Structure, Vol. 11, 1049–1050, September, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0969-2126(03)00183-7

Letter to the Editor

Homology Among Telomeric End-Protection Proteins

Telomere maintenance and end protection are essential for the survival and proliferation of eukaryotic cells, leading to the prediction that components of this system would be highly conserved. In practice, however, evidence for homology among these factors has been elusive, and, in the case of the known end-protection proteins, evolutionary relationships have been postulated largely on the basis of protein structural and functional similarity alone. Here we report support from sequence profile analyses for a significant and specific evolutionary relationship among OB-fold telomeric end-protection factors.

Several proteins have been identified that specifically bind to the telomeric single-strand overhang, such as *Oxytricha nova* telomere end binding protein (TEBP), the metazoan, fission yeast, and plant protection of telomeres 1 (Pot1) proteins, and *Saccharomyces cerevisiae* Cdc13. Cdc13 is an essential protein, and fission yeast Pot1, TEBP and Cdc13 all participate in telomere length regulation (Baumann and Cech, 2001; Froelich-Ammon et al., 1998; Gottschling and Zakian, 1986; Nugent et al., 1996). Although the Pot1 proteins were originally identified from weak sequence similarity to the amino-terminal portion of the TEBP α subunit, no similarity was apparent between any of these proteins and Cdc13 (Baumann and Cech, 2001). High-resolution crystal structures of TEBP complexes and the solution structure of Cdc13 have been solved (Horvath et al., 1998; Mitton-Fry et al., 2002; Theobald and Schultz, 2003). Strikingly, both proteins recognize telomeric ssDNA with OB-fold domains, an evolutionarily ancient protein fold frequently used for specific recognition of single-stranded nucleic acids and renowned for its sequence heterogeneity (Murzin, 1993; Theobald et al., 2003). Based upon the combination of structural and functional similarity among these proteins, it has been proposed that yeast Cdc13 shares a telomeric common ancestor with ciliate TEBP (Horvath et al., 1998), as well as with the fungal, plant, and metazoan Pot1 proteins (Cervantes and Lundblad, 2002; Mitton-Fry et al., 2002).

To test this proposed homology, we performed sequence-based profile-profile analyses to evaluate the extent of sequence similarity detectable among the OB-fold domains of these telomeric end-protection factors. Profile-based methods, which evaluate similarities shared between families of sequences rather than between single sequences, significantly outperform pairwise methods in detecting remote homology (Karplus et al., 1997; Park et al., 1998; Sadreyev and Grishin, 2003). Of these methods, recently developed profile-profile comparison algorithms, such as those implemented in COMPASS (used in this study) (Sadreyev and Grishin, 2003) and prof_sim (Yona and Levitt, 2002), are currently the most sensitive for remote homolog detection.

To probe for distant homology among telomeric OB-fold domains, a database of 81 OB-fold sequence alignments was constructed. We obtained from the ASTRAL/SCOP protein structure database (Chandonia et al., 2002; Murzin et al., 1995) a collection of sequences representing all known OB-fold domains with less than 40% sequence identity, including 45 nontelomeric nucleic acid binding OB-fold domains and 30 OB-fold domains that do not bind nucleic acids. Specific multiple sequence alignments were made for each SCOP OB-fold domain by searching the nonredundant protein database with BLAST, aligning sequences with BLAST E-values $<10^{-10}$, and cropping the alignments to the limits of the original query OB-fold domain. Additionally, close fungal homologs of *Saccharomyces cerevisiae* Cdc13p were identified via BLAST searches of five yeast genomes (Cliften and Johnston, 2003). Close homologs of the Pot1 proteins were likewise obtained from the NCBI nonredundant protein database. As negative controls, alignments of all SCOP RNA recognition motif domains (RRMs), eukaryotic and prokaryotic KH domains, and canonical dsRNA binding domains (RBDs) were also constructed and included in this database. Both KH domains and RRMs are structurally and functionally similar to OB-folds, being β sheet, single-stranded nucleic acid binding domains.

Using COMPASS, the telomeric OB-fold alignments were then scored against this database. The results are summarized in Table 1. Although some of the nontelomeric OB-fold domains yielded significant scores (E-value < 0.05), in all cases the highest scoring significant alignments were from other telomeric OB-fold domains. For example, the Cdc13 OB-fold domain scores highest (E-value $= 5.3 \times 10^{-3}$) against the telomeric Pot1 OB-fold, a value four orders of magnitude greater than the average significance found between the Cdc13 OB-fold and nontelomeric OB-folds. Consistent with previous analyses, the N-terminal OB-fold domains in Pot1 and TEBP α score highly. Because the similarity found in this analysis is distributed throughout the OB-fold domains and is not restricted to DNA binding interfaces, the sequence similarity is unlikely to be due to convergence for binding to GT-rich single-stranded telomeric DNA.

As a positive control, when alignments of SSB proteins were scored against the database, the highest scoring significant alignments were against other SSBs, as expected for these functionally related proteins. The lone exception is the second OB-fold domain from human RPA70, which interestingly scores significantly (E-value $= 7.4 \times 10^{-4}$) against Pot1. None of the RRM, KH-domain, and RBD alignments yielded significant E-values with the telomeric OB-fold alignments.

Thus, using profile-profile sequence analysis, we are able to detect distant homology between telomere end binding proteins. The sequence similarity is significantly

Table 1. Lowest COMPASS E-Value Scores for Telomeric End-Binding Protein OB-Folds

Cdc13 OB		Pot1 OB		*On*TEBP α OB1	
pot1 OB	5.3×10^{-3}	TEBP α1	1.4×10^{-14}	pot1 OB	2.0×10^{-14}
*On*TEBP αOB1	9.0	rpa70 OB2	7.4×10^{-4}	NAD DNA ligase	7.4×10^{-3}
		cdc13 OB	4.0×10^{-3}	*S.c.* AspRS	2.0×10^{-2}
Nontelomeric[a]	40	Nontelomeric[a]	7.7	Nontelomeric[a]	37

Unless otherwise indicated, alignments scoring <0.05 are shown. Queries were excluded from the database for each search, and thus E-values are not exactly reciprocal. The database of alignments used, complete scores, and additional technical details are available from the authors.
[a] Geometric average of E-values for indicated OB-fold scored against all nontelomeric OB-fold alignments.

in excess of what is found between the telomeric proteins and nearly all other OB-fold proteins, including other OB-fold nucleic acid binding domains. These results support the relevance of the *Saccharomyces* telomere as a model system for eukaryotic telomere biology and further support the usefulness of profile-profile-based methods for clarifying the evolutionary relationships of proteins with known structure.

Douglas L. Theobald,[1,*] Rachel B. Cervantes,[2] Victoria Lundblad,[2] and Deborah S. Wuttke[1]
[1]Department of Chemistry and Biochemistry
University of Colorado at Boulder
Boulder, Colorado 80309
[2]Department of Molecular and Human Genetics
Baylor College of Medicine
Houston, Texas 77030

*Correspondence: theobal@colorado.edu

References

Baumann, P., and Cech, T.R. (2001). Science *292*, 1171–1175.

Cervantes, R.B., and Lundblad, V. (2002). Curr. Opin. Cell Biol. *14*, 351–356.

Chandonia, J.M., Walker, N.S., Lo Conte, L., Koehl, P., Levitt, M., and Brenner, S.E. (2002). Nucleic Acids Res. *30*, 260–263.

Cliften, P., and Johnston, M. (2003). Washington University Genome Sequencing Center (http://genome.wustl.edu/).

Froelich-Ammon, S.J., Dickinson, B.A., Bevilacqua, J.M., Schultz, S.C., and Cech, T.R. (1998). Genes Dev. *12*, 1504–1514.

Gottschling, D.E., and Zakian, V.A. (1986). Cell *47*, 195–205.

Horvath, M.P., Schweiker, V.L., Bevilacqua, J.M., Ruggles, J.A., and Schultz, S.C. (1998). Cell *95*, 963–974.

Karplus, K., Sjolander, K., Barrett, C., Cline, M., Haussler, D., Hughey, R., Holm, L., and Sander, C. (1997). Proteins Suppl. *1*, 134–139.

Mitton-Fry, R.M., Anderson, E.M., Hughes, T.R., Lundblad, V., and Wuttke, D.S. (2002). Science *296*, 145–147.

Murzin, A.G. (1993). EMBO J. *12*, 861–867.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). J. Mol. Biol. *247*, 536–540.

Nugent, C.I., Hughes, T.R., Lue, N.F., and Lundblad, V. (1996). Science *274*, 249–252.

Park, J., Karplus, K., Barrett, C., Hughey, R., Haussler, D., Hubbard, T., and Chothia, C. (1998). J. Mol. Biol. *284*, 1201–1210.

Sadreyev, R., and Grishin, N. (2003). J. Mol. Biol. *326*, 317–336.

Theobald, D.L., Mitton-Fry, R.M., and Wuttke, D.S. (2003). Annu. Rev. Biophys. Biomol. Struct. *32*, 115–133.

Theobald, D.L., and Schultz, S.C. (2003). EMBO J. *22*, 4314–4324.

Yona, G., and Levitt, M. (2002). J. Mol. Biol. *315*, 1257–1275.