intermediate-fluorescence state'. Thus, both the wild-type H/ACA RNP and the mutant H/ACA RNP lacking L7Ae were effectively shifted from the low-fluorescence state to another conformational state in an active site–dependent manner by L7Ae (**Fig. 5**).

DISCUSSION

The complexity of H/ACA RNP pseudouridylase assembly has been well documented[27–29]. Despite the close structural homology between Cbf5 and the stand-alone pseudouridylase TruB, *in vitro* activity assays using purified archaeal components have shown that three additional proteins are required for efficient modification by Cbf5[27,29]. This knowledge has led to an intense and ongoing search for the molecular basis of the requirement for the accessory factors. As part of this effort, we report a crystal structure of the H/ACA RNP subcomplex containing Cbf5, Nop10, Gar1, an H/ACA RNA and a bound target RNA. Notably, we found that the guide RNA occupies a substantially different position in the absence of L7Ae than in an assembled apo H/ACA RNP structure[39]. The lack of L7Ae affects the locations of the upper stem and the pseudouridine pocket and thereby of the bound target RNA. Indeed, in this subcomplex the target uridine is found about 11 Å from the catalytic aspartate residue. On the basis of the assumption that the target RNA should be fully engaged in the active site in the holoenzyme, we propose that L7Ae has an important role in the placement of the target RNA. The concurrent interaction of L7Ae with the kink-turn motif of the guide RNA and with the composite surface formed by Nop10 and Cbf5 leads to anchoring of the upper stem of the guide RNA and consequently to functional docking of the target RNA (via the complementarity to the 5′ half of the pseudouridine pocket).

This proposal is supported by the results of our 2-AP fluorescence assay, which indicate that an L7Ae-induced conformational change occurs near the target uridine. In the absence of L7Ae, both the wild-type H/ACA RNP and that containing Cbf5 with an active-site mutation (D85A) stay in a low-fluorescence state that is likely to correspond to the crystal structure that we solved. L7Ae titration of both H/ACA RNPs resulted in marked increases in the fluorescence intensity, which is interpreted as a L7Ae-induced conformational change near the target uridine. However, we found that a complete L7Ae-induced conformational change (to the high-fluorescence state) requires the presence of the catalytic aspartate residue. One potential explanation for the sensitivity of the target RNA conformation to the catalytic aspartate is that conversion of 5-FU to 5-FhΨ causes an additional change in the conformation of the target uridine. Alternatively, a critical interaction established between the catalytic aspartate of the wild-type Cbf5 and the target uridine is required for the completion of the conformational change in the target RNA.

In this work, we have demonstrated that H/ACA RNP function requires remodeling of the guide RNA structure by L7Ae, assisted by Nop10, along with Cbf5 (and its catalytic residue), to place the target uridine at the active site for pseudouridylation.

METHODS

Design and purification of the components used in crystallization. To ensure that the D85A mutant of Cbf5 could bind Pf9 RNA, electrophoretic mobility shift assays were carried out as described[28]. These data indicated that the D85A mutant bound guide RNA with similar affinity as the wild-type protein (D. Baker, R.M.T. and M.P.T., unpublished data). We also formed crystals with the wild-type Cbf5 and the target RNA containing 5-fluorouridine and found that they were isomorphous to those containing the D85A mutant (**Supplementary Table 1** online), suggesting that the mutant-containing complex has structural features grossly similar to those of the wild-type complex. Consistent with our observations, an aspartate-to-asparagine substitution in Cbf5's homolog, TruB, did not impair binding of its substrate RNA[45]. Proteins used in crystallization were purified according to published protocols[38].

The model guide RNA used in crystallization contains a pseudouridine pocket identical to that of Pf9 but altered upper and lower stems, and it was assembled by hybridization of two RNA oligonucleotides. RNA oligos were purchased from Dharmacon and purified according to the manufacturer's recommended protocols.

Crystallization. The two guide strands and the target RNA, at a 1:1:1 molar ratio, were annealed by heating the solution for 1 min at 70 °C and then slowly cooling it. The RNA–protein complex was formed at a 1:1.2 molar ratio with a total concentration of 21.8 mg ml^{-1}. The crystals were obtained by vapor diffusion methods in hanging drops at 30 °C. The RNA-protein mixture was mixed in an equal volume before being equilibrated with a reservoir solution containing 50 mM MES, pH 6.0, 100 mM NH_4COOCH_3, 5 mM $MgSO_4$ and 1.0 M NaCl. Diamond-shaped crystals were obtained in 3–5 d. Crystals were soaked stepwise in cryosolutions containing the mother liquor plus 10% (v/v) and 15% (v/v) glycerol, respectively, before being flash cooled in a liquid nitrogen stream for data collection. Data were collected at the South Eastern Consortium Access Team (SER-CAT) beamline 22ID and were processed using HKL2000 (ref. 49). The crystals are in a primitive tetragonal space group with

Table 1 Data collection and refinement statistics

	Cbf5(D85A)–Nop5–Gar1–RNA complex
Data collection	
Space group	$P4_12_12$
Cell dimensions	
a, *b*, *c* (Å)	96.56, 96.56, 240.98
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	42.5–2.8 (2.9–2.8)
R_{sym}	11.0 (50.1)
$I/\sigma I$	41.9 (2.7)
Completeness (%)	83.1 (33.2)
Redundancy	15.1 (8.1)
Refinement	
Resolution (Å)	42.5–2.87 (2.97–2.87)
Total number of reflections	29,512 (2873)
R_{work}/R_{free}	24.2 (35.8)/30.1 (44.3)
No. atoms	
Protein	3,711
RNA	1,293
Water/ion	1 Zn
B-factors	
Protein	49.7/50.6/40.7
Guide RNA/target RNA	68.6/84.3
r.m.s. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.769

Data were collected from a single crystal. Values in parentheses are for highest-resolution shell.

cell dimensions $a = 96.56$ Å, $b = 96.56$ Å, $c = 240.98$ Å. The solvent content of the crystals was determined to be 65.4%, which was consistent with the presence of one RNP in each asymmetric unit.

Structure determination. The structure was determined by molecular replacement methods using MOLREP[50] through the CCP4i interface to the CCP4 programs[51], which combines rotation and translation searches in a single step. The previously determined structure of the Cbf5–Nop10–Gar1 complex was used as a search model. A single and outstanding solution was found in the space group $P4_12_12$. The trimeric protein complex transformed to the correct solution was subjected to successive rigid body, minimization and simulated annealing refinement. At this stage, SigmaA-weighted $3F_o - 2F_c$ and $F_o - F_c$ electron densities were computed, which revealed the bound RNA molecules in the front surface of Cbf5–Nop10–Gar1 trimer. A molecular mask generated using a preliminary RNA model and the protein complex was then used to perform solvent flattening and flipping using initial phases computed from protein coordinates only in CNS[52]. The density-modified map was of sufficient quality (**Supplementary Fig. 2** online) for tracing most RNA nucleotides. The real space correlation coefficients between the final model and the density modified map were 0.46 and 0.76 for RNA and proteins, respectively. Refinement of the RNP was carried out using CNS[52] and REFMAC5[53]. To account for rigid-body displacements of the complex with a reduced parameter set, the final stage of the refinement used translation-libration-screw-motion (TLS) parameters as implemented in REFMAC5. The three individual proteins and RNA strands were treated as single 'rigid-body' groups, and the final TLS parameters are listed in **Supplementary Table 2** online. The refined structure has 0.45-Å coordinate error based on the maximum-likelihood method. A composite omit, cross-validated, SigmaA-weighted $3F_o - 2F_c$ map was computed using the final model and displayed around the RNA model (**Fig. 2a**) with 5% of the mode omitted at each time. The real space correlation coefficients between the final model and the composite omit $3F_o - 2F_c$ map are 0.87 and 0.88 for proteins and RNA, respectively. The final protein structure was assessed with Procheck[54] and found to be consistent with stereochemically valid models (**Table 1**).

Steady-state fluorescence studies. The target RNA designed for fluorescence studies contains 2-AP at position tG11 and 5-FU at position tU10. The 2-AP– and 5-FU–labeled target RNA was purchased from Dharmacon. The Pf9 H/ACA guide RNA was obtained by T7 transcription. The annealed target–guide RNA complex at 1 μM concentration was then incubated with 10 μM of the Cbf5–Nop10–Gar1 or Cbf5(D85A)–Nop10–Gar1 protein complex in a total volume of 100 μl. L7Ae was titrated in small increments until its final concentration reached 10–15 μM, when fluorescence intensities no longer increased. Fluorescence measurements were performed in a Cary Eclipse fluorescence spectrophotometer (Varian). The sample cuvette was maintained at 50 °C using a circulating water bath. The excitation wavelength was 325 nm and the fluorescence intensity was measured at the peak of fluorescence (375 nm). The excitation and emission bandwidths were both 5 nm. For fluorescence intensity measurement after each titration, the sample was incubated for 20 min and then 8–10 accumulative scans were taken. Each titration was repeated in at least three independent experiments.

Accession codes. Protein Data Bank: Coordinates have been deposited with accession code 2RFK.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

This work was supported by US National Institutes of Health (NIH) grant R01 GM66958-01 (H.L.) and NIH grant RO1 GM54682 (M.T. and R.T.). B. Liang is a predoctoral fellow of the American Heart Association, Florida/Puerto Rico Affiliate (0615182B). X-ray diffraction data were collected from the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions for APS beamlines are listed at http://necat.chem.cornell.edu/ and www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

AUTHOR CONTRIBUTIONS

B.L. designed and carried out crystallographic studies of the wild-type complex, acquired fluorescence data, and contributed to manuscript preparation; S.X. carried out crystallographic studies of the D85A mutant complex and contributed to manuscript preparation; M.P.T. and R.M.T. supplied plasmids encoding H/ACA RNP proteins and contributed to manuscript preparation; H.L. supervised the project and contributed to manuscript preparation.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Hannon, G.J., Rivas, F.V., Murchison, E.P. & Steitz, J.A. The expanding universe of noncoding RNAs. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 551–564 (2006).
2. Matera, A.G., Terns, R.M. & Terns, M.P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Genet.* **8**, 209–220 (2007).
3. Huttenhofer, A. & Schattner, P. The principles of guiding by RNA: chimeric RNA-protein enzymes. *Nat. Rev. Genet.* **7**, 475–482 (2006).
4. Decatur, W.A. & Fournier, M.J. RNA-guided nucleotide modification of ribosomal and other RNAs. *J. Biol. Chem.* **278**, 695–698 (2003).
5. Yu, Y.T., Terns, R.M. & Terns, M.P. in *Fine-tuning of RNA Functions by Modification and Editing*. (ed. Grosjean, H.) 223–262 (Springer, New York, 2005).
6. Kiss, T. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* **20**, 3617–3622 (2001).
7. Decatur, W.A. & Fournier, M.J. rRNA modifications and ribosome function. *Trends Biochem. Sci.* **27**, 344–351 (2002).
8. King, T.H., Liu, B., McCully, R.R. & Fournier, M.J. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol. Cell* **11**, 425–435 (2003).
9. Ni, J., Tien, A.L. & Fournier, M.J. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565–573 (1997).
10. Ganot, P., Bortolin, M.L. & Kiss, T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799–809 (1997).
11. Darzacq, X. *et al.* Cajal body-specific small nuclear RNAs: a novel class of 2′-*O*-methylation and pseudouridylation guide RNAs. *EMBO J.* **21**, 2746–2756 (2002).
12. Eliceiri, G.L. The vertebrate E1/U17 small nucleolar ribonucleoprotein particle. *J. Cell. Biochem.* **98**, 486–495 (2006).
13. Collins, K. The biogenesis and regulation of telomerase holoenzymes. *Nat. Rev. Mol. Cell Biol.* **7**, 484–494 (2006).

14. Omer, A.D., Ziesche, S., Decatur, W.A., Fournier, M.J. & Dennis, P.P. RNA-modifying machines in archaea. *Mol. Microbiol.* **48**, 617–629 (2003).
15. Arnez, J.G. & Steitz, T.A. Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry* **33**, 7560–7567 (1994).
16. Davis, D.R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **23**, 5020–5026 (1995).
17. Newby, M.I. & Greenbaum, N.L. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA* **7**, 833–845 (2001).
18. Yarian, C.S. *et al.* Structural and functional roles of the N1- and N3-protons of psi at tRNA's position 39. *Nucleic Acids Res.* **27**, 3543–3549 (1999).
19. Newby, M.I. & Greenbaum, N.L. Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. *Proc. Natl. Acad. Sci. USA* **99**, 12697–12702 (2002).
20. Yang, C., McPheeters, D.S. & Yu, Y.T. Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**, 6655–6662 (2005).
21. Zhao, X. & Yu, Y.T. Pseudouridines in and near the branch site recognition region of U2 snRNA are required for snRNP biogenesis and pre-mRNA splicing in *Xenopus* oocytes. *RNA* **10**, 681–690 (2004).
22. Donmez, G., Hartmuth, K. & Luhrmann, R. Modified nucleotides at the 5′ end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA* **10**, 1925–1933 (2004).
23. Yu, Y.T., Shu, M.D. & Steitz, J.A. Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J.* **17**, 5783–5795 (1998).
24. Valadkhan, S. & Manley, J.L. Characterization of the catalytic activity of U2 and U6 snRNAs. *RNA* **9**, 892–904 (2003).
25. Yu, Y.T. The most complex pseudouridylase. *Structure* **14**, 167–168 (2006).
26. Reichow, S.L., Hamma, T., Ferre-D'Amare, A.R. & Varani, G. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res.* **35**, 1452–1464 (2007).
27. Wang, C. & Meier, U.T. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.* **23**, 1857–1867 (2004).
28. Baker, D.L. *et al.* RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev.* **19**, 1238–1248 (2005).
29. Charpentier, B., Muller, S. & Branlant, C. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res.* **33**, 3133–3144 (2005).
30. Koonin, E.V. Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res.* **24**, 2411–2415 (1996).
31. Marrone, A., Walne, A. & Dokal, I. Dyskeratosis congenita: telomerase, telomeres and anticipation. *Curr. Opin. Genet. Dev.* **15**, 249–257 (2005).
32. Marrone, A. & Mason, P.J. Dyskeratosis congenita. *Cell. Mol. Life Sci.* **60**, 507–517 (2003).
33. Hamma, T. & Ferre-D'Amare, A.R. Pseudouridine synthases. *Chem. Biol.* **13**, 1125–1135 (2006).
34. Normand, C. *et al.* Analysis of the binding of the N-terminal conserved domain of yeast Cbf5p to a box H/ACA snoRNA. *RNA* **12**, 1868–1882 (2006).
35. Bortolin, M.L., Ganot, P. & Kiss, T. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *EMBO J.* **18**, 457–469 (1999).
36. Manival, X. *et al.* Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res.* **34**, 826–839 (2006).
37. Hamma, T., Reichow, S.L., Varani, G. & Ferre-D'Amare, A.R. The Cbf5–Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat. Struct. Mol. Biol.* **12**, 1101–1107 (2005).
38. Rashid, R. *et al.* Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol. Cell* **21**, 249–260 (2006).
39. Li, L. & Ye, K. Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**, 302–307 (2006).
40. Wu, H. & Feigon, J. H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc. Natl. Acad. Sci. USA* **104**, 6655–6660 (2007).
41. Jin, H., Loria, J.P. & Moore, P.B. Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol. Cell* **26**, 205–215 (2007).
42. Schattner, P. *et al.* Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res.* **32**, 4281–4296 (2004).
43. Ganot, P., Caizergues-Ferrer, M. & Kiss, T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.* **11**, 941–956 (1997).
44. Walne, A.J. *et al.* Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. *Hum. Mol. Genet.* **16**, 1619–1629 (2007).
45. Hoang, C., Hamilton, C.S., Mueller, E.G. & Ferre-D'Amare, A.R. Precursor complex structure of pseudouridine synthase TruB suggests coupling of active site perturbations to an RNA-sequestering peripheral protein domain. *Protein Sci.* **14**, 2201–2206 (2005).
46. Baker, D. *et al.* Determination of protein-RNA interaction sites in the CBF5-H/ACA guide RNA complex by mass spectrometric protein footprinting. *Biochemistry* (in the press).
47. Hoang, C. & Ferre-D'Amare, A.R. Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**, 929–939 (2001).
48. Pan, H., Agarwalla, S., Moustakas, D.T., Finer-Moore, J. & Stroud, R.M. Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc. Natl. Acad. Sci. USA* **100**, 12648–12653 (2003).
49. Otwinowski, Z., & Minor, W. in *Processing of X-ray Diffraction Data Collected in Oscillation Mode Methods in Enzymology* Vol. 276 (eds. Carter, C.W. & Sweet, R.M.) 307–326 (Academic Press, San Diego, 1997).
50. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
51. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **D50**, 760–763 (1994).
52. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905–921 (1998).
53. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **53**, 240–255 (1997).
54. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).

Substrate RNA Positioning in the Archaeal H/ACA Ribonucleoprotein Complex

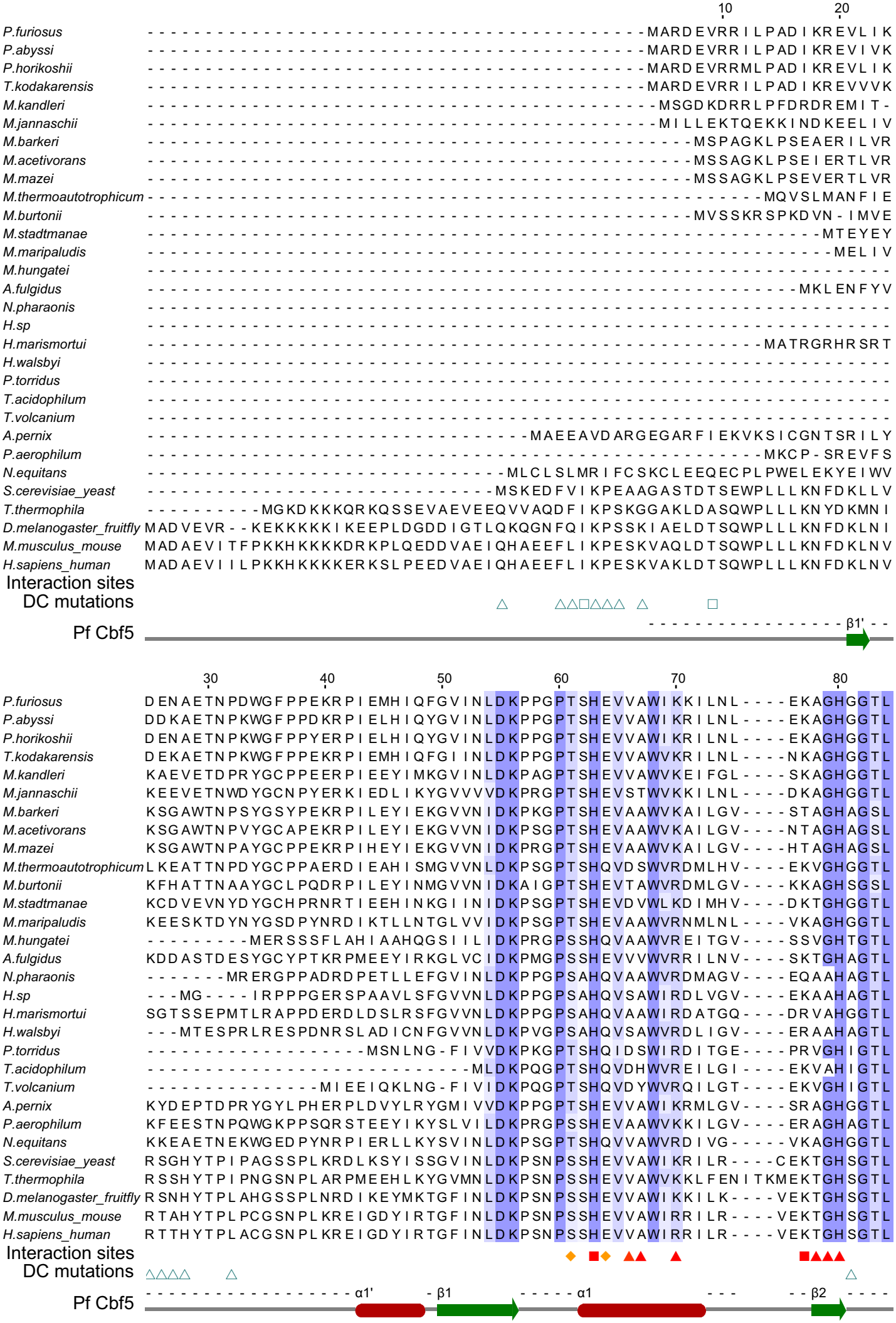
Bo Liang[†2], *Song Xue*[†1,2], *Rebecca M. Terns*[3], *Michael P. Terns*[3,4] *and Hong Li*[*1,2]

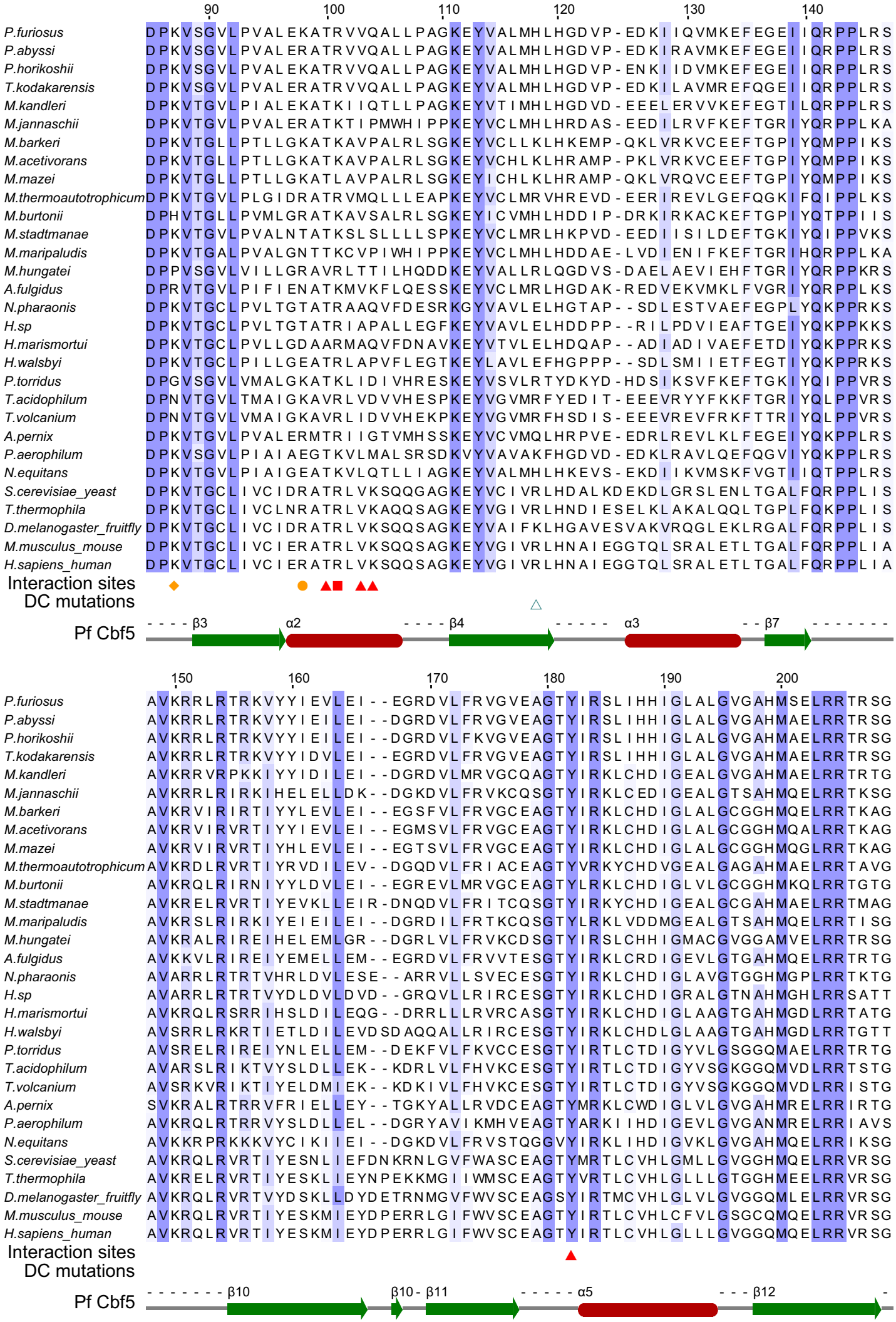
[1]*Department of Chemistry and Biochemistry,* [2]*Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306*

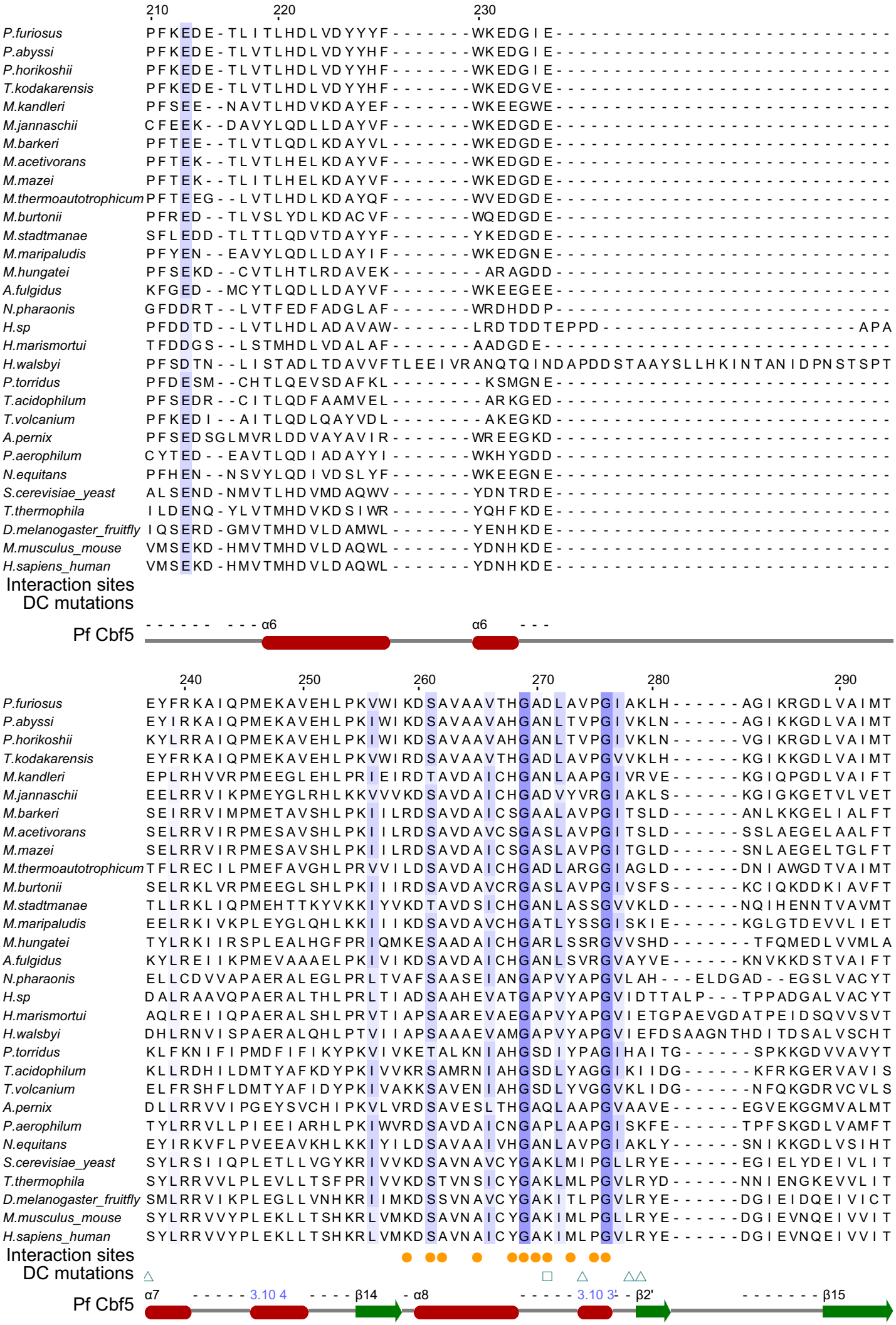
[3]*Departments of Biochemistry and Molecular Biology, and* [4]*Genetics, University of Georgia, Athens, Georgia 30602*

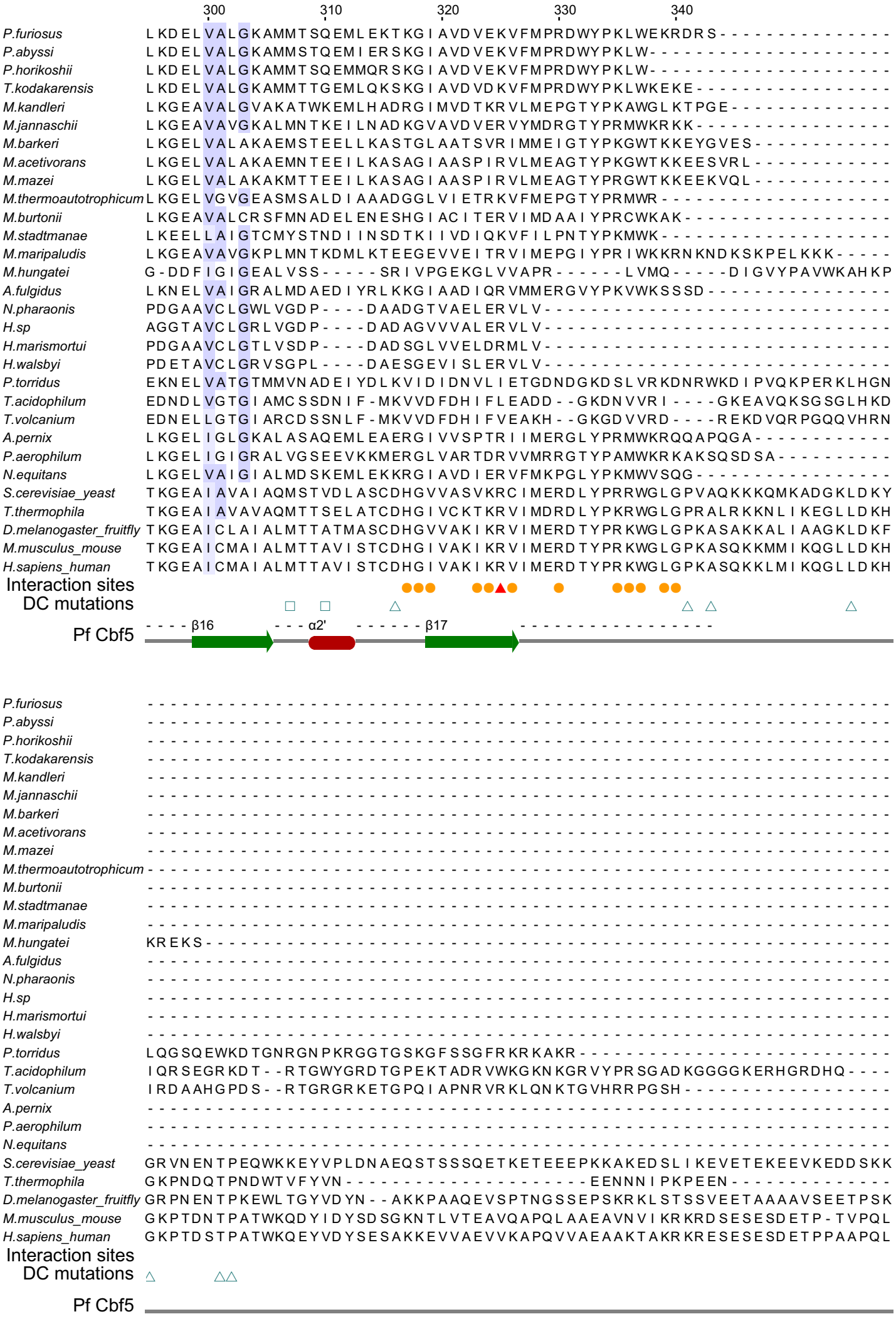
[†]authors contributed to the work equally.

[*]Corresponding author. Email: hongli@sb.fsu.edu,
Telephone 850-644-6785, Fax: 850-644-7244


10
20
P.furiosus
P.abyssi
P.horikoshii
T.kodakarensis
M.kandleri
M.jannaschii
M.barkeri
M.acetivorans
M.mazei
M.thermoautotrophicum
M.burtonii
M.stadtmanae
M.maripaludis
M.hungatei
A.fulgidus
N.pharaonis
H.sp
H.marismortui
H.walsbyi
P.torridus
T.acidophilum
T.volcanium
A.pernix
P.aerophilum
N.equitans
S.cerevisiae_yeast
T.thermophila
D.melanogaster_fruitfly
M.musculus_mouse
H.sapiens_human
Interaction sites
DC mutations
Pf Cbf5
β1'
30
40
50
60
70
80
P.furiosus
P.abyssi
P.horikoshii
T.kodakarensis
M.kandleri
M.jannaschii
M.barkeri
M.acetivorans
M.mazei
M.thermoautotrophicum
M.burtonii
M.stadtmanae
M.maripaludis
M.hungatei
A.fulgidus
N.pharaonis
H.sp
H.marismortui
H.walsbyi
P.torridus
T.acidophilum
T.volcanium
A.pernix
P.aerophilum
N.equitans
S.cerevisiae_yeast
T.thermophila
D.melanogaster_fruitfly
M.musculus_mouse
H.sapiens_human
Interaction sites
DC mutations
Pf Cbf5
α1'
β1
α1
β2


90
100
110
120
130
140
P.furiosus DPKVSGVLPVALEKATRVVQALLPAGKEYVALMHLHGDVP-EDKIIQVMKEFEGEIIQRPPLRS
P.abyssi DPKVSGVLPVALERATRVVQALLPAGKEYVALMHLHGDVP-EDKIRAVMKEFEGEIIQRPPLRS
P.horikoshii DPKVSGVLPVALEKATRVVQALLPAGKEYVALMHLHGDVP-ENKIIDVMKEFEGEIIQRPPLRS
T.kodakarensis DPKVSGVLPVALERATRVVQALLPAGKEYVALMHLHGDVP-EDKILAVMREFQGEIIQRPPLRS
M.kandleri DPKVTGVLPIALEKATKIIQTLLPAGKEYVTIMHLHGDVD-EEELERVVKEFEGTILQRPPLRS
M.jannaschii DPKVTGVLPVALERATKTIPMWHIPPKEYVCLMHLHRDAS-EEDILRVFKEFTGRIYQRPPLKA
M.barkeri DPKVTGLLPTLLGKATKAVPALRLSGKEYVCLLKLHKEMP-QKLVRKVCEEFTGPIYQMPPIKS
M.acetivorans DPKVTGLLPTLLGKATKAVPALRLSGKEYVCHLKLHRAMP-PKLVRKVCEEFTGPIYQMPPIKS
M.mazei DPKVTGLLPTLLGKATLAVPALRLSGKEYICHLKLHRAMP-QKLVRQVCEEFTGPIYQMPPIKS
M.thermoautotrophicum DPKVTGVLPLGIDRATRVMQLLLEAPKEYVCLMRVHREVD-EERIREVLGEFQGKIFQIPPLKS
M.burtonii DPHVTGLLPVMLGRATKAVSALRLSGKEYICVMHLHDDIP-DRKIRKACKEFTGPIYQTPPIIS
M.stadtmanae DPKVTGVLPVALNTATKSLSLLLLSPKEYVCLMRLHKPVD-EEDIISILDEFTGKIYQIPPVKS
M.maripaludis DPKVTGALPVALGNTTKCVPIWHIPPKEYVCLMHLHDDAE-LVDIENIFKEFTGRIHQRPPLKA
M.hungatei DPPVSGVLVILLGRAVRLTTILHQDDKEYVALLRLQGDVS-DAELAEVIEHFTGRIYQRPPKRS
A.fulgidus DPRVTGVLPIFIENATKMVKFLQESSKEYVCLMRLHGDAK-REDVEKVMKLFVGRIYQRPPLKS
N.pharaonis DPKVTGCLPVLTGTATRAAQVFDESRKGYVAVLELHGTAP--SDLESTVAEFEGPLYQKPPRKS
H.sp DPKVTGCLPVLTGTATRIAPALLEGFKEYVAVLELHDDPP--RILPDVIEAFTGEIYQKPPKKS
H.marismortui DPKVTGCLPVLLGDAARMAQVFDNAVKEYVTVLELHDQAP--ADIADIVAEFETDIYQKPPRKS
H.walsbyi DPKVTGCLPILLGEATRLAPVFLEGTKEYLAVLEFHGPPP--SDLSMIIETFEGTIYQKPPRKS
P.torridus DPGVSGVLVMALGKATKLIDIVHRESKEYVSVLRTYDKYD-HDSIKSVFKEFTGKIYQIPPVRS
T.acidophilum DPNVTGVLTMAIGKAVRLVDVVHESPKEYVGVMRFYEDIT-EEEVRYYFKKFTGRIYQLPPVRS
T.volcanium DPNVTGVLVMAIGKAVRLIDVVHEKPKEYVGVMRFHSDIS-EEEVREVFRKFTTRIYQLPPVRS
A.pernix DPKVTGVLPVALERMTRIIGTVMHSSKEYVCVMQLHRPVE-EDRLREVLKLFEGEIYQKPPLRS
P.aerophilum DPKVSGVLPIAIAEGTKVLMALSRSDKVYVAVAKFHGDVD-EDKLRAVLQEFQGVIYQKPPLRS
N.equitans DPKVTGVLPIAIGEATKVLQTLLIAGKEYVALMHLHKEVS-EKDIIKVMSKFVGTIIQTPPLRS
S.cerevisiae_yeast DPKVTGCLIVCIDRATRLVKSQQGAGKEYVCIVRLHDALKDEKDLGRSLENLTGALFQRPPLIS
T.thermophila DPKVTGCLIVCLNRATRLVKAQQSAGKEYVGIVRLHNDIESELKLAKALQQLTGPLFQKPPLIS
D.melanogaster_fruitfly DPKVTGCLIVCIDRATRLVKSQQSAGKEYVAIFKLHGAVESVAKVRQGLEKLRGALFQRPPLIS
M.musculus_mouse DPKVTGCLIVCIERATRLVKSQQSAGKEYVGIVRLHNAIEGGTQLSRALETLTGALFQRPPLIA
H.sapiens_human DPKVTGCLIVCIERATRLVKSQQSAGKEYVGIVRLHNAIEGGTQLSRALETLTGALFQRPPLIA
Interaction sites
DC mutations
Pf Cbf5
β3
α2
β4
α3
β7
150
160
170
180
190
200
P.furiosus AVKRRLRTRKVYYIEVLEI--EGRDVLFRVGVEAGTYIRSLIHHIGLALGVGAHMSELRRTRSG
P.abyssi AVKRRLRTRKVYYIEILEI--DGRDVLFRVGVEAGTYIRSLIHHIGLALGVGAHMAELRRTRSG
P.horikoshii AVKRRLRTRKVYYIEILEI--DGRDVLFKVGVEAGTYIRSLIHHIGLALGVGAHMAELRRTRSG
T.kodakarensis AVKRRLRTRKVYYIDVLEI--EGRDVLFRVGVEAGTYIRSLIHHIGLALGVGAHMAELRRTRSG
M.kandleri AVKRRVRPKKIYYIDILEI--DGRDVLMRVGCQAGTYIRKLCHDIGEALGVGAHMAELRRTRTG
M.jannaschii AVKRRLRIRKIHELELLDK--DGKDVLFRVKCQSGTYIRKLCEDIGEALGTSAHMQELRRTKSG
M.barkeri AVKRVIRIRTIYYLEVLEI--EGSFVLFRVGCEAGTYIRKLCHDIGLALGCGGHMQELRRTKAG
M.acetivorans AVKRVIRVRTIYYIEVLEI--EGMSVLFRVGCEAGTYIRKLCHDIGLALGCGGHMQALRRTKAG
M.mazei AVKRVIRVRTIYHLEVLEI--EGTSVLFRVGCEAGTYIRKLCHDIGLALGCGGHMQGLRRTKAG
M.thermoautotrophicum AVKRDLRVRTIYRVDILEV--DGQDVLFRIACEAGTYVRKYCHDVGEALGAGAHMAELRRTAVG
M.burtonii AVKRQLRIRNIYYLDVLEI--EGREVLMRVGCEAGTYLRKLCHDIGLVLGCGGHMKQLRRTGTG
M.stadtmanae AVKRELRVRTIYEVKLLEIR-DNQDVLFRITCQSGTYIRKYCHDIGEALGCGAHMAELRRTMAG
M.maripaludis AVKRSLRIRKIYEIEILEI--DGRDILFRTKCQSGTYLRKLVDDMGEALGTSAHMQELRRTISG
M.hungatei AVKRALRIREIHELEMLGR--DGRLVLFRVKCDSGTYIRSLCHHIGMACGVGGAMVELRRTRSG
A.fulgidus AVKKVLRIREIYEMELLEM--EGRDVLFRVVTESGTYIRKLCRDIGEVLGTGAHMQELRRTRTG
N.pharaonis AVARRLRTRTVHRLDVLESE--ARRVLLSVECESGTYIRKLCHDIGLAVGTGGHMGPLRRTKTG
H.sp AVARRLRTRTVYDLDVLDVD--GRQVLLRIRCESGTYIRKLCHDIGRALGTNAHMGHLRRSATT
H.marismortui AVKRQLRSRRIHSLDILEQG--DRRLLLRVRCASGTYIRKLCHDIGLAAGTGAHMGDLRRTATG
H.walsbyi AVSRRLRKRTIETLDILEVDSDAQQALLRIRCESGTYIRKLCHDLGLAAGTGAHMGDLRRTGTT
P.torridus AVSRELRIREIYNLELLEM--DEKFVLFKVCCESGTYIRTLCTDIGYVLGSGGQMAELRRTRTG
T.acidophilum AVARSLRIKTVYSLDLLEK--KDRLVLFHVKCESGTYIRTLCTDIGYVSGKGGQMVDLRRTSTG
T.volcanium AVSRKVRIKTIYELDMIEK--KDKIVLFHVKCESGTYIRTLCTDIGYVSGKGGQMVDLRRISTG
A.pernix SVKRALRTRRVFRIELLEY--TGKYALLRVDCEAGTYMRKLCWDIGLVLGVGAHMRELRRIRTG
P.aerophilum AVKRQLRTRRVYSLDLLEL--DGRYAVIKMHVEAGTYARKIIHDIGEVLGVGANMRELRRIAVS
N.equitans AVKKRPRKKKVYCIKIIEI--DGKDVLFRVSTQGGVYIRKLIHDIGVKLGVGAHMQELRRIIKSG
S.cerevisiae_yeast AVKRQLRVRTIYESNLIEFDNKRNLGVFWASCEAGTYMRTLCVHLGMLLGVGGHMQELRRVRSG
T.thermophila AVKRELRVRTIYESKLIEYNPEKKMGIIWMSCEAGTYVRTLCVHLGLLLGTGGHMEELRRVRSG
D.melanogaster_fruitfly AVKRQLRVRTVYDSKLLDYDETRNMGVFWVSCEAGSYIRTMCVHLGLVLGVGGQMLELRRVRSG
M.musculus_mouse AVKRQLRVRTIYESKMIEYDPERRLGIFWVSCEAGTYIRTLCVHLCFVLGSGCQMQELRRVRSG
H.sapiens_human AVKRQLRVRTIYESKMIEYDPERRLGIFWVSCEAGTYIRTLCVHLGLLLGVGGQMQELRRVRSG
Interaction sites
DC mutations
Pf Cbf5
β10
β10-β11
α5
β12


210
220
230
P.furiosus PFKEDE-TLITLHDLVDYYYF-------WKEDGIE
P.abyssi PFKEDE-TLVTLHDLVDYYHF-------WKEDGIE
P.horikoshii PFKEDE-TLVTLHDLVDYYHF-------WKEDGIE
T.kodakarensis PFKEDE-TLVTLHDLVDYYHF-------WKEDGVE
M.kandleri PFSEE--NAVTLHDVKDAYEF-------WKEEGWE
M.jannaschii CFEEK--DAVYLQDLLDAYVF-------WKEDGDE
M.barkeri PFTEE--TLVTLQDLKDAYVL-------WKEDGDE
M.acetivorans PFTEK--TLVTLHELKDAYVF-------WKEDGDE
M.mazei PFTEK--TLITLHELKDAYVF-------WKEDGDE
M.thermoautotrophicum PFTEEG--LVTLHDLKDAYQF-------WVEDGDE
M.burtonii PFRED--TLVSLYDLKDACVF-------WQEDGDE
M.stadtmanae SFLEDD-TLTTLQDVTDAYYF-------YKEDGDE
M.maripaludis PFYEN--EAVYLQDLLDAYIF-------WKEDGNE
M.hungatei PFSEKD--CVTLHTLRDAVEK--------ARAGDD
A.fulgidus KFGED--MCYTLQDLLDAYVF-------WKEEGEE
N.pharaonis GFDDRT--LVTFEDFADGLAF-------WRDHDDP
H.sp PFDDTD--LVTLHDLADAVAW-------LRDTDDTEPPD----------APA
H.marismortui TFDDGS--LSTMHDLVDALAF-------AADGDE
H.walsbyi PFSDTN--LISTADLTDAVVFTLEEIVRANQTQINDAPDDSTAAYSLLHKINTANIDPNSTSPT
P.torridus PFDESM--CHTLQEVSDAFKL--------KSMGNE
T.acidophilum PFSEDR--CITLQDFAAMVEL--------ARKGED
T.volcanium PFKEDI--AITLQDLQAYVDL--------AKEGKD
A.pernix PFSEDSGLMVRLDDVAYAVIR-------WREEGKD
P.aerophilum CYTED--EAVTLQDIADAYYI-------WKHYGDD
N.equitans PFHEN--NSVYLQDIVDSLYF-------WKEEGNE
S.cerevisiae_yeast ALSEND-NMVTLHDVMDAQWV-------YDNTRDE
T.thermophila ILDENQ-YLVTMHDVKDSIWR-------YQHFKDE
D.melanogaster_fruitfly IQSERD-GMVTMHDVLDAMWL-------YENHKDE
M.musculus_mouse VMSEKD-HMVTMHDVLDAQWL-------YDNHKDE
H.sapiens_human VMSEKD-HMVTMHDVLDAQWL-------YDNHKDE
Interaction sites
DC mutations
Pf Cbf5
α6
α6
240
250
260
270
280
290
P.furiosus EYFRKAIQPMEKAVEHLPKVWIKDSAVAAVTHGADLAVPGIAKLH------AGIKRGDLVAIMT
P.abyssi EYIRKAIQPMEKAVEHLPKIWIKDSAVAAVAHGANLTVPGIVKLN------AGIKKGDLVAIMT
P.horikoshii KYLRRAIQPMEKAVEHLPKIWIKDSAVAAVAHGANLTVPGIVKLN------VGIKRGDLVAIMT
T.kodakarensis EYFRKAIQPMEKAVEHLPKVWIRDSAVAAVTHGADLAVPGVVKLH------KGIKKGDLVAIMT
M.kandleri EPLRHVVRPMEEGLEHLPRIEIRDTAVDAICHGANLAAPGIVRVE------KGIQPGDLVAIFT
M.jannaschii EELRRVIKPMEYGLRHLKKVVVKDSAVDAICHGADVYVRGIAKLS------KGIGKGETVLVET
M.barkeri SEIRRVIMPMETAVSHLPKIILRDSAVDAICSGAALAVPGITSLD------ANLKKGELIALFT
M.acetivorans SELRRVIRPMESAVSHLPKIILRDSAVDAVCSGASLAVPGITSLD------SSLAEGELAALFT
M.mazei SELRRVIRPMESAVSHLPKIILRDSAVDAICSGASLAVPGITGLD------SNLAEGELTGLFT
M.thermoautotrophicum TFLRECILPMEFAVGHLPRVVILDSAVDAICHGADLARGGIAGLD------DNIAWGDTVAIMT
M.burtonii SELRKLVRPMEEGLSHLPKIIIRDSAVDAVCRGASLAVPGIVSFS------KCIQKDDKIAVFT
M.stadtmanae TLLRKLIQPMEHTTKYVKKIYVKDTAVDSICHGANLASSGVVKLD------NQIHENNTVAVMT
M.maripaludis EELRKIVKPLEYGLQHLKKIIIKDSAVDAVCHGATLYSSGISKIE------KGLGTDEVVLIET
M.hungatei TYLRKIIRSPLEALHGFPRIQMKESAADAICHGARLSSRGVVSHD-------TFQMEDLVVMLA
A.fulgidus KYLREIIKPMEVAAAELPKIVIKDSAVDAICHGANLSVRGVAYVE------KNVKKDSTVAIFT
N.pharaonis ELLCDVVAPAERALEGLPRLTVAFSAASEIANGAPVYAPGVLAH---ELDGAD--EGSLVACYT
H.sp DALRAAVQPAERALTHLPRLTIADSAAHEVATGAPVYAPGVIDTTALP---TPPADGALVACYT
H.marismortui AQLREIIQPAERALSHLPRVTIAPSAAREVAEGAPVYAPGVIETGPAEVGDATPEIDSQVVSVT
H.walsbyi DHLRNVISPAERALQHLPTVIIAPSAAAEVAMGAPVYAPGVIEFDSAAGNTHDITDSALVSCHT
P.torridus KLFKNIFIPMDFIFIKYPKVIVKETALKNIAHGSDIYPAGIHAITG------SPKKGDVVAVYT
T.acidophilum KLLRDHILDMTYAFKDYPKIVVKRSAMRNIAHGSDLYAGGIKIIDG------KFRKGERVAVIS
T.volcanium ELFRSHFLDMTYAFIDYPKIVAKKSAVENIAHGSDLYVGGVKLIDG------NFQKGDRVCVLS
A.pernix DLLRRVVIPGEYSVCHIPKVLVRDSAVESLTHGAQLAAPGVAAVE------EGVEKGGMVALMT
P.aerophilum TYLRRVLLPIEEIARHLPKIWVRDSAVDAICNGAPLAAPGISKFE------TPFSKGDLVAMFT
N.equitans EYIRKVFLPVEEAVKHLKKIYILDSAVAAIVHGANLAVPGIAKLY------SNIKKGDLVSIHT
S.cerevisiae_yeast SYLRSIIQPLETLLVGYKRIVVKDSAVNAVCYGAKLMIPGLLRYE------EGIELYDEIVLIT
T.thermophila SYLRRVVLPLEVLLTSFPRIVVKDSTVNSICYGAKLMLPGVLRYD------NNIENGKEVVLIT
D.melanogaster_fruitfly SMLRRVIKPLEGLLVNHKRIIMKDSSVNAVCYGAKITLPGVLRYE------DGIEIDQEIVICT
M.musculus_mouse SYLRRVVYPLEKLLTSHKRLVMKDSAVNAICYGAKIMLPGLLRYE------DGIEVNQEIVVIT
H.sapiens_human SYLRRVVYPLEKLLTSHKRLVMKDSAVNAICYGAKIMLPGVLRYE------DGIEVNQEIVVIT
Interaction sites
DC mutations
Pf Cbf5
α7
3.10 4
β14
α8
3.10 3
β2'
β15


300
310
320
330
340
P.furiosus
P.abyssi
P.horikoshii
T.kodakarensis
M.kandleri
M.jannaschii
M.barkeri
M.acetivorans
M.mazei
M.thermoautotrophicum
M.burtonii
M.stadtmanae
M.maripaludis
M.hungatei
A.fulgidus
N.pharaonis
H.sp
H.marismortui
H.walsbyi
P.torridus
T.acidophilum
T.volcanium
A.pernix
P.aerophilum
N.equitans
S.cerevisiae_yeast
T.thermophila
D.melanogaster_fruitfly
M.musculus_mouse
H.sapiens_human
Interaction sites
DC mutations
β16
α2'
β17
Pf Cbf5
Interaction sites
DC mutations
Pf Cbf5

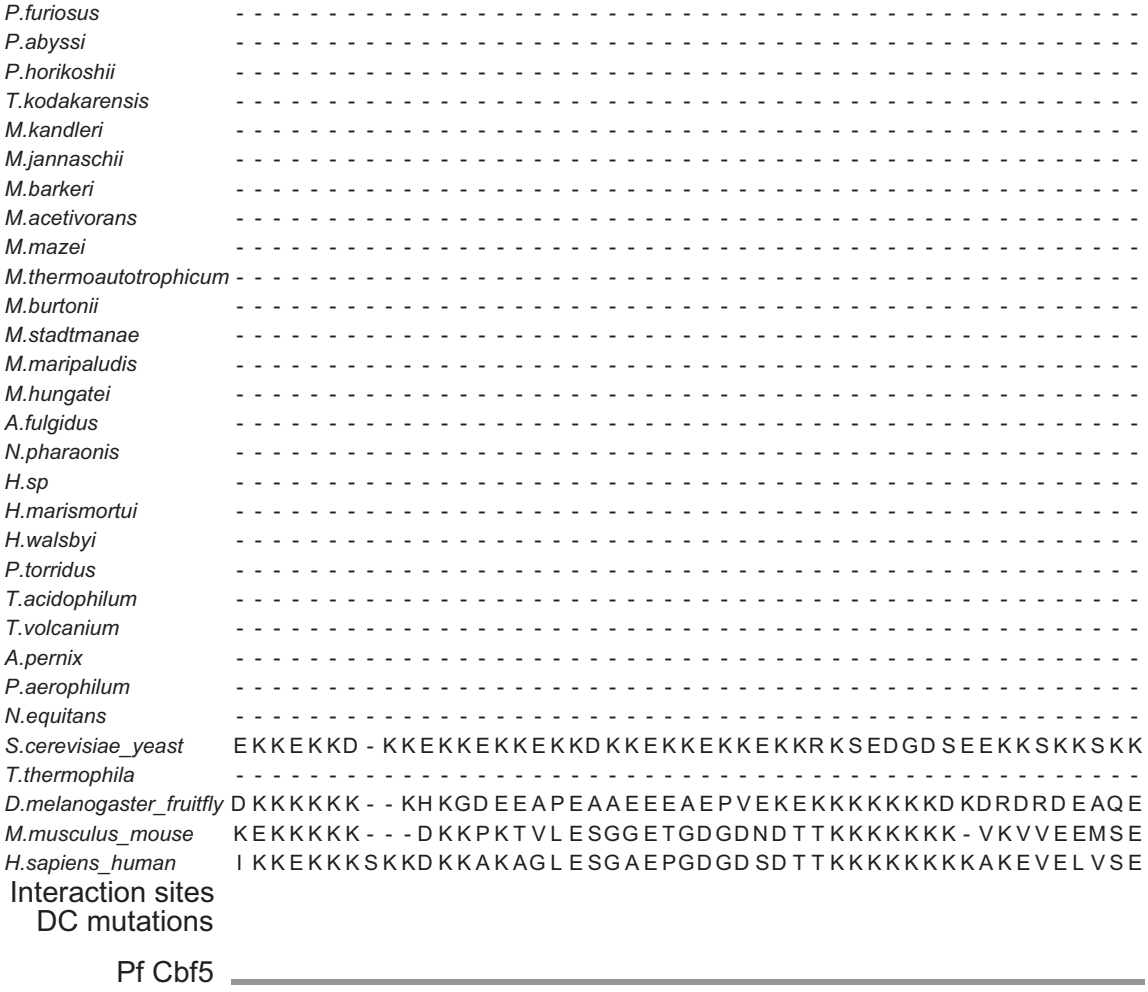

P.furiosus
P.abyssi
P.horikoshii
T.kodakarensis
M.kandleri
M.jannaschii
M.barkeri
M.acetivorans
M.mazei
M.thermoautotrophicum
M.burtonii
M.stadtmanae
M.maripaludis
M.hungatei
A.fulgidus
N.pharaonis
H.sp
H.marismortui
H.walsbyi
P.torridus
T.acidophilum
T.volcanium
A.pernix
P.aerophilum
N.equitans
S.cerevisiae_yeast E K K E K K D - K K E K K E K K E K K D K K E K K E K K E K K R K S E D G D S E E K K S K K S K K
T.thermophila
D.melanogaster_fruitfly D K K K K K K - - K H K G D E E A P E A A E E E A E P V E K E K K K K K K K D K D R D R D E A Q E
M.musculus_mouse K E K K K K K - - - D K K P K T V L E S G G E T G D G D N D T T K K K K K K K - V K V V E E M S E
H.sapiens_human I K K E K K K S K K D K K A K A G L E S G A E P G D G D S D T T K K K K K K K K A K E V E L V S E
Interaction sites
DC mutations
Pf Cbf5

Figure S1. Sequence alignment of Cbf5 and sites of interactions with the pseudouridine pocket (red triangle for SH1 or SH2 stem) and with the guide RNA alone (yellow diamond for P2 and yellow circle for P1 stem). Residues interacting with both guide RNA (P2 or P1) and the pseudouridine pocket (SH1 or SH2) are labeled by red squares. Sites of interactions are defined as those that have more than 2 $Å^2$ decrease in solvent accessible surface as result of RNA binding. Mutations in human Cbf5 that cause dyskeratosis congenita (DC) are also indicated with open squares (multiple families) or triangles (single family).

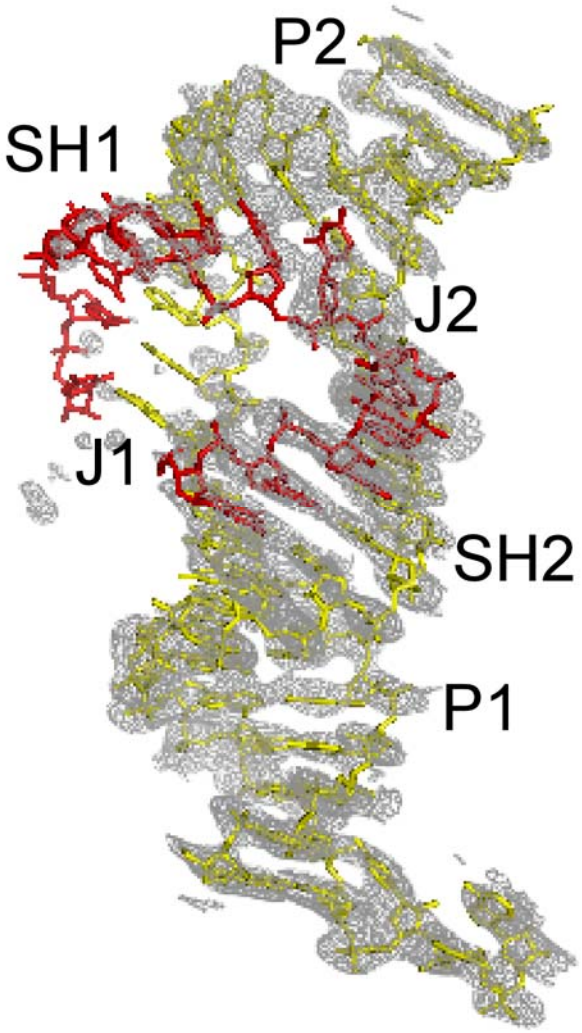


Figure S2. Density modified map ($3F_o$-$2F_c$) computed prior to modeling of RNA and using protein coordinates for phases only. The map is displayed around the RNA as indicated by the stick models (guide RNA is in yellow and target RNA is in red).

Table S1. Diffraction data statistics of the crystals containing the wild-type Cbf5-Nop10-Gar1 complex bound with the bi-molecular guide RNA and the target RNA containing 5-fluorouridine at the target site. Values in parentheses refer to those of the highest resolution shell.

Space group*	$P4_12_12$
Unit-cell parameters (Å)	
a	96.629
b	96.629
c	238.951
Resolution range (Å)	50-8.0 (8.3-8.0)
No. of unique reflections	2309 (230)
Redundance	6.4 (6.7)
Completeness (%)	98.7 (100.0)
$I/\sigma(I)$	41.1 (21.8)
R_{sym}(%)	7.9 (15.5)

*the space group is assumed to be the same as the Asp85Ala mutant-containing complex

Table S2. Elements of **T**, **L**, **S** tensors for the six groups in the orthogonal coordinate system used in TLS refinement

Group 1. Cbf5									
T(Å^2)	-0.0205	-0.0707	0.3184	-0.3764	-0.2188	0.1242			
L(°^{2})	1.876	1.7205	4.9607	0.1261	0.5403	-1.2868			
S(Å2)	0.0846	-0.4581	0.3734	-0.1793	-0.2675	-0.515	0.4228	-0.5061	1.0077
Group 2. Nop10									
T(Å^2)	-0.1567	-0.273	0.1119	-0.2459	-0.1013	-0.1515			
L(°^{2})	7.6412	3.1016	11.2289	1.7167	1.1447	-3.4687			
S(Å2)	0.0143	-0.1582	0.1439	1.0323	-0.3755	-0.2104	-0.2068	0.2505	0.6775
Group 3. Gar1									
T(Å^2)	0.5526	0.9082	0.8904	-0.6972	-0.2446	0.4869			
L(°^{2})	14.0046	6.7073	10.2546	2.7106	1.6742	-4.3765			
S(Å2)	-0.3504	0.2738	0.0766	1.7614	1.3078	0.2559	-0.4432	-1.0284	1.1234
Group 4. 5' guide RNA strand									
T(Å^2)	0.5429	0.4862	0.2118	-0.19	-0.0595	0.2529			
L(°^{2})	1.9926	4.708	17.4169	-0.8017	5.7773	-4.0332			
S(Å2)	0.1349	0.1396	-0.2746	-0.6752	0.5955	0.4962	0.8044	-2.6099	-0.9471
Group 5. 3' guide RNA strand									
T(Å^2)	0.5869	0.0952	-0.0135	-0.5015	-0.3204	-0.0524			
L(°^{2})	2.6505	2.7531	7.8963	-1.5719	0.9959	-3.3112			
S(Å2)	0.0247	0.0135	-0.0381	-0.6623	0.0124	-0.1304	0.7577	-1.2759	-0.0453
Group 6. target RNA									
T(Å^2)	0.671	0.1768	0.2631	-0.7094	-0.4597	-0.0996			
L(°^{2})	17.8313	30.3146	0	12.1335	0	0			
S(Å2)	-0.4621	0.3104	0.1516	-0.3801	1.0859	0.6757	0.1528	-0.7902	-1.283