Journal of Biochemistry and Molecular Biology, Vol. 38, No. 5, September 2005, pp. 550-554

JOURNAL OF Biochemistry and Molecular Biology

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

Introduction

The primary objective of structural genomics/proteomics is to elucidate as many structures as possible to better understand the relationship between protein sequence, 3D structure, and function (Kim, 1998; Yee *et al.*, 2002). Two approaches may be used to archive these goals. One involves the identification of new protein folds and the other solving the structure of singleton sequences with no sequence homologues. Both approaches increase information about sequence-structure relationships and help to develop protein structure prediction methods.

YKR049C (also named FMP46) is a mitochondrial protein derived from the *Saccharomyces cerevisiae* genome sequence. Its cellular presence and localization at mitochondria has been elucidated by florescence microscopy, GFP-fused protein expression and proteome analysis (Huh *et al.*, 2003; Sickmann *et al.*, 2003). However, although it is known to be localized in mitochondria, its molecular and cellular functions remain unknown. BLAST (Altschul *et al.*, 1990) search results suggested several sequence homologues of YKR049C (Fig. 1). However, since all were annotated as hypothetical proteins from yeast-like eukaryotes, they yielded no information about the biological function of YKR049C. In this report, we deduce the putative function of YKR049C from its solution structure and by comparing this with homologous proteins.

Materials and Methods

Protein purification The hypothetical ORF of YKR049C from *Saccharomyces cerevisiae* was sub-cloned into the pET-15b 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 ^{15}N-ammonium chloride (1 g/L) and ^{13}C-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

*To whom correspondence should be addressed.
Tel: 82-2-2123-2706; Fax: 82-2-363-2706
E-mail: wlee@spin.yonsei.ac.kr

containing 25 mM sodium phosphate (pH 6.5), 150 mM NaCl, 1 mM DTT, 90% H_2O/10% D_2O.

NMR spectroscopy The majority of NMR spectra were collected at 25°C on Varian Inova 500 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 CCC(CO)NH-TOCSY, H(CCCO)NH-TOCSY, and HCCH-TOCSY spectra (Bax *et al.*, 1994; Kay, 1997). For distance restraints, simultaneous CN-NOESY (Pascal *et al.*, 1995) and ^{13}C-edited NOESY were acquired with a 150ms mixing time. ^{13}C-edited NOESY and HCCH-TOCSY spectra were acquired using a Bruker DRX600 equipped with a Cryoprobe. Data from ^{13}C-edited NOESY experiments were used to assign aromatic protons.

Structure calculation Distance restraints for structure calculations were derived from cross-peaks in simultaneous CN-NOESY and ^{13}C-edited NOESY (τ_m = 150 ms) spectra. Slowly exchanging amide protons were identified by acquiring a series of ^{15}N-^{1}H HSQC spectra after dissolving lyophilized protein into 100% D_2O. Structure calculations were performed using CYANA (Herrmann *et al.*, 2002). The molecular topology of the β-sheet was determined using unambiguous H_α-H_α NOEs and automatic NOESY assignments without human-interpretation. After determining β-sheet topology, hydrogen bond restraints were applied. Angle constraints were obtained from TALOS predictions (Cornilescu *et al.*, 1999). NMR-derived experimental restraints contained 1168 unambiguous NOEs (175 intraresidue, 296 sequential, 206 medium-range ($2 <= |i-j| < 5$), and 490 long-range ($|i-j| >= 5$) NOE constraints), 86 distance restraints for 43 backbone hydrogen bonds, and 168 dihedral angle restraints were used for the structure calculation. The 20 structures calculated were subjected to molecular dynamics simulation in explicit water using the CNS program (Brunger *et al.*, 1998). Originally, the water shell refinement protocol was a part of ARIA (Linge *et al.*, 2003), but in the present study this protocol was performed separately using a home-made python script in our laboratory. At this stage of the structure elucidation, all dihedral angle constraints were fully relaxed in the range ±60 degrees. Final structures with lowest NOE energies were retained and validated using Procheck (Laskowski *et al.*, 1996). Structures were analyzed and visualized using PyMOL (DeLano Scientific LLC, San Carlos, USA) and MOLMOL (Koradi *et al.*, 1996), and electrostatic potential surface was calculated using APBS (Baker *et al.*, 2001). The elucidated structure was deposited to the Protein Data Bank under accession code 1WPI.

Result and Discussion

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 ^{1}H-^{15}N 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 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

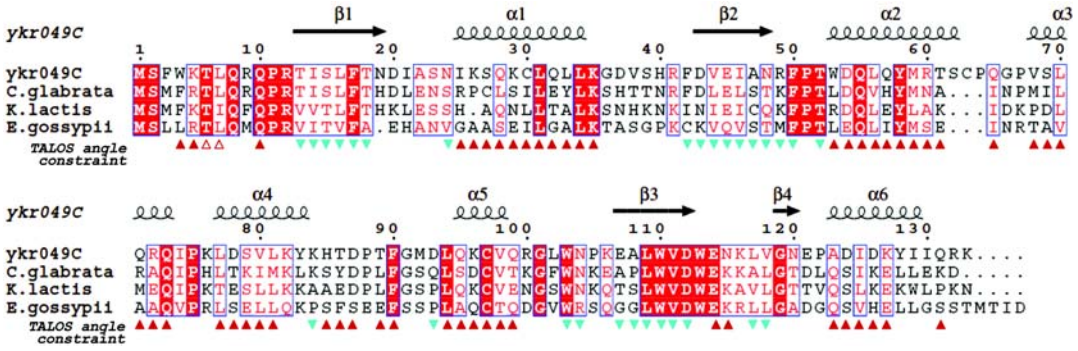

Fig. 1. Sequence alignment of YKR049C with homologous proteins. Sequences were obtained from the Swiss Protein Database or from the trEMBL, and have the following accession numbers: *Saccharomyces cerevisiae* (P36141), *Candida glabrata* (Q6FJI9), *Kluyveromyces lactis* (Q6CN57), *Eremothecium gossypii* (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.

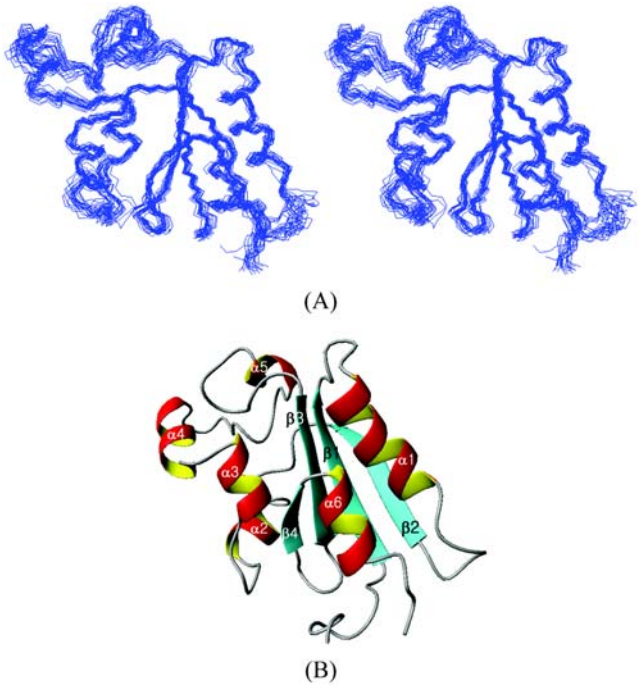

Fig. 2. Solution Structure of YKR049C. (A) A stereo plot of the backbone atoms for 20 lowest target function structures are optimally superimposed with respect to the average coordinates of the residues 10-131 (N, Cα, C', O). The figure was generated using MOLMOL. (B) A ribbon representation of YKR049C. Helices are displayed in red and β-strands in cyan.

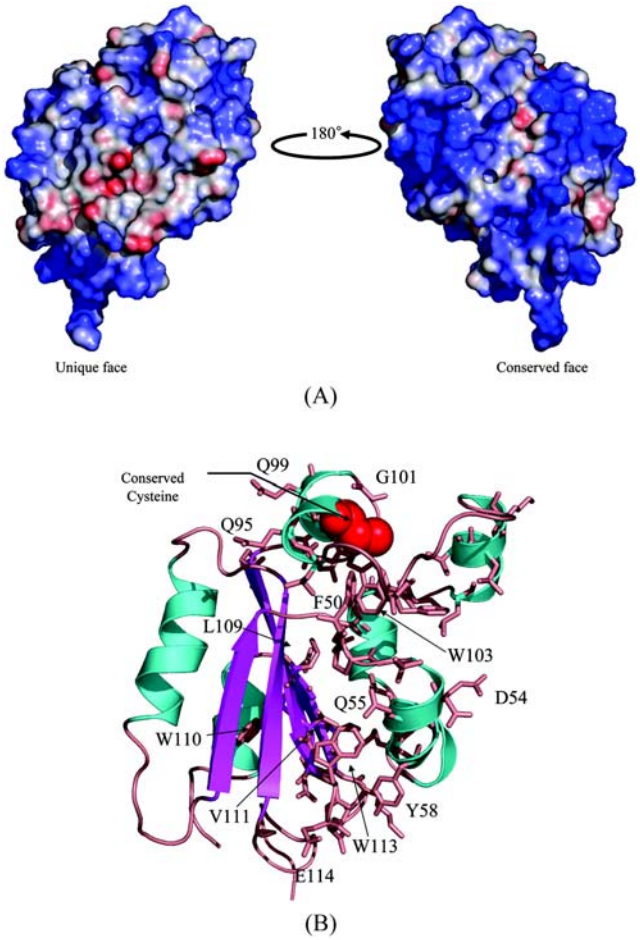

Fig 3. Molecular properties of YKR049C. (A) Electrostatic potential surface calculated by APBS. Blue means a positive and red a negative charge. (B) Distribution of conserved residues among YKR049C homologues listed in Fig 1. All conserved residues are displayed as stick models. A single active cysteine residue (Cys97) is presented as a CPK model.

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 1.177Å 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-γ1 fragment, and glutaredoxin-like protein

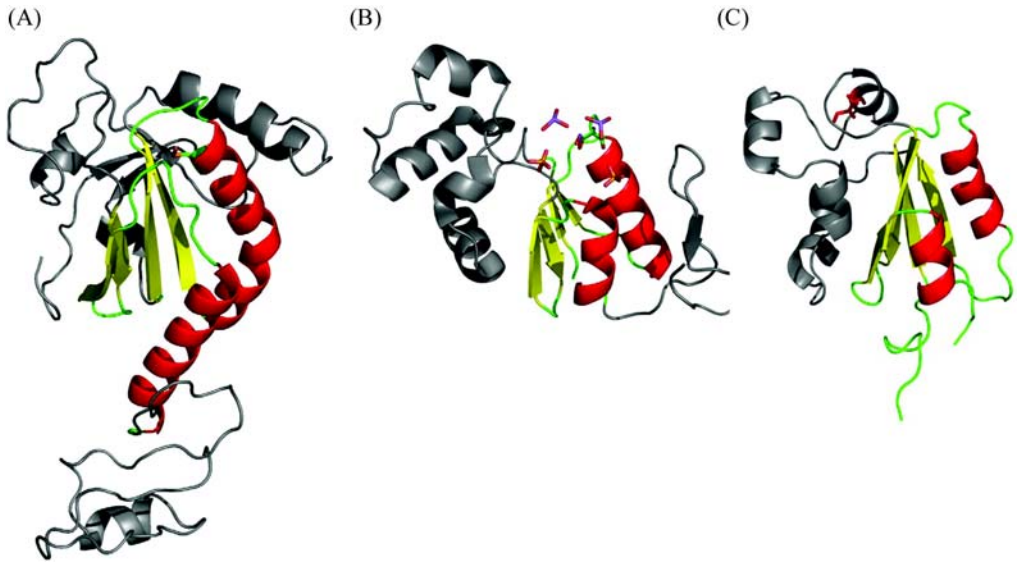

Fig. 4 Structural comparison of thioredoxin-fold proteins containing a single active cysteine. (A) Human peroxidase Prx vi (PDB:1prx), (B) Arsenate reductase ArsC (PDB:1j9b), and (C) YKR049C. The active sites of all proteins are displayed as stick models.

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 nonchoromosomal 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 H_2O_2, yielding two H_2O 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 H_2O_2.

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

Table 1. Structural statistics for the final calculated structure of YKR049C

Distance restraints	
All	1168
Intraresidue	175
Sequential (|i-j| = 1)	297
Medium (1<|i-j| < 5)	206
Long-range (|i-j|> = 5)	490
Hydrogen bonds[a]	86
Dihedral restraints	
All	168
Phi	84
Psi	84
Mean CYANA target function	40±1.53
van der Waals energy	-449±30.03
Mean RMS deviations from the average coordinate[b]	
Backbone atoms (N,Cα,C',O)	0.681 Å
All heavy atoms	1.333 Å
Ramachandran statistics	
Residues in most favored region	64.7%
Residues in additional allowed regions	25.2%
Residues in generously allowed regions	7.6%
Residues in disallowed regions	2.5%

[a]Two restraints per one hydrogen bond
[b]Secondary structured region only.

YKR049C is a 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 (M1-0203-00-0020), by the Northeast Structural Genomics Consortium (NIH PSI grant P50 GM26413), and by the Ontario Research and Development Challenge Fund and Genome Canada.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037-10041.

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. and Zhu, G. (1994) Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* **239**, 79-105.

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., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998) Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr.* **54**, 905-921.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289-302.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293.

Goddard, T. D. and Kneller, D. G. SPARKY 3, University of California, San Francisco.

Hermann, T., Guntert, P. and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209-227.

Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

Holmgren, A. (1995) Thioredoxin structure and mechanism: conformational changes on oxidation of the active-site sulfhydryls to a disulfide. *Structure* **3**, 239-243.

Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S. and O'Shea, E. K. (2003) Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691.

Kay, L. E. (1997) NMR methods for the study of protein structure and dynamics. *Biochem. Cell. Biol.* **75**, 1-15.

Kim, S. H. (1998) Shining a light on structural genomics. *Nat. Struct. Biol.* **5**, 643.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51-55.

Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477-486.

Linge, J. P., Williams, M. A., Spronk, A. E. M., Bonvin, A. M. and Nilges, M. (2003) Refinement of protein structures in explicit solvent. *Proteins* **50**, 496-506.

Martin, J. L. (1995) Thioredoxin - a fold for all reasons. *Structure* **3**, 245-250.

Martin, P., DeMel, S., Shi, J., Gladysheva, T., Gatti, D. L., Rosen, B. P. and Edwards, B. F. P. (2001) Insights into the structure, salvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. *Structure* **9**, 1071-1081.

Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Forman-Kay, J. D. and Kay, L. E. (1994) Simultaneous acquisition of ^{15}N-edited and ^{13}C-edited NOE spectra of proteins dissolved in H_2O. *J. Magn. Reson.* **103**, 197-201.

Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H. E., Schonfisch, B., Perschil, I., Chacinska, A., Guiard, B., Rehling, P., Pfanner, N. and Meisinger, C. (2003) The proteome of Saccharomyces cerevisiae mitochondria. *Proc. Natl. Acad. Sci. USA* **100**, 13207-13212.

Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G. M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C. H., Cort, J. R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L. P., Gehring, K., Kennedy, M. A., Edward, A. M. and Arrowsmith, C. H. (2002) An NMR approach to structural proteomics. *Proc. Natl Acad. Sci. USA* **99**, 1825-1830.

Zimmerman, D. E., Kulikowski, C. A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C., Powers, R. and Montelione, G. T. (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. *J. Mol. Biol.* **269**, 592-610.