Articles

Sequence-Specific and 3′-End Selective Single-Strand DNA Binding by the *Oxytricha nova* Telomere End Binding Protein α Subunit†

Scott Classen,*,‡ Dan Lyons,§ Thomas R. Cech, and Steve C. Schultz||

Department of Chemistry and Biochemistry and Howard Hughes Medical Institute, University of Colorado, Boulder, Colorado 80309-0215

Received December 17, 2002; Revised Manuscript Received June 3, 2003

ABSTRACT: *Oxytricha nova* telomere end binding protein (*On*TEBP) specifically recognizes and caps single-strand $(T_4G_4)_2$ telomeric DNA at the very 3′-ends of *O. nova* macronuclear chromosomes. The discovery of proteins homologous to the N-terminal domain of the *On*TEBP α subunit in *Euplotes crassus*, *Schizosaccharomyces pombe*, and *Homo sapiens* suggests that related proteins are widely distributed in eukaryotes. Previously reported crystal structures of the ssDNA binding domain of the *On*TEBP α subunit both uncomplexed and complexed with telomeric ssDNA have suggested specific mechanisms for sequence-specific and 3′-end selective recognition of the single-strand telomeric DNA. We now describe comparative binding studies of ssDNA recognition by the N-terminal domain of the *On*TEBP α subunit. Addition of nucleotides to the 3′-end of the TTTTGGGG telomere repeat decreases the level of α binding by up to 7-fold, revealing a modest specificity for a 3′-terminus relative to an internal DNA binding site. Nucleotide substitutions at specific positions within the $t_1t_2t_3T_4G_5G_6\mathbf{G}_7G_8$ repeat show that base substitutions at some sites do not substantially decrease the binding affinity (<2-fold for lowercase letters), while substitutions at other sites dramatically reduce the binding affinity (>20-fold decrease for the uppercase bold letter). Comparison of the structural and binding data provides unique insights into the ways in which proteins recognize and bind single-stranded DNA.

Telomeres are the specialized nucleoprotein complexes that compose the very ends of linear eukaryotic chromosomes (*1*, *2*). Telomeres serve a variety of critical functions in cells. They form protective caps at the ends of chromosomes that prevent nucleolytic degradation, recombination, and end-to-end fusion (*3*). Telomeres serve as substrates for the ribonucleoprotein enzyme telomerase, which adds telomeric DNA to the 3′-ends of chromosomes to offset the loss that results from the primer dependence of DNA replication (*4*−*6*). Telomeres also participate in chromosome−chromosome interactions and chromosomal organizations within the nucleus forming condensed chromatin structures that are localized near the nuclear periphery in many organisms (*7*−*9*).

*On*TEBP[1] is composed of two protein subunits, a 56 kDa α subunit and a 41 kDa *β* subunit, which together bind to the 16-nucleotide ssDNA extension at the 3′-ends of *Oxytricha nova* macronuclear chromosomes (*10*, *11*). The macronuclear genes for *O. nova* telomere end binding protein (*On*TEBP) were the first genes encoding single-strand telomere end binding proteins to be cloned, sequenced, and characterized (*12*). The α subunit of *On*TEBP by itself is also able to bind telomeric ssDNA, but it binds differently than in the context of the α−*β*−ssDNA ternary complex and likely performs its own distinct functions (*13*). The 41 kDa *β* subunit by itself does not interact strongly with telomeric ssDNA, but modifies the way in which the α subunit interacts with $(T_4G_4)_n$ ssDNA (*14*). Although the α and *β* subunits do not interact appreciably in the absence of ssDNA, in the presence of telomeric ssDNA they form a very stable α−*β*−ssDNA complex with a stoichiometry of 1:1:1 (*15*).

The α subunit is composed of two domains. The 35 kDa N-terminal domain (the subject of this paper, subsequently

† This research was supported by grants to S.C.S. from the American Cancer Society (RPG-93-011-04-NP) and the NIH (1R01CA81109) and an NIH grant to T.R.C. (GM28039). D.L. was a fellow of the Jane Coffin Childs Memorial Fund for Medical Research.

* To whom correspondence should be addressed. Phone: (510) 643-9491. E-mail: classen@uclink.berkeley.edu.

‡ Present address: Department of Molecular and Cell Biology, 237 Hildebrand Hall #3206, University of California at Berkeley, Berkeley, CA 94720-3206.

§ Present address: Janus, 100 Fillmore St., Denver, CO 80206.

|| Present address: Math and Science Division, Diné College, Tsaile, Navajo Nation (AZ) 86556.

[1] Abbreviations: *On*TEBP, *O. nova* telomere end binding protein; ssDNA, single-strand deoxyribonucleic acid; DMS, dimethyl sulfate; EMSA, electrophoretic mobility shift assay; BIAcore, surface plasmon resonance; NCFB, nitrocellulose filter binding; α, *On*TEBP α subunit; *β*, *On*TEBP *β* subunit.

10.1021/bi0273718 CCC: $25.00
Published on Web 07/16/2003

termed α35) contains two OB folds (*16*) (reviewed in refs *17* and *18*) and binds ssDNA in a manner similar to that of full-length α. The 21 kDa C-terminal domain contains one OB fold, mediates interactions with the β subunit (*19*), and is involved in dimerization of the α subunit (*13*). The discovery of proteins homologous to the N-terminal domain of the *On*TEBP α subunit in *Euplotes crassus* (*20*), *Schizosaccharomyces pombe*, and *Homo sapiens* (*21*) suggests that proteins related to this domain of α are widely distributed in eukaryotes.

Dimethyl sulfate (DMS) protection assays on whole cells or isolated macronuclei from *O. nova* showed a protein-specific pattern of methylation protection at the 3′-end of DNA (*11*). These experiments showed that α has a preference for the 3′-most G_4 repeat when synthetic telomeric DNA was used, but similar experiments with native macronuclear DNA showed a slight preference for internal G_4 repeats (*14*). Native *On*TEBP has been shown to bind synthetic $(T_4G_4)_n$ sequences with $T_4 \rightarrow U_4$ substitutions, but not telomeric sequences with $G_4 \rightarrow C_4$ or $T_4 \rightarrow A_4$ substitutions (*22*).

An investigation of DNA binding by the related *E. crassus* α protein (no β has been identified in this organism) revealed that the full-length 51 kDa α protein binds tightly to synthetic telomeres containing the canonical *Euplotes* $T_4G_4T_4G_2$ single-strand 3′-overhang. Sequences with $T_4 \rightarrow U_4$ substitutions are tolerated; sequences ending in T_4G_4 instead of T_4G_2 bind less tightly, and sequences ending in T_4 or containing $G_4 \rightarrow C_4$ substitutions are not bound at all (*23*). A stable 35 kDa proteolytic fragment, homologous to the α35 protein investigated in this paper, behaves in a manner identical to that of the 51 kDa α protein, but it can better tolerate oligonucleotides ending in G_4 as opposed to the natural G_2 sequence, suggesting that the 35 kDa subunit contributes sequence specificity and salt stability and that 3′-end recognition is contributed by the C-terminal 15 kDa subdomain of the *E. crassus* α protein (which is not homologous to the *O. nova* protein) (*23*).

Previously reported crystal structures of *On*TEBP complexes, including an α−β−ssDNA ternary complex, an $(\alpha56-ssDNA)_2$ dimer complex, an α35−ssDNA complex, and uncomplexed α35, have provided detailed insights into the rich variety of interactions that might give rise to sequence- and 3′-end-specific ssDNA binding (*13*, *24*, *25*). However, an understanding of the thermodynamics of these interactions is required together with structural studies to uncover the sources of sequence specificity and 3′-end specificity of ssDNA recognition by *On*TEBP.

The α35−ssDNA complex may be ideally suited for this type of study. First, the α35−ssDNA system behaves as a simple two-component system, unlike the α−β−ssDNA and $(\alpha56-ssDNA)_2$ complexes in which intermediate and competing complexes complicate analysis. Second, α35 is less positively charged (+14 at pH 7.0) than α56 (+20 at pH 7.0), facilitating electrophoretic mobility shift assay (EMSA) experiments and, potentially, reducing the number of nonspecific interactions with longer DNAs. Third, crystal structures are now available for both the complexed and uncomplexed form of α35, revealing a rich variety of molecular interactions unique to single-strand nucleic acid−protein complexes (*25*). Described here are the first detailed comparative binding studies of the nature of sequence-specific and 3′-end selective recognition of telomeric ssDNA by *On*TEBP using surface plasmon resonance (BIAcore), electrophoretic mobility shift assays (EMSAs), and nitrocellulose filter binding (NCFB).

MATERIALS AND METHODS

Protein Purification. α35 was expressed in *Escherichia coli* BL21(DE3)pLysS cells harboring plasmid pKKT7-α35. Cells were grown in 2xYT medium supplemented with 5 mM glucose, 30 mM KP_i (pH 7.8), 0.5 mg/mL ampicillin, and 0.017 mg/mL chloramphenicol. The cells were grown at 37 °C to an optical density at 600 nm (OD_{600}) of ∼0.6, cooled to room temperature, and induced with 0.5 mM isopropyl thiogalactoside (IPTG) for an additional 8−9 h. Cells were lysed by sonication in 25 mM HEPES (pH 7.5), 150 mM NaCl, 2 mM DTT, 1 mM EDTA, and 0.02% NaN_3. The protein was precipitated by addition of solid ammonium sulfate to 65% of saturation and purified by ion exchange chromatography using S-Sepharose, Q-Sepharose, and a Superdex-75 gel filtration column (Amersham-Pharmacia). Typical yields were ∼80 mg of α35 per 6 L of cell culture (≈60 g of cell wet weight). α35 was stored in 5 mM Tris (pH 7.5), 100 mM NaCl, 2 mM DTT, 0.1 mM EDTA, and 0.02% NaN_3. α35 was serially diluted into buffer containing 50 mM MES (pH 6.5), 100 mM NaCl, 4 μM yeast tRNA, 0.5 mg/mL lysozyme, 2 mM DTT, 0.1 mM EDTA, and 0.02% NaN_3 for the EMSA and the NCFB assays. Lysozyme (NEB) was omitted for the NCFB assays because it saturated the capacity of the nitrocellulose membrane. For the BIAcore experiments, α35 was serially diluted into PBS buffer [10 mM phosphate (pH 7.5), 136 mM NaCl, 2 mM KCl, and 0.005% P20].

Electrophoretic Mobility Shift Assays (EMSAs). Protein dilutions were freshly prepared for each experiment. Typical binding reactions included 10 μL of protein and 10 μL of a binding mixture containing 50 mM MES (pH 6.5), 100 mM NaCl, 4% glycerol, 2 mM DTT, 0.02% NaN_3, 4 μM yeast tRNA, 0.5 mg/mL lysozyme, a trace of bromophenol blue, and 100−200 pM 5′-^{32}P-labeled DNA oligonucleotide. Reaction mixtures were allowed to equilibrate for several hours at 4 °C and then run for 2−3 h through 8 to 11% polyacrylamide gels in 0.5× BTPAE (9 g of Bis-Tris propane, 32 mL of 1 M sodium acetate, and 1.6 mL of 0.5 M EDTA, brought up to 800 mL and adjusted to pH 6.5 with acetic acid). The gels were dried and then visualized and quantitated on a phosphorimager. Apparent equilibrium binding constants (K_{app}s) were determined by plotting the log of the protein concentration versus the fraction of radiolabeled oligonucleotide shifted. Data were fit to a two-component model with baseline offset and scaling factor adjustments to correct for background counts and incomplete DNA binding, respectively.

A variety of experimental conditions were explored to improve the quality of the gel shift data. Initial gels had yielded smeared bands, low apparent affinities, and a total loss of labeled DNA at high protein concentrations. Varying the salt concentration in the binding mixture from 0 to 300 mM NaCl had little effect. A scan of polyanionic competitors (polyglutamic acid, tRNA, poly-T, and a short 20-base nontelomeric ssDNA oligonucleotide), however, revealed that tRNA greatly improved the way in which the α35−ssDNA complex migrated into the gel, yielding sharp distinct bands.

Labeled DNA continued to be lost, however, at high protein concentrations, so BSA and lysozyme were tested as protein competitors. Inclusion of lysozyme in the reaction mixture eliminated the loss of label and also greatly improved the K_{app} from ~640 to ~40 nM for $N_8T_4G_4$ ssDNA. All subsequent gel shift experiments included 4 μM tRNA and 0.4 mg/mL lysozyme in the binding mixture.

The length of the DNA was also explored. ssDNAs longer than ~16 nucleotides were necessary for the complex to migrate into the gel since the net charge of α35 is +13. For ssDNAs ending in the cognate TTTTGGGG sequence, longer DNAs bound with nearly the same affinity (40 nM for both N_8TTTTGGGG and N_{10}TTTTGGGG and 80 nM for N_{18}TTTTGGGG). For ssDNAs ending with an inverted GGGGTTTT sequence, however, the affinity was much tighter for longer DNAs (>5 μM for N_{10}GGGGTTTT DNA and 310 nM for N_{18}GGGGTTTT DNA). Therefore, all subsequent EMSA and NCFB experiments were carried out with 16-nucleotide (N_8 5′-tails) ssDNAs to minimize the effects of nonspecific binding.

Nitrocellulose Filter Binding Assays. Binding reactions were set up as described for the EMSA studies. Binding reactions included 15 μL of protein and 15 μL of a binding mixture containing 50 mM MES (pH 6.5), 100 mM NaCl, 4% glycerol, 2 mM DTT, 0.02% NaN_3, 4 μM yeast tRNA, a trace of bromophenol blue, and 100−200 pM 5′-^{32}P-labeled DNA oligonucleotide. However, since 0.5 mg/mL lysozyme in the reaction mixture would overwhelm the binding capacity of the nitrocellulose membrane (80−150 μg/cm^2), lysozyme was left out of all mixtures used in the nitrocellulose filter binding experiments.

The solutions were mixed in 96-well plates using multipipettors and allowed to equilibrate for several hours on ice. The mixtures were filtered using a Schleicher & Schuell 96-well minifold dot-blot filter apparatus through a membrane sandwich containing a top layer of Protean nitrocellulose membrane (BA85 0.45 μM pore size, Schleicher & Schuell), a middle layer of filter paper (Whatman No. 1), and a bottom layer of Hi-bond positively charged nylon membrane (Amersham) all placed on top of a thick support paper (GB002, Schleicher & Schuell). The membrane sandwich was equilibrated in a wash solution containing the same components as the binding buffer without tRNA. The absence of tRNA in the buffer used to wet the membranes had no effect on measured K_{app}s. The membrane sandwich was clamped into the filter apparatus, and a vacuum was applied to filter the samples. The membranes were first prewashed with a 100 μL aliquot of binding buffer; then 25 μL of the reaction mixture was loaded into the well and filtered, and finally, the filters were washed twice with 100 μL of binding buffer. The membranes were separated and dried at room temperature for 30−60 min before analysis on a phosphorimager was carried out. Apparent equilibrium binding constants were determined by plotting the log of the protein concentration versus the fraction of radiolabeled oligonucleotide retained by the nitrocellulose membrane. Data were fit to a two-component model with baseline offset and scaling factor adjustments.

Binding as a function of NaCl and tRNA concentration was explored. Salt had very little effect on binding to either the N_8TTTTGGGG or N_8GGGGTTTT sequence (2-fold increase in K_{app} between 100 and 400 mM NaCl). Addition of 4 μM competitor tRNA had no effect on binding to N_8TTTTGGGG DNA but greatly reduced the level of binding to the N_8GGGGTTTT sequence, indicating that this competitor helped to reduce the level of nonspecific DNA binding in the NCFB assay. Retention of the ssDNA and the α35−ssDNA complex was not affected by wash volumes varying from 0 to 1 mL, indicating that the complex was stably retained by the nitrocellulose membrane.

BIAcore Assays. ssDNAs with a 5′-biotin group attached via a 15-atom triethylene glycol (TEG) linker were attached to neutravidin-modified BIAcore CM5 chips as follows. A BIAcore CM5 chip was equilibrated in PBS buffer [10 mM phosphate (pH 7.5), 136 mM NaCl, 2 mM KCl, and 0.005% P20] at a rate of 5 μL/min. The chip was activated by injecting 35 μL of an NHS/EDC mixture (50 and 200 mM, respectively) at a rate of 5 μL/min and then modified with neutravidin by injecting 35 μL of a 5 μg/mL neutravidin solution in pH 5.0 acetate buffer. A typical immobilization resulted in 1200−1600 RUs (response units) of neutravidin binding to the chip. The chip, with neutravidin covalently coupled to the surface, was ready for addition of the 5′-biotinylated ssDNA. To immobilize the ssDNA, 2 μL of 200 nM ssDNA in PBS buffer was injected over the neutravidin chip at a rate of 10 μL/min. Typically, only 30−50 RUs of biotinylated DNA was immobilized on each chip (since low surface density was found to be important for eliminating mass transfer-influenced kinetics). To minimize nonspecific binding, the 5′-leaders of the ssDNA used in all BIAcore experiments had a four-base 5′-leader sequence, which was identical to the four 3′-most nucleotides of the eight-base 5′-leader used in the EMSA and NCFB experiments.

The affinities of α35 for various telomeric ssDNAs were measured using the kinetic method of measurement which involved real time measurement of the association (k_{on}) and dissociation (k_{off}) rate constants. The values of k_{on} and k_{off} were determined at five protein concentrations (2.5, 3.75, 5, 7.5, and 10 nM). All measurements were taken at room temperature in PBS buffer. Each k_{on} was determined by injecting 60 μL of α35 at a rate of 50 μL/min, providing a 72 s window for measurement. A postinjection period of 360 s at a rate of 50 μL/min followed for measurement of the k_{off}. Data from the five different concentrations were superimposed in the Biaevaluation software package, and the k_{on} and k_{off} values were fit globally using a simple 1:1 Langmuir model with a correction for drifting baseline. Within the equation describing the 1:1 Langmuir model, the R_{max} value should be proportional to the number of ssDNA molecules immobilized on the surface of the BIAcore chip. However, it was observed that the globally refined values for R_{max} varied much more than the amount of ssDNA known to be immobilized. Because the value for k_{on} is highly correlated with R_{max}, the data showed drastic differences in k_{on} and thus K_{app}. To address this problem, R_{max} was fixed for each ssDNA at a value proportional to the known amount of immobilized ssDNA. The simple 1:1 Langmuir model was modified to contain a fixed R_{max} value instead of the globally refined R_{max} in the original model. The fixed R_{max} value used for each curve fit was determined by the equation describing R_{max} in the Biatechnology handbook [R_{max} = (analyte MW/ligand MW)(ligand response)(valence)].

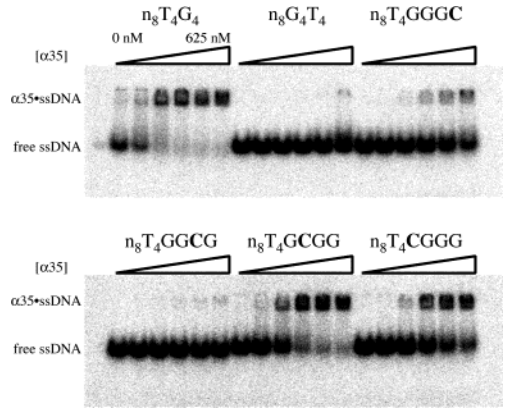

FIGURE 1: Representative electrophoretic mobility shift data. Gel shift experiments in which each of the G nucleotides in the G_4 tract was individually substituted with a C and the effect on binding to α35 was evaluated. n_8 denotes the nontelomeric nucleotide sequence given in Table 1. α35 concentrations were varied from 0 to 625 nM.

Table 1: K_{app} Values from Electrophoretic Mobility Shift Assays

DNA sequence	rel K_{app} [a]	ranking [b]
ACATCGCTCAGCCAGACA**TTTTGGGG**	**2**	
ACATCGCTCAGCCAGACA**GGGGTTTT**	**8**	•
CGCTCAGCCAGACA**TTTTGGGG**	**2**	
CGCTCAGCCAGACA**GGGGTTTT**	**8**	•
CTCAGCCAGACA**TTTTGGGG**	**3**	
CTCAGCCAGACA**GGGGTTTT**	**16**	••
CAGCCAGACA**TTTTGGGG**	**1**	
CAGCCAGACA**GGGGTTTT**	**63**	••
GCCAGACATTTTGGGG	**1**	
GCCAGACAGGGGTTTT	weak	•••
GCCAGACATTTTGGGG**T**	**4**	•
GCCAGACATTTTGGGG**TT**	**6**	•
GCCAGACATTTTGGGG**TTT**	**6**	•
AGCCAGACA**TTT**GGGG	**2**	
CAGCCAGACA**TT**GGGG	**4**	•
ACAGCCAGACA**T**GGGG	**8**	•
AACAGCCAGACAGGGG	**8**	•
GCCAGACATTTTGGG**C**	**16**	••
GCCAGACATTTTGG**C**G	weak	•••
GCCAGACATTTTG**C**GG	**3**	
GCCAGACATTTT**C**GGG	**6**	•

[a] rel $K_{app} = K_{app}$(mut)/K_{app}(wt) [K_{app}(wt) = 40 nM]. [b] rel K_{app} of 1−3 (blank), 4−9 (•), ≥10 (••), or too weak to measure (•••).

RESULTS

3′-End Selective ssDNA Binding by α35. Initial electrophoretic mobility shift assays (EMSAs) were performed using 18-nucleotide ssDNAs ending in TTTTGGGG and GGGGTTTT to determine the degree of preference for terminal GGGG tracts versus internal sites (Figure 1). EMSA experiments revealed a decrease in K_{app} of at least 125-fold for the internal relative to the terminal site (Table 1). Note that in addition to eliminating the G_4 sequence at the 3′-end, the GGGGTTTT sequence lacks an internal TTTT tract 5′ to the GGGG tract, which we will see is also important for specific binding.

To more fully explore 3′-end specificity, a series of ssDNAs with one, two, three, or four T nucleotides added to the 3′-end of the TTTTGGGG sequence were evaluated in EMSA (Table 1), NCFB (Table 2), and BIAcore experiments (Table 3). All three methods showed that adding T nucleotides to the 3′-end of the ssDNA indeed weakened interactions with α35. EMSA and NCFB experiments showed the TTTTGGGGT and TTTTGGGGTT ssDNAs bound approximately 4−6-fold more weakly than TTTTGGG ssDNA. BIAcore experiments showed slightly greater decreases with the TTTTGGGGT and TTTTGGGGTTTT

Table 2: K_{app} Values from Nitrocellulose Filter Binding Assays

DNA sequence	rel K_{app} [a]	ranking [b]
GCCAGACATTTTGGGG	**1**	
GCCAGACAGGGGTTTT	**24**	••
GCCAGACATTTTGGGG**T**	**4**	•
GCCAGACATTTTGGGG**TT**	**5**	•
GCCAGACATTTTGGGG**TTT**	**4**	•
GCCAGACATTTTGGGG**TTTT**	**2**	
AGCCAGACA**TTT**GGGG	**1**	
CAGCCAGACA**TT**GGGG	**4**	•
ACAGCCAGACA**T**GGGG	**10**	••
AACAGCCAGACAGGGG	**12**	••
GCCAGACATTTTGGG**C**	**17**	••
GCCAGACATTTTGG**C**G	weak	•••
GCCAGACATTTTG**C**GG	**6**	•
GCCAGACATTTT**C**GGG	**5**	•
GCCAGACATTTTGGG**A**	**10**	••
GCCAGACATTTTGGG**T**	**8**	•
GCCAGACATTTTGG**A**G	weak	•••
GCCAGACATTTTGG**T**G	weak	•••
GCCAGACATTTTG**A**GG	**2**	
GCCAGACATTTTG**T**GG	**7**	•
GCCAGACATTTT**A**GGG	**3**	
GCCAGACATTTT**T**GGG	**5**	•

[a] rel $K_{app} = K_{app}$(mut)/K_{app}(wt) [K_{app}(wt) = 1.8 μM]. [b] rel K_{app} of 1−3 (blank), 4−9 (•), ≥10 (••), or too weak to measure (•••).

Table 3: K_{app} Values from BIAcore Assays

DNA sequence	rel k_{on} [a]	rel k_{off} [b]	rel K_{app} [c]	ranking [d]
TACATTTTGGGG	1.0	1.0	**1.0**	
TACATTTTGGGG**T**	0.8	3.4	**4.3**	•
TACATTTTGGGG**TTTT**	0.7	5.0	**7.2**	•
TACA**C**TTTGGGG	0.4	0.8	**2.1**	
TACAT**C**TTGGGG	0.6	1.0	**1.5**	
TACATT**C**TGGGG	2.2	1.7	**0.8**	
TACATTT**C**GGGG	1.5	5.3	**3.4**	•
TACATTTT**C**GGG	1.1	3.5	**3.4**	•
TACATTTTG**C**GG	1.2	3.5	**3.0**	
TACATTTTGG**C**G	0.5	10.8	**24**	••
TACATTTTGGG**C**	0.8	7.7	**9.5**	•
TACA_TTTGGGG[e]	1.3	1.2	**1.0**	
TACAT_TTGGGG	0.9	1.5	**1.6**	
TACATT_TGGGG	1.2	0.9	**0.8**	
TACATTT_GGGG	0.8	2.7	**3.2**	
TACATTTT_GGG	1.0	3.4	**3.6**	•
TACATTTTG_GG	0.7	4.9	**6.8**	•
TACATTTTGG_G	0.6	6.5	**10**	••

[a] rel $k_{off} = k_{off}$(mut)/k_{off}(wt) [k_{off}(wt) = 5.23×10^{-2} s^{-1}]. [b] rel $k_{on} = k_{on}$(mut)/k_{on}(wt) [k_{on}(wt) = 4.95×10^5 M^{-1} s^{-1}]. [c] rel $K_{app} = K_{app}$(mut)/K_{app}(wt) [K_{app}(wt) = 10 nM]. [d] rel K_{app} of 1−3.3 (blank), 3.4−9.9 (•), and ≥10 (••). [e] _ indicates an abasic substitution at this position.

ssDNAs binding 4- and 7-fold more weakly, respectively, than TTTTGGGG ssDNA (Table 3). Therefore, all three methods demonstrate that α35 exhibits a small but consistent preference for a 3′-G residue.

Sequence-Specific Recognition of the G_4 Tract of TTTTGGGG ssDNA. DNAs containing a variety of single-nucleotide or abasic substitutions at each G residue in the $T_1T_2T_3T_4G_5G_6G_7G_8$ sequence were evaluated for interaction with α35 by EMSA, NCFB, and BIAcore methods (Tables 1−3). The results of this analysis revealed a recognition pattern of $t_1t_2t_3T_4G_5G_6\mathbf{G}_7G_8$, where the lowercase letters indicate that substitutions at this position resulted in <2-fold reductions in relative affinity, the uppercase letters indicate 2−20-fold reductions in relative affinity, and the uppercase bold letter indicates a >20-fold reduction in relative affinity.

All three methods revealed that position 8 (G_8) is critical for high-affinity binding. NCFB analysis showed a 10−20-fold decrease in affinity when G_8 was changed to a C, T, or A; EMSA experiments showed a 16-fold decrease in the level of binding for the $G_8 \rightarrow$ C substitution, and BIAcore experiments showed a 9.5-fold decrease in affinity for $G_8 \rightarrow$ C substitution primarily as a result of an increase in k_{off} (7.7-fold faster). The increase in k_{off} may explain the greater reductions in affinities as measured by EMSA and NCFB, since this could exaggerate the differences observed by these methods. The values for k_{off} as measured by BIAcore analysis are in a range that is significant compared to the time required for loading gels and for filtering the samples through nitrocellulose.

G_7 appears to be the most critical nucleotide for high-affinity interactions between α35 and TTTTGGGG ssDNA. In fact it was not possible to measure a K_D for $G_7 \rightarrow$ C, $G_7 \rightarrow$ A, or $G_7 \rightarrow$ T substituted ssDNAs by NCFB assays. In the EMSA studies, no shift in the $G_7 \rightarrow$ C DNA band was observed even at α35 concentrations of 5 μM. Because of the limited protein binding capacity of the nitrocellulose membrane used in these experiments (80−150 μg/cm^2), the α35 concentration was limited to 10−20 μM. At 10 μM α35, modest binding was observed (~30%) with the $G_7 \rightarrow$ C substituted ssDNA, indicating a K_{app} of greater than 10 μM. BIAcore experiments with the $G_7 \rightarrow$ C substituted ssDNA gave a relative K_{app} that was 24-fold weaker than that of TTTTGGGG ssDNA (Figure 2). As with the $G_8 \rightarrow$ C substituted DNA, this resulted primarily from an increase in k_{off} (11 times faster). The change in k_{on} was only slightly affected (2 times slower). Similar results were obtained for the $G_7 \rightarrow$ abasic substitution. BIAcore experiments revealed a 10.2-fold reduction in K_{app} for the $G_7 \rightarrow$ abasic substitution, resulting from a k_{on} rate constant that was 2 times slower and k_{off} rate constant that was 6.5 times faster than those for TTTTGGGG ssDNA. Once again, the results of the EMSA and NCFB assays were likely exaggerated due to the increased k_{off} for the $G_7 \rightarrow$ C and $G_7 \rightarrow$ abasic substituted ssDNAs. Nevertheless, all three methods are consistent in that they indicate that G_7 is the most critical position for sequence-specific recognition of TTTTGGGG ssDNA by α35.

Nucleotides G_6 and G_5 are also important for binding, but less so than G_8 and G_7. Binding is 6−7-fold weaker for $G_6 \rightarrow$ C or $G_6 \rightarrow$ T substituted ssDNAs and 2-fold weaker for $G_6 \rightarrow$ A substituted ssDNAs in NCFB experiments.

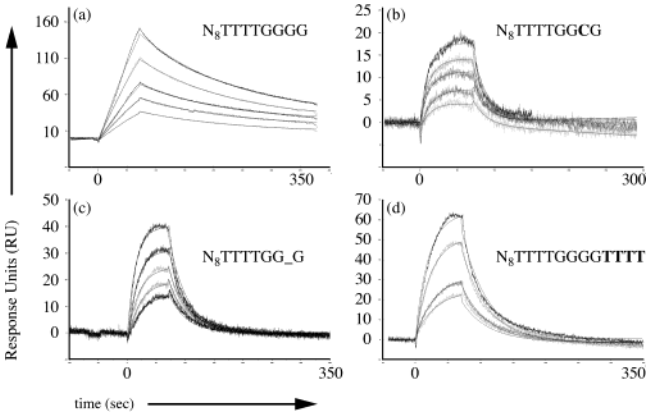

FIGURE 2: Representative BIAcore sensograms. Each sensogram shows overlays for five different injections of α35 at concentrations from 2.5 to 10.0 nM determined during a 72 s window for measuring k_{on} followed by a 360 s window for measuring k_{off}. (a) Notice the slow k_{on} and k_{off} rates for the N_8TTTTGGGG ssDNA as compared with those of the G → C (b) and G → abasic (c) substituted DNAs. Data in panel d are from a ssDNA with four extra T nucleotides added to the 3′-end.

EMSA studies gave a 3-fold decrease in the extent of binding for the $G_6 \rightarrow$ C substituted ssDNA. For $G_5 \rightarrow$ C or $G_5 \rightarrow$ T substituted ssDNAs, NCFB showed a 5-fold decrease in affinity and a 3-fold decrease for the $G_5 \rightarrow$ A substitution. EMSA experiments gave similar results with a 6-fold decrease in the extent of binding for a $G_5 \rightarrow$ C substituted ssDNA. BIAcore experiments again gave results consistent with those of the EMSA and NCFB assays (3−7-fold reductions in affinities for substitutions at G_5 and G_6) and, furthermore, revealed these reductions resulted largely from differences in k_{off}. For $G_6 \rightarrow$ C or abasic and $G_5 \rightarrow$ C or abasic substitutions, k_{off} was 3−5 times faster and, therefore, accounts for the majority of the increase in K_{app}.

Sequence-Specific Recognition of the T_4 Tract of TTTTGGGG ssDNA. A series of ssDNAs in which the T nucleotides on the 5′-side of the G_4 tract were progressively removed were evaluated for interaction with α35 by EMSA and NCFB experiments. The four ssDNAs used in this study ended in ATTTGGGG, CATTGGGG, ACATGGGG, and GACAGGGG. For the ATTTGGGG ssDNA, the EMSA showed 2-fold weaker binding while the NCFB assay showed no distinguishable difference. The CATTGGGG ssDNA had 4-fold weaker binding as measured by EMSA and NCFB experiments. The ACATGGGG and GACAGGGG ssDNAs bound 8-fold weaker than the TTTTGGGG ssDNA as measured by the EMSA and 10- and 12-fold weaker, respectively, as measured by the NCFB assay. These results indicate that the TTTT tract is indeed important for binding and that the effect, although small for each individual site, is cumulative throughout the T_4 tract.

To explore the contributions of individual nucleotides at these positions, DNAs containing T → C or T → abasic substitutions at each T in the TTTTGGGG sequence were evaluated for interaction with α35 by BIAcore analysis (Table 3). For the $T_4 \rightarrow$ C and $T_4 \rightarrow$ abasic substitutions the K_{app} values were 3.4- and 3.3-fold weaker, respectively, and this resulted largely from increases in k_{off} of 5.3- and 2.7-fold, respectively. Interestingly, T_4 was the only T position for which substitution resulted in a k_{off} significantly faster and a K_{app} significantly weaker than those of TTTTGGGG

ssDNA. For the remaining T positions (T_3, T_2, and T_1), the T → C and T → abasic substitutions had only small effects (<2-fold) on k_{on}, k_{off}, and K_{app} (Table 3).

Comparisons of Results from EMSA, NCFB, and BIAcore Experiments. All three of these methods are considered heterophasic and, therefore, are not thermodynamic. Some of the limitations of these methods are discussed below. However, the relative values obtained for K_{app} by all three independent methods agree very well. Our NCFB experiments, however, gave higher K_{app} values than the EMSA and BIAcore experiments. The higher K_{app} values from NCFB experiments may be due, in part, to the effects of lysozyme. Including lysozyme in the reaction mixtures greatly improved the K_{app} values of α35 in EMSA experiments and produced sharper, cleaner bands that did not disappear with increased α35 concentrations. Lysozyme perhaps saturates and blocks the various experimental surfaces that contact the sample, thus allowing for more reproducible EMSA experiments and lower K_{app} values. As mentioned in Materials and Methods, lysozyme could not be included in the NCFB studies since it would overwhelm the binding capacity of the nitrocellulose membrane. Nevertheless, the relative values for K_{app} are consistent among all three methods, which provides confidence in the measured values.

A 1:1 binding model was assumed for all three methodologies. At least three lines of evidence indicate a 1:1 binding model. First, the crystal structure of α35 bound to a ssDNA reveals a 1:1 complex. Second, EMSA experiments show a single shifted band indicating a discreet 1:1 complex. Third, BIAcore data preferentially fit a 1:1 binding model. Analysis of the active fraction of the protein was not performed, so we have focused the discussion entirely on the relative binding affinities.

An important consideration when analyzing BIAcore data is the recognition and proper treatment of mass transport-influenced kinetics since this can complicate data analysis. Very low levels of immobilized ligand have been shown to limit the effects of mass transport-influenced kinetics to acceptable levels (*26*). Therefore, the density of ssDNA on the BIAcore chips in our experiments was kept to an extremely low level (30−50 RUs of immobilized ssDNA per chip). The resulting sensograms showed no signs of mass transport-influenced kinetics as determined by fitting the data to multiple binding models and comparing the results. Invariably, the data did not fit models containing mass transport corrections any better than the simple 1:1 model.

Although we cannot completely rule out the effects of nonspecific binding, we have tried to minimize the effects by using ssDNA oligonucleotides with short tails (N_8 in the case of all EMSA and NCFB assays and N_4 in the case of BIAcore experiments). Competitor tRNA was added to both the EMSA and NCFB experiments to further reduce the extent of nonspecific binding to the ssDNA oligonucleotides. Although various caveats of each method raise a variety of concerns (especially for absolute K_D values), the relative K_{app} values for all the DNA sequences that were tested from all three methods give comparable results. Since these methods have different potential sources of error, this provides confidence in the validity of the measurements in accurately revealing the sites most critical for specific binding of α35 to telomeric ssDNA.

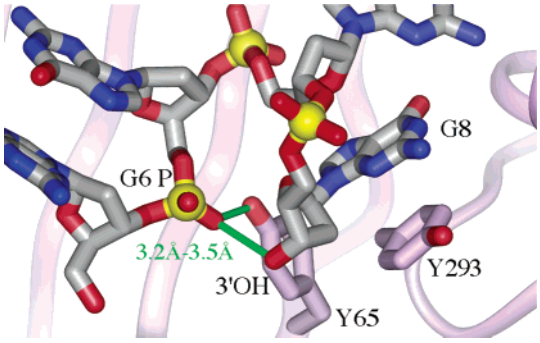

FIGURE 3: 3′-Hydroxyl group in the α35−ssDNA complex. The ssDNA is shown as stick representations (colored according to atom types, carbon in gray, oxygen in red, nitrogen in blue, and phosphorus in yellow). The 3′-hydroxyl groups in the α35−ssDNA complexes are all between 3.2 and 3.5 Å from the phosphodiester group of G_6 in the five different observations of the α35−ssDNA complex (*25*).

DISCUSSION

3′-End Selective Binding of ssDNA by α35. An important characteristic of the *O. nova* telomere end binding protein is its ability to specifically recognize and bind the 3′-ends of single-strand telomeric DNA. Interestingly, the 3′-end of the DNA is positioned differently in each of the reported *On*TEBP complexes. In the α56−*β*28−$G_4T_4G_4$ ternary complex, the 3′-end of the ssDNA is bound deep within the complex in the α portion of the α*β* site (*24*). Only a 3′-end can fit into this site. In the $(\alpha56-T_4G_4)_2$ complex and in the α35−ssDNA complex, however, the 3′-end of the ssDNA has shifted by precisely one T_4G_4 repeat relative to its position in the α56−*β*28−$G_4T_4G_4$ complex and is now positioned in the α site (*13*, *25*). This shift is necessary in the $(\alpha56-T_4G_4)_2$ dimer complex because dimerization physically blocks the α portion of the α*β* site. This shift in the 3′-end of the DNA is also observed in the α35−T_2G_4 crystal structure, however, suggesting that α alone in the absence of *β* and without dimerization inherently prefers the 3′-end of the DNA in the α site. These structural data agree well with the results of methylation protection studies in solution (*14*) and the binding studies described here.

The structure of the α35−T_2G_4 complex suggests a mechanism for 3′-end discrimination within the α site. In the α35−T_2G_4 complex, the 3′-hydroxyl is only 3.2 Å from the phosphodiester group between G_5 and G_6 (Figure 3). If a phosphodiester group were present at this site rather than a 3′-hydroxyl, electrostatic repulsion would replace the potentially favorable hydrogen bond existing between the 3′-hydroxyl and the phosphate oxygen and would result in destabilization of the complex. All three binding methods show an approximate 4-fold decrease in affinity when a single T residue is added to the 3′-end of a TTTTGGGG sequence. Further additions of T nucleotides to the 3′-end of TTTTGGGG ssDNA result in 6−7-fold decreases in affinity as observed by EMSA and BIAcore experiments. The magnitudes of these reductions in affinity are consistent with an electrostatic model for 3′-end discrimination. The even larger decreases in affinity that accompany the addition of more T nucleotides to the 3′-end are perhaps the result of both steric effects and the accumulation of additional charge. Recent studies of the 3′-end binding properties of *S. pombe* Pot1p showed a similar 4-fold decrease in the extent of

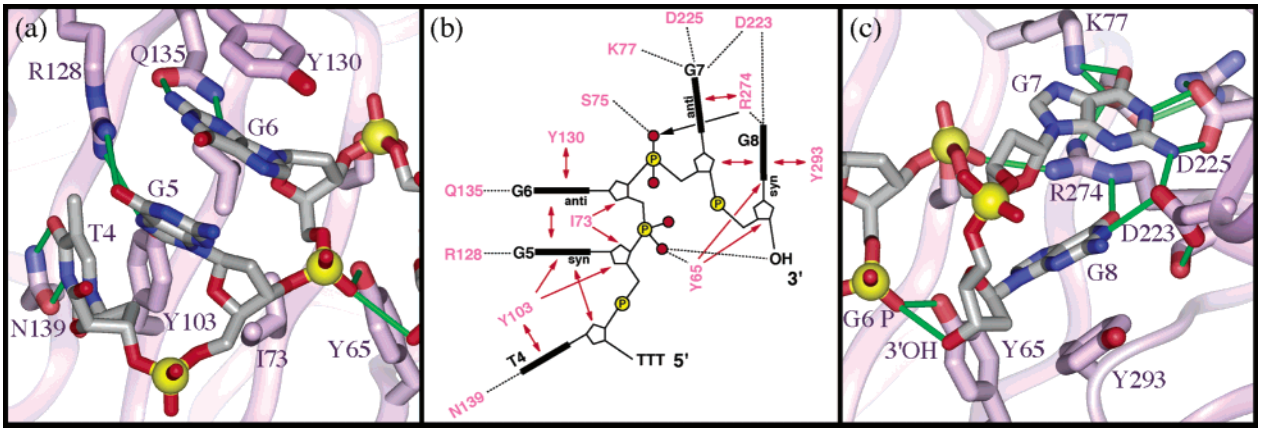

FIGURE 4: Details of protein−ssDNA interactions in the α35−ssDNA complex. α35 is shown as a light purple ribbon with relevant side chains shown as stick representations with carbons colored light purple, oxygens red, and nitrogens blue. The DNA is shown as a stick representation and colored as described in the legend of Figure 3. Hydrogen bonds are shown as solid green lines. (a) Protein−ssDNA interactions among T_4, G_5, G_6, and α35. Bases are stacked on aromatic side chains or on other bases, and in the case of G_5, its base is sandwiched between the base of G_6 and the deoxyribose group of T_4. (b) Schematic of α35−ssDNA interactions. Bases of the ssDNA are shown as long rectangles; DNA is colored as described in the legend of Figure 3. Hydrogen bonds are shown as dotted lines; ionic interactions are shown as solid black arrows, and close contacts between hydrophobic groups are shown as red arrows. If aromatic groups are stacked together, this is indicated with a red, double-headed arrow. (c) Protein−ssDNA interactions between G_7 and G_8 and α35. Notice the central role of R274 in these interactions. R274 hydrogen bonds with O4 of G_8 and forms an ionic interaction with the phosphodiester group of G_7 while simultaneously stacking against the base of G_7 (*25*).

binding when a telomeric repeat was moved from the 3′-end to an internal site (*27*). The telomere end binding proteins from *S. pombe*, *H. sapiens*, and *O. nova* all contain or are predicted to contain OB folds (*21*, *24*), which suggests that an electrostatic model for 3′-end discrimination facilitated by OB folds might be an evolutionarily conserved feature of telomere 3′-end recognition.

Sequence-Specific Recognition of TTTTGGGG ssDNA. The nucleotides most important for telomeric ssDNA recognition by α35 are within the G_4 tract of the $t_1t_2t_3T_4G_5G_6\mathbf{G_7}G_8$ telomeric repeat. The nucleotide that contributes the most to sequence discrimination is G_7. In the crystal structure of the α35−T_2G_4 complex, the base of G_7 adopts an *anti* conformation and stacks flat against the guanidinium group of R274, potentially forming a cation−π stacking interaction (Figure 4c). Around the edge of the base of G_7, the exocyclic amino group (N2) hydrogen bonds to both D223 and D225, the N1 group hydrogen bonds to D225, and O6 hydrogen bonds to K77. Note that these interactions primarily involve the Watson−Crick (W/C) hydrogen bond donor and acceptor groups of G which are perhaps the most effective sites for nucleotide base discrimination. The ribose ring of G_7 does not interact with the protein, but instead, its bridging oxygen packs against the base of G_8 in a manner that is frequently observed in crystal structures in the context of psuedo-continuously stacked double-strand nucleic acids (*28*, *29*). The phosphodiester group between G_6 and G_7 hydrogen bonds to S75 and forms a salt bridge with R274. These peripheral packing interactions may be important in positioning the base in the binding pocket, thus allowing optimal stacking and hydrogen bonding interactions.

There is a growing body of literature indicating that cation−π stacking interactions such as this impart both stability and specificity to protein−DNA interactions (*30*, *31*). The arginine−guanine pair is not only the most frequently observed pair in the structure database but also the most stable as monitored by *ab initio* energy calculations. The average energy of stabilization is reported to be −8.6 kcal/mol for arginine−guanine pairs, −1.4 kcal/mol for arginine−adenine pairs, −0.2 kcal/mol for arginine−thymine pairs, and 2.4 kcal/mol for arginine−cytosine pairs (*32*). Therefore, a C_7−R274 cation−π stack is likely not as thermodynamically favorable as a G_7−R274 cation−π stack, consistent with the decreased affinity upon $G_7 \rightarrow C$ substitution. It is also noteworthy that the $G_7 \rightarrow C$ mutation results in a K_{app} that is greater than that for the $G_7 \rightarrow$ abasic mutation, which is consistent with the *ab initio* calculations that showed arginine−cytosine pairs to be slightly destabilizing (*32*).

The second most important position for sequence discrimination is G_8, the 3′-most nucleotide of the TTTTGGGG repeat. In the structure of the α35−T_2G_4 complex, the base of G_8 is stacked between the aromatic face of Y293 and the bridging oxygen of the deoxyribose ring of G_7 (Figure 4b,c). Around the edge of this base, O6 hydrogen bonds to the ϵ amino group of R274, N1 hydrogen bonds to D223, and the hydrogen of C8 points into the aromatic face of Y65 in an edge-to-face interaction. The deoxyribose ring of G_8 packs against Y65 which appears to help position this nucleotide in its binding pocket. Together, these interactions appear to constrain G_8 in a *syn* conformation. All base substitutions tested at this position caused approximately the same decrease in affinity, suggesting that the loss of specific interactions with G is most critical for specific versus nonspecific binding at this position.

The EMSA and nitrocellulose filter binding studies indicate that G_5 and G_6 are also important for binding, although to a lesser extent than G_7 and G_8. The base of G_6 is sandwiched between the base of G_5 and the aromatic face of Y130 (Figure 4a,b). Around the edge of the base of G_6, the N2 hydrogens and N3 lone pair electrons form hydrogen bonds with Q135 in a bidentate manner. The base of G_5 stacks against G_6 and makes bidentate hydrogen bonds to R128. The interactions with G_5 and G_6 are less extensive and less interlinked than those with G_7 and G_8.

The bases of G_7 and G_8 use their W/C hydrogen bond donor−acceptor groups (O6, N1, and N2) to hydrogen bond with amino acid side chains in the ssDNA binding pocket. In contrast, G_5 and G_6 primarily interact via the peripheral non-W/C hydrogen bond donor−acceptor groups. Since W/C hydrogen bonding groups present a more unique pattern of hydrogen bonding interactions, it follows that W/C hydrogen bond donor−acceptor interactions potentially offer a more effective means of sequence discrimination than interactions with non-W/C groups in sequence-specific protein−ssDNA complexes.

For example, the non-W/C N3 lone pair acceptor group of a G residue could potentially be mimicked by the N3 lone pair electrons from an A residue or the O2 lone pair electrons from C or T, or hydrogen bonding with the lone pair electrons from N7 of G could be mimicked by the N7 from an A. This hypothesis predicts that in the α35−ssDNA complex, specificity for G at positions 7 and 8 would be greater than at positions 5 and 6, since G_5 and G_6 do not fully satisfy the hydrogen bonding potential of their W/C hydrogen bond donor−acceptor groups. The binding data from the EMSA, NCFB, and BIAcore experiments are in agreement with this idea.

On the basis of the binding data reported in this paper, the T_1−T_3 nucleotides of the TTTTGGGG telomeric repeat contribute only modestly to specificity. The T_4 nucleotide, however, appears to be as important as G_5 and G_6. It is interesting to note that for T_4 the base-specific W/C hydrogen bond donor−acceptor groups form hydrogen bonds to N139 in a bidentate manner much the same as observed in an A·T base pair. This would suggest a high degree of specificity for T at this position. This prediction is again born out in our BIAcore binding analysis where T → C and abasic substitutions resulted in 3.4- and 3.2-fold increases in K_{app}, respectively. The remainder of the T nucleotides, although required for crystallization, were not observed in the experimental electron density, and substitutions at these positions do not significantly affect binding by α35.

The results from the comparative binding studies reported here indicate that both the presence of a 3′-end terminating in a ...$GGGG_{OH}$ tract and the presence of certain nucleotides within the telomeric $t_1t_2t_3T_4G_5G_6\mathbf{G}_7G_8$ repeat is important for ssDNA binding and telomere formation. It appears that the α subunit plays a key role in recognizing single-strand DNA by utilizing both 3′-end and sequence-specific discrimination to preferentially incorporate telomeric ssDNA into mature α−*β*−ssDNA complexes.

Implications for Sequence-Specific Single-Strand DNA Recognition. Specific recognition of dsDNA is thought to involve matching of complementary arrays of hydrogen bonds in the major groove of the dsDNA and is primarily entropically driven (*33*, *34*). Sequence-specific binding of ssDNA is a fundamentally different situation the sources of which likely result from contributions of substrate rigidity, hydrogen bonding, stacking, and electrostatic interactions different from those occurring in dsDNA recognition (*35*−*37*). The simplest description of specific base recognition depends solely on the number of W/C-like hydrogen bonds formed between the edges of the bases and the interacting amino acid side chains. The number of W/C-like hydrogen bonds formed by each base in the TGGGG sequence as observed in the α35−ssDNA crystal structure is 2, 1, 1, 4, and 2, respectively. Thus, this simple model is qualitatively a fairly good indicator of the relative importance of the bases for the stability of the α35−ssDNA complex.

The number of W/C-like hydrogen bonds provides a starting point for explaining sequence specificity; however, as with DNA duplex stability, it is probably not the driving force for complex formation. A more accurate description of the stability of the protein−ssDNA complex should also take into account stacking effects (*38*). To gain insight into the contributions of stacking effects, calculations of the buried surface area of the DNA bases in the α35−ssDNA complex were performed. The accessible surface area of the ssDNA which is buried upon binding α35 is difficult to determine because ssDNA free in solution likely samples many different conformations. However, there is evidence that base stacking effects in ssDNA cause an inherent sequence-dependent rigidity, thus preferentially favoring certain ssDNA conformations (*36*). With that said, if calculations are performed while fixing the ssDNA in the conformation as observed in the crystal structure, we can measure the difference in accessible surface area between the DNA alone and the DNA in the context of α35. Similar calculations can be carried out with each individual base. These calculations give an indication of how "tight" each base is buried within the α35−ssDNA complex. In the calculation with the entire DNA removed all at once, the calculated buried surface area is 68 Å^2 for G_5, 212 Å^2 for G_6, 259 Å^2 for G_7, and 164 Å^2 for G_8. When calculations were carried out by removing each base individually, the calculated buried surface areas were 194, 432, 444, and 314 Å^2, respectively. The values are larger for the second method of calculation because it includes accessible surface area buried between DNA bases. Together with the hydrogen bonding data, the degree of stacking correlates reasonably well with the relative thermodynamic contributions as measured by the K_{app} values reported in this paper.

In the crystal structure of the α35−ssDNA complex, the presence of an uncomplexed α35 molecule as well as two α35−ssDNA complexes enabled a comparison of the uncomplexed and complexed forms (*25*). Interestingly, the DNA binding site was observed to be largely preformed in the absence of ssDNA with the exception of the region that interacts with the 3′-end in the α site. Three residues (R274, D223, and S275) that interact with G_7 and G_8 are in different rotomer conformations in the uncomplexed and complexed structures. The rearrangements observed in these three residues show that, although much of the ssDNA binding site is preformed in α35 and ready to dock with the ssDNA, a slight but apparently critical refolding of part of the ssDNA binding site accompanies complex formation. There are no observable differences in the side chain rotomers of the three residues (K77, R272, and D225) which form hydrogen bonds to the W/C hydrogen bond donor and acceptor groups of the base of G_7. An intricate network of hydrogen bonds holds these side chains in a rigid conformation in both the complexed and uncomplexed forms. This is in contrast to the hydrogen bond acceptor−donor groups of the amino acid side chains that interact with G_8, which adopt different conformations in the absence of DNA. Perhaps formation of the α35−ssDNA complex involves an ordered docking and folding mechanism. The preformed parts of the interface could form an initial docking site for interactions with the

flexible ssDNA chain, and as these regions dock, the rest of the binding site might refold to accommodate a 3′-G while discriminating between the other nucleotides. This model is consistent with the observation that mutations of G_7 have the largest effect on the measured binding constant. We propose that the preformed binding site presented to G_7 enables a larger free energy of binding to be expended on that particular base than on the other bases, whose binding sites may not be as well formed (i.e., G_8) or whose binding sites are partly composed of the surfaces of adjacent bases (i.e., T_4, G_5, and G_6) and require ordered binding of the individual bases for their binding sites to be fully formed. Additionally, a refolding event like this might explain the relatively slow k_{on} rate constants measured by BIAcore experiments.

The strong preference for runs of four G nucleotides in combination with the thermodynamically unfavorable effects of additional T nucleotides on the 3′-end results in a pattern of relative information content (tttTGG**G**G) that suggests how the telomere end binding protein of *O. nova* could help control telomere extension. The protein precisely measures out four G nucleotides separated by four T nucleotides and preferentially binds at the 3′-ends of these sequences, which could contribute to the relatively homogeneous chromosome 3′-end sequence observed in *O. nova* (*39*).

Binding of telomeric ssDNA by the *On*TEBP heterodimer is thought to proceed via an ordered, cooperative process involving cofolding of the protein and ssDNA, which would require refolding and rearrangement of the ssDNA and various loops. Potentially, there are several steps along the assembly and disassembly pathway ($\alpha + \beta +$ ssDNA $\rightarrow$ $\alpha-\beta-$ssDNA) which integrate sequence-specific ssDNA recognition. The α35−ssDNA complex, whose DNA binding properties have been investigated in this paper, represents a first binding and recognition step on this assembly and disassembly pathway.

ACKNOWLEDGMENT

We thank D. L. Theobald, O. B. Peersen, and M. P. Horvath for their support and advice during the course of these experiments.

REFERENCES

1. Blackburn, E. H. (1991) *Nature 350*, 569−573.
2. Zakian, V. A. (1995) *Science 270*, 1601−1607.
3. McClintock, B. (1941) *Genetics 26*, 234−282.
4. Watson, J. D. (1972) *Nat. New Biol. 239*, 197−201.
5. Olovnikov, A. M. (1973) *J. Theor. Biol. 41*, 181−190.
6. Greider, C. W., and Blackburn, E. H. (1985) *Cell 43*, 405−413.
7. Rabl, C. (1885) *Morphologisches Jarhbuch 10*, 214−330.
8. Comings, D. E. (1980) *Hum. Genet. 53*, 131−143.
9. Dandjinou, A. T., Dionne, I., Gravel, S., LeBel, C., Parenteau, J., and Wellinger, R. J. (1999) *Histol. Histopathol. 14*, 517−524.
10. Gottschling, D. E., and Zakian, V. A. (1986) *Cell 47*, 195−205.
11. Price, C. M., and Cech, T. R. (1987) *Genes Dev. 1*, 783−793.
12. Hicke, B. J., Celander, D. W., MacDonald, G. H., Price, C. M., and Cech, T. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1481−1485.
13. Peersen, O. B., Ruggles, J. A., and Schultz, S. C. (2002) *Nat. Struct. Biol. 9*, 182−187.
14. Gray, J. T., Celander, D. W., Price, C. M., and Cech, T. R. (1991) *Cell 67*, 807−814.
15. Fang, G., and Cech, T. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6056−6060.
16. Murzin, A. G. (1993) *EMBO J. 12*, 861−867.
17. Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003) *Annu. Rev. Biophys. Biomol. Struct. 32*, 45−133.
18. Arcus, V. (2002) *Curr. Opin. Struct. Biol. 12*, 794−801.
19. Fang, G., Gray, J. T., and Cech, T. R. (1993) *Genes Dev. 7*, 870−882.
20. Wang, W., Skopp, R., Scofield, M., and Price, C. (1992) *Nucleic Acids Res. 20*, 6621−6629.
21. Baumann, P., and Cech, T. R. (2001) *Science 292*, 1171−1175.
22. Raghuraman, M. K., Dunn, C. J., Hicke, B. J., and Cech, T. R. (1989) *Nucleic Acids Res. 17*, 4235−4253.
23. Price, C. M., Skopp, R., Krueger, J., and Williams, D. (1992) *Biochemistry 31*, 10835−10843.
24. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A., and Schultz, S. C. (1998) *Cell 95*, 963−974.
25. Classen, S., Ruggles, J. A., and Schultz, S. C. (2001) *J. Mol. Biol. 314*, 1113−1125.
26. Schuck, P., and Minton, A. P. (1996) *Anal. Biochem. 240*, 262−272.
27. Lei, M., Baumann, P., and Cech, T. R. (2002) *Biochemistry 41*, 14560−14568.
28. Correll, C. C., Wool, I. G., and Munishkin, A. (1999) *J. Mol. Biol. 292*, 275−287.
29. Ryter, J. M., and Schultz, S. C. (1998) *EMBO J. 17*, 7505−7513.
30. Dougherty, D. A., and Stauffer, D. A. (1990) *Science 250*, 1558−1560.
31. Gallivan, J. P., and Dougherty, D. A. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 9459−9464.
32. Wintjens, R., Lievin, J., Rooman, M., and Buisine, E. (2000) *J. Mol. Biol. 302*, 395−410.
33. von Hippel, P. H., and Berg, O. G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1608−1612.
34. Ha, J. H., Spolar, R. S., and Record, M. T., Jr. (1989) *J. Mol. Biol. 209*, 801−816.
35. Ackroyd, P. C., Cleary, J., and Glick, G. D. (2001) *Biochemistry 40*, 2911−2922.
36. Goddard, N. L., Bonnet, G., Krichevsky, O., and Libchaber, A. (2000) *Phys. Rev. Lett. 85*, 2400−2403.
37. Zhang, Y., Zhou, H., and Ou-Yang, Z. C. (2001) *Biophys. J. 81*, 1133−1143.
38. Petersheim, M., and Turner, D. H. (1983) *Biochemistry 22*, 256−263.
39. Klobutcher, L. A., Swanton, M. T., Donini, P., and Prescott, D. M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3015−3019.

BI0273718