doi:10.1016/j.jmb.2006.03.037 J. Mol. Biol. (2006) 359, 1–9

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

COMMUNICATION

Crystal Structure of the C-terminal Domain of *S. cerevisiae* eIF5

Zhiyi Wei[1,2]†, Yanyan Xue[1,2]†, Hang Xu[1] and Weimin Gong[1,2]*

[1]*National Laboratory of Biomacromolecules, USTC-IBP Joint Laboratory for Protein Sciences, Institute of Biophysics Chinese Academy of Sciences Beijing 100101, People's Republic of China*

[2]*School of Life Sciences University of Science and Technology of China, Hefei Anhui 230026, People's Republic of China*

**Corresponding author*

eIF5, a GTPase-activating protein (GAP) specific for eIF2, plays a critical role in pre-initiation complex assembly and correct AUG selection during eukaryotic translation initiation. eIF5 is involved in the formation of the multifactor complex (MFC), an important intermediate of the 43 S pre-initiation complex. The C-terminal domain (CTD) of eIF5 functions as the structural core in the MFC assembly. Here we report the 1.5 Å crystal structure of eIF5-CTD, confirming that eIF5-CTD contains an atypical HEAT motif. In addition, analyzing the electrostatic potential and the distribution of conserved residues on the protein surface, we confirm and suggest some potential regions of interactions between eIF5-CTD and other eIFs. The structure of eIF5-CTD provides useful information in understanding the mechanism of the MFC assembly.

Keywords: crystal structure; eIF5-CTD; multifactor complex; translation initiation; pre-initiation complex

Translation initiation is a sophisticated, well regulated and highly coordinated cellular process in eukaryotes,[1–3] in which at least 11 eukaryotic initiation factors (eIFs) are included. The first step of translation initiation involves formation of a ternary complex (TC) of eIF2·GTP·Met – $tRNA_i^{Met}$. Such a ternary complex is able to interact with the 40 S ribosomal subunit, in the presence of eIF3 and eIF1A, to constitute a stable 43 S pre-initiation complex. Meanwhile, the mRNA molecule to be translated is prepared by eIF4F, eIF4A and eIF4B to remove unfavorable secondary structures. Subsequently, the active mRNA is loaded onto the 43 S complex and the 40 S ribosomal subunit starts scanning along the mRNA to search for the AUG start codon. The loading process is facilitated by eIF3, probably due to its simultaneous interactions with eIF2, mRNA and 40 S subunit. Once the anticodon of the Met – $tRNA_i^{Met}$ and the AUG codon correctly pair at the P site of the 40 S subunit, GTPase activity in eIF2 is triggered and the GTP in TC is converted to GDP. This conversion is stimulated by a GTPase-activating protein (GAP), eIF5 and results in an inactive GDP bound form of eIF2. The eIF2·GDP complex is then released and must be re-activated before entering the next cycle of translation initiation. The reactivation of eIF2 is accomplished by eIF2B, a guanine-nucleotide exchange factor that facilitates the exchange of GDP to GTP in eIF2.

As mentioned above, the assembly of the 43 S pre-initiation complex involves several important initiation factors. These initiation factors, including eIF2 (presenting as eIF2 TC), eIF3, eIF5 and eIF1, are all components of a stable intermediate complex named multifactor complex (MFC).[4] The MFC is not only a structural unit but also a functional moiety, as its integrity is required for the 43 S complex assembly[5–7] and the following selection for the AUG start codon.[4,8,9] In addition, the MFC plays an important role in TC recruitment during the re-initiation events.[10] Within the MFC, eIF2 is a trimeric protein composed of α, β, and γ subunits, eIF3 is a multisubunit complex while eIF1 and eIF5 are both composed of a single polypeptide. A recent investigation on the MFC revealed its minimal components, mini-MFC, which is constituted by the C-terminal domain of eIF5 (residues 201–405), the N-terminal domain (residues 1–140) containing three lysine-rich segments (K boxes) of the eIF2-β subunit (eIF2β), the N-terminal serine-rich acidic segment (residues 1–156) of the eIF3-c subunit

† Z.W. & Y.X. contributed equally to this work.

Abbreviations used: GAP, GTPase-activating protein; MFC, multifactor complex; CTD, C-terminal domain; NTD, N-terminal domain; IF, initiation factor; TC, ternary complex; AA boxes, aromatic/acidic residue-rich regions.

E-mail address of the corresponding author: wgong@ibp.ac.cn

(eIF3c), and eIF1.[11] The interactions between the above domains of the various eIFs have been identified and proposed to initiate the assembly of the entire MFC.[11] However, it is uncertain whether the MFC assembly is a common feature in eukaryotic translation initiation, because the evidence of the MFC existence has only recently been obtained in yeast.

eIF5, the GAP for eIF2, is ubiquitously expressed in yeast, plants, mammals and other eukaryotes. eIF5 binds directly to eIF2β, and stimulates GTP hydrolysis only in the eIF2 bound to 40 S ribosomal subunits.[12,13] Genetic and biochemical evidence demonstrates that eIF5 is divided into two domains with distinct functions, the N-terminal domain (eIF5-NTD) that is required for the eIF5-induced GTP hydrolysis [14,15] and the C-terminal domain (eIF5-CTD) that is responsible for interactions with other eIFs.[9,16,17] The eIF5-NTD is highly conserved between eukaryotic species[14] while the eIF5-CTD is only moderately conserved (this study), indicating slightly different interactions between eIF5 and other eIFs in different organisms.

It has been shown that eIF5-CTD plays a critical role in the MFC formation and acts as the core of pre-initiation complex assembly.[4,7,11] Analysis of the eIF5-CTD sequence revealed two aromatic/acidic residue-rich regions (AA boxes), which are also included in the C-terminal domains of eIF2B-ϵ and eIF4G.[18] The AA boxes in eIF5 and eIF2Bϵ are required for their interactions with the K boxes of eIF2β.[18] In addition to the interaction with eIF2β, eIF5-CTD also interacts with eIF3c and such interactions between these eIFs are cooperative. It is reported that the interaction between eIF5-CTD and eIF3c is greatly enhanced by the interaction between eIF5-CTD and the K boxes of eIF2β, suggesting that MFC formation is coordinately controlled during translation initiation.[11] In order to elucidate the structural basis of the binding activity of eIF5-CTD, we crystallized *Saccharomyces cerevisiae* eIF5-CTD and report the 1.5 Å structure. Sequence analysis and structural comparison indicate that eIF5-CTD and the C-terminal domains of eIF2Bϵ and eIF4G belong to an atypical HEAT motif. Our studies also confirm and suggest some potential regions of interactions between eIF5-CTD and other eIFs by analysis of electrostatic potential and conserved residue distribution on the protein surface, which is consistent with a previous research.[17]

Structure determination

The cDNA fragment encoding the eIF5 C-terminal domain (241–405) was sub-cloned from a *S. cerevisiae* cDNA library (a generous gift from Dr Congzhao Zhou, University of Science and Technology of China) into pGEX4T-1 expression vector (GE Healthcare) to express glutathione *S*-transferase (GST)-eIF5-CTD fusion protein. The eIF5-CTD protein was purified by glutathione affinity chromatography followed by thrombin cleavage and Superdex-G75 gel filtration.

The hanging drop vapor diffusion method was employed to obtain high quality crystals diffracting to 1.5 Å. The phase information was provided by a $KAu(CN)_2$ derivative, using the single-wavelength anomalous diffraction (SAD) method. The structure of eIF5-CTD was refined to an *R* factor of 17.7% and an R_{free} factor of 20.5% (Table 1).

Structure overview

eIF5-CTD was purified and crystallized as a recombinant protein, including residues Val241–Glu405 (numbered after full-length eIF5). The eIF5-CTD also contained some extra vector-encoded residues at both N (GSPEF, numbered as 236*–240*) and C (LERPHRD, numbered as 406*–412*) termini. The final model contained six eIF5-CTD protomers (named protomer A–F) per asymmetric unit and most residues were well modeled except for some at the termini. A disordered region has been observed at the C terminus of eIF-CTD, including Glu407*–Asp412* in each protomer as well as 396–404 in A, 395–405 in B, 396–402 in C, 395–401 in D, 397–402 in E, and 396–405 in F, respectively. Although Leu406* was observed in all six protomers, only five of them could be accurately modeled, since the electron density surrounding the backbone of Leu406* in B was too weak. At the N terminus of eIF5-CTD, both Gly236* and Ser237* are invisible in the electron density map.

General foldings are identical in the six eIF5-CTD protomers except for their N-terminal tails (241–245). It is not surprising that the N terminus of eIF5-CTD is flexible and able to adopt slightly different conformations in crystallographic packing, since these residues probably act as a hinge between eIF5 N-terminal and C-terminal domains. Within each eIF5-CTD protomer, the structure is essentially helical, containing ten α-helices (α1–α10) and one 3_{10}-helix (η1) with an exception in E (the fourth α-helix is replaced by a 3_{10}-helix). These helices form four antiparallel helical repeats (R_I–R_{IV}), which are packed against each other in a counterclockwise rotational manner. The angle between two adjacent repeats is 25∼35° (Figure 1(a)), calculated by the program interhlx†. This arrangement results in a left-handed twist in eIF5-CTD. Further examination revealed that each repeat of eIF5-CTD can be divided into an N-terminal part and a C-terminal part. For the N-terminal parts (R_IN–$R_{IV}N$), all of them are composed of an α-helix (Figures 1(a) and 2(a)). For the C-terminal parts (R_IC–$R_{IV}C$), R_IC consists of one α-helix (α2) while the others ($R_{II}C$–$R_{IV}C$) contain two helices (α4 and α5 in $R_{II}C$, η1 (3_{10}-helix) and α7 in $R_{III}C$, α9 and α10 in $R_{IV}C$) (Figure 2(a)).

† Yap, K. (1998). Interhlx—Interhelical angle program. University of Toronto.

Table 1. Statistics of data collection, SAD phasing and model refinement

A. *Data collection*			
Data sets	Native	Au-derivative	
Space group	*P*1		
Unit cell parameters	$a=44.85$ Å, $b=64.54$ Å, $c=108.92$ Å $\alpha=88.6°$, $\beta=86.6°$, $\gamma=74.9°$	$a=44.83$ Å, $b=64.73$ Å, $c=109.01$ Å $\alpha=88.7°$, $\beta=87.0°$, $\gamma=74.5°$	
Resolution range (Å)	50–1.5 (1.55–1.5)	50–2.1 (2.18–2.1)	
No. of total reflections	601,081	261,803	
No. of unique reflections	173,580 (12,895)	65,439 (6364)	
I/σ	25.4 (4.1)	30.3 (12.9)	
Completeness (%)	92.8 (68.9)	94.7 (92.0)	
R_{merge} (%)[a]	4.4 (24.5)	6.1 (12.9)	
B. *SAD phasing (20–2.1Å)*	Mean FOM[b] Mean FOM[b] after density modification	0.46 0.80	
C. *Structure refinement*			
Resolution (Å)	50–1.5 (1.54–1.5)		
R_{cryst}/R_{free} (%)[c]	17.7 (25.3)/20.5 (30.5)	RMSD bonds (Å)/angles (deg.)	0.010/1.2
No. of reflections		No. of atoms	
Working set	164,931	Protein atoms	8135
Test set	8647	Water molecules	916
		Sulfate ions	20[d]
Average *B*-factor (Å^2)		Ramachandran plot	
Main–chain	13.1/15.1/15.5/16.6/16.8/17.5[e]	Most favored regions (%)	93.7
Side-chain	14.9/16.9/17.4/17.4/18.0/18.7[e]	Additionally allowed (%)	6.1
Water	32.0	Generously allowed (%)	0.1

Crystallization of eIF5-CTD protein was achieved by hanging drop vapor diffusion method. 1 μl of such concentrated protein solution was then mixed with 1 μl of crystallization solution (2.8 M $(NH_4)_2SO_4$, 0.1 M Mes (pH 6.4)) to set up a hanging drop. High quality crystals were obtained within two days at 4 °C. To prepare heavy-atom derivatives, crystals were soaked in the crystallization solution containing 5 mM $KAu(CN)_2$ (Hampton Research) for four days. The diffraction data of the native crystal and its Au derivative were collected at the beamline 22-ID (SER-CAT) at the Advanced Photon Source (APS) using a MAR 300 CCD detector and at the home facility using a Rigaku R-AXIS IV^{++} imaging-plate system with a Rigaku FRE Cu rotating-anode generator, respectively. All data were collected at 100 K and processed with HKL2000.[29] SHELXD[30] was used to find initial ten Au sites. Fourteen additional sites were found by autoSHARP,[31] which also performed heavy-atom site refinement and phase calculation for the total 24 Au sites. Automatic model building by ARP/wARP[32] yielded an initial model with ~90% residues and ~70% side-chains placed correctly. The initial model containing six eIF5-CTD protomers was refined in Refmac5[33] against the 1.5 Å native data set. Iterative manual adjustments and rebuilding of the model was finished in Coot.[34] NCS restraints were applied at the beginning of the refinement and were released at the very last stage. After several cycles of positional and individual *B* factor refinement, water molecules and sulfate ions were added to the model. At the final step, TLS refinement[35] was performed in Refmac5 with 29 TLS groups as suggested by TLSMD.[36] The model was refined to an *R* factor of 17.7% and an R_{free} factor of 20.5%. The stereochemical quality of the final model was checked by PROCHECK.[37] The mean temperature factor for all atoms is 18.2 Å^2, which is consistent with the suggested value of 18.5 Å^2 by Wilson plot. Numbers in parentheses represent the value for the highest resolution shell.

[a] $R_{merge}=\sum|I_i-I_m|/\sum I_i$, where I_i is the intensity of the measured reflection and I_m is the mean intensity of all symmetry related reflections.

[b] Mean FOM (figure of merit)$=<|\Sigma P(\alpha)e^{i\alpha}/\Sigma P(\alpha)|$, where α is the phase and $P(\alpha)$ is the phase probability distribution.

[c] $R_{cryst}=\Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$, where F_{obs} and F_{calc} are observed and calculated structure factors. $R_{free}=\sum_T||F_{obs}|-|F_{calc}||/\sum_T|F_{obs}|$, where T is a test data set of about 5% of the total reflections randomly chosen and set aside prior to refinement.

[d] Number of sulfate ions.

[e] Values for six protomers (A/B/C/D/E/F) in one asymmetric unit.

Structure comparison with C-terminal domains of eIF2Bϵ and eIF4GI

It is predicted by SMART† and some other research based on computational analysis of sequences of proteins[19,20] that eIF5-CTD and the C-terminal domains of eIF2Bϵ (eIF2Bϵ-CTD) and eIF4G (eIF4G-CTD) all contain a similar fold, "eIF5C". Results of structural comparison of eIF5-CTD with yeast eIF2Bϵ-CTD (PDB id: 1PAQ) and human eIF4GI-CTD (PDB id: 1UG3) using DALI [21] confirmed the prediction. The DALI results showed that when compared to eIF5-CTD, eIF2Bϵ-CTD (residues 544–703) and the second HEAT domain of eIF4GI-CTD (eIF4GI-CTD-H2, residues 1439–1566) had *Z* scores of 12.1 (136 C^α atoms aligned, with RMSD of 2.8Å) and 11.7 (115 C^α atoms aligned, with RMSD of 2.1Å), respectively. The eIF2Bϵ-CTD structure contains the four antiparallel repeats similar to eIF5-CTD, while eIF4GI-CTD-H2 only contains the first three repeats (R_I–R_{III}) and $R_{IV}N$. However, the full-length eIF4GI-CTD-H2 domain (residues 1439–1600) structure should include a complete R_{IV}, as indicated by sequence alignment (Figure 2(a)). Consistent with structural similarity, the three proteins all contain two conserved AA boxes. The AA boxes are composed of aromatic and acidic residues, which are highly conserved in eIF5, eIF2Bϵ and eIF4G families (Figure 2(a)).

† Simple Modular Architecture Research Tool, http://smart.embl.de/

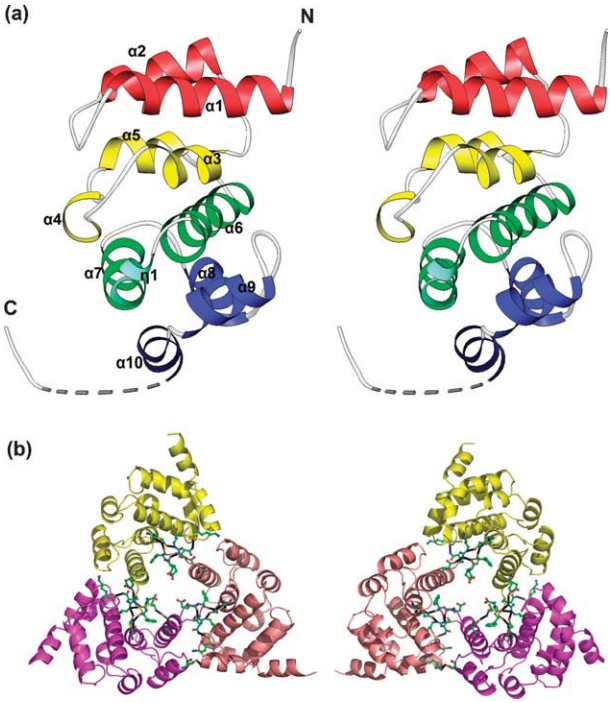

Figure 1. (a) Stereo view of the overall structure of the eIF5-CTD protomer. The four antiparallel helical repeats (R_I–R_{IV}) are colored in red, yellow, green, and blue, respectively. The 3_{10}-helix η1 is colored in light green. The secondary structural elements are labeled in the left diagram. The disordered region is indicated by a broken line. No extra vector-encoded residues are included. See the text for details. (b) Diagrams of the eIF5-CTD trimer. The trimeric eIF5-CTD structure is constituted by protomers A, B, and C colored in yellow, purple, and salmon, respectively. The residues forming hydrogen bonds in the interfaces are presented in sticks and balls while the hydrogen bonds are shown as broken lines. The right diagram is a vertically inverted view of the left.

Although similar in overall foldings, the three domain structures described above display some obvious structural deviations. The major difference is located in the C-terminal regions of R_{II}–R_{IV}, showing different bending angles (calculated by the program HELANAL[22]). In eIF2Bϵ-CTD and eIF4GI-CTD-H2, $R_{III}C$ is a straight α-helix while $R_{II}C$ and $R_{IV}C$ (not available in the present eIF4GI-CTD-H2 structure) are kinked. The eIF2Bϵ-CTD $R_{II}C$ and $R_{IV}C$ are kinked by ∼45° and ∼55°, respectively, while the eIF4GI-CTD-H2 $R_{II}C$ is kinked by ∼54° (Figure 2(b)). However, in eIF5-CTD, $R_{II}C$–$R_{IV}C$ are all bent, among which $R_{III}C$ is kinked by ∼54° while $R_{II}C$ and $R_{IV}C$ are kinked by ∼87° and ∼56°, respectively. Although the $R_{II}C$ and $R_{IV}C$ in the three domain structures show different kinking angles, they do bend at nearly the same positions. These results indicate that $R_{II}C$–$R_{IV}C$ in eIF5-CTD are bent more heavily than those in the other two structures so that each repeat is separated into two helices in eIF5-CTD. The structural alignment shows that the R_{II} fragments in eIF2Bϵ-CTD and eIF4GI-CTD-H2 are longer than that in eIF5-CTD. Such differences are reflected in structures as a longer $R_{II}N$ α-helix in eIF2Bϵ-CTD and an additional β-hairpin connecting $R_{II}N$ and $R_{II}C$ in eIF4GI-CTD-H2 (Figure 2). The additional β-hairpin, on which the residues are not highly conserved in eIF4G (data not shown), is involved in the interface formation between the two HEAT domains in the eIF4GI-CTD structure. Since eIF5-CTD has been proved to bind to a HEAT domain of eIF4G,[23] it is possible that the interaction between the eIF5-CTD and eIF4G HEAT domain is similar to that between two eIF4G HEAT domains. Therefore, the loops connecting $R_{II}N$-$R_{II}C$ and $R_{III}N$-$R_{III}C$ might be involved in eIF5-CTD binding to eIF4G.

According to Boesen *et al.*[24] R_I is more mobile than the other repeats in eIF2Bϵ-CTD because of significantly higher temperature factors. However, this phenomenon is not observed in the eIF5-CTD and eIF4GI-CTD-H2 structures. In eIF5-CTD protomer A, the average *B* factors are 13.48 Å^2 for all atoms and 14.57 Å^2 for all atoms in R_I (residues 241–280). Similarly, the values are 29.36 Å^2 for all atoms in eIF4GI-CTD-H2 and 29.42 Å^2 for all atoms in its R_I (residues 1438–1470). Therefore, due to the compact association between R_I and the other three repeats, it would be more suitable to define the four repeats as one functional domain instead of two. Although similar to HEAT repeat, eIF5-CTD as well as eIF2Bϵ-CTD and eIF4GI-CTD-H2 show a few unique structural features different from canonical HEAT repeats. For instance, several helices in eIF5-CTD ($R_{II}C$–$R_{IV}C$) and eIF2Bϵ-CTD ($R_{II}C$ and $R_{IV}C$) are bent so heavily that they are separated into two helices, in contrast to most helices in other HEAT motifs that are kinked by up to ∼45° without separation.[25] In addition, searching in PDB database based on eIF5-CTD (performed by DALI), did not result in any other similar structures with *Z* scores larger than 7, much lower than those between eIF5-CTD, eIF2Bϵ-CTD and eIF4GI (>11). Combined with the unique conserved AA boxes, it suggests that eIF5-CTD, eIF2Bϵ-CTD and eIF4GI-CTD-H2 form a novel fold class as an atypical HEAT motif. Some other atypical HEAT motifs were found in the HAM (HEAT analog motif) domain of human eIF3k26 and the middle portion of human eIF4GII.[27] The eIF5-CTD structure again supports the idea[24] that such atypical HEAT motifs are common among eIFs and important for protein–protein interactions.

Surface properties of eIF5-CTD

Analysis of the electrostatic potential distribution, performed by PISA server,[28] shows that eIF5-CTD has a bipartite surface, on which acidic and basic residues form distinct clusters (Figure 3). Therefore, the charged surface of eIF5-CTD can be divided into three areas, one positive and two negative. Intriguingly, analysis of the primary sequence of eIF5-CTD demonstrates conserved residues within these surface areas. The first area (area I) contains highly conserved acidic residues

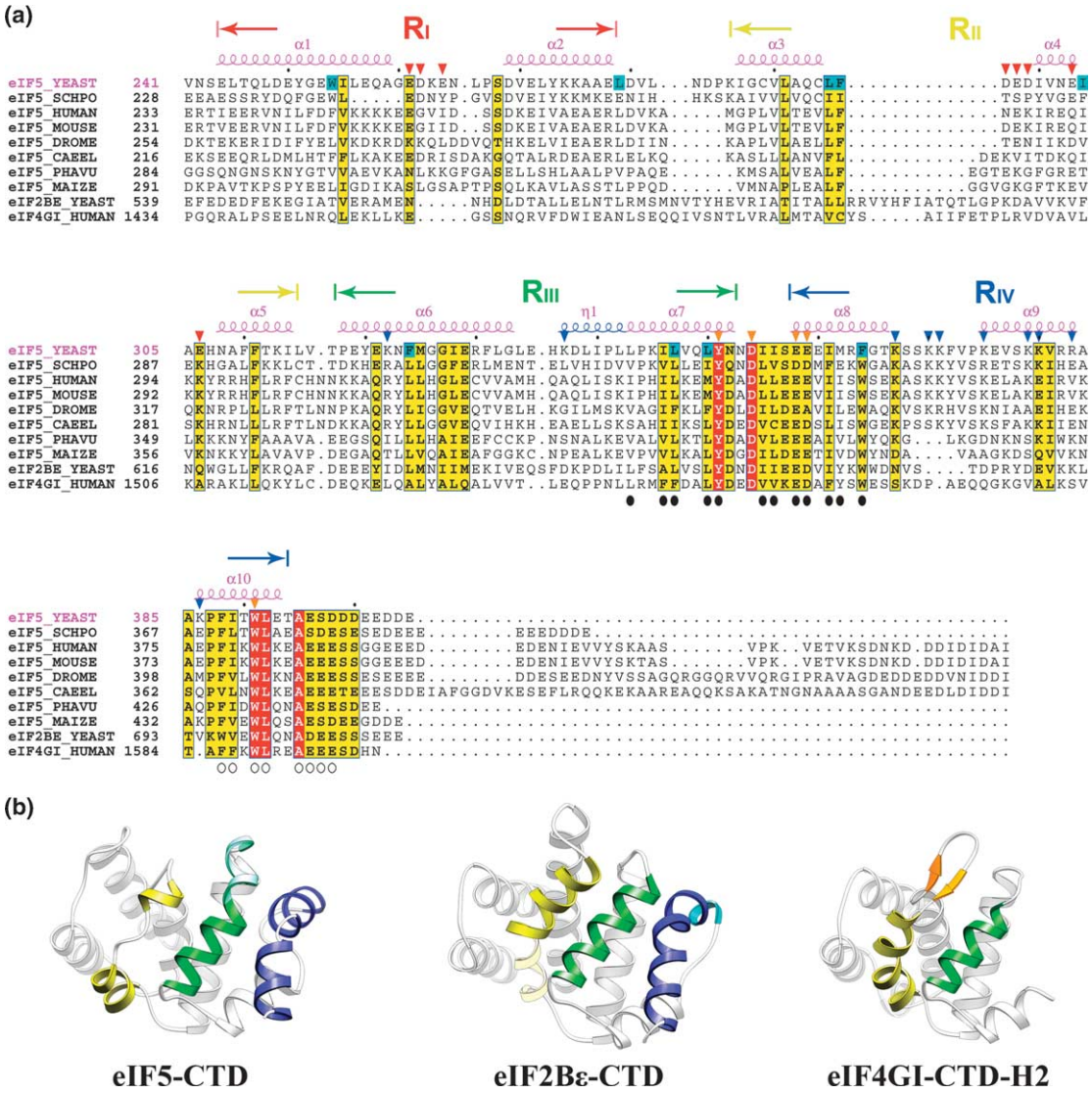

Figure 2. (a) Multiple-sequence alignment of the C termini of eIF5 homologues, eIF2Bϵ-CTD and eIF4GI-CTD-H2. The eIF5 sequences of *Saccharomyces cerevisiae* (Baker's yeast, Swiss-Prot accession number: P38431), *Schizosaccharomyces pombe* (fission yeast, Q09689), *Homo sapiens* (human, P55010), *Mus musculus* (mouse, P59325), *Drosophila melanogaster* (fruit fly, Q9VXK6), *Caenorhabditis elegans* (Q22918), *Phaseolus vulgaris* (kidney bean, P48724), *Zea mays* (maize, P55876), are aligned using Clustal W.[38] Yeast eIF2Bϵ and human eIF4GI are aligned with yeast eIF5-CTD by DALI.[21] The secondary structures of eIF5-CTD, which is defined by the DSSP program,[39] are indicated above the alignment diagram. The four repeats (R_I–R_{IV}) are labeled above the secondary structure in the colors corresponding to Figure 1(a). Identical and chemically similar residues in the alignment are boxed in red and yellow, respectively. Filled and open circles denote residues in AA box 1 and 2, respectively. The residues constituting potential binding sites for other eIFs on area I, II and III (see the text and Figure 3) are indicated by orange, blue and red, respectively. Cyan boxes highlighted in the sequence of yeast eIF5-CTD correspond to the residues that are mutated in a temperature-sensitive (Ts^-) phenotype.[40] It is worth noticing that these nine hydrophobic residues are involved in the formation of the eIF5-CTD hydrophobic core, confirming the previous prediction.[17] (b) Structural comparison of eIF5-CTD, eIF2Bϵ-CTD and eIF4GI-CTD-H2. eIF5-CTD and eIF2Bϵ-CTD contain four repeats while eIF4GI-CTD-H2 contains only three and a half repeats. Residues after Glu396 in the eIF5-CTD are not included in the comparison because they are for the most part disordered. In each of the three structures, the R_{II}C–R_{IV}C repeats are shown as front view and labeled in yellow, green and blue, respectively. The light colored region in these repeats indicates the 3_{10}-helices. The β hairpin of eIF4GI-CTD-H2 is shown in orange.

(Asp354, Glu358 and Glu359) and aromatic residues (Tyr351 and Trp391), most of which are the residues of the two AA boxes (Figure 2(a)). The second area (area II) includes some lysine residues (Lys322, Lys367, Lys370, Lys371, Lys375 and Lys379), which are conserved in eIF5 homologues from yeast, insects and mammals but not conserved in those from plants (Figure 2(a)). In addition to these conserved lysine residues, some non-conserved basic residues (Lys337, Lys380, Arg383, and Lys386) also contribute to the positive charge in area II. The third area (area III) is highly negatively charged and resides on the opposite side of eIF5-CTD to area II (Figure 3). The acidic residues in area

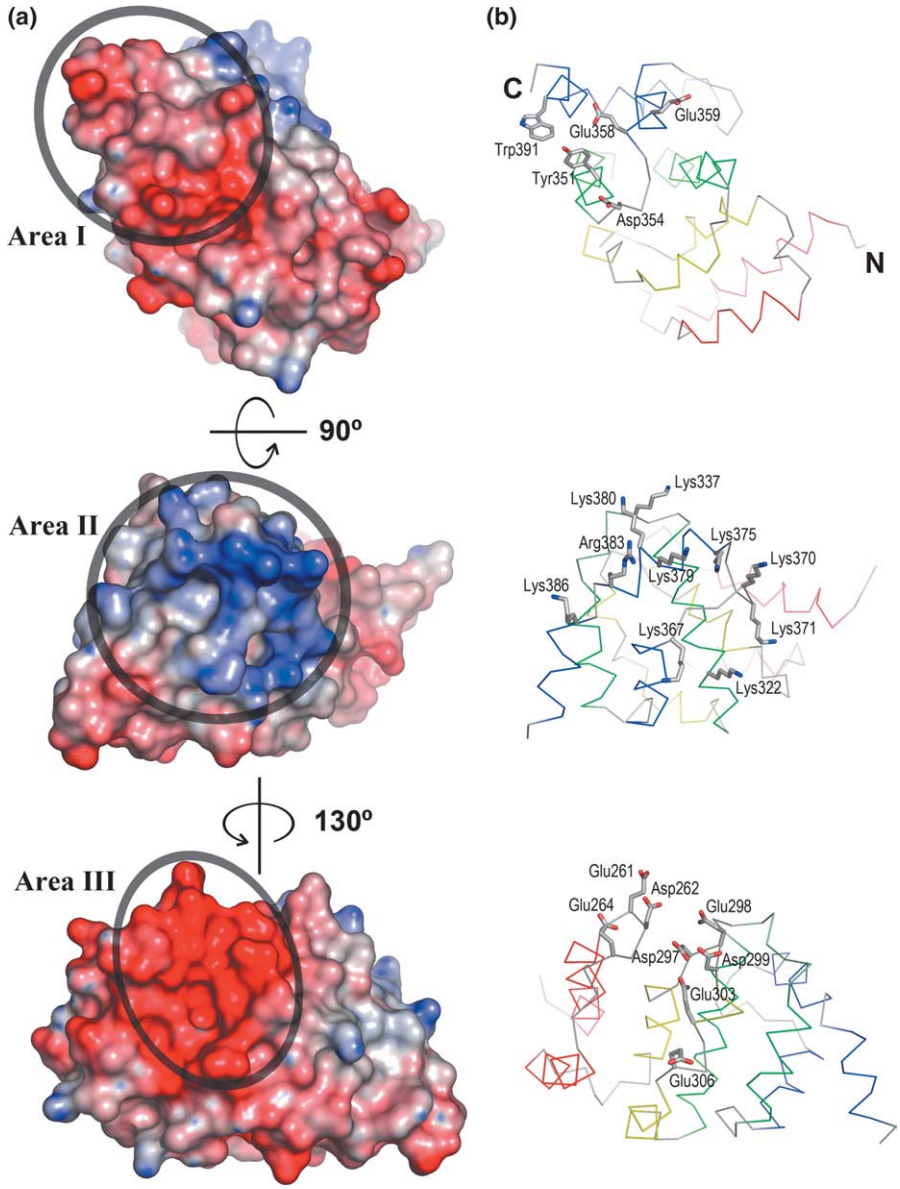

Figure 3. (a) Electrostatic potential distribution on the surface of eIF5-CTD in different views. The surface is colored red for negative potential and blue for positive potential. Areas I, II and III are circled. (b) The CA model of eIF5-CTD with the same views of (a). The four repeats are colored as in Figure 1(a). The residues (marked in Figure 2(a) by triangles) on areas I, II and III are shown with their side-chains in top, middle and bottom diagrams, respectively.

III are located in two clusters on the eIF5-CTD primary sequence. One is the loop connecting α1 and α2 (Glu261, Asp262 and Glu264), while the other contains residues in and around α4 (Asp297, Glu298, Glu299, Glu303, and Glu306). Although these acidic residues are non-conserved, abundant acidic residues in the regions described above are obvious in the other eIF5 homologues except for the plant eIF5′s, which lack acidic residues in the loop connecting α1 and α2 (Figure 2(a)). The residue conservation within the charged areas suggests that all but plants eIF5 homologues share a common surface electrostatic potential distribution and thus might employ a similar mechanism in binding to other protein factors.

Interactions between the eIF5-CTD protomers

Interactions between eIF5-CTD protomers are observed in crystal packing. In the asymmetric unit, the six protomers form two eIF5-CTD trimers of high similarity, while the protomers within each trimer are related by a non-crystallographic 3-fold axis (Figure 1(b)). The two trimers are related by a non-crystallographic translation of about $\Delta x = 1/2$, $\Delta y = 1/2$, $\Delta z = 1/2$, which has been confirmed by the fact that the crystal seems to belong to $R3$ space group at low resolution. The trimer interfaces are formed between the C terminus of one protomer and the $R_I N$–$R_{III} N$ (i.e. α1, α3, and α6) as well as the loop connecting $R_{IV} N$ and $R_{IV} C$ of the adjacent one

(Figure 1(b)), with a buried area of 573Å^2–657 Å^2 calculated by the PISA server.[28] The PISA results also demonstrate that the interfaces are mainly stabilized by hydrogen bonds (Figure 1(b)). Meanwhile Phe240* accounts for some hydrophobic interactions in the interfaces. The trimerization of eIF5-CTD is not a result of selective crystal packing, since such a trimer does exist in solution as confirmed by analytical gel filtration (data not shown). However, no oligomerization of eIF5 has been reported before and the trimerization involves several vector-encoded residues, such as Glu239* and Phe240*. Therefore, whether full-length eIF5 forms a trimer and whether the eIF5-CTD trimer is biologically relevant requires further investigation.

New insights of the binding surfaces on eIF5

The MFC assembly involves interactions between various eIFs. Our study clearly demonstrates three areas on eIF5-CTD as potential binding sites for other eIFs. As the AA boxes of eIF5-CTD are required for binding to the lysine-rich segments (K boxes) of eIF2β-NTD,[18] the area I of eIF5-CTD that contains all exposed residues of the AA boxes is believed to be involved in such interactions. Recently, some binding surfaces of eIF5-CTD for other eIFs were characterized by surface residue mutagenesis based on a homology-modeled structure.[17] Using *in vitro* (GST pull-down) assays, area I together with an acidic residue-rich fragment of eIF5-CTD C terminus (mostly disordered in the eIF5-CTD structure) was identified as a potential binding site for the K boxes of eIF2β-NTD while area II was identified as a binding site for eIF3c and eIF1. In general, our electrostatic potential distribution analysis of the eIF5-CTD crystal structure is consistent with these observations. Nevertheless, careful examination of the surface residues in the eIF5-CTD structure revealed some contradictory aspects. According to Yamamoto *et al.*[17] the acidic residue Glu360 and the basic residue Arg382 are modeled to interact with the K boxes on area I and with eIF3c-NTD on area II, respectively. However, in the eIF5-CTD structure, Glu360 has a very small solvent accessible area of 8.33 Å^2 and forms two salt bridges with Arg363, using both carboxyl oxygen atoms. Thus, it is unlikely that Glu360 directly interacts with the K box residues. Likewise, Arg382 seems not to be exposed enough and also forms two salt bridges with another buried residue Glu329. Because Glu329 is strictly conserved in all eIF5 homologues, it is very possible that Arg382 is involved in the stabilization of the eIF5-CTD structure rather than the interaction with eIF3c or eIF1. Our eIF5-CTD structure also reveals some new candidates for the interactions between eIF factors. Some basic residues (Lys322, Lys380, Arg383, and Lys386) contribute to the positive surface charge in area II and are possibly involved in the interactions with eIF3c and/or eIF1, although most of these residues (Lys380, Arg383, and Lys386) are not conserved.

In addition to the previously characterized areas I and II, we found a highly negatively charged surface, area III, which is expected to exist in all eIF5 homologues except for plant eIF5s (Figure 2(a)). Area III is composed of the acidic residues in the loop connecting α1 and α2, and the acidic residues in and around α4. Because of its negative charges area III may serve as another binding site for the K boxes. As being ~20Å apart, area III and area I plus the C terminus are probably responsible for binding to different K boxes. Another K box binding protein is eIF2Bϵ-CTD, which contains fewer acidic residues on the corresponding region compared to area III of eIF5-CTD. This indicates that eIF2Bϵ-CTD and eIF5-CTD might bind to eIF2β-NTD in different ways and/or with different affinities. Plant eIF5s might possess a different mechanism of binding to its cognate eIF partners, since the predicted electrostatic potential distributions on area II and area III of plant eIF5-CTD are quite different from other eIF5 homologues. Whether these hypotheses based on the structure analysis are correct or not remains to be investigated.

Protein Data Bank accession code

The coordinates and structure factors have been deposited in the RCSB Protein Data Bank with accession code 2FUL.

Acknowledgements

This work is supported by National Foundation of Talent Youth (grant no. 30225015), 973 programs (no. 2004CB520801), National Natural Science Foundation of China (grant no.10490913) and Chinese Academy of Sciences (KSCX2-SW-224). We appreciate Professor Zhijie Liu for his generous help in data collection at the beamline 22-ID, Southeast Regional Collaborative Access Team (SER-CAT), Advanced Photon Source, Argonne National Laboratory. 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. We also thank Professor Congzhao Zhou for providing the yeast cDNA library.

References

1. Merrick, W. C. (2003). Initiation of protein biosynthesis in eukaryotes. *Biochem. Mol. Biol. Ed.* **31**, 378–385.
2. Hershey, J. W. B. & Merrick, W. C. (2000). The pathway and mechanism of initiation of protein synthesis. In *Translational Control of Gene Expression. Translational Control of Gene Expression* (Sonenberg, N., Hershey, J. W. B. & Mathews, M. B., eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

3. Hinnebusch, A. G. (2000). Mechanism and regulation of initiator methionyl-tRNA binding to ribosomes. In *Translational Control of Gene Expression* (Sonenberg, N., Hershey, J. W. B. & Mathews, M. B., eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
4. Asano, K., Clayton, J., Shalev, A. & Hinnebusch, A. G. (2000). A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate *in vivo*. *Genes Dev.* **14**, 2534–2546.
5. Phan, L., Schoenfeld, L. W., Valasek, L., Nielsen, K. H. & Hinnebusch, A. G. (2001). A subcomplex of three eIF3 subunits binds eIF1 and eIF5 and stimulates ribosome binding of mRNA and tRNA(i)Met. *EMBO J.* **20**, 2954–2965.
6. Valasek, L., Nielsen, K. H. & Hinnebusch, A. G. (2002). Direct eIF2-eIF3 contact in the multifactor complex is important for translation initiation *in vivo*. *EMBO J.* **21**, 5886–5898.
7. Valasek, L., Mathew, A. A., Shin, B. S., Nielsen, K. H., Szamecz, B. & Hinnebusch, A. G. (2003). The yeast eIF3 subunits TIF32/a, NIP1/c, and eIF5 make critical connections with the 40 S ribosome *in vivo*. *Genes Dev.* **17**, 786–799.
8. Huang, H. K., Yoon, H., Hannig, E. M. & Donahue, T. F. (1997). GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev.* **11**, 2396–2413.
9. Asano, K., Shalev, A., Phan, L., Nielsen, K., Clayton, J., Valasek, L. *et al*. (2001). Multiple roles for the C-terminal domain of eIF5 in translation initiation complex assembly and GTPase activation. *EMBO J.* **20**, 2326–2337.
10. Valasek, L., Nielsen, K. H., Zhang, F., Fekete, C. A. & Hinnebusch, A. G. (2004). Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection. *Mol. Cell. Biol.* **24**, 9437–9455.
11. Singh, C. R., Yamamoto, Y. & Asano, K. (2004). Physical association of eukaryotic initiation factor (eIF) 5 carboxyl-terminal domain with the lysine-rich eIF2beta segment strongly enhances its binding to eIF3. *J. Biol. Chem.* **279**, 49644–49655.
12. Chaudhuri, J., Das, K. & Maitra, U. (1994). Purification and characterization of bacterially expressed mammalian translation initiation factor 5 (eIF-5): demonstration that eIF-5 forms a specific complex with eIF-2. *Biochemistry*, **33**, 4794–4799.
13. Das, S., Maiti, T., Das, K. & Maitra, U. (1997). Specific interaction of eukaryotic translation initiation factor 5 (eIF5) with the beta-subunit of eIF2. *J. Biol. Chem.* **272**, 31712–31718.
14. Das, S., Ghosh, R. & Maitra, U. (2001). Eukaryotic translation initiation factor 5 functions as a GTPase-activating protein. *J. Biol. Chem.* **276**, 6720–6726.
15. Paulin, F. E., Campbell, L. E., O'Brien, K., Loughlin, J. & Proud, C. G. (2001). Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Curr. Biol.* **11**, 55–59.
16. Das, S. & Maitra, U. (2000). Mutational analysis of mammalian translation initiation factor 5 (eIF5): role of interaction between the beta subunit of eIF2 and eIF5 in eIF5 function *in vitro* and *in vivo*. *Mol. Cell. Biol.* **20**, 3942–3950.
17. Yamamoto, Y., Singh, C. R., Marintchev, A., Hall, N. S., Hannig, E. M., Wagner, G. & Asano, K. (2005). The eukaryotic initiation factor (eIF) 5 HEAT domain mediates multifactor assembly and scanning with distinct interfaces to eIF1, eIF2, eIF3, and eIF4G. *Proc. Natl Acad. Sci. USA*, **102**, 16164–16169.
18. Asano, K., Krishnamoorthy, T., Phan, L., Pavitt, G. D. & Hinnebusch, A. G. (1999). Conserved bipartite motifs in yeast eIF5 and eIF2Bepsilon, GTPase-activating and GDP-GTP exchange factors in translation initiation, mediate binding to their common substrate eIF2. *EMBO J.* **18**, 1673–1688.
19. Koonin, E. V. (1995). Multidomain organization of eukaryotic guanine nucleotide exchange translation initiation factor eIF-2B subunits revealed by analysis of conserved sequence motifs. *Protein Sci.* **4**, 1608–1617.
20. Aravind, L. & Koonin, E. V. (2000). Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. *Genome Res.* **10**, 1172–1184.
21. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
22. Bansal, M., Kumar, S. & Velavan, R. (2000). HELANAL: a program to characterize helix geometry in proteins. *J. Biomol. Struct. Dynam.* **17**, 811–819.
23. He, H., von der Haar, T., Singh, C. R., Ii, M., Li, B., Hinnebusch, A. G., McCarthy, J. E. & Asano, K. (2003). The yeast eukaryotic initiation factor 4G (eIF4G) HEAT domain interacts with eIF1 and eIF5 and is involved in stringent AUG selection. *Mol. Cell. Biol.* **23**, 5431–5445.
24. Boesen, T., Mohammad, S. S., Pavitt, G. D. & Andersen, G. R. (2004). Structure of the catalytic fragment of translation initiation factor 2B and identification of a critically important catalytic residue. *J. Biol. Chem.* **279**, 10584–10592.
25. Kobe, B., Gleichmann, T., Horne, J., Jennings, I. G., Scotney, P. D. & Teh, T. (1999). Turn up the HEAT. *Struct. Fold. Des.* **7**, R91–R97.
26. Wei, Z., Zhang, P., Zhou, Z., Cheng, Z., Wan, M. & Gong, W. (2004). Crystal structure of human eIF3k, the first structure of eIF3 subunits. *J. Biol. Chem.* **279**, 34983–34990.
27. Marcotrigiano, J., Lomakin, I. B., Sonenberg, N., Pestova, T. V., Hellen, C. U. & Burley, S. K. (2001). A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol. Cell*, **7**, 193–203.
28. Krissinel, E. & Henrick, K. (2005). Detection of Protein Assemblies in Crystals. In *CompLife 2005* (Berthold, M. R., Glen, R., Diederichs, K., Kohlbacher, O. & Fischer, I., eds), LNBI 3695, pp. 163–174, Springer.
29. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
30. Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallog. sect. D*, **58**, 1772–1779.
31. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2005). Automated structure solution with autoSHARP. In *Crystallographic Methods* (Doublie, S., ed.), Humana Press, Totowa, NJ.
32. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
33. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.

34. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. sect. D*, **60**, 2126–2132.
35. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321.
36. Painter, J. & Merritt, E. A. (2005). A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. *Acta Crystallog. sect. D*, **61**, 465–471.
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
38. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res*, **22**, 4673–4680.
39. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
40. Singh, C. R., Curtis, C., Yamamoto, Y., Hall, N. S., Kruse, D. S., He, H. *et al*. (2005). Eukaryotic translation initiation factor 5 is critical for integrity of the scanning preinitiation complex and accurate control of GCN4 translation. *Mol. Cell. Biol.* **25**, 5480–5491.

Edited by J. Doudna

(Received 1 February 2006; received in revised form 8 March 2006; accepted 15 March 2006)
Available online 31 March 2006