Site-Directed Mutagenesis Reveals the Thermodynamic Requirements for Single-Stranded DNA Recognition by the Telomere-Binding Protein Cdc13†

Emily M. Anderson, Wayne A. Halsey, and Deborah S. Wuttke*

Department of Chemistry and Biochemistry, UCB 215, University of Colorado at Boulder, Boulder, Colorado 80309-0215

Received October 22, 2002; Revised Manuscript Received January 21, 2003

ABSTRACT: The essential *Saccharomyces cerevisiae* protein Cdc13 binds the conserved single-stranded overhang at the end of telomeres and mediates access of protein complexes involved in both end-capping and telomerase activity. The single-stranded DNA-binding domain (ssDBD) of Cdc13 exhibits both high affinity (K_d of 3 pM) and sequence specificity for the GT-rich sequences present at yeast telomeres. We have used the ssDBD of Cdc13 to understand the sequence-specific recognition of extended single-stranded DNA (ssDNA). The recent structure of the Cdc13 DNA-binding domain revealed that ssDNA is recognized by a large protein surface containing an oligonucleotide/oligosaccharide-binding fold (OB-fold) augmented by an extended 30-amino acid loop. Contacts to ssDNA occur via a contiguous surface of aromatic, hydrophobic, and basic residues. A complete alanine scan of the binding interface has been used to determine the contribution of each contacting side chain to binding affinity. Substitution of any aromatic or hydrophobic residue at the interface was deleterious to binding (20 to >700-fold decrease in binding affinity), while tolerance for replacement of basic residues was observed. The important aromatic and hydrophobic contacts are spread throughout the extended interface, indicating that the entire surface is both structurally and thermodynamically required for binding. While all of these contacts are important, several of the individual alanine substitutions that abolish binding cluster to one region of the protein surface. This region is vital for recognition of four bases at the 5′ end of the DNA and constitutes a "hotspot" of binding affinity.

The recognition of single-stranded nucleic acids plays an important role in many essential cellular processes, including telomere regulation (*1*), DNA replication and repair (*2*, *3*), transcription (*4*, *5*), translation (*6*), and RNA processing (*7*). Single-stranded DNA (ssDNA)[1] is often bound by proteins in a sequence-independent manner, as in the case of *Escherichia coli* single-stranded DNA-binding protein (SSB) (*8*) or replication protein A (RepA) (*9*). However, for some processes, such as transcriptional regulation (*10*) or telomere replication and end protection (*1*, *11*−*14*), ssDNA recognition is sequence specific. Single-stranded DNA-binding proteins often recognize an extended nucleic acid conformation involving numerous contacts to the accessible bases. Comparatively little is known about the thermodynamic basis for this mode of recognition relative to our understanding of the protein recognition of double-stranded DNA (*15*−*17*) and structured RNA (*18*−*21*).

The *Saccharomyces cerevisiae* telomere-binding protein Cdc13 is used in this study as a model for high affinity, sequence-specific recognition of ssDNA (*1*). Telomeres contain repetitive tracts of noncoding DNA containing a GT-rich strand with 5′-3′ polarity, terminating in a 3′ single-stranded overhang (*22*, *23*). Cdc13 specifically recognizes this single-stranded overhang and has at least two genetically distinct roles in the cell. The essential role of Cdc13 is a telomere capping function. Cdc13 is a component of a telomere end−protection complex that prevents chromosomal degradation and end-to-end fusion (*24*). Loss of this capping activity results in severe resection of the C-rich strand of the chromosome end, resulting in cell-cycle arrest (*25*, *26*). In addition to capping the telomere, Cdc13 also mediates the recruitment/activation of the replicative enzyme, telomerase, at the telomere in vivo (*27*−*29*). Loss of this activity results in gradual shortening of the telomere, and leads to a delayed lethal phenotype known as replicative senescence.

In contrast to both vertebrates and ciliates, budding yeast telomeres contain a heterogeneous sequence represented by the consensus sequence $G_{2-3}(TG)_{1-6}$ (*30*, *31*). Cdc13 specifically binds yeast single-stranded telomeric DNA with high affinity, but does not appreciably bind RNA or double-stranded DNA (*1*). In vitro competition binding experiments have demonstrated that closely related telomeric sequences, such as *Oxytricha nova* (T_4G_4) and human (T_2AG_3), bind with reduced affinity compared to yeast telomeric ssDNA (*1*). Functionally, Cdc13 must discriminate between other

† This work was supported by the NIH (GM59414), The Arnold and Mabel Beckman Foundation, and a predoctoral fellowship from the U.S. Army Breast Cancer Research Program (EMA, DAMD17-99-1-9150).

* To whom correspondence should be addressed. Phone: 303-492-4576. Fax: 303-492-5894. E-mail: Deborah.Wuttke@colorado.edu.

[1] Abbreviations: ATP, adenosine triphosphate; BSA, bovine serum albumin; DTT, dithiothreitol; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; K_d, apparent equilibrium dissociation constant; Na_2EDTA, disodium ethylenediaminetetraacetic acid; OB-fold, oligonucleotide/oligosaccharide-binding fold; ssDBD, single-stranded DNA-binding domain; ssDNA, single-stranded DNA; TEL-11, single-stranded yeast telomeric DNA of sequence dGTGTGGGTGTG.

10.1021/bi027047c CCC: $25.00
Published on Web 03/18/2003

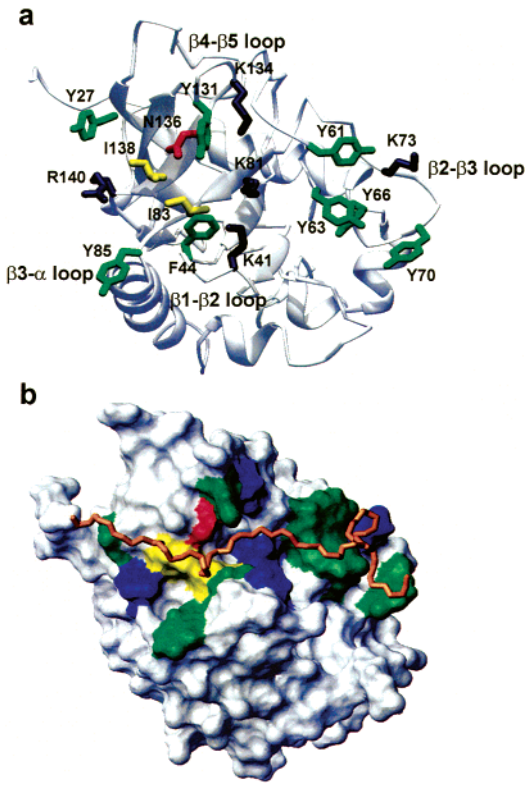

FIGURE 1: A ribbon diagram of the Cdc13 ssDBD with the amino acids that define the DNA-binding interface displayed. (a) The side chains of the residues chosen for alanine mutagenesis are indicated with aromatic groups in green, hydrophobic groups in yellow, and positively charged residues in blue. The loops between $\beta1-\beta2$, $\beta2-\beta3$, $\beta3-\alpha$, and $\beta4-\beta5$ are marked. (b) Space-filling model with the same color scheme and protein orientation as in panel a. The DNA backbone is in gold.

ssDNAs and RNAs transiently present in the nucleus. We are investigating the thermodynamic strategy for achieving this specificity.

The single-stranded DNA-binding domain (ssDBD) of Cdc13 exhibits the same specificity for telomeric DNA as the full-length protein (unpublished data). This domain binds with 100-fold greater affinity (K_d of 3 pM) than the full-length protein (K_d of 310 pM) (*32*). The Cdc13 ssDBD uses an OB-fold topology for DNA recognition (*33*). The oligonucleotide/oligosaccharide-binding fold (OB-fold) superfamily includes several other single-stranded nucleic acid-binding proteins such as *E. coli* SSB, Rho, CspA, CspB, and RepA (*8*, *9*, *34*−*36*). Notably, the structure of the Cdc13 ssDBD is similar to the OB-folds found in the telomere end-binding protein (TEBP) from the ciliate *O. nova* (*37*). While there is no obvious sequence relationship between Cdc13 and *O. nova* TEBP, the *O. nova* protein exhibits sequence similarity to the telomere associated Pot1 proteins found in several species, including humans (*14*), implying that the OB-fold is widely used for recognition of ssDNA at telomeres.

In the context of nucleic-acid binding proteins, the OB-fold domain typically interacts with a small ligand (2−5 nucleotides) through interactions with the loops between $\beta1-\beta2$, $\beta3-\alpha$, and $\beta4-\beta5$ (*38*). This mode of recognition is also observed in the Cdc13 DBD-ssDNA interaction (Figure 1a). However, this canonical OB-fold interaction surface is expanded significantly in Cdc13 by the addition of a 30 amino acid, structurally well-defined loop between $\beta2-\beta3$ (*32*). The extended interface accommodates the entire DNA molecule, explaining the requirement for at least an 11-mer of cognate ssDNA for full binding affinity. Recently, we have determined all of the contacts between the protein and DNA and solved the structure of the protein/DNA complex (*33*) (Figure 1). While extensive contacts are made to the bases, extended, continuous aromatic stacking interactions as seen in the structure of *O. nova* TEBP are not observed between Cdc13 ssDBD and ssDNA (*37*, *39*).

To better understand the molecular determinants of binding in this system and the role of the unusual $\beta2-\beta3$ loop, we have used site-directed alanine mutagenesis to investigate the thermodynamic contributions to ssDNA binding by each of the side chains identified at the Cdc13/DNA interface. Using this method, contacts that contribute energetically to binding can be distinguished from amino acids that are simply proximal to bound DNA in the structure. This complete thermodynamic characterization of the binding interface reveals a large contribution to binding from all of the hydrophobic and aromatic residues at the interface. Four amino acids that recognize the 5′ dGTGT of the minimal 11-mer (dGTGTGGGTGTG) required for high affinity binding contribute disproportionately to the total binding energy.

MATERIALS AND METHODS

Materials. Primers for site-directed mutagenesis were purchased from Invitrogen. *E. coli* cell stocks were obtained from Novagen. All solutions and media were prepared with chemicals purchased from Fisher Scientific and Sigma-Aldrich Company.

Site-Directed Mutagenesis of the Cdc13 DBD. The Cdc13 single-stranded telomeric DNA-binding domain (corresponding to amino acids 497−694) was expressed using the vector pET21a (Novagen) and constructed as described previously (*32*). Site-directed mutagenesis was carried out using suitable primers and the QuickChange mutagenesis kit (Stratagene). The presence of the desired mutation was verified by DNA sequencing of the gene construct.

Expression and Purification of Recombinant Wild-Type and Mutant Proteins. Proteins were expressed and purified as descibed previously (*32*). All mutant proteins expressed in soluble form and were purified in yields comparable to the wild-type protein (5−12 mg/L) except for the R140A mutant, which produced lower yields (0.7 mg/L), due to a reduction in solubility that resulted in protein inclusion bodies.

Nitrocellulose Filter-Binding Assay of Apparent Equilibrium Dissociation Constants. Equilibrium binding reactions were performed with ^{32}P-dGTGTGGGTGTG (TEL-11) at concentrations at least 20-fold below the apparent dissociation constant (50 or 200 pM) using serial dilutions of protein over a broad concentration range. Protein dilutions were prepared in 5 mM HEPES, pH 7.8, 750 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM Na_2EDTA, 2 mM DTT, and 0.1 mg/mL BSA (bovine serum albumin). The reactions were incubated on ice for 60 min to equilibrate. Filter binding assays were performed using 96-well MultiScreen MAHA N4550 filter plates with a MultiScreen Resist Vacuum manifold (Millipore). The wells were prewet for 1 min with 80 μL of binding

buffer without BSA or DTT. Samples (80 μL) were loaded and incubated for 1 min before filtering. While still under vacuum, the wells were washed twice with 200 μL of binding buffer lacking BSA and DTT. The filter was air-dried before being exposed to a PhosphorImager screen. Filter-bound counts were quantified (Imagequant) and plots were fit to a standard two-state binding model: $y = C1(x/(x + K_d)) + C2$, where y is the number of counts of bound DNA, x is the concentration of protein, C1 is the maximum plateau of binding, C2 is the baseline or background counts, and K_d is the apparent equilibrium dissociation constant. K_d values were determined on different days and/or with different protein preparations and are reported as the average value plus or minus the standard deviation of three separate measurements. The large standard deviations observed for the weak binders reflect the upper limit in sensitivity of the assay due to protein saturation of the filter.

Determination of Salt (KCl) Dependence of Wild-Type Cdc13 DBD Binding. The salt dependence of DNA-binding by the Cdc13 DBD to TEL-11 was determined by nitrocellulose filter binding using a range of KCl concentrations. K_d values were obtained in 5 mM HEPES, pH 7.8, 2.5 mM $MgCl_2$, 0.1 mM Na_2EDTA, 2 mM DTT, 0.1 mg/mL BSA, and salt concentrations of 0.6 M KCl, 0.75 M KCl, 0.8 M KCl, 1.0 M KCl, and 1.2 M KCl. As a control to determine that protein binding by the filter was not salt dependent, samples of a range of BSA concentrations were loaded onto filters at each salt concentration followed by staining the filter with Ponceau S. By visual inspection, the filter binding capacity was not affected by salt in the concentration range used in this study.

Gel-Shift Assay of Equilibrium Binding. The 11-mer TEL-11 was 5′-end-labeled using 10 U of T4 DNA kinase with 5 μM DNA and 150 mCi/mL γ^{32}P-ATP. A 25 μL reaction was incubated at 37 °C for 30 min, and unincorporated γ^{32}P-ATP was removed using a microspin G25 column (Pharmacia). Equilibrium binding reactions were performed with ^{32}P 11-mer at 200 pM and proteins at 200 nM in 5 mM HEPES, pH 7.8, 75 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM Na_2-EDTA, 2 mM DTT, and 0.1 mg/mL BSA. The reactions were incubated on ice for 60 min to equilibrate. Ten microliters of each reaction with a small amount of bromophenol blue tracking dye was loaded on a 20 cm × 20 cm × 1.5 mm, 5% acrylamide, nondenaturing gel prerun at a constant 200 V for 45 min. The samples were loaded while running, and the gel was run for another 20 min to separate the free and bound species. The gel was dried and visualized by PhosphorImager.

RESULTS

Single-Stranded DNA Binding by the Cdc13 Exhibits Log−Linear Salt Dependence. The ssDBD of Cdc13 has previously been shown to have a very high affinity for single-stranded telomeric DNA, with a K_d of 3 pM at 75 mM KCl, pH 7.8, and 4 °C (*32*). Binding was characterized to the TEL-11 oligonucleotide, a sequence of DNA that is complementary to the yeast telomerase RNA template and is representative of yeast telomeric sequences (*40*). Natural sequence variants (*30*) bind with comparable affinity (Supporting Information, Table 1). Measurement of a low picomolar K_d requires extremely low concentrations of DNA probe and extended exposure times for PhosphorImager analysis, and is therefore plagued by low signal-to-noise. We used the ionic strength dependence of binding affinity to identify conditions under which screening a large number of mutants would be experimentally tractable. Figure 2a displays representative filter-binding data obtained at 750 mM KCl, along with a plot of these data fit to a standard two-state binding model in Figure 2b. Under these conditions, serial dilutions of protein ranging from 100 fM to 2.5 μM can be used to determine differences in binding affinity that are stronger (K_d of subnanomolar) or substantially weaker than wild-type protein (K_d of up to low micromolar).

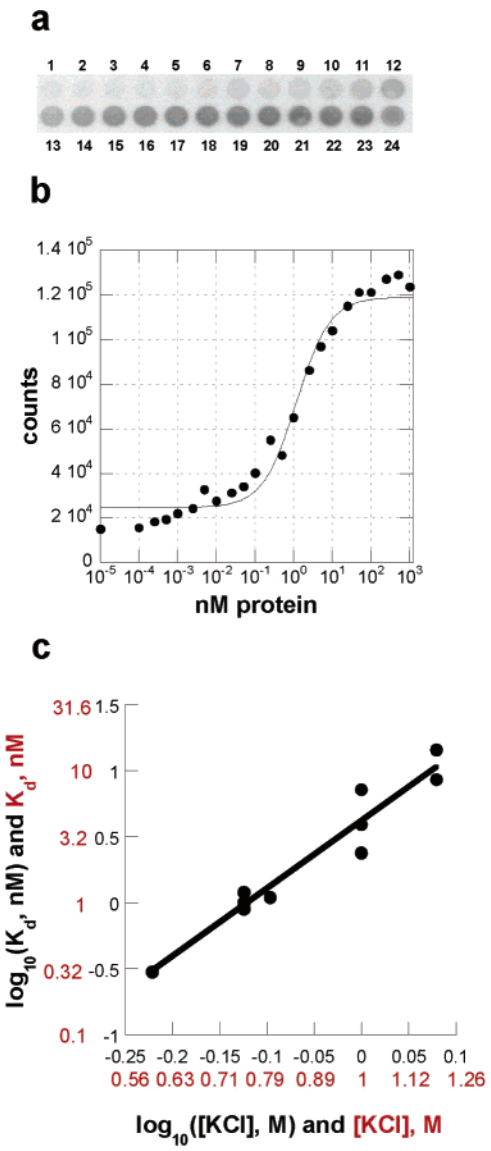

FIGURE 2: A representative nitrocellulose filter-binding experiment for wild-type Cdc13 ssDBD. (a) Raw filter-binding data. The first well contained ^{32}P-labeled DNA probe alone (50 pM), while the remaining wells (labeled 2−24 in order of increasing protein concentration) contained serial dilutions of Cdc13 ssDBD from 100 fM to 2.5 μM. (b) A plot of counts bound to the filter (bound DNA) as a function of the concentration of protein. The data were fit to a standard two-state binding model as described in Materials and Methods. The calculated K_d is 1 nM. (c) A plot of log (K_d, nM) versus log ([KCl], M) for wild-type Cdc13 ssDBD. The data were fit to the line $y = 0.62 + 5.2x$, $R = 0.97$. For convenience, the logarithmic scale markers have been converted back to a linear scale, shown in red. Each point represents a K_d measurement derived from a separate filter binding experiment.

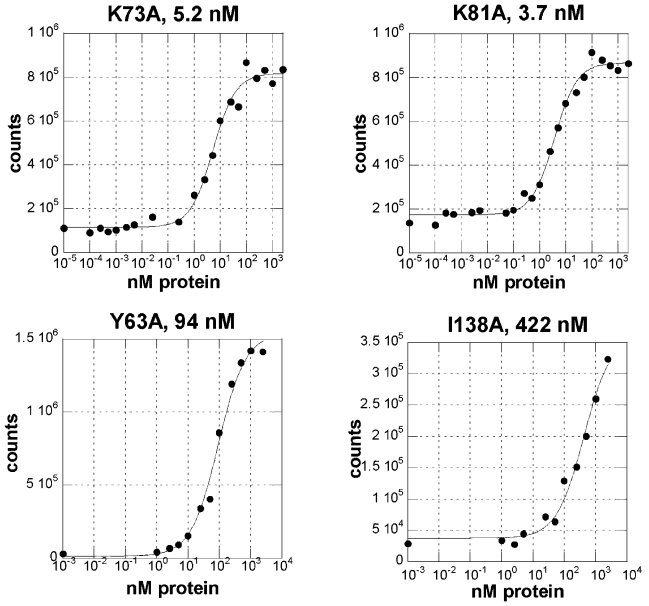

FIGURE 3: Representative binding curves of site-specific alanine mutants of Cdc13 ssDBD with TEL-11. Mutations tested and their calculated K_d values are listed above each graph. The experiments were performed and curves fit as described in Materials and Methods.

The binding of Cdc13 ssDBD and TEL-11 decreased as the concentration of KCl was increased, as reflected by an increase in the apparent equilibrium dissociation constant. The log−log plot of the apparent equilibrium dissociation constant versus the KCl concentration is linear with a slope of approximately 5 (Figure 2c). Thus, the recognition of DNA in this system contains an electrostatic component that can be attenuated by salt to reduce binding affinity to a conveniently measurable range (nM). The linearity of the plot suggests that salt does not perturb the binding interaction in deleterious ways such as denaturation of the protein or formation of alternate structures. Binding conditions were chosen (750 nM KCl) such that the affinity of wild-type Cdc13 DBD for TEL-11 was 1 nM. All subsequent measurements were conducted under the same binding conditions so that extrapolation to different salt concentrations was not necessary for comparison of K_d values.

Site-Directed Alanine Mutagenesis. The DNA-binding interface defined by amino acids that directly contact DNA (*33*) was probed by a panel of single-site alanine mutants (Figure 1). The mutants exhibited a range of affinities to TEL-11 spanning 3 orders of magnitude, and notably none of the alanine point mutants had higher affinity for DNA than wild-type protein. Representative binding curves for both tightly and weakly binding mutants are presented in Figure 3. The binding affinities of the various mutations are presented by amino acid type in Table 1.

Eight aromatic to alanine mutations (amino acid side chains displayed in green on the structure in Figure 1) span the 44 Å long interface, containing Y27A, F44A, Y61A, Y63A, Y66A, Y70A, Y85A, and Y131A. All of these side chains were identified as contacting DNA by intermolecular NOEs (*33*) except Y85, whose side chain could not be assigned but was predicted to contact DNA on the basis of its location in the protein−DNA interface (unpublished experiments). These aromatic to alanine substitutions had dramatic effects on DNA binding (Table 1 and Figure 4),

Table 1: Apparent Equilibrium Dissociation Constants and Differences in Binding Energy for Point Mutants of Cdc13 ssDBD Binding to TEL-11

mutant	average K_d (nM)	std deviation	$\Delta\Delta G$ (kcal/mol)[a]
WT	1.0	0.2	
Y27A	650	210	3.6
F44A	170	68	2.8
Y61A	170	40	2.8
Y61F	14	5	1.5
Y63A	110	36	2.6
Y66A	240	140	3.0
Y70A	23	4.7	1.7
Y85A	680	180	3.6
Y131A	150	63	2.8
I83A	67	11	2.3
I138A	710	260	3.6
N136A	5.4	2.3	1.0
K41A	7.9	4.4	1.1
K73A	8.0	2.5	1.2
K81A	2.8	1.2	0.6
K134A	5.7	2.7	1.0
R140A	540	400	3.5
R140K	1.0	0.2	0

[a] Calculated using $\Delta\Delta G = RT \ln(K_{d1}/K_{d2})$ where K_{d1} is the K_d of the mutant and K_{d2} is the K_d of wild-type protein, $T = 277.15$ K.

ranging from a 23-fold reduction (Y70A) to severe loss of binding (Y85A, 680-fold). Mutation at positions F44, Y61, Y63, Y66, and Y131 had an intermediate effect on binding (110−240-fold). Thus, the elongated binding surface contains many side chains thermodynamically important for binding, containing elements located both in the classical OB-fold region and the elongated $\beta2-\beta3$ loop.

Two hydrophobic residues are located at the protein−DNA interface, isoleucine 83 and isoleucine 138. Point mutation of these hydrophobic residues, which are relatively close in space on the protein surface (Figure 1), results in differential effects on DNA binding. The binding affinity of I83A is reduced about 70-fold, while binding affinity is effectively completely lost in I138A (>700-fold reduction).

As would be expected from the salt dependence of binding, the DNA-binding surface of Cdc13 contains multiple positively charged residues, including K41, K73, K81, R140, and a single polar residue, N136. With the prominent exception of R140, mutation of these residues had relatively modest effects on DNA-binding affinity (<10-fold). The higher salt conditions used in this study might be expected to attenuate the effect of a charge to alanine substitution. However, binding of these mutant proteins under low-salt conditions as measured by gel-shift assay ([KCl] = 75 mM, Supporting Information, Figure 1) qualitatively supported the high-salt data. The charge-to-alanine mutants had wild-type or near wild-type affinity in sharp contrast to the aromatic substitutions described above which bound weakly or not at all. One basic residue (K134) is in the vicinity of the DNA-binding interface (~6 Å) but does not directly contact the DNA. Mutation of lysine 134 to alanine had a similar effect on binding affinity as the other lysine mutations (affinity reduced 3−8-fold), supporting the role of nearby charged residues in providing a positively charged surface that generally favors DNA binding without providing specific strong contacts.

Secondary Substitutions. The role of one of the tyrosines (Y61) and the arginine (R140) at the interface were probed in more detail by two conservative mutations. First, a

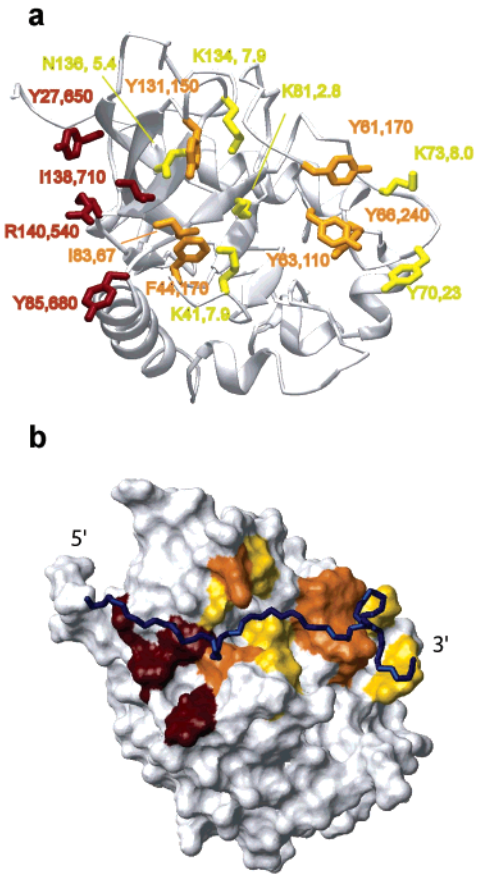

FIGURE 4: Relative binding effects of point mutants mapped on the Cdc13 ssDBD structure. The structures are in the same orientation as Figure 1. (a) The side chains of the residues chosen for alanine mutagenesis and their relative K_d values for binding TEL-11 are indicated. Dark red represents extreme effects (>500-fold), orange represents large effects (70−250-fold), and yellow represents modest effects (3−20-fold). (b) Surface representation of the Cdc13 ssDBD with effects of point mutants scaled according to color as in panel a; the DNA backbone is in blue.

substitution of tyrosine to phenylalanine at position 61 was constructed, which resulted in a 15-fold decrease in affinity. By contrast, the Y61A mutation reduced the apparent binding affinity by close to 200-fold. This significant rescue of binding activity by conservative substitution of one aromatic by another aromatic side chain indicates that an aromatic amino acid is required for DNA binding at this position.

In sharp contrast to the other charged residues, replacement of arginine 140 with alanine had a dramatic effect on binding. An R140K mutant was used to test whether the guanido group of arginine was mediating a specific interaction with DNA. Surprisingly, binding was completely restored in the R140K mutant, thus indicating that at this position only a charged residue is necessary, possibly for a very specific network of hydrogen bonds or as a determinant of protein structural integrity.

DISCUSSION

To investigate the individual thermodynamic contributions of amino acids located throughout the protein−DNA interface, a complete alanine scan of the DNA contact residues of the Cdc13 ssDBD was performed. Extended, single-stranded GT-rich DNA is recognized by the protein across an elongated cleft formed by one side of the OB-fold barrel and an unusually long (30 amino acid) loop between β-strands 2 and 3 (*33*) (Figure 1a). The binding surface identified by direct contacts to the DNA (Figure 1b) is composed primarily of aromatic, hydrophobic, and positively charged residues. These aromatic and hydrophobic residues, as well as arginine 140, are well conserved in Cdc13 homologues from closely related species of yeast (R. Cervantes and V. Lundblad, personal communication). Lysine residues at the interface are less well conserved, although the overall basic character of the interface is preserved. While the structural and phylogenetic data provide insight into the nature of the Cdc13 ssDBD/ssDNA interface, mutagenesis data are necessary to determine which interactions significantly contribute to ssDNA-binding affinity.

To deconvolute the thermodynamic components underlying ssDNA binding, we first determined the ionic strength dependence of Cdc13 ssDBD binding to the minimal oligonucleotide TEL-11. The dependence of binding on concentration of KCl is log−linear with a slope of approximately five (Figure 2c). The attenuation of binding with increasing ionic strength is consistent with the presence of several positively charged protein residues at the protein−DNA interface that contribute to the positive electrostatic surface potential of the Cdc13 ssDBD in the region of the ssDNA-binding interface (Figure 5). A log−linear salt dependence of DNA binding is typical of protein−DNA interactions and has been well characterized with double-stranded DNA-binding proteins (*41*−*44*).

Point mutants of the Cdc13 ssDBD across this interface exhibited a wide range (3 orders of magnitude) of binding affinities for TEL-11 (Table 1 and Figure 4). These comparative binding assays were conducted at 750 mM KCl, where the wild-type interaction was experimentally tractable (K_d of 1 nM, $\Delta G = 11.5$ kcal/mol at 4 °C). In general, the charge-to-alanine substitutions had uniformly small effects on binding, with one exception discussed below. These small effects are not unique to our high salt conditions, as the relative binding affinities of the mutants at low salt (75 mM KCl) revealed similar trends in binding affinity (Figure 1, Supporting Information). It remains to be determined if the charged residues at the interface contribute more to specificity of binding than affinity. The energetic effects of the charge substitutions are most likely additive and contributing to an overall electrostatic surface, as the slope of the salt dependence of the K81A mutant protein was attenuated by a value close to one (data not shown).

A significant exception to the generally small energetic effects of charge-to-alanine mutations was observed at the conserved arginine 140. Replacement of the arginine with alanine at this position caused a decrease in binding of over 500-fold, which was completely restored by reversion to another charged residue, lysine. Therefore, a positively charged residue is essential at this position, possibly to nucleate an essential recognition element by hydrogen bonding either to protein itself or to the DNA. In the structure of the complex, R140 interacts with the base of T2 and may contribute both to specificity and affinity at this position. Arginine is often present at the positively charged surfaces of proteins that bind nucleic acids and is critical for the base-specific recognition of nucleic acids, for example in the Tat/Tar complex (*45*), the U1A protein (*46*), and in zinc-finger protein/DNA complexes (*47*, *48*). It is unusual, however, for a critical arginine residue to tolerate replacement by lysine.

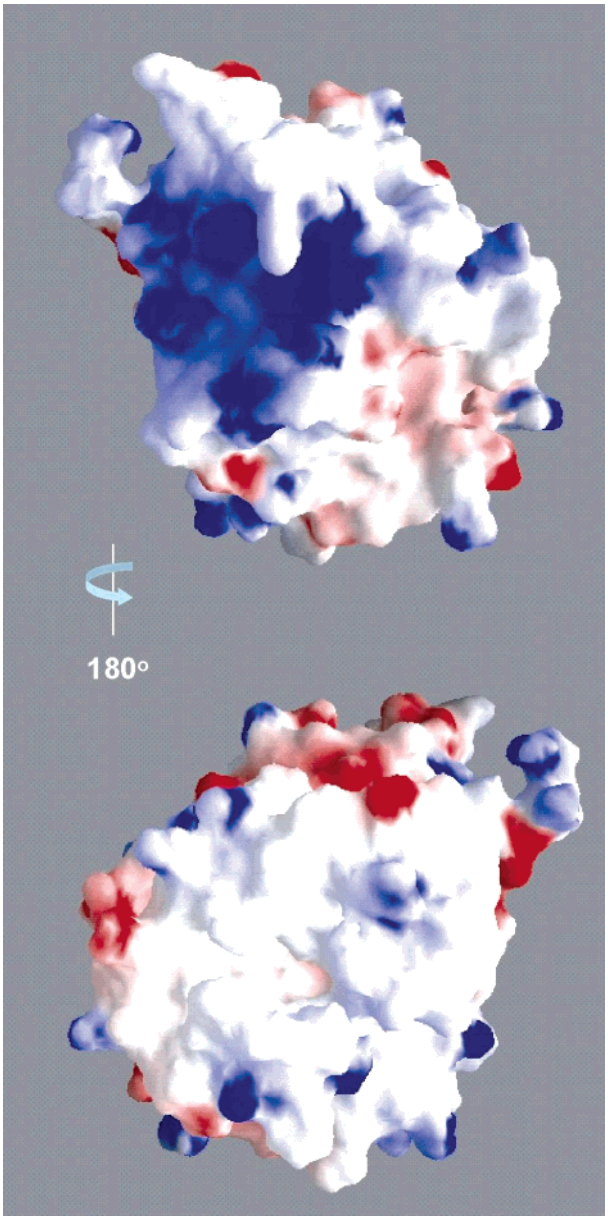

FIGURE 5: An electrostatic surface representation of Cdc13 ssDBD. The top orientation of the domain is the same as for Figures 1 and 3. Positively charged residues are in blue, and negatively charged residues are in red. This figure was generated using the program GRASP (*69*).

In contrast to the mostly modest effects observed in the mutation of charged residues, every individual aromatic and hydrophobic substitution had a pronounced effect on binding affinity. Our structural data clearly revealed intimate interactions directly with DNA nucleotides for every aliphatic and aromatic amino acid chosen for mutational analysis ((*33*) and Supporting Information, Table 2). The DNA-binding site is bounded by Y27 and Y70, which recognize bases at the 5′ and 3′ end, respectively. The remaining structurally defined aromatics, F44, Y131, Y63, and Y61, contact the bases of T4, G5, and multiple 3′ nucleotides, respectively. On the basis of the rescue of binding activity by Y61F, aromatic character is required for appropriate recognition at this site, consistent with a stacking interaction. Most of the aromatic contact sites are strictly conserved with Y or F when compared to Cdc13 sequences from closely related yeasts (R. Cervantes and V. Lundblad, personal communication). The large contribution of aromatic residues to DNA binding observed in Cdc13 is similar to other ssDNA- and RNA-binding proteins. Mutation of several conserved aromatics to alanine in the OB-fold domains of RepA protein was deleterious to binding (*49*). Mutation of a conserved tryptophan (W54S) in *E. coli* SSB reduced both ssDNA binding and inter-tetramer cooperativity (*50*). In both RepA and SSB the conserved critical aromatic residues stack with DNA bases. Additionally, an alanine point mutant of the phenylalanine providing the bulk of the recognition energy of aspartyl tRNA synthetase N-terminal domain reduced binding by 35-fold for the anticodon loop (*51*). This mode of interaction is present in RRM (RNA recognition motif) domains as well—mutation of F56 in U1A protein destabilized its interaction with RNA by 5.5 kcal/mol (*52*).

The effects of the aromatic and hydrophobic point mutations are not additive. The $\Delta\Delta G$ for each mutant summed for all sites at the interface greatly exceeds the total wild-type binding energy (Table 1), suggesting that these interactions are not governed by simple shape complementarity or lock-and-key type interactions. In contrast, separable interactions are more typically features of the recognition of a scaffold such as B-form helical DNA (*47*, *53*−*56*) and ordered RNA (*57*). The significant cooperativity observed between protein side chains in the Cdc13 ssDBD is indicative of their strong physical linkage (*58*). Upon binding, the flexible, ssDNA could adopt a more ordered state to be recognized by Cdc13, similar to a folding event. This implies that aromatic amino acid/base stacking could provide the large enthalpic contribution required for a loss of DNA conformational entropy. The hydrophobic nature of the Cdc13/ssDNA interface is more characteristic of a protein/protein interface than a protein/dsDNA interface (*17*), and is similar to proteins which recognize ssRNA, such as RRM domains (*59*−*61*).

The point mutants that have large effects on binding are spread throughout the DNA-binding interface (Figure 4), confirming that this unusually large interface is required for high-affinity binding of the minimal DNA 11-mer, including the large $\beta2-\beta3$ loop. However, from comparison of the relative effects of the aromatic and hydrophobic alanine mutants (Figure 4b), the effects on binding can be grouped into two subtypes. One group exhibits large effects, with four mutants essentially eliminating binding. Interestingly, these mutants cluster on the side of the protein/ssDNA interface that interacts with 5′-most 4 nucleotides (GTGT) of the DNA (Figure 4b). This region can be considered a "hotspot" of binding energy (*62*) and suggests a mechanism via which Cdc13 recognizes the heterogeneous sequences present at yeast telomeres. This hotspot contains recognition regions (strands $\beta1$, $\beta3$, and $\beta5$, along with loops $\beta1-\beta2$, $\beta3-\alpha$, and $\beta4-\beta5$) that are classical for OB-fold recognition of a small stretch of nucleic acid. Thus, Cdc13 primarily recognizes the GTGT sequence at the 5′ end of its target. This region also exhibits the greatest specificity for ssDNA, as it is the least tolerant to base substitution (unpublished data). This interaction is augmented by inclusion of the 30 amino acid $\beta2-\beta3$ loop containing numerous tyrosines (Y61, Y63, Y66, Y70) that add to the total binding energy in recognizing the 3′ end, which is also an alternating GT sequence (TGTG). From this model, an inexact number of intervening G bases might be tolerated as long as the total length of the oligomer

is 10−11 nucleotides, consistent with sequences observed at yeast telomeres (Supporting Information, Table 1) (*30*). Experiments are in progress to further elucidate the exquisite specificity of Cdc13 for yeast telomeric DNA.

The structural and thermodynamic importance of direct aromatic interaction supplanted by hydrophobic contacts with exposed bases has been observed in some of the other nucleic acid-binding members of the OB-fold family, including AspRS (*52*), LysRS (*63*), CspB (*36*, *64*, *65*), and RepA (*9*, *49*, *66*). While it is common to observe some aromatic residues involved in recognition, a total of eight in Cdc13 is quite unusual, considering that they are all contributing thermodynamically to binding. In contrast, *O. nova* TEBP uses three OB-fold domains containing a total of 10 aromatic residues to specifically recognize a ssDNA 12-mer (*37*). This is consistent with the observation that the minimal ssDNA requirement for Cdc13 is 11 nucleotides (*67*), while for most other single OB-folds it is 2−5. The aromatic and hydrophobic nature of the Cdc13 interface differs from what is used by classical dsDNA binding proteins, but it is more similar to extended RNA-binding proteins, such as U1A (*59*) and TRAP (*68*), or to protein/peptide interfaces (*17*). In the case of Cdc13, expansion of the classical OB-fold binding motif with a large $\beta2-\beta3$ loop has been suitably adapted for specific interaction with a relatively long, extended ssDNA. This study suggests how Cdc13 can specifically recognize the heterogeneous G-rich single-stranded sequences found in yeast telomeric DNA.

ACKNOWLEDGMENT

We thank Prof. James Goodrich, Prof. Olke Uhlenbeck, Aimee Eldridge, and David Atkins for careful reading and comments on the manuscript, Rachel Mitton-Fry and Douglas Theobald for helpful insights on the structure of the Cdc13 ssDBD/DNA complex, and Bryn Weaver for assistance in performing some of the binding assays. We also thank Rachel Cervantes and Prof. Vicki Lundblad for providing Cdc13 homologue sequence information prior to publication.

SUPPORTING INFORMATION AVAILABLE

Figure 1 is a Phosphorimage of a gel-shift assay conducted in low salt (75 mM KCl, see Materials and Methods) of WT Cdc13 ssDBD and point mutants performed with 200 pM DNA and 200 nM proteins. Table 1 presents K_d values measured for Cdc13 ssDBD binding to common yeast telomere sequence variants (*30*). Table 2 lists assigned protein side chain/DNA nucleotide contacts (unpublished experiments). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Nugent, C. I., Hughes, T. R., Lue, N. F., and Lundblad, V. (1996) Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science 274*, 249−252.
2. Lohman, T. M., and Ferrari, M. E. (1994) *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities, *Annu. Rev. Biochem. 63*, 527−570.
3. Wold, M. S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem. 66*, 61−92.
4. Swamynathan, S. K., Nambiar, A., and Guntaka, R. V. (1998) Role of single-stranded DNA regions and Y-box proteins in transcriptional regulation of viral and cellular genes. *FASEB J. 12*, 515−522.
5. Raveh, S., Vinh, J., Rossier, J., Agou, F., and Véron, M. (2001) Peptidic determinants and structural model of human NDP kinase B (Nm23-H2) bound to single-stranded DNA. *Biochemistry 40*, 5882−5893.
6. Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell 98*, 835−845.
7. Varani, G., and Nagai, K. (1998) RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct. 27*, 407−445.
8. Raghunathan, S., Kozlov, A. G., Lohman, T. M., and Waksman, G. (2000) Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nat. Struct. Biol. 7*, 648−652.
9. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997) Structure of the single-stranded DNA-binding domain of replication protein A bound to DNA. *Nature 385*, 176−181.
10. Braddock, D. T., Baber, J. L., Levens, D., and Clore, G. M. (2002) Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J. 21*, 3476−3485.
11. Price, C. M., and Cech, T. R. (1987) Telomeric DNA-protein interactions of *Oxytricha* macronuclear DNA. *Genes Dev. 1*, 783−793.
12. Gottschling, D. E., and Zakian, V. A. (1986) Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell 47*, 195−205.
13. Lin, J.-J., and Zakian, V. A. (1996) The *Saccharomyces CDC13* protein is a single-strand TG_{1-3} telomeric DNA-binding protein in vitro that affects telomere behavior in vivo. *Proc. Natl. Acad. Sci. U.S.A. 93*, 13760−13765.
14. Baumann, P., and Cech, T. R. (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science 292*, 1171−1175.
15. Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol. 1*, 1−37.
16. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) Amino acid−base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res. 29*, 2860−2874.
17. Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999) Protein-DNA interactions: a structural analysis. *J. Mol. Biol. 287*, 877−896.
18. Draper, D. E. (1999) Themes in RNA-protein recognition. *J. Mol. Biol. 293*, 255−270.
19. Doudna, J. A., and Rath, V. L. (2002) Structure and function of the eukaryotic ribosome: the next frontier. *Cell 109*, 153−156.
20. Moore, P. B. (2001) The ribosome at atomic resolution. *Biochemistry 40*, 3243−3250.
21. Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell 108*, 557−572.
22. Shore, D. (2001) Telomeric chromatin: replicating and wrapping up chromosome ends. *Curr. Opin. Genet. Dev. 11*, 189−198.
23. Blackburn, E. H. (2001) Switching and signaling at the telomere. *Cell 106*, 661−673.
24. Lustig, A. J. (2001) Cdc13 subcomplexes regulate multiple telomere functions. *Nat. Struct. Biol. 8*, 297−299.
25. Garvik, B., Carson, M., and Hartwell, L. (1995) Single-stranded DNA arising at telomeres in *cdc13* mutants may constitute a specific signal for the *RAD9* checkpoint. *Mol. Cell. Biol. 15*, 6128−6138.
26. Booth, C., Griffith, E., Brady, G., and Lydall, D. (2001) Quantitative amplification of single-stranded DNA (QAOS) demonstrates that *cdc13−1* mutants generate ssDNA in a telomere to centromere direction. *Nucleic Acids Res. 29*, 4414−4422.
27. Evans, S. K., and Lundblad, V. (1999) Est1 and Cdc13 as comediators of telomerase access. *Science 286*, 117−120.
28. Pennock, E., Buckley, K., and Lundblad, V. (2001) Cdc13 delivers separate complexes to the telomere for end protection and replication. *Cell 104*, 387−396.
29. Taggart, A. K. P., Teng, S.-C., and Zakian, V. A. (2002) Est1p as a cell cycle-regulated activator of telomere-bound telomerase. *Science 297*, 1023−1026.
30. Förstemann, K., Höss, M., and Lingner, J. (2000) Telomerase-dependent repeat divergence at the 3′ ends of yeast telomeres. *Nucleic Acids Res. 28*, 2690−2694.

31. Förstemann, K., and Lingner, J. (2001) Molecular basis for telomere repeat divergence in budding yeast. *Mol. Cell. Biol. 21*, 7277−7286.
32. Anderson, E. M., Halsey, W. A., and Wuttke, D. S. (2002) Delineation of the high-affinity single-stranded telomeric DNA-binding domain of *Saccharomyces cerevisiae* Cdc13. *Nucleic Acids Res. 30*, 4305−4313.
33. Mitton-Fry, R. M., Anderson, E. M., Hughes, T. R., Lundblad, V., and Wuttke, D. S. (2002) Conserved structure for single-stranded telomeric DNA recognition. *Science 296*, 145−147.
34. Bogden, C. E., Fass, D., Bergman, N., Nichols, M. D., and Berger, J. M. (1999) The structural basis for terminator recognition by the Rho transcription termination factor. *Mol. Cell 3*, 487−493.
35. Schindelin, H., Jiang, W., Inouye, M., and Heinemann, U. (1994) Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 91*, 5119−5123.
36. Schindelin, H., Marahiel, M. A., and Heinemann, U. (1993) Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature 364*, 164−168.
37. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A., and Schultz, S. C. (1998) Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell 95*, 963−974.
38. Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct. 32*, 115−133.
39. Classen, S., Ruggles, J. A., and Schultz, S. C. (2001) Crystal structure of the N-terminal domain of *Oxytricha nova* telomere end-binding protein α subunit both uncomplexed and complexed with telomeric ssDNA. *J. Mol. Biol. 314*, 1113−1125.
40. Singer, M. S., and Gottschling, D. E. (1994) TLC1: template RNA component of *Saccharomyces cerevisiae* telomerase. *Science 266*, 404−409.
41. Barkley, M. D., Lewis, P. A., and Sullivan, G. E. (1981) Ion effects on the lac repressor−operator equilibrium. *Biochemistry 20*, 3842−3851.
42. Misra, V. K., Hecht, J. L., Sharp, K. A., Friedman, R. A., and Honig, B. (1994) Salt effects on protein-DNA interactions. The λcI repressor and *Eco*RI endonuclease. *J. Mol. Biol. 238*, 264−280.
43. Fogolari, F., Elcock, A. H., Esposito, G., Viglino, P., Briggs, J. M., and McCammon, J. A. (1997) Electrostatic effects in homeodomain-DNA interactions. *J. Mol. Biol. 267*, 368−381.
44. Record, M. T., Jr., Zhang, W., and Anderson, C. F. (1998) Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. *Adv. Protein Chem. 51*, 281−353.
45. Aboul-ela, F., Karn, J., and Varani, G. (1995) The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol. 253*, 313−332.
46. Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature 348*, 515−520.
47. Pavletich, N. P., and Pabo, C. O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science 252*, 809−817.
48. Wolfe, S. A., Nekludova, L., and Pabo, C. O. (2000) DNA recognition by Cys_2His_2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct. 29*, 183−212.
49. Bastin-Shanower, S. A., and Brill, S. J. (2001) Functional analysis of the four DNA binding domains of replication protein A. The role of RPA2 in ssDNA binding. *J. Biol. Chem. 276*, 36446−36453.
50. Ferrari, M. E., Fang, J., and Lohman, T. M. (1997) A mutation in *E. coli* SSB protein (W54S) alters intra-tetramer negative cooperativity and inter-tetramer positive cooperativity for single-stranded DNA binding. *Biophys. Chem. 64*, 235−251.
51. Eriani, G., and Gangloff, J. (1999) Yeast aspartyl-tRNA synthetase residues interacting with $tRNA^{Asp}$ identity bases connectively contribute to $tRNA^{Asp}$ binding in the ground and transition-state complex and discriminate against noncognate tRNAs. *J. Mol. Biol. 291*, 761−773.
52. Blakaj, D. M., McConnell, K. J., Beveridge, D. L., and Baranger, A. M. (2001) Molecular dynamics and thermodynamics of protein-RNA interactions: mutation of a conserved aromatic residue modifies stacking interactions and structural adaptation in the U1A-stem loop 2 RNA complex. *J. Am. Chem. Soc. 123*, 2548−2551.
53. Luscombe, N. M., and Thornton, J. M. (2002) Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. *J. Mol. Biol. 320*, 991−1009.
54. Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994) DNA recognition by beta-sheets in the Arc repressor-operator crystal structure. *Nature 367*, 754−757.
55. Ghosh, G., van Duyne, G., Ghosh, S., and Sigler, P. B. (1995) Structure of NF-κB p50 homodimer bound to a κB site. *Nature 373*, 303−310.
56. Shi, Y., and Berg, J. M. (1995) A direct comparison of the properties of natural and designed zinc-finger proteins. *Chem. Biol. 2*, 83−89.
57. Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W., and Ramakrishnan, V. (1999) A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell 97*, 491−502.
58. Wyman, J., and Gill, S. J. (1990) *Binding and Linkage: Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA.
59. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature 372*, 432−438.
60. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999) Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature 398*, 579−585.
61. Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R.-M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev. 13*, 1102−1115.
62. Clackson, T., and Wells, J. A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science 267*, 383−386.
63. Commans, S., Plateau, P., Blanquet, S., and Dardel, F. (1995) Solution structure of the anticodon-binding domain of *Escherichia coli* lysyl-tRNA synthetase and studies of its interaction with $tRNA^{Lys}$. *J. Mol. Biol. 253*, 100−113.
64. Schnuchel, A., Wiltscheck, R., Czisch, M., Herrler, M., Willimsky, G., Graumann, P., Marahiel, M. A., and Holak, T. A. (1993) Structure in solution of the major cold-shock protein from *Bacillus subtilis*. *Nature 364*, 169−171.
65. Schröder, K., Graumann, P., Schnuchel, A., Holak, T. A., and Marahiel, M. A. (1995) Mutational analysis of the putative nucleic acid-binding surface of the cold-shock domain, CspB, revealed an essential role of aromatic and basic residues in binding of single-stranded DNA containing the Y-box motif. *Mol. Microbiol. 16*, 699−708.
66. Brill, S. J., and Bastin-Shanower, S. (1998) Identification and characterization of the fourth single-stranded-DNA binding domain of replication protein A. *Mol. Cell. Biol. 18*, 7225−7234.
67. Hughes, T. R., Weilbaecher, R. G., Walterscheid, M., and Lundblad, V. (2000) Identification of the single-strand telomeric DNA binding domain of the *Saccharomyces cerevisiae* Cdc13 protein. *Proc. Natl. Acad. Sci. U.S.A. 97*, 6457−6462.
68. Swairjo, M. A., Morales, A. J., Wang, C.-C., Ortiz, A. R., and Schimmel, P. (2000) Crystal structure of Trbp111: a structure-specific tRNA-binding protein. *EMBO J. 19*, 6287−6298.
69. Antson, A. A., Dodson, E. J., Dodson, G., Greaves, R. B., Chen, X., and Gollnick, P. (1999) Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature 401*, 235−242.
70. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins 11*, 281−296.

BI027047C