Structure of the Conserved Transcriptional Repressor Enhancer of Rudimentary Homolog[†,‡]

Cheng Wan,[§] Wolfram Tempel,[∥] Zhi-Jie Liu,[∥] Bi-Cheng Wang,[∥] and Robert B. Rose*[,§]

Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina 27695-7622, and Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602-7229

Received October 16, 2004; Revised Manuscript Received December 28, 2004

ABSTRACT: erh (enhancer of rudimentary homolog) is a ubiquitously expressed transcriptional coregulator that is highly conserved among eukaryotes, from humans to plants to protozoa. Functions attributed to erh include enhancement of pyrimidine biosynthesis, a role in cell cycle regulation, and repression of the tissue-specific transcription factor HNF-1 (hepatocyte nuclear factor-1) through binding the coactivator DCoH (dimerization cofactor of HNF1). No homologous sequences, other than erh orthologs, have been identified, and little is known about the interactions of erh. To further elucidate its function, we determined the crystal structure of erh to 2.0 Å resolution. The erh structure is a novel $\alpha + \beta$ fold consisting of a four-stranded antiparallel β sheet with three amphipathic α helices situated on one face of the β sheet. Structure-based searches of the Protein Data Bank, like sequence-based searches, failed to identify paralogs. We present structural and biochemical evidence that erh functions as a dimer. The dimer interface consists of a β sandwich composed of the β sheet from each monomer. Many of the surface residues of erh are conserved, including patches of hydrophobic and charged residues, suggesting protein−protein interaction interfaces. Two putative CKII phosphorylation sites are highly ordered in the structure and are predicted to disrupt dimerization and protein−protein interactions.

erh[1] (enhancer of rudimentary homolog) is a highly conserved 104-amino acid protein: the frog and human proteins are identical in sequence and differ from zebra fish erh by a single amino acid (*1*, *2*). Expression profile studies indicate that erh is over-expressed in rapidly dividing cells such as carcinoma cell lines, is enriched in ectodermally derived tissue during development, and is ubiquitously expressed in the adult (*1*, *3*−*5*). Despite its pervasive presence, little is known of the function of this protein.

erh localizes in the cytoplasm and nucleus (*1*). It has been characterized as an enhancer and repressor of transcription in a number of different contexts. erh was initially identified in a genetic screen in *Drosophila* as a mutation that augmented the phenotype of rudimentary gene mutations, observed as wing truncations (*5*). The protein was proposed to be an enhancer of the rudimentary gene, which encodes the first three enzymatic activities of the pyrimidine biosynthetic pathway. Because of the importance of the rudimentary gene during development, it was suggested that the *Drosophila* enhancer of rudimentary (DROER) protein played a developmental role. A later study demonstrated that erh is a maternal factor in *Xenopus* that is expressed developmentally in ectodermally derived tissues such as the eyes, spinal cord, and parts of the brain, supporting the development role of erh (*1*). This study demonstrated that erh could function as a repressor in a cell-type specific manner when fused to the GAL4 DNA binding domain. erh was further proposed to regulate the cell cycle, being expressed at higher levels in proliferating cells than nondividing cells (*2*). erh is phosphorylated by CKII (casein kinase II), a cell cycle regulated kinase (*2*).

The protein interaction partners of erh are mostly unknown. The one reported interaction was identified by a yeast two-hybrid screen of the coactivator DCoH (dimerization cofactor of HNF1) (*1*). In a transient transfection assay, erh appeared to repress the coactivator function of DCoH, preventing it from enhancing the activity of the transcription factor HNF-1 (hepatocyte nuclear factor-1). Other binding partners are predicted since erh is expressed more widely than DCoH or HNF-1.

As noted in the first report of DROER, erh is highly conserved with no close sequence paralogs (*5*). The lack of sequence similarity to proteins of known structure has prevented structure prediction through homology modeling. We therefore determined the structure of erh as a step in understanding its function. Although erh is a small protein, the structure indicates that it folds into a unique dimer with no known structural homologues.

[†] This work was funded by an American Diabetes Association Junior Faculty Award 7-03-JF-34.

[‡] Coordinates and structure factors for erh were deposited in the Protein Data Bank under entry code 1W9G.

* To whom correspondence should be addressed. Phone: (919) 513-4191. Fax: (919) 515-2047. E-mail: bob_rose@ncsu.edu.

[§] North Carolina State University.

[∥] University of Georgia.

[1] Abbreviations: erh, enhancer of rudimentary homolog; CE, combinatorial extension; CKII, casein kinase II; DCoH, dimerization cofactor of HNF-1; HNF-1, hepatocyte nuclear factor-1; DROER, *Drosophila* enhancer of rudimentary; LB, Luria−Bertani media; PDB, protein data bank; SDS−PAGE, sodium dodecyl sulphate−polyacrylamide gel electrophoresis; SAD, single wavelength anomalous dispersion; SERCAT, Southeast Regional Collaborative Access Team.

10.1021/bi047785w CCC: $30.25
Published on Web 03/03/2005

Table 1: Data Collection, Phasing, and Refinement Statistics

space group unit cell lengths (Å)	$P2_1$ $a = 45.36$, $b = 62.50$, $c = 48.45$, $\beta = 117.92$	
data collection	SAD	high resolution
resolution range (Å)	42.9−2.84	42.9−2
highest resolution bin (Å)	3.11−2.84	2.11−2.0
observations[a]	33252	115950 (15847)
unique reflections[a]	5616	15585 (2180)
completeness[a] (%)	98.3 (98.0)	95.3 (93.9)
Rsym(I)[a,b] (%)	7.3 (13.1)	11.0 (71.2)
Wilson *B*-factor (Å^2)		22.4
R_{anom}[a,c] (%)	4	
$\langle I\rangle/\langle\sigma(I)\rangle$[a]	13.2 (3.0)	16.9 (2.8)
FOM[a,d,e]	0.34	
anomalous signal/noise[e]	0.94	
refinement statistics		
resolution range (Å)		43.03−2.00
protein atoms in asu		1595
water molecules		104
no. of reflections for R_{free}[f]		765
R-factor/R_{free} (%)		21.1/25.9
average *B*-factor (Å^2)		28.4
residues in most/least favored Ramachandran regions (%)		94.5/0.0
RMS deviations		
bonds (Å)		0.02
angles (deg)		1.28

[a] Numbers in parantheses refer to data in the highest resolution shell. [b] $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$. [c] $R_{anom} = \Sigma|\langle I^+\rangle - \langle I^-\rangle|/\Sigma\langle I\rangle$. [d] FOM (figure of merit). [e] From SOLVE. [f] Based on 5% reflections.

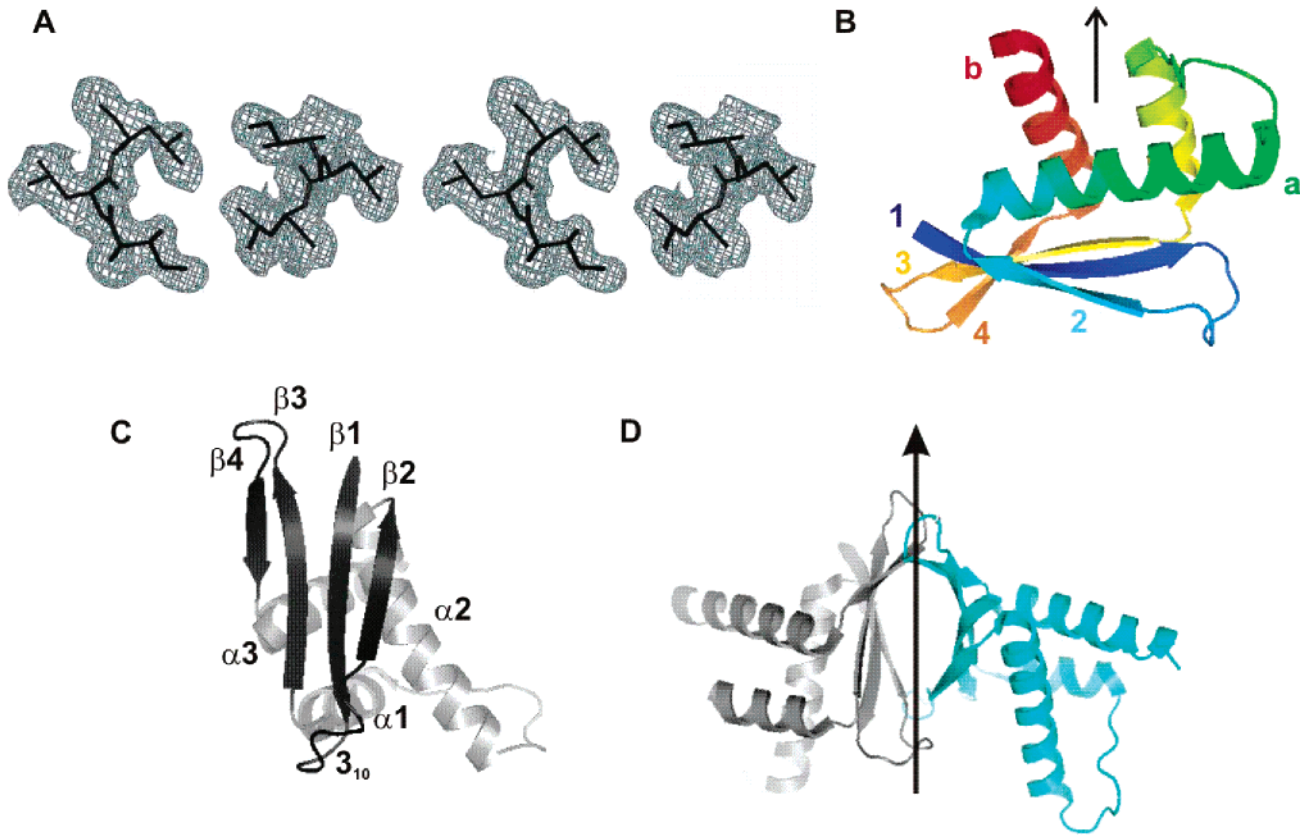

FIGURE 1: Structure of erh. Panel A shows a stereodiagram of a $F_o - F_c$ simulated anneal omit map to 2.0 Å resolution generated with residues Ile5, Leu6, and Leu7 omitted from each monomer at the dimer interface. Simulated annealing was carried out in CNS, and the temperature was raised to 2500 K to eliminate model bias in the map. A stick representation of the final model is shown in the density (wall-eyed stereo). Panel B shows a ribbon diagram of the erh monomer, colored by residue number from blue (N-terminus) to red (C-terminus). The arrow indicates the pseudo-2-fold axis, superimposing β-strands 1−2 onto strands 3−4 (numbered). The attached helices (labeled a and b, respectively) are in different orientations. Panel C shows the topology of the monomer, as viewed from the β-sheet: β1-3$_{10}$-β2-α1-α2-β3-β4-α3. Panel D shows a ribbon diagram of the erh dimer. The monomers (cyan and black) are related by a 2-fold axis (arrow) that is almost coincident with the crystallographic *x*-axis.

MATERIALS AND METHODS

Sample Preparation. cDNA of the human enhancer of rudimentary homolog was obtained from the Mammalian Gene Collection (MGC) through the Lawrence Livermore National Laboratory (http://image.llnl.gov), IMAGE number 3507241. The cDNA was subcloned for overexpression into pET24b (Novagen) designed as a thrombin-cleavable 6-His-N-terminal fusion. *Escherichia coli* BL21(DE3) (Novagen) transfected cells were grown in LB (Luria−Bertani) media, induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside, and harvested after 6 h. Cell pellets were lysed in 50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole; bound to Ni−NTA resin (Ni^{2+} chelated nitrilotriacetate resin) (Amersham Pharmacia); and eluted with 250 mM imidazole, 300 NaCl, and 1 mM β-mercaptoethanol. The His-tag was removed with thrombin (Sigma) according to the manufacturer's protocol. The protein was further purified by size-exclusion chromatography with a Sephacryl S100 column (Amersham Pharmacia) and concentrated to 11 mg/mL using a centrifugal filter device (Millipore).

Structure Determination. Initial crystallization conditions were identified using a sparse matrix screen (Hampton Research) (*6*). The final optimized conditions were 10% 2-propanol, 20% PEG4000, and 0.1 M HEPES (pH 7.5). Prism-shaped crystals grew after 48 h by the hanging drop vapor diffusion method to about 0.3 mm in the longest dimension. Crystals were frozen in liquid nitrogen after a quick soak in cryoprotectant composed of 5% 2-propanol, 20% PEG4000, 5% PEG400, and 0.1 M HEPES (pH 7.5). Selenomethionine-incorporated protein was generated for phase determination by the Single Anomalous Dispersion (SAD) method. The erh-containing plasmid was expressed in the methionine auxotroph B834 (DE3) (Novagen) in the presence of selenomethionine (Fisher). Mass spectrometry indicated 100% selenomethionine substitution of the four methionine sites of erh.

Data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory at 0.9793 Å (near the selenium anomalous peak) (Table 1). The data were processed using the ELVES interface (*7*). The data were indexed with MOSFLM (*8*) and scaled and merged with SCALA keeping Bijvoet reflections separate (*9*, *10*). *F* values were calculated with Truncate (*11*).

The initial phases were determined by the single wavelength anomalous scattering (SAD) method (*12*). Using the SAD mode of the program SOLVE (*13*), and diffraction data to 2.8 Å resolution, eight (out of eight expected) putative selenium atoms were located in the asymmetric unit (Table 1). Initial phases were improved by density modification with RESOLVE (*14*), setting the solvent content to 52%, consistent with the presence of a dimer in the asymmetric unit. Noncrystallographic symmetry operators were obtained from the partial model output by RESOLVE. Phases were then extended to 2.0 Å using the program DM (*15*) in a combination of averaging and solvent flattening modes. Further automated model building with the program ARP/wARP (*16*) was followed by manual model adjustments using XFIT (*17*). Final rounds of refinement were accomplished using CNS (*18*) and refmac5 (*19*, *20*) alternated with manual model building using the program O (*21*). CNS was also used to calculate the buried surface area.

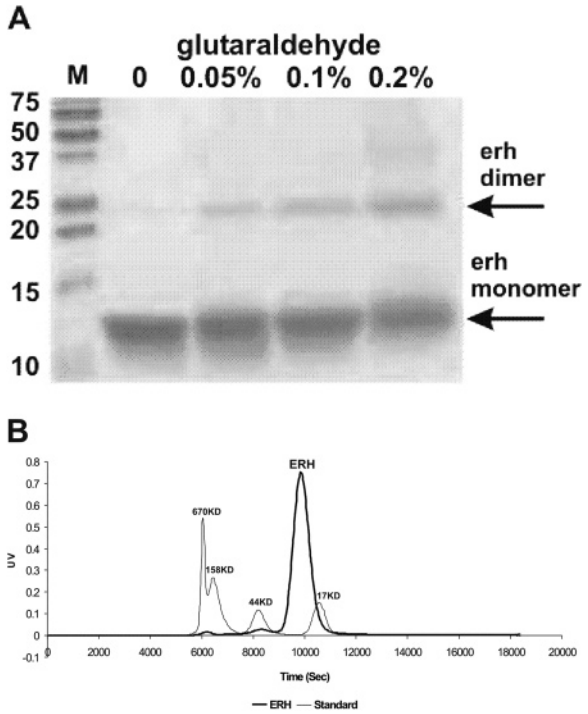

FIGURE 2: erh is a dimer in solution. Panel A shows the results of glutaraldehyde cross-linking of erh. Purified erh was reacted with varying concentrations of glutaraldehyde as indicated and analyzed by SDS−PAGE. Molecular weight markers (Precision Plus, Biorad) are shown in the first lane (M), labeled in units of kDa. The arrowheads indicate the monomer and dimer of erh. Panel B shows a chromatogram of erh after gel filtration. The chromatogram of the protein standards (Biorad) was superimposed on the chromatogram of erh for comparison. The molecular weights of the peaks from the standards are marked. The retention time of erh indicates it is a dimer in solution; the molecular weight of an erh monomer is 12 kDa.

erh orthologs were identified with PSI BLAST searching with the human erh sequence (*22*). Sequences representing 15 distinct eukaryotic genera were identified and aligned with CLUSTAL W (*23*): *Aedes aegypti*, *Anopheles gambiae*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Danio rerio*, *Dictyostelium discoideum*, *Drosophila melanogaster*, *Echinococcus multilocularis*, *Homo sapiens*, *Mus musculus*, *Oryza sativa*, *Plasmodium falciparum*, *Rattus norvegicus*, soybean, and *Xenopus laevis*.

Chemical Cross-Linking and Gel Filtration. For glutaraldehyde cross-linking, a 0.04 mg/mL solution of purified erh was incubated with varying concentrations of glutaraldehyde in PBS (phosphate buffered saline) on ice. After 1 h, the solution was quenched with the addition of a final concentration of 100 mM glycine at pH 8.0. The mixture was kept on ice for 30 min and analyzed with a 15% SDS−PAGE gel. For gel filtration, the retention time of erh was measured with a Hi-prep 26/60 Sephacryl S100 column (Pharmacia) in PBS and compared with gel filtration standards (Biorad).

RESULTS AND DISCUSSION

Structure of erh. We report the high-resolution crystal structure of human erh. Although the data could be indexed in the space group *C*222, the high R_{sym} values indicated a lower symmetry space group, and the data were processed

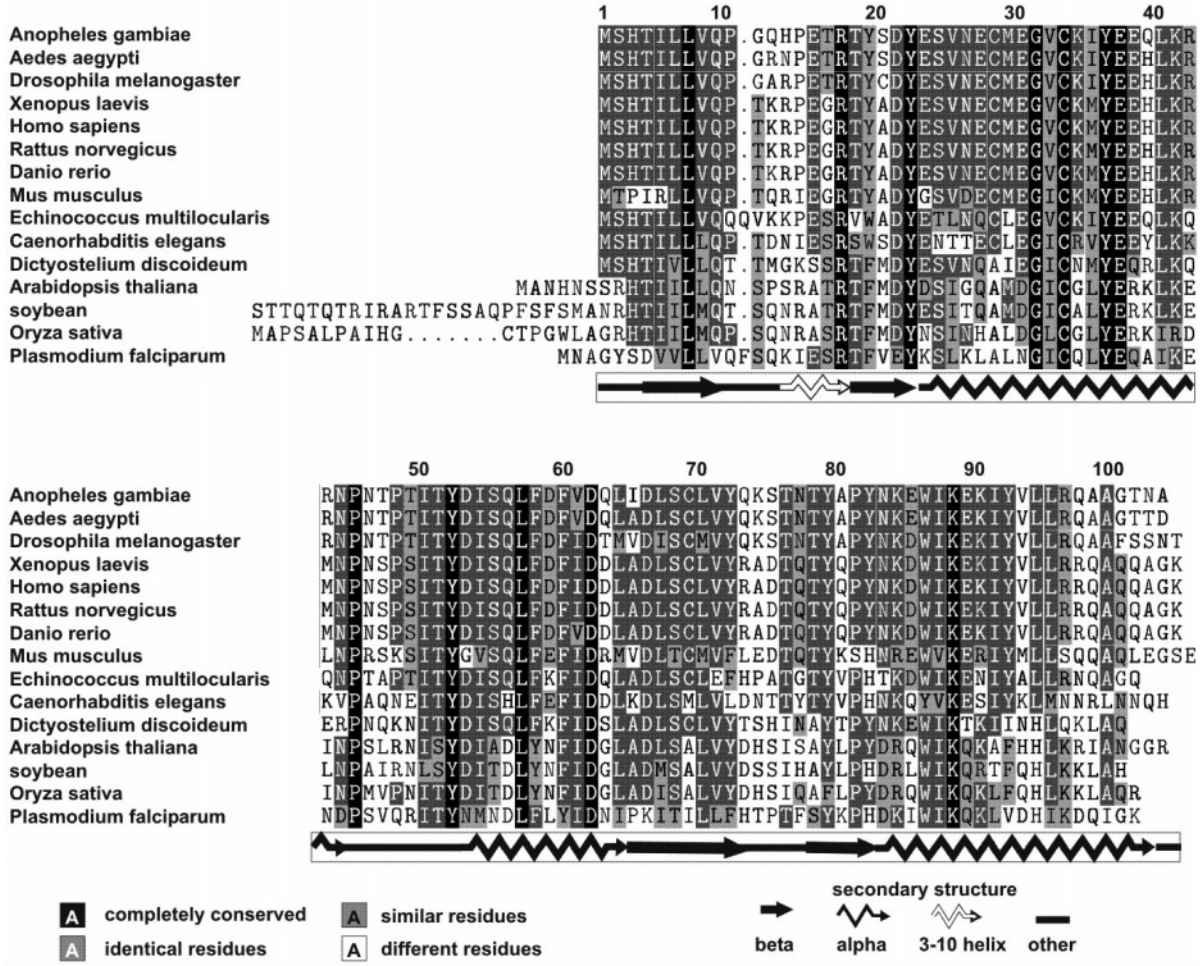

FIGURE 3: Alignment of erh sequences. The erh sequences were identified from a PSI-BLAST search of the nonredundant database and aligned with ClustalW. The alignment was drawn with Textshade (*41*) (http://workbench.sdsc.edu) with a similarity threshold of 50% using the default similarities (FYW, IVLM, RK, DE, GA, TS, and NQ). The sequence is numbered according to human erh. The secondary structure is depicted below the sequence and was derived from the determined tertiary structure.

in space group $P2_1$. With two monomers in the asymmetric unit, the resulting Matthews coefficient (V_m) was 2.5 Å^3/Da and 51% solvent content (*24*). The final refined model gave R/R_{free} values of 0.211/0.259 for all data to 2.0 Å (Table 1 and Figure 1A). Noncrystallographic symmetry restraints were not used during refinement because they increased the *R*-factor and did not improve the quality of the density. The root-mean-square deviation between Cα atoms of the two monomers in the final model is 0.26 Å. The final model includes residues 1−102 for each monomer and 104 water molecules. Residues 103 and 104 of each monomer are disordered, as are three residues of the loop between helix 1 and helix 2 of one monomer (Pro48B, Ser49B, and Ile50B; B corresponds to the monomer with segid B in the PDB file). This loop is the most mobile region of the protein, with an average *B*-factor of 46.7 Å^2 for the Cα atoms (residues 45A−52A, monomer with segid A), as compared with an average of 27.5 Å^2 for the Cα atoms of the overall protein dimer. The side chains of residues Arg42A and B, Asn46B, Ser49A, Ile50A, and T51B in the mobile loop and Lys12A and B, Glu23B, Lys34B, Lys41A and B, Asp75B, Gln77A and B, Gln100A and B, and Gln101A and B were disordered and omitted from the final model.

The erh structure is a novel α + β fold consisting of a four-stranded antiparallel β sheet with three amphipathic α helices situated on one face of the β sheet (Figure 1B) (*25*). The topology of the sheet is 2−1−3−4 (Figure 1C) with two α helices situated between β strands 2 and 3. The α1 helix packs against strands 1 and 2 with its axis nearly parallel to the strands, reminiscent of a zinc finger. The axes of the other two helices are situated perpendicular to the β sheet strands. The final α helix is at the C-terminus. The monomer is pseudo-symmetric, with β1−β2 superimposing on β3−β4 with an rms deviation of 1.5 Å for the backbone atoms and with one helix (α1 and α3, respectively) following each set of β strands, although in different orientations (Figure 1B). There is minimal sequence identity between the two β−β−α motifs.

Both faces of the β sheet contain hydrophobic residues, suggesting the outer β sheet surface is involved in protein−protein interactions (see Figure 4). erh is a dimer in solution, as indicated by glutaraldehyde cross-linking and size-exclusion chromatography (Figure 2). The asymmetric unit of the crystal contains the most obvious erh dimer (Figure 1D). The dimer interface consists of a β sandwich, with a 4-stranded β sheet contributed from each monomer. The overall topology of the sandwich is 2−1−3−4−4′−3′−1′−2′ (primed and unprimed strands designate different monomers). The hydrogen-bonding network between adjacent dimer-related antiparallel strands is not extensive: between strand 4 of each monomer, two main-chain hydrogen bonds are formed (between the carbonyl and nitrogen atoms of Tyr

79 of each monomer). There are no main-chain−main-chain interactions between strands 2 and 2′. Instead, the carbonyl oxygen of Arg17 is hydrogen bonded to the side-chain hydroxyl of Thr18′, the carbonyl oxygen of Tyr19 may hydrogen bond with the Nϵ of Arg17′, and the carbonyl oxygen of Leu7 interacts with the hydroxyl group of Tyr19′ through an ordered water molecule. The primary interactions between monomers result from van der Waals contacts between the hydrophobic side chains of residues Ile5, Leu7, Leu70, and Pro81 and the hydrophobic regions of Arg17 and Tyr19. A total of 1375 Å^2 of surface area is buried between monomers in the dimer, less buried surface than expected for a typical homodimer, which has been estimated to be 1685 Å^2 (*26*). In addition, there is a salt bridge interaction between Arg17 and Asp21′ and a hydrogen bond between the side chain hydroxyls of Thr78 and Thr78′.

The overall fold of erh is not represented in the Protein Data Bank (*27*). Sequence analysis by multiple prediction methods, as implemented by the Polish Bioinformatics Web Site Meta-Server (http://bioinfo.pl/ (*28*)), failed to identify close homologues, other than erh orthologs. These methods included 1-D sequence recognition, for example, with Blast (*22*), and 3-D fold recognition, for example, with 3D-PSSM (*29*), GenTHREADER (*30*, *31*), and Pmodeler (*32*). Searches of known structures with the DALI (http://www.ebi.ac.uk/dali/ (*33*)) and VAST (http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml (*34*)) servers did not identify significant structural homologues. We also examined the top 10 solutions identified through the University of Glasgow Topology of Protein Structure (TOPS) website (http://www.tops.leeds.ac.uk/Results.html (*35*)) and the top three solutions from the Combinatorial Extension (CE) method (http://cl.sdsc.edu/ce.html (*36*)).

The highest scoring alignments identified through searches with the erh structure were large structures that aligned to pieces of erh: VAST identified photosystem I (PDB 1jbo) overlapping 40 residues (score 5.7, *P*-value 0.03), and DALI identified phosphatase 2c (PDB 1a6q) overlapping 57 amino acids (*Z*-score 3.7). CE identified the α helical protein retinoblastoma tumor suppressor domain A (PDB 1A06) and the α + *β* protein DNA primase/helicase chain A (PDB 1CR4), overlapping 71−73 residues (*Z*-score 3.9 with an rmsd of 4.8−4.9 Å). The topology of erh is similar to PsaD (DALI, *Z*-score 3.4, rmsd of 3.5 Å for 58 amino acids), a 138 amino acid protein that is a stromal subdomain of cyanobacterial photosystem I (PDB 1jb0 chain D). A number of proteins identified through DALI and Meta-Server overlapped the *β*-*β*-α motif, with the helix axis oriented parallel to the *β* strands. No homologues of erh were identified with the two helices perpendicular to the *β* strands. Searching with the monomer or dimer of erh resulted in the same set of structures. Searching with the eight-stranded dimer interface of erh using DALI identified the seven-stranded *β* barrel in the ku70 monomer as the closest match (PDB 1jeq, *Z*-score 2.3). Although the topologies differ and the ku70 strands form a more continuous barrel, the two domains overlap with an rmsd of 3.9 Å for 56 Cα residues.

Evolutionary Conservation. erh is highly conserved among eukaryotes, as has been noted previously (Figure 3) (*2*). A PSI BLAST search starting with the human erh sequence converged after two cycles and identified erh sequences from 15 genera with an expectation value of 8×10^{-7} or less

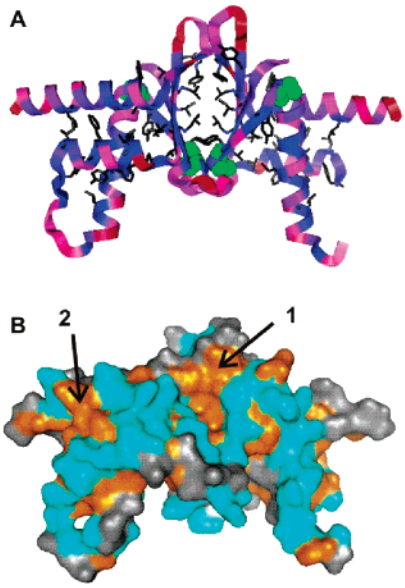

FIGURE 4: erh is evolutionarily highly conserved in eukaryotes. Panel A shows a ribbon representation of the human erh dimer colored by conservation among 15 orthologs identified by PSI-BLAST (*22*), from red (least conserved) to magenta to blue (most conserved). One sequence from each identified genus was aligned by CLUSTALW (*23*). The score of the most highly conserved similarity group at each position was used, according to the similarity groupings: DE, NQ, ND, EQ, KR, AS, YW, IVLM, and ST. Hydrophobic side chains are displayed of residues conserved in 13 or more sequences (hydrophobic residues: VLIFMWY). The sites phosphorylated by casein kinase II, Thr18 and Ser24, are represented by green van der Waals surfaces. (The gap in one of the monomers is due to disorder of three residues in the loop between helices 1 and 2.) Panel B shows the solvent-accessible (Conolly) surface of erh colored by conservation among animal sequences: orange, hydrophobic (amino acids Val, Phe, Leu, Ile, Trp, Met, Tyr, Ala, and Pro); cyan, hydrophilic (the other 11 amino acids); and gray, not conserved. Conservation is defined as similarity among at least eight of the nine animal sequences at each position. 1 and 2 mark hydrophobic cavities formed at the dimer interface (1) or between the parallel helices (2).

(*22*). Figure 4A indicates the distribution of the conserved residues on the erh structure. Conserved hydrophobic residues are found in three regions of the protein: (1) core residues within erh monomers: between helices and between the helices and *β* sheet; (2) at the dimer interface, between *β* sheets of each monomer; and (3) surface residues. Many of the surface residues are conserved within the animal kingdom, from tapeworms to humans (Figure 4B). A number of conserved hydrophobic residues are solvent-exposed, for example, in the cavity generated between the *β* sandwich interface and on the surface of helices 2 and 3. The axes of these helices are parallel and may represent a site for interactions with other proteins, in particular, DCoH. The interaction between DCoH and HNF-1, which erh disrupts (*1*), consists of two antiparallel helices from the dimer domain of HNF-1 contacting two antiparallel helices from DCoH (*37*). Helices 2 and 3 of erh superimpose well with the helices of HNF-1 and may compete with HNF-1 for binding to DCoH.

Phosphorylation Sites. The cell-cycle regulator casein kinase II (CKII) was shown to phosphorylate *Drosophila* erh in vitro at two consensus sequences, S/T X X D/E: at Thr18 and Ser24 (*2*). Figure 4A indicates that Thr18 is located at the beginning of *β* strand 2, while Ser24 is located

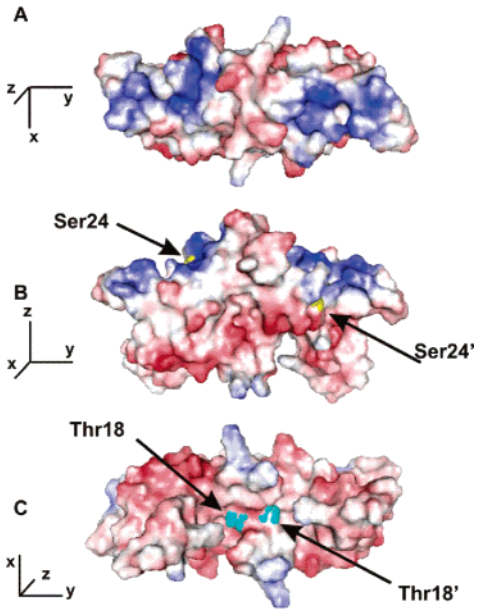

FIGURE 5: Electrostatic potential surface of the erh dimer. Three perpendicular views of the solvent-accessible surface of erh are shown, colored according to electrostatic potential: blue is positive, red is negative, and white is uncharged. An orthogonal coordinate system in each panel indicates the relative orientation of the three views: all rotations are 90° around the *y*-axis. Panel A shows that the top surface is primarily positively charged. Panel B shows that the side surface is divided into three zones: positive, uncharged, and negative. The CKII phosphorylation sites at Ser24 (yellow) are situated on the edge of the positive region. Panel C shows the bottom surface is primarily negatively charged. The CKII phosphorylation sites at Thr18 (cyan) are located at the bottom of a deep negative pocket. The electrostatic potential was generated with the APBS plugin of PyMol (version 0.96). Because a number of charged residues were disordered in the X-ray structure, all side chains were added to the model of erh in favorable rotamers before the electrostatic potential was calculated.

at the beginning of helix 1. Both sites are highly conserved among all erh sequences and are solvent exposed. The acidic residue of the CKII consensus sequence for Thr18 (Asp/Glu21) is similarly conserved, while the acidic residue for the Ser24 consensus sequence (Glu27) is conserved in the animal kingdom but not in *Dictyostelium*, *Plasmodium*, or plants. Gelsthorpe et al. proposed that phosphorylation at these sites altered the secondary structure of erh (*2*). Both putative phosphorylation sites are highly ordered in the structure. As described previously, the γ hydroxyl of Thr18 forms a hydrogen bond across the dimer interface with the carbonyl oxygen of Arg17′. The hydroxyl substituents of the dimer-related Thr18 residues are 6 Å apart. Therefore, phosphorylation at this site may disrupt dimerization. Ser24 caps the amino terminus of helix 1 by hydrogen bonding with the backbone nitrogen of Glu27. CKII has been shown to phosphorylate other Ser residues in the N-cap position (*38*). Phosphorylation of N-cap sites stabilize the α-helix (*26*, *39*).

The surface charge potential of erh is not evenly distributed, with positively charged, neutral, and negatively charged bands that may influence binding interactions (Figure 5). Overall, erh is negatively charged at physiological pH, with a predicted pI of 5.6. While Ser18 is located in a negative region of the surface (Figure 5C), Ser24 is situated in a region that is neutral or positively charged. Therefore, phosphorylation of Ser24 might disrupt charge−charge interactions associated with protein binding.

Other sites are predicted to be good phosphorylation sites by the NetPhos server, a neural-network-based prediction approach (http://us.expasy.org/tools/ (*40*)). Ser47, Thr11, Tyr19, Tyr22, and Tyr92 score well above the threshold (threshold of 0.4, scores of 0.74 or better). Tyr22 and Tyr92 are among five conserved Tyr residues of the animal erh sequences. Human erh contains eight Tyr residues, all of which are solvent-exposed. This suggests the possibility that other sites of erh are phosphorylated.

CONCLUSIONS

The structure of erh reveals an as-yet uncharacterized fold, despite the small size of the protein. Structure-based searches of the Protein Data Bank, like previously reported sequence-based searches, failed to identify evolutionarily related proteins that might provide clues of the function of erh.

The structure of the erh monomer consists of a four-stranded antiparallel β sheet with three helices on one surface. We present structural and biochemical evidence that erh is a functional dimer. The dimer interface forms a β sandwich with a hydrophobic core. Much of the surface of erh is conserved. The structure highlights likely binding surfaces, including a hydrophobic cavity situated at the dimer interface and a set of helices with conserved solvent-exposed hydrophobic residues. The charge distribution on the surface of erh is clustered into positive and negative bands. Further studies will be required to determine how erh interacts with DCoH, as well as other binding partners.

ACKNOWLEDGMENT

X-ray data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID (or 22-BM) beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at 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 W-31-109-Eng-38. We thank Dr. Michael Goshe for his assistance with mass spectrometry.

REFERENCES

1. 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*, 1379−85.
2. 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*, 189−95.
3. 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*, 125−7.
4. Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R., and Hogenesch, J. B. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes, *Proc. Natl. Acad. Sci. U.S.A. 101*, 6062−7.
5. Wojcik, E., Murphy, A. M., Fares, H., Dang-Vu, K., and Tsubota, S. I. (1994) Enhancer of rudimentary p1, e(r)p1, a highly conserved enhancer of the rudimentary gene, *Genetics 138*, 1163−70.
6. Jancarik, J., and Kim, S. H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins, *J. Appl. Crystallogr. 24*, 409−11.

7. Holton, J., and Alber, T. (2004) Automated protein crystal structure determination using ELVES, *Proc. Natl. Acad. Sci. U.S.A. 101*, 1537−42.
8. Leslie, A., Brick, P., and Wonacott, A. (1986) *Daresbury Lab. Information Quart. Protein Crystallogr. 18*, 33−39.
9. Collaborative Computation Project (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr. D50*, 760−3.
10. Evans, P. (1993) *Proceedings of CCP4 Study Weekend on Data Collection & Processing*, pp 114−122, CLRC Daresbury Laboratory, http://www.ccp4.ac.uk/courses/procs.html.
11. French, G., and Wilson, K. (1978) On the treatment of negative intensity observations, *Acta Crystallogr. A34*, 517.
12. Wang, B. C. (1985) Resolution of phase ambiguity in macromolecular crystallography, *Methods Enzymol. 115*, 90−112.
13. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr., Sect. D 55* (Pt 4), 849−61.
14. Terwilliger, T. C. (2002) Automated structure solution, density modification, and model building, *Acta Crystallogr., Sect. D 58*, 1937−40.
15. Cowtan, K. D., and Zhang, K. Y. (1999) Density modification for macromolecular phase improvement, *Prog. Biophys. Mol. Biol. 72*, 245−70.
16. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458−63.
17. McRee, D. E. (1999) XtalView/Xfit−A versatile program for manipulating atomic coordinates and electron density, *J. Struct. Biol. 125*, 156−65.
18. Brunger, A., Adams, P., Clore, G., Delano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N., Read, R., Rice, L., Simonson, T., and Warren, G. (1998) Crystallography and NMR system (CNS): a new software system for macromolecular structure determination, *Acta Crystallogr., Sect. D 54*, 905−21.
19. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite, *Acta Crystallogr., Sect. D 59*, 1131−7.
20. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S., and Dodson, E. J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT, *Acta Crystallogr. Sect. D 55* (Pt 1), 247−55.
21. Jones, T., Zhou, J., Cowan, S., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr., Sect. A 47*, 110−9.
22. Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−402.
23. 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*, 4673−80.
24. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491−7.
25. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol. 247*, 536−40.
26. Jones, S., and Thornton, J. M. (1995) Protein−protein interactions: a review of protein dimer structures, *Prog. Biophys. Mol. Biol. 63*, 31−65.
27. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235−42.
28. Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003) 3D-Jury: a simple approach to improve protein structure predictions, *Bioinformatics 19*, 1015−8.
29. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM, *J. Mol. Biol. 299*, 499−520.
30. Jones, D. T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences, *J. Mol. Biol. 287*, 797−815.
31. McGuffin, L. J., and Jones, D. T. (2003) Improvement of the GenTHREADER method for genomic fold recognition, *Bioinformatics 19*, 874−81.
32. Wallner, B., Fang, H., and Elofsson, A. (2003) Automatic consensus-based fold recognition using Pcons, ProQ, and Pmodeller, *Proteins 53 Suppl 6*, 534−41.
33. Holm, L., and Sander, C. (1995) Dali: a network tool for protein structure comparison, *Trends Biochem. Sci. 20*, 478−80.
34. Gibrat, J. F., Madej, T., and Bryant, S. H. (1996) Surprising similarities in structure comparison, *Curr. Opin. Struct. Biol. 6*, 377−85.
35. Gilbert, D., Westhead, D., Nagano, N., and Thornton, J. (1999) Motif-based searching in TOPS protein topology databases, *Bioinformatics 15*, 317−26.
36. Shindyalov, I. N., and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, *Protein Eng. 11*, 739−47.
37. Rose, R. B., Bayle, J. H., Endrizzi, J. A., Cronk, J. D., Crabtree, G. R., and Alber, T. (2000) Structural basis of dimerization, coactivator recognition, and MODY3 mutations in HNF-1alpha, *Nat. Struct. Biol. 7*, 744−8.
38. Brownlie, P., Ceska, T., Lamers, M., Romier, C., Stier, G., Teo, H., and Suck, D. (1997) The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control, *Structure 5*, 509−20.
39. Andrew, C. D., Warwicker, J., Jones, G. R., and Doig, A. J. (2002) Effect of phosphorylation on α-helix stability as a function of position, *Biochemistry 41*, 1897−905.
40. Blom, N., Gammeltoft, S., and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, *J. Mol. Biol. 294*, 1351−62.
41. Beitz, E. (2000) TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e, *Bioinformatics 16*, 135−9.

BI047785W