Journal of Biomolecular NMR (2005) 32: 329–334
DOI 10.1007/s10858-005-7959-z

NMR structure note

Solution structure of the mouse enhancer of rudimentary protein reveals a novel fold

Hua Li[a], Makoto Inoue[a], Takashi Yabuki[a], Masaaki Aoki[a], Eiko Seki[a], Takayoshi Matsuda[a], Emi Nunokawa[a], Yoko Motoda[a], Atsuo Kobayashi[a], Takaho Terada[a,b], Mikako Shirouzu[a,b], Seizo Koshiba[a], Yi-Jan Lin[a], Peter Güntert[a], Harukazu Suzuki[a], Yoshihide Hayashizaki[a], Takanori Kigawa[a] & Shigeyuki Yokoyama[a,b,c,*]

[a]*RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, 230-0045, Japan;* [b]*RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo, 679-5148, Japan;* [c]*Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan*

Received 4 April 2005; Accepted 18 May 2005

Key words: enhancer of rudimentary, NMR structure, pyrimidine biosynthesis

Biological context

In most eukaryotes, the first three enzymatic activities of the pyrimidine biosynthesis pathway, glutamine-dependent carbamylphosphate synthetase (CPS), aspartate transcarbamylase (ATC) and dihydroorotase (DHO), are contained within a single polypeptide of 200–240 kDa. This polypeptide is encoded by the *CAD* gene in mammals, by the *rudimentary* gene, *r*, in *Drosophila melanogaster*, and by the *Pyr1-3* gene in *Dictyostelium discoideum*. These enzymatic activities play a central role in controlling the pyrimidine level in the cell. Thus, the *Drosophilar* mutants are pyrimidine auxotrophs, and they have a characteristic truncation of the wings. In mammals, this important enzymatic role has been demonstrated by the direct relationship between the expression of the *CAD* gene and cellular proliferation. Moreover, the correlation between the *CAD* gene activity and the cellular growth rate is dramatically apparent in the cases of tumors.

The *enhancer of rudimentary* (ER) gene was originally identified as interacting with the *r* gene by a genetic screen. It encodes a small protein, ER, which is evolutionarily highly conserved in organisms as diverse as vertebrates, invertebrates, and plants. The regulatory or enzymatic activity of ER has been implicated in pyrimidine biosynthesis and the cell cycle (Wojcik et al., 1994), but its molecular function remains unknown. In this study, we describe the solution structure of the mouse ER protein, determined by using heteronuclear NMR spectroscopy.

*To whom correspondence should be addressed. E-mail: yokoyama@biochem.s.u-tokyo.ac.jp

Methods and results

The gene encoding full-length murine ER, 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, Carlsbad, CA) as a fusion with an N-terminal His-tag and a thrombin protease cleavage site. The ^{13}C, ^{15}N-labeled ER protein was synthesized by the cell-free protein expression system described previously (Kigawa et al., 1999, 2004). For the NMR structure determination, a 1.5 mM sample of uniformly ^{13}C, ^{15}N-labeled ER protein was prepared in 20 mM sodium phosphate buffer (pH 6.0) containing 100 mM NaCl, 1 mM d-DTT, 0.02% NaN_3, and 10% ^2H_2O/90% ^1H_2O. The protein sample for the NMR measurements consisted of 111 amino acid residues. The first seven amino acid residues at the N-terminus (GSEGAAT)

were derived from the linker sequence used in the expression and purification system, while the remaining 104 amino acid residues encompass the full-length ER protein. The amino acid numbering in the article is consistent with that of the ER protein.

All of the NMR spectra were recorded at 25°C on a Bruker AVANCE 600 or 800 spectrometer equipped with a pulse-field gradient triple-resonance probe. Sequence-specific resonance assignments were made using the standard triple-resonance techniques (Wüthrich, 1986; Bax, 1994). The backbone assignment was achieved by the combined analysis of HNCO, HNCACO, HNCA, HNCOCA, CBCACONH, HNCACB and CCCONNH spectra. The aliphatic side chain resonances were identified by the combinational use of HCCCONNH, HCCH-TOCSY, HCCH-COSY and ^{15}N-NOESY-HSQC spectra. The aromatic ring resonances were assigned mainly by analyzing the HCCH-COSY and ^{13}C-NOESY-HSQC spectra in the aromatic region. All of the spectra were processed using the NMRPipe software package (Delaglio et al., 1995) and were analyzed with the program Kujira (version 0.913) (Kobayashi et al., personal communications).

The sequence-specific chemical shift assignments of the nearly complete ^{1}H, ^{15}N and ^{13}C resonances were obtained by applying the standard triple-resonance NMR measurements (Wüthrich, 1986; Bax, 1994). The five Pro residues in ER (Pro10, Pro14, Pro45, Pro48, and Pro81) are all in the *trans* conformation, as revealed by the chemical shift differences in C^{β} and C^{γ} (Schubert et al., 2002) and by the strong $d_{\alpha\delta}(i, i+1)$ NOEs (Wüthrich, 1986).

During the NMR measurements, we found that the linewidths of the NMR spectra of ER were broader than those of a protein with a similar molecular weight, indicating that ER behaved like an oligomeric protein. The ratio of the apparent molecular weight to the calculated molecular weight obtained in the gel filtration experiment was 2.6, showing that ER behaved as a dimer or a trimer in solution. In this report, we determined the structure of the ER monomer in solution.

In order to distinguish the intramonomeric NOEs from the intermonomeric NOEs, the ^{12}C-filtered and ^{13}C-edited 3D NOE spectra were measured with a sample of uniformly ^{13}C, ^{15}N-labeled ER as well as a sample of non-labeled ER and ^{13}C, ^{15}N-labeled ER mixed at a 1:1 ratio. However, because of the low signal-to-noise ratio, we could not obtain any unambiguous intermonomer NOEs. Then, ^{15}N-NOESY-HSQC spectra were measured using three kinds of samples as follows, (a) ^{15}N-labeled ER, (b) ^{2}H, ^{15}N-labeled ER, and (c) a mixture of ^{2}H, ^{15}N-labeled and non-labeled ER at a 1:1 molar ratio. In each sample, the concentration of the ER protein was 0.78 mM, and each sample was dissolved in 20 mM sodium phosphate buffer (pH 6.0) containing 100 mM NaCl, 1 mM d-DTT, 0.02% NaN_3, and 90% $^{1}H_2O$/10% $^{2}H_2O$. Ideally, if the deuteration rate is 100%, then no NOE peak should be observed for sample (b), and therefore, the NOE peaks appearing in sample (c) can all be regarded as the intermonomer NOEs (Walters et al., 1997). However, in fact, the actual deuteration rate for the sample was about 90%, so NOE peaks will be observed not only for sample (c) but also for sample (b). Since an increase in the NOE peak intensity was observed for sample (c), those NOE-related proton pairs could be attributed to the subunit interface. The results of the peak intensity comparison are shown in Figure 1a. While most of the peaks showed similar intensities in samples (b) and (c), an increase in the peak intensity was indeed observed for some of the NOE peaks. The residues with increased NOE peak intensities are indicated by the residue number and the amino acid type in Figure 1a. For example, the intensity of the NOEs between the amide proton of V8 and the aromatic ring protons (*e.g.*, HD1/2 and HE1/2) of Y19 was greatly increased in sample (c), as compared with that in sample (b) (Figure 1b). Besides the intensity increase of the NOE peaks, some NOE peaks newly appeared in sample (c), and were attributed to the subunit interface. In total, the residues of V8, R17, T18, Y19, L70, T78, Y79, Q80, and P81 gave either a peak intensity increase or the appearance of a new peak, indicating that these residues potentially form the subunit interface.

NOE restraints were obtained from the ^{13}C-edited NOESY-HSQC and ^{15}N-edited NOESY-HSQC spectra with the mixing time of 80 ms. The NOEs of the subunit interface identified above were excluded from the set of restraints used in the structure calculation. Dihedral angle restraints were derived using the program TALOS (Cornilescu et al., 1999). The stereo-specific signal assignments were determined for 12 Val and Leu residues when the two methyl groups were distinguishable from their NOE patterns. Automated

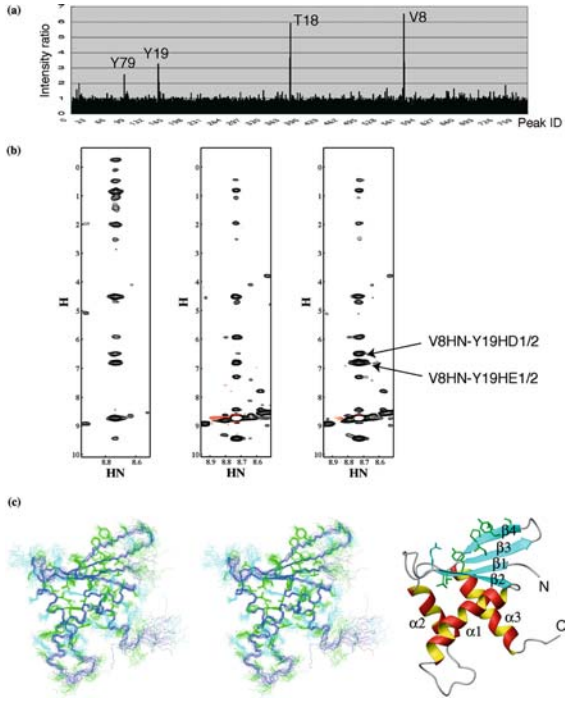

Figure 1. Identification of the subunit interface and the structure of the ER monomer. (a) Plot of the peak intensity ratio *versus* the peak ID. The residues with large intensity ratios (*e.g.*, V8, T18, Y19, Y79) are labeled. (b) ^{15}N-NOESY-HSQC spectra strips of the amide group of V8 in samples: ^{15}N-labeled ER (left), ^{2}H, ^{15}N-labeled ER (middle), and the mixture of ^{2}H, ^{15}N-labeled and non-labeled ER at a 1:1 molar ratio (right). NOE peaks between the amide proton of V8 and the aromatic ring protons, HD1/2 and HE1/2, of Y19 are indicated. (c) Stereo view of the wire models, illustrating the ensemble of the 20 structures with the lowest CYANA target function (left and middle), and the ribbon diagram of ER (right). In the wire model, the backbone is shown in blue, the heavy atoms of the side chains of the charged residues (Arg, Lys, Glu, Asp) are shown in light blue, and the heavy atoms of the side chains of the other residues are shown in green. In the ribbon model, the α-helices and the β-strands are depicted in red and cyan, respectively. The heavy atoms of the side chains of residues V8, R17, T18, Y19, L70, T78, Y79, Q80, and P81, which may be involved in the putative subunit interface, are indicated. The figures were drawn with MOLMOL (Koradi et al., 1996).

NOE cross-peak assignments (Herrmann et al., 2002) and structure calculations were performed with torsion angle dynamics (Güntert et al., 1997) using the CYANA software package (version 1.0.7). The 20 conformers with the lowest target function in cycle7 of CYANA were chosen. The programs PROCHECK-NMR (Laskowski et al., 1996) and MOLMOL (Koradi et al., 1996) were used to validate and to visualize the final structures, respectively. The statistics of the structures, as well as the distance and torsion angle constraints used for the structure calculation, are summarized in Table 1.

The NMR structure of ER is composed of a four-stranded antiparallel β-sheet, in the strand order 2-1-3-4 (β1: 4-8, β2: 18–22, β3: 68–73, β4: 78–82), and three α-helices (α1: 25–41, α2: 54–63, α3: 84–96) packed against the same side of the sheet (Figure 1c). The overall topology is $\beta1\beta2\alpha1\alpha2\beta3\beta4\alpha3$. The secondary structural elements were identified on the basis of the chemical shift indices (Wishart et al., 1992, 1994) and the NOE patterns. Within the ER structure, it is interesting that the residues (V8, R17, T18, Y19, L70, T78, Y79, Q80, and P81), which may potentially form the subunit interface, are all located on the four-stranded β-sheet (Figure 1c, right), suggesting that ER forms a multimer by using the β-sheet as an interface.

Table 1. Summary of conformational restraints and statistics of the final 20 best structures

NOE upper distance limits	
Total	1608
Intra residue and sequential ($\|i-j\| \leq 1$)	901
Medium range (1)	355
Long range ($\|i-j\| \geq 5$)	352
Torsion angle restraints	121
Stereo assignments	12
CYANA target function value	0.81
Distance restraint violations	
Number > 0.1 Å	10
Number > 0.3 Å	0
Maximum	0.21
Torsion angle restraint violations	
Number > 5°	0
Maximum	0.99
PROCHECK Ramachandran plot analysis	
(Res. 4–43, 53–98)	
Residues in favored regions	85.2%
Residues in additionally allowed regions	14.0%
Residues in generously allowed regions	0.8%
Residues in disallowed regions	0.0%
RMS deviation to the averaged coordinates	
All regions (Res. 4–98)	
Backbone atoms	0.50
Heavy atoms	0.88
Ordered regions	
Backbone atoms	0.36
Heavy atoms	0.76

Discussion and conclusions

The ER protein is evolutionarily highly conserved in plants and animals (Figure 2a). More than half of the residues are well conserved, especially those located at the secondary structure regions. The highly conserved residues of Y36, L40, Y52, and F60 form the hydrophobic contact between α1 and α2 (Figure 2b, upper right). The highly conserved K88 electrostatically interacts with the highly conserved D62 and stacks with the conserved aromatic F58, and both interactions serve as important contacts between α2 and α3 (Figure 2b, lower right). The conserved residues of Y82 and W86 stack their aromatic rings between β4 and α3 (Figure 2b, lower left). Besides the conserved residues located in the hydrophobic core, it is interesting that many of the highly and intermediately conserved residues (*e. g.*, H3, I5, L7, Q9, R17, Y19, D21, D66, S68, L70, Y79, and P81) exist on the four-stranded β-sheet. Furthermore, most of these residues are hydrophobic, but appear on the monomer surface (Figure 2c). As described above, the residues of V8, R17, T18, Y19, L70, T78, Y79, Q80, and P81 may be involved in the subunit interface. Most of the residues overlap very well, implying that the subunit interface formation by the β-sheet may be important for the function of ER, although it remains unknown whether ER forms a multimer *in vivo*.

The gel filtration experiment revealed that ER behaves as a dimer or a trimer in solution. In order to determine the exact oligomeric state of ER, first we prepared a ^{13}C, ^{15}N-labeled ER mutant, R13G (designated as ER(R13G)*). Next, we prepared a series of mixed samples composed of ER(R13G)* and non-labeled ER (designated as ER(wt)) at different molar ratios. Conceptually, in each mixed sample, several kinds of multimeric states with different compositions of ER(wt) and ER(R13G)* would appear at different molar ratios when the chemical exchange between the monomer and the multimer reaches equilibrium. The concept of this oligomeric state identification procedure is shown in Figure 3a. In Figure 3b, portions of the ^{15}N-HSQC spectra are shown for a series of samples, where the Y19 residue giving the large mutation-induced chemical shift is highlighted to show the spectral changes that occur with the different mixing ratios. When ER(wt) was added into ER(R13G)*, the single peak of Y19 in ER(R13G)* split into two peaks, which represented the two species of dimer, ER(R13G)*/ER(R13G)* and ER(R13G)*/ER(wt), respectively. Therefore, ER behaves as a homodimer in solution. The fact that

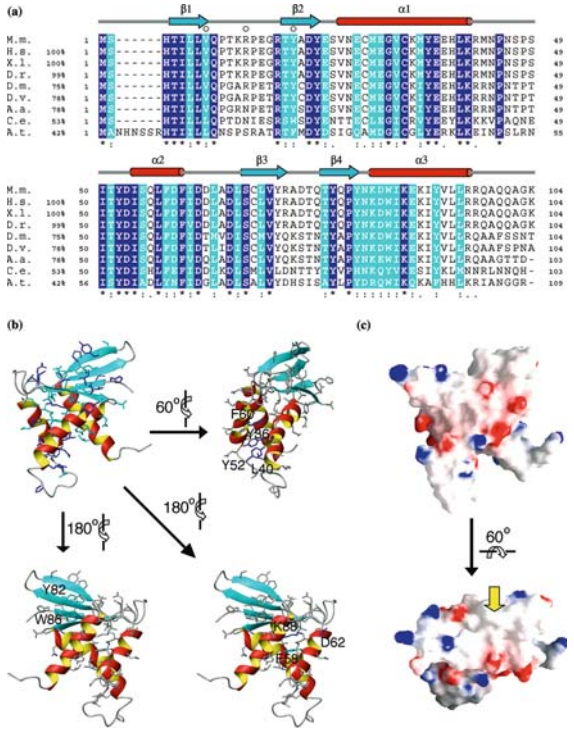

Figure 2. The structural features of ER. (a) Multiple sequence alignment of ER proteins from various species. Secondary structural elements of murine ER, defined by NMR spectroscopy, are shown above the sequence. The sequence identity between the ER proteins from other species and the mouse ER protein is indicated. The V8, R13, and Y19 residues, which are discussed in the text, are labeled with 'o'. The highly conserved and intermediately conserved amino acid residues are highlighted in blue and light blue, respectively. The abbreviations of the species and the accession numbers (Genbank or SWISS-PROT) are: M. m., *Mus musculus*, AAH83141; H. s., *Homo sapiens*, AAH14301; X. l., *Xenopus laevis*, AAF28892; D. r., *Danio rerio*, AAH59528; D. m., *Drosophila melanogaster*, Q24337; D. v., *Drosophila virilis*, Q94554; A. a., *Aedes aegypti*, Q93104; C. e., *Caenorhabditis elegans*, Q22640; A. t., AAC49667, *Arabidopsis thaliana*. (b) The highly and intermediately conserved residues are mapped on the ribbon model of ER, using the same color codes as in panel A (upper left). In the other sub-panels, the residues discussed in the text are highlighted in blue or light blue, according to their conservation, while the other residues are colored grey. (c) The electrostatic potential surface of ER is shown in the same direction as in Figure 1c (upper). To emphasize the hydrophobic patch (indicated by a yellow arrow) formed by the four-stranded β-sheet, the electrostatic potential surface of ER was rotated by 60° along the *x*-axis (bottom). The figures were drawn with GRASP (Nicholls et al., 1991).

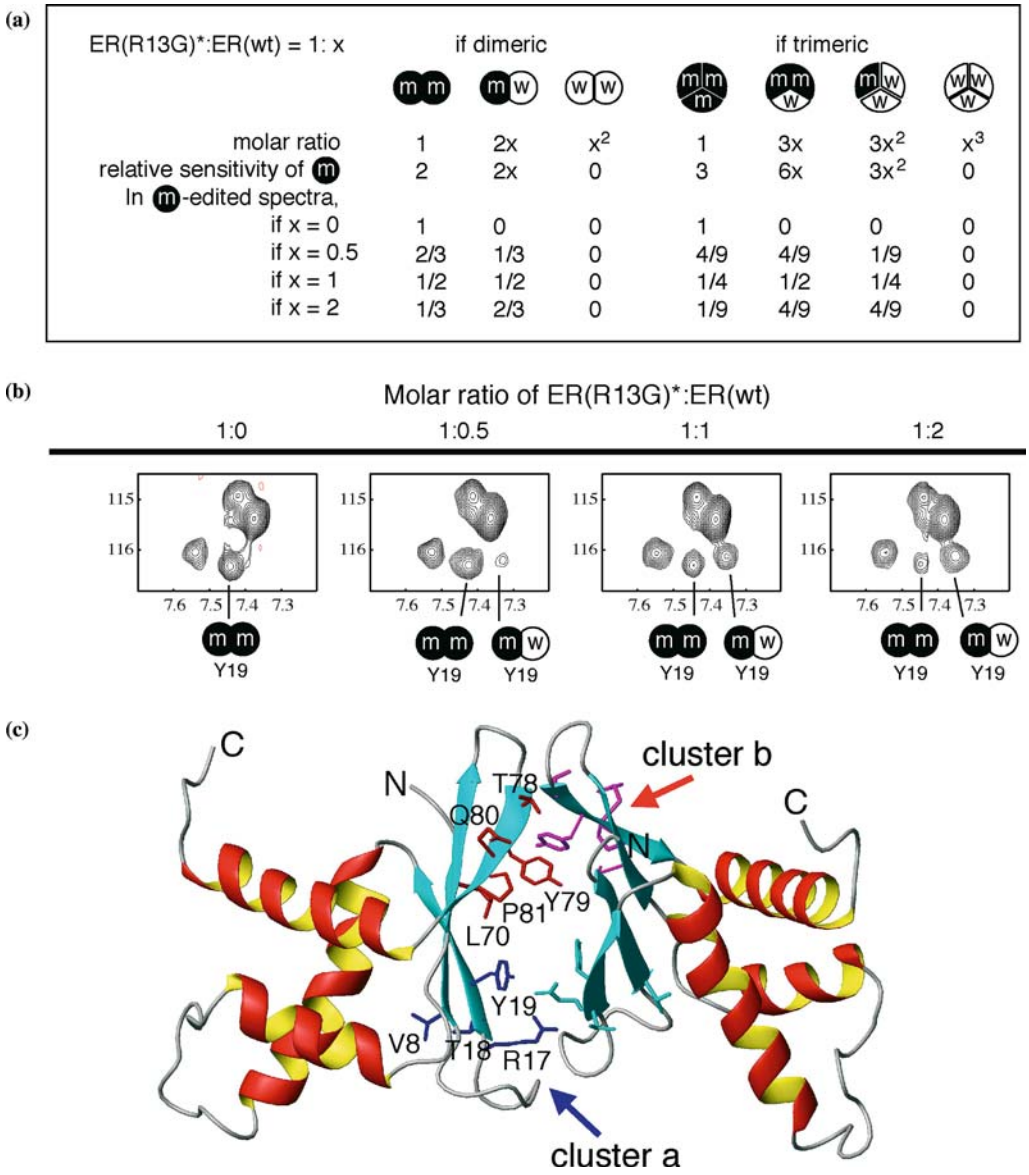

Figure 3. ER behaves as a homodimer in solution. (a) Concept of the identification procedure for the oligomeric state of ER. Here 'm' in the black background represents ER(R13G)*, and 'w' in the white background represents ER(wt). (b) Portions of the ^{15}N-HSQC spectra of the series of samples with different mixing ratios. The Y19 residue, which undergoes a large mutation-induced chemical shift, is highlighted to show the spectral changes with the different mixing ratios. The two split peaks, representing the two species of dimer ER(R13G)*/ER(R13G)* and ER(R13G)*/ER(wt), respectively, were observed in each sample mixture. (c) A proposed dimer model for ER. The two ER monomer structures were handled manually, according to the identified intermolecular NOEs. The intermolecular NOEs could be classified into two clusters: cluster a consists of the NOEs related to V8, R17, T18, and Y19, shown in blue and cyan, respectively, in each monomer, and cluster b consists of the NOEs related to L70, T78, Y79, Q80, and P81, shown in red and magenta, respectively. The intermolecular NOEs were only observed within the same cluster. Therefore, ER probably behaves as a homodimer in solution, using the β-sheet as an interface in a head-to-head fashion.

ER forms a dimer in solution is also supported by the result of our recent analytical ultracentrifugation experiment (Arai et al., 2005). What does the ER dimer model look like? We found that the intermolecular NOEs could be classified into two clusters: cluster a consists of the NOEs related to V8, R17, T18, and Y19, and cluster b consists of the NOEs related to L70, T78, Y79, Q80, and P81

(Figure 3c). The intermolecular NOEs were only observed within the same cluster. Therefore, ER probably behaves as a dimer in solution using the β-sheet as an interface in a head-to-head fashion.

A search for structural homologs, using the coordinates of the monomeric ER structure on the DALI server (Holm and Sander, 1996), showed that no protein shared significant structural similarity with ER (Z-scores ranging from 3.1 to 2.0). The three proteins with the highest Z-scores (>3.0) were the photosystem I subunit Psad from cyanobacteria ($Z=3.1$, PDB ID: 1jb0, sequence identity 12%), Cre recombinase from Bacteriophage p1 ($Z=3.0$, PDB ID: 4cre, sequence identity 8%) and serine/threonine phosphatase 2C from human ($Z=3.0$, PDB ID: 1a6q, sequence identity 6%). In these proteins, the regions sharing structural similarity with ER are either a subdomain of a large protein or a partial region of a domain. Therefore, the DALI search results revealed that ER forms a novel fold.

In summary, the ER protein, which is highly conserved among organisms as diverse as vertebrates, invertebrates, and plants, has been implicated as functioning in pyrimidine biosynthesis and the cell cycle. In this study, ER was found to behave as a dimer in solution, and the solution structure of the ER monomer was determined by heteronuclear NMR spectroscopy. The ER monomer consists of a four-stranded antiparallel β-sheet, with a strand order of $\beta2\beta1\beta3\beta4$, and three α-helices ($\alpha1$, $\alpha2$, and $\alpha3$) packed against one side of the sheet, with an overall topology of $\beta1\ \beta2\alpha1\alpha2\beta3\beta4\alpha3$. A structural homology search revealed that ER forms a novel fold. These structural features of ER will shed light on its functional mechanism at the molecular level.

The coordinates have been deposited in the Protein Data Bank (accession code 1WWQ).

Acknowledgements

The authors thank Yukiko Fujikura, Natsuko Matsuda, Miyuki Saito, Yukako Miyata, Masaomi Ikari, Fumiko Hiroyasu, Yasuko Tomo, Megumi Watanabe, Miyuki Sato, Satoko Yasuda, Hiroshi Hirota, and Mayumi Yoshida for help with sample preparations and maintaining the NMR facility. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Arai, R., Kukimoto-Niino, M., Uda-Tochio, H., Morita, S., Uchikubo-Kamo, T., Akasaka, R., Etou, Y., Hayashizaki, Y., Kigawa, T., Terada, T. et al. (2005) *Protein Sci.*, **14,** 1888–1893.

Bax, A. (1994) *Curr. Opin. Struct. Biol.*, **4**, 738–744.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Herrmann, T., Güntert, P. and Wüthrich, K. (2002) *J. Mol. Biol.*, **319**, 209–227.

Holm, L. and Sander, C. (1996) *Science*, **273**, 595–602.

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y. and Konno, H. et al. (2001), *Nature*, **409**, 685–690.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. and Yokoyama, S. (1999) *FEBS Lett.*, **442**, 15–19.

Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A. and Yokoyama, S. (2004) *J. Struct. Funct. Genom.*, **5**, 63–68.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–55.

Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptain, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.

Nicholls, A., Sharp, K.A. and Honig, B. (1991) *Proteins: Struct. Funct. Genet.*, **11**, 281–296.

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R. and Suzuki, H. et al. (2002), *Nature*, **420**, 563–573.

Schubert, M., Labudde, D., Oschkinat, H. and Schmieder, P. (2002) *J. Biomol. NMR*, **24**, 149–154.

Walters, K.J., Matsuo, H. and Wagner, G. (1997) *J. Am. Chem. Soc.*, **119**, 5958–5959.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.

Wojcik, E., Murphy, A.M., Fares, H., Dang-Vu, K. and Tsubota, S.I. (1994) *Genetics*, **138**, 1163–1170.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* Wiley, New York, NY.