Solution Structure of YKR049C, a Putative Redox Protein from *Saccharomyces cerevisiae*

Jin-Won Jung†, Adelinda Yee‡, Bin Wu‡, Cheryl H. Arrowsmith‡ and Weontae Lee†,§,*

†Department of Biochemistry and Protein Network Research Center, Yonsei University, 134 Shinchon-Dong Seodaemoon-Gu, Seoul 120-749, Korea
‡Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, ON, M5G 2M9, Canada
§Northeast Structural Genomics Consortium

Received 12 May 2005, Accepted 8 June 2005

YKR049C is a mitochondrial protein in *Saccharomyces cerevisiae* that is conserved among yeast species, including *Candida albicans*. However, no biological function for YKR049C has been ascribed based on its primary sequence information. In the present study, NMR spectroscopy was used to determine the putative biological function of YKR049C based on its solution structure. YKR049C shows a well-defined thioredoxin fold with a unique insertion of helices between two β-strands. The central β-sheet divides the protein into two parts; a unique face and a conserved face. The 'unique face' is located between β2 and β3. Interestingly, the sequences most conserved among YKR049C families are found on this 'unique face', which incorporates L109 to E114. The side chains of these conserved residues interact with residues on the helical region with a stretch of hydrophobic surface. A putative active site composed by two short helices and a single Cys97 was also well observed. Our findings suggest that YKR049C is a redox protein with a thioredoxin fold containing a single active cysteine.

**Keywords:** NMR spectroscopy, *Saccharomyces cerevisiae*, Structural proteomics, Thioredoxin fold

**Materials and Methods**

**Protein purification** The hypothetical ORF of YKR049C from *Saccharomyces cerevisiae* was sub-cloned into the pET15b expression vector with an N-terminal His tag. The protein was expressed in *E. coli* strain BL21(DE3) grown in M9-minimal medium supplemented with 8-Nammonium chloride (1 g/L) and 13C-glucose (1 g/L) and the protein was purified using standard immobilized metal affinity chromatography and polished by gel filtration using Superdex-75. The concentrations of protein sample obtained ranged from 1.0 mM to 1.5 mM in aqueous solution.
NMR spectroscopy  The majority of NMR spectra were collected at 25°C on Varian Inova 600 and 600 MHz spectrometers equipped with pulsed field gradient triple-resonance probes. Chemical shifts were referenced to internal DSS. All NMR data were processed using the NMRPipe program (Delaglio et al., 1995) and analyzed using SPARKY (Goddard and Kneller). Backbone resonance assignments were made using HNCO, CBCA(CO)NH, HNCACB and HNCA spectra. The initial assignment was made using AUTOASSIGN (Zimmerman et al., 1997) and confirmed manually. Aliphatic side chain assignments were made using SPARKY (Goddard and Kneller). Backbone resonance assignments were made using HNCO, CBCA(CO)NH, HNCACB and HNCA spectra. The initial assignment was made using AUTOASSIGN (Zimmerman et al., 1997) and confirmed manually. Aliphatic side chain assignments were made using SPARKY (Goddard and Kneller). Chemical shifts were referenced to internal DSS. All NMR data were processed using the NMRPipe program (Delaglio et al., 1995) and analyzed using SPARKY (Goddard and Kneller). Backbone resonance assignments were made using HNCO, CBCA(CO)NH, HNCACB and HNCA spectra. The initial assignment was made using AUTOASSIGN (Zimmerman et al., 1997) and confirmed manually. Aliphatic side chain assignments were made using SPARKY (Goddard and Kneller). Backbone resonance assignments were made using HNCO, CBCA(CO)NH, HNCACB and HNCA spectra. The initial assignment was made using AUTOASSIGN (Zimmerman et al., 1997) and confirmed manually. Aliphatic side chain assignments were made using SPARKY (Goddard and Kneller).

Structure calculation  Distance restraints for structure calculations were derived from cross-peaks in simultaneous CN-NOESY and 13C-edited NOESY (τm = 150 ms) spectra. Slowly exchanging amide protons were identified by acquiring a series of 15N-H HSQC spectra after dissolving lyophilized protein into 100% D2O containing 25 mM sodium phosphate (pH 6.5), 150 mM NaCl, 1 mM TDDT, 90% H2O/10% D2O.

NMR resonance assignments and structure determination  All backbone resonance assignments of YKR049C were completed except for the N-terminal 10 residues due to resonance overlap or absence of amide proton signals. The 1H-15N HSQC spectrum of the protein without a His-tag was acquired and compared with that of uncleaved sample, but showed no significant differences (data not shown).

Since YKR049C is found in mitochondria, the N-terminal region of YKR049C might be a mitochondria localizing signal sequence. Moreover, because we were not able to assign complete NMR resonances to the N-terminal region due to resonance overlap, the TALOS-predicted dihedral angle constraints of some assigned fragments could only give structural information. However, we believe that the signal peptide could have α-helical structure because the N-terminal sequence had the angle constraints of an α-helical structure (Fig 1). The predicted values of W4 and K5 appeared on α-helix region. T6 and L7 did not produce the predicted result. The most proper result from TALOS is comprised of 10 predicted values in the same region, but T6 and L7 only showed eight or nine predicted values for the α-helix region and one or two values just outside the α-helix.

**Fig. 1.** Sequence alignment of YKR049C with homologous proteins. Sequences were obtained from the Swiss Protein Database or from the trEMBL, and have the following accession numbers: *Saccharomyces cerevisiae* (P36141), *Candida glabrata* (Q6EJ19), *Kluyveromyces lactis* (Q6C5N7), *Ermosohalium gossypii* (Q755M5). Secondary structures derived using NMR data are also shown. The bottom line shows dihedral constraints predicted by TALOS. The red triangles are α-helix constraints and the cyan triangles β-strand constraints. Open triangles indicate a low relevance prediction.
region. These could be accepted as indicators of an α-helix after careful inspection (Cornilescu et al., 1999). Hence, we assume that the N-terminal portion has an α-helical structure.

The NMR structures of YKR049C were calculated using CYANA and further refined using CNS with the water shell refinement protocol. The mean RMSD value for the backbone atoms in the structural region (11-132) was defined as 0.91Å and 1.17Å for all heavy atoms, respectively (Fig. 2A).

Solution structure of YKR049C

The overall structure of YKR049C shows a thioredoxin fold, which is a distinct structural motif found among proteins that interact with cysteine-containing substrates (Martin, 1995). All helices are located at opposite sides of the central β-sheet (Fig. 2B). The anti-parallel β-sheet consists of four β-strands (β1-β4). β3 and β4 form a standard β-hairpin structure. This β-sheet divides the protein structure into two parts. One is comprised of two long parallel α-helices, H1 and H6, which are anchored by the central β-sheet that has a typical thioredoxin fold topology. We refer to this region as the ‘common face’ because its structure is common among thioredoxin fold and there is no uniqueness of YKR049C family in comparison with other thioredoxin fold proteins. The remainder of YKR049C has four helices, H2, H3, H4, and H5, which are located on the other side of the thioredoxin fold from common face. We refer to this region as the ‘unique face’ because residues that compose this region are conserved among YKR049C homologues and unique against other thioredoxin fold proteins.

The solution structure of YKR049C shows a hydrophobic surface, which is conserved among thioredoxins, that is mainly comprised of H1 and H6 (Fig. 3A). The orientations of these two helices are well conserved in the thioredoxin fold. The most conserved sequence stretches from L109 to E114, which is found in the central core region (Fig. 1). This stretch expands to the middle of the β-sheet, and it makes hydrophobic interactions with helices (Fig. 3B). Therefore, we believe that these residues are likely to have been conserved because they retain a common molecular topology for YKR049C homologues.

Structure-function of YKR049C

The fold of YKR049C was determined to be a thioredoxin fold based on a DALI (Holm and Sander, 1993) search. Proteins with similar topology, namely, Ure2, arsenate reductase, glutaredoxin, elongation factor 1-γ fragment, and glutaredoxin-like protein...
were compared. All have a typical thioredoxin fold with a similar molecular topology; moreover, they all have redox-related molecular functions. The CXXC motif has been reported to be conserved among redox family proteins, and is located between β1 and α1. Moreover, the redox proteins usually possess active residues at this location (Martin, 1995). However, although YKR049C has a thioredoxin fold, no residues corresponding to the CXXC motif were found, except Cys30, which is not conserved among homologue proteins based on available sequence information (Fig. 1). Instead, Cys97, which is located on a loop comprised of β2 and α2, is well conserved among the YKR049C family. The two cysteine residues of the CXXC motif in thioredoxin proteins are easily oxidized to form an intra-molecular disulfide bond between two cysteines. A previous report suggested that these residues in glutaredoxin proteins are essential for redox-related activity. Moreover, it was proposed that ArsC is a non-chromosomal protein that possesses arsenate reductase activity (Martin et al., 2001). ArsC reduces arsenate to arsenite using reduced glutathione. Although ArsC protein has a thioredoxin fold, it possesses a single cysteine residue at its active site (Fig. 4). In the presence of arsenate, a single active cysteine easily forms a thiarsahydoxyl intermediate. Glutaredoxin1 enzyme converts this intermediate to free cysteine residue and an arsenite molecule by GSH. Therefore, ArsC protein could efficiently catalyze the reduction of arsenate with a single cysteine at its active site. Prx6 provides another example of a single active cysteine protein. Peroxiredoxin is an enzyme that catalyzes the reduction of H2O2, yielding two H2O molecules. It usually has two active cysteine residues to counteract oxidative stress. Although most peroxiredoxin proteins have two active cysteines, Prx6 has only one active cysteine at its active site, which is sufficient to reduce H2O2.

The structure of YKR049C shows that its putative active site is composed of two short helices, H4 and H5. These two helices wrap around a single Cys97 residue to form a small active site pocket (Fig. 3B). Another unique feature of YKR049C is a small hydrophobic patch formed by a number of hydrophobic residues, which might serve as a binding surface for other proteins. Therefore, we propose that
YKR049C is a redox protein and that it could be involved in the reduction of small toxic molecules.

Acknowledgments This research was supported by the Korean Science and Engineering Foundation (KOSEF) through the NRL program of MOST NRDP (MI-0200-00-0020), by the Northeast Structural Genomics Consortium (NIH PSI grant P50 GM26413), and by the Ontario Research and Development Challenge Fund and Genome Canada.

References


