PROTEIN STRUCTURE REPORT

Crystal structure of an enhancer of rudimentary homolog (ERH) at 2.1 Å resolution

RYOICHI ARAI,^{1,3} MUTSUKO KUKIMOTO-NIINO,¹ HIROKO UDA-TOCHIO,¹
SATOSHI MORITA,¹ TOMOMI UCHIKUBO-KAMO,¹ RYOGO AKASAKA,¹
YUUKA ETOU,¹ YOSHIHIDE HAYASHIZAKI,² TAKANORI KIGAWA,¹
TAKAHO TERADA,^{1,3} MIKAKO SHIROUZU,^{1,3} AND
SHIGEYUKI YOKOYAMA^{1,3,4}

¹Protein Research Group and ²Genome Exploration Research Group, RIKEN Genomic Sciences Center, Tsurumi, Yokohama 230-0045, Japan

³RIKEN Harima Institute at SPring-8, Mikazuki, Sayo, Hyogo 679-5148, Japan

⁴Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan

(RECEIVED March 30, 2005; FINAL REVISION March 30, 2005; ACCEPTED April 1, 2005)

Abstract

The *enhancer of rudimentary* gene, *e(r)*, of *Drosophila melanogaster* encodes an enhancer of rudimentary (ER) protein with functions implicated in pyrimidine biosynthesis and the cell cycle. The ER homolog (ERH) is highly conserved among vertebrates, invertebrates, and plants. *Xenopus laevis* ERH was reported to be a transcriptional repressor. Here we report the 2.1 Å crystal structure of murine ERH (Protein Data Bank ID 1WZ7), determined by the multiwavelength anomalous dispersion (MAD) method. The monomeric structure of ERH comprises a single domain consisting of three α -helices and four β -strands, which is a novel fold. In the crystal structure, ERH assumes a dimeric structure, through interactions between the β -sheet regions. The formation of an ERH dimer is consistent with the results of analytical ultracentrifugation. The residues at the core region and at the dimer interface are highly conserved, suggesting the conservation of the dimer formation as well as the monomer fold. The long flexible loop (44~53) is also significantly conserved, suggesting that this loop region may be important for the functions of ERH. In addition, the putative phosphorylation sites are located at the start of the β 2-strand (Thr18) and at the start of the α 1-helix (Ser24), implying that the phosphorylation might cause some structural changes.

Keywords: enhancer of rudimentary homolog (ERH); novel fold; pyrimidine biosynthesis; cell cycle, transcriptional repressor, cell-free protein synthesis; structural genomics

The *enhancer of rudimentary* gene, *e(r)*, which encodes the ER protein, was originally cloned in a genetic screen to isolate genes interacting with the *r* (rudimentary gene)

phenotype in *Drosophila melanogaster* (Wojcik et al. 1994). The *r* gene encodes the first three enzymatic activities of the pyrimidine biosynthetic pathway, and the *r* mutants are pyrimidine auxotrophs with a characteristic truncation of the wings. The mutation of *e(r)* enhanced the wing truncation in the *r* background.

The ER homologs (ERHs) are highly conserved proteins among vertebrates, invertebrates, and plants (Fig. 1A). The *Mus musculus* ERH protein (Kuwano et al. 1996), composed of 104 amino acids, is completely identical

Reprint requests to: Shigeyuki Yokoyama, Protein Research Group, Genomic Sciences Center, RIKEN Yokohama Institute, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan; e-mail: yokoyama@biochem.s.u-tokyo.ac.jp; fax: +81-45-503-9195.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.051484505>.

to the *Homosapiens* (Isomura et al. 1996) and *Xenopus laevis* (Pogge von Strandmann et al. 2001) ERH proteins, and shares 76% identity to the *D. melanogaster* ER protein. The murine ERH protein also shares 52% and 42% identities to the *Caenorhabditis elegans* and *Arabidopsis thaliana* ERH proteins, respectively (Gelsthorpe et al. 1997). The *Xenopus* ERH protein interacts with DCoH/PCD (dimerization cofactor of HNF1/pterin-4 α -carbinolamine dehydratase) and acts as a cell type-specific transcriptional

repressor, which probably interferes with HNF1-dependent gene regulation via DCoH/PCD (Pogge von Strandmann et al. 2001). To analyze the molecular functions of ERH in more detail, the three-dimensional structural information is needed. Here we report the crystal structure of murine ERH at 2.1 Å resolution, determined by the multiwavelength anomalous dispersion (MAD) method (Hendrickson 1991), and discuss the structural aspects of the ERH protein.

Figure 1. (A) Sequence alignment of the enhancer of rudimentary homolog (ERH) proteins. Mouse/human: ERH from *M. musculus*, *H. sapiens*, and *X. laevis* (identical); *Drosophila*: ERH from *D. melanogaster*; *C. elegans*: ERH from *C. elegans*; and *Arabidopsis*: ERH from *A. thaliana*. The alignment was generated by ESPript (Gouet et al. 1999) with CLUSTALW (Thompson et al. 1994). The secondary structures of the murine ERH protein, as determined by DSSP (Kabsch and Sander 1983), are shown above the sequences. (B) Ribbon representation of the murine ERH monomer (stereo view). The α -helices and the β -strands are shown in red and in yellow, respectively. (C) The superimposition of the main-chain structures of the three ERH monomers in the asymmetric unit (stereo view). Chains A, B, and C are colored red, blue, and green, respectively. The superimposition was carried out with the program lsqkab (Kabsch 1976) in CCP4.

Results and Discussion

The ERH crystal belongs to the *C*-centered monoclinic space group *C*2, with unit cell constants of $a = 85.25$ Å, $b = 43.26$ Å, $c = 92.06$ Å, $\alpha = 90.00^\circ$, $\beta = 99.95^\circ$, and $\gamma = 90.00^\circ$, and contains three protein molecules per asymmetric unit. The structure was refined to 2.1 Å by the MAD method. The crystallographic data are summarized in Table 1. The final model includes 295 amino acid residues of three ERH monomers, and 79 water molecules in the asymmetric unit. The C-terminal residues (chain A, 101~104; B, 99~104; C, 99~104) and the loop region of chain C (46~48) are invisible, due to disorder. The ERH monomer comprises a single domain consisting of an anti-parallel β -sheet (β 1, β 2, β 3, and β 4) and three α -helices (α 1, α 2, and α 3) (Fig. 1A,B).

Table 1. X-ray data collection, phasing, and refinement statistics

	Peak	Edge	Remote
<i>Data collection</i>			
Wavelength (Å)	0.97904	0.97939	0.96000
Resolution (Å)	50–2.1	50–2.1	50–2.1
Unique reflections	19,595	19,567	19,612
Redundancy	3.6 (3.2)	3.6 (3.3)	3.6 (3.3)
Completeness (%)	99.7 (99.2)	99.7 (99.2)	99.7 (98.6)
$I / \sigma(I)$	9.5 (4.6)	12.0 (4.6)	11.1 (4.6)
R_{sym}^a (%)	8.7 (27.1)	7.6 (26.7)	7.9 (27.0)
<i>MAD analysis</i>			
Resolution (Å)		24.8–2.1	
No. of Se sites ^b		11	
FOM _{MAD} ^c		0.55	
FOM _{RESOLVE} ^d		0.71	
<i>Refinement</i>			
Resolution (Å)		24.73–2.10	
No. of reflections		19,221	
No. of protein atoms		2456	
No. of water molecules		79	
R_{work} (%)		21.9	
R_{free} (%) ^e		27.2	
RMSD bond length (Å)		0.010	
RMSD bond angles (°)		1.6	
Average B factor (Å ²)		46.4	
<i>Ramachandran plot</i>			
Most favored regions (%)		87.3	
Additional allowed regions (%)		11.9	
Generously allowed regions (%)		0.7	
Disallowed regions (%)		0.0	

All numbers in parentheses represent last outer shell (2.18–2.10 Å) statistics. RMSD indicates root mean square deviation.

^a $R_{\text{sym}} = \sum |I_{\text{avg}} - I_i| / \sum I_i$, where I_i is the observed intensity and I_{avg} is the average intensity.

^b Number of selenium sites located with SOLVE.

^c Figure of merit after SOLVE phasing.

^d Figure of merit after RESOLVE.

^e R_{free} is calculated for 9.9% of randomly selected reflections excluded from refinement.

The residues of the hydrophobic core region are highly conserved among vertebrates, invertebrates, and a plant (eight identical and 14 similar residues out of 25 residues). Figure 1C shows the superimposition of the main-chain structures of the three ERH monomers in the asymmetric unit. The regions consisting of α -helices and β -strands superimpose completely, except for the α 2-helix. The loop regions at residues 10~17 (β 1- β 2 loop) and 44~53 (α 1- α 2 loop) and the C termini overlap slightly, suggesting that the structures of these regions are relatively flexible. When reviewing the structural homology search results of the ERH monomer by the programs DALI (Holm and Sander 1993), VAST (Madej et al. 1995), and MATRAS (Kawabata 2003), we did not find any structure that was significantly similar to the ERH monomer, suggesting that the ERH structure is a novel fold.

Figure 2A shows the crystal packing of the ERH protein chains. The chains B (blue) and C (green) in the asymmetric unit form a dimer through interactions between the β -sheet regions. The two A chains dimerize by the crystallographic twofold symmetry axis, similar to the dimerization of chains B and C. According to the analytical ultracentrifuge analysis, the molecular weight of ERH was ~26.5 kDa, indicating that the ERH protein forms a dimer in solution (the molecular weight of the ERH monomer is 12.8 kDa). The analytical ultracentrifugation results are consistent with the dimer formation of ERH proteins in the crystal. At the dimer interface between the β -sheet regions of the two molecules, there are hydrophobic interactions among Ile5, Leu7, Tyr19, Leu70, and Tyr79; electrostatic interactions between Arg17 and Asp21; and hydrogen bonds between the amido proton and the carbonyl oxygen of each Tyr79, suggesting that the dimer forms stably at this region (Fig. 2B). In addition, the residues of the dimer interface are significantly conserved among mouse/human, *D. melanogaster*, *C. elegans*, and *A. thaliana* (five identical and two similar residues out of seven) (Fig. 1A), suggesting that the dimer formation is strictly conserved and may be important for the functions.

The reported putative phosphorylation sites (Thr18 and Ser24) (Gelsthorpe et al. 1997) for casein kinase II (the consensus sequence, S/T X X X D/E) (Meggio et al. 1994) are shown in Figure 2C. The putative phosphorylation sites are located at the start of the β 2-strand (Thr18) and at the start of the α 1-helix (Ser24). Although the effect of the phosphorylation remains unclear, the phosphorylation might cause some structural changes, since there are possibilities of electrostatic repulsion between each phospho-Thr18 and between Glu23, phospho-Ser24, and Glu27.

Figure 2. (A) The crystal packing of the ERH protein chains (ribbon representations). Chains A, B, and C are colored red, blue, and green, respectively. (B) Ribbon representation of the interface of the ERH dimer (chains A and A'; stereo view). The residues at the dimer interface are shown in a stick model. (C) Ribbon representation of the ERH dimer (chains A and A'; stereo view). The putative phosphorylation sites for casein kinase II (Thr18 and Ser24) are shown as a blue stick model. (D) Electrostatic surface representation of the ERH dimer. Blue and red surfaces represent positive and negative potentials, respectively. The view is the same as in C. (E) Residue conservation mapping on the surface of the ERH dimer. Red and orange surfaces represent completely and highly conserved residues among mouse/human, *D. melanogaster*, *C. elegans*, and *A. thaliana*, respectively. The view is the same as in C. The negatively charged, highly conserved region, consisting of Asp21, Glu23, Glu27, and Glu30, is shown as a gray oval in D and in E.

The electrostatic potential distribution (Fig. 2D) and the highly conserved residues (Fig. 2E) on the solvent-accessible surface of the ERH dimer show the presence of charged and highly conserved regions. A significant negatively charged, conserved region, consisting of Asp21, Glu23, Glu27, and Glu30, may be important for the functions (Fig. 2D,E). In addition, the residues at the long loop region (44~53) are highly conserved (Fig. 1A), although the loop is structurally flexible (Fig. 1C), suggesting that it is probably essential for the functions, rather than the formation of the structure. One possibility is that this loop region is used for recognition by a binding partner protein. The crystal structure of ERH is

useful for further functional analyses of ERH, such as a mutational analysis. The present study will contribute toward understanding the molecular functions of ERH in the universal biological processes among higher eukaryotes.

Materials and methods

Protein expression and purification

The gene encoding full-length murine ERH, obtained from the FANTOM RIKEN full-length cDNA clones (Kawai et al. 2001; Okazaki et al. 2002), was cloned into the plasmid vector pCR2.1 (Invitrogen) as a fusion with an N-terminal His-tag

and a thrombin protease cleavage site. The selenomethionine (SeMet)-substituted ERH protein was synthesized by a cell-free protein expression system (Kigawa et al. 1999, 2004). The protein was first purified by a HisTrap column (Amersham Biosciences), and then the His-tag was cleaved by thrombin protease and removed by chromatography on a HisTrap column (Amersham Biosciences). Furthermore, the protein was purified by a HiTrap Q and Superdex 75 column (Amersham Biosciences) chromatography steps. The yield of the purified ERH protein was 14 mg per 27 mL of the cell-free synthesis reaction solution. The construct that was used for crystallization contained the cloning artifact sequence GSEGAAT at the N terminus.

Crystallization and data collection

The preliminary crystals of ERH (SeMet) were obtained under condition no. 42 (0.1 M Bis-Tris buffer at pH 5.5 containing 25% PEG 3350) of the Index crystal screening kit (Hampton Research) by the 96-well sitting drop vapor diffusion method. The crystals of ERH (SeMet) used for structure determination were obtained in drops composed of 1 μ L of 20.9 mg/mL protein solution (20 mM Tris-HCl buffer at pH 8.0 containing 150 mM NaCl, 2 mM DTT) and 1 μ L of reservoir solution (0.1 M Bis-Tris buffer at pH 4.6 containing 22% PEG 3350; Hampton Research) by the hanging drop vapor diffusion method against 500 μ L of the reservoir solution. Some clusters of plate-like crystals were obtained within a few days. A single crystal segment ($\sim 300 \times 300 \times 20 \mu\text{m}^3$) was isolated from the crystal cluster and used for data collection. The data collection was carried out at 100 K with the reservoir solution containing 18% PEG 400 as a cryoprotectant. The MAD data were collected at three different wavelengths at BL26B1 (Yamamoto et al. 2002) and SPring-8 (Harima), and were recorded on a Jupiter210 CCD detector (Rigaku). All diffraction data were processed with the HKL2000 program (Otwinowski and Minor 1997).

Structure determination and refinement

The program SOLVE (Terwilliger and Berendzen 1999) was used to locate the selenium sites and to calculate the phases, and RESOLVE was used for the density modification and partial model building (Terwilliger 2002). The rest of the model was built with the program O (Jones et al. 1991) and was refined with the programs Refmac5 (Murshudov et al. 1997) in CCP4 (Collaborative Computational Project 1994) and CNS (Brunger et al. 1998). Refinement statistics are presented in Table 1. The quality of the model was inspected by the program PROCHECK (Laskowski et al. 1993). Most of the graphic figures were created using the program PyMol (DeLano Scientific), except for the electrostatic surface representation, which was created with the program GRASP (Nicholls et al. 1991). The atomic coordinates and the structure factors have been deposited in the Protein Data Bank, with the accession code 1WZ7.

Acknowledgments

We thank H. Hamana, N. Ohbayashi, Y. Kamewari, and Y. Saito for protein preparations; M. Idaka, T. Wada, and C. Takemoto-Hori for their help in crystallization; S. Kamo for computer maintenance; and H. Wang, S. Kishishita, and

K. Murayama for their help in data collection. We are grateful to the Protein Preparation Screening Team members for construction of the expression vector, and K. Yajima and T. Nakayama for clerical assistance. We also thank M. Yamamoto and his colleagues for data collection at RIKEN Structural Genomics beamline BL26B1 at SPring-8. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI); the National Project on Protein Structural and Functional Analyses; and the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

- 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., et al. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**: 905–921.
- Collaborative Computational Project. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**: 760–763.
- Gelsthorpe, M., Pulumati, M., McCallum, C., Dang-Vu, K., and Tsubota, S.I. 1997. The putative cell cycle gene, enhancer of rudimentary, encodes a highly conserved protein found in plants and animals. *Gene* **186**: 189–195.
- Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. 1999. ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**: 305–308.
- Hendrickson, W.A. 1991. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**: 51–58.
- Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**: 123–138.
- Isomura, M., Okui, K., Fujiwara, T., Shin, S., and Nakamura, Y. 1996. Cloning and mapping of a novel human cDNA homologous to DROER, the enhancer of the *Drosophila melanogaster* rudimentary gene. *Genomics* **32**: 125–127.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**: 110–119.
- Kabsch, W. 1976. Solution for best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**: 922–923.
- Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**: 2577–2637.
- Kawabata, T. 2003. MATRAS: A program for protein 3D structure comparison. *Nucleic Acids Res.* **31**: 3367–3369.
- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., et al. 2001. Functional annotation of a full-length mouse cDNA collection. *Nature* **409**: 685–690.
- Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., and Yokoyama, S. 1999. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* **442**: 15–19.
- Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A., and Yokoyama, S. 2004. Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression. *J. Struct. Funct. Genomics* **5**: 63–68.
- Kuwano, Y., Kawamura, T., Sugahara, S., Watanabe, H., Ogata, K., and Abo, T. 1996. cDNA cloning of a novel mouse protein “Mer,” a homologue of enhancer of rudimentary gene of *Drosophila melanogaster*. *Biomed. Res.* **17**: 305–309.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. Procheck: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**: 283–291.
- Madej, T., Gibrat, J.F., and Bryant, S.H. 1995. Threading a database of protein cores. *Proteins* **23**: 356–369.
- Meggio, F., Marin, O., and Pinna, L.A. 1994. Substrate specificity of protein kinase CK2. *Cell. Mol. Biol. Res.* **40**: 401–409.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**: 240–255.

- Nicholls, A., Sharp, K.A., and Honig, B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**: 281–296.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**: 307–326.
- Pogge von Strandmann, E., Senkel, S., and Ryffel, G.U. 2001. ERH (enhancer of rudimentary homologue), a conserved factor identical between frog and human, is a transcriptional repressor. *Biol. Chem.* **382**: 1379–1385.
- Terwilliger, T.C. 2002. Automated structure solution, density modification and model building. *Acta Crystallogr. D Biol. Crystallogr.* **58**: 1937–1940.
- Terwilliger, T.C., and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 849–861.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Wojcik, E., Murphy, A.M., Fares, H., Dang-Vu, K., and Tsubota, S.I. 1994. Enhancer of rudimentaryp1, e(r)p1, a highly conserved enhancer of the rudimentary gene. *Genetics* **138**: 1163–1170.
- Yamamoto, M., Kumasaka, T., Ueno, G., Ida, K., Kanda, H., Miyano, M., and Ishikawa, T. 2002. RIKEN structural genomics beamlines at SPring-8. *Acta Crystallogr. A* **58**: C302.