THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 33, Issue of August 13, pp. 34983–34990, 2004
Printed in U.S.A.

Crystal Structure of Human eIF3k, the First Structure of eIF3 Subunits*

Received for publication, May 10, 2004
Published, JBC Papers in Press, June 4, 2004, DOI 10.1074/jbc.M405158200

Zhiyi Wei‡§¶, Ping Zhang‡§¶, Zhaocai Zhou‡§, Zhongjun Cheng‡§, Mao Wan‡§, and Weimin Gong‡§‖

From the ‡National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China and the §School of Life Sciences, Key Laboratory of Structural Biology, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China

eIF3k, the smallest subunit of eukaryotic initiation factor 3 (eIF3), interacts with several other subunits of eIF3 and the 40 S ribosomal subunit. eIF3k is conserved among high eukaryotes, including mammals, insects, and plants, and it is ubiquitously expressed in human tissues. Interestingly, eIF3k does not exist in some species of yeast. Thus, eIF3k may play a unique regulatory role in higher organisms. Here we report the crystal structure of human eIF3k, the first high-resolution structure of an eIF3 component. This novel structure contains two distinct domains, a HEAT (named for Huntington, elongation factor 3, A subunit of protein phosphatase 2A, target of rapamycin) repeat-like HAM (HEAT analogous motif) domain and a winged-helix-like WH domain. Through structural comparison and sequence conservation analysis, we show that eIF3k has three putative protein-binding surfaces and has potential RNA binding activity. The structure provides key information for understanding the structure and function of the eIF3 complex.

Translation initiation is a sophisticated cellular process, especially in eukaryotes. In general, translation initiation in eukaryotic organisms involves three steps (1); first, the methionyl-initiator tRNA (Met-tRNA$_i^{Met}$) binds to the 40 S ribosomal subunit to form a 43 S preinitiation complex; second, the preinitiation complex binds to mRNA and scans to the AUG start codon in the mRNA; and third, the 60 S ribosomal subunit joins the mRNA-bound preinitiation complex to form an 80 S initiation complex, ready to commence translation. Each of these steps is stimulated by a number of proteins called eukaryotic initiation factors (eIFs).[1] At least 11 eIFs have been identified, comprising over 25 polypeptides (2). In contrast, only three to five initiation factors are known in prokaryotes. This difference in protein complexity suggests that more protein-RNA and protein-protein interactions rather than RNA-RNA interactions are required for efficient translation initiation in eukaryotic cells.

In mammalian cells, eIF3 is the largest initiation factor with an apparent molecular mass of about 600 kDa. It plays a central role in steps 1 and 2 of the translation initiation process (1, 3). For instance, eIF3 can bind to dissociated 40 S subunits and delay the reassociation with the 60 S ribosomal subunit for a long enough time to permit initiation. eIF3 also stabilizes the binding of the Met-tRNA$_i^{Met}$·eIF2·GTP ternary complex to the 40 S subunits and promotes the formation of a 43 S preinitiation complex comprised of the 40 S subunit, the ternary complex, eIF1, eIF1A, and eIF3. In addition, eIF3 stimulates the binding of 5′-m^{7}G-capped mRNA by interaction with the mRNA-associated factor eIF4G.

eIF3 is a multisubunit protein complex. Various genes encoding eIF3 subunits have been cloned from mammals, plants, and yeasts. Twelve different subunits (eIF3a/p170, b/p116, c/p110, d/p66, e/p48, f/p47, g/p44, h/p40, i/p36, j/p45, k/p28, l/p69) have been identified in mammals, while in yeast only six subunits (eIF3a/TIF32, b/PRT1, c/NIP1, i/TIF34, g/TIF35, j/HCR1) have been found and are all homologous to the corresponding mammalian subunits. The five yeast subunits, eIF3a, eIF3b, eIF3c, eIF3g, and eIF3I, can form a core complex (4). Mammalian and yeast eIF3s differ not only in the number of subunits but also in the structure of some subunits. For example, mammalian eIF3a contains a repeat region, but this region is absent in the yeast ortholog (2). These differences suggest that mammalian and plant eIF3 have evolved to include additional subunits, which are likely to function as regulatory factors and the extra structural motifs provide the capacity to mediate extra protein-protein or protein-RNA interactions required for the tighter regulation in higher eukaryotes. Some eIF3 subunits (eIF3a, eIF3c, eIF3e, eIF3f, and eIF3h) may serve as structural scaffolds or to provide docking sites for other proteins, since they contain PCI or MPN domains, which are found in components of large protein complexes and have been implicated in protein-protein interactions (2, 5).

To understand the structure and function of the eIF3 complex in eukaryotic translation initiation, we have begun a systematic structural study of the eIF3 components. Here we describe the high-resolution crystal structure of human eIF3k, the smallest non-core subunit of eIF3. Mammalian eIF3k has recently been characterized (6) to co-express with the five core subunits of eIF3 and form a stable co-immunoprecipitatable complex with the core complex. eIF3k also interacts directly

* This work was supported by the Foundation for Authors of National Excellent Doctoral Dissertation of the People's Republic of China (Project No. 200128), National Foundation of Talent Youth (Grant No. 30225015), the National High Technology Research and Development Program of China (Grant No. 2001AA233021), the 863 Special Program of China (Grant No. 2002BA711A13), the Key Important Project and other projects from the National Natural Science Foundation of China (Grant Nos. 30121001, 30070170, 30130080, and 30121001), and the Chinese Academy of Sciences (Grant No. KSCX1-SW-17). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1RZ4) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ These authors contributed equally to this work.

‖ To whom the correspondence should be addressed. E-mail: wgong@sun5.ibp.ac.cn.

[1] The abbreviations used are: eIF, eukaryotic initiation factor; HTH, helix-turn-helix; HEAT, Huntington, elongation factor 3, A subunit of protein phosphatase 2A, target of rapamycin; HAM, HEAT analogous motif; WH, winged-helix; PHAT, pseudo-HEAT analogous topology; PDB, Protein Data Bank.

TABLE I
Summary of data collection and processing
Numbers in parentheses represent the value for the highest resolution shell. MAD, multiwavelength anomalous dispersion.

MAD data collection	Edge	Peak	Remote
Wavelength (Å)	0.97992	0.97969	0.9
Resolution range (Å)	30–2.1 (2.15–2.1)	30–2.1 (2.15–2.1)	30–2.3 (2.35–2.3)
No. of total reflections	116,044	218,413	189,782
No. of unique reflections	12,780 (828)	12,597 (822)	9,690 (638)
I/σ	12.0 (3.7)	21.7 (9.5)	16.6 (7.3)
Completeness (%)	98.7 (97.6)	100 (100)	100 (99.8)
R_{merge}	0.080 (0.300)	0.079 (0.211)	0.108 (0.269)

with eIF3c, eIF3g, and eIF3j. The crystal structure of eIF3k reveals a novel ear-like protein structure containing two domains. Structural comparisons and sequence conservation analysis suggest that eIF3k is likely to act as a structural scaffold for protein-protein and protein-RNA interactions. It has three putative protein-binding regions and has potential RNA binding activity associated with its HTH (helix-turn-helix) motif.

EXPERIMENTAL PROCEDURES

Expression and Purification of Human eIF3k—The complete cDNA fragment encoding human eIF3k protein was subcloned from a human brain cDNA library into a pET-22b expression vector, and human eIF3k was expressed highly as a soluble protein in *Escherichia coli* strain BL21 (DE3) with a 6-residue histidine tag fused to its C terminus. Purification of the eIF3k protein was carried out through affinity chromatography with a Chelating Sepharose™ Fast Flow (Amersham Biosciences). For phase determination, the recombinant plasmid was transferred into Met-auxotrophic strain B834 to obtain the selenomethionyl derivative of human eIF3k protein.

Crystallization and X-ray Data Collection—Crystals of seleno-methionine-substituted eIF3k protein were grown using the hanging drop vapor diffusion method at 4 °C. The initial crystallization conditions (2.0 M ammonium sulfate, 0.1 M Hepes at pH 7.5, 0.1 M NaCl) were obtained using Crystal Screen kits I and II from Hampton Research. By optimizing the concentration of ammonium sulfate to 1.6 M, crystals of larger size and better quality were obtained. A multiwavelength anomalous dispersion data set was collected from a single seleno-methionine-substituted eIF3k crystal at 100 K on beamline 3W1A of the Beijing Synchrotron Radiation Facility at the Institute of High Energy Physics, Chinese Academy of Sciences. The data were collected at three wavelengths (λ_{edge} = 0.97992 Å, λ_{peak} = 0.97969 Å, and λ_{remote} = 0.9 Å). All data were processed and scaled with the DENZO and SCALEPACK (7), respectively. Data collection statistics are presented in Table I.

Phasing, Model Building, and Refinement—Six of seven expected selenium positions were found by SOLVE (8) using the three data sets, and the initial phase was calculated to 2.5 Å. RESOLVE (9, 10) was used for density modification and building of the initial model of human eIF3k. The initial model containing about 75% of the residues was refined against the peak data set in the 30–2.5-Å resolution range with maximum-likehood amplitude targets by using the Crystallography and NMR System (11). Subsequently, the refinement was extended to resolution bins of 2.3 Å with iterative manual adjustments and rebuilding of the model using the program O (12) and $2F_o - F_c$ and $F_o - F_c$ electron density maps as references. A sulfate ion and water molecules were added to the model when the resolution extended to 2.1 Å, and the value of R_{free} is about 30%. After that, individual atomic B factors were refined. Finally, the model was checked for errors with simulated annealing omit maps covering a 10-residue segment of the structure at a time. The stereochemical quality of the final model was checked by PROCHECK (13), and the final refinement statistics and geometry are excellent (Table II).

Sequence and Structure Analysis—Seven eIF3k-related sequences were identified in the sequence data base using BLAST (14); multiple sequences alignment was done with T-coffee (15); the similarity score of residues were calculated using Blosum62 matrix by ESPript (16); structural similarity searches were performed with DALI (17); and electrostatic surface potentials were calculated with MOLMOL (18). Figs. 1, 3, 4, 5, and 7 were prepared using Ribbons (19); Fig. 2 was prepared with ESPript; and Fig. 6 was prepared using MOLMOL.

TABLE II
Structural refinement statistics
Numbers in parentheses represent the value for the highest resolution shell.

Space group	$P2_12_12$
Unit cell dimensions (Å)	83.074 × 44.671 × 55.524
Resolution (Å)	30–2.1 (2.13–2.1)
R_{work} (%)[a]	18.7 (18.7)
R_{free} (%)[b]	22.2 (27.3)
No. of reflections	
Working set	11,090
Test set	1,276
No. of atoms	
Protein atoms	1,714 (including 6 selenium atoms)
Water molecules	149
Sulfate ions	5 in 1 sulfate ions
r.m.s.d. from ideality	
Bond lengths (Å)	0.005
Bond angles (°)	1.1
Average B factor (Å^2)	
Main chain	19.1
Side chain	22.8
Water	29.0
Sulfate ion	43.6
Ramachandran plot	
Most favored regions (%)	94.9
Additionally allowed (%)	5.1

[a] $R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$, where F_{obs} and F_{calc} are observed and calculated structure factors.
[b] $R_{free} = \Sigma_T||F_{obs}| - |F_{calc}||/\Sigma_T|F_{obs}|$, where T is a test data set of 10% of the total reflections randomly chosen and set aside prior to refinement.

RESULTS

Overall Structure—The human eIF3k structure was determined by multiwavelength anomalous dispersion with selenomethionine substituted protein. The model was refined to 2.1-Å resolution with a good agreement with diffraction data and has high stereochemical quality (Table II). The final model includes an eIF3k monomer with 213 residues (residue 2–183 and 186–216), 6 selenium atoms, 1 sulfate ion, and 149 water molecules. The C-terminal 2 residues and the polyhistidine tag are disordered. No electron density for residues Gly^{184} and Ser^{185} and the side chains of Gln^{119}, Asp^{182}, Glu^{183}, and Ser^{216} was observed. In addition, residues Phe^{56}, Glu^{123}, Trp^{156}, Ile^{196}, Lys^{197}, and Lys^{204} have weak electron density.

The overall structure of eIF3k has an ear-like shape with dimensions of 35 Å × 35 Å × 60Å. It has 16 helices (14 α helices and two 3_{10} helices) and a three-strand β-sheet (Fig. 1). The secondary structure elements are indicated in Fig. 2. Although sequence analysis did not suggest any known structural motifs, a three-dimensional structural comparison of the crystal structure of eIF3k determined here using DALI (17) clearly showed that eIF3k consists of two distinct domains. The N-terminal region of eIF3k is a HAM domain, named for HEAT (20) analogous motifs, in a mostly right-handed superhelical arrangement, formed by a leading α helix and three HEAT analogous repeats, H1 (consisting of α3 and α4), H2 (α6 and α7), and H3

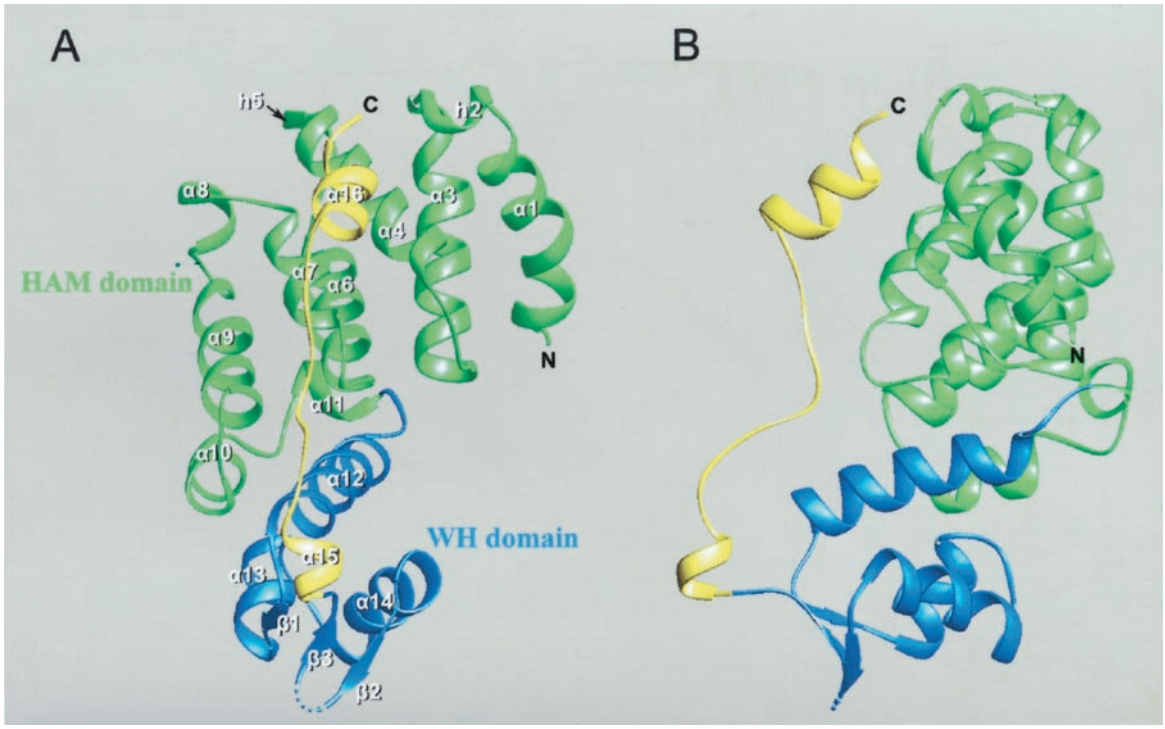

FIG. 1. **The overall structure of human eIF3k.** Human eIF3k contains a HAM domain (*green*) and a WH domain (*blue*) followed by a C-terminal tail (*yellow*). The HAM domain consists of eight α helices (α) and two 3_{10} helices (h); the WH domain consists of three α helices and three β strands (β). The disordered region is indicated by a *broken line* connecting β2 and β3. The loop between h5 and α6 is overlapped by α16. The view in *B* is rotated 90° around a vertical axis from the view in *A*.

(α9 and α10). All of the repeats and the leading α helix are followed by a short helix. The C-terminal half contains a WH (winged-helix) domain, which is followed by a long C-terminal tail flanked by an α helix at both ends. The WH domain is a compact α/β structure containing three α helices (α12, α13, and α14) and three β strands (β1, β2 and β3). Approximately 1500 E^2 of solvent-accessible surface area is buried by the HAM-WH interdomain interactions. The combination of the two domains and C-terminal tail results in two distinct faces of the protein surface: a concave face and a convex face (Fig. 6).

The HAM Domain; Comparison with HEAT-repeat-containing Proteins—HEAT-repeat-containing proteins such as the PR65/A subunit of protein phosphatase 2A (21), nuclear transport protein karyopherin-β2 (22), and α-adaptin C, a subunit of the AP2 complex (23), play important roles in assembling multiprotein complexes in various cellular life activities (24). The HEAT repeat motif is formed by two antiparallel α helices (named αA and αB), and it is usually 37–43 residues long. The motif occurs in blocks of at least 3 and up to 22 tandem repeats (20). HEAT repeat family members have similar patterns of hydrophobic residues, and a highly conserved proline is frequently found in certain αA helices located in the middle of many HEAT repeats to facilitate the kink of the helices (20).

The human eIF3k HAM domain (residues 24–121) is generally similar to that in HEAT-repeat-containing proteins, but it also shows some unusual features. Its three HEAT analogous motifs have 21 (H3) to 32 (H1 and H2) residues. Adjacent HEAT analogous motifs are connected by a short 3_{10} helix (h5, between H1 and H2) or a short α helix (α8, between H2 and H3). Compared with the canonical HEAT-repeat-containing protein PR65/A, two and a half HEAT analogous repeats (α3, α4, α6, α7, and α9) in the HAM domain resemble the HEAT repeats 13 and 14 and αA of repeat 15 of PR65/A, and they can be superimposed with a r.m.s.d. (root mean square deviation) of 3.2 E for 84 residues in the region (Fig. 3). An analysis of the protein sequences (data not shown) shows that several highly conserved hydrophobic residues in the HEAT analogous motifs of the HAM domain occupy the same positions as the corresponding residues in HEAT repeats (20, 25). However, unlike most other HEAT repeats, no proline residues are found in the eIF3k helices (α3, α6, and α9) corresponding to the αA of HEAT repeats. As a consequence, the α helices in the HAM domain are not kinked. These differences between the HAM domain and other HEAT domains suggest that the HEAT analogous motifs in the HAM domain are related but distinct, and they may belong to the new subfamily of ADB (named for β-adatin; Ref. 25). The absence of a consensus proline residue in αA is a unique feature of the ADB repeat in the HEAT repeats family. This is exemplified by the high similarity (Z score = 6.4 by DALI) between the HAM domain and the HEAT motifs (residues Ala^{340} to Ala^{434}) in α-adaptin C, an ADB subfamily member. Nevertheless, it is possible that the HAM domain may belong to a novel class in HEAT repeats as H1, H2, and H3 are structurally different when these three HEAT analogous motifs are compared. The intra-repeat connecting helices in the HAM domain do not exist in the PR65/A HEAT repeats and most HEAT repeats of other proteins (Fig. 3). This unusual topology is also found in a few heat-repeat-containing proteins, such as the PHAT (pseudo-HEAT analogous topology) domain of Smaug (26) (Fig. 3). The two and a half HEAT analogous repeats of the HAM domain are more similar to the PHAT domain (r.m.s.d. = 2.7 Å for 84 residues) than to the PR65/A HEAT repeats (Fig. 3). Another unusual conformation in the HAM domain is the packing between H3 and H2, which is almost side-by-side instead of face-to-face, as in normal HEAT repeats, due to its unusual left-handedness. The unusual H3 turn causes α6, α9, α10, and α12 to form a hydrophobic core (Fig. 4). A similar type of conformation also appears in the structure of the C-terminal region of the vesicular transport protein Sec17 (27) (Fig. 3).

In addition, the eIF3k HAM domain also shows some similarity to several other proteins containing antiparallel α helical

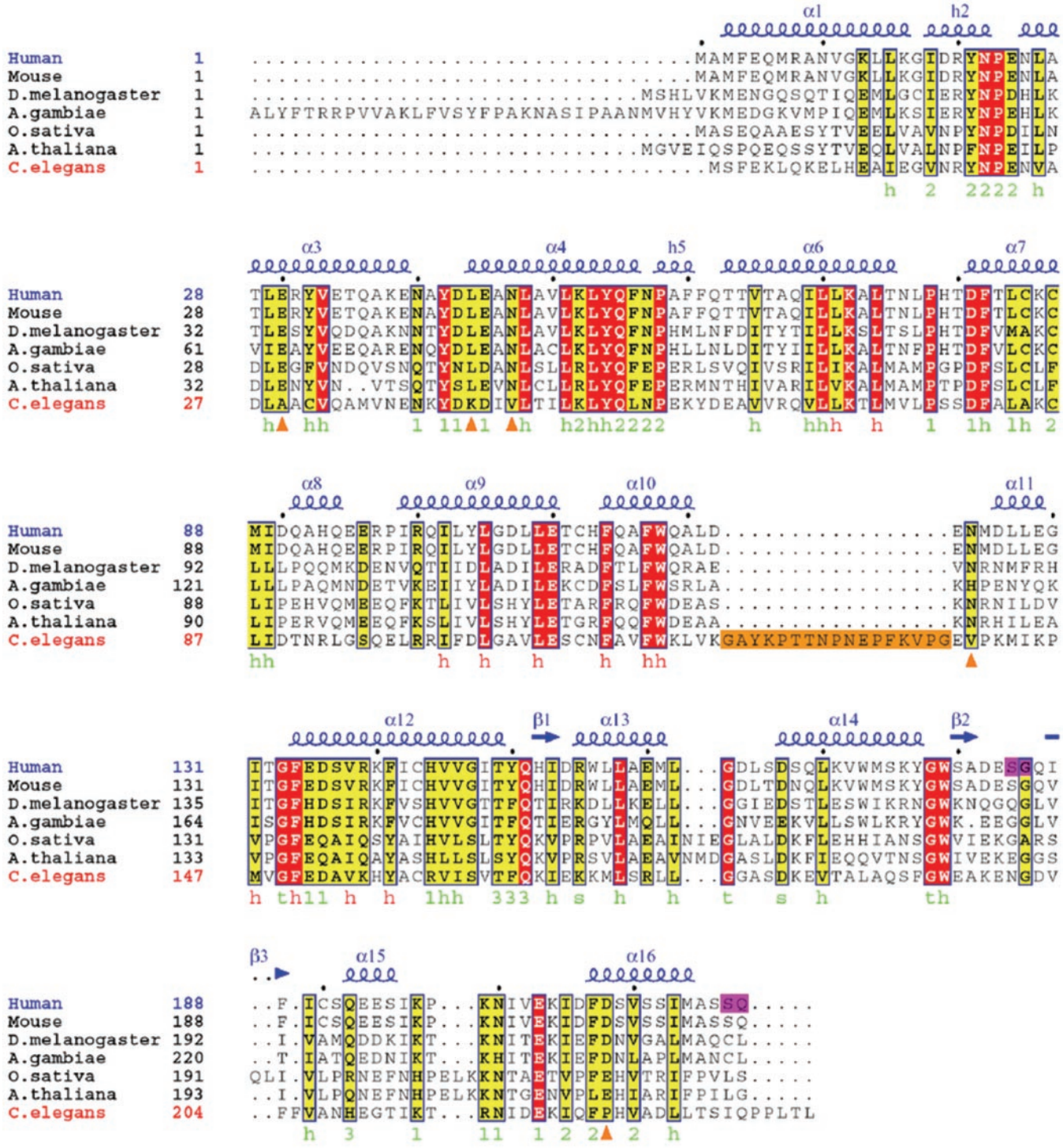

FIG. 2. **Sequence alignment of eIF3k homologues.** Protein sequences were from *Homo sapiens* (human), *Mus musculus* (house mouse), *Drosophila melanogaster*, *Anopheles gambiae*, *Caenorhabditis elegans*, *Oryza sativa*, and *Arabidopsis thaliana*. Sequence accession numbers of the Swiss-Prot data base are Q9UBQ5 (for human), Q9DBZ5 (for mouse), Q9W2D9 (for *D. melanogaster*), Q7QGK4 (for *A. gambiae*), Q9XUP3 (for *C. elegans*), Q94HF1 (for *O. sativa*), and Q9SZA3 (for *A. thaiana*). The sequence of *A. gambiae* is derived from an EMBL/GenBank™/DDBJ whole genome shotgun entry, which should be considered preliminary data. The secondary structure of human eIF3k, which is mainly defined by the analysis of the structure using DSSP program (47), is indicated above the alignment. Residues in the alignment that are identical are shown in *red boxes*; those that are similar are shown in *yellow boxes*. The sequences highlighted in a *pink box* correspond to the disordered regions (Ser^{184}, Gly^{185}, Ser^{217}, and Gln^{218}) missing from the human eIF3k structure. The characters under the alignment denote the predicted function of the conserved residues. The elucidations for these characters are listed as follows: *1*, *2*, and *3*, the residues on the binding surface I, II, and III, respectively (Fig. 6); *h*, intra-molecular hydrophobic interaction (*h colored* in *red* denotes residues that participate in forming the hydrophobic core between the HAM domain and the WH domain shown in Fig. 4); *s*, the two highly conserved residues that form a salt bridges (Fig. 7); *t*, the conserved glycines that make turns between the secondary structure elements; *orange triangle*, the residues that are special in *C. elegans* compared with other organisms (hydrophobic residues in other organisms were replaced by hydrophilic residues in *C. elegans* or in reverse). The sequence of *C. elegans* eIF3k has distinct property in some position. For instance, the residues that are indicated by the *orange triangle* are very different between *C. elegans* and other organisms; and an additional sequence in the *orange box* between α10 and α11 that belong to the HAM domain and WH domain, respectively, exist in *C. elegans* only.

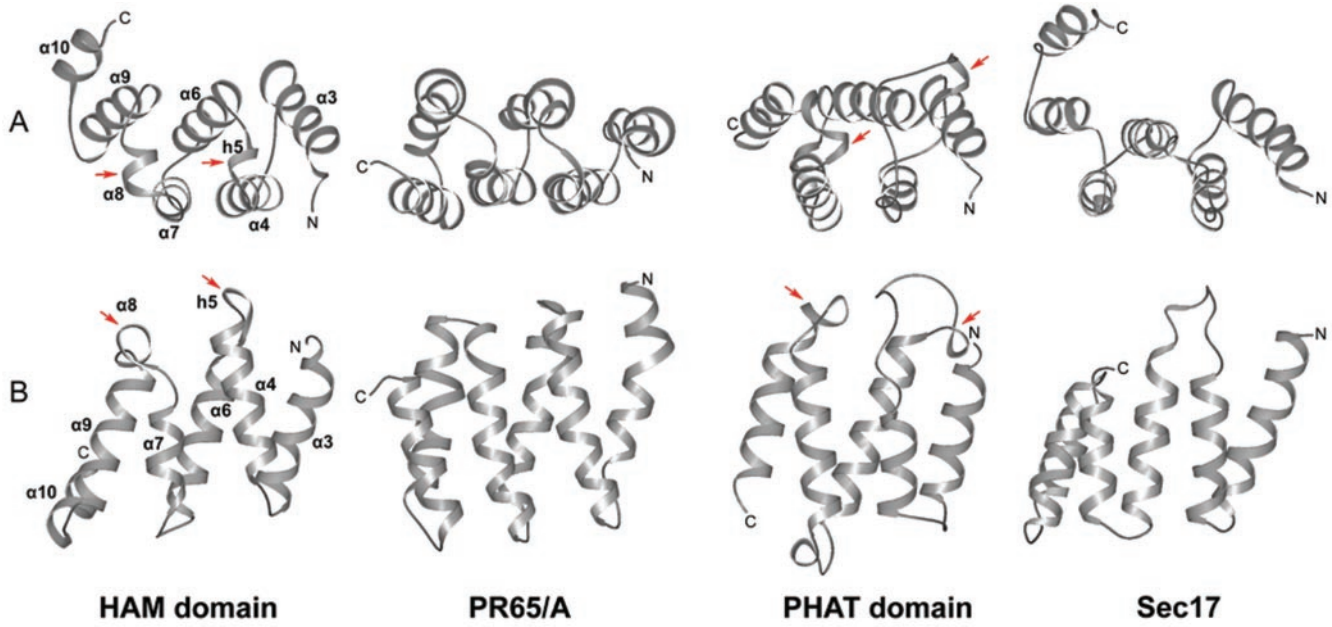

FIG. 3. **Comparison of the HAM domain with other HEAT or HEAT-like repeats.** The HAM domain, PR65/A HEAT repeats, and the Sec17 C-terminal region all contain three antiparallel α-helical repeats, while the PHAT domain of Smaug contains only two and a half antiparallel α-helical repeats. *A*, a top view of four HEAT/HEAT analogous repeats; *B*, a lateral view. The *red arrows* denote the places of the locations of the helices connecting HEAT analogous motifs in the HAM domain and in the PHAT domain. The left-handed superhelix at the C termini of the HAM domain and the Sec17 is seen very clearly from the top view (*A*).

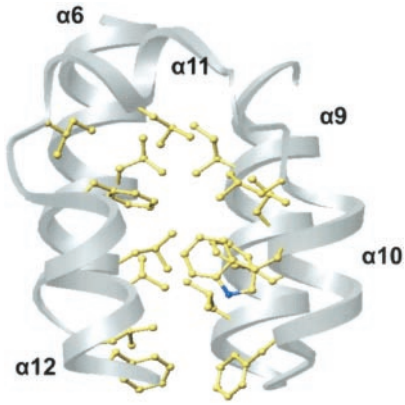

FIG. 4. **The hydrophobic core between the HAM and WH domains.** Highly conserved hydrophobic residues in α6, α9, α10, and α12 form a main part of the core. The core-forming residues are showed in a ball-and-stick models; they include Leu^{128}, Ile^{131}, Leu^{71}, Phe^{134}, Leu^{74}, Val^{138}, Leu^{77}, and Phe^{141} on the *left side* in an order from the top down and on the *right side* are Ile^{102}, Leu^{121}, Leu^{105}, Trp^{118}, Phe^{117}, Leu^{109}, and Phe^{114}. All of these residues are highly conserved (Fig. 2) except Leu^{77}, Leu^{121}, and Leu^{128} (in α11), although the residues corresponding to these three are also hydrophobic among eIF3k homologues in other species (Fig. 2).

repeats, including tetratricopeptide repeats and armadillo repeats that are both involved in protein-protein interactions (24), and to part of the α helix domain (residues 44–164) of chondroitin AC lyase (28).

The WH Domain; Conformation of the WH Domain—The WH domain has the appearance of an earlobe in an ear-like structure (Fig. 1). This domain is stably packed against the HAM domain via the hydrophobic core formed mainly by the conserved hydrophobic residues in α6, α9, α10, and α12 (Fig. 4). The WH domain comprises three α helices and three β strands in the order of α12-β1-α13-α14-β2-β3. α12 and α13 are antiparallel to each other, while α14 is almost perpendicular to α13 (Fig. 5). The latter two helices are connected by a short turn of three amino acids. This architecture belongs to a HTH motif (29). The three short β strands, each containing only 2 residues, form a twisted antiparallel β-sheet, with β1 bonding to the β hairpin formed between β2 and β3, which are bonded by exactly two hydrogen bonds. This arrangement of β strands is a common feature of HTH motifs with a α/β topology (named winged-helix motifs). The hydrophobic residues in the α helices, together with Trp^{179} and Ile^{189}, interdigitate to form another hydrophobic core stabilizing the architecture of the WH domain. Canonical winged-helix motifs commonly have two large loops or wings (w1 and w2). Wing w1 connects β2 and β3, and wing w2 extends from strand β3 to the C terminus of the winged-helix domain (30). In the WH domain of eIF3k, wing w1 is very short (residues 189–193) (Fig. 5); and two residues, 191 and 192, are disordered in the structure. In addition, wing w2 is replace by the α helix (α15) of the C-terminal tail.

Comparison with Other Winged-helix Motifs—A structural similarity search in the PDB (Protein Data Bank) with DALI using the WH domain (residues 132–191) resulted in a number of matches to proteins containing winged-helix motifs. The most similar proteins found by DALI are the selenocysteine-specific elongation factor SelB (PDB code: 1LVA; Z score = 6.5 and r.m.s.d. = 2.4 for 56 residues) (31), the purine operon repressor of *Bacillus subtilis* (PDB code: 1P41; Z score = 6.1 and r.m.s.d. = 1.9 for 54 residues) (32), Rap30 DNA-binding domain (PDB code: 2BBY; Z score = 5.4 and r.m.s.d. = 2.4 for 53 residues) (33), IclR transcriptional factor (PDB code: 1JMR; Z score = 5.3 and r.m.s.d. = 2.7 for 55 residues) (34), endonuclease FokI (PDB code: 2FOK; Z score = 5.2 and r.m.s.d. = 2.4 for 58 residues) (35), double-stranded RNA-specific adenosine deaminase (PDB code: 1QBJ; Z score = 5.2 and r.m.s.d. = 2.7 for 54 residues) (36), and the Esa1 histone acetyltransferase domain (PDB code: 1FY7; Z score = 5.2 and r.m.s.d. = 2.6 for 56 residues) (37). All of winged-helix motifs in these proteins, including the WH domain, differ in the length of the third helix (corresponding to α14 in the WH domain) (Fig. 5), which is associated with major-groove DNA binding and is called the recognition helix. The β hairpin (the wing w1) is also involved in DNA binding via the DNA backbone and/or the minor groove.

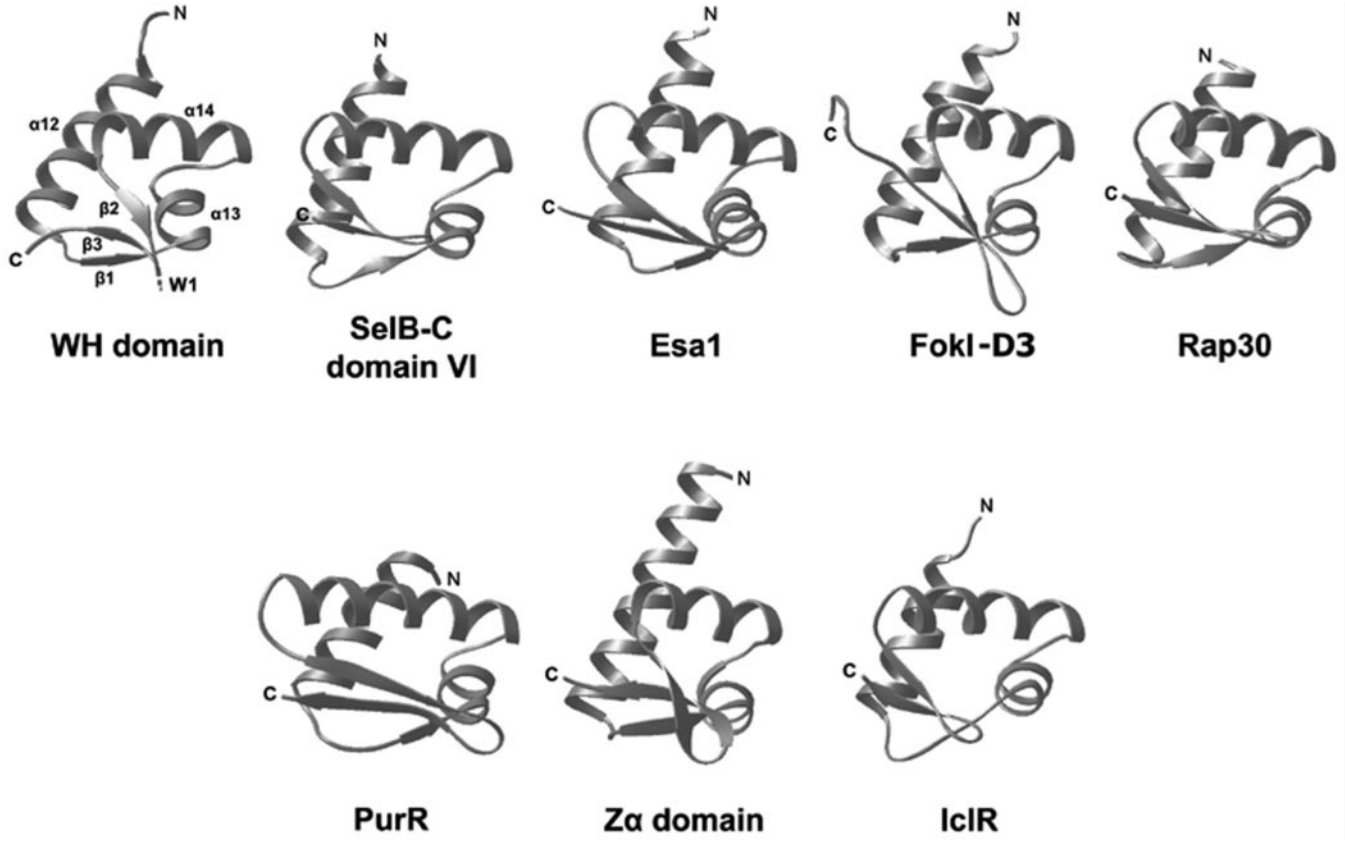

FIG. 5. **A comparison of winged-helix domains.** The recognition helix in the WH domain is α14, and *W1* indicates wing w1 in the WH domain. The two winged-helix domains on the *upper left side* (SelB-C and Esa1) do not have DNA/RNA binding activity, while the two winged-helix domains on *upper rightside* (FokI-D3 and Rap30) interact weakly with DNA and are also involved in protein-protein interactions. The lower three winged-helix domains (PurR, Zα domain, and IclR) are all involved in DNA binding. The disordered region in the WH domain is indicated by a *broken line*.

DISCUSSION

The HAM Domain Can Supply Binding Surfaces for Protein-Protein Interactions—HEAT repeats have been proposed to mediate protein-protein interactions (38, 39), and the internal repetition enlarges the available protein binding surface area. Another eukaryotic translation initiation factor, eIF4GII, also contains a HEAT domain that interacts with eIF4A (40) and eIF1 and eIF5 (41). The detailed binding regions in HEAT motifs have not yet been delineated, but the loops connecting αA with αB (AB loop) may be the binding surface based on sequence conservation, mutational analysis, and oncogenic mutation studies (21). The structure of the tetra-protein complex AP2 (23) provides a model for understanding how proteins bind the AB loop. In the HAM domain of human eIF3k, the AB loops help the formation of the concave face that might be suitable for protein-protein interactions (see below). It is worth mentioning that in the protein PR65/A, the intra-repeat turns of repeats 13–15, which are similar to that in the HAM domain (Fig. 3), were found to be the protein-binding site (21, 42). Although the intra-repeat turns in the HEAT repeats are replaced by α helices (h2, h5, and α8) in the HAM domain, the location of conserved hydrophobic residues on these helices (Fig. 2) suggests that they could still serve as the protein binding surface in eIF3k. These differences might be helpful to enlarge the protein-binding surface of eIF3k. The presence of HEAT analogous motifs in eIF3k strongly suggests that the HAM domain is essential for interactions between eIF3k and other proteins, especially other eIF3 subunits that have been shown to interact with eIF3k directly by glutathione S-transferase pull-down assays (6).

Conserved Residues in eIF3k Suggest Three Binding Surfaces—To gain insights into the mechanisms by which eIF3k interacts with other proteins, we combined the structural knowledge with sequence conservation to identify potential protein-protein interaction sites on eIF3k. We aligned the sequences of seven eIF3k homologues from different organisms that are available in public protein databases (Fig. 2). And we found that many highly conserved residues are distributed throughout the whole protein. The conserved residues can be divided into three classes: hydrophilic amino acids (class I), hydrophobic amino acids involved in inside hydrophobic interactions (class II), and hydrophobic amino acids exposed on the surface (class III). The residues in class II are mainly located at the interface of α helices that interdigitate to maintain the structure of eIF3k. Most of these three classes of residues are marked in Fig. 2 with their predicted function labeled.

Previous experiments indicated that eIF3k could interact with several other eIF3 subunits (eIF3c, eIF3g, and eIF3j), as well as the 40 S ribosomal subunit (6). Thus, the aforementioned class I and class III residues can serve as potential binding sites for the other proteins. Fig. 6 shows the distribution of the conserved residues on the solvent-accessible protein surface. And a surface representation showing the surface electronic potential distribution is display side-by-side for comparison. It is interesting to note that the concave side of the protein could be a putative binding surface, which we call binding surface I, for other proteins. On the binding surface I, there are 14 conserved residues, including 6 electronegative residues, Asp^{43}, Glu^{45}, Asp^{81}, Asp^{135}, Asp^{136}, and Glu^{203}; 2 electropositive residues, Lys^{197} and Lys^{199}; 3 class III residues, Tyr^{42}, Pro^{78}, and Leu^{84}; and some hydrophilic residues (Fig. 6, *A* and *B*). Thus, this concave face may bind a protein through hydrogen bonds (or salt bridges) and hydrophobic interactions. Considering this is the largest potential binding surface, we propose that this region be

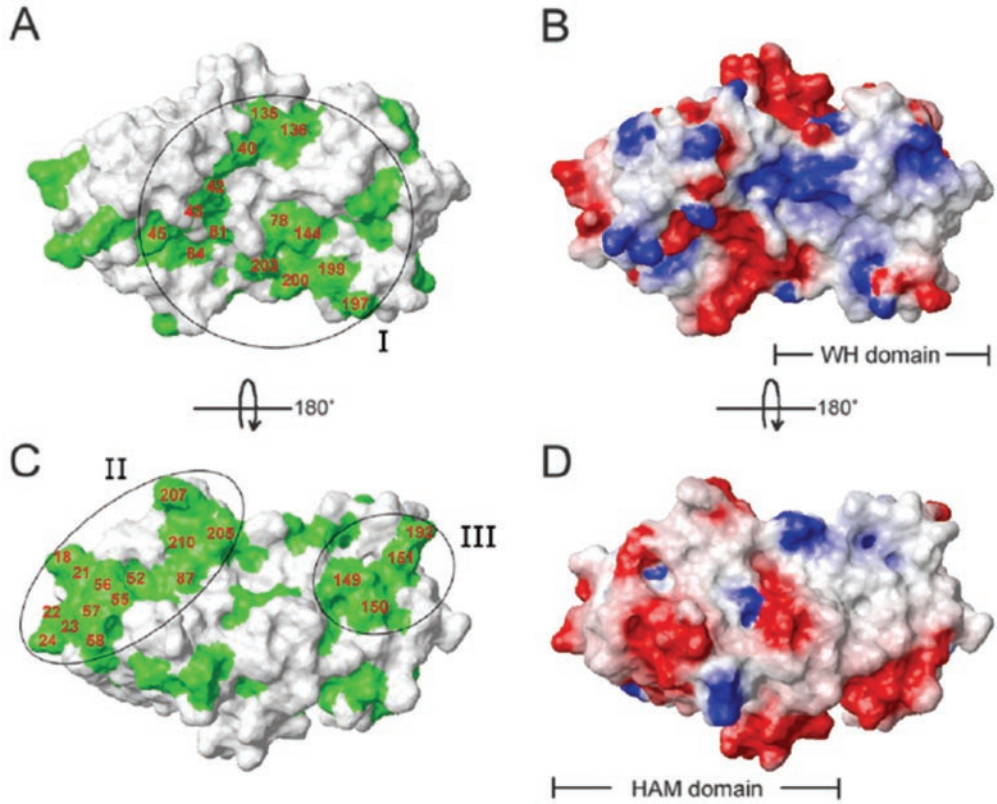

FIG. 6. **Surface distributions of conserved residues and electrostatic potential of eIF3k.** *A*, conserved residues are mapped onto the concave face (colored *light green*). The conserved residues on the binding surface I are marked by *red numbers*. The binding surface I was *circled* in brief, and it appears to be capable of interacting with a rather large protein due to its large area. *B*, the electrostatic potential surface on the concave face of eIF3k is colored *red* for electronegative residues and *blue* for electropositive residues. *C*, conserved residues are mapped onto the convex face. There are two putative binding surfaces (binding surface II and III) for other proteins on the convex face. *D*, the electrostatic potential surface on the convex face of eIF3k. The charged residues are fewer in this face than in concave face. Obviously, there are more electronegative residues than electropositive residues on molecular surface. This is the reason why human eIF3k is an acidic protein with pI = 4.8 (6). All the conserved residues located on the surface could be found in Fig. 2 (labeled with *numbers 1*, *2*, and *3*).

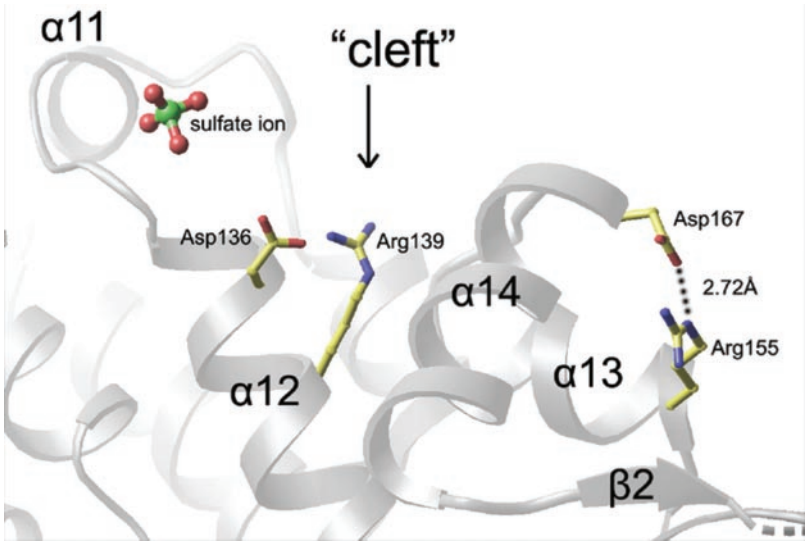

FIG. 7. **The cleft in the WH domain.** The cleft is composed mainly of α12 and α14, the first two helices of the WH domain. The salt bridge between the Nϵ of Arg^{155} and a carboxyl oxygen atom of Asp^{167} is on the *left side* of the cleft (shown as a *dotted line*) with an interatomic distance of 2.72 Å. There are two conserved residues, Asp^{136} and Arg^{139}, at the bottom of the cleft and a sulfate ion at the hatch of the cleft.

involved in the interaction between eIF3k and eIF3c, the biggest subunit among eIF3c, eIF3g, and eIF3j.

The other class III residues are located in a bipartite region of the convex face (Fig. 6*C*). The residues in or around h2 and h5 connecting α1 and H1, H1, and H2, respectively, are reasonably well conserved (Fig. 2). Among these residues, the hydrophobic residues Ile^{18}, Tyr^{21}, Pro^{23}, Phe^{56} (at the C-terminal end of α4), and Pro^{58} belong to class III. The class III residues Ile^{205}, Phe^{207}, and Val^{210} are located in the C-terminal tail of eIF3k whose surface connects with the surface of h2 and h5 to form a rather large protein-binding surface (binding surface II, Fig. 6*C*). The distribution of these highly conserved residues on the protein surface is consistent with their potential involvement in protein-protein interactions mainly through hydrophobic interactions.

The third binding surface (binding surface III) is in the WH domain comprising Thr149, Tyr^{150}, Gln^{151}, and Gln^{192} (Fig. 6*C*). Binding surface III may be responsible for the protein

binding function of the WH domain.

In the human eIF3k structure, there is a salt bridge between the Nϵ of Arg^{155} and a carboxyl oxygen atom of Asp^{167} (Fig. 7), both of which are highly conserved with basic residues (arginine and lysine) and acidic residues (asparagine and glutamic acid) respectively. To our knowledge, this is unique to eIF3k, as no salt bridges are found in the similar region of other winged-helix motifs. The stabilization of this salt bridge may affect the shape of a "cleft" where a sulfate ion and two unidentified electron density masses are present. The two conserved residues, Asp^{136} and Arg^{139}, are at the bottom of the cleft (Fig. 7). The significance of this conformation is still unknown.

The Winged-helix in eIF3k; RNA Binding or Protein Docking—The winged-helix motif is a subfamily of the HTH family, which includes many DNA-binding proteins. However, the structure of the phylogenetically conserved core of the signal recognition particle revealed that a HTH motif could also bind RNA (43). The C-terminal fragment of the elongation factor SelB (SelB-C) provided another example of RNA binding by winged-helix domains (31). As pointed out earlier, eIF3 is required for the binding of the 40 S ribosomal subunit at the 5′-end of mRNA, and it stimulates the binding of the ternary complex, comprised of eIF2, Met-tRNA, and GTP, to the 40 S subunit to form the preinitiation complex (44, 45). These properties of eIF3 provide an intriguing possibility that the eIF3k WH domain may be an RNA binding domain.

Our surprising structural finding that eIF3k contains a winged-helix domain points to a potentially interesting scenario that it may be involved in RNA binding. However, the RNA binding activity of eIF3k has not yet been demonstrated experimentally. Alternative activities other than RNA binding involving the eIF3k WH domain are also possible. A careful examination of the eIF3k WH domain shows that it differs from canonical winged-helix motifs in details. For instance, in well characterized winged-helix motifs, the recognition helix interacts with DNA, and it contains conserved basic residues within the helix or regions immediately surrounding the helix. The corresponding eIF3k helix does not contain conserved basic residues in or around α14 (Fig. 2). Furthermore, not all winged-helix motifs have documented nucleic acids binding activities. For instance, the winged-helix motif in the yeast Esa1 histone acetyltransferase domain is known to be involved in the binding of CoA (37), but its DNA or RNA binding activity has yet to be demonstrated; the D3 domain of FokI, also a winged-helix motif that is highly similar to the WH domain, was found to barely touch DNA, and its putative recognition helix did not interact with the major groove of DNA, thought to be important for protein-protein interactions (46); the Rap30 DNA-binding domain may also be involved in the assembly of the transcription preinitiation complex (33). Thus, the question of whether the recognition helix in eIF3k interacts with RNA, or its WH domain is a protein-docking domain, remains to be clarified.

CONCLUSIONS

We have determined the crystal structure of human eIF3k, which is the first eIF3 subunit structure reported. It contains two distinct domains: a HAM domain and a WH domain. From the results of structure comparisons and residues conservation analysis, we postulate that eIF3k might serve as a structural scaffold, with potential surfaces for binding to other eIF3 subunits and to RNA. Our work provides a starting point for further systematic biochemical, genetic, and structural studies aimed at understanding the structure and function of eIF3 complex and at elucidation of the mechanism of translation initiation.

Acknowledgments—We thank Dr. Peng Liu, Dr. Yuhui Dong, and Deqiang Yao from the Institute of High Energy Physics, Chinese Academy of Sciences for their kind help in diffraction data collection. We also thank Prof. R. Xu and Prof. J. W. B. Hershey for their constructive criticism and suggestions.

REFERENCES

1. Pain, V. M. (1996) Eur. *J. Biochem.* **236,** 747–771
2. Hershey, J. W. B. & Merrick, W. C. (2000) in *Translational Control of Gene Expression* (Sonenberg N., Hershey, J. W. B. & Mathews, M. B., eds) pp. 33–88, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
3. Hinnebusch, A. G. (2000) in *Translational Control of Gene Expression* (Sonenberg, N., Hershey, J. W. B. & Mathews, M. B., eds) pp. 185–243, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
4. Phan L., Zhang X., Asano K., Anderson J., Vornlocher H.-P., Greenberg J. R., Goldfarb D. S., Qin J. & Hinnebusch A. G. (1998) *Mol. Cell. Biol.* **18,** 4935–4946
5. Kim T.-H., Hofmann K., Arnim A. G. & Chamovitz D. A. (2001) *Trends Plant Sci.* **6,** 379–386
6. Mayeur, G. L., Fraser, C. S., Peiretti, F., Block, K. L. & Hershey, J. W. B. (2003) *Eur. J. Biochem.* **270,** 4133–4139
7. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
8. Terwilliger, T. C. & Berendzen, J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 849–861
9. Terwilliger, T. C. (2000) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56,** 965–972
10. Terwilliger, T. C. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59,** 34–44
11. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
12. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
13. Laskowski, R. A., MacArthur, M. W., Moss D. S. & Thornton J. M. J. (*1993*) *Appl. Crystallogr.* **26,** 283–291
14. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
15. Notredame, C., Higgins, D. G. & Heringa, J. (2000) *J. Mol. Biol.* **302,** 205–217
16. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999) *Bioinformatics* **15,** 305–308
17. Holm, L. & Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
18. Koradi, R. & Billeter, M. (1998) *PDB Newslett.* **84,** 5–7
19. Carson, M. (1997) *Methods Enzymol.* **277,** 493–505
20. Kobe, B., Gleichmann, T., Horne, J., Jennings, I. G., Scotney, P. D. & Teh, T. (1999) *Structure* (*Lond.*) **7,** 91–97
21. Groves, M. R., Hanlon, N., Turowski, P., Hemmings, B. A. & Barford, D. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 8669–8672
22. Chook, Y. M. & Blobel, G. (1999) *Nature* **399,** 230–237
23. Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R. & Owen, D. J. (2002) *Cell* **109,** 523–535
24. Peifer, M., Berg, S. & Reynolds, A. B. (1996) *Cell* **76,** 789–791
25. Andrade, M. A., Petosa, C., O'Donoghue, S. I., Müller, C. W. & Bork, P. (2001) *J. Mol. Biol.* **309,** 1–18
26. Green, J. B., Gardner, C. D., Wharton, R. P. & Aggarwal, A. K. (2003) *Mol. Cell* **11,** 1537–1548
27. Rice, L. M. & Brunger, A. T. (1999) *Mol. Cell* **4,** 85–95
28. Féthière, J., Eggimann B. & Cygler, M. (1999) *J. Mol. Biol.* **288,** 635–647
29. Brennan, R. G. & Matthews, B. W. (1989) *J. Biol. Chem.* **264,** 1903–1906
30. Gajiwala, K. S. & Burley, S. K. (2000) *Curr. Opin. Struct. Biol.* **10,** 110–116
31. Selmer, M. & Su, X.-D. (2002) *EMBO J.* **21,** 4145–4153
32. Sinha, S. C., Krahn, J., Shin, B. S., Tomchick, D. R., Zalkin, H. & Smith J. L. (2003) *J. Bacteriol.* **185,** 4087–4098
33. Groft, C. M., Uljon, S. N., Wang, R. & Werner M. H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 9117–9122
34. Zhang, R.-G., Kim, Y., Skarina, T., Beasley, S., Laskowski, R., Arrowsmith, C., Edwards, A., Joachimiak, A. & Savchenko, A. (2002) *J. Biol. Chem.* **277,** 19183–19190
35. Wah, D. A., Bitinaite, J., Schildkraut, I. & Aggarwal, A. K. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 10564–10569
36. Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A. & Rich, A. (1999) *Science* **284,** 1841–1845
37. Yan, Y., Barlev, N. A., Haley, R. H., Berger, S. L. & Marmorstein, R. (2000) *Mol. Cell* **6,** 1195–1205
38. Blatch, G. L. & Lässle, M. (1999) *BioEssay* **21,** 932–939
39. Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. (2001) *J. Struct. Biol.* **134,** 117–131
40. Marcotrigiano, J., Lomakin, I. B., Sonenberg, N., Pestova, T. V., Hellen, C. U. T. & Burley, S. K. (2001) *Mol. Cell* **7,** 193–203
41. He, H., von der Haar, T., Singh, C. R., Ii, M., Li, B., Hinnebusch, A. G., McCarthy, J. E. G. & Asano, K. (2003) *Mol. Cell. Biol.* **23,** 5431–5445
42. Ruediger, R., Hentz, Fait, J., Mumby, M. & Gernot, W. (1994) *J. Virol.* **68,** 123–129
43. Batey, R. T., Rambo, R. P., Lucast, L., Rha, B. & Doudna, J. A. (2000) *Science* **287,** 1232–1239
44. Chaudhuri, J., Chowdhury, D. & Maitra, U. (1999) *J. Biol. Chem.* **274,** 17975–17980
45. Pestova, T. V., Borukhov, S. I. & Hellen, C. U. T. (1998) *Nature* **394,** 854–859
46. Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1997) *Nature* **388,** 97–100
47. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22,** 2577–2637