



Published in final edited form as:

*Proteins*. 2010 May 15; 78(7): 1803–1806. doi:10.1002/prot.22703.

## Crystal structure of an eIF4G-like protein from *Danio rerio*

Euiyoung Bae<sup>1,2</sup>, Eduard Bitto<sup>1,3</sup>, Craig A. Bingman<sup>1</sup>, Jason G. McCoy<sup>1</sup>, Gary E. Wesenberg<sup>1</sup>, and George N. Phillips Jr.<sup>1,\*</sup>

<sup>1</sup>Center for Eukaryotic Structural Genomics, University of Wisconsin-Madison, Madison, Wisconsin

<sup>2</sup>Department of Agricultural Biotechnology and Center for Agricultural Biomaterials, Seoul National University, Seoul, Korea

<sup>3</sup>Department of Chemistry and Biochemistry, Georgian Court University, Lakewood, New Jersey

### Introduction

The gene LOC791917 *Danio rerio* (zebrafish) encodes a protein annotated in the UniProt knowledgebase<sup>1</sup> as the “middle domain of eukaryotic initiation factor 4G domain containing protein b” (MIF4Gdb). Its molecular weight is 25.8 kDa, and it comprises 222 amino acid residues. BLAST searches revealed homologues of *D. rerio* MIF4Gdb in many eukaryotes including humans.<sup>2</sup> The homologues and MIF4Gdb were identified as members of the Pfam family, MIF4G (PF02854), which is named after the middle domain of eukaryotic initiation factor 4G (eIF4G).<sup>3–5</sup> eIF4G is a component of eukaryotic translational initiation complex, and contains binding sites for other initiation factors, suggesting its critical role in translational initiation.<sup>6</sup> The MIF4G domain also occurs in several other proteins involved in RNA metabolism, including the Nonsense-mediated mRNA decay 2 protein (NMD2/UPF2), and the nuclear cap-binding protein 80-kD subunit (CBP80).<sup>5</sup> Sequence and structure analysis of the MIF4G domains in many proteins indicates that the domain assumes all helical fold and has tandem repeated motifs.<sup>5,7</sup> The zebrafish protein described here has homology to domains of other proteins variously referred to as NIC-containing proteins (NMD2, eIF4G, CBP80). The biological function of *D. rerio* MIF4Gdb has not yet been experimentally characterized, and the annotation is based on amino acid sequence comparison. *D. rerio* MIF4Gdb did not share more than 25% sequence identity with any protein for which the three-dimensional structure is known and was selected as a target for structure determination by the Center for Eukaryotic Structural Genomics (CESG). Here, we report the crystal structure of *D. rerio* MIF4Gdb (UniGene code Dr.79360, UniProt code Q5EAQ1, CESG target number GO.79294).

### Materials and Methods

The gene coding for MIF4Gdb was selected as a target as part of a group of genes chosen to code for proteins that are as dissimilar as possible to structures previously deposited in the Protein Data Bank, and also as dissimilar as possible to targets that CESG had previously selected. It was assigned a project database identifier of GO.79294. Complete, detailed protocols for the production of this protein can be found in PepcDB.<sup>8</sup> Briefly, the gene was cloned into pVP33K, the first production Flexi@Vector<sup>9</sup> used on our project, and selenomethionyl protein was purified following the standard CESG pipeline protocol for cloning,<sup>10</sup> protein expression,<sup>11</sup> protein purification<sup>12</sup> and overall information management.

\*Correspondence to: Center for Eukaryotic Structural Genomics, University of Wisconsin-Madison, Madison, Wisconsin 53706. Telephone: 608-263-6142, Fax: 608-262-3453, phillips@biochem.wisc.edu..

<sup>13</sup> Initial crystallization screens were conducted at 4 and 20 °C, in Corning 3775 plates, using a local screen called UW-192. Crystal growth was monitored using Bruker Nonius Crystal Farms at 4°C and 20°C, and scored using Crystal Farm Navigator (Nexus Biosystems, Inc.). A Tecan Genesis RSP 150 robot assembled precipitant solutions for optimization experiments. Diffraction quality crystals were grown in hanging drop batch experiments. Crystals were grown from 10 mg/ml protein solution in buffer (50 mM NaCl, 3 mM NaN<sub>3</sub>, 0.3 mM TCEP, 5 mM BisTris pH 7.0) mixed with an equal amount of reservoir solution containing 7% (w/v) PEG 4K, 0.4 M NaCl, 100 mM MES/ acetate pH 5.5 at 20 °C. The crystals were cryoprotected in 15% (w/v) ethylene glycol, 10% (w/v) PEG 4K, 100 mM MES/ acetate pH 5.5 and were flash-frozen in liquid nitrogen.

Diffraction data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source (APS), Argonne National Laboratory at 100 K. The diffraction images were processed with HKL2000.<sup>14</sup> The selenium substructure of the crystal was determined by using SHELXD<sup>15</sup> and HySS from PHENIX,<sup>16,17</sup> and the selenium positions were used for single wavelength anomalous diffraction phasing in autoSHARP.<sup>18</sup> The initial model was built by the automatic tracing procedure of ARP/wARP,<sup>19</sup> and the structure was completed using alternating cycles of manual building in COOT<sup>20</sup> and refinement in REFMAC5.<sup>21</sup> The stereochemical quality of the final model was assessed using MolProbity.<sup>22</sup> PyMol was used to generate figures.<sup>23</sup> The final coordinates were deposited in the RCSB Protein Data Bank<sup>24</sup> with accession number 2I2O.

## Results and Discussion

The crystal structure of MIF4Gdb from *D. rerio* was determined to a resolution of 1.92 Å using single wavelength anomalous diffraction. Data collection and refinement statistics are summarized in Table 1. The asymmetric unit of the structure contains two MIF4Gdb chains (residues 7–217 for chain A; residues 8–217 for chain B). Several N- and C-terminal residues were not included in the model due to insufficient electron density.

The structure of the MIF4Gdb monomer reveals a crescent shaped molecule consisting entirely of helices (13  $\alpha$ - and two  $3_{10}$ -helices) and connecting loops (Figure 1). Except for the two terminal ones (h1 and h15), the 13 helices (h2–h14) are arranged as four HEAT-like (huntingtin-elongation-A subunit-TOR-like) motifs containing armadillo repeats.<sup>25</sup> Each HEAT-like motif consists mainly of two antiparallel  $\alpha$ -helices (termed A and B) that are held together by hydrophobic interactions along their adjacent sides. The eight longer  $\alpha$ -helices (h3, h5, h6, h8, h10, h12–h14) serve as the main constituents of the four HEAT-like motifs forming the characteristic antiparallel  $\alpha$ -helical pairs, and the five shorter helices (h2, h4, h7, h9 and h11) are located either within a motif or between motifs mediating turns between adjacent helices. Each subunit has a nickel (or perhaps zinc) atom bound on the concave side of the crescent-shaped dimer. Whether there is a physiological role for bound metal ion is unknown.

The consecutive HEAT-like motifs are stacked on each other and the polypeptide chain forms a right-handed solenoid. The stacking of the four HEAT-like motifs is parallel. That is, the helices of the same type (A or B) in the motifs are located side by side. This parallel arrangement of the four HEAT-like motifs generates a double layer of  $\alpha$ -helices in which the four A helices forms one face and the four B helices comprise the other. The structure has an extended hydrophobic core, which is essentially the area between the two layers formed by the A and B helices. The hydrophobic core is stabilized by salt bridges and Van der Waals interactions between the conserved nonpolar residues.

Tandem arrays of HEAT-like motifs are found in a wide variety of proteins where they serve as scaffolding modules for assembly of large multi-protein complexes.<sup>25</sup> They include huntingtin protein, protein phosphatase 2A (pp2A), importin  $\beta$ , elongation factor 3, and many others. The closest structural neighbors of the MIF4Gdb monomer identified by VAST search are the middle segment of eukaryotic initiation factor 4GI (eIF4GI) from *Saccharomyces cerevisiae* (PDB ID: 2VSX) and the middle domain of human eIF4GII (PDB ID: 1HU3).<sup>26,27</sup> The MIF4Gdb structure described here superposes onto the homologous chain of the yeast structure with VAST score 12.4, 2.5 Å RMSD and 20.4 % sequence identity over 167 aligned residues. The overlay with the human protein is comparable, with a score of 12.5, 2.4 Å RMSD Å, and 24.7% identity over 154 amino acids. The next closest structure in the list is the pp2A (PDB ID: 3GFA) with only 8.9% sequence identity. The comparison of these structures reveals the same overall fold with slight differences in the orientation of the N-terminal and C-terminal helices. All three appear to be dimers in the physiological state. The dimeric interfaces amongst these three most similar proteins are also conserved, based on analysis with the PISA server<sup>28</sup>, giving a Q score of 0.371 and 0.370 for 1HU3 and 2VSX with our structures, respectively.

The yeast eIF4GI is one of the two isoforms (eIF4GI and eIF4GII) of the translation initiation factor eIF4G, a modular adaptor protein that recruits the components necessary for the initiation of protein synthesis in eukaryotes.<sup>29,30</sup> In the yeast complex, the helical domains of eIF4G serve to orient DEAD-box sequence motifs of an RNA helicase in a way that they become active. Some eIF4G proteins also bind other eukaryotic initiation factors and picornaviral IRES (internal ribosome entry site) elements.<sup>31</sup> Previous structural and mutational studies identified several residues of eIF4G involved in eIF4A and IRES binding.<sup>27,32</sup> Although some of these residues are also present in MIF4Gdb, it is not straightforward to make a prediction about eIF4G-like function of MIF4Gdb on this basis since the ligand-binding residues are poorly conserved even among the eukaryotic eIF4G homologues.<sup>32</sup>

In a more recent study, a human protein with 72% sequence identity to the zebrafish eIF4Gdb was shown to be involved in stem loop rather than polyA mediated translation<sup>33</sup>, which is common for histone mRNA's. This human protein has been named SLIP1, for SLBP (Stem Loop Binding Protein) interacting protein 1. Given the sequence identity and the known role of similar proteins, it seems quite possible that the purpose of the protein whose structure is defined here is also to act as a scaffold that helps assemble components of translation machinery.

## Acknowledgments

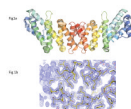
We thank all the members of the CESC. This work was supported by National Institutes of Health/National Institute for General Medical Sciences Grants P50 GM64598 and U54 GM074901 (J. L. Markley, PI). Euiyoung Bae was supported by the Korea Research Foundation (KRF) grant funded by the Korea government (MEST) (No. 2009-0067791) and Research Settlement Fund for the new faculty of Seoul National University. Data were collected at SER-CAT 22-ID beamline at the APS, Argonne National Laboratory. We thank John Chrzas and John Gonczy for support during our data collection. Supporting institutions may be found at [www.ser-cat.org/members.html](http://www.ser-cat.org/members.html). Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

## References

1. TheUniprotConsortium. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* 2009; 37:D169–D174. [PubMed: 18836194]
2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25(17):3389–3402. [PubMed: 9254694]

3. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. *Nucleic Acids Res.* 2004; 32(Database issue):D138–141. [PubMed: 14681378]
4. Ponting CP. Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. *Trends Biochem Sci.* 2000; 25(9):423–426. [PubMed: 10973054]
5. Aravind L, Koonin EV. Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. *Genome Res.* 2000; 10(8):1172–1184. [PubMed: 10958635]
6. Keiper BD, Gan W, Rhoads RE. Protein synthesis initiation factor 4G. *Int J Biochem Cell Biol.* 1999; 31(1):37–41. [PubMed: 10216942]
7. Mazza C, Ohno M, Segref A, Mattaj IW, Cusack S. Crystal structure of the human nuclear cap binding complex. *Mol Cell.* 2001; 8(2):383–396. [PubMed: 11545740]
8. Berman HM, Westbrook JD, Gabanyi MJ, Tao W, Shah R, Kouranov A, Schwede T, Arnold K, Kiefer F, Bordoli L, Kopp J, Podvinec M, Adams PD, Carter LG, Minor W, Nair R, La Baer J. The protein structure initiative structural genomics knowledgebase. *Nucleic Acids Res.* 2009; 37(Database issue):D365–368. [PubMed: 19010965]
9. Blommel PG, Martin PA, Seder KD, Wrobel RL, Fox BG. Flexi vector cloning. *Methods Mol Biol.* 2009; 498:55–73. [PubMed: 18988018]
10. Thao S, Zhao Q, Kimball T, Steffen E, Blommel PG, Ritters M, Newman CS, Fox BG, Wrobel RL. Results from high-throughput DNA cloning of *Arabidopsis thaliana* target genes using site-specific recombination. *J Struct Funct Genomics.* 2004; 5(4):267–276. [PubMed: 15750721]
11. Sreenath HK, Bingman CA, Buchan BW, Seder KD, Burns BT, Geetha HV, Jeon WB, Vojtik FC, Aceti DJ, Frederick RO, Phillips GN Jr. Fox BG. Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expr Purif.* 2005; 40(2):256–267. [PubMed: 15766867]
12. Jeon WB, Aceti DJ, Bingman CA, Vojtik FC, Olson AC, Ellefson JM, McCombs JE, Sreenath HK, Blommel PG, Seder KD, Burns BT, Geetha HV, Harms AC, Sabat G, Sussman MR, Fox BG, Phillips GN Jr. High-throughput purification and quality assurance of *Arabidopsis thaliana* proteins for eukaryotic structural genomics. *J Struct Funct Genomics.* 2005; 6(2-3):143–147. [PubMed: 16211511]
13. Zolnai Z, Lee PT, Li J, Chapman MR, Newman CS, Phillips GN Jr. Rayment I, Ulrich EL, Volkman BF, Markley JL. Project management system for structural and functional proteomics: Sesame. *J Struct Funct Genomics.* 2003; 4(1):11–23. [PubMed: 12943363]
14. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.* 1997; 276:307–326.
15. Uson I, Sheldrick GM. Advances in direct methods for protein crystallography. *Curr Opin Struct Biol.* 1999; 9(5):643–648. [PubMed: 10508770]
16. Grosse-Kunstleve RW, Adams PD. Substructure search procedures for macromolecular structures. *Acta Crystallogr D Biol Crystallogr.* 2003; 59(Pt 11):1966–1973. [PubMed: 14573951]
17. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr.* 2002; 58(Pt 11):1948–1954. [PubMed: 12393927]
18. de la Fortelle E, Bricogne G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Method Enzymol.* 1997; 276:472–494.
19. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol.* 1999; 6(5):458–463. [PubMed: 10331874]
20. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–2132. [PubMed: 15572765]
21. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 1997; 53(Pt 3):240–255. [PubMed: 15299926]

22. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*. 2003; 50(3):437–450. [PubMed: 12557186]
23. DeLano, WL. The PYMOL Molecular Graphic System. DeLano Scientific LLC; San Carlos, CA, USA: 2002.
24. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, Westbrook J. The Protein Data Bank and the challenge of structural genomics. *Nat Struct Biol*. 2000; 7(Suppl):957–959. [PubMed: 11103999]
25. Andrade MA, Bork P. HEAT repeats in the Huntington's disease protein. *Nat Genet*. 1995; 11(2): 115–116. [PubMed: 7550332]
26. Madej T, Gibrat JF, Bryant SH. Threading a database of protein cores. *Proteins*. 1995; 23(3):356–369. [PubMed: 8710828]
27. Schutz P, Bumann M, Oberholzer AE, Bieniossek C, Trachsel H, Altmann M, Baumann U. Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci U S A*. 2008; 105(28):9564–9569. [PubMed: 18606994]
28. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007; 372(3):774–797. [PubMed: 17681537]
29. Sonenberg N, Dever TE. Eukaryotic translation initiation factors and regulators. *Curr Opin Struct Biol*. 2003; 13(1):56–63. [PubMed: 12581660]
30. Merrick WC. Cap-dependent and cap-independent translation in eukaryotic systems. *Gene*. 2004; 332:1–11. [PubMed: 15145049]
31. Fraser CS, Doudna JA. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat Rev Microbiol*. 2007; 5(1):29–38. [PubMed: 17128284]
32. Marcotrigiano J, Lomakin IB, Sonenberg N, Pestova TV, Hellen CU, Burley SK. A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol Cell*. 2001; 7(1):193–203. [PubMed: 11172724]
33. Cakmakci NG, Lerner RS, Wagner EJ, Zheng L, Marzluff WF. SLIP1, a factor required for activation of histone mRNA translation by the stem-loop binding protein. *Mol Cell Biol*. 2008; 28(3):1182–1194. [PubMed: 18025107]



**Figure 1.** Crystal structure of *D. rerio* MIF4Gdb. (A) Ribbon diagram of *D. rerio* MIF4Gdb structure. A rainbow gradient is used to color each polypeptide chain from its N-terminus (blue) to C-terminus (red). Pairs of  $\alpha$ -helices can be seen to form an extended sheet. The C-termini form dimer contacts and the N-terminal regions are available for interaction with other proteins, in keeping with other homologs. (B) Representative electron density of the MIF4Gdb structure. Amino acid residues are colored by atom types (carbon: green; oxygen: red; nitrogen: blue), and the  $2F_{\text{obs}} - F_{\text{calc}}$  map is contoured at  $1.5\sigma$ .

**Table 1**

Summary of crystal parameters and data collection and refinement statistics (values in parentheses are for the highest-resolution shell).

Space group	C2
Unit cell parameters	a=317.80Å, b=40.95Å, c=40.92Å, β=90.26°
Wavelength (Å)	0.97925
Data collection statistics	
Resolution range (Å)	40.61 – 1.92 (1.99 – 1.92)
Number of reflections, measured/unique	254650/39789
Completeness (%)	97.5 (95.5)
R <sub>merge</sub> <sup>a</sup>	0.074 (0.245)
Redundancy	6.40 (5.40)
Mean I/σ(I)	13.38 (5.40)
Refinement statistics	
Resolution range (Å)	40.61 – 1.92 (1.97 – 1.92)
Number of reflections, total/test	39773/1998
R <sub>cryst</sub> <sup>b</sup> /R <sub>free</sub> <sup>c</sup>	0.190/0.235 (0.231/0.277)
RMSD bonds (Å)	0.015
RMSD angles (deg)	1.357
Average B factor (Å <sup>2</sup> )	8.56
Number of water molecules	418
Ramachandran favored (%)	97.7
Ramachandran allowed (%)	99.8

<sup>a</sup>R<sub>merge</sub> =  $\frac{\sum_h \sum_i |I_i(h) - \langle I(h) \rangle|}{\sum_h \sum_i I_i(h)}$ , 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.

<sup>b</sup>R<sub>cryst</sub> =  $\frac{\sum_h \|F_{obs} - F_{calc}\|}{\sum_h F_{obs}}$ , where  $F_{obs}$  and  $F_{calc}$  are the observed and calculated structure factor amplitudes, respectively.

<sup>c</sup>R<sub>free</sub> was calculated as R<sub>cryst</sub> using 5.0% of the randomly selected unique reflections that were omitted from structure refinement.