Cell, Vol. 95, 963–974, December 23, 1998, Copyright ©1998 by Cell Press

Crystal Structure of the *Oxytricha nova* Telomere End Binding Protein Complexed with Single Strand DNA

Martin P. Horvath*, Viloya L. Schweiker*, Joanne M. Bevilacqua†, James A. Ruggles‡, and Steve C. Schultz*§

*Department of Chemistry and Biochemistry
University of Colorado
Boulder, Colorado 80309-0215

†Department of Chemistry
The Pennsylvania State University
University Park, Pennsylvania 16802

‡Department of Clinical Pharmacology
H. Lundback A/S
Copenhagen
Denmark

Summary

Telomeres are specialized protein–DNA complexes that compose the ends of eukaryotic chromosomes. Telomeres protect chromosome termini from degradation and recombination and act together with telomerase to ensure complete genome replication. We have determined the crystal structure of the two-subunit *Oxytricha nova* telomere end binding protein (*On*TEBP) complexed with single strand telomeric DNA at 2.8 Å resolution. The structure reveals four oligonucleotide/oligosaccharide–binding folds, three of which form a deep cleft that binds the ssDNA, and a fourth that forms an unusual protein–protein interaction between the α and β subunits. This structure provides a molecular description of how the two subunits of *On*TEBP recognize and bind ssDNA to form a sequence-specific, telomeric nucleoprotein complex that caps the very 3′ ends of chromosomes.

Introduction

Telomeres are the specialized nucleoprotein structures that make up the ends of eukaryotic chromosomes (Blackburn, 1991; Zakian, 1995). Telomeres are essential for chromosome stability because they protect natural chromosome ends from degradation and end-to-end fusion (McClintock, 1941) and because they serve as substrates for the ribonucleoprotein enzyme telomerase, which ensures complete genome replication (Greider and Blackburn, 1985). When telomere length and structure are not properly maintained, cells lose their ability to divide and become senescent (Lundblad and Szostak, 1989; Yu et al., 1990). Lack of telomere maintenance in normal human cells appears to limit their proliferative potential, indicating that telomeres are important players in the processes of aging and cancer (Harley et al., 1990; Bodnar et al., 1998).

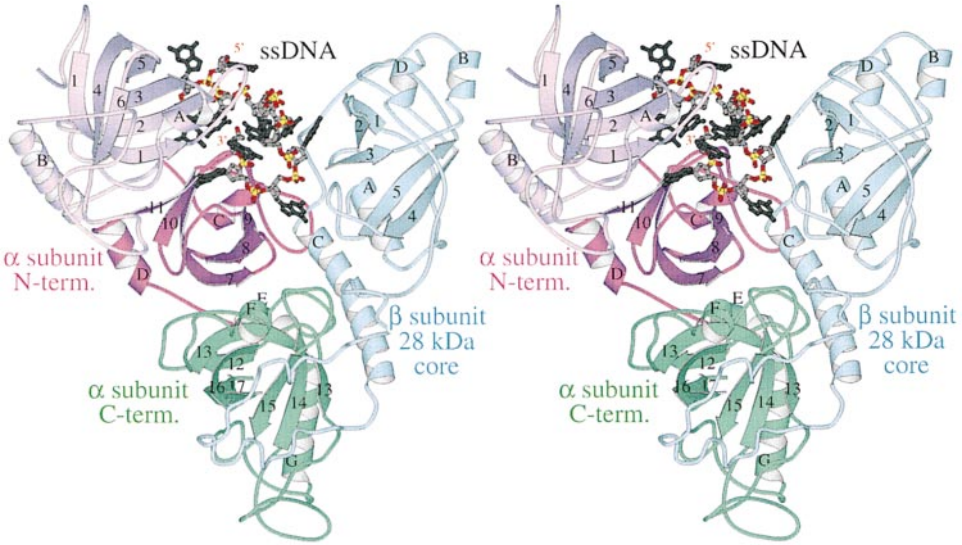
The DNA component of telomeres in nearly all eukaryotes consists of tracts of a short, tandemly repeated TG-rich sequence oriented 5′→3′ toward the chromosome terminus: $(T_4G_4)_n$ in *Oxytricha nova*; $(T_2G_4)_n$ in *Tetrahymena*; $(T_3AG_3)_n$ in *Arabidopsis*; and $(T_2AG_3)_n$ in all vertebrates and some fungi (reviewed in Zakian, 1995). The lengths of telomeric DNA vary greatly in different organisms: 7–10 kb in humans, up to 150 kb in certain strains of mice, ~350 nucleotides in *Saccharomyces cerevisiae*, and a precisely regulated 36 nucleotides in the macronuclei of *Oxytricha nova*. Although most of the telomeric DNA is double stranded, the 3′ ends are single stranded in all organisms investigated to date, including *Oxytricha*, *Tetrahymena*, *Didymium*, *Saccharomyces*, and humans (Klobutcher et al., 1981; Henderson and Blackburn, 1989; McElligott and Wellinger, 1997). Having a single-stranded 3′ end appears to be critical, since only this form can act as a substrate for the telomere-replicating enzyme telomerase (Lee et al., 1993; Lingner and Cech, 1996).

The protein components of telomeres include factors specific for double strand telomeric DNA, such as yeast Rap1p and human TRF1 and TRF2, and factors specific for single strand telomeric DNA, such as Cdc13p and Est1p from *Saccharomyces cerevisiae* and the telomere end binding proteins from *Oxytricha nova* and *Euplotes crassus* (Shore, 1997). The first of these telomere-binding proteins to be isolated and characterized was the telomere end binding protein from *Oxytricha nova* (Gottschling and Zakian, 1986; Price and Cech, 1987).

The *O. nova* telomere end binding protein (*On*TEBP) contains two protein subunits that tenaciously bind the 16-nucleotide, 3′-terminal, single strand $T_4G_4T_4G_4$ DNA extension found at the ends of macronuclear chromosomes. The 56 kDa α subunit alone binds single strand $(T_4G_4)_n$ DNA in a sequence-specific manner, although differently than in the context of the α:β:ssDNA ternary complex (Gray et al., 1991). The 41 kDa β subunit alone interacts only weakly with DNA, but modifies the DNA-binding properties of α (Gray et al., 1991). α and β do not interact with each other in the absence of DNA, but together with $T_4G_4T_4G_4$ single strand telomeric DNA they form a very stable α:β:ssDNA ternary complex with 1:1:1 stoichiometry (Fang and Cech, 1993b). α contains a 35 kDa N-terminal domain that binds single strand DNA similarly to full-length α and a 21 kDa C-terminal domain that mediates interaction with β (Fang et al., 1993; J. A. R. and S. C. S., unpublished observation). β contains a 28 kDa N-terminal core domain that is sufficient for ternary complex formation and a highly basic C-terminal "tail" that promotes the formation of G quartet DNA structures in vitro, although the existence and function of such structures in vivo is uncertain (Fang and Cech, 1993a).

Here we report the crystal structure of the *Oxytricha nova* telomere end binding protein complexed with its telomeric single strand DNA-binding site at 2.8 Å resolution. The structure provides a molecular description of a nucleoprotein complex that composes the very end of a chromosome and provides insights into the mechanism by which *On*TEBP recognizes, binds, and protects telomeric DNA and how *On*TEBP may regulate telomerase (Froelich-Ammon et al., 1998).

§To whom correspondence should be addressed (e-mail: schultz@mmol.colorado.edu).



α subunit		β subunit	
$S1_\alpha$	52_α-62_α	HA_β	13_β-22_β
$S2_\alpha$	72_α-79_α	$S1_\beta$	41_β-45_β
$S3_\alpha$	98_α-104_α	$S2_\beta$	50_β-54_β
HA_α	106_α-111_α	$S3_\beta$	57_β-61_β
$S4_\alpha$	119_α-129_α	HB_β	68_β-74_β
$S5_\alpha$	133_α-139_α	$S4_\beta$	92_β-99_β
$S6_\alpha$	145_α-149_α	$S5_\beta$	112_β-119_β
HB_α	185_α-202_α	HC_β	142_β-164_β
HC_α	213_α-219_α	HD_β	216_β-221_β
$S7_\alpha$	225_α-235_α		
$S8_\alpha$	240_α-246_α		
$S9_\alpha$	252_α-258_α		
$S10_\alpha$	270_α-274_α		
$S11_\alpha$	295_α-299_α		
HD_α	304_α-312_α		
HE_α	317_α-323_α		
$S12_\alpha$	335_α-338_α		
HF_α	350_α-357_α		
$S13_\alpha$	368_α-379_α		
$S14_\alpha$	412_α-418_α		
$S15_\alpha$	429_α-434_α		
HG_α	455_α-469_α		
$S16_\alpha$	475_α-482_α		
$S17_\alpha$	485_α-494_α		

Prefixes: S, β-strand; H, α-helix. Elements which together compose an OB fold are enclosed by a box.

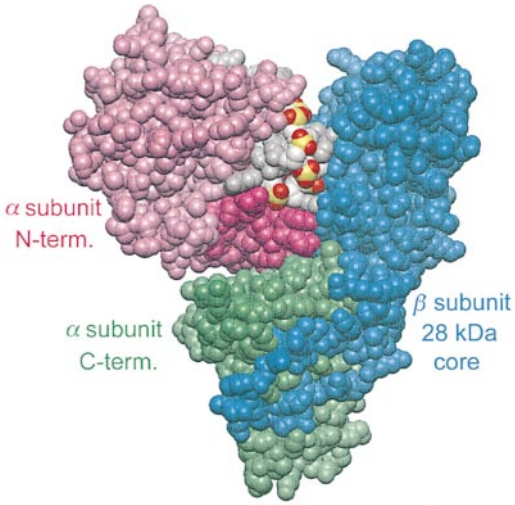


Figure 1. Structure of *On*TEBP

The *On*TEBP–ssDNA complex is shown as an α-carbon ribbon representation in stereo (top), and as a space-filling representation (bottom, right). The DNA is drawn with the base and deoxyribose groups in gray, the phosphorous atoms in yellow, and the phosphate oxygens in red. The α subunit is shown in purple and green, and the 28 kDa N-terminal core domain of the β subunit is shown in blue. The secondary structure elements of the protein are labeled, and residue numbers for the β strands and α helices are listed. α is composed of two domains, an N-terminal domain (dark and light purple) that interacts with the DNA and a C-terminal domain (green) that interacts with β. β contains two parts, a globular portion that interacts with the DNA and an unusual extended peptide loop that interacts with α. The DNA is bound between the N-terminal domain of α and the globular region of β and is folded into an irregular, nonhelical structure with the bases largely buried within the complex and the phosphate groups largely exposed on the surface. *On*TEBP contains four OB folds: two in the α N-terminal domain, one in the α C-terminal domain, and one in the globular portion of β. These four OB folds are found in the four differently colored parts of the structure: light purple, dark purple, green, and blue.

Results

Structure of *On*TEBP Complexed with ssDNA

The structure of *On*TEBP complexed with single strand telomeric DNA is shown in Figure 1. The α subunit contains two structural domains, an N-terminal domain (shown in purple) that interacts with ssDNA and a C-terminal domain (shown in green) that interacts with the β subunit. The β subunit (shown in blue) contains a globular region that interacts with the single strand DNA and a dramatically extended peptide loop that interacts with the C-terminal domain of α. The DNA is bound between the N-terminal domain of α and the globular portion of β and is folded into an irregular, nonhelical structure with the bases largely buried within the complex and the phosphate groups largely exposed on the surface. The 3′ terminal deoxyribose group is buried deep within the complex, and the 3′ hydroxyl hydrogen bonds to a peptide amide of α. Although single strand G-rich DNAs are capable of forming very stable four-stranded structures referred to as G quartets (Sundquist and Klug, 1989; Williamson et al., 1989), such a structure

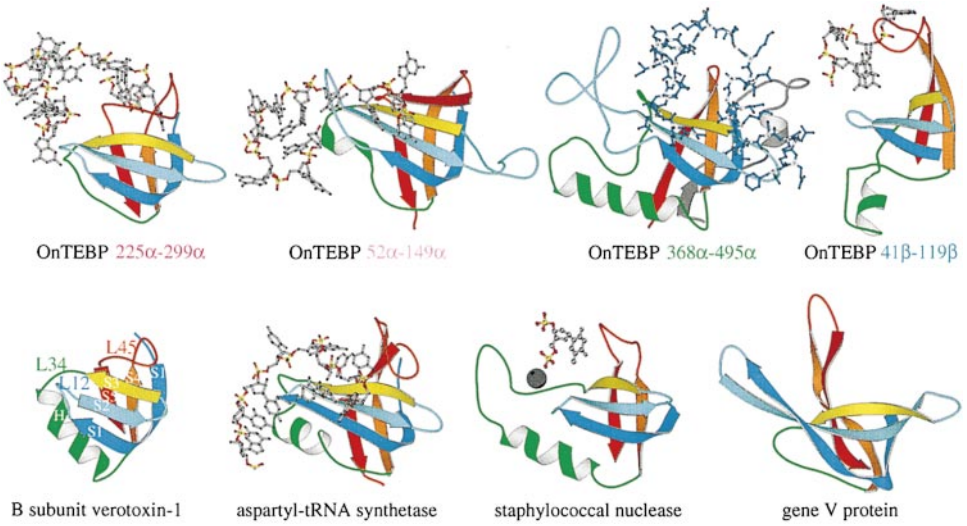


Figure 2. Oligonucleotide/Oligosaccharide–Binding (OB) Folds within *On*TEBP

The four OB folds of *On*TEBP are shown in the top row. Residue limits for each OB fold of *On*TEBP are indicated and the labels are colored as in Figure 1: light purple and dark purple for the two OB folds of the α N-terminal domain, green for the α C-terminal domain, and blue for β. The OB folds from the α N-terminal domain and β subunit each interact with the single strand telomeric DNA, shown as a gray stick model. The OB fold from the α C-terminal domain is shown complexed with the extended peptide loop of the β subunit, shown as a blue stick model. For comparison, the originally described (Murzin, 1993) examples of the OB fold are shown in the bottom row. These are the B subunit of verotoxin-1 from *E. coli*, the anticodon-binding domain of aspartyl-tRNA synthetase complexed with tRNA, staphylococcal nuclease complexed with the Ca^{2+}-pTp inhibitor, and *fd* gene V protein. The strands and loops of the OB fold are colored as follows: strand S1 in blue; strand S2 in cyan; strand S3 in yellow; loop L_{3-4} and the intervening helix H, if present, in green; strand S4 in orange; and strand S5 in red.

is not seen in the *On*TEBP complex. This observation is consistent with previous experiments in solution that demonstrate that G quartets do not bind to *On*TEBP (Raghuraman and Cech, 1990).

*On*TEBP Contains Four Oligonucleotide/Oligosaccharide–Binding Folds

Although amino acid sequence comparisons were only able to relate *On*TEBP to telomere end binding proteins of closely related ciliated protozoa (Fang and Cech, 1991; Wang et al., 1992), the structure reveals that *On*-TEBP is also related to a large class of proteins that contain oligonucleotide/oligosaccharide–binding (OB) folds. *On*TEBP contains four of these OB folds (Figure 2). The OB fold, as originally described for staphylococcal nuclease, the B subunit of *E. coli* verotoxin-1, and the anticodon-binding domain of aspartyl-tRNA synthetase (Murzin, 1993), and as subsequently identified in many other proteins, including a domain of human replication protein A (Bochkarev et al., 1997), consists of five β strands that form two orthogonally packed antiparallel β sheets. The topology of these strands is S1-S2-S3 for sheet 1 and S1-S4-S5 for sheet 2, with the first strand, S1, forming the outer edge of both sheets. An α helix is often found between strands S3 and S4, capping the open edge of this β barrel–like fold.

Substrate binding and ligand recognition involves the loops connecting β strands S1 and S2 (L_{1-2}), strands S3 and S4 (L_{3-4}), and strands S4 and S5 (L_{4-5}) for all currently known OB folds. The corresponding loops found in the four OB folds of *On*TEBP likewise participate in ligand recognition (Figure 2). This scheme of presenting loops at the edge of orthogonally packed β sheets appears to be a general and effective means of generating an extended recognition surface.

The two OB folds that compose the α N-terminal domain (colored light purple and dark purple in Figure 1) are tightly associated with each other, and the ligand recognition loops from these two OB folds cooperate to interact with many of the same nucleotides of the ssDNA. The manner in which the two OB folds of the α N-terminal domain form a single recognition surface is reminiscent of how pairs of immunoglobulin folds recognize ligands; both the OB folds of the α N-terminal domain and immunoglobulins utilize variable loops connecting β strands in two sets of crossing β sheets to form a complex ligand-binding surface.

The OB fold that composes the C-terminal domain of the α subunit (colored green in Figure 1) is found in a novel context complexed not with oligonucleotide or oligosaccharide, but with the oligopeptide loop of the β subunit, thus extending the repertoire of the OB fold to the oligonucleotide/oligosaccharide/oligopeptide (or simply oligomer)–binding fold. Although this interaction involves the same loops that other OB folds utilize to interact with ligands, interactions also extend to additional regions (e.g., L_{2-3}) that are not normally involved with ligand binding. Thus, the C-terminal domain of the α subunit extends our understanding of what is possible for an OB fold both in terms of the type of ligand recognized and the manner in which the binding surface is constructed.

The OB fold contained within the β subunit (colored blue in Figure 1) is interesting because, although the β strands of the canonical OB fold are present in the correct topology, the core is more loosely folded and the loops L_{4-5} and L_{3-4} are extended so that the structure begins to resemble the monomer subunit of the gene V protein (Figure 2). The helix connecting β strands S3 and S4 (helix B_β) is also in a modified position, rotated

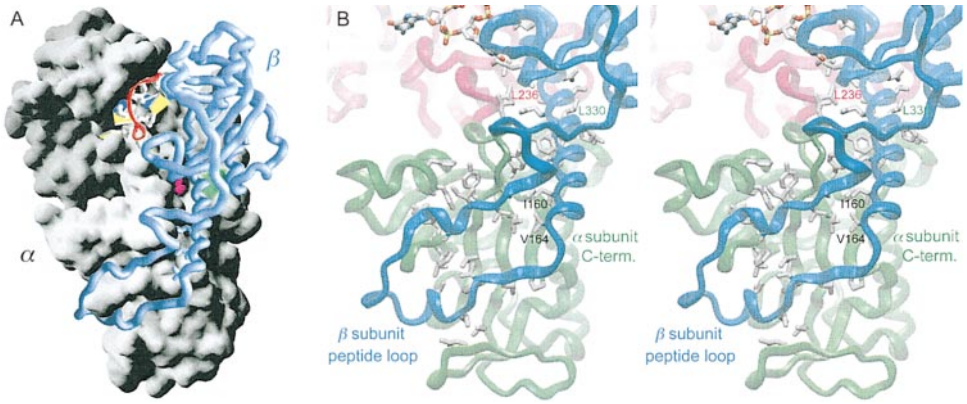


Figure 3. Protein–Protein Interactions in *On*TEBP

(A) Surface representation of the protein–protein association. α is shown as a solvent-accessible surface, β is shown as an α-carbon trace in blue, and the ssDNA is shown as a ribbon with boxes for the base groups (yellow, G; blue, T). A large groove that associates with helix C_β of β is apparent, as are the surfaces that interact with the extended peptide loop of β. Two leucine residues at the surface of α that interact with the N-terminal portion of helix C_β are indicated by the colored patches on α, with purple showing the location of $L236_\alpha$ and green showing the location of $L330_\alpha$. Three residues at the C-terminal region of helix C_β form a hydrophobic ridge, and these are shown in blue. From the middle of helix C_β to the C-terminal end these residues are $L156_\beta$, $I160_\beta$, and $V164_\beta$.

(B) Detailed view of the residues at the protein–protein interface. Notice how helix C_β of β (blue) bridges the N-terminal (purple) and C-terminal (green) domains of α. The peptide loop of β follows directly after the last turn of this helix (residue $V164_\beta$). Hydrophobic side chains (gray) close to the surface of the α C-terminal domain create two hydrophobic patches that align with clusters of hydrophobic residues located along this extended peptide loop of β.

90° relative to the orientation normally observed in OB folds. These structural variations may relate to the special properties of β, which makes critical contacts with both ssDNA and with α in the *On*TEBP–ssDNA complex, but which interacts neither with ssDNA nor with α on its own. The less-than-ideal structure of this OB fold may ensure that incorporation of β into the complex is a cooperative event dependent on the presence of both ssDNA and α, as has been observed in solution studies (Fang et al., 1993).

The Protein–Protein Interface

The interface between the α and β protein subunits involves 52 residues of α and 49 residues of β and buries a total surface area of 5400 Å^2. Two structural features of β are utilized for interaction with α: (1) a long helix, helix C_β, that spans the N- and C-terminal domains of α, and (2) an extended peptide loop that interacts along the surface of the C-terminal domain of α (Figure 3).

The N-terminal region of helix C_β fits into a groove formed by two loops of the α N-terminal domain and one loop from the α C-terminal domain (Figures 1 and 3). The exclusively charged and polar residues of helix C_β in this part of the protein–protein interface establish a network of hydrogen bonds and ionic interactions with charged and polar residues of α. Hydrophobic interactions also contribute, however, in that two leucine side chains, one from the α N-terminal domain ($L236_\alpha$) and one from the α C-terminal domain ($L330_\alpha$), are each inserted beyond the charged and polar ends of lysine, glutamate, aspartate, glutamine, and threonine residues of helix C_β to interact with the aliphatic portions of these side chains. Interaction between the C-terminal portion of helix C_β and the α C-terminal domain is distinctly more hydrophobic, with a ridge consisting of $L156_\beta$, $I160_\beta$, and $V164_\beta$, each placed on consecutive turns of helix C_β, packing into a hydrophobic groove in the α C-terminal domain.

The extended peptide loop of β follows immediately after the end of helix C_β and fits snugly along the surface of the α C-terminal domain (Figure 3). This interaction is unusual in that the peptide loop is not folded together with other portions of β, but folds with the α C-terminal domain, perhaps providing unique properties to the assembly and stability of the α:β:ssDNA complex. Although polar and charged residues predominate along this extended peptide loop of β and hydrogen bond with residues of α, two local regions of hydrophobic interactions also occur. In these two regions, hydrophobic leucine, isoleucine, and valine residues of the β peptide loop are clustered in two short helical structures so as to interact with patches of hydrophobic residues of the α C-terminal domain. These strategically placed hydrophobic contacts appear to anchor the extended peptide loop of β onto the surface of the α C-terminal domain.

The ssDNA–Protein Interface

Each nucleotide of the telomeric single strand DNA interacts with the protein and with the other nucleotides in a unique way, yielding a rich variety of stacking, hydrophobic, hydrogen-bonding, and ionic interactions. These interactions are shown in Figure 4. In general, the phosphodiester groups of the DNA are exposed to solvent, while the deoxyribose groups and the bases are partially or completely buried. Nearly every base stacks face to face with an aromatic amino acid side chain or with another base of the DNA. The bases of nucleotides G2, G3, G4, T5, T8, G9, G10, and G11 hydrogen bond with residues of the protein, often with one hydrogen bond per base, but G3, G4, and G10 each make two or more hydrogen bonds. The protein–DNA interactions can be divided into three regions: (1) a 3′

DNA "loop" that folds between the α and β protein subunits (Figure 4A), (2) an extended stack of nucleotide bases and aromatic protein side chains (Figure 4B), and (3) a folded protein–DNA structure at the 5′ end (Figure 4C). These are detailed below.

3′ DNA Loop Region

In the 3′ DNA loop (Figure 4A), the most 3′ nucleotide, G12, is bound between the α subunit and other portions of the DNA. The deoxyribose group of G12 is completely buried, surrounded by the bases of T8 and T6 on one side and by aromatic and hydrophobic side chains of the α subunit on the other side. The 3′ hydroxyl group hydrogen bonds to the peptide amide of $K66_{\alpha}$. The phosphodiester 5′ of G12 is also completely buried and forms an ion pair with the side chain of $K261_{\alpha}$. The base of G12 stacks against the base of T8 and, on the other face, packs against the deoxyribose group of G11.

The nucleotide G11 is bound within a pocket between the two OB folds of the α N-terminal domain. The deoxyribose group is covered on one side by the base of G12 and on the other side by a leucine residue and the aliphatic carbons of $K261_{\alpha}$, the same lysine side chain that ion pairs with the phosphodiester group of G12. The base of G11 packs between two phenylalanines and an isoleucine on one side and a leucine on the other side. One edge of the base points down into the binding pocket so that N2 hydrogen bonds to the peptide carbonyl of residue $K261_{\alpha}$. The other edge of the base of G11 points out of the binding pocket with N7 partly exposed to solvent. The phosphodiester group 5′ to G11 is also exposed to solvent.

Whereas nucleotides G12 and G11 interact exclusively with residues in α, G10 is bound between the α and β subunits. The base of G10 stacks between a tyrosine of α ($Y239_{\alpha}$) and the guanidinium group of an arginine of β ($R140_{\beta}$). The $Y239_{\alpha}$:G10 interaction is consistent with solution studies that show that DNAs containing 5-bromodeoxyuridine at this position photocross-link with $Y239_{\alpha}$ (Hicke et al., 1994). The guanidinium group of $R140_{\beta}$ makes several interesting interactions, stacking flat against the base of G10 and also hydrogen bonding with a nonbridging oxygen of the phosphodiester 5′ to G10. This same arginine side chain also hydrogen bonds with a loop in β that contains a carbonyl oxygen that hydrogen bonds with N2 of the base. Apparently, arginine is uniquely suited to recognition of single strand DNA since it is able to ion pair with the phosphodiester group, stack with the base (this may be considered a cation-π-type interaction [Dougherty, 1996]), and simultaneously participate in hydrogen bond networks that lock the base into a specific position. Such a remarkable pattern of interactions is seen again for arginine $R274_{\alpha}$, as is described below. The N7 of G10 hydrogen bonds with a buried lysine side chain of β ($K145_{\beta}$). The deoxyribose of G10 packs against a leucine side chain of α ($L258_{\alpha}$). Clearly, G10 mediates an intricate array of interactions that link the two protein subunits.

Of all the nucleotides in the DNA, G9 is the only one that does not interact with the α subunit but interacts exclusively with β. The base and deoxyribose of nucleotide G9 pack against a hydrophobic patch on β composed of two phenylalanines, a proline, and a leucine side chain. One of these phenylalanines, $F106_{\beta}$, stacks with the 5-membered ring of the base and also makes close contact with the deoxyribose group. A glutamate side chain of β ($E45_{\beta}$) hydrogen bonds with N1 and N2 of G9 to precisely position this nucleotide.

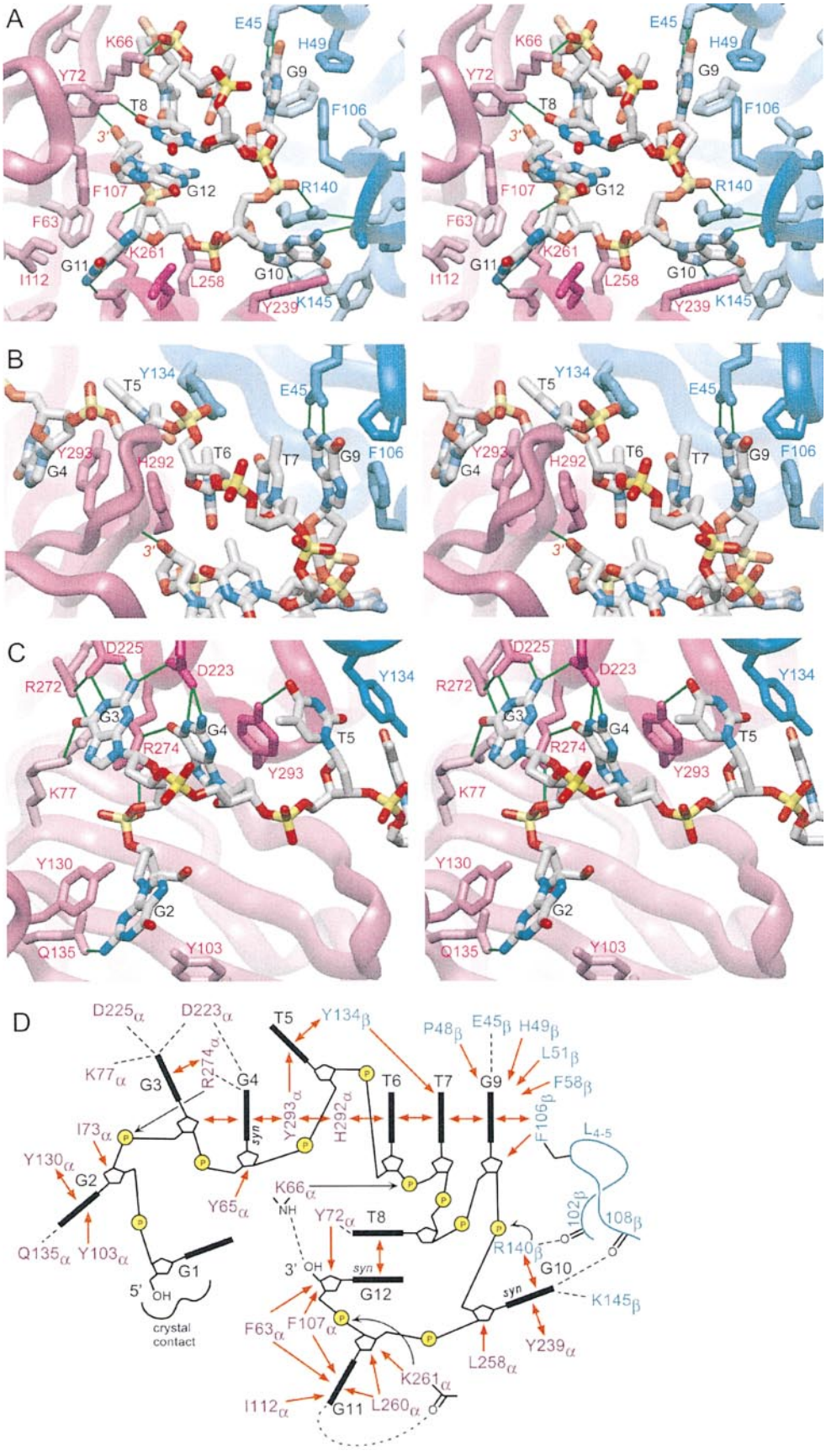
Nucleotide T8 completes the 3′ DNA loop with its base stacked with the base of G12. The methyl group of the base packs into a pocket in the DNA formed by the deoxyribose group of T7 and the edges of bases T6 and T7. The base of T8 hydrogen bonds with a tyrosine residue of α ($Y72_{\alpha}$), but otherwise this nucleotide makes no other contact with either α or β.

Extended Base–Amino Acid Stacking Array

The second region of the ssDNA–protein interface involves an unusually extended stack of nucleotide bases and aromatic amino acid side chains (Figure 4B). One end of this stack involves the $F106_{\beta}$:G9 stack described above. The other face of the base group of G9 stacks with the base of T7, which stacks against T6, which in turn stacks on the imidazole ring of a histidine side chain of the α subunit ($H292_{\alpha}$). The T6:$H292_{\alpha}$ stack observed in the structure is consistent with photocross-linking studies that show that DNAs containing 5-bromodeoxyuridine in place of T6 cross-link with $H292_{\alpha}$ (Hicke et al., 1994). $H292_{\alpha}$ stacks with the edge of a tyrosine side chain also in the α subunit ($Y293_{\alpha}$), and this tyrosine side chain, in turn, stacks squarely with the base of G4. The other face of the base of G4 stacks with the bridging oxygen of the deoxyribose of G3. Interestingly, while the base of G9 and the base of G4 hydrogen bond with protein side chains at each end of the aromatic stack, the intervening bases of T7 and T6 do not. The intervening aromatic side chains also do not hydrogen bond with any other group in the protein. The precisely oriented ends of the extended stack, therefore, apparently contribute to the stabilization of groups throughout the stack in a manner that would be expected to be highly cooperative: one face-to-face contact stabilizes an adjacent contact, which in turn favors formation of the next contact, until all of the aromatic groups are aligned.

Folded Protein–ssDNA Structure at the 5′ End

The third region of the ssDNA–protein interface involves an intricate set of hydrogen bonds and stacking contacts that are folded together at the 5′ end of the ssDNA (Figure 4C). At the center of this folded structure, an arginine residue of the α subunit, $R274_{\alpha}$, simultaneously makes three different interactions with the ssDNA: (1) a hydrogen bond with O6 of G4, (2) a hydrogen bond with the buried phosphodiester group 5′ of G3, and (3) a stacking contact of the guanidinium group flat against the base of G3. The bases of G3 and G4 also form specific hydrogen bonds with two aspartate side chains and a lysine side chain of α. The base of T5 makes a face-to-face stack with tyrosine $Y134_{\beta}$ of β and a face-to-edge contact and a hydrogen bond with $Y293_{\alpha}$, the same tyrosine of α that stacks face to face with the base of G4. Whereas the phosphodiester 3′ of G2 interacts with $R274_{\alpha}$, the base of G2 binds into its own separate binding pocket on the surface of α, where it packs between two tyrosine side chains and hydrogen bonds with a glutamine side chain. The remarkably interrelated manner in which the DNA and protein fold together in this region of the structure is likely important for specificity, as is discussed below.


A
K66
E45
H49
G9
Y72
T8
F106
3'
F107
G12
R140
F63
K261
G11
I112
L258
G10
Y239
K145
B
Y134
T5
E45
Y293
G4
H292
T6
T7
G9
F106
3'
C
D225
D223
Y134
R272
G3
G4
T5
R274
Y293
K77
Y130
G2
Q135
Y103
D
D225α
D223α
T5
Y134β
P48β
E45β
H49β
L51β
K77α
G3
R274α
G4
T6
T7
G9
F58β
F106β
L4-5
I73α
Y293α
H292α
syn
Y130α
G2
K66α
NH
Y65α
Y72α
T8
102β
108β
Q135α
Y103α
G1
3'
OH
syn
R140β
G10
5'
OH
G12
syn
K145β
crystal
contact
F63α
F107α
L258α
Y239α
K261α
I112α
G11
L260α

The positions of G1 and the phosphodiester 3′ of G1 appear to be disrupted by crystal packing contacts, which would explain why crystals of *On*TEBP do not grow with ssDNAs containing additional nucleotides at the 5′ end. There are subtle differences in the binding of 12-nucleotide $G_4T_4G_4$ DNA versus full-length $T_4G_4T_4G_4$ DNA; however, both of these DNAs form stable complexes with *On*TEBP in solution (J. M. B. and S. C. S., unpublished). Just beyond the crystal packing contact are three strands of a β sheet of α that present several aromatic, hydrophobic, and positively charged side chains that likely interact with the remaining 5′ TTTT nucleotides of the natural 16-nucleotide single strand telomeric DNA. Positioned at the far edge of the β sheet is $Y142_\alpha$, which photocross-links to DNA if one of the 5′ Ts is replaced with 5-bromodeoxyuridine (Hicke et al., 1994). On the basis of these observations, we expect that the native ssDNA continues across this β sheet and that the entire 16-nucleotide length of single strand telomeric DNA is recognized by one *On*TEBP molecule.

Discussion

Motifs for Sequence-Specific Recognition of Single Strand Nucleic Acid

The structure of *On*TEBP complexed with ssDNA provides a detailed look at how a protein can recognize single strand nucleic acid in a sequence-specific manner. This structure adds to and complements information obtained from structures of sequence-nonspecific protein–ssDNA complexes, including those of the large (Klenow) fragment of DNA polymerase I (Freemont et al., 1988), gene 32 protein (Shamoo et al., 1995), and a domain of human replication protein A (Bochkarev et al., 1997), as well as protein–ssRNA complexes including those of ssRNA viruses (Chen et al., 1989), amino acid–tRNA synthetases (Rould et al., 1989; Ruff et al., 1991), splicesomal proteins U1A (Oubridge et al., 1994) and U2B''-U2A' (Price et al., 1998), and the mRNA cap protein (Marcotrigiano et al., 1997). Although a rich variety of interactions are observed in each of these complexes, certain themes are emerging. The bases are partially or completely buried and often stack with aromatic amino acid side chains. The phosphodiester-sugar backbone is largely solvent exposed, and any phosphodiester groups that are buried ion pair with lysine or arginine side chains. And finally, the protein residues interacting with the nucleic acid are generally presented at the surface of β sheets or in loops connecting β strands.

The *On*TEBP–ssDNA complex adds to these themes in several ways. First, certain nucleotides of the telomeric DNA are more deeply buried than nucleotides in other protein–single strand nucleic acid complexes, a feature that likely relates to sequence-specific recognition and the function of capping and protecting the ends of chromosomes. These deep pockets in *On*TEBP are formed by close associations between protein folds or domains. At nucleotide G11, the deep pocket that encloses the base is found between the two OB folds that compose the α N-terminal domain. At G10, the deep pocket forms at the interface between the N-terminal domains of α and β. From these examples, it would appear that pairs of protein folds or domains can more readily form intimate and specific interactions with single strand nucleic acid than can a single protein fold.

Another unique feature is the use of arginine side chains for specific interaction with G nucleotides. At both G3 and G10, an arginine side chain forms a salt bridge with the 5′ phosphodiester group, stacks flat against the base, and participates in a network of hydrogen bonds that potentially contributes to specific base recognition. The remarkable properties of arginine apparently make this amino acid especially well suited to the recognition of single strand DNA since the different functionalities of the guanidinium group can form multiple interactions with the nucleic acid base and phosphodiester groups and with other protein functional groups. An arginine has also been observed to stack with a C nucleotide in the structure of a domain of replication protein A complexed with ssDNA (Bochkarev et al., 1997). This arginine does not participate in a hydrogen

Figure 4. Protein–ssDNA Interactions in *On*TEBP

The α subunit is shown in purple, the β subunit is shown in blue, and the DNA is colored according to atom types: carbon in gray, oxygen in red, nitrogen in blue, and phosphorous in yellow. Hydrogen bonds are shown as solid green lines.

(A) Protein–ssDNA interactions in the 3′ DNA loop region. Stacking of the bases of T8 and G12 define a loop in the DNA that binds in a cleft between α and β. Each nucleotide interacts with the protein in a unique way. The 3′ hydroxyl group of G12 is deeply buried in the complex and hydrogen bonds with a backbone amide of α. G11 is bound in a hydrophobic pocket in α. G10 is intricately positioned between α and β. G9 interacts with a hydrophobic patch of β and is the only nucleotide that does not interact with α.

(B) Extended DNA–protein stacking interactions. Bases of the ssDNA and amino acid side chains of both α and β are aligned in a dramatic array of stacked aromatic groups including, from right to left, the side chain of $F106_\beta$, the base of G9, the base of T7, the base of T6, the side chain of $H292_\alpha$, the side chain of $Y293_\alpha$, and the base of G4. The other face of the base of G4 stacks with the O1' oxygen of the deoxyribose of G3.

(C) Folded protein–DNA structure at the 5′ end. In this region of the structure, multiple hydrogen bonds and stacking interactions have been assembled in a highly interrelated manner. At the center of this folded structure, an arginine residue, $R274_\alpha$, simultaneously makes three different interactions with the ssDNA: (1) a hydrogen bond with O6 of G4, (2) a hydrogen bond with the phosphodiester 5′ of G3, and (3) a stacking contact of the guanidinium group flat against the base of G3. The nucleotide G1 is omitted from this picture because it is most likely in a nonnative position as the consequence of a crystal packing contact.

(D) Schematic representation of the protein–ssDNA interactions. Bases of the ssDNA are shown as black rectangles; phosphate groups are shown as yellow circles; residues in α are indicated with purple labels; and residues in β are indicated with blue labels. Hydrogen bonds are shown as dotted lines, as for the peptide amide of $K66_\alpha$, which hydrogen bonds with the 3′ hydroxyl group of G12. Ionic interactions are shown as solid, black arrows, as for lysine $K66_\alpha$, which forms a salt bridge with the phosphate group 5′ of T7. Close contacts between hydrophobic groups are drawn as orange arrows, as for leucine $L258_\alpha$, which packs against the deoxyribose group of G10. If aromatic groups are stacked together, this is indicated by an orange, double-headed arrow, as for $F106_\beta$, which stacks with the base of G9.

bond network that could distinguish one base from another, however, as is consistent with sequence-independent DNA binding by replication protein A.

Cooperativity among different protein–ssDNA contacts probably also contributes to sequence-specific recognition in the *On*TEBP–ssDNA complex. For instance, the interactions of $R140_{\beta}$ with the 5′ phosphate and base of nucleotide G10 position this side chain to hydrogen bond with loop L_{4-5} of β, and this loop, in turn, hydrogen bonds with the base of G10 (Figure 4A). Together with these interactions, the side chain of $F106_{\beta}$ is presented at the tip of L_{4-5} so as to stack with the base of G9. On the basis of these observed interactions, it seems plausible that the conformation of L_{4-5} seen in the structure results from a process of local folding, as multiple contacts with the DNA ligand are established. In this region of the structure, the DNA has also adopted a very distinctive fold, the 3′ DNA "loop," as the bases are placed into individual binding pockets on α and β. Thus, it appears that the protein and the ssDNA finish folding only as they make contact with each other in a process of "cofolding."

The evidence for "cofolding" of the protein and ssDNA is indirect but compelling: there is no pathway by which the ssDNA could enter into its binding site if the protein surfaces were fixed in a preformed structure. Furthermore, it is difficult to imagine that the protein loops found at the binding surface and the particular fold of the ssDNA would be stable in the conformations seen in the complex without the observed protein–ssDNA interactions to hold them together. The 3′ DNA loop region discussed above provides but one example of this. The array of stacking interactions (Figure 4B) and the hydrogen bonds and stacking interactions at the 5′ DNA structure (Figure 4C) also support the idea that the structure we see in the crystal results from a process of cofolding of DNA and protein. Thermodynamic and structural evaluation of double strand DNA-binding proteins suggest that local and global folding of protein likely participates in the recognition process (Spolar and Record, 1994). The rich array of interwoven interactions observed in the complex of *On*TEBP with ssDNA suggests that folding of both protein and DNA may contribute to recognition of single strand telomeric DNA.

Sequence-specific recognition of G-rich ssDNA may also involve the conformation about the C1'-N glycosyl bond. In the *On*TEBP complex, three of the G nucleotides adopt the *syn* orientation. While the A, C, and T nucleotides prefer the *anti* conformation, 5′ guanylic acid readily adopts the *syn* form in solution, as observed by CD (Guschlbauer et al., 1972) and NMR (Son et al., 1972). Quantum chemical calculations have shown that van der Waals and electrostatic attractions between the exocyclic amino group of guanosine and the 5′ phosphate are associated with delocalization of the lone pair electrons of the $-NH_2$ group and are responsible for the preferred *syn* conformation (Saenger, 1984). Indeed, the N2 amino group and the 5′ phosphate are in close proximity in all three nucleotides with the *syn* orientation in the *On*TEBP–ssDNA complex (G4, G10, and G12). Specificity of *On*TEBP for G at certain positions may depend, in part, on the exceptional properties of guanosine, which naturally favors the *syn* orientation more so than the other three nucleotides. The ability to specifically recognize G in ssDNA on the basis of the *syn/anti* equilibrium may also relate to the conserved G-rich character of telomeric sequence repeats.

DNA-Mediated Protein–Protein Interactions

Why does the association of α and β depend on the presence of single strand DNA as has been observed in solution studies (Fang and Cech, 1993b)? The structure suggests a direct and an indirect mechanism for DNA-dependent protein–protein association. In the direct mechanism, DNA acts as the "mortar," holding portions of the two proteins together. An example of this type of DNA-mediated protein–protein interaction is the base of nucleotide G10, which makes contact with a tyrosine side chain of α and with an arginine side chain from β (Figure 4A). Similarly, the base of nucleotide T5 is sandwiched between a tyrosine of β and a tyrosine of α (Figure 4B), and the extended stacking array clearly links the two subunits via the DNA (Figure 4B). In the indirect mechanism, protein–ssDNA interactions assist to structure loops of one protein, which in turn interact with the second protein. Specifically, two loops of the α N-terminal domain are in simultaneous contact with the ssDNA and with the β subunit. Interestingly, these two loops correspond to the "recognition loops" of the second OB fold in the α N-terminal domain. If the contacts with ssDNA stabilize loop conformations that favor formation of contacts with β, then a cooperative interdependence of protein–protein and protein–ssDNA association is established.

There also exists at the protein–protein interface a large number of favorable contacts that do not appear to depend on DNA. Indeed, direct protein–protein interactions bury a surface that is as large or larger than those observed in other multimeric proteins (Janin et al., 1988). Therefore, while DNA contacts would certainly be expected to promote protein association, it is still curious that α and β absolutely require DNA for association. Perhaps the cavities that remain at the DNA recognition surface upon removal of DNA are sufficiently destabilizing that the proteins respond by adopting some other fold or oligomerization state that precludes association of α and β. This answer is a plausible extension of the concept that the DNA and regions of protein structure "cofold" upon complex formation. The key to understanding the DNA requirement for protein association may thus be hidden within the structures of the uncomplexed subunits.

Implications for Telomere Biology

The structure of *On*TEBP seems well suited to a protein that recognizes and caps the 3′ terminus of the chromosome. The binding pockets in the protein fold the 3′ end of the ssDNA into a conformation that only a ssDNA with a free 3′ end can adopt. In this way, the very terminus of the single strand telomeric DNA is identified. Furthermore, the 3′ end is completely buried within the complex, consistent with the observations that in solution, *On*TEBP protects the telomeric ssDNA from Bal31 digestion (Gottschling and Zakian, 1986) and prevents extension of the ssDNA by telomerase (Froelich-Ammon

Table 1. Data Collection and Refinement Statistics

Data Collection

	Native	Hg-1	Hg-2	Hg-3	Hg-4	Hg-S1	Hg-S2	Hg-S3
Resolution limits (Å)	30-2.8	12-2.8	12-2.8	12-3.1	12-3.1	12-3	12-2.9	12-2.95
Total observations	70,514	36,295	49,072	26,000	27,598	34,063	19,869	32,146
Unique observations	24,061	19,505	15,106	9,297	18,386	17,716	15,075	18,098
Completeness (%)	82.4	67.7	52.5	43.5	86.2	75.4	57.6	73.3
R_{sym}	0.084	0.108	0.108	0.105	0.097	0.098	0.106	0.109
R_{ano}	---	---	0.126	0.152	---	---	---	---
R_{iso}	---	0.103	0.106	0.110	0.157	0.109	0.193	0.213
R_{Cullis}	---	0.64	0.65	0.65	0.69	0.65	0.75	0.73
Phasing power	---	1.84	1.76	1.68	1.53	1.76	1.34	1.35
Sites occupied [a][b][c]	---	●○○	●○○	●○○	●◉○	●○○	●◉◉	●●◉

Model Refinement

Data collection	Native	r.m.s. deviations from target values		Ramachandran statistics	(%, N)
Resolution limits (Å)	20-2.8	Bond lengths (Å)	0.008	Most favored	85.0, 514
Data selected	all data	Bond angles (°)	1.41	Additionally allowed	14.0, 84
Bulk solvent mask	yes	Dihedral angles (°)	25.3	Generously allowed	1.0, 6
R_{cryst}	0.255	Improper angles (°)	1.13	Disallowed	0.0, 0
R_{free}	0.301				

Average *B*-factors (Å^2) 28.2 (21.8 for α; 42.1 for β, and 24.0 for the ssDNA)

$R_{sym} = \Sigma|Ih - \langle Ih\rangle|/\Sigma Ih$, where $\langle Ih\rangle$ is the average intensity over symmetry equivalents. $R_{ano} = \Sigma|Ih^+ - Ih^-|/\Sigma Ih$, where Ih^+ and Ih^- are Friedel reflection mates. $R_{iso} = \Sigma|I_{native} - I_{derivative}|/\Sigma I_{native}$. $R_{cullis} = \Sigma||F_{PH}| - |F_P + F_H||/\Sigma F_H$ where F_{PH} and F_P are the measured derivative and native structure factors and F_H is the calculated heavy atom structure factor. Phasing power = $(\Sigma|F_H|^2/\Sigma|E|^2)^{1/2}$, where $\Sigma|E|^2 = \Sigma(|F_{PH}|_{obs} - |F_{PH}|_{calc})^{1/2}$. Occupancy at three mercury binding sites is indicated by differently shaded circles: black, fully occupied; gray, partially occupied; open, unoccupied. Site *a* was in the β subunit, site *b* was in the α subunit, and site *c* was engineered by introducing a non-bridging sulfur at the phosphate between G11 and G12. For the derivative Hg-S2, two additional weak sites were found. R_{cryst} and $R_{free} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes. R_{free} was calculated with 10% of the reflections not used during refinement. Ramachandran plot statistics were calculated by PROCHECK.

et al., 1998). Since the entire 16-nucleotide extension of telomeric ssDNA binds with one α:β heterodimer, *On*TEBP could act as the "measuring caliper" that ensures that an exact length of ssDNA is placed at the end of each chromosome: by sequestering the 3′ end of the DNA, *On*TEBP could turn off telomere extension by telomerase once the 16-nucleotide single strand DNA extension is recognized and capped.

Generality for Other Telomeric ssDNA-Binding Factors

Telomere-binding proteins from *Euplotes crassus* (Wang et al., 1992) and *Stylonychia mytilis* (Fang and Cech, 1991) share homology with the *Oxytricha nova* telomere end binding protein, suggesting that recognition and protection of telomere ends in these other ciliates resembles the capping of 3′ ends by *On*TEBP. Whereas the *Stylonichia* proteins are very closely related to α (79% amino acid sequence identity) and β (77% identity) (Fang and Cech, 1991), the *Euplotes* proteins exhibit strong homology only with the N-terminal DNA-binding domain of α, and there does not appear to be a homolog of β (Wang et al., 1992). The sequence of the 3′ terminal extension in telomeres from *Euplotes* is shorter by two nucleotides relative to those of *Oxytricha* and *Stylonychia*, implying that *Euplotes* may have dispensed with β by shortening the 3′ terminal extension by two nucleotides. This suggestion is consistent with a role of telomere end binding proteins in determining the precise length of the 3′ terminal extension of the telomere (Vermeesch and Price, 1994).

Recently, two proteins of *S. cerevisiae*, Cdc13p (Lin and Zakian, 1996; Nugent et al., 1996) and Est1p (Virta-Pearlman et al., 1996), have also been shown to specifically bind single strand telomeric DNA, and Est1p appears to recognize specifically the 3′ end of such single strand DNA. It is interesting to speculate that the structural principles uncovered for the *O. nova* telomere end binding protein might also be involved in other systems, such as the telomeric ssDNA-binding factors from *S. cerevisiae*. Although amino acid sequence comparisons between *On*TEBP, Cdc13p, and Est1p do not reveal any obvious similarities, OB folds, which compose the DNA recognition surfaces in *On*TEBP, are reliably identified only by structural homology and not by sequence homology (Murzin, 1993). Thus, it remains to be seen

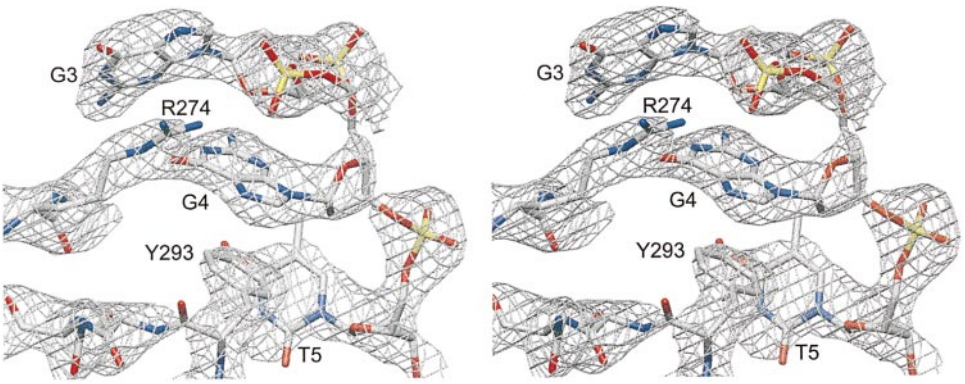


Figure 5. Simulated Annealed Omit Electron Density Map of a Portion of the *On*TEBP–ssDNA Complex Calculated at 2.8 Å Resolution and Contoured at 1.8σ

whether the OB fold is a general solution for telomeric ssDNA-binding factors, or whether other telomeric proteins achieve similar functions by using different structures.

Experimental Procedures

Sample Preparation

The α and β protein subunits were expressed separately in *E. coli*, strain BL21(DE3)pLysS, using T7-based expression systems (Gray et al., 1991; J. A. R. and S. C. S., unpublished). For the α subunit, cells were grown at 37°C to an optical density at 600 nm (OD_{600}) of ~0.3 and then cooled to room temperature and grown to an OD_{600} of ~1.0. Production of protein was induced by addition of isopropyl-β-D-thiogalactoside to 0.5 mM, and the cells were incubated overnight at room temperature. Cells were collected by centrifugation and frozen at −20°C. The cells were resuspended and lysed by sonication in a solution containing 50 mM 4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid (HEPES) (pH 7.5), 50 mM NaCl, 1 mM ethyldiamine tetraacetic acid (EDTA), 0.02% NaN_3, and 2 mM dithiothreitol (DTT). The protein was precipitated by addition of ammonium sulfate to 70% of saturation, purified by ion exchange chromatography using S-Sepharose and Q-Sepharose columns (Pharmacia, Uppsala, Sweden), and further purified using a Superdex-75 gel filtration column (Pharmacia). Typical yields were 10–15 mg of pure protein from 10 g of *E. coli* (wet weight). The 28 kDa N-terminal core domain of β was expressed and purified using an identical procedure, except that the Q-Sepharose column step was omitted. Typical yields for the 28 kDa N-terminal core domain of β were 15–20 mg of pure protein from 10 g of *E. coli* (wet weight).

The single strand 12-nucleotide DNA ($G_4T_4G_4$) was obtained by chemical synthesis and purified by HPLC with the use of a reverse phase C_4 column (Vydac, Hesperia CA 92345).

Crystallization

Crystals of the ternary complex containing the α subunit, the 28 kDa N-terminal core domain of the β subunit, and the 12-nucleotide single strand DNA were grown by the hanging drop vapor diffusion method at room temperature. Three microliters of the α:β28:ssDNA complex (200 μM α subunit with 1.0–1.5 equivalents each of the β subunit and the ssDNA in 50 mM NaCl, 5 mM Tris (pH 7.5), 0.1 mM EDTA, 0.02% NaN_3, 2 mM DTT) plus three microliters of precipitant (30% ethylene glycol, 3%–8% polyethylene glycol 4000, 40 mM 2-(4-morpholino)-ethane sulfonic acid [MES] [pH 5.5–6.5], 0.02% NaN_3, 2 mM DTT) were equilibrated against 1 ml of precipitant with 50 mM NaCl added. Hexagonal crystals appeared within 2–3 days and grew to full size (0.4 mm × 0.2 mm) over a 2–4 week period. The crystals were hexagonal, space group $P6_122$ with a, b = 94.6 Å and c = 426.6 Å.

Data Collection

Crystals were harvested by transfer to a solution containing 12% polyethylene glycol 4000, 30% ethylene glycol, 40 mM MES (pH 6.5), 0.02% NaN_3, and 2 mM DTT. Heavy atom derivatives were prepared by soaking crystals in harvesting solutions that contained 1 mM mercury (II) acetate or methylmercury (II) chloride but did not contain DTT. Native and derivative data were collected from crystals maintained at room temperature on an R-AXIS II/Rigaku RH2 system equipped with the Yale focusing mirrors. For two of the derivative crystals, anomolous data was collected at ±180° for each 1.5° oscillation. Data were reduced to reflection intensities using either the Molecular Structure Corporation software or the programs DENZO and SCALEPACK (Otwinowski and Minor, 1996).

Structure Determination and Refinement

The structure was solved by multiple isomorphous replacement methods with the use of seven mercury-containing heavy atom derivatives that exhibited differential occupancies for three major mercury-binding sites (Table 1). Heavy atom positions and occupancies were refined, and phases were calculated using MLPHARE (Otwinowski, 1991). The MIR phases were modified by solvent flattening and histogram matching using DM (Cowtan and Main, 1998). The initial model was built into the solvent-flattened MIR electron density maps using the program O (Jones et al., 1991). Positioning of the ssDNA was assisted by the location of a bound mercury atom in crystals containing a phosphorothioate-modified ssDNA. Complete modeling of the β subunit required several rounds of phase combination with SIGMAA (Read, 1986) with model-derived phases weighted by 0.1 relative to experimentally derived MIR phases. The structure was refined first with X-PLOR (Brünger, 1993) and subsequently with CNS (Brünger et al., 1998). The current structure consists of 5687 nonhydrogen atoms: residues 37_α–495_α of the α subunit, residues 10_β–222_β of the β subunit, nucleotides G1–G12 of the ssDNA, and 30 solvent molecules. A simulated annealed omit electron density for a portion of the protein–ssDNA interface is shown in Figure 5.

Acknowledgments

We thank Tom Cech for many insightful and helpful contributions throughout the course of this work. We thank John Gray, Guowei Fang, and Brian Hicke for their assistance during the early stages of this work; the members of the Schultz lab for their ongoing assistance; and Tom Cech and Olke Uhlenbeck for their critical reading of this manuscript. This work was supported by the American Cancer Society (research project grant numbers NP-844A and RPG-93-011-04-NP). M. P. H. is a research fellow of the Helen Hay Whitney Foundation. J. A. R. was a fellow of the Damon Runyon-Walter Winchell Cancer Research Fund. J. M. B. was supported by an NIH postdoctoral fellowship (grant number 5 F32 GM17155-02).

Received September 23, 1998; revised November 2, 1998.

References

Blackburn, E.H. (1991). Structure and function of telomeres. Nature *350*, 569–573.

Bochkarev, A., Pfuetzner, R.A., Edwards, A.M., and Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. Nature *385*, 176–181.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of life-span by introduction of telomerase into normal human cells. Science *279*, 349–352.

Brünger, A.T. (1993). X-PLOR Manual Version 3.1: A System for X-Ray Crystallography and NMR (New Haven: Yale University Press).

Brünger, 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., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.

Chen, Z.G., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G., and Johnson, J.E. (1989). Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution. Science *245*, 154–159.

Cowtan, K., and Main, P. (1998). Miscellaneous algorithms for density modification. Acta Crystallogr. D *54*, 487–493.

Dougherty, D.A. (1996). Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. Science *271*, 163–168.

Fang, G.W., and Cech, T.R. (1991). Molecular cloning of telomere-binding protein genes from *Stylonychia mytilis*. Nucleic Acids Res. *19*, 5515–5518.

Fang, G., and Cech, T.R. (1993a). The beta subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. Cell *74*, 875–885.

Fang, G., and Cech, T.R. (1993b). *Oxytricha* telomere-binding protein: DNA-dependent dimerization of the alpha and beta subunits. Proc. Natl. Acad. Sci. USA *90*, 6056–6060.

Fang, G., Gray, J.T., and Cech, T.R. (1993). *Oxytricha* telomere-binding protein: separable DNA-binding and dimerization domains of the alpha-subunit. Genes Dev. *7*, 870–882.

Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R., and Steitz, T.A. (1988). Cocrystal structure of an editing complex of Klenow fragment with DNA. Proc. Natl. Acad. Sci. USA *85*, 8924–8928.

Froelich-Ammon, S.J., Dickinson, B.A., Bevilacqua, J.M., Schultz, S.C., and Cech, T.R. (1998). Modulation of telomerase activity by telomere DNA-binding proteins in *Oxytricha*. Genes Dev. *12*, 1504–1514.

Gottschling, D.E., and Zakian, V.A. (1986). Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. Cell *47*, 195–205.

Gray, J.T., Celander, D.W., Price, C.M., and Cech, T.R. (1991). Cloning and expression of genes for the *Oxytricha* telomere-binding protein: specific subunit interactions in the telomeric complex. Cell *67*, 807–814.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. Cell *43*, 405–413.

Guschlbauer, W., Fric, I., and Holy, A. (1972). Oligonucleotide conformations. Optical studies on GpU analogues with modified-uridine residues. Eur. J. Biochem. *31*, 1–13.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458–460.

Henderson, E.R., and Blackburn, E.H. (1989). An overhanging 3′ terminus is a conserved feature of telomeres. Mol. Cell. Biol. *9*, 345–38.

Hicke, B.J., Willis, M.C., Koch, T.H., and Cech, T.R. (1994). Telomeric protein-DNA point contacts identified by photo-cross-linking using 5-bromodeoxyuridine. Biochemistry *33*, 3364–3373.

Janin, J., Miller, S., and Chothia, C. (1988). Surface, subunit interfaces and interior of oligomeric proteins. J. Mol. Biol. *204*, 155–164.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Klobutcher, L.A., Swanton, M.T., Donini, P., and Prescott, D.M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3′ terminus. Proc. Natl. Acad. Sci. USA *78*, 3015–3019.

Lee, M.S., Gallagher, R.C., Bradley, J., and Blackburn, E.H. (1993). *In vivo* and *in vitro* studies of telomeres and telomerase. Cold Spring Harbor Symp. Quant. Biol. *58*, 707–718.

Lin, J.J., and Zakian, V.A. (1996). The *Saccharomyces* CDC13 protein is a single-strand TG1-3 telomeric DNA-binding protein *in vitro* that affects telomere behavior *in vivo*. Proc. Natl. Acad. Sci. USA *93*, 13760–13765.

Lingner, J., and Cech, T.R. (1996). Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3′ overhang. Proc. Natl. Acad. Sci. USA *93*, 10712–10717.

Lundblad, V., and Szostak, J.W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. Cell *57*, 633–643.

Marcotrigiano, J., Gingras, A.C., Sonenberg, N., and Burley, S.K. (1997). Cocrystal structure of the messenger RNA 5′ cap-binding protein (eIF4E) bound to 7-methyl-GDP. Cell *89*, 951–961.

McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. Genetics *26*, 234–282.

McElligott, R., and Wellinger, R.J. (1997). The terminal DNA structure of mammalian chromosomes. EMBO J. *16*, 3705–3714.

Murzin, A.G. (1993). OB (oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J. *12*, 861–867.

Nugent, C.I., Hughes, T.R., Lue, N.F., and Lundblad, V. (1996). Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. Science *274*, 249–252.

Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In Isomorphous Replacement and Anomalous Scattering, W. Wolf, P.R. Evans, and A.G.W. Leslie, eds. (Daresbury UK: Daresbury Laboratory).

Otwinowski, Z., and Minor, W. (1996). Processing of x-ray diffraction data collectin in oscillation mode. Methods Enzmol. *276*, 307–326.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature *372*, 432–438.

Price, C.M., and Cech, T.R. (1987). Telomeric DNA-protein interactions of *Oxytricha* macronuclear DNA. Genes Dev. *1*, 783–793.

Price, S.R., Evans, P.R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645–650.

Raghuraman, M.K., and Cech, T.R. (1990). Effect of monovalent cation-induced telomeric DNA structure on the binding of *Oxytricha* telomeric protein. Nucleic Acids Res. *18*, 4543–4552.

Read, R.J. (1986). Improved fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. A *42*, 140–149.

Rould, M.A., Perona, J.J., Soll, D., and Steitz, T.A. (1989). Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. Science *246*, 1135–1142.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., and Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). Science *252*, 1682–1689.

Saenger, W. (1984). Principles of Nucleic Acid Structure (New York: Springer-Verlag).

Shamoo, Y., Friedman, A.M., Parsons, M.R., Konigsberg, W.H., and Steitz, T.A. (1995). Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. Nature *376*, 362–366.

Shore, D. (1997). Telomerase and telomere-binding proteins: controlling the endgame. Trends Biochem. Sci. *22*, 233–235.

Son, T.D., Guschlbauer, W., and Gueron, M. (1972). Flexibility and conformations of guanosine monophosphates by the Overhauser effect. J. Am. Chem. Soc. *94*, 7903–7911.

Spolar, R.S., and Record, M.T., Jr. (1994). Coupling of local folding to site-specific binding of proteins to DNA. Science *263*, 777–784.

Sundquist, W.I., and Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature *342*, 825–829.

Vermeesch, J.R., and Price, C.M. (1994). Telomeric DNA sequence and structure following de novo telomere synthesis in *Euplotes crassus*. Mol. Cell. Biol. *14*, 554–566.

Virta-Pearlman, V., Morris, D.K., and Lundblad, V. (1996). Est1 has the properties of a single-stranded telomere end-binding protein. Genes Dev. *10*, 3094–3104.

Wang, W., Skopp, R., Scofield, M., and Price, C. (1992). *Euplotes crassus* has genes encoding telomere-binding proteins and telomere-binding protein homologs. Nucleic Acids Res. *20*, 6621–6629.

Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell *59*, 871–880.

Yu, G.L., Bradley, J.D., Attardi, L.D., and Blackburn, E.H. (1990). *In vivo* alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. Nature *344*, 126–132.

Zakian, V.A. (1995). Telomeres: beginning to understand the end. Science *270*, 1601–1607.

Brookhaven Protein Data Bank ID Code

The coordinates for the *On*TEBP–ssDNA complex have been deposited in the Brookhaven Protein Data Bank under ID code 1otc.