Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

The structure at 2.5 Å resolution of human basophilic leukemia-expressed protein BLES03

Eduard Bitto,[a] Craig A. Bingman,[a] Howard Robinson,[b] Simon T. M. Allard,[a] Gary E. Wesenberg[a] and George N. Phillips Jr[a]*

[a]Center for Eukaryotic Structural Genomics, Department of Biochemistry, University of Wisconsin-Madison, USA, and [b]Brookhaven National Laboratory, Upton, NY 11973-5000, USA

Correspondence e-mail: phillips@biochem.wisc.edu

Received 9 June 2005
Accepted 25 July 2005
Online 31 August 2005

PDB Reference: BLES03, 1ztp, r1ztpsf.

The crystal structure of the human basophilic leukemia-expressed protein (BLES03, p5326, Hs.433573) was determined by single-wavelength anomalous diffraction and refined to an R factor of 18.8% ($R_{\rm free}$ = 24.5%) at 2.5 Å resolution. BLES03 shows no detectable sequence similarity to any functionally characterized proteins using state-of-the-art sequence-comparison tools. The structure of BLES03 adopts a fold similar to that of eukaryotic transcription initiation factor 4E (eIF4E), a protein involved in the recognition of the cap structure of eukaryotic mRNA. In addition to fold similarity, the electrostatic surface potentials of BLES03 and eIF4E show a clear conservation of basic and acidic patches. In the crystal lattice, the acidic amino-terminal helices of BLES03 monomers are bound within the basic cavity of symmetry-related monomers in a manner analogous to the binding of mRNA by eIF4E. Interestingly, the gene locus encoding BLES03 is located between genes encoding the proteins DRAP1 and FOSL1, both of which are involved in transcription initiation. It is hypothesized that BLES03 itself may be involved in a biochemical process that requires recognition of nucleic acids.

1. Introduction

The human gene locus 11q13.1 encodes basophilic leukemia-expressed protein (BLES03, p5326, UniGene code Hs.433573) with a molecular weight of 27.5 kDa (residues 1–251) and a predicted pI of 5.4. The function of this protein has not yet been established. BLES03 does not show a detectable sequence-family relationship to any previously established protein family based on a *SUPERFAMILY* server search (Gough *et al.*, 2001). A search of the Pfam database revealed that BLES03 and two other closely related sequences form a core of the Pfam-B 63280 family (Bateman *et al.*, 2004). BLES03 thus represents a valuable structural genomics fold-space target. Here we report the three-dimensional structure of BLES03 protein at 2.5 Å determined by single-wavelength anomalous diffraction (SAD). We show that the fold of BLES03 is similar to that of eukaryotic transcription initiation factor 4E (eIF4E), with some topological variations. Furthermore, BLES03 and eIF4E present an analogous electrostatic surface potential despite minimal sequence conservation. The structure was determined under the National Institutes of Health NIGMS Protein Structure Initiative.

2. Materials and methods

The gene encoding the BLES03 protein was cloned and a selenomethionine-labeled protein was purified following the standard Center For Eukaryotic Structural Genomics (CESG) pipeline protocol for cloning (Thao *et al.*, 2004), protein expression (Sreenath *et al.*, 2005), protein purification (Jeon *et al.*, 2005) and overall information management (Zolnai *et al.*, 2003). Crystals of BLES03 were grown by the hanging-drop method from 10 mg ml^{-1} protein solution in buffer (50 m*M* NaCl, 3 m*M* NaN_3, 0.3 m*M* TCEP, 5 m*M* bis-tris pH 6.0) mixed with an equal amount of well solution containing 1.2 *M* sodium citrate, 100 m*M* Tris pH 8.5 at 293 K. Crystals grew as extended rods with dimensions of approximately

200 × 30 × 30 µm. The selenomethionyl crystals of BLES03 belong to space group $P2_12_12_1$, with unit-cell parameters a = 62.5, b = 116.8, c = 123.6 Å. Crystals were cryoprotected by soaking in a solution containing 1.2 *M* sodium citrate, 100 m*M* Tris pH 8.5 supplemented with increasing concentrations of glycerol up to a final concentration of 20%. X-ray diffraction data were collected at the X29 beamline at Brookhaven National Laboratory. The diffraction images were integrated and scaled using *HKL*2000 (Otwinowski & Minor, 1997). The selenium substructure of SeMet-labeled BLES03 crystals was determined using *HySS* and *SHELXD* (Grosse-Kunstleve & Adams, 2003; Schneider & Sheldrick, 2002). The protein structure was phased in *CNS* to 3.7 Å using SAD data (Brünger *et al.*, 1998); phase information was further improved by density modification and phase extention to 2.5 Å resolution. Detailed inspection of the preliminary electron-density maps and selenium substructure of the crystal revealed that one of the three monomers in the asymmetric unit adopted a different conformation that consequently altered the position of the selenium sites. Therefore, we partially traced one of the monomers and manually placed its trace into the electron density of the remaining two monomers using the program *O* (Jones *et al.*, 1991). Operators relating the monomers were obtained using *PDBSET* (Collaborative Computational Project, Number 4, 1994). A partial polyalanine model of the traced monomer was generated in *Xfit* (McRee, 1999) and was used to create a protein mask with the program *MAMA* (Kleywegt & Jones, 1999). A real-space electron-density correlation search with the program *IMP* was then used to optimize the transformation operators (Kleywegt & Read, 1997). The improved operators were used to regenerate all BLES03 monomers. Four equivalent C^{α} atoms were selected from each monomer. These positions, along with the original phase information to 3.7 Å from *CNS*, were supplied to the automatic density modification and NCS-averaging as implemented in *RESOLVE* (Terwilliger, 2000). The resulting electron-density map was of very high quality and the automatic tracing procedure of *ARP*/*wARP* (Perrakis *et al.*, 1999) produced an inital model with approximately 80% of all possible residues placed, of which 93% had side chains assigned. The structure

Table 1
Summary of crystal parameters, data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 62.547, b = 116.809, c = 123.635, α = 90.00, β = 90.00, γ = 90.00
Data-collection and phasing statistics	
Energy (keV)	12.664
Wavelength (Å)	0.97900
Resolution range (Å)	43.97–2.50 (2.56–2.50)
No. of reflections (measured/unique)	390157/31664
Completeness (%)	98.6 (93.1)
$R_{\rm merge}$†	0.106 (0.597)
Redundancy	12.3 (9.0)
Mean $I/\sigma(I)$	13.35 (2.68)
Mean FOM from *CNS*	0.30
Refinement and model statistics	
Resolution range (Å)	84.82–2.50
Data set used in refinement	Selenium peak
No. reflections (total/test)	29882/1607
$R_{\rm cryst}$‡	0.18831
$R_{\rm free}$§	0.24524
R.m.s.d. bonds (Å)	0.013
R.m.s.d. angles (°)	1.422
Average B factor (Å^2)	39.453
No. of water molecules	217
Ramachandran plot, residues in	
Most favorable region (%)	92.3
Additional allowed region (%)	7.7
Generously allowed region (%)	0.0
Disallowed region (%)	0.0

† $R_{\rm merge} = \sum_h \sum_i |I_i(h) - \langle I(h)\rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h)\rangle$ is the mean intensity of the reflection. ‡ $R_{\rm cryst} = \sum_h ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_h |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes, respectively. § $R_{\rm free}$ was calculated as $R_{\rm cryst}$ using 5.1% of the randomly selected unique reflections that were omitted from structure refinement.

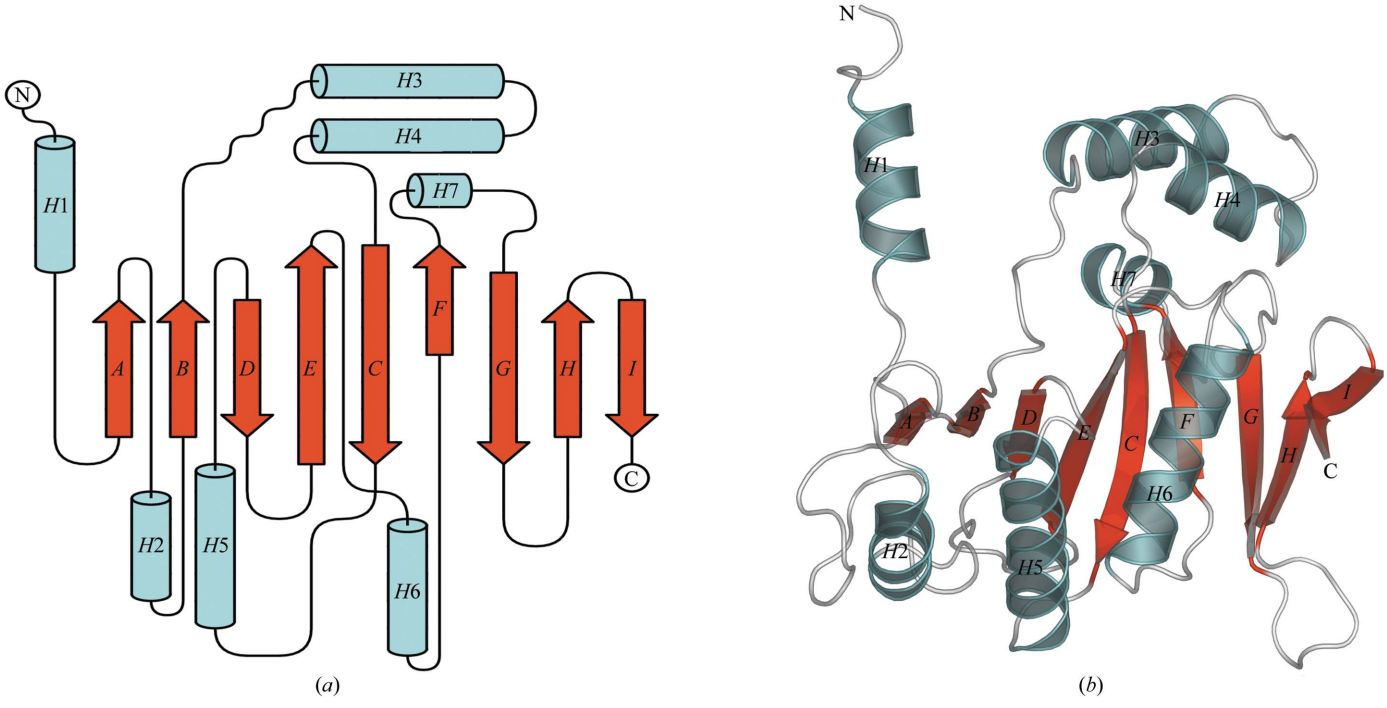


Figure 1
(*a*) A topology diagram of the BLES03 structure (PDB code 1ztp). The central nine-stranded β-sheet (red) is surrounded by several helices (cyan). The figure was prepared using *TopDraw* based on a topology analysis of the BLES03 structure by the *TOPS* server (Bond, 2003; Westhead *et al.*, 1999). (*b*) A ribbon diagram of the BLES03 structure. The structure is labeled and colored to match the topology diagram. The figure was generated using *PyMol* (DeLano, 2002).

was completed using alternate cycles of manual building in *Xfit* and *COOT* and refinement in *REFMAC*5 (McRee, 1999; Emsley & Cowtan, 2004; Murshudov *et al.*, 1997). Loose positional and thermal restraints between the three monomers in the asymmetric unit were applied during the refinement. All refinement steps were monitored using an R_{free} value based on 5.1% of the independent reflections. The stereochemical quality of the final model was assessed using *PROCHECK* and *MolProbity* (Laskowski *et al.*, 1993; Lovell *et al.*, 2003).

3. Results and discussion

The structure of BLES03 has been refined to a resolution of 2.5 Å. Data-collection, refinement and model statistics are summarized in Table 1. The final model describes three monomers, containing residues 17–231 and 238–250 in molecule *A*, residues 17–251 in molecule *B* and residues 30–250 in molecule *C*. In addition, 217 water molecules were built into the final model.

The three-dimensional structure of BLES03 revealed that this protein belongs to the α/β-class of proteins with a two-layer sandwich architecture (see Fig. 1*a*). The central feature of the structure is a nine-stranded β-sheet formed by eight antiparallel β-strands (*B*–*I*) and strand *A* that is parallel to strand *B*. On one side of the central β-sheet are helices *H*2, *H*5 and *H*6 that run in an approximately parallel direction to the β-strands (see Fig. 1*b*). On the other side are helices *H*3, *H*4 and *H*7 that form a smaller satellite domain and a solvent-exposed amino-terminal helix *H*1.

To classify the fold of BLES03, a structural homology search was conducted using the *DALI* and *VAST* servers (Holm & Sander, 1993; Madej *et al.*, 1995). Both *DALI* and *VAST* identified a range of structural homologs of BLES03. The strongest hits identified by both servers were those of eIF4E. Specifically, the top homolog found by DALI was a mouse eIF4E with Z = 9.7, r.m.s.d. 3.2 Å and 15% sequence identity over 136 aligned C^{α} residues (PDB code 1ejh; Marcotrigiano *et al.*, 1999). The *VAST* server also identified the same protein as a top homolog, with a *VAST* score of 14.8, r.m.s.d. 2.3 Å and 16% sequence identity over 119 aligned residues (PDB code 1ej1; Marcotrigiano *et al.*, 1997). Other significant hits from the *VAST* server included eIF4E proteins from mouse (PDB codes 1l8b and 1ejh; Niedzwiecka *et al.*, 2002; Marcotrigiano *et al.*, 1999), human (PDB code 1ipb; Tomoo *et al.*, 2002) and yeast (PDB codes 1rf8 and 1ap8; Gross *et al.*, 2003; Matsuo *et al.*, 1997). The analysis of fold similarity between eIF4E and BLES03 revealed that the overall fold is quite similar. The structures overlap in seven β-strands (one of which, however, runs in an opposite direction) and helices *H*2, *H*5, *H*6 and *H*7 (see Fig. 2).

A structural alignment of all eIF4E proteins with BLES03 also revealed several conserved residues that may be important for fold stability (see Fig. 3*a*). Conserved residues map onto the structure of BLES03 in two clusters. The first cluster involves the fully conserved residues Trp129, Gly127 and Thr176 and the less conserved Tyr203 and Tyr226 (see Fig. 3*b*). Residues of this cluster may be involved in the stabilization of the helix *H*6 interaction with the central β-sheet. The second cluster involves the conserved residues Asp140, Trp143 and Leu133 and the less conserved residue Ser60 (see Fig. 3*b*). Residue Asp140 forms two hydrogen bonds to the amide N atom and hydroxyl O atom of Ser60. Trp143 resides nearby within a hydrophobic pocket and stabilizes the apposition of helix *H*5 against the central β-sheet. Leu133 resides within another hydrophobic pocket formed between helix *H*5 and the central β-sheet strands *C* and *E*. Interestingly, the structurally important residues Gly127, Trp129 and Trp143 are also absolutely conserved among proteins homologous to BLES03 that were identified by *PSI-BLAST* (Altschul et al., 1997) (with threshold $E < 2 \times 10^{-27}$ after three *PSI-BLAST* cycles). On the other hand, several functionally important residues found in eIF4E are not conserved in BLES03 proteins. Most importantly, residues Trp56 and Trp102 of mouse eIF4E that are involved in a stacking interaction with the guanosine ring of the mRNA cap are not conserved in BLES03 (Marcotrigiano *et al.*, 1997). If a favorable conformational change occurred within helix *H*7 of BLES03, Tyr210 of BLES03 could possibly substitute for Trp102 of mouse eIF4E. However, no aromatic residue of BLES03 could take on the role of Trp56 as found in mouse eIF4E. We suspect that BLES03 is not tuned to recognize the mRNA-cap structure.

Several notable topological differences exist between the folds of BLES03 and eIF4E. Firstly, the amino-terminal portion of BLES03 adopts an entirely different topology. Specifically, an antiparallel β-strand that in eIF4E is formed between β-strands analogous to strands *A* and *B* of BLES03 is completely missing in BLES03. A parallel β-strand *A* is formed at this position instead. This topological arrangement results in the positioning of the amino-terminal part of BLES03 on the opposite face of the protein compared with eIF4E. Secondly, BLES03 and eIF4E are topologically different at their

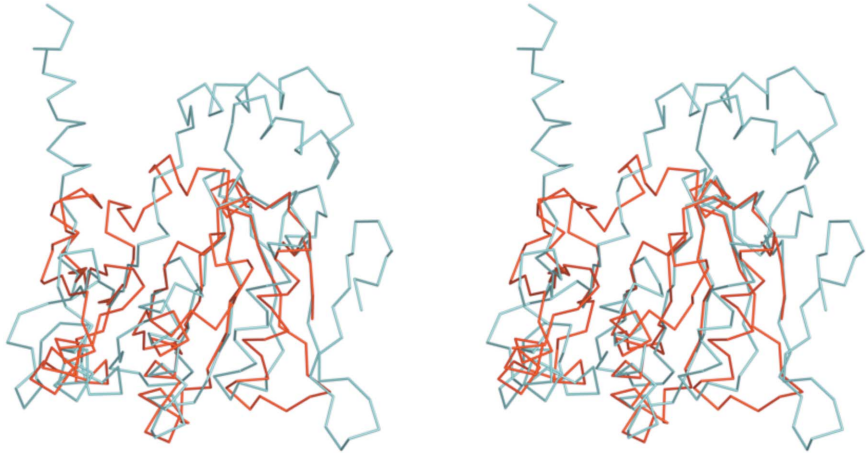


Figure 2
Structural superposition of monomer *B* of human BLES03 (cyan; PDB code 1ztp) and mouse eIF4E (red; PDB code 1eh1). The orientation of the structures is consistent with that introduced in Fig. 1. The figure was generated using *PyMol* based on the structural alignment of proteins by *DALI* (DeLano, 2002; Holm & Sander, 1993).

carboxy-termini. The longer BLES03 protein (251 *versus* 217 residues in mouse eIF4E) forms two additional antiparallel β-strands *H* and *I*. The third significant difference between the folds of BLES03 and eIF4E is the presence of helices *H*3 and *H*4 in BLES03 that are connected to the central β-sheet strands *B* and *C* by an extensive linker. At the same topological location in eIF4E the respective β-strands of the central β-sheet are connected by a much shorter linker containing a single-turn helix (see Fig. 2).

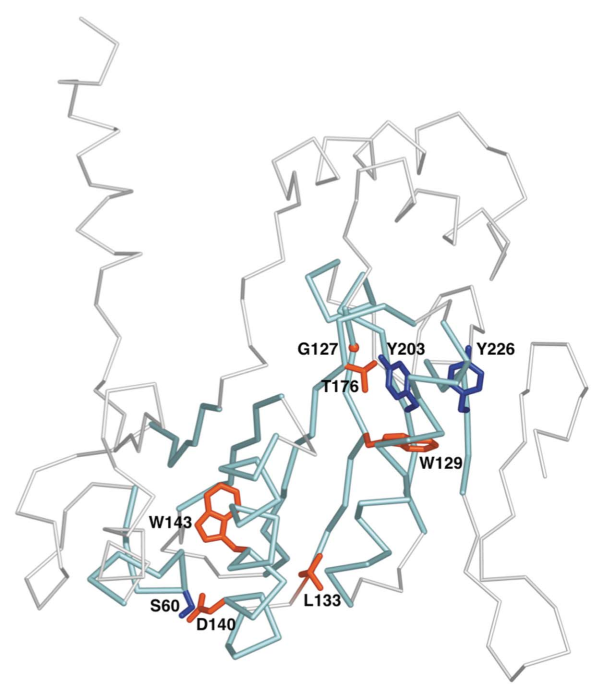
The *PSI-BLAST* search did not detect any sequence similarity between BLES03 and eIF4E. The profile–profile sequence alignment tool *FFAS*03 (Jaroszewski *et al.*, 2000) revealed a putative distant homology between BLES03 and human and yeast eIF4E. The alignment of BLES03 with human eIF4E by FFAS03 spans BLES03 residues 120–195. These residues form the most conserved portion of the protein and include helices *H*5, *H*6 and β-strands *C*, *D* and *E* (see Fig. 1). The alignment with yeast eIF4E spans (with several gaps) the first 205 residues of BLES03. However, the *FFAS*03 scores for both of these alignments fell below the threshold value for the confident prediction of family relationship. The *FFAS*03 predition can therefore be classified as a false negative. Finding that the BLES03 fold is similar to that of eIF4E is therefore significant and will extend our capability to identify additional sequences of proteins likely to adopt the eIF4E/BLES03-like fold.

Despite very low sequence similarity, BLES03 and eIF4E proteins show a significant conservation of electrostatic surface potential. Specifically, extensive basic and acidic patches are located in analogous positions on both proteins (see Fig. 4). In the case of eIF4E, the basic patch is responsible for the binding of the phosphate backbone of a capped mRNA molecule (Tomoo *et al.*, 2002). In crystallized BLES03, the acidic amino-terminal helix *H*1 is bound within the basic cavity located on the surface of a symmetry-related BLES03 molecule (see Fig. 4*a*). We find this observation intriguing and speculate that this patch may also be involved in an interaction with a natural binding partner, possibly an RNA or DNA strand. It is interesting to note that the automatic protein clustering of the BLES03 sequence by *ProtoNet* assigned this protein to the 271979 cluster represented by the DNA-repair protein XPGC/yeast RAD (Kaplan *et al.*, 2004). Many of the 147 members of this cluster are established endonucleases, exonucleases or DNA-repair proteins. It is also intriguing that the human gene encoding the BLES03 protein is located between two genes encoding proteins involved in transcription regulation. Namely, these are the FOSL1 protein of the Fos family and the DR1-associated protein 1 (DRAP1). Fos-family proteins dimerize with proteins of the JUN family to form the transcription-factor complex AP-1 (Wagner, 2002). DRAP1, on the other hand, acts as a co-repressor of transcription by interacting with DR1 and enhancing DR1-mediated repression. DR1 acts as a repressor of transcription by interacting with the TATA-binding protein TFIID and thus preventing the formation of the preinitiation complex (Kim *et al.*, 1997; Yeung *et al.*, 1997). Taken together, these findings lead us to hypothesize that BLES03 may be involved in a biochemical process that requires recognition of nucleic acids.

While inspecting the results of the *PSI-BLAST* search we noticed that two versions of rat BLES03 exist in the sequence databases. One of these proteins contains an additional 41 amino-terminal residues. On inspection of the sequence of human locus 11q13.1 it became clear that the spliced mRNA product of this locus could be transcribed into a BLES03 protein with an additional 42 amino-terminal residues. These residues show about 75% identity to the amino-terminal residues of the extended rat BLES03 protein. The first 20 of these residues form a sequence of low complexity. It is not clear whether the currently annotated or the extended version of the

```
            ....*....|....*....|....*....|....*....|....*....|....*....|
1ZTP   17 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~edgftaehlaaeamaadmdp 36
1EJ1   28 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~vanpehyikhplqnrwalwffkndksktwqan 59
1IPB    1 ~matvepettptpnpptteeektesnqevanpehyikhplqnrwalwffkndksktwqan 59
1RF8    1 msveevskkfeenvsvddttatpktvlsdsahfdvkhplntkwtlwytkpavdkseswsd 60

            ....*....|....*....|....*....|....*....|....*....|....*....|
1ZTP   37 wlvfdarttpateldawlAKYPPSQVTRygdpgspnsepvGWIAVYGqgyspnsgdvqgl 96
1EJ1   60 lrliskfdtvedfwalynHIQLSSNLMPg~~~~~~~~~~~CDYSLFKdgiepmw~~~~~~ 102
1IPB   60 lrliskfdtvedfwalynHIQLSSNLMPg~~~~~~~~~~~CDYSLFKdgiepmw~~~~~~ 102
1RF8   61 llrpvtsfqtveefwaiiQNIPEPHELPl~~~~~~~~~~kSDYHVFRndvrpew~~~~~~ 104

            ....*....|....*....|....*....|....*....|....*....|....*....|
1ZTP   97 qaawealqtsgrpitpgtlrqlaithhvlSGKWLMHL~~~A~~~PGFKLDHAWAGIARAV 150
1EJ1  103 ~~~~~~~~~~~~~~~~~~~~~~edeknkrGGRWLITL~~~NkqqRRSDLDRFWLETLLCL 137
1IPB  103 ~~~~~~~~~~~~~~~~~~~~~~edeknkrGGRWLITLnkqQ~~~RRSDLDRFWLETLLCL 137
1RF8  105 ~~~~~~~~~~~~~~~~~~~~~~edeanakGGKWSFQLc~~G~~~KGADIDELWLCTLLAV 137

            ....*....|....*....|....*....|....*....|....*....|....*....|
1ZTP  151 VEGr~~~~~~~lQVAKVSPrakegGRQVICVYTDDFtDRLGVLEADSAIRAAGik~~~cL 200
1EJ1  138 IGEsfddysddvCGAVVNVr~~~aKGDKIAIWTTECeNRDAVTHIGRVYKERLglppkiV 194
1IPB  138 IGEsfddysddvCGAVVNVr~~~aKGDKIAIWTTECeNREAVTHIGRVYKERLglppkiV 194
1RF8  138 IGEtideddsqiNGVVLSIr~~~kGGNKFALWTKCE~DKEPLLRIGGKFKQVLkltddgH 193

            ....*....|....*....|....*....|....*....|....*....|.
1ZTP  201 LTYKPDVYTYLgiyranrwhlcPTLYESrfqlggsargsrvldrannvelt 251
1EJ1  195 IGYQSHADTATksg~~~~~sttKNRFVV~~~~~~~~~~~~~~~~~~~~~~~ 217
1IPB  195 IGYQSHADTATksg~~~~~sttKNRFVV~~~~~~~~~~~~~~~~~~~~~~~ 217
1RF8  194 LEFFPHCSANGrh~~~~~~~~pQPSITL~~~~~~~~~~~~~~~~~~~~~~~ 213
```

(*a*)



(*b*)

Figure 3

(*a*) A structural sequence alignment of human BLES03 (PDB code 1ztp) and mouse (PDB code 1ej1), human (PDB code 1ipb) and yeast (PDB code 1rf8) eIF4e proteins. Sequences highlighted in upper case cyan letters represent structurally aligned residues among all the proteins. Red upper case letters indicate residues that are fully conserved among the proteins and blue upper case letters indicate additional residues possibly involved in fold stabilization of these proteins. Gray lower case letters represent residues that do not align structurally. The structural alignment was performed using *VAST* (Madej *et al.*, 1995). (*b*) Fully conserved (red) and other interesting (blue) residues mapped onto a C^{α} trace of the BLES03 structure. The color coding of the backbone is consistent with the structural alignment from Fig. 3(*a*). Cyan color highlights segments that are structurally aligned among BLES03 and mouse, human and yeast eIF4e. The figure was prepared using *PyMol* (DeLano, 2002).

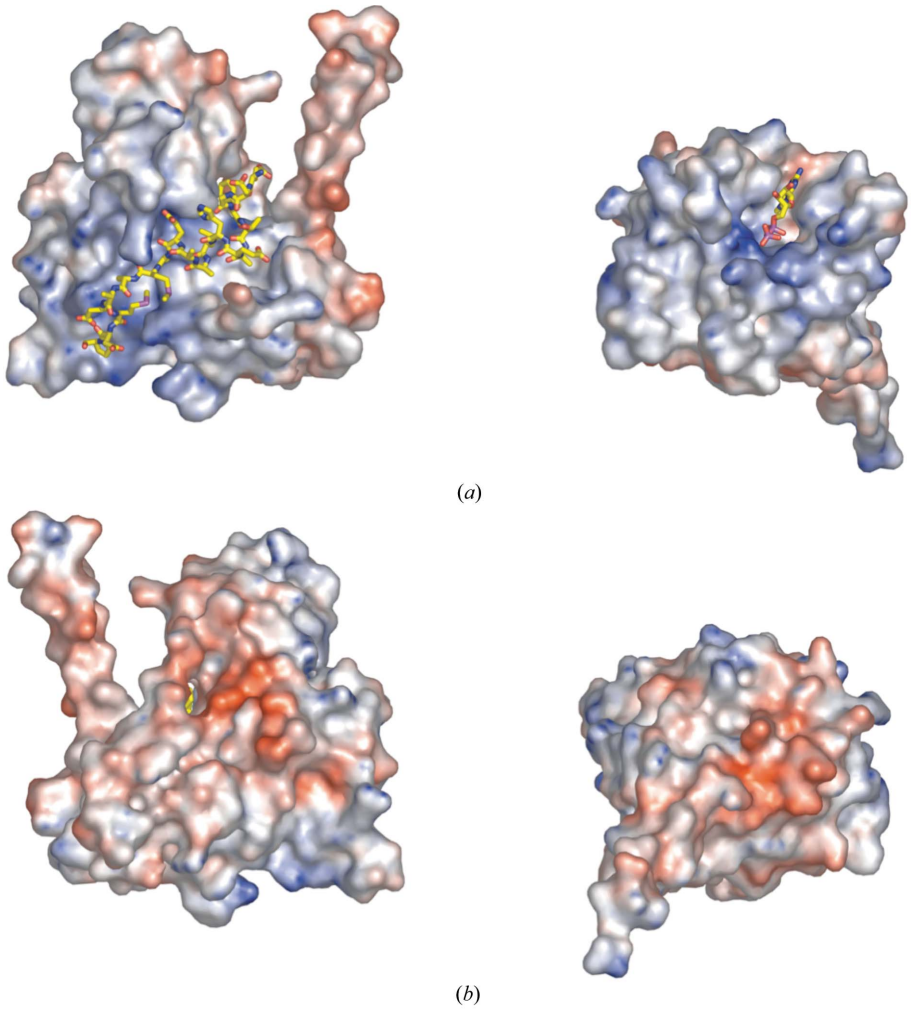


Figure 4
The electrostatic surface potential of BLES03 (left) and mouse eIF4e (right; PDB code 1ej1) contoured from $-10kT$ (red) to $10kT$ (blue). Proteins were structurally superposed as depicted in Fig. 2 and are depicted in the same orientation. (*a*) The extended basic patch on the surface of BLES03 (left) is involved in electrostatic interactions with the acidic helix *H*1 (yellow sticks) of the symmetry-related BLES03 monomer. The basic patch on the surface of mouse eIF4e (right) is involved in the recognition of phosphates of mRNA-cap structure (sticks). (*b*) The acidic patch with the strongest electronegative potential in BLES03 is localized mainly on the surface of a loop connecting the β-strand *E* and the helix *H*6 and the first two turns of the helix *H*6 (see Fig. 1). A similar acidic patch is located on the surface of mouse eIF4E in approximately the same position. The molecules are rotated 180° from the view presented in Fig. 4(*a*). The figure was generated using *PyMol* using electrostatic potential generated by *APBS* (DeLano, 2002; Baker *et al.*, 2001).

human BLES03 protein represents an actual species found *in vivo*. Additional amino-terminal residues could be involved in the physiological regulation of the BLES03 protein function. It is of interest that an amino-terminal extension of about 40 residues found in yeast eIF4E facilitates the folding transition of eIF4G to form a right-handed helical ring which wraps around the amino-terminus of eIF4E. The formation of the complex between eIF4E and eIF4G allosterically enhances the association of eIF4E with the cap of mRNA. eIF4G can then interact with the small ribosomal subunit-interacting protein eIF3 in order to load the ribosome onto mRNA during cap-dependent translation (Gross *et al.*, 2003).

In conclusion, the crystal structure of BLES03 revealed an unexpected structural similarity of this protein to eIF4E. The observed similarity extends beyond the shared fold and encompasses the conservation of both the electrostatic surface potential and the location of a binding cavity.

We acknowledge financial support from NIH National Institute for General Medical Sciences grant P50 GM64598. Data for this study were measured at beamline X29 of the National Synchrotron Light Source. Financial support comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy and from the National Center for Research Resources of the National Institutes of Health. Special thanks goes to all members of the CESG team including Todd Kimball, John Kunert, Nicholas Dillon, Rachel Schiesher, Juhyung Chin, Megan Riters, Andrew C. Olson, Jason M. Ellefson, Janet E. McCombs, Brendan T. Burns, Blake W. Buchan, Holalkere V. Geetha, Zhaohui Sun, Ip Kei Sam, Eldon L. Ulrich, Janelle Warrick, Bryan Ramirez, Zsolt Zolnai, Peter T. Lee, Jianhua Zhang, David J. Aceti, Russell L. Wrobel, Ronnie O. Frederick, Hassan Sreenath, Frank C. Vojtik, Won Bae Jeon, Craig S. Newman, John Primm, Michael R. Sussman, Brian G. Fox and John L. Markley.

References

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.
Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). *Nucleic Acids Res.* **32**, D138–D141.
Bond, C. S. (2003). *Bioinformatics*, **19**, 311–312.
Brünger, 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 Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gough, J., Karplus, K., Hughey, R. & Chothia, C. (2001). *J. Mol. Biol.* **313**, 903–919.
Gross, J. D., Moerke, N. J., von der Haar, T., Lugovskoy, A. A., Sachs, A. B., McCarthy, J. E. & Wagner, G. (2003). *Cell*, **115**, 739–750.
Grosse-Kunstleve, R. W. & Adams, P. D. (2003). *Acta Cryst.* D**59**, 1966–1973.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Jaroszewski, L., Rychlewski, L. & Godzik, A. (2000). *Protein Sci.* **9**, 1487–1496.
Jeon, W., Aceti, D. J., Bingman, C., Vojtik, F., Olson, A., Ellefson, J., McCombs, J., Sreenath, H., Blommel, P., Seder, K., Buchan, B., Burns, B., Geetha, H., Harms, A., Sabat, G., Sussman, M., Fox, B. & Phillips, G. (2005). In the press.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kaplan, N., Friedlich, M., Fromer, M. & Linial, M. (2004). *BMC Bioinformatics*, **5**, 196.
Kim, S., Na, J. G., Hampsey, M. & Reinberg, D. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 820–825.
Kleywegt, G. J. & Jones, T. A. (1999). *Acta Cryst.* D**55**, 941–944.
Kleywegt, G. J. & Read, R. J. (1997). *Structure*, **5**, 1557–1569.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Madej, T., Gibrat, J. F. & Bryant, S. H. (1995). *Proteins*, **23**, 356–369.
Marcotrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. (1997). *Cell*, **89**, 951–961.
Marcotrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. (1999). *Mol. Cell*, **3**, 707–716.
Matsuo, H., Li, H., McGuire, A. M., Fletcher, C. M., Gingras, A. C., Sonenberg, N. & Wagner, G. (1997). *Nature Struct. Biol.* **4**, 717–724.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Niedzwiecka, A., Marcotrigiano, J., Stepinski, J., Jankowska-Anyszka, M., Wyslouch-Cieszynska, A., Dadlez, M., Gingras, A. C., Mak, P., Darzynkiewicz, E., Sonenberg, N., Burley, S. K. & Stolarski, R. (2002). *J. Mol. Biol.* **319**, 615–635.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sreenath, H. K., Bingman, C. A., Buchan, B. W., Seder, K. D., Burns, B. T., Geetha, H. V., Jeon, W. B., Vojtik, F. C., Aceti, D. J., Frederick, R. O., Phillips, G. N. Jr & Fox, B. G. (2005). *Protein Expr. Purif.* **40**, 256–267.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Thao, S., Zhao, Q., Kimball, T., Steffen, E., Blommel, P. G., Riters, M., Newman, C., Fox, B. & Wrobel, R. (2004). *J. Struct. Funct. Genomics*, **5**, 255–265.
Tomoo, K., Shen, X., Okabe, K., Nozoe, Y., Fukuhara, S., Morino, S., Ishida, T., Taniguchi, T., Hasegawa, H., Terashima, A., Sasaki, M., Katsuya, Y., Kitamura, K., Miyoshi, H., Ishikawa, M. & Miura, K. (2002). *Biochem. J.* **362**, 539–544.
Wagner, E. F. (2002). *Ann. Rheum. Dis.* **61**, *Suppl.* 2, 40–42.
Westhead, D. R., Slidel, T. W., Flores, T. P. & Thornton, J. M. (1999). *Protein Sci.* **8**, 897–904.
Yeung, K., Kim, S. & Reinberg, D. (1997). *Mol. Cell. Biol.* **17**, 36–45.
Zolnai, Z., Lee, P. T., Li, J., Chapman, M. R., Newman, C. S., Phillips, G. N. Jr, Rayment, I., Ulrich, E. L., Volkman, B. F. & Markley, J. L. (2003). *J. Struct. Funct. Genomics*, **4**, 11–23.