Identification of the single-strand telomeric DNA binding domain of the *Saccharomyces cerevisiae* Cdc13 protein

Timothy R. Hughes[†‡], Rodney G. Weilbaecher[§], Monika Walterscheid[§], and Victoria Lundblad[†§¶]

[†]Program in Cell and Molecular Biology and [§]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030

Edited by Thomas R. Cech, Howard Hughes Medical Institute, Chevy Chase, MD, and approved March 28, 2000 (received for review July 30, 1999)

The *CDC13* gene of *Saccharomyces cerevisiae* is required both to protect telomeric DNA and to ensure proper function of yeast telomerase *in vivo*. We have previously demonstrated that Cdc13p has a high affinity single-strand telomeric DNA binding activity, although the primary amino acid sequence of Cdc13p has no previously characterized DNA binding motifs. We report here mapping of the Cdc13 DNA binding domain by a combination of proteolysis mapping and deletion cloning. The DNA binding domain maps to residues 557–694 of the 924-amino acid Cdc13 polypeptide, within the most basic region of Cdc13p. A slightly larger version of this domain can be efficiently expressed in *Escherichia coli* as a soluble small protein, with DNA binding properties comparable to those of the full-length protein. A single amino acid missense mutation within this domain results in thermolabile DNA binding and conditional lethality in yeast, consistent with the prediction that DNA binding should be essential for *CDC13* function. These results show that Cdc13p contains a discrete substructure responsible for DNA binding and should facilitate structural characterization of this telomere binding protein.

CDC13 | telomeres | DNA binding protein | telomerase

Telomeres, the ends of linear eukaryotic chromosomes, serve multiple functions in the cell. They have long been known to be distinguished from DNA ends generated by double-strand breaks (1–3) and are required for stability of chromosomes (4, 5); telomeres are thus believed to act as "caps" at chromosomal termini. In addition, the telomere plays a specialized role in DNA replication as the substrate for the enzyme telomerase, which elongates the 3′ G-rich overhang found at chromosome ends (reviewed in ref. 6). Telomerase therefore counters the sequence loss that would otherwise accumulate because of the inability of the lagging strand polymerase to completely replicate DNA ends (7–9).

The *Saccharomyces cerevisiae* Cdc13 protein has emerged as a key component in yeast telomere metabolism, with biochemical and genetic features suggesting a role as a chromosome "capping" factor. Cdc13 is a single-strand telomeric DNA binding protein (10, 11) with at least two discrete roles at the yeast telomere: it protects telomeres from DNA damage (12, 13), and it also mediates the *in vivo* function of telomerase (10, 14). *CDC13* is required for immediate cell viability (12), thereby distinguishing it from other components of the telomerase pathway (4, 15–17). However, the identification of a novel allele of *CDC13* (*cdc13-2*), which has the *in vivo* phenotypes of a telomerase defect (10), yet is competent for yeast telomerase catalytic activity (18), has led to the proposal that Cdc13p mediates telomerase access while bound to the telomere (10, 14). Binding of Cdc13p to the telomere also is postulated to provide protection of chromosome ends (12, 13). This role may go beyond passive protection, in that Cdc13p has been proposed to regulate a nuclease responsible for creating the long single-stranded overhangs observed transiently at *S. cerevisiae* telomeres in late S phase (19). This stems from the observation that the ultimately lethal DNA damage suffered in the absence of *CDC13* function is specific to the telomeric C-strand (12, 13), which is also the target of the mechanism that creates G-rich single-strand overhangs at wild-type telomeres (19). These observations predict that Cdc13 DNA binding activity is required for cell viability, but this has not yet been tested, as neither of the two known alleles of *CDC13* (*cdc13-1* and *cdc13-2*) alter DNA binding (ref. 10; this work).

In several species of ciliated protozoans, chromosome ends are complexed with telomere-specific binding proteins, which recognize the 3′ overhanging structure and protect bound DNA from nuclease treatment (20–22), although it has not been demonstrated that these proteins mediate chromosome capping *in vivo*. The single-strand telomeric DNA binding activity from *Oxytricha nova* is composed of two subunits, α (56 kDa) and β (41 kDa), which bind telomeric substrates cooperatively in a 1:1:1 ratio, with both subunits required for high affinity binding (23, 24). The crystal structure of the *O. nova* proteins complexed with single-stranded DNA has revealed that three OB (oligonucleotide/oligosaccharide-binding) folds collaborate to form the cleft that binds DNA, providing the first molecular detail of how a single-strand telomere binding protein interacts with its substrate (25).

Cdc13 is notable in that it is a sequence-specific DNA binding protein with extremely high affinity for its single-strand telomeric substrate (10), comparable to that of many sequence-specific duplex DNA binding proteins. However, the primary sequence of Cdc13p shows no overt similarity to other DNA binding proteins, including the ciliate telomere DNA binding proteins. Therefore, this biochemical property of Cdc13p is largely uncharacterized, as neither its functional significance nor its structural requirements have been investigated. To conduct a more detailed examination of the role of DNA binding in *CDC13* function, we have mapped the Cdc13 DNA binding domain by a combination of proteolysis mapping and deletion cloning. This 130-amino acid domain can be readily expressed as a small recombinant protein in *Escherichia coli* with DNA binding properties similar to the full length Cdc13 protein. A single amino acid missense mutation that maps within this domain results in thermolabile DNA binding *in vitro* and temperature-sensitive growth in yeast, consistent with an essential role for DNA binding in *CDC13* function. We also show that successive Cdc13 proteins can bind along oligonucleotides of sufficient length; if multimeric complexes are capable of forming on a

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: DBD, DNA binding domain.

[‡]Present address: Rosetta Inpharmatics, Kirkland, WA 98034.

[¶]To whom reprint requests should be addressed. E-mail: lundblad@bcm.tmc.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

single telomeric overhang *in vivo*, this suggests a potential role for Cdc13p as a sensor of overhang length.

Materials and Methods

Strains and Plasmids. Proteins produced in *E. coli* were expressed in BL21-DE3 cells, containing the T7 polymerase gene integrated into the genome, from plasmids expressing all or part of the 924-amino acid *CDC13* coding sequence from the T7 promoter. A schematic diagram of the constructs used in this work to express protein in *E. coli* is shown in Fig. 2*A*. The parent plasmid, pVL427, expressed the full-length Cdc13 protein with a His_6 tag at the N terminus. Deletion constructs were created by deleting sequences between convenient restriction sites in pVL427; constructs not starting with amino acid 1 have an engineered AUG start codon but are missing the His_6 tag. pVL610 and pVL683 encode additional recombinant material (WNSKLDPAANKARKEAELAAATAEQ*) at the C terminus, which is derived from the polylinker because of the absence of the Cdc13 stop codon. pVL683 encodes the 34-kDa Cdc13 DNA binding domain (DBD) protein (consisting of amino acids 1–19 and 452–694 of Cdc13p, plus a His_6 tag inserted immediately after the start codon); pVL1294, derived from pVL683, encodes the mutant Cdc13-6 DBD protein. pVL773, constructed by a PCR strategy, encodes the 28-kDa Cdc13 DBD protein (amino acids 450–694 of Cdc13p bracketed by a start methionine and a stop codon, without a His_6 tag). Full-length Cdc13-1 and Cdc13-2 proteins were expressed in *E. coli* from plasmids pVL518 and pVL517, respectively, which are identical to pVL427 with the exception of the relevant single amino acid mutations (10). Full-length Cdc13 and Cdc13-6 proteins, expressed in yeast in single copy by the *CDC13* promoter from plasmids pVL440 and pVL1295, respectively, were introduced into a *cdc13*-Δ yeast strain for *in vivo* analysis, as described (14).

Protein Production and Purification. Full-length His_6-tagged Cdc13 protein was produced from a baculovirus expression system and was purified by using Ni/nitrilotriacetic acid affinity chromatography exactly as described in ref. 10. Proteins produced in *E. coli* were purified by essentially the same protocol. Cultures were grown at 37°C in LB Amp (50 mg/liter) to OD_{600} = 0.5, and expression was induced by addition of IPTG to 1 mM, followed by incubation at 27°C at 250 rpm for 2–3 h. All subsequent steps were at 4°C. Cells were pelleted, were washed and resuspended in sonication buffer (50 mM sodium phosphate, pH 8.0/300 mM NaCl/10% glycerol/0.5% Tween 20/10 mM imidazole/1 mM PMSF), and were sonicated for 3–4 repetitions of 10–30 sec, and extracts were clarified by centrifugation in a microcentrifuge for 10 min. His_6-tagged proteins were purified on Ni/nitrilotriacetic acid agarose (Qiagen, Chatsworth, CA) by batch method: Ni/nitrilotriacetic acid agarose beads (20–50 μl) were added to each ml of extract and were rotated for 30–60 min at 4°C. Beads were harvested by brief centrifugation (≈1 sec) and were washed once with sonication buffer, twice with sonication buffer adjusted to pH 6.1, and finally with sonication buffer lacking NaCl; all washes were for ≈10 min with >20× bead volume. His_6-tagged proteins were eluted by two to three sequential 15-min incubations with 1× bead volume of 50 mM sodium phosphate (pH 8.0) and 250 mM imidazole. The 28-kDa Cdc13 DBD protein produced by pVL773 does not bear a His_6 tag but nonetheless exhibited an affinity for the Ni/nitrilotriacetic acid agarose resin; affinity purification was carried out as described above, with the exception that imidazole was omitted from binding and wash buffers. For experiments shown in Fig. 2*A*, several of the deletion constructs lacked a His_6 tag; therefore, *E. coli* extracts were used (after clarification by microcentrifugation) in assessing the DNA binding ability of this panel of deletion constructs. These clarified crude extracts contained a Cdc13p-dependent binding activity that was not present in mock extracts (Fig. 3*C*). Extracts prepared from BL21-DE3 *E. coli* cells expressing derivatives of Cdc13p were compared with the activities produced by pVL427 and empty vector, as positive and negative controls.

DNA Gel Shift Assays. Binding reactions and DNA gel shift assays were performed as described in ref. 10, except for the experiments presented in Table 1, which used a slight modification of the nondenaturing gel conditions (50 mM Tris/20 mM borate/1% glycerol/5% polyacrylamide). Protein and oligonucleotide were added sequentially, were incubated for 15 min on ice, and were electrophoresed at 4°C through nondenaturing polyacrylamide gels at 200 volts. Longer incubations (up to 90 min) or variations in the concentration of the DNA probe (from 10 to 50 pM) did not alter the observed binding constants. Single-stranded oligomers were heat-denatured and snap-cooled on ice just before the binding assay, to minimize the formation of secondary structures. Gel percentages were varied as indicated in figure legends to optimize resolution of small complexes. Binding curves used for the results shown in Table 1 were quantitated on a PhosphorImager and were analyzed by using the IMAGEQUANT software (Molecular Dynamics). Protein concentrations were quantitated by comparison to BSA standards (New England Biolabs) on Coomassie-stained gels, a protocol that has previously agreed with amino acid analysis to within 15% (10). Apparent binding constants were determined as the protein concentration resulting in half-binding to an oligonucleotide substrate at 20 pM. The exact value of the half-binding point was determined from interpolation on a Hill plot. Each value is the average of at least two protein titration series giving results within 10%. All constants for each protein were determined by using the same protein preparation, to ensure internal consistency.

Purification of a Cdc13-Proteolysed Fragment. Binding reactions identical to those used for DNA gel shift assays were assembled and were left on ice for 15 min, followed by addition of proteinase K (United States Biochemical) at 0.25 mg/ml and incubation at either 37°C or 4°C for an additional 15 min. For gel purification of DNA-bound proteolysis products, polyacrylamide-TBE gels were exposed to film to locate the proteolyzed complexes, which were excised from the gel with a scalpel. Protein and oligonucleotide were eluted from the gel slice by three successive elutions in 200 μl of 0.5× TBE (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3), incubating for 30 min at 60°C. The combined lyophilized elutions were resolved on a 15% SDS/tricine gel. The ≈16-kDa major band was transferred to poly(vinylidene difluoride), and N-terminal sequence was obtained at the Protein Chemistry Core Facility at Baylor College of Medicine.

Results

Cdc13 Protein Contains a Discrete DNA Binding Substructure. To ask whether a stable substructure of the Cdc13 protein could be identified that was capable of binding telomeric DNA, full-length Cdc13 protein was subjected to limited digestion with proteinase K, a nonspecific protease, either before or after binding to a telomeric oligonucleotide. DNA-protein complexes were resolved by electrophoresis on a nondenaturing acrylamide gel and were visualized by autoradiography. Several discrete digestion products with increased mobility were generated (Fig. 1*A*). The faster migrating products, which appeared with increased proteolysis, occurred at the expense of, and are presumably derived from, a larger proteolysis product. The stability of the proteolysis products was enhanced if the full-length protein was bound to oligonucleotide before proteolysis (Fig. 1*A*; data not shown). Therefore, Cdc13 protein potentially undergoes a conformational change when bound to DNA,

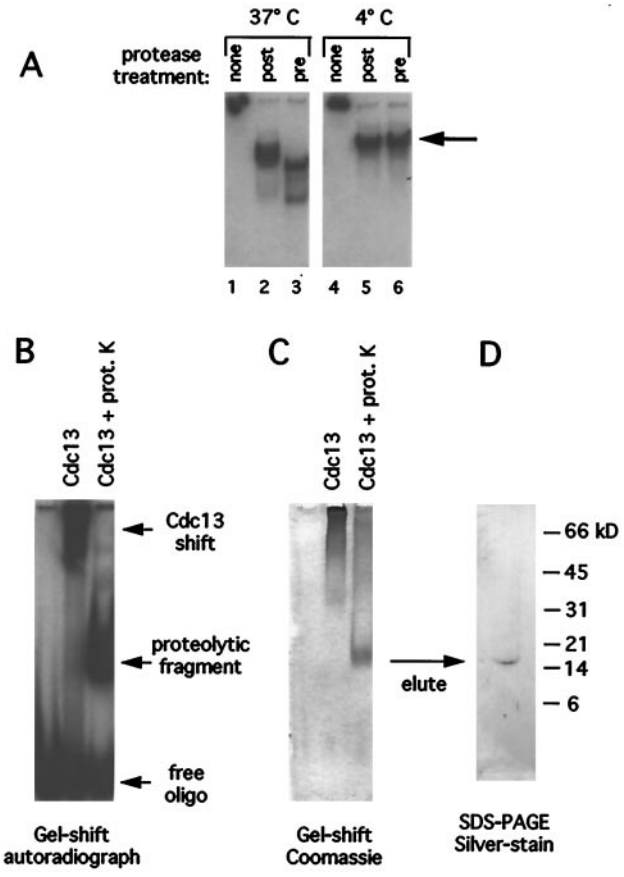

Fig. 1. Proteolysis mapping of the Cdc13 DBD. (*A*) Telomeric DNA [1 nM end-labeled d(TGTGTGGG)$_3$ oligomer], 5 μg of Cdc13 protein, and proteinase K [added either before (lanes 3 and 6) or after (lanes 2 and 5) addition of oligonucleotide] were assembled on ice, and reactions were incubated at 37°C (lanes 1–3) or 4°C (lanes 4–6) for 15 min. Reaction products were run on 10% polyacrylamide/TBE gels at 4°C; under these conditions, the full length Cdc13p:DNA complex does not enter the gel and is retained in the well (lanes 1 and 4). The large arrow indicates the proteolysis product examined in more detail in Fig. 2. (*B*) Autoradiograph of a Cdc13p-DNA gel shift [2 μg of Cdc13p protein and 2 μM 5′ end-labeled d(TGTGTGGG)$_3$ oligomer] with or without subsequent proteinase K digestion at 0°C, resolved on a nondenaturing 8% polyacrylamide-TBE gel. prot K, proteinase K. (*C*) The same gel as in *B*, fixed and Coomassie-stained. (*D*) Silver-stained 15% SDS/PAGE tricine gel showing protein eluted from a gel slice containing a proteolytic fragment identical to the one visible in *C*, with the reaction volume scaled up by a factor of 10.

although a protease-resistant substructure capable of binding nucleic acid apparently exists even in the absence of DNA.

When the extent of digestion was limited by performing the reactions at 4°C in the absence of SDS, a single proteolysis product was readily observed. In scaled-up experiments (Fig. 1*B*), this proteolyzed complex could be visualized by subsequent Coomassie staining of the DNA gel shift (Fig. 1*C*). The increased mobility complex was excised from the gel, was eluted, and was resolved by SDS/PAGE. A single band was visible, migrating at ≈16 kDa (Fig. 1*D*). Sequencing of the amino terminus of this peptide identified (x)YDRYLIDYE, an exact 9-amino acid match starting with residue 557 of the full length Cdc13 protein. This suggests that a peptide consisting of amino acid 557 to ≈690 (predicted molecular mass of ≈16 kDa) can exist in a stable association with DNA and predicts that this portion of the Cdc13 protein encompasses the DNA binding domain. Because this 16-kDa fragment is not the fastest migrating proteolytic complex (Fig. 1*A*), it is likely that there are even smaller internal portions that can retain contact with DNA. However, attempts to identify the smaller domains capable of DNA binding by peptide sequencing were unsuccessful.

Expression of Recombinant Cdc13 DBD. To confirm the results of proteolysis experiments and define more precisely the C-terminal boundary of the Cdc13 DBD, a series of deletions constructs were expressed in *E. coli* and were tested for DNA binding activity by using nondenaturing gels. All plasmids expressing proteins that included amino acids 452–694 produced a DNA binding activity similar to that of the full-length Cdc13 protein, whereas constructs expressing other portions of the Cdc13 polypeptide did not (Fig. 2*A*; data not shown). Two different constructs expressing small recombinant proteins encompassing amino acids 452–694 of Cdc13p (referred to as 34-kDa Cdc13 DBD and 28-kDa Cdc13 DBD) efficiently produced soluble protein in *E. coli* and bound telomeric DNA substrates (Figs. 2*C* and 4*C*) and therefore were used for subsequent analysis.

Limited proteinase K treatment of the recombinant 34-kDa Cdc13 DBD while bound to telomeric DNA also resulted in a complex with a change in mobility. This protease-treated complex roughly comigrated on nondenaturing gels with the product resulting from treatment of full-length Cdc13 protein with proteinase K (Fig. 2*C*), suggesting that the structure excised from the full-length protein is contained within the recombinant Cdc13 DBD protein. In addition, the predicted molecular weight of amino acids 557–694 is 16.4-kDa, in good agreement with the observed migration of the fragment produced by treatment with proteinase K (Fig. 1*D*). Therefore, using peptide fragment sequencing and deletion mapping to define the N and C termini of the DBD, respectively, residues 557–694 comprise the portion of Cdc13p that physically associates with DNA. Even after localization of DNA binding to a restricted region of the Cdc13 polypeptide, no primary sequence similarity to known DNA binding proteins, including the DBD of the ciliate *Oxytricha* α telomere binding protein, could be detected.

Identification of a Thermolabile Mutation Within the DBD of Cdc13. To characterize the *in vivo* requirements for DNA binding, the region from amino acids 452–694 of the intact Cdc13 protein was subjected to PCR mutagenesis. A library of plasmids carrying the mutagenized *CDC13* gene was introduced back into a *cdc13-*Δ yeast strain, and colonies were identified that exhibited temperature-sensitive growth [the details of this mutagenesis are described elsewhere (M.W. and V.L., unpublished work)]. One mutant allele, *cdc13-6*, exhibited a strong temperature-sensitive growth phenotype, because growth was completely abolished at the nonpermissive temperature of 36°C (Fig. 3*A*). Sequence analysis revealed a single missense mutation, V543F, located within the DBD of the Cdc13-6 protein. This domain was subcloned into the 34-kDa Cdc13 DBD construct for protein expression in *E. coli*. As shown in Fig. 3*B*, the Cdc13-6 DBD exhibited a thermolabile defect in DNA binding *in vitro*; binding was reduced approximately 10-fold when protein-DNA reaction mixtures were incubated at 36°C before analysis of complex formation by gel shift analysis, whereas binding of the wild-type Cdc13 DBD was not diminished at the elevated temperature. Therefore, the conditional lethality of the *cdc13-6* mutation *in vivo* is consistent with the prediction that DNA binding is essential for *CDC13* function.

A previously identified temperature-sensitive mutation, *cdc13-1*, contains the missense mutation P371S (10), which maps outside the Cdc13 DBD, although an earlier report had suggested that binding activity of the Cdc13-1 protein was compromised at high temperatures (11). We re-examined this mutant protein, by comparing the DNA binding activity of the Cdc13-1 protein to that of wild-type Cdc13 protein; also included in this analysis was the Cdc13-2 protein (10). All three proteins were

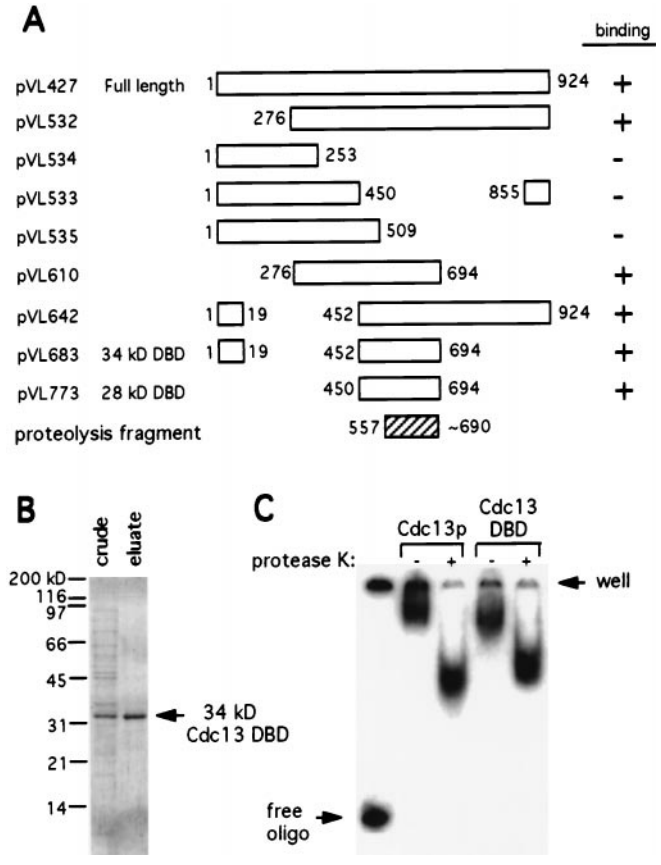

Fig. 2. Deletion mapping and expression of the Cdc13 DBD. (*A*) DNA binding ability of deletion derivatives of Cdc13, as assessed by DNA gel shift assays with *E. coli* crude extracts and 5′ end labeled d(TGTGTGGG)$_3$ oligomer; all constructs indicated as ''+'' exhibited telomeric DNA binding activity comparable to that of the full length protein (data not shown). The N-terminal and predicted C-terminal boundaries of the proteolysis fragment characterized in Fig. 1 (shaded box) are shown for comparison. pVL610 and pVL683 encode additional recombinant material at the C terminus not indicated in the figure (see *Materials and Methods* for details). (*B*) Expression of the 34-kDa recombinant His$_6$-tagged Cdc13 DBD (pVL683) in *E. coli*. Crude extract and eluate samples were resolved on 12% SDS/PAGE and were stained with Coomassie. (*C*) DNA gel shift reactions, with either the 34-kDa Cdc13 DBD or full-length Cdc13p, with or without proteinase K treatment, were resolved on 8% polyacrylamide/TBE.

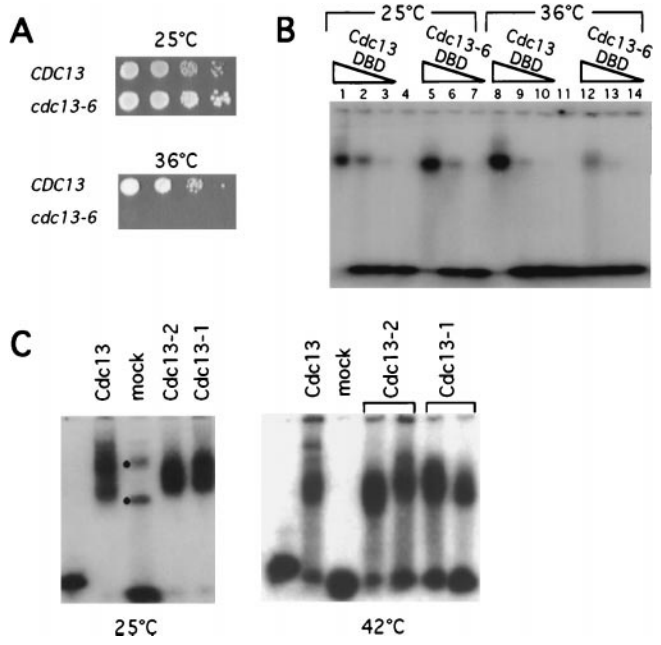

Fig. 3. Identification of a thermolabile DNA binding mutation in Cdc13p. (*A*) Ten-fold serial dilutions of a *CDC13*$^+$ yeast strain (*cdc13*-Δ/pVL440) or a *cdc13-6* strain (*cdc13*-Δ/pVL1295) were spotted onto rich media plates and were grown for 48 h at either 25°C or 36°C. (*B*) Telomeric DNA gel shift assays with 20 nM (TGTGTGGG)$_2$ and purified Cdc13 DBD (lanes 1–3 and lanes 8–10, at 3 μg, 0.75 μg, and 0.15 μg for each set of three lanes) or Cdc13-6 DBD (lanes 5–7 and lanes 12–14, at 2 μg, 0.5 μg, and 0.1 μg for each set of three lanes); reaction mixtures containing protein and reaction buffer were incubated at 25°C (lanes 1–7) or 37°C (lanes 8–14) for 5 min, followed by addition of end-labeled telomeric substrate and incubation for an additional 15 min at the relevant temperature, and were loaded immediately onto 5% polyacrylamide/TBE gel run at 25°C. Incubation at 37°C did not result in degradation of the Cdc13-6 DBD protein (data not shown). (*C*) Telomeric DNA gel shift assays with equivalent volumes of partially purified Cdc13, Cdc13-2, Cdc13-1, or mock protein preparations and 1 nM end-labeled yeast d(TGTGTGGG)$_3$ oligomer. Binding reactions and 5% polyacrylamide/TBE gels were either incubated and run at 25°C (*Left*) or were incubated and run at 42°C, using a prewarmed gel (*Right*). Dots in the mock protein lane of the 25°C gel correspond to contaminating *E. coli* DNA binding activity (or activities).

expressed in *E. coli*, although with low efficiency and resulting in a mixture of full-length and truncation products after purification (data not shown). This gave rise to a somewhat smeared appearance on DNA gel shift assays. Nonetheless, a DNA binding activity comparable to that of the wild type was observed with the Cdc13-1 mutant protein (Fig. 3*C Left*). This activity was present even when assayed at 42°C (Fig. 3*C Right*), a temperature that exceeds by ≈15°C the *in vivo* maximum permissive temperature for growth of a *cdc13-1* yeast strain (12). Therefore, the defect in the Cdc13-1 protein does not appear to be attributable to thermolabile DNA binding, in agreement with the fact that this mutation maps outside the DBD.

Cdc13p Binds a Minimum 11-Base Telomeric Substrate. Binding of full-length Cdc13p and the 34-kDa Cdc13 DBD protein to a panel of telomeric substrates of varying lengths was measured, to determine the minimum binding site and to compare the binding properties of the two proteins. All substrates 11 bases or longer were bound with comparable efficiency, whereas binding efficiency dropped rapidly with substrate length under 11 bases (Table 1). This reduction was primarily caused by a loss of the minimum binding site, rather than a decrease in the overall length of the DNA substrate, as constant length oligonucleotides with 11, 10, or 9 nucleotides of embedded telomeric sequence showed a similar behavior (Table 1). Thus, the minimum Cdc13p binding site appears to be 11 nucleotides. This is in reasonable

Table 1. Characteristics of full-length Cdc13p and Cdc13 DBD binding to substrates of varying lengths

Substrate	Length, bases	K_d^{app} Cdc13p	K_d^{app} Cdc13 DBD
(TGTGTGGG)$_3$	24	0.50 nM	0.39 nM
TGTGTGGGTGTG	12	0.46	0.37
GTGGGTGTGTG	11	0.46	0.37
TGTGGGTGTG	10	55.0	48.0
caaGTGTGGGTGTGaac	11/17	n.t.	0.25
caaaTGTGGGTGTGaac	10/17	n.t.	12.0
caaaTGTGGGTGTaaac	9/17	n.t.	66.0

The apparent binding constant (K_d^{app}) for each substrate was determined from a compilation of three separate but identical gel shift experiments, each with a dilution series of protein and oligonucleotide concentration fixed at 20 pM. n.t., not tested.

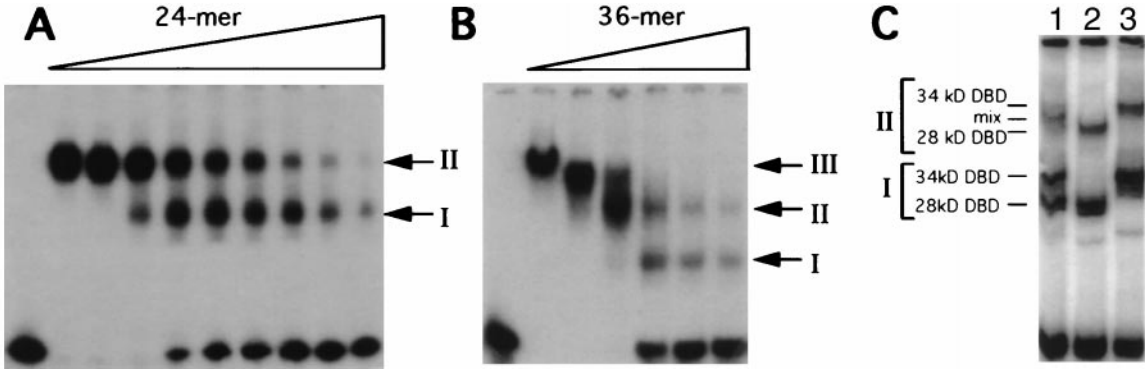

Fig. 4. Cdc13 protein forms multiple complexes with DNA. (*A*) 150 nM 34-kDa recombinant Cdc13 DBD incubated with differing amounts of unlabeled yeast d(TGTGTGGG)$_3$ oligomer (1.5-fold dilutions ranging from 24 to 620 nM), plus 1 nM 5′ end-labeled yeast d(TGTGTGGG)$_3$ oligomer. The two complexes are indicated by arrows. (*B*) 150 nM 34-kDa recombinant Cdc13 DBD incubated with differing amounts of unlabeled d(TG)$_{18}$ oligomer (25 nM, 50 nM, 100 nM, 250 nM, 500 nM, and 1,000 nM, from left to right), plus 1 nM 5′ end-labeled d(TG)$_{18}$ oligomer. (*C*) The 28-kDa Cdc13-DBD (lane 2), the 34-kDa Cdc13-DBD (lane 3), or an equal mixture of each (lane 1) was incubated with 6 μM unlabeled d(TG)$_{18}$ oligomer, plus 1 nM 5′ end-labeled yeast d(TG)$_{18}$ oligomer; proteins, 5 μM. Note that, in the left lane, the mixed complex II band is expected to be present at a 2-fold higher ratio relative to complex II bands that result from solely the 28-kDa protein or the 34-kDa protein.

agreement with the results of Lin and Zakian (11), who used single-stranded substrates with adjacent 5′ duplex telomeric DNA. Table 1 also demonstrates that the 34-kDa Cdc13 DBD bound each telomeric substrate with an affinity indistinguishable from that observed for the full-length protein, further supporting the argument that this construct expresses a domain with properties comparable to that of the intact protein.

Multiple Cdc13 Proteins Can Bind Longer Telomeric DNA Substrates. Previous studies have demonstrated that Cdc13p DNA binding does not require free DNA ends (10, 11). This implies that long single-strand telomeric DNA substrates may be able to serve as tandem Cdc13p binding sites. To examine whether multiple Cdc13 proteins could bind to the same oligonucleotide substrate, a fixed amount of protein was incubated with varying amounts of telomeric oligonucleotide, and the complexes formed were examined by DNA gel shift analysis. In these experiments, both protein and oligonucleotide concentrations were above the binding constant, with reaction conditions proceeding from protein excess to substrate excess. Two discrete complexes, labeled I and II, were formed with a 24-base substrate, using either recombinant Cdc13 DBD (Fig. 4*A*) or full-length Cdc13 protein (data not shown). The smaller complex was formed predominantly under substrate-excess conditions and presumably represents one Cdc13 molecule bound per oligonucleotide (we have not determined whether Cdc13 binds DNA as a monomer or a multimeric complex; for simplicity, this discussion assumes binding as a monomer). The slower migrating complex was favored under protein-excess conditions, consistent with the hypothesis that it is composed of two proteins per oligonucleotide substrate. These data suggest that two Cdc13p molecules can bind adjacent to one another along a 24-base oligonucleotide substrate, which is consistent with the demonstration in Table 1 that the minimum binding site is 11 nucleotides. The number of discernible complexes also could be modulated by changing the length of the substrate oligonucleotide. At least three complexes were observed with a 36-base long oligonucleotide (indicated by arrows in Fig. 4*B*), whereas only a single discrete complex was resolved with an 11-base substrate (data not shown).

As an alternative means of assessing whether slower mobility complexes were caused by binding of more than one Cdc13 protein, a DNA binding experiment was conducted in which proteins of two different sizes (the 34- and 28-kDa recombinant Cdc13 DBD constructs) were mixed together before the addition of oligonucleotide. Complexes were again resolved on a nondenaturing polyacrylamide gel and were examined for the presence of novel complexes, indicative of simultaneous binding of two proteins of different size by a single oligonucleotide molecule (Fig. 4*C*). As predicted, complex I did not form intermediate structures in the mixing experiment; the two complex I bands observed in the "mixed" lane (Fig. 4*C*, lane 1) each comigrated with a complex formed when either protein was added alone to telomeric DNA (lanes 2 and 3). This indicates that complex I contains a single bound Cdc13 protein. However, a new band formed in the complex II region of the gel, when the two proteins were co-incubated with DNA (Fig. 4*C*, lane 1). This intermediate structure was not present in the unmixed lanes, regardless of the protein/DNA ratio (data not shown), indicating that it contained a mixture of the two differently sized Cdc13 proteins. The simplest explanation is that complexes I and II represent one or two Cdc13 proteins, respectively, binding along the length of each oligonucleotide.

Discussion

In the present work, we have localized the Cdc13 DBD to an internal 16-kDa domain. This region can be excised by a nonspecific protease, and a slightly larger version can be expressed autonomously in *E. coli* to produce a DNA binding activity with properties similar to that of the full-length Cdc13 protein. Therefore, the Cdc13 DBD likely forms a functionally distinct module, consistent with the demonstration that this domain, when fused to a subunit of telomerase, increases access of telomerase to the telomere (14). The relatively high resistance of the ≈16-kDa fragment to proteolysis suggests that it also constitutes a single structural domain. The fact that a small DNA binding substructure can be both released by proteolysis and synthesized as a functional recombinant protein is a property shared by the *O. nova* α telomere binding protein (24), as well as the human hnRNP A1 splicing factor (26). hnRNP A1 regulates telomere length *in vivo* (27) and binds both RNA and single-stranded telomeric DNA substrates *in vitro* via highly conserved RRM motifs (28), motifs that are not present in Cdc13p. No primary amino acid sequence similarities were detected between the Cdc13 DBD and the *O. nova* telomere binding protein, although this does not rule out the possibility that these two proteins are structurally similar. The DBDs of these two proteins do in fact share some features: although Cdc13 has an overall predicted pI of 6.55, the Cdc13 DBD (amino acids 557–694) has a pI of 9.95, surprisingly close to that of the *O. nova* α DBD, which has a pI of 9.78. Furthermore, the

Cdc13 DBD is rich in aromatic amino acids, which are involved in a number of DNA base stacking interactions in the *O. nova* crystal structure (25).

This work also has shown that successive Cdc13 proteins can bind along DNA substrates of sufficient length, a property also shared by the *O. nova* α protein (23). Multiple complexes can be observed in DNA gel shift assays, with the number of complexes dictated by the length of the oligonucleotide substrate. This suggests that Cdc13 protein could conceivably bind the entirety of telomeric 3′ single-strand overhangs *in vivo*. In budding yeast, long terminal overhangs arise transiently at the end of S phase, roughly coincident with DNA replication (29); the estimated length of these extended overhangs would be capable of binding multiple Cdc13 molecules. A role for such multimeric complexes might be simply protective, but the potential for multiple proteins to bind to the same terminus also suggests a potential function as a sensor of overhang length.

The identification of a discrete and readily expressed Cdc13 DBD demonstrates that, like many other DNA binding proteins, Cdc13p has a modular structure. This simplifies study of the DNA binding activity and provides a framework in which to evaluate the mechanistic consequences of mutations. We anticipate that the results described here will lead to a more complete understanding of the relationships between the biochemical and *in vivo* behavior of the Cdc13 protein and will aid in determining whether Cdc13 protein shares structural features with other single-strand nucleic acid binding proteins.

We thank Richard Cook and the Protein Chemistry Core Facility at Baylor College of Medicine for peptide sequencing, and Deborah Wuttke and Thomas Cech for comments on the manuscript. This work was supported by National Institutes of Health Grants GM55867 (to V.L.) and F32 GM19755 (to R.G.W.).

1. McClintock, B. (1941) *Genetics* **26,** 234–282.
2. Muller, H. J. (1938) *Collect. Net Woods Hole* **13,** 181–198.
3. Sandell, L. L. & Zakian, V. A. (1993) *Cell* **75,** 729–739.
4. Lundblad, V. & Szostak, J. W. (1989) *Cell* **57,** 633–643.
5. Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. B. & Bacchetti, S. (1992) *EMBO J.* **11,** 1921–1929.
6. Nugent, C. I. & Lundblad, V. (1998) *Genes Dev.* 1073–1085.
7. Watson, J. D. (1972) *Nat. New Biol.* **239,** 197–201.
8. Olovnikov, A. M. (1973) *J. Theor. Biol.* **41,** 181–190.
9. Lingner, J., Cooper, J. P. & Cech, T. R. (1995) *Science* **269,** 1533–1534.
10. Nugent, C. I., Hughes, T. R., Lue, N. F. & Lundblad, V. (1996) *Science* **274,** 249–252.
11. Lin, J.-J. & Zakian, V. A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 13760–13765.
12. Garvik, B., Carson, M. & Hartwell, L. (1995) *Mol. Cell. Biol.* **15,** 6128–6138.
13. Diede, S. J. & Gottschling, D. E. (1999) *Cell* **99,** 723–733.
14. Evans, S. K. & Lundblad, V. (1999) *Science* **286,** 117–120.
15. Lendvay, T. S., Morris, D. K., Sah, J., Balasubramanian, B. & Lundblad, V. (1996) *Genetics* **144,** 1399–1412.
16. Singer, M. S. & Gottschling, D. E. (1994) *Science* **266,** 404–409.
17. Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V. & Cech, T. R. (1997) *Science* **276,** 561–567.
18. Lingner, J., Cech, T. R., Hughes, T. R. & Lundblad, V. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 11190–11195.
19. Wellinger, R. J., Ethier, K., Labrecque, P. & Zakian, V. A. (1996) *Cell* **85,** 423–433.
20. Gottschling, D. E. & Zakian, V. A. (1986) *Cell* **47,** 195–205.
21. Price, C. M. & Cech, T. R. (1987) *Genes Dev.* **1,** 783–793.
22. Price, C. M. (1990) *Mol. Cell. Biol.* **10,** 3421–3431.
23. Gray, J. T., Celander, D. W., Price, C. M. & Cech, T. R. (1991) *Cell* **67,** 807–814.
24. Fang, G., Gray, J. T. & Cech, T. R. (1993) *Genes Dev.* **7,** 870–882.
25. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A. & Schultz, S. C. (1998) *Cell* **95,** 963–974.
26. Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. & Burd, C. G. (1993) *Annu. Rev. Biochem.* **62,** 289–321.
27. LaBranche, H., Dupuis, S., Ben-David, Y., Bani M. R., Wellinger, R. J. & Chabot B. (1998) *Nat. Genet.* **19,** 199–202.
28. Ding J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R. & Xu, R. M. (1999) *Genes Dev.* **13,** 1102–1115.
29. Wellinger, R. J., Wolf, A. J. & Zakian, V. A. (1993) *Cell* **72,** 51–60.