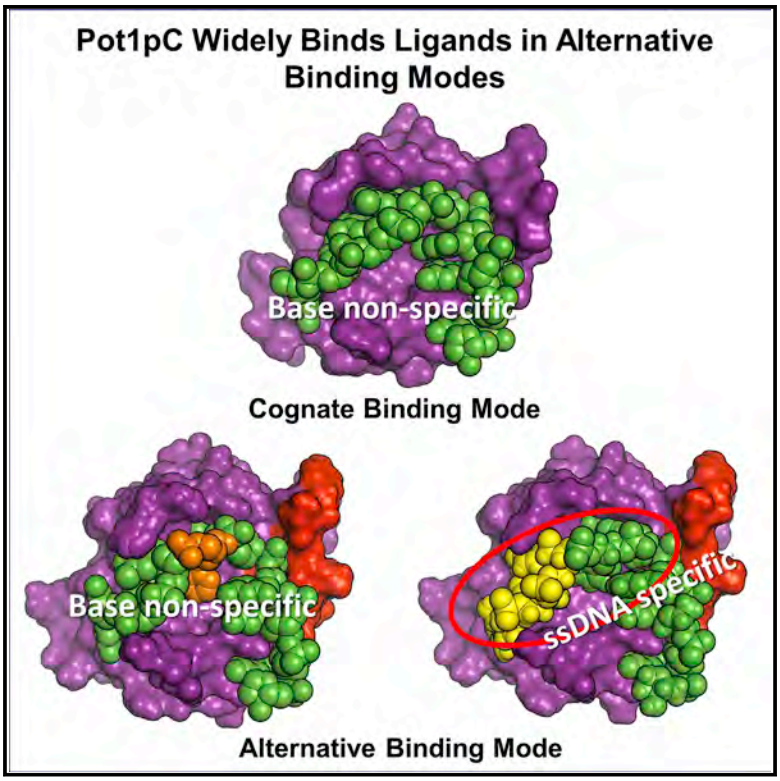
Article

Structure

Discrimination against RNA Backbones by a ssDNA Binding Protein

Graphical Abstract



Authors

Neil R. Lloyd, Deborah S. Wuttke

Correspondence

deborah.wuttke@colorado.edu

In Brief

Lloyd et al. determine the nucleotide-position specificity of the C-terminal domain of *S. pombe* Pot1 for ssDNA over ssRNA and report three ligand-bound structures that reveal RNA-chimeric ligands bind in a widely utilized non-compensatory binding mode that features significant rearrangement of the binding interface to achieve thermodynamically equivalent binding.

Highlights

- ssDNA-binding Pot1pC discriminates against ssRNA at the 5′ end of the ligand
- Structural insight obtained into ribose specificity in the context of base tolerance
- Alternative binding mode enforces suboptimal binding geometries for RNA
- Universally employed alternative binding mode accommodates modifications

Lloyd & Wuttke, 2018, Structure *26*, 722–733
May 1, 2018
https://doi.org/10.1016/j.str.2018.03.016

CellPress

CellPress

Structure
Article

Discrimination against RNA Backbones by a ssDNA Binding Protein

Neil R. Lloyd[1] and Deborah S. Wuttke[1,2,*]
[1]Department of Chemistry and Biochemistry, University of Colorado, UCB 596, Boulder, CO 80309-0596, USA
[2]Lead Contact
*Correspondence: deborah.wuttke@colorado.edu
https://doi.org/10.1016/j.str.2018.03.016

SUMMARY

Pot1 is the shelterin component responsible for the protection of the single-stranded DNA (ssDNA) overhang at telomeres in nearly all eukaryotic organisms. The C-terminal domain of the DNA-binding domain, Pot1pC, exhibits non-specific ssDNA recognition, achieved through thermodynamically equivalent alternative binding conformations. Given this flexibility, it is unclear how specificity for ssDNA over RNA, an activity required for biological function, is achieved. Examination of the ribose-position specificity of Pot1pC shows that ssDNA specificity is additive but not uniformly distributed across the ligand. High-resolution structures of several Pot1pC complexes with RNA-DNA chimeric ligands reveal Pot1pC discriminates against RNA by utilizing non-compensatory binding modes that feature significant rearrangement of the binding interface. These alternative conformations, accessed through both ligand and protein flexibility, recover much, but not all, of the binding energy, leading to the observed reduction in affinities. These findings suggest that intermolecular interfaces are remarkably sophisticated in their tuning of specificity toward flexible ligands.

INTRODUCTION

Protein ligand-binding specificity, the ability to strongly interact with binding partners while preventing undesired interactions with similar molecules, plays a vital role for the proper function of many proteins. Proteins that bind to double- or single-stranded nucleic acids are a classic system for studying ligand specificity, as they are known to discriminate based on sequence, structure, shape, or a combination thereof, to perform their biological functions (Rohs et al., 2010). Properly tuned binding specificity is needed for processes such as transcription factor activation of a particular set of genes, the amino-acylation of tRNAs by their respective synthetases (Giegé et al., 1998), and the TATA binding protein as a general transcription factor (Kim et al., 1993a, 1993b). For sequence-specific interactions with nucleic acids, specificity is widely believed to be achieved through hydrogen bond donor and acceptor patterns that provide shape complementarity not readily satisfied by other species (Messias and Sattler, 2004; Croy and Wuttke, 2006; Cléry et al., 2008). In contrast, sequence-indiscriminate recognition of nucleic acids, important for the function of proteins such as replication protein A (RPA) (Wold, 1997; Bochkarev and Bochkareva, 2004), single-strand break protein (Lohman and Ferrari, 1994), DNA polymerase (Beese et al., 1993), and others, are thought to be largely driven by non-specific stacking/hydrophobic interactions and/or electrostatic interactions with the sugar-phosphate backbone (Record et al., 1976; Wilson et al., 2014).

Recent data suggest that this nucleic acid specificity paradigm, however, deserves broadening. One example is the structure-function specificity exhibited by the second oligonucleotide/oligosaccharide binding (OB)-fold of *Schizosaccharomyces pombe* telomere protection protein Pot1 (Pot1pC). This domain tightly binds the degenerate telomere sequence, $(GGTTAC)(A/AC)_{0-1}(G)_{0-7}$, of *S. pombe* in concert with the first OB-fold of Pot1 (Pot1pN), and together these two OB-folds comprise the DNA-binding domain (DBD) of Pot1 (Lei et al., 2003; Croy et al., 2009; Altschuler et al., 2011). In the full-length DBD, sequence specificity is conferred primarily by the six nucleotide sequence d(GGTTAC) bound by Pot1pN, and, remarkably, full substitution of the nine nucleotides bound by Pot1pC to polyT or the sequence complement results in no binding loss as long as the Pot1pN sequence remains unaltered (Lei et al., 2003; Altschuler et al., 2011). This suggests that the functional flexibility conferred by the Pot1 DBD is achieved through a specific N-terminal domain, Pot1pN, and an equivalently important, but non-specific, C-terminal domain, Pot1pC (Lei et al., 2003; Croy et al., 2009; Altschuler et al., 2011; Dickey et al., 2013).

Surprisingly, the high-resolution structure of the cognate single-stranded DNA (ssDNA)-Pot1pC complex d(GGTTACGGT) reveals hydrogen bond contacts primarily to the bases in a manner commonly associated with the canonical sequence-specific interactions, but the accompanying biochemistry indicates that these Pot1pC interactions do not actually confer specificity (Dickey et al., 2013). Individual substitutions to the complement base at most nucleotide positions for Pot1pC in isolation are fully thermodynamically tolerated (Dickey et al., 2013). This apparent contradiction was reconciled with additional structures of base-substituted ligands, which reveal how Pot1pC non-specificity is accomplished with apparently specific interactions (Dickey et al., 2013). Pot1pC employs plasticity at the interface between the protein and ssDNA ligand by which a chemically rich interface, rearrangement of a flexible protein loop, and flexibility in the ligand, allow the formation of compensatory hydrogen bond networks despite disruptions of cognate

Table 1. Thermodynamic Impact of Ribose Substitutions

dNTP/rNTP[a]	K_D (nM)[b]	Fold Change[c]	ΔH (kcal/mol)[b]	TΔS (kcal/mol)[b]
Cognate **GGTTA CGGT**[d]	24[d]	–	−29[d]	−18[d]
(349)dU **GGUUACGGU**	60 ± 16	2.5	−30 ± 1.7	−20 ± 2
1-3R **GGUTACGGT**	228 ± 24	9.5	−34 ± 1	−24 ± 1
1R **G**GUTACGGT	92 ± 2	3.8	−32 ± 0.3	−22 ± 0.3
2R **GG**TTACGGT	56 ± 10	2.3	−30 ± 1	−20 ± 1
3R **GGU**TACGGT	48 ± 6	2.0	−28 ± 0.7	−18 ± 0.8
4-6R **GGTUACGGT**	154 ± 2	6.4	−29 ± 0.7	−20 ± 0.7
4R **GGTU**ACGGT	57 ± 9	2.4	−29 ± 2	−19 ± 2
5R **GGTT**ACGGT	53 ± 14	2.2	−30 ± 1	−20 ± 1
6R **GGTTA**CGGT	60 ± 4	2.5	−26 ± 0.7	−16 ± 0.7
7-9R **GGTTACGGU**	44 ± 7	1.8	−30 ± 2	−20 ± 2
1-9R **GGUUACGGU**	1930 ± 110	80	−16 ± 2	−8.1 ± 2

[a]Substituted nucleotides are underlined.
[b]Apparent K_D, ΔH, and TΔS are averaged from triplicate isothermal titration calorimetry experiments with SEM. K_D, ΔH, and binding sites (N) were fit using a one-site binding model in MicroCal Origin 7.0 and TΔS calculated from fit values. Representative data shown in Figure S1.
[c]Fold change is relative to cognate DNA K_D (24 nM).
[d]Values from Dickey et al. (2013).

interactions upon ligand base substitutions. These changes in interface conformations range from subtle structural rearrangements to much more dramatic structural rearrangements in both protein and ligand. In *S. pombe*, this plasticity may play dual roles by allowing recognition of degenerate telomere sequences and may provide differentially accessible 3′ substrates for telomere extension (Dickey and Wuttke, 2014).

As the sole autonomous ssDNA binding protein of the shelterin complex (Baumann and Cech, 2001; Palm and de Lange, 2008), Pot1 is critical for end protection of telomeres. Disruptions in the human *POT1* or *S. pombe* Pot1 genes results in DNA-damage response pathways, chromosomal fusion, and cell death (Churikov et al., 2006; Denchi and de Lange, 2007; de Lange, 2009; Rai et al., 2010). In humans, deletion of the Pot1 DBD leads to rapid and extensive telomere elongation (Loayza and De Lange, 2003), whereas full-length deletion in *S. pombe* results in widespread cell death, with the exception of rare survivors that manage to evade mortality by circularizing their chromosomes (Baumann and Cech, 2001).

Given that binding ssDNA is needed for proper telomere maintenance (Denchi and de Lange, 2007), the question arises of how Pot1 discriminates against the vast pool of cellular RNAs. If Pot1 also bound single-stranded RNA (ssRNA), then, at cellular concentrations of each, virtually all Pot1 would be predicted to be sequestered into non-productive Pot1/RNA complexes simply as a result of RNA containing a Pot1 binding sequence by random chance (Lei et al., 2003). Moreover, there is an abundance of RNAs present with the same sequence as the ssDNA at telomeres. Direct transcription of the telomere in *S. pombe* results in a population of RNAs that each contain ~30 potential Pot1 binding sites rendered in RNA (Bah et al., 2012) compared with the 6 total telomere binding sites. Thus, strong discrimination against RNA ligands by *S. pombe* Pot1 (at least ~10^5 based on the concentration of spurious binding sites in all RNA and the low expression of Pot1) is necessary to prevent Pot1 sequestration by RNA. Failure to discriminate against RNA would likely recapitulate deletion phenotypes and their catastrophic effects on genome stability. Accomplishing the necessary discrimination to prevent this is a challenging prospect as ssDNA and ssRNA are chemically similar, adopt similar conformational structures, and are both highly flexible.

Consistent with the biochemical prediction that Pot1 must strongly disfavor RNA binding, experimental results demonstrate that the full-length DBD of *S. pombe* Pot1 indeed discriminates against RNA of the same cognate sequence by at least a factor 10^6 (Altschuler, 2011). This discrimination is conferred in part by the specificity determining first OB-fold, Pot1pN, which alone disfavors RNA by a factor of >200 using interactions at two positions, primarily through the loss of a methyl-hydrophobic interaction (Lei et al., 2003). However, to explain the discrimination observed by the full-length protein, Pot1pC must also contribute to the discrimination against RNA. The observed flexibility in ssDNA base recognition raises the question how Pot1pC achieves this specificity for ssDNA, as the differences between RNA and DNA are more chemically subtle than base replacement.

To resolve how Pot1pC discriminates against ssRNA despite the remarkable structural plasticity of its interface, we characterized the ribose-position specificity of Pot1pC by measuring binding affinities of RNA and chimeric RNA-DNA ligands containing ribose nucleotide substitutions in the cognate sequence and found that specificity for DNA over RNA is not evenly distributed across the ligand. We also solved three high-resolution crystal structures of Pot1pC bound to these chimeric RNA-DNA ligands, revealing a widely utilized cryptic binding mode of Pot1pC characterized by the rearrangement of T4 into an alternative binding pocket and substantial rearrangement of the 3′ portion of the interface. These rearrangements allow full thermodynamic accommodation of RNA nucleotide substitutions at positions near the 3′ end of the ligand, facilitated by a long and flexible protein loop, but not fully at the 5′ end due to suboptimal binding conformations for RNA ligands.

RESULTS

Pot1pC Discriminates against RNA Additively by Ribose Position

Pot1pC minimally recognizes a 9 nucleotide-long sequence (9mer) of ssDNA of the sequence GGTTACGGT with an apparent binding dissociation constant (K_D) of 24 nM (Dickey et al., 2013). Full substitution of this cognate ssDNA sequence with ribose nucleotides (1-9R) results in a substantial loss of affinity by 80-fold as measured by isothermal titration calorimetry (Table 1), demonstrating that both the Pot1pN and Pot1pC subdomains of Pot1-DBD contribute to RNA discrimination. One possibility is that this discrimination is achieved through recognition of the specific chemical differences between DNA and RNA of the same sequence. However, in the ssDNA complex, the methyl groups of the three thymine bases of the cognate sequence, as well as most of the ribose 2′ hydroxyl positions, are primarily solvent exposed (Figure S1), with little

Table 2. Data Collection and Refinement Statistics for Ribose Chimeric Pot1pC Complexes

	1R rGGTTACGGT	1-3R rGrGrUTACGGT	7-9R GGTTACrGrGrU
RCSB PDB ID	5USB	5USN	5USO
Data Collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a, b, c (Å)	41.29, 58.01, 65.91	41.64, 59.8, 66.08	44.56, 57.61, 66.76
α, β, γ (°)	90 90 90	90 90 90	90 90 90
Resolution (Å)	33.64–1.615 (1.673–1.615)	44.339–1.9 (1.968–1.9)	43.62–2.0 (2.072–2.0)
R_{merge}	0.064	0.131	0.163
I/σ	119.88 (6.00)	101.80 (10.01)	131.41 (6.77)
Completeness (%)	96.58 (90.87)	98.97 (96.98)	94.39 (92.36)
Redundancy	7.4 (5.5)	12.7 (12.9)	22.1 (17.8)
Refinement			
Resolution	33.64–1.615 (1.673–1.615)	44.339–1.9 (1.968–1.9)	43.62–2.0 (2.072–2.0)
No. of reflections	20,220 (1,852)	13,388 (1,284)	12,121 (1,177)
R_{work}/R_{free}	0.1759/0.2027	0.1829/0.2231	0.2102/0.2457
No. of atoms	1,697	1,557	1,495
Protein	1,300	1,194	1,188
Ligand/ion	186	187/1	187
Water	211	175	120
B factors			
Protein	21.61	21.31	31.86
Ligand/ion	29.09	31.69/35.12	40.58
Water	34.28	30.17	36.5
RMSD			
Bond lengths (Å)	0.016	0.005	0.009
Bond angles (°)	1.49	0.66	0.99
Crystallization Conditions	50 mM Tris, 0.2 mM sodium formate, 20% PEG 8K	100 mM Tris pH 8.4, 0.2 mM sodium formate, 15% PEG 4K	100 mM Tris pH 7.5, 0.2 mM sodium formate, 15% PEG 4K

RNA nucleotides in are underlined. Each structure determined by one crystal. Highest-resolution shell in parentheses. RMSD, root-mean-square deviation.

to no interaction with the protein. In this context and given the multiple binding modes observed for non-cognate ssDNA ligands, predicting how Pot1pC discriminates against RNA cannot be confidently discerned from the ssDNA complex structure.

To resolve how the differences between ssRNA and ssDNA contribute to Pot1pC RNA discrimination, we probed affinity changes as a function of nucleotide position by using chimeric DNA/RNA ligands containing combinations of deoxyribose and ribose nucleotides, first by nucleotide triplets and then by individual nucleotides for the triplets that exhibited significant discrimination. In addition, the protein specificity for thymine methyl groups was examined through deoxyuridine substitutions. Binding with ligands that group ribose substitutions in triplets at the first three nucleotides (1-3R), fourth through sixth (4-6R) nucleotides, and the last three nucleotides (7-9R) of the cognate ssDNA sequence reveal that Pot1pC discriminates against ssRNA primarily at the first 6 nucleotides, with modest binding reductions observed in the 1-3R and 4-6R ligands and little to no affinity change for the 7-9R ligand (Table 1). Together, the sum of the energetic differences between the cognate ligand and the triplets is consistent with the energetic loss for 1-9R, suggesting that the binding perturbations are additive and not cooperative.

Individually, the nucleotide position that shows the most significant impact on binding affinity with the addition of the ribose 2′ hydroxyl is G1 (1R). Substitution of this position with rG leads to a ~4-fold reduction in affinity (Table 1). In contrast to the methyl specificity exhibited by Pot1pN (Lei et al., 2003), neither full substitution of the cognate sequence with uracil (T3-4-9dU) nor the ligands containing dT to rU substitutions (positions 3, 4, and 9; tested by ligands 3R, 4R, and 7-9R, respectively) significantly impact affinity, suggesting the addition of hydroxyl groups on the ribose moiety are primarily responsible for the loss of Pot1pC affinity. In the structural context of the cognate ssDNA Pot1pC complex, this is not unexpected as these methyl groups are primarily solvent exposed except for intramolecular base stacking between T3 and T4 and a limited contribution to the aromatic stack between T9 and Trp27 and Tyr28 (Figure S1). Because dU and rU substitutions did not exhibit significant affinity loss at these positions, binding with rT substitutions was not tested.

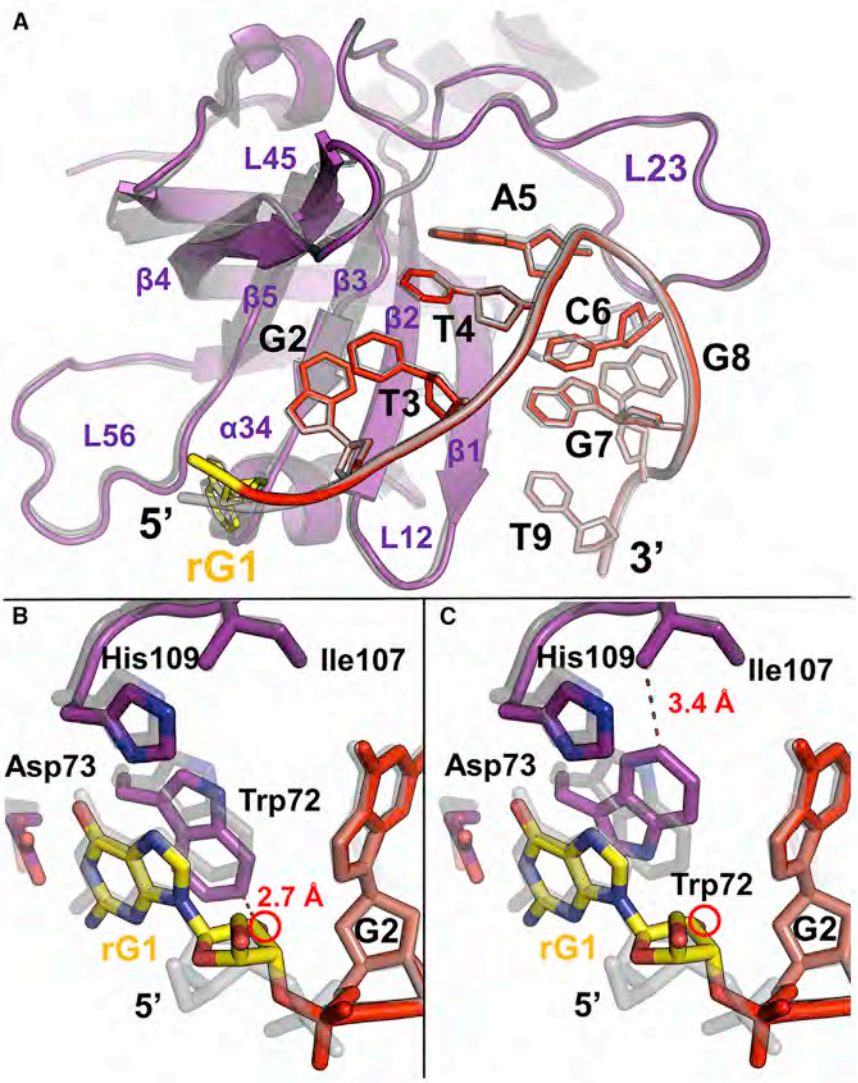


Figure 1. An Unfavorable Interaction between rG1 Hydroxyl, Trp72, and G2 Base is the Strongest Individual Discrimination Determinant

(A) 1R bound Pot1pC (PDB: 5USB) shows high similarity to cognate Pot1pC (PDB: 4HIK). Overlay shown for cognate (DNA; white) Pot1pC (gray) complex and 1R (red, rG1 substitution yellow) Pot1pC (purple) complex.

(B and C) Enlarged view of the rG1 binding site reveals most cognate binding features are maintained in the 1R complex. However, the rG1 2′ hydroxyl forces Trp72 into two alternative conformations, shown separately in (B and C) for clarity, with unfavorably close contact with either the 2′ hydroxyl of rG1 (B) or Ile107 (C). The 2′ hydroxyl is highlighted by a red circle in (B and C).

A Suboptimal Binding Geometry between rG1 hydroxyl, Trp72, and the G2 Base Is the Strongest Individual Discrimination Determinant

Following the lessons learned from the ssDNA-Pot1pC structures, whereby unexpected interactions were formed upon base substitution, we sought to resolve the underlying structural mechanisms responsible for Pot1pC 2′ hydroxyl discrimination by solving high-resolution structures of Pot1pC bound to chimeric ligand species. Diffraction-quality crystals of the 1-9R bound complex remained elusive, likely due to its weak affinity. Instead, noting the additivity observed in the thermodynamics of the chimeric ligands, we solved chimeric complexes where individual or groups of sites in the ssDNA were replaced with their ribose equivalent (solved structures and statistics summarized in Table 2). For individual site replacement, we targeted the position that displays the most discrimination (1R, which contains a dG to rG substitution at position 1). Using conditions established previously, we were able to obtain diffraction-quality crystals and solved this structure (PDB: 5USB) to 1.62 Å resolution, R_{work}/R_{free} (0.1759/0.2027) using molecular replacement with the cognate bound Pot1pC structure (PDB: 4HIK) (Dickey et al., 2013). The structure of the 1R complex overlays closely to the cognate structure aligning with a root-mean-square deviation (RMSD) of 1.16 Å (protein) and 0.75 Å (ligand) (Figure 1A).

Curiously, despite binding with a ~4-fold weaker affinity, the 1R complex maintains all the hydrogen bond and stacking interactions observed in the cognate structure (Figure 1B). Closer examination of the 2′ hydroxyl suggests that the loss of affinity is the result of an unfavorably close contact between the ribose moiety of rG1 and the cognate positioning of Trp72 (Figure 1B/1C). While the conformation of the ribose is somewhat ambiguous due to weaker electron density and higher B factor than the neighboring groups, its relative position is constrained by the strong electron density for the base and phosphate moieties (Figure S3A). Thus, favorable ribose conformations place the 2′ hydroxyl (as modeled) or alternatively the ring O4′ oxygen (not shown) into an unfavorably close contact to Trp72 (2.7 or 2.3 Å between heavy atoms, respectively; Figure 1B) and close enough for the 2′ hydroxyl to form a hydrogen bond with the ribose ring oxygen of G2. Careful examination of the electron density of Trp72 (Figure S3A) suggests that this residue partially alleviates this steric clash by adopting another conformation (Figure 1C). However, the alternative conformation also has an unfavorably close contact to Ile107 (3.4 Å between heavy atoms), which is likely why Trp72 adopts a conformation that is positioned to clash with a 2′ hydroxyl at position 1 (Figure 1B) in the cognate ligand bound structure. Together, these suboptimal conformations enforced by the rG1 base and phosphate likely comprise the mechanism of discrimination for this ligand, as it can only be bound in a non-ideal configuration resulting in the observed loss in affinity. Although the rG1 substitution can force accommodation, it does not recapitulate the full binding energy of the

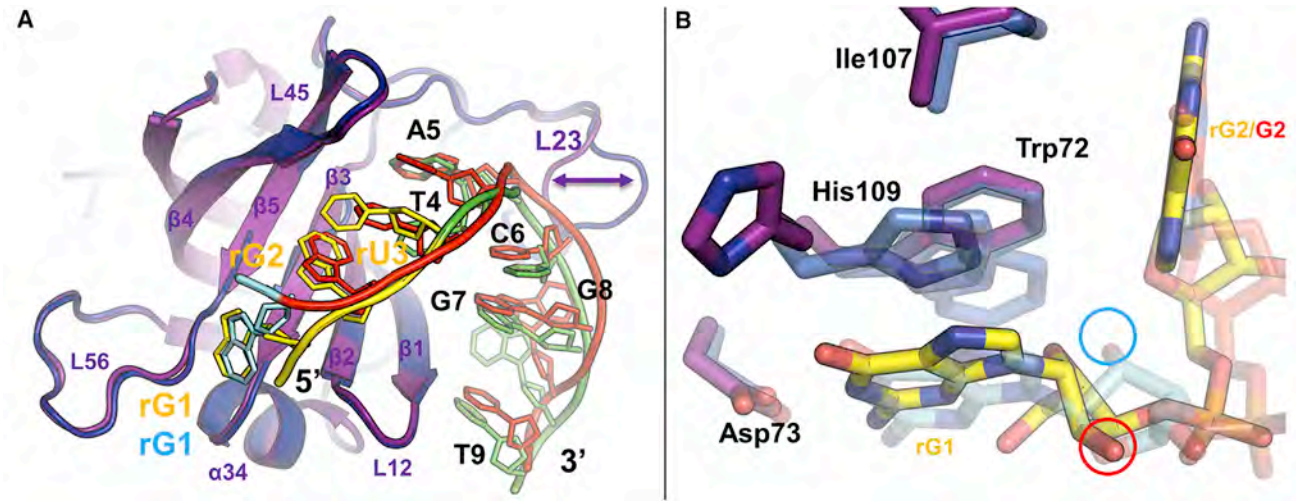


Figure 2. 1-3R Binds Pot1pC in an Alternative Binding Mode, Recapitulates the Unfavorable Interactions Seen in the 1R Structure, with the Additional Loss of the His109-G1 Interaction

(A) 1-3R bound Pot1pC (PDB: 5USN) shows significant conformational changes compared with 1R bound Pot1pC (PDB: 5USB). Overlay shown for the 1R (red, 1R substitution cyan) bound Pot1pC (blue) compared with the 1-3R (green, 1-3R substitutions yellow) bound Pot1pC (purple). The conformational change of L23 is highlighted by a purple arrow.

(B) Comparison of the 1R and 1-3R G1 binding sites reveals a shared steric clash between Trp72 and Ile107 to accommodate the rG1 2′ hydroxyl despite different ribose conformation models (2′ hydroxyls highlighted with circles; 1R cyan, 1-3R red). In addition, the 1-3R complex loses the interaction between His109 and G1.

cognate DNA due to unfavorable interactions between Trp72 and the ribose moiety or the neighboring Ile107.

Pot1pC Plasticity Partially, but Not Fully, Compensates for Lost Interactions in Presence of the 2′ Hydroxyl at Positions 1–3

To better understand the above structure and expand our understanding to additional positions, we solved the structure of the first triplet complex (1-3R; d(GGT) to r(GGU); PBD: 5USN, 1.9 Å, R_{work}/R_{free}: 0.1829/0.2231). This complex shows substantial rearrangement of the protein and dramatic rearrangement of the ligand relative to the 1R (1.53 Å protein, 2.34 Å ligand RMSD; Figure 2A) and cognate structures (1.55 Å protein, 2.39 Å ligand RMSD; Figure 3A). The conformational changes are especially significant in the 3′ portion of the ligand and β2-β3 loop (L23), a region distant from the site of modification, suggesting that, despite the propagation of conformational changes in the ligand backbone, the protein interface is able to make similarly large adjustments to compensate.

In this structure, relative to both the cognate and 1R structures, the rG1 ribose (Figure 2B) is flipped, roughly swapping the positions of the 2′ hydroxyl and 5′ carbons. Without further reorganization, this ribose conformation would have put the ring O4′ oxygen within a predicted 2.4 Å of the cognate conformation of Trp72 (Figure 3B) based on structural alignments. Presumably to avoid this unfavorable contact, the conformation of Trp72 rotates roughly 180°, placing it into close proximity to Ile107 and forming a hydrogen bond between the rG1 ring O4′ oxygen and the indole nitrogen of Trp72.

The 1-3R structure contains additional changes near G1 (Figure 2B) relative to those observed in the 1R structure. The 1-3R rG1 maintains the hydrogen bonds to the amide backbone of Gly110 and the side chain of Asp73, but, relative to the 1R and cognate DNA complexes (Figures 2B and 3B), His109 rotates ~120° out the binding pocket and no longer forms the hydrogen bond with N7 of G1. Thus, the 1-3R complex supports the steric clash mechanism observed in the 1R complex and potentiates the discrimination at rG1 through the lost interaction with His109.

The 1-3R ligand dramatically diverges in conformation far beyond the sites of substitution, with conformational changes encompassing the entire length of the nucleic acid (Figure 3A). The disruptions begin near the site of substitution and appear to be due to a chain of conformational changes that occur to alleviate close contacts. The shift in the positions of rG2 and rU3 disrupts a cognate G2-Lys97 interaction (Figure 3B). The combination of a predicted close contact between the rU3 2′ hydroxyl and the cognate position T4 and the conformational change in the sugar-phosphate backbone results in T4 swinging into an alternative binding pocket (Figure 3C) and disrupting most T4-protein interactions. In this pocket, T4 clashes with the cognate position of C6, forcing it to shift position (Figure 3C), while A5 maintains its cognate interactions and binding pocket. Rearrangement of the C6 base positions the G7 phosphate into the cognate position of G8 (Figure 3D). The rearrangement of G7 is accommodated by shifts in L23 residues Ser56 and Arg57, while disrupted G8-protein interactions are compensated by G8 interactions with Thr26 and Glu85. The changes in C6 and G8 also shift G7 and T9 downward (Figure 3D), but these changes are accommodated by a corresponding movement of Trp27 and Tyr28, which stack with G7 and T9. A summary of the ligand-protein interactions for cognate and the 1-3R complexes are summarized schematically in Figures 4A and 4B, respectively. Notably, similar structural changes for positions 6–9 and L23 are seen in several other Pot1pC complexes that exhibit no binding defect, suggesting that the unfavorable

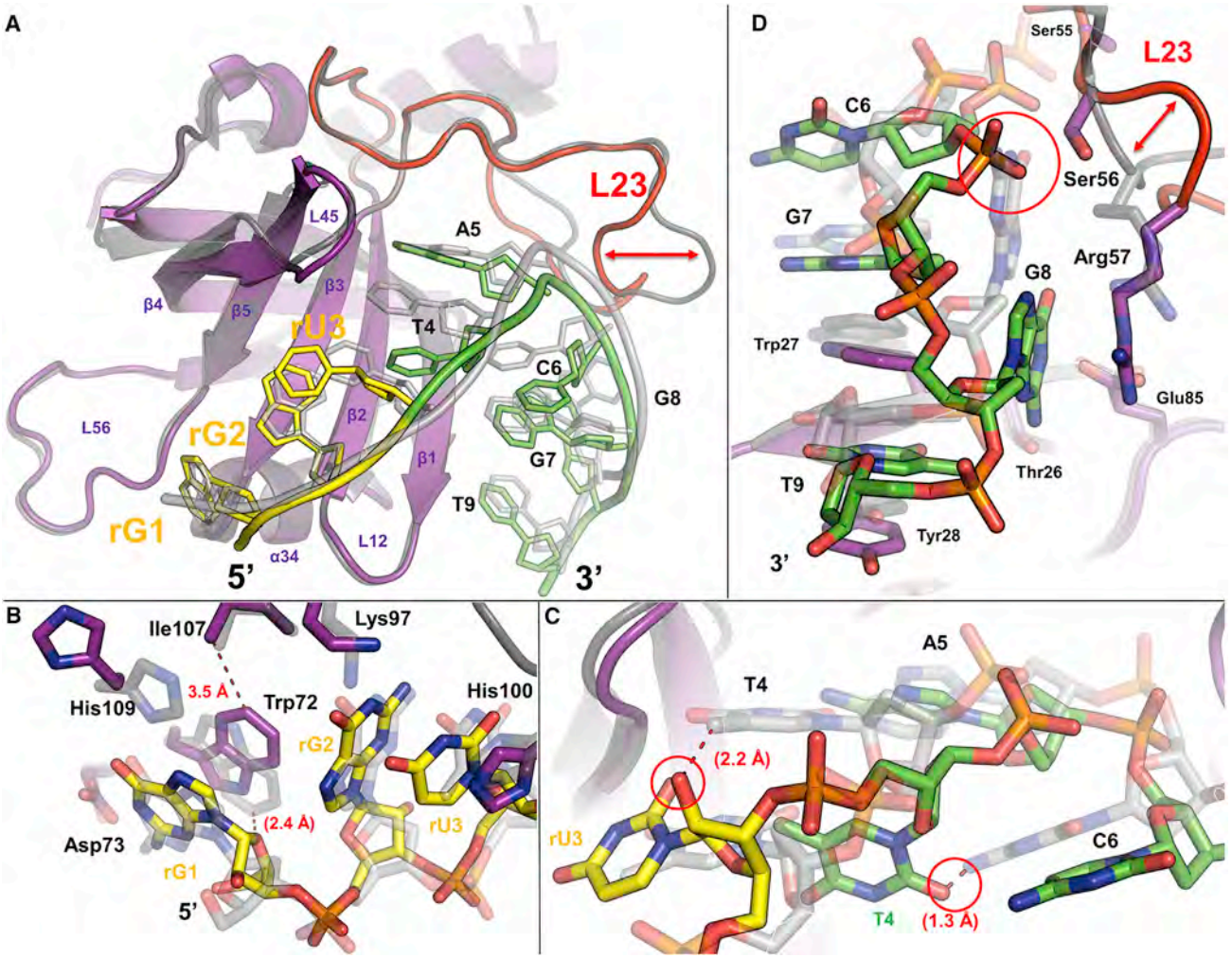


Figure 3. Structural Rearrangement of the 1-3R Ligand Driven by Conformational Changes in the Sugar Phosphate Backbone and Alleviation of Steric Clashes at the Bases

(A) 1-3R bound Pot1pC (PDB: 5USN) reveals significant conformational changes compared with the cognate bound Pot1pC (PDB: 4HIK). Overlay shown for the cognate DNA (white) bound Pot1pC (gray) compared with the 1-3R (green, 1-3R substitutions yellow) bound Pot1pC (purple). The shift of the protein backbone near the 3′ portion of the ligand (red arrow) illustrates the conformational plasticity of L23 (red).

(B) Comparison of the 5′ portion of the 1-3R complex (1-3R substitutions yellow; Pot1pC, purple) with the cognate Pot1pC (DNA, white; Pot1pC, gray) reveals several lost interactions and an unfavorable steric clash. Predicted distances between cognate and 1-3R complexes are in red parentheses, observed distance in red without parentheses.

(C) Comparison of nucleotides 3–6 of the 1-3R ligand (ligand green, substitutions yellow) to cognate reveals substantial rearrangement of non-substituted nucleotides. A would-be steric clash between rU3 2′ hydroxyl (red circle) and T4 and rearrangement of the ligand backbone shifts T4 into an alternative binding pocket. This position clashes with the cognate position of C6 (red circle) and results in another nucleotide shift. Predicted distances between cognate and 1-3R complexes are in red in parentheses.

(D) Comparison of the 3′ portion of the 1-3R ligand reveals additional changes compared with the cognate complex. The shift of C6 positions the phosphate of G7 into the cognate position of G8 (red circle). G8 flips and L23 residues shift to accommodate the positions of the phosphate groups, and the G8 base forms hydrogen bonds with Glu85 and Thr26. Corresponding shifts of Trp27 and Tyr28 maintains the cognate stacking interactions. The L23 conformational change is highlighted with a red arrow.

interactions at rG1 and lost interactions for rG1 and T4 are the driving mechanisms behind the 1-3R ligand discrimination.

Cryptic Secondary Binding Mode Is Widely Used to Provide Partial Thermodynamic Compensation

Even though the interface is quite different than in the cognate structure, the binding mode of the 1-3R structure shows striking similarity to the structural changes observed in two completely different non-cognate complexes, the T4A Pot1pC complex and G2C (1-3R versus T4A: 0.86 Å protein, 1.30 Å ligand; 1-3R versus G2C: 0.88 Å protein, 1.83 Å ligand; 1-3R versus cognate:1.55 Å protein, 2.39 Å ligand RMSD) (Dickey et al., 2013). These structures were characterized by the base at position 4 flipping 55° away from the β4-β5 loop (L45) and toward the β barrel. In the 1-3R complex, T4 rotates into the same binding pocket observed for the adenine of the T4A DNA substitution

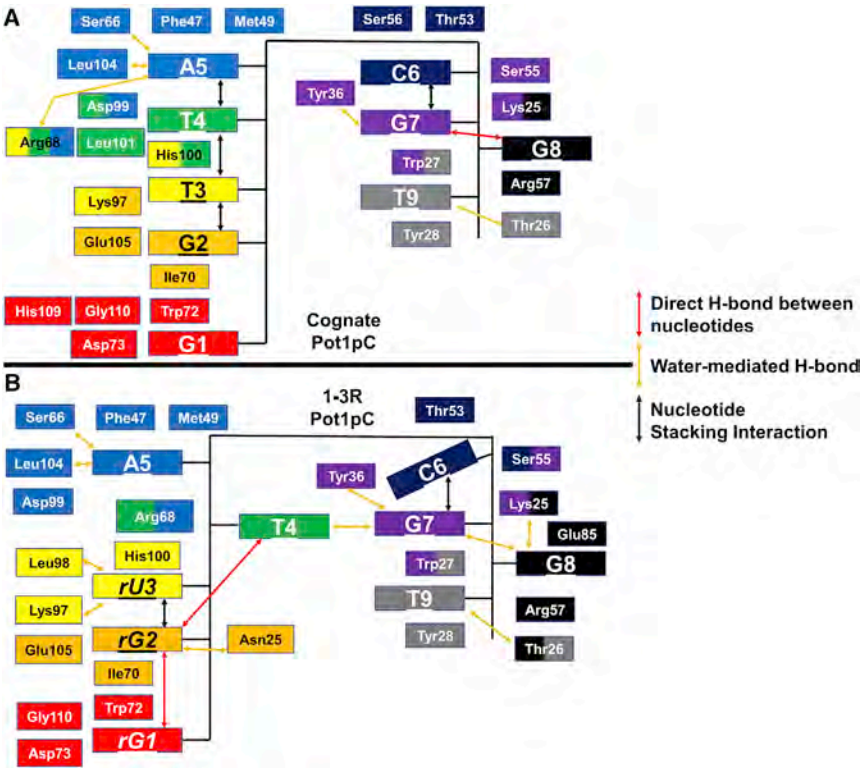


Figure 4. Rearrangement of the 1-3R Ligand Results in a Mix of Lost and Compensatory Interactions

(A and B) Schematic summarizing the cognate (A) and the 1-3R ligand-protein interactions (B). Nucleotide interacting residues are color coded by interaction partner. Multicolored residues indicate interactions with more than one nucleotide. Red arrows indicate hydrogen bonds between nucleotides; orange arrows, a water-mediated interaction; and black arrows, stacking interactions.

(Figure 5A and 5B) as well as T4 in the G2C complex (Figure 5C). Like 1-3R, both T4A and G2C diverge in ligand conformation relative to the cognate ligand well beyond the sites of substitution. The large rearrangement of the 3′ portion of the ligand likely results from an overlap in the cognate position of C6 and the alternative position of nucleotide 4. While the T4A DNA binding mode has an equivalent affinity to the cognate DNA ligand, both the 1-3R and G2C complexes have reduced affinity despite ostensibly binding in the same binding mode (1-3R, 9-fold; G2C, 36-fold). The common difference is that these ligand complexes flip T4 into the adenine binding pocket of T4A. The use of the T4A adenine binding pocket appears to not fully compensate for the interactions lost upon T4 binding in this pocket as thymine is not large enough to reach the β barrel residues in this pocket-like T4A, resulting in the observed net loss of binding affinity due to suboptimal complex geometries.

Backbone Alterations at the 3′ End of the Ligand Are Readily Accommodated by Ligand and L23 Structural Rearrangement

In contrast to the thermodynamic consequences of introducing riboses at positions 1–6, ribose incorporation at positions 7–9 has no affect on binding affinity. To address how this full accommodation is achieved, we solved the crystal structure (7-9R; d(GGT) to r(GGU); PBD: 5USO, 2.0 Å, R_{work}/R_{free}: 0.2102/0.2457). Instead of simple non-specific recognition of the backbone at these positions, the crystal structure of 7-9R reveals rearrangement of the ligand for nucleotides 6–9 compared with the cognate structure (1.31 Å protein, 1.63 Å ligand RMSD; Figure 6A). As expected from the cognate structure in which the simple addition of the 2′ hydroxyl at G7 would clash with the base position of G8, the sugar orientation of rG7 shifts such that the 2′ hydroxyl is instead solvent exposed (Figure 6B). This rearrangement of the sugar-phosphate backbone has consequences elsewhere in the ligand, altering the orientation of C6 as well as the rG7 phosphate resulting in the phosphate clashing with the cognate position of the G8 base. In turn, the rG8 base rotates 180° around the glycosidic bond, removing the intramolecular hydrogen bond between G8 and the G7 phosphate. As seen in the 1-3R complex, the L23 residues Ser56 and Arg57 rearrange to accommodate these shifts while the side chains of Thr26 and Glu85 are statically poised to interact with the alternative position of rG8. In addition, the rG8/rU9 backbone shifts downward relative to the cognate backbone. Despite these changes, the rU9 base is readily accommodated in essentially the same relative positioning as the cognate T9 hydrophobic stack between Trp27 and Tyr28 (Figure 6B). In large part, the flexibility of the sugar phosphate backbone and L23 facilitates this accommodation and contributes to the plasticity of the interface. The solution flexibility of this interface may be even greater than that seen in these crystal structures due to the contacts between the ligand ribose-phosphate backbone and neighboring symmetry mates which constrain the observable flexibility by crystallography.

Model for 4-6R and 1-9R Pot1pC Complex Structures

The structures we have solved allows us to develop a model for how 1-9R binds. Comparison of the ligand conformations for the 1-3R and 7-9R complexes reveals that the binding mode of nucleotides 5–9 is highly similar (Figure 7). This apparent compatibility of these binding modes, which also share many features with the binding mode of the T4A and G2C DNA ligands, suggests that 1-9R may also bind in this major binding mode. Based on the overlap of positions 5 and 6 in all Pot1pC structures, the speculation that they occupy similar conformations as the 1-3R and 7-9R structures would be consistent with the apparent energetic additivity for the triplets compared with 1-9R. If this were the case, we can speculate on the mechanisms of discrimination for the 4-6R ligand based on the ligand overlap of these complexes. At nucleotide rU4, the 2′ hydroxyl would clash with the rA5 position in the 1-3R binding position while being solvent

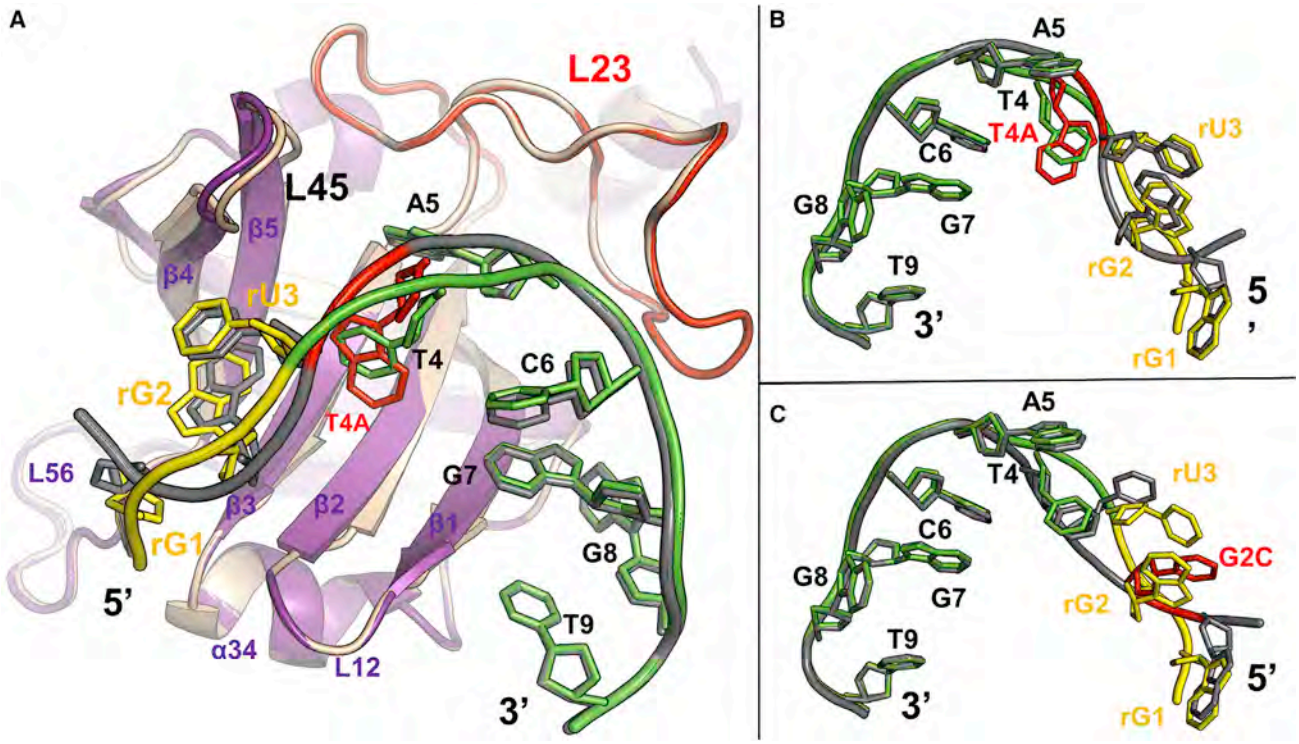


Figure 5. 1-3R Binds Pot1pC More Like the T4A and G2C DNA Ligands than the Cognate DNA Ligand

(A) Comparison of the 1-3R (PDB: 5USN) and T4A (PDB: 4HIO) complexes reveals high similarity. Overlay shown for 1-3R bound Pot1pC (1-3R green, substitutions yellow; Pot1pC, purple) and T4A bound Pot1pC (ligand gray, T4A substitution red). T4 of 1-3R occupies the same binding pocket as A4 of T4A, L23 (red).

(B) Overlay of 1-3R and T4A ligands shown rotated ~180° relative to (A) shows high similarity between the T4A and 1-3R binding modes, with greater agreement in the 3′ portion of the ligands.

(C) Overlay of the 1-3R and G2C (PDB: 4HID, gray, G2C substitution red) ligands shows high similarity in the G2C and 1-3R binding modes.

exposed in the cognate/7-9R T4 position. For position 5, the rA5 2′ hydroxyl would unfavorably occupy a hydrophobic pocket created by Phe47 and Met49 (Figure 7). In addition, rC6 presents an apparent steric clash between the would-be 2′ hydroxyl and the base of G7 in the 1-3R binding conformation (Figure 7) but could form a potential hydrogen bond with Lys25 in the 7-9R conformation. While this speculation for the 4-6R mechanisms of discrimination should be taken with caution in the absence of definitive structural information for these substitutions, features of L23 suggest that the plasticity used to bind the 3′ portion of the ligand does not extend to Phe47 and Met49. The neighboring residue Phe46 is buried in the hydrophobic core of the protein and nearby prolines, Pro48 and Pro51, severely limit the backbone conformations favorable for Phe47 and Met49, suggesting accommodation is likely to resemble the suboptimal geometry of Trp72 rather than wholesale rearrangement of Arg57 and Ser56.

DISCUSSION

S. pombe Pot1 has the challenge of recognizing an inherently degenerate telomere sequence, $G_{2\text{-}8}$TACGGT(A) (Leonardi et al., 2008; Lee et al., 2014), with both high specificity and affinity. In doing so, it must discriminate against ssDNA with similar sequence as well as the much more abundant RNA ligands containing the identical sequence. *Sp*Pot1 accomplishes this DNA specificity by having evolved a modular DBD comprising a sequence-specific binding domain, Pot1pN (Lei et al., 2003), and a non-specific binding domain, Pot1pC (Trujillo et al., 2005; Altschuler et al., 2011; Dickey et al., 2013), which contribute equally to the full-length binding affinity. Surprisingly, though, both the sequence-specific and sequence-non-specific domains discriminate against the telomeric sequence rendered in RNA (Lei et al., 2003; Altschuler, 2011). An *a priori* rationale to explain the full extent of how Pot1pC achieves this is difficult to formulate as the bulk of the interaction between Pot1pC and its ligands are primarily base-mediated hydrogen bond interactions and comparatively few interactions with the sugar-phosphate backbone (Dickey et al., 2013). Moreover, this protein/nucleic acid interface exhibits remarkable plasticity capable of both subtle and dramatic structural rearrangements of the protein and ligand to form thermodynamically equivalent complexes. At first glance, the three thymine bases in the cognate ligand could suggest that uridine substitution may explain ssRNA discrimination, but the methyl groups only interact in limited aromatic stacking interactions and are partially solvent exposed (Figure S1D). Likewise, the 2′ hydroxyl groups are mostly solvent exposed in the cognate and non-cognate structures (Figures S1A–S1C). Therefore, neither the substitution of thymines to uridine nor the addition of 2′ hydroxyl groups into the cognate structural conformation provides a satisfactory explanation for the 80-fold reduction in affinity observed for ssRNA.

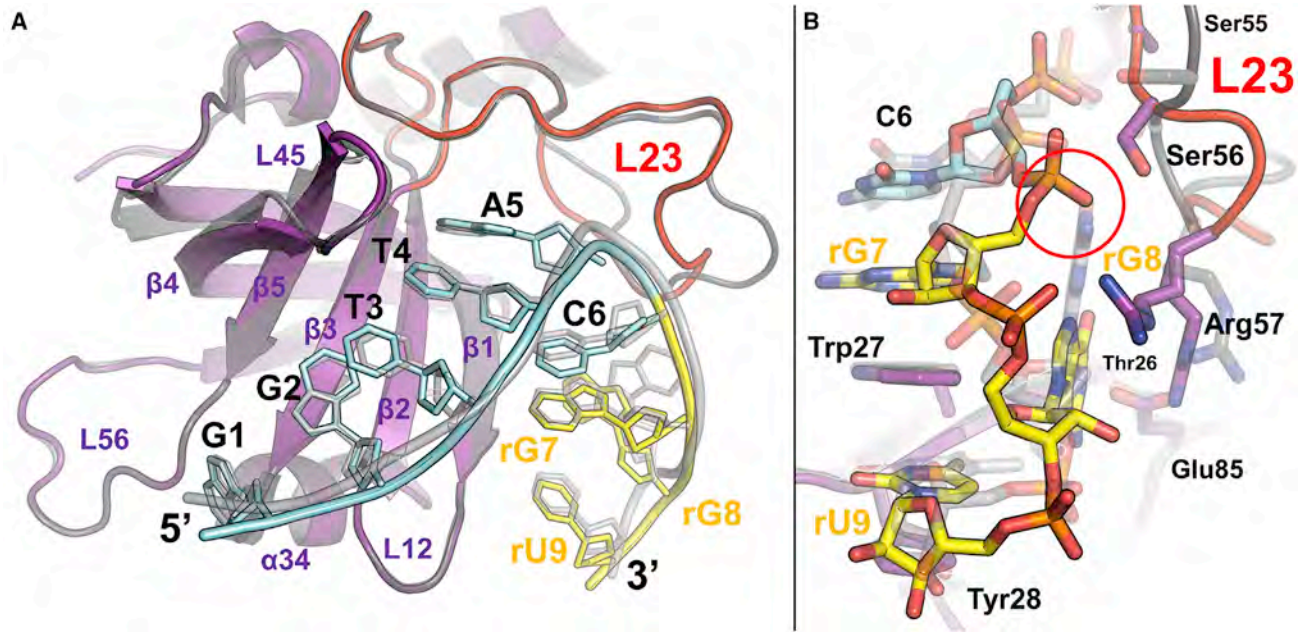


Figure 6. 7-9R Binding Causes Rearrangement of L23 to Form Thermodynamically Compensatory Interactions in Response to Ligand Backbone Conformational Changes

(A) 7-9R bound Pot1pC (PBD: 5USO) reveals rearrangement of L23 and G8 in response to substitutions relative to cognate (PDB: 4HIK). Overlay shown for the cognate DNA (white) bound Pot1pC (gray) compared with the 7-9R (cyan, 7-9R substitutions yellow) Pot1pC (purple).

(B) Comparison of the binding pocket of the 7-9R substitutions (7-9R ligand cyan, substitutions yellow; Pot1pC, purple) with the cognate DNA ligand (white; Pot1pC, gray). Despite all 2′ hydroxyls being solvent exposed, rearrangement of the sugar phosphate backbone forces a shift of rG8 due to rG7 phosphate moving into the cognate position of G8 (red circle). Rearrangement of the L23 (red) residues accommodate the shift in ligand backbone, while Glu85 and Thr25 form hydrogen bonds with the flipped G8.

Our set of chimeric ssRNA-ssDNA Pot1pC complex structures allows us to identify the features of the Pot1pC binding interface that provide specificity for ssDNA over ssRNA while binding ssDNA with a surprising level of non-specificity. Overall, the underlying Pot1pC specificity for DNA ligands appears to result from forcing RNA ligands into suboptimal binding geometries in regions of the protein with less conformational flexibility. In these regions of the protein, which are responsible for interacting with the first five nucleotides, base substitutions are accommodated by residues poised for compensatory interactions with alternative bases, but unlike L23 are limited primarily to rotameric rearrangements of side chains. The 1R and 1-3R structures reveal that the first nucleotide of the 9mer ligand is responsible for the largest individual discrimination observed for Pot1pC through an unfavorable interaction between the ribose moiety and residues on the protein surface (Figures 1B and 1C). Speculatively, 4-6R also places the 2′ hydroxyl of rA5 in a similar clash based on the ligand conformations seen in all solved Pot1pC complexes (Figure 7). In addition, differential ligand flexibility appears to underlie the mechanism of discrimination at other positions. While ssDNA and ssRNA are both highly flexible ligands and ssDNA can generally adopt the same conformations as ssRNA, the reverse is not strictly true—the presence of the 2′ hydroxyls in RNA change the energy landscape favorable for different backbone conformations. In the case of the 1-3R structure, the sugar orientation of nucleotide 3 switches from the cognate binding mode to that seen in the T4A and G2C binding modes with the accompanying rearrangement of T4 into a non-compensatory binding pocket, whereby the side chains that interact with the T4A substitution are unable to reach the smaller thymine. The 7-9R structure shows a similar structural impact of differential flexibility but exemplifies the difficulty of predicting the biochemical impact of ribose nucleotides at specific positions based on the structure of the Pot1pC cognate complex. The cognate orientation of the G7 ribose suggests that a steric clash with the G8 base would result in the loss of binding affinity. However, this clash is alleviated through rearrangement of the sugar-phosphate backbone, rotation of the G8 base about the glycosidic bond, and concomitant changes in loop conformation which underscore the ability of the protein to exhibit a flexibility as dramatic as the more obviously flexible ligand. The features of the rearrangement in the 3′ portion of the 7-9R ligand and L23 are largely recapitulated in the 1-3R, G2C, and T4A complexes. Our ability to solve a range of structures has revealed that, rather than adopting a mashup of many conformations, Pot1pC has a widely utilized alternative binding mode that is employed partially or in full in response to myriad chemical modifications.

Structures of related ssDNA-binding proteins reveal other mechanisms utilized for RNA discrimination. In the case of mammalian Pot1 (mPot1), the underlying mechanism of discrimination occurs in the region of the protein homologous to *S. pombe* Pot1pN, as the C-terminal domain of the human protein, analogous to Pot1pC, barely engages with the ligand (Lei et al., 2004). Discrimination in this case appears to result primarily from losing a hydrophobic interaction from a thymine methyl

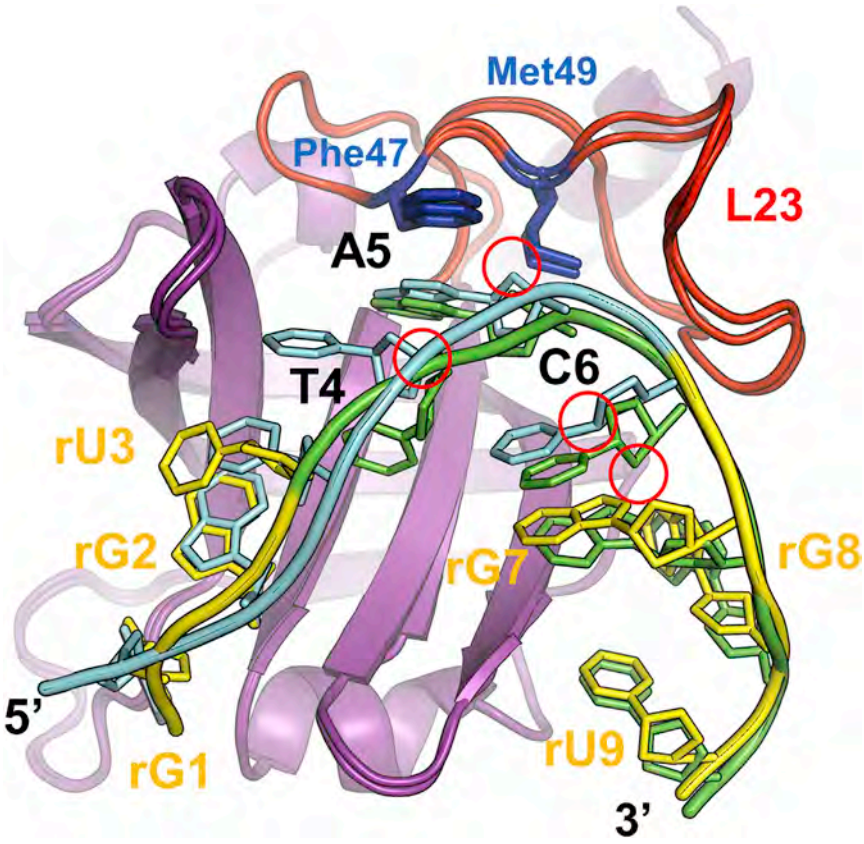


Figure 7. Comparison of Sugar rG1 Sugar Conformations in 1R and 1-3R Complexes and Model for Discrimination at Positions 4–6

Comparison of the 1-3R (PDB: 5USN) and 7-9R (PDB: 5USO) complexes (all substitutions yellow; Pot1pC purple, 1-3R ligand green, and 7-9R ligand cyan) suggest compatibility between the binding modes of these complexes. Both L23 (red) and the 3′ portion of the ligands agree reasonably well, suggesting the full 1-9R complex may bind in a similar binding mode. The predicted positions of the 2′ hydroxyl of positions 4–6 are highlighted with red circles for this binding mode. The U4 and A5 2′ hydroxyls are both poised to force a steric clash in the 1-3R binding mode. U4 with the A5 base while A5 is positioned in a clash with Phe47 and Met49 (blue). The C6 2′ hydroxyl shows more ambiguity of position with a close contact with G7 in the 1-3R complex, but is positioned to form a potential hydrogen bond with Lys25 in the 7-9R complex.

as well as forcing the 2′ hydroxyl at that position into a sterically unfavorable interaction in that same hydrophobic pocket (Nandakumar et al., 2010) in a manner similar to our speculated mechanism of discrimination at A5. However, other positions throughout the ligand also contributed to ssDNA specificity, especially when placed near the discriminating position—suggesting that nearby ribose nucleotides reinforce the suboptimal binding geometries at discriminating positions by further limiting the conformational flexibility of the ligand (Nandakumar et al., 2010). While an RNA bound or chimeric complex for Pot1pN has not been solved, predictions for the mechanisms of specificity can be made based on the available structures and existing biochemical data. If the RNA adopted the same conformation as the DNA bound complex, there would be readily apparent steric clashes for hydroxyls at two positions similar to the 2′ hydroxyl clash with Trp72 at rG1 and energetically unfavorable hydrophobic pockets left empty by the loss of thymine methyl groups (Lei et al., 2003). At both positions (GG**TT**AC, cognate), a greater than 200-fold discrimination is observed at T3 attributable entirely to the loss of the methyl group, while T4 exhibits ~100-fold affinity reduction due to the methyl and ~7-fold reduction due to the 2′ hydroxyl (Lei et al., 2003). The presumed 2′ hydroxyl steric clash at Pot1pN's T4 results in a binding defect in line with our observations for the 1R and 4-6R substitutions, whereas the loss of methyl interactions has a much greater effect for *S. pombe* Pot1pN but not mPot1. *S. pombe* Pot1pN more strongly engages with the methyl groups of the thymines, with empty space left behind in hydrophobic pockets in their absence, whereas mPot1 and Pot1pC have far fewer close contacts to the methyl groups of their ligands. Together, these suggest that strong engagement with the methyl groups of thymines in hydrophobic pockets provides the strongest mechanism for favoring ssDNA over ssRNA, while discrimination against 2′ hydroxyls appears to primarily result from steric clashes with the ribose moiety or forcing suboptimal binding geometries resulting from rearrangement of the sugar-phosphate backbone in response to the differential conformational flexibility of ribose and deoxyribose moieties.

Radical conformational adjustment to confer non-specific binding has been observed in other systems. The atypical Puf domain of Puf5 allows for the specific binding of RNA sequences of variable length (8–12) through rearrangement of the ligands, although without the concomitant rearrangement of the protein (Wilinski et al., 2015). The change in RNA ligand conformation places spacer nucleotides into non-specific pockets at the protein interface or arranges them in stacking interactions with other nucleotides. A similar mechanism of binding is seen for the *Oxytricha nova* telomere end-binding protein in which non-cognate bases are flipped out of the binding interface and neighboring nucleotides are shuffled into "cognate-like" conformational register (Theobald and Schultz, 2003). In other systems, the ability of plastic interfaces to accommodate cryptic ligand specificities through similar mechanisms may play important, but currently unappreciated, biological roles. For example, the plasticity exhibited by SH2 domains, PLCγ1 and SH2B1, for phosphotyrosine ligands divergent from cognate specificities reveal the potential for these proteins to play roles in additional signaling pathways (McKercher et al., 2017).

A key result of these studies of altered complexes is that the degree of specificity exhibited for flexible ligands is carefully tuned through the use of alternative conformations. In particular, the extended L23 of Pot1pC, which provides key structural

rearrangement for ligand accommodation, may provide hints for the mechanisms of specificity at other protein interfaces. Long loops, while often showing poor electron density, may well play general roles in ligand accommodation, especially for chemically diverse ligands. These sophisticated mechanisms of conformational malleability observed for Pot1pC are likely shared by other single-stranded nucleic acid binding proteins that are either fully non-specific, such as RPA (Chen and Wold, 2014), or address the same non-degenerate specificity requirements as *S. pombe* Pot1, such as the *S. cerevisiae* Cdc13-Stn1-Ten1 complex (Anderson et al., 2003) and human CST (Wan et al., 2015; Hom and Wuttke, 2017). Moreover, recognition of intrinsically disordered peptides shares many of the same fundamental features and challenges of single-stranded nucleic acid recognition in that disordered proteins are highly flexible and are often comprised of similarly low complexity sequences. As a result, some proteins may recognize some degenerate low-complexity peptide sequence by providing an interface rich with potential favorably interacting residues and discriminate against other similar disordered peptides through differential flexibility preventing full interface utilization by those ligands.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Reagents
  - Protein Expression and Purification
  - Isothermal Titration Calorimetry
  - Crystallization
  - Data Collection and Refinement
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and one table and can be found with this article online at https://doi.org/10.1016/j.str.2018.03.016.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation MCB1121842 and MCB1716425 (to D.S.W.) and NIH Training Grant T32 GM-065103 (to N.R.L.). We would also like to thank Marissa McKercher and Leslie Glustrom for manuscript comments, David McKay for assistance in X-ray data collection and refinement, and Robert Batey for manuscript comments and structure refinement assistance. We also acknowledge the wonderful staff at The Advanced Light Source for help with X-ray data collection. The Berkeley Center for Structural Biology is supported in part by the NIH, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

AUTHOR CONTRIBUTIONS

Conceptualization, N.R.L. and D.S.W.; Methodology, N.R.L. and D.S.W.; Investigation, N.R.L.; Writing, N.R.L. and D.S.W.; Visualization, N.R.L. and D.S.W.; Funding Acquisition, D.S.W.; Supervision: D.S.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 17, 2017
Revised: January 10, 2018
Accepted: March 27, 2018
Published: April 19, 2018

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with *phenix.refine*. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367.

Altschuler, S.E. (2011). Characterization of Single-Stranded DNA Binding and Small Molecule Inhibition of *S. pombe* Pot1 (University of Colorado Boulder).

Altschuler, S.E., Dickey, T.H., and Wuttke, D.S. (2011). *Schizosaccharomyces pombe* protection of telomeres 1 utilizes alternate binding modes to accommodate different telomeric sequences. Biochemistry *50*, 7503–7513.

Anderson, E.M., Halsey, W.A., and Wuttke, D.S. (2003). Site-directed mutagenesis reveals the thermodynamic requirements for single-stranded DNA recognition by the telomere-binding protein Cdc13. Biochemistry *42*, 3751–3758.

Bah, A., Wischnewski, H., Shchepachev, V., and Azzalin, C.M. (2012). The telomeric transcriptome of *Schizosaccharomyces pombe*. Nucleic Acids Res. *40*, 2995–3005.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). *iMOSFLM*: a new graphical interface for diffraction-image processing with *MOSFLM*. Acta Crystallogr. D Biol. Crystallogr. *67*, 271–281.

Baumann, P., and Cech, T.R. (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. Science *292*, 1171–1175.

Beese, L.S., Derbyshire, V., and Steitz, T.A. (1993). Structure of DNA polymerase I Klenow fragment bound to duplex DNA. Science *260*, 352–355.

Bochkarev, A., and Bochkareva, E. (2004). From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. Curr. Opin. Struct. Biol. *14*, 36–42.

Chen, R., and Wold, M.S. (2014). Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. BioEssays *36*, 1156–1161.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). *MolProbity*: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Churikov, D., Wei, C., and Price, C.M. (2006). Vertebrate POT1 restricts G-overhang length and prevents activation of a telomeric DNA damage checkpoint but is dispensable for overhang protection. Mol. Cell. Biol. *26*, 6971–6982.

Cléry, A., Blatter, M., and Allain, F.H. (2008). RNA recognition motifs: boring? Not quite. Curr. Opin. Struct. Biol. *18*, 290–298.

Croy, J.E., and Wuttke, D.S. (2006). Themes in ssDNA recognition by telomere-end protection proteins. Trends Biochem. Sci. *31*, 516–525.

Croy, J.E., Altschuler, S.E., Grimm, N.E., and Wuttke, D.S. (2009). Nonadditivity in the recognition of single-stranded DNA by the *Schizosaccharomyces pombe* protection of telomeres 1 DNA-binding domain, Pot1-DBD. Biochemistry *48*, 6864–6875.

Denchi, E.L., and de Lange, T. (2007). Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. Nature *448*, 1068–1071.

Dickey, T.H., and Wuttke, D.S. (2014). The telomeric protein Pot1 from *Schizosaccharomyces pombe* binds ssDNA in two modes with differing 3' end availability. Nucleic Acids Res. *42*, 9656–9665.

Dickey, T.H., McKercher, M.A., and Wuttke, D.S. (2013). Nonspecific recognition is achieved in Pot1pC through the use of multiple binding modes. Structure *21*, 121–132.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of *Coot*. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D Biol. Crystallogr. *67*, 282–292.

Giegé, R., Sissler, M., and Florentz, C. (1998). Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. *26*, 5017–5035.

Hom, R.A., and Wuttke, D.S. (2017). Human CST prefers G-rich but not necessarily telomeric sequences. Biochemistry *56*, 4210–4218.

Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993a). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520–527.

Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993b). Crystal structure of a yeast TBP/TATA-box complex. Nature *365*, 512–520.

de Lange, T. (2009). How telomeres solve the end-protection problem. Science *326*, 948–952.

Lee, M., Hills, M., Conomos, D., Stutz, M.D., Dagg, R.A., Lau, L.M.S., Reddel, R.R., and Pickett, H.A. (2014). Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. Nucleic Acids Res. *42*, 1733–1746.

Lei, M., Podell, E.R., Baumann, P., and Cech, T.R. (2003). DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. Nature *426*, 198–203.

Lei, M., Podell, E.R., and Cech, T.R. (2004). Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. Nat. Struct. Mol. Biol. *11*, 1223–1229.

Leonardi, J., Box, J.A., Bunch, J.T., and Baumann, P. (2008). TER1, the RNA subunit of fission yeast telomerase. Nat. Struct. Mol. Biol. *15*, 26–33.

Loayza, D., and De Lange, T. (2003). POT1 as a terminal transducer of TRF1 telomere length control. Nature *423*, 1013–1018.

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.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). *Phaser* crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

McKercher, M.A., Guan, X., Tan, Z., and Wuttke, D.S. (2017). Diversity in peptide recognition by the SH2 domain of SH2B1. Proteins. https://doi.org/10.1002/prot.25420.

Messias, A.C., and Sattler, M. (2004). Structural basis of single-stranded RNA recognition. Acc. Chem. Res. *37*, 279–287.

Nandakumar, J., Podell, E.R., and Cech, T.R. (2010). How telomeric protein POT1 avoids RNA to achieve specificity for single-stranded DNA. Proc. Natl. Acad. Sci. USA *107*, 651–656.

Palm, W., and de Lange, T. (2008). How shelterin protects mammalian telomeres. Annu. Rev. Genet. *42*, 301–334.

Rai, R., Zheng, H., He, H., Luo, Y., Multani, A., Carpenter, P.B., and Chang, S. (2010). The function of classical and alternative non-homologous end-joining pathways in the fusion of dysfunctional telomeres. EMBO J. *29*, 2598–2610.

Record, T.M., Lohman, T.M., and De Haseth, P. (1976). Ion effects on ligand-nucleic acid interactions. J. Mol. Biol. *107*, 145–158.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. Annu. Rev. Biochem. *79*, 233–269.

Terwilliger, T. (2004). SOLVE and RESOLVE: automated structure solution, density modification and model building. J. Synchrotron Radiat. *11*, 49–52.

Theobald, D.L., and Schultz, S.C. (2003). Nucleotide shuffling and ssDNA recognition in *Oxytricha nova* telomere end-binding protein complexes. EMBO J. *22*, 4314–4324.

Trujillo, K.M., Bunch, J.T., and Baumann, P. (2005). Extended DNA binding site in Pot1 broadens sequence specificity to allow recognition of heterogeneous fission yeast telomeres. J. Biol. Chem. *280*, 9119–9128.

Wan, B., Tang, T., Upton, H., Shuai, J., Zhou, Y., Li, S., Chen, J., Brunzelle, J.S., Zeng, Z., Collins, K., et al. (2015). The *Tetrahymena* telomerase p75–p45–p19 subcomplex is a unique CST complex. Nat. Struct. Mol. Biol. *22*, 1023–1026.

Wilinski, D., Qiu, C., Lapointe, C.P., Nevil, M., Campbell, Z.T., Tanaka Hall, T.M., and Wickens, M. (2015). RNA regulatory networks diversified through curvature of the PUF protein scaffold. Nat. Commun. *6*, 8213.

Wilson, K.A., Kellie, J.L., and Wetmore, S.D. (2014). DNA-protein π-interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. Nucleic Acids Res. *42*, 6726–6741.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
(198-339) V199D Pot1pC	Dickey et al., 2013	N/A
Chitin Resin	New England Biolabs	S6651S
Deposited Data		
1R Pot1pC complex structure	This paper	PDB: 5USB
1-3R Pot1pC complex structure	This paper	PDB: 5USN
7-9R Pot1pC complex structure	This paper	PDB: 5USO
Cognate Pot1pC complex structure	Dickey et al., 2013	PDB: 4HIK
T4A Pot1pC complex structure	Dickey et al., 2013	PDB: 4HIO
G2C Pot1pC complex structure	Dickey et al., 2013	PDB: 4HID
Experimental Models: Organisms/Strains		
Escherichia coli: BL21 (DE3)	New England Biolabs	C2527I
Oligonucleotides: Please see Table S1 for oligo/primer sequences.		
Recombinant DNA		
(198-339) V199D Pot1pC-intein fusion (NEB: pTXB1)	Dickey et al., 2013	N/A
Software and Algorithms		
MicroCal Origin 7.0	OriginLab	http://www.originlab.com/; RRID: SCR_002815
iMosflm and SCALA	Winn et al., 2011	www.ccp4.ac.uk/; RRID: SCR_014217
PHASER	McCoy et al., 2007	http://www.phaser.cimr.cam.ac.uk/; RRID: SCR_014219
PHENIX	Adams et al., 2010	http://www.phenix-online.org/; RRID: SCR_014224
Phenix.refine	Afonine et al., 2012	http://www.phenix-online.org/
Coot	Emsley et al., 2010	http://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/; RRID: SCR_014222
MolProbity	Chen et al., 2010	http://www.phenix-online.org/; RRID: SCR_014226
PyMOL	Schrödinger	https://pymol.org/2/; RRID: SCR_000305

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the Lead Contact Deborah Wuttke (Deborah.wuttke@colorado.edu).

METHOD DETAILS

Reagents

All oligonucleotides used in this study were obtained from Integrated DNA Technologies.

Protein Expression and Purification

Pot1pC was expressed and purified using essentially the same method described in Dickey et al. (2013). Briefly, V199D Pot1pC was expressed as an intein-chitin-binding domain fusion in BL21 (DE3) *E. coli* at 18°C for 20 hours. Following bacterial cell harvesting and lysis, the fusion construct was bound to chitin beads (New England Biolabs) and Pot1pC was cleaved from the intein-chitin binding domain by incubation with 100 mM beta mercaptoethanol (βME) for 20-40 hrs at 4°C. Following elution, Pot1pC was concentrated and injected onto a Superdex 75 column (GE) in 100 mM Tris pH 8.0, 100 mM KCl, 0.1% (w/v) deoxycholate, 3 mM βME, and 5% (v/v) glycerol. After elution from the size exclusion column, ~99% pure protein was concentrated to 450-600 μM, snap frozen in liquid nitrogen, and stored at -70°C.

Isothermal Titration Calorimetry

Pot1pC stored at -70°C was thawed and dialyzed overnight at 4°C in buffer containing 20 mM potassium phosphate pH 8.0, 150 mM NaCl, and 3 mM βME. Oligonucleotides obtained from Integrated DNA Technologies were resuspended in the same dialysis buffer. Heats of dilution experiments showed no detectable heat evolved and thus were not subtracted from binding experiments. All experiments were performed in triplicate on a MicroCal iTC200 (GE Healthcare) at 25°C. The sample cell was loaded with 230 μL of 5-100 μM Pot1pC into which buffer matched nucleic acid at approximately 10-fold higher concentration was titrated as follows: one 0.2 μL dummy injection, followed by nineteen 2 μL injections, and a final 1.3 μL injection. Data were integrated and fit by nonlinear least-squares fitting to a single binding site model using MicroCal Origin 7.0 (OriginLab, Northampton, MA).

Ultraviolet (280 and 260 nm) absorbance measurements were used to calculate protein and nucleic acid concentrations, using extinction coefficients provided by ExPASy ProtParam and Integrated DNA Technologies, respectively.

Crystallization

Crystals were grown using the hanging drop vapor diffusion method at 4°C. Drops contained 1 μL of mother liquor and 1 μL of a solution of 1:1 protein:ssRNA/DNA (5-15 mg/mL). Crystallization conditions for each complex are listed in Table 2. The 1-3R and 7-9R crystals were obtained by two-step seeding with the cognate ssDNA complex crystals providing the initial seeds and then the resulting low-quality 1-3R and 7-9R crystals as the seeds for a second round of seeding. Seeds were generated by vortexing seed crystals in mother liquor (Seed Bead crystal kit, Hampton Research) and resulting microcrystals were transferred to the hanging drop by dipping a cat whisker into the seed solution and swiping it through the drop. Crystals were cryoprotected by sequentially transferring the crystal in mother liquor solutions supplemented with 5%, 10%, 15%, and 20% (v/v) ethylene glycol and flash frozen in liquid nitrogen.

Data Collection and Refinement

X-ray diffraction data for 1R was collected at the Advanced Light Source (ALS) Beamline 8.2.1 and the data sets for 1-3R and 7-9R were collected at the ALS Beamline 8.2.2. Reflections were indexed using iMOSFLM (Battye et al., 2011) and scaled using Scala within the CCP4 program suite (Evans, 2011; Winn et al., 2011). The phases were solved through molecular replacement using the coordinates of cognate Pot1pC without ssDNA (4HIK) (Dickey et al., 2013) as a starting model using PHASER in the PHENIX suite (McCoy et al., 2007; Adams et al., 2010) followed by rigid body refinement using PHENIX Refine (Terwilliger, 2004; Afonine et al., 2012). The non-cognate RNA ligands were built into the electron density manually in Coot (Emsley et al., 2010) and subsequent refinement was performed in the PHENIX program suite with manual adjustment in Coot. The final models were validated using PHENIX.validate and MolProbity (Chen et al., 2010) to assess quality (statistics for final models can be found in Table 2).

QUANTIFICATION AND STATISTICAL ANALYSIS

For all ligand interactions, K_D, ΔH, and binding sites (N) were fit using a one-site binding model in MicroCal Origin 7.0, and TΔS was calculated from fitted values. Reported K_D, ΔH, and TΔS values were averaged from at least three ITC experiments, and the error reported is standard error of the mean. Representative ITC data and curve fitting is shown in Figure S1.

DATA AND SOFTWARE AVAILABILITY

The coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 5USB (1R), 5USN (1-3R), and 5USO (7-9R).