

Tagawa, T. Tomoya, K. Nomura, T. Hanagaki, and T. Matsuyama. We thank T. Takase and M. Matsui for sampling of the dark-grown plants. Supported by a grant for Genome Research from RIKEN; the Program for Promotion of Basic Research Activities for Innovative Biosciences; the Special Coordination Fund of the Science and Technology Agency; and a

Grant-in-Aid from the Ministry of Education, Science and Culture of Japan to K.S. Also supported in part by a Grant-in-Aid for Scientific Research on Priority Areas (C) "Genome Science" from the Ministry of Education, Science, Sports and Culture of Japan to M.S. and by a Research Grant for the RIKEN Genome Exploration Research Project from

the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government to Y.H.

19 February 2002; accepted 12 March 2002
Published online 21 March 2002;
10.1126/science.1071006

Include this information when citing this paper.

Conserved Structure for Single-Stranded Telomeric DNA Recognition

Rachel M. Mitton-Fry,¹ Emily M. Anderson,¹
Timothy R. Hughes,^{2*} Victoria Lundblad,^{2,3} Deborah S. Wuttke^{1†}

The essential Cdc13 protein in the yeast *Saccharomyces cerevisiae* is a single-stranded telomeric DNA binding protein required for chromosome end protection and telomere replication. Here we report the solution structure of the Cdc13 DNA binding domain in complex with telomeric DNA. The structure reveals the use of a single OB (oligonucleotide/oligosaccharide binding) fold augmented by an unusually large loop for DNA recognition. This OB fold is structurally similar to OB folds found in the ciliated protozoan telomere end-binding protein, although no sequence similarity is apparent between them. The common usage of an OB fold for telomeric DNA interaction demonstrates conservation of end-protection mechanisms among eukaryotes.

Telomeres are the specialized nucleoprotein complexes that cap eukaryotic chromosomes, protecting chromosome ends from unregulated degradation and end-to-end fusion. Telomeric DNA is typically composed of repetitive, noncoding sequence terminating in a single-stranded TG-rich overhang. Several mechanisms have been identified for capping this overhang, ranging from sequestration through protein binding in ciliates and yeasts to t-loop formation in mammals (1–3). Proteins that specifically bind to this single-stranded overhang, such as the *Oxytricha nova* telomere end-binding protein (TEBP) (4, 5), the *Schizosaccharomyces pombe* protection of telomeres 1 (Pot1) and human Pot1 (6), and the *Saccharomyces cerevisiae* Cdc13 (7, 8), are involved in telomeric end protection. For example, depletion of Cdc13 activity causes extensive resection of the 5' strand of the yeast telomere and DNA damage-dependent cell cycle arrest (9–12), whereas deletion of the *pot1* gene leads to complete telomere loss and cell death (6). Cdc13 is also required for telomere elongation as a positive regulator of telomerase (7, 13).

Cdc13 is believed to fulfill both of these important, yet disparate, roles through localization to the 3' single-stranded telomeric end, followed by recruitment of relevant complexes to the telomere through protein-protein interactions (14–16).

Evidence for conservation of telomeric end-protection proteins among distantly related eukaryotes has been elusive. Although the Pot proteins were originally identified on the basis of weak sequence similarity to the NH₂-terminal portion of the α subunit of the heterodimeric *O. nova* TEBP (6), no similarity was apparent between any of these proteins and Cdc13. To investigate the requirements for telomeric end protection and sequence-specific interaction with single-stranded DNA (ssDNA), we determined the solution structure of the Cdc13 DNA binding domain (DBD) in complex with telomeric ssDNA. This 23.5-kD domain retains DNA binding activity and specificity (17–19), and fusions of the DBD with other components of the end-protection or telomerase machinery eliminate the need for full-length protein in vivo (14, 15). The ssDNA 11-nucleotide (nt) oligomer dGTGTGGGTGTG in the complex is the minimal Cdc13 binding site (17) and the complement to the center of the coding region of the telomerase RNA template (20).

The high-resolution Cdc13 DBD structure in complex with ssDNA (Fig. 1) was calculated from a total of 2865 nuclear magnetic

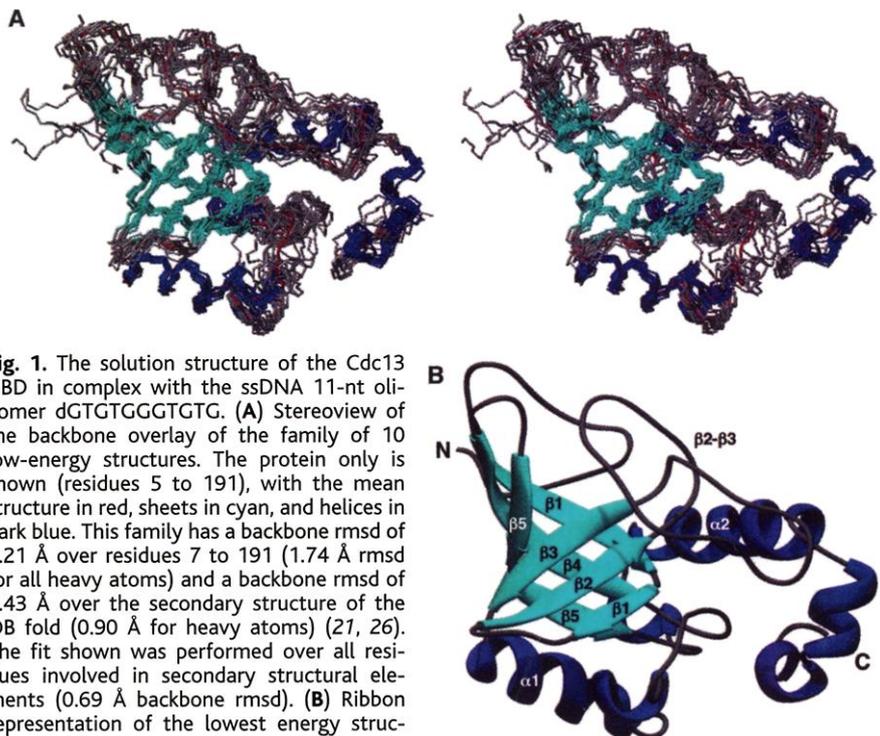


Fig. 1. The solution structure of the Cdc13 DBD in complex with the ssDNA 11-nt oligomer dGTGTGGGTGTG. (A) Stereoview of the backbone overlay of the family of 10 low-energy structures. The protein only is shown (residues 5 to 191), with the mean structure in red, sheets in cyan, and helices in dark blue. This family has a backbone rmsd of 1.21 Å over residues 7 to 191 (1.74 Å rmsd for all heavy atoms) and a backbone rmsd of 0.43 Å over the secondary structure of the OB fold (0.90 Å for heavy atoms) (21, 26). The fit shown was performed over all residues involved in secondary structural elements (0.69 Å backbone rmsd). (B) Ribbon representation of the lowest energy structure, residues 7 to 191. Figures were prepared with MOLMOL (33) and RIBBONS (34).

¹Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA. ²Interdepartmental Program in Cell and Molecular Biology, ³Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.

*Present address: University of Toronto, Banting and Best Department of Medical Research, Toronto, Ontario M5G 1L6, Canada.

†To whom correspondence should be addressed. E-mail: deborah.wuttke@colorado.edu

REPORTS

resonance (NMR) restraints (21). Comparison of the Cdc13 DBD to the structural database unequivocally places Cdc13 in the oligonucleotide-binding superfamily of OB

fold proteins (22, 23). The OB fold is a small structural motif used for oligonucleotide, oligosaccharide, and oligopeptide binding (24). This fold, exemplified by

verotoxin-1, staphylococcal nuclease, and the anticodon-binding domain of asp-tRNA synthetase, cannot yet be predicted on the basis of sequence comparisons. The canonical OB fold, also seen in the Cdc13 DBD, consists of a β barrel formed by two orthogonally packed three-stranded antiparallel β sheets. Sheet 1 is composed of β 1, β 2, and β 3, and sheet 2 comprises β 5, β 4, and β 1. In the Cdc13 DBD, numerous NOEs (nuclear Overhauser effects) between β 3 and β 5 close the barrel, and an α helix between β 3 and β 4 caps the bottom of the barrel. The Cdc13 DBD has an unusually long, 30-residue loop between β 2 and β 3 that packs tightly over the β 2 and β 3 strands. This loop is structurally well-defined, as indicated by heteronuclear NOE measurements, although it has no regular secondary structure. An α -helical region extends the domain COOH-terminally beyond the OB fold.

The DNA binding site on the protein surface was identified by comparison of NMR chemical shifts in the presence and absence of DNA and analysis of intermolecular NOEs between the protein and DNA. DNA binding induced extensive chemical shift changes throughout the OB fold region of the domain and were most pronounced in β 3, β 4, and β 5, and across the loop between β 2- β 3 (Fig. 2). The COOH-terminal helical region showed no evidence of involvement in DNA binding. We next directly localized the protein-DNA interface on the basis of more than 50 intermolecular NOEs observed between the protein and DNA (Fig. 3) (25). These protein-DNA contacts unambiguously define an extensive intermolecular interface that coincides with the binding surface indicated by the chemical shift changes. This interaction surface extends across sheet 2, through the cleft defined by the loops between β 3- α 1 and β 4- β 5, over strand β 3, and across the length of the long loop between β 2- β 3. Interestingly, 9 of the 11 nucleotides in the DNA molecule showed NOE contacts to at least one amino acid in the protein. Thirteen residues

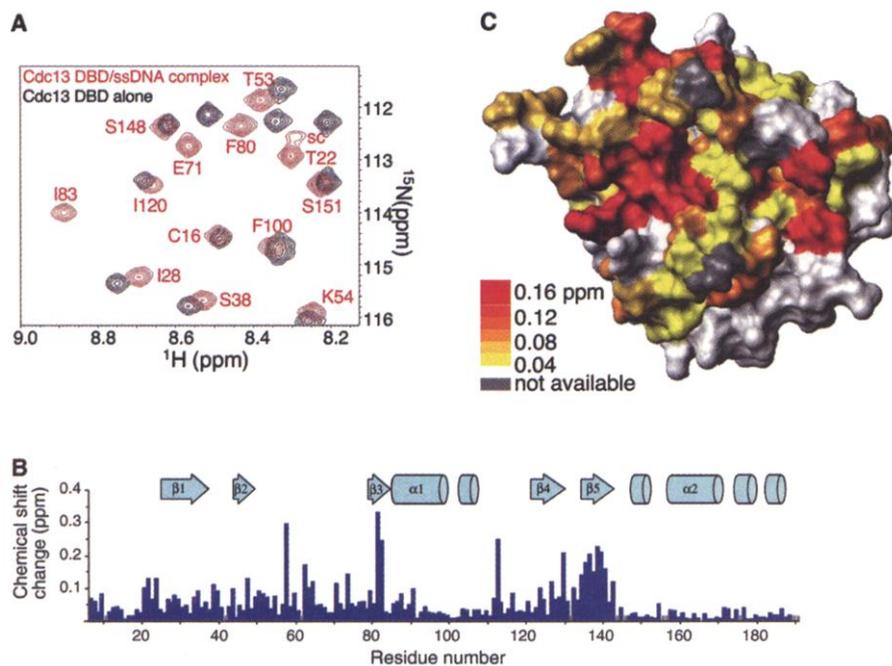


Fig. 2. Interaction of the Cdc13 DBD with single-stranded telomeric DNA determined by chemical shift perturbation. (A) Comparison of DBD chemical shifts in the presence and absence of DNA. Overlay of a region of the ^{15}N - ^1H HSQC (heteronuclear single-quantum coherence) spectra of protein-ssDNA complex (red) and of protein alone (black). Crosspeaks from the protein/DNA complex have been labeled on the spectrum (sc, side chain) (26, 35). Because the complex binds in the slow-exchange time regime, assignments for protein alone cannot be determined by titration. (B) Minimal chemical shift perturbation upon DNA binding and secondary structural elements mapped on the protein sequence. Substantial chemical shift changes occur throughout the OB fold portion of the DBD, concentrating in the β -barrel and loop regions. Perturbation values have been calculated according to the equation: $\text{perturbation} = \sqrt{(\Delta\text{ppm}_{\text{H,min}})^2 + (0.17 \times \Delta\text{ppm}_{\text{N,min}})^2}$, where $\Delta\text{ppm}_{\text{H,min}}$ and $\Delta\text{ppm}_{\text{N,min}}$ are the minimal chemical shift differences (in parts per million) for proton and nitrogen, respectively (36). Gray shading indicates residues for which no backbone information is available (predominantly prolines). This method assumes that the crosspeak in the spectrum of protein alone with the least chemical shift change from any given peak in the complex spectrum corresponds to the same residue. Thus, this analysis is an underestimation of the true perturbation upon DNA binding. (C) Minimal chemical shift perturbation upon DNA binding mapped on the DBD structure, residues 7 to 191. Yellow to red shading indicates residues with increasing chemical shift perturbation upon DNA binding. Gray shading indicates residues for which no backbone information is available (predominantly prolines).

Fig. 3. Interaction of the Cdc13 DBD with single-stranded telomeric DNA as seen by direct NOE contacts. (A) Amino acid residues involved in intermolecular NOE contacts with the DNA are mapped on the protein structure (26). Tyrosine residues are highlighted in yellow, basic residues in blue, and hydrophobic residues in magenta. (B) The above contact residues mapped on the protein surface, shaded as in (A).

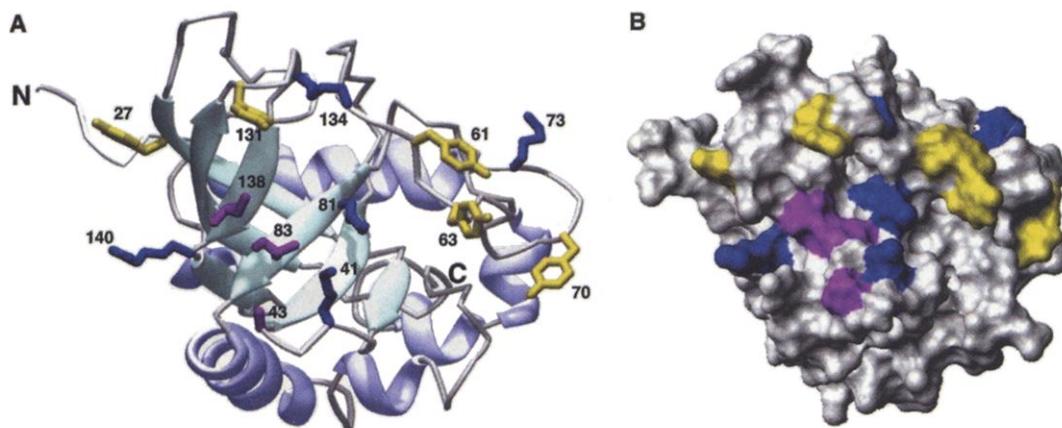
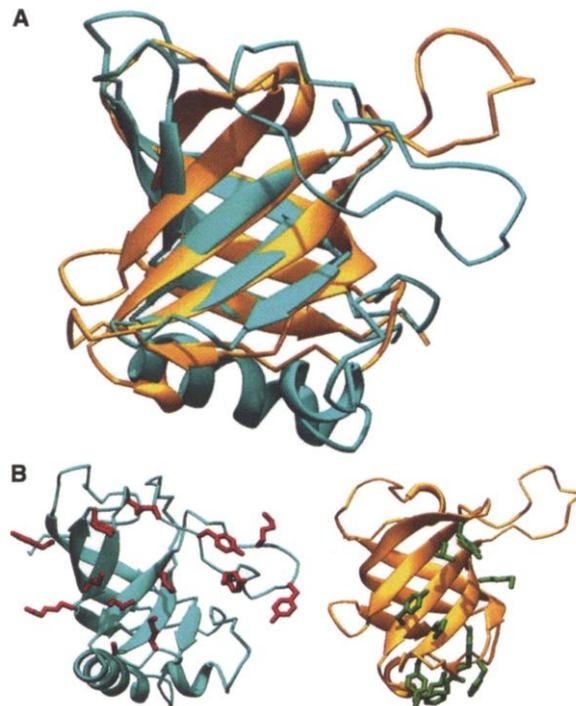


Fig. 4. Structural comparison of the Cdc13 DBD OB fold with the NH₂-terminal *O. nova* α OB fold. **(A)** Overlay of the two OB folds. The Cdc13 DBD (residues 10 to 148) is shown in cyan (26), and the *O. nova* α OB fold (residues 37 to 150) is shown in gold. Fits were performed with LSQMAN (37). **(B)** Comparison of the DNA binding interfaces of the two OB folds. Colors are as described in (A), with contact residues of the Cdc13 DBD OB fold (left) highlighted in red and those of the NH₂-terminal *O. nova* α OB fold (right) in green. This figure was rotated 40° from (A) (the same orientation as seen in Fig. 3) to illustrate the size difference between the two interfaces. *O. nova* α contact residues are taken from (28).



were unambiguously identified at the protein-DNA interface, including five aromatic (Y27, Y61, Y63, Y70, and Y131), three hydrophobic (A43, I83, and I138), and five basic amino acids (K41, K73, K81, K134, and R140) (26). The predominance of aromatic, hydrophobic, and basic residues in the Cdc13 DBD interface suggests that the aromatic stacking, hydrophobic interactions, and phosphate contacts typically seen in OB fold-ligand structures are also critical for ssDNA binding in the Cdc13 DBD.

OB folds typically interact with a small ligand (e.g., 2 to 5 nucleotides for OB folds that recognize nucleic acids) through interactions with the loops between β1-β2, β3-α, and β4-β5 (24). This mode of recognition is also observed in the Cdc13 DBD-ssDNA interaction. However, the Cdc13 DBD markedly expands its interaction surface by using a large loop between β2-β3 for DNA binding (27). The extended interface can accommodate the entire DNA molecule, explaining the requirement for at least an 11-nt oligomer of cognate ssDNA for full binding affinity (17). This exploitation of the Cdc13 DBD loop for ligand recognition illustrates the substantial malleability and adaptability of the OB fold.

Although the Pot proteins have not yet been structurally characterized, the structure of the ternary complex of the related *O. nova* TEBP bound to a 12-nt oligomer of cognate single-stranded telomeric DNA (G₄T₄G₄) has been solved at high resolution (28, 29). The protein complex contains four OB folds, three of which are integral to DNA binding. The Cdc13 DBD exhibits a high degree of structural similarity to each of these OB folds with superpositions of

less than 3 Å root mean square deviation (rmsd) over the secondary structural elements of the OB fold (30). Notably, the Cdc13 DBD superimposes with a 2.2 Å rmsd to the NH₂-terminal OB fold of the α subunit, the region of TEBP that was used to identify the Pot proteins (Fig. 4). Structure-based sequence alignments revealed no appreciable sequence similarity between the Cdc13 DBD and the *O. nova* OB folds over the region of structural superposition or over the amino acids that make direct DNA contacts. This lack of sequence similarity demonstrates the critical need for structure-based comparisons for assessment of homology among divergent proteins. The close structural relationship observed here suggests that despite its sequence divergence, Cdc13 shares a common ancestor with the *O. nova* TEBP, and therefore with the Pot proteins as well.

The structural similarity between the Cdc13 DBD and other proteins involved in telomere end protection shows that the OB fold is a broadly conserved structural element for binding single-stranded telomeric termini. Functional similarities are also seen in the cellular responses to Cdc13 and Pot1 depletion with regard to chromosome end protection. This combination of structural and functional similarity between Cdc13 and telomere end-binding proteins from other distantly related eukaryotes indicates that mechanisms of telomeric end protection are widely conserved throughout evolution.

References and Notes

1. D. Shore, *Curr. Opin. Genet. Dev.* **11**, 189 (2001).
2. E. H. Blackburn, *Cell* **106**, 661 (2001).
3. T. de Lange, *Science* **292**, 1075 (2001).
4. D. E. Gottschling, V. A. Zakian, *Cell* **47**, 195 (1986).

5. C. M. Price, T. R. Cech, *Genes Dev.* **1**, 783 (1987).
6. P. Baumann, T. R. Cech, *Science* **292**, 1171 (2001).
7. C. I. Nugent, T. R. Hughes, N. F. Lue, V. Lundblad, *Science* **274**, 249 (1996).
8. J.-J. Lin, V. A. Zakian, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13760 (1996).
9. B. Garvik, M. Carson, L. Hartwell, *Mol. Cell. Biol.* **15**, 6128 (1995).
10. T. A. Weinert, L. H. Hartwell, *Genetics* **134**, 63 (1993).
11. C. Booth, E. Griffith, G. Brady, D. Lydall, *Nucleic Acids Res.* **29**, 4414 (2001).
12. S. J. Diede, D. E. Gottschling, *Cell* **99**, 723 (1999).
13. J. Lingner, T. R. Cech, T. R. Hughes, V. Lundblad, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11190 (1997).
14. S. K. Evans, V. Lundblad, *Science* **286**, 117 (1999).
15. E. Pennock, K. Buckley, V. Lundblad, *Cell* **104**, 387 (2001).
16. A. J. Lustig, *Nature Struct. Biol.* **8**, 297 (2001).
17. T. R. Hughes, R. G. Weilbaecher, M. Walterscheid, V. Lundblad, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6457 (2000).
18. E. M. Anderson, W. A. Halsey, D. S. Wuttke, in preparation.
19. Furthermore, as has been previously observed for an NH₂-terminal proteolytic fragment of *S. pombe* Pot1 (6) and the α35 subunit of the *O. nova* TEBP (31), the Cdc13 DBD binds telomeric DNA more tightly than the full-length protein (18).
20. M. S. Singer, D. E. Gottschling, *Science* **266**, 404 (1994).
21. Supplementary Web material is available on Science Online at www.sciencemag.org/cgi/content/full/296/5565/145/DC1.
22. S. Dietmann *et al.*, *Nucleic Acids Res.* **29**, 55 (2001).
23. S. Dietmann, L. Holm, *Nature Struct. Biol.* **8**, 953 (2001).
24. A. G. Murzin, *EMBO J.* **12**, 861 (1993).
25. Although our NMR data have allowed for the identification of 11 DNA spin systems, the dearth of NOEs between the spin systems precludes unambiguous assignment of each nucleotide. The low number of internucleotide NOEs is consistent with the DNA adopting a largely extended conformation.
26. Residue numbers listed for the Cdc13 DBD correspond to those of the full-length protein minus 495.
27. The NH₂-terminal OB fold of the *O. nova* α subunit contains a large (18 residue) loop between β2 and β3 that is distant from the site of nucleic acid interaction.
28. M. P. Horvath, V. L. Schweiker, J. M. Bevilacqua, J. A. Ruggles, S. C. Schultz, *Cell* **95**, 963 (1998).
29. M. P. Horvath, S. C. Schultz, *J. Mol. Biol.* **310**, 367 (2001).
30. The *O. nova* α NH₂-terminal domain (35 kD) bound to ssDNA shows no substantial structural differences from the corresponding OB folds in the ternary complex (32), validating the use of the ternary complex for structural comparison.
31. G. Fang, J. T. Gray, T. R. Cech, *Genes Dev.* **7**, 870 (1993).
32. S. Classen, J. A. Ruggles, S. C. Schultz, *J. Mol. Biol.* **314**, 1113 (2001).
33. R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **14**, 51 (1996).
34. M. Carson, *J. Appl. Crystallogr.* **24**, 958 (1991).
35. R. M. Mitton-Fry, D. S. Wuttke, *J. Biomol. NMR*, in press.
36. B. T. Farmer *et al.*, *Nature Struct. Biol.* **3**, 995 (1996).
37. G. J. Kleywegt, T. A. Jones, *CCPA/ESF-EACBM Newslett.* **31**, 9 (1994).
38. We thank T. Cech, A. Pardi, and the members of the Wuttke lab for helpful discussions and critical reading of the manuscript; R. Batey, K. Bjorkman, E. Fry, W. Halsey, B. Kelly, C. Mandel, and H. Zhou for technical assistance; and M. Summers and C. Tang for access to an 800-MHz spectrometer. Funded by NIH grants GM59414 (D.S.W.) and GM55867 (V.L.), the American Cancer Society (D.S.W.), a University of Colorado Junior Faculty Development Award (D.S.W.), a Beckman Young Investigator Award (D.S.W.), and the U.S. Army Breast Cancer Research Program (E.M.A.). R.M.M.-F. is a Howard Hughes Medical Institute Pre-doctoral Fellow.

7 December 2001; accepted 15 February 2002