Specific Recognition of the C-rich Strand of Human Telomeric DNA and the RNA Template of Human Telomerase by the First KH Domain of Human Poly(C)-Binding Protein-2

Zhihua Du*, Jinghua Yu*, Yihong Chen*, Raul Andino†, and Thomas L. James‡*

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-2280; Department of Microbiology and Immunology, University of California, San Francisco, California 94143-2280

Running Title: Structure and Nucleic Acid Binding of PCBP2 KH1 Domain

‡ To whom correspondence should be addressed: Department of Pharmaceutical Chemistry, 600 16th Street – Genentech Hall, University of California, San Francisco, California 94143-2280, U.S.A.; e-mail: james@picasso.ucsf.edu; tel.: (415) 476-1916; fax: (415)-502-8298

* Department of Pharmaceutical Chemistry

† Department of Microbiology and Immunology

Copyright 2004 by The American Society for Biochemistry and Molecular Biology, Inc.
Abbreviations: PCBP, poly(C)-binding protein; KH-domain, hnRNP-K homology domain; KH1, the first KH domain of PCBP2; RNA, ribonucleic acid; 5’UTR, 5’-untranslated region; 3’UTR, 3’-untranslated region; RNP, ribonucleoprotein; IRES, internal ribosomal entry site; PV1, poliovirus type 1; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; DQF-COSY, double-quantum-filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; HSQC, heteronuclear single quantum coherence spectroscopy; ppm, parts per million; RMSD, root-mean-square deviation; C12E6, hexaethylene glycol monododecyl ether.
ABSTRACT

Poly(C)-binding proteins (PCBPs) constitute a family of nucleic acid-binding proteins that play important roles in a wide spectrum of regulatory mechanisms. The diverse functions of PCBPs are dependent on the ability of the PCBPs to recognize poly(C) sequences with high affinity and specificity. PCBPs contain three copies of KH (hnRNP K homology) domains, which are responsible for binding nucleic acids. We have determined the NMR structure of the first KH domain (KH1) from PCBP2. PCBP2 KH1 domain adopts a structure with three \( \alpha \)-helices packed against one side of a three-stranded antiparallel \( \beta \)-sheet. Specific binding of PCBP2 KH1 to a number of poly(C) RNA and DNA sequences, including the C-rich strand of the human telomeric DNA repeat, the RNA template region of human telomerase, and regulatory recognition motifs in the poliovirus-1 5’UTR, was established by monitoring chemical shift changes in protein \( ^{15}N \)-HSQC spectra. The nucleic acid binding groove was further mapped by chemical shift perturbation upon binding to a six-nucleotide human telomeric DNA. The binding groove is an \( \alpha/\beta \) platform formed by the juxtaposition of two \( \alpha \)-helices, one \( \beta \)-strand, and two flanking loops. While there is a groove in common with all of the DNA and RNA binders with a hydrophobic floor accommodating a three-residue stretch of C residues, nuances in recognizing flanking residues are provided by hydrogen bonding partners in the KH domain. Specific interactions of PCBP2 KH1 with telomeric DNA and telomerase RNA suggest that PCBPs may participate in mechanisms involved in the regulation of telomere/telomerase functions.
Poly(C)-binding proteins (PCBPs) constitute a family of nucleic acid-binding proteins characterized by their ability to interact with poly(C) sequences with high affinity and specificity (for reviews, see (1,2)). To date, five major members of the PCBP family have been identified in mammalian cells, namely PCBP1-4 (also known as αCP1-4; PCBP1 and 2 are also known as hnRNP E1 and E2) and hnRNP K. These PCBP proteins are evolutionarily related, as revealed by comparative analyses of the protein and gene structures (1). Except PCBP1, whose mRNA is intron-less, all PCBPs are known to have different isoforms due to alternative splicing. Most often, the PCBPs contain three copies of a conserved RNA-binding motif known as the KH domain (hnRNP K homology), one of the most frequently occurring RNA-binding modules. The arrangement of the three KH domains is similar in each of the PCBPs: two consecutive KH domains at the N-terminus and a third KH domain at the C-terminus, with an intervening sequence of variable length between the second and third domains (Figure 1a).

PCBP proteins play important roles in a wide spectrum of regulation mechanisms in mammalian cells, at both the transcriptional control and posttranscriptional control levels, as well as in apoptotic and developmental pathways (1). It has been shown that binding of PCBPs to pyrimidine-rich sequences within the 3'UTRs of a number of cellular mRNAs confers unusual stability to these mRNAs, including α-globin (3,4), collagen-α1 (5,6), tyrosine hydroxylase (7), and erythropoietin (8) mRNAs. Interaction between PCBPs and 3'UTR pyrimidine-rich sequences can also result in translational silencing, as seen in 15-lipoxygenase (LOX) mRNA (9-11). Interestingly, the PCBP-binding sequences in the stabilized mRNAs and the silenced mRNAs are very similar, each containing three C-rich patches. The difference lies in that the PCBP-binding motifs always appear in tandem in the silenced mRNAs; a minimum of two repeats is required for function. Correspondingly, the stabilizing complexes involve only a single
molecule of PCBP1 or PCBP2, while in the silencing complexes, two PCBPs are involved: hnRNP K and PCBP1 or 2.

Besides cellular RNAs, PCBPs also critically participate in the regulation of viral RNA functions. It was shown that the 5’-UTR of Poliovirus mRNA (which is also the genomic RNA) harbors two binding sites for PCBP1 or 2. Binding of PCBP2 to one of the sites, a C-rich internal bulge sequence known as loop B RNA within the IRES (Internal Ribosomal Entry Site) element, is required for cap-independent translation of the viral RNA (12),(13) (14,15); (16). Binding of PCBP1 or 2 to the other site, a C-rich loop in the stem-loop B domain within a so-called cloverleaf-like RNA structure, achieves high affinity only in the presence of the viral protein 3CD (precursor of the viral protease 3C and the viral polymerase 3D) bound to a different part of the cloverleaf. Formation of the PCBP-cloverleaf-3CD ternary complex inhibits translation and switches the viral RNA to a template for RNA replication (17); (15). Unlike PCBP-binding sites within 3’UTRs of the cellular mRNAs, the two viral C-rich recognition sequences are presented in the context of a structured RNA, as demonstrated by our recent NMR structures of two sequence variants of the loop B RNA (18).

Involvement of PCBPs in regulating RNA functions is certainly not limited to the cases mentioned above. A recent study has identified 160 mRNA species that associate in vivo with PCBP2 from a human hematopoietic cell line (19), suggesting that the contribution of PCBPs in posttranscriptional gene regulation may be far more profound than currently known. Moreover, PCBP actions may not be limited to mRNAs. For example, an interaction between a PCBP4 isoform known as MCG10 or PCBP4b and the C-rich RNA template region of human telomerase has been proposed as a possible mechanism for the functional role of PCBP4b as a p53-induced regulator in apoptosis and cell cycle arrest at G2-M (20).
Biological functions of PCBPs are further diversified by their ability to interact specifically with not only RNA but also DNA sequences. Specific binding of hnRNP K to the single-stranded pyrimidine-rich sequence in the promoter of human \textit{c-myc} gene activates transcription (21). It was also shown that hnRNP K could recognize the C-rich strand of human telomeric DNA with high affinity (22). This property of high-affinity binding to telomeric repeats has recently been extended to PCBP1. Using affinity chromatography and mass spectroscopy, Bandiera and coworkers had identified PCBP1 as the most specific C-rich telomeric DNA binder in the K562 human cell line (23). The authors also implied that PCBP2 should share similar binding property based on the high degree of homology between PCBP2 and PCBP1 (82% identical, 88% positive; the three nucleic acid-binding KH domains share an even higher degree of homology, which is 93% identical and 97% positive collectively –see Figure 1b), and the remarkable similarity between high affinity SELEX RNA target sequences for PCBP2, $W_{2.6}[C_{3.5}W_{2.6}]_{3.4}$, W is A or U (24), and the C-rich human telomeric repeats, (CCCTAA)$_n$. Further studies are needed to confirm the interaction between the PCBPs and human telomeric DNA, and to assign the functional importance of such an interaction.

Given all the current knowledge about PCBPs, it is very intriguing to note that the highly diverse functional roles of PCBPs are nearly all dependent on the ability to recognize C-rich RNA or DNA sequences with high specificity and affinity. It is known that individual KH domains from PCBPs can bind nucleic acids as discrete and independent entities (25). With the presence of three KH domains in the PCBPs, it would be very interesting to discover which KH domain, or synergy of domains, contributes to the nucleic-acid-binding specificity and affinity. It would also be important to reveal the detailed interactions defining specificity and affinity, and
how the seemingly similar events of binding to single-stranded C-rich motifs give rise to the very broad spectrum of functional diversities.

In an effort to address these fundamentally important questions regarding structures and functions of PCBPs, we have begun a program to study the nucleic acid-protein interactions with a focus on PCBP2. So far, a large number of nucleic acid targets have been established or proposed for PCBP2. This portfolio covers both RNA and DNA sequences, both naturally occurring and SELEX motifs, both non-structured and highly structured RNAs. Moreover, there is relatively good knowledge about the nucleic acid-binding property of the three KH domains from PCBP2; it was shown that both the KH1 and KH3 domains as independent entities could bind strongly to poly(C) ribohomopolymers (25); however, only the KH1 domain was shown to interact specifically with the two C-rich sequences presented within the cloverleaf and the IRES structures from the Poliovirus 5’UTR. With all these data, there is no doubt that PCBP2 represents one of the best biochemically characterized proteins not only in the PCBP subfamily but also in the superfamily of KH domain-containing proteins (considering that no bona fide nucleic acid targets have been identified for most of the known KH domain-containing proteins). Detailed structural study on the PCBP2 system should provide valuable insights into the molecular mechanisms of PCBP functions in particular, and KH-domain mediated processes in general. As part of our continuing research effort, we have determined the solution structure of the first KH1 domain from PCBP2 and characterize its interaction with DNA and RNA targets including sequences from cellular mRNA, Poliovirus RNA, human C-rich strand telomeric DNA, and the RNA template of human telomerase.
EXPERIMENTAL PROCEDURES

Sample Preparation. The gene that encodes the first KH domain (KH1) of human Poly(C) Binding Protein-2 (PCBP2) was amplified by PCR using appropriate primers and a plasmid containing the gene for full-length PCBP2. The amplified gene was cloned between Nde I and Xho I sites of the plasmid vector pET 24a, so the protein was overexpressed with a C-terminal His-tag. The cloned plasmid was transformed into a BL21(DE3) strain of Escherichia coli (Strategene), and the bacteria were grown at 37°C in Luria-Bertani medium for unlabeled samples, or in M9 minimal medium with 15N-NH4Cl as the sole nitrogen source and 13C-glucose as the sole carbon source for 15N-labeled and 15N,13C-labeled samples. Protein expression was induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at OD600 = 0.6-0.8, and cells were harvested after a further growth of 3-4 hours at 37°C. Cell pellets were resuspended in a buffer containing 50 mM sodium phosphate (pH 8.0) and 50 mM of sodium chloride and lysed by sonication. The lysate was cleared by centrifugation at 12,000g for 30 minutes. The supernatant containing the His-tagged protein was added to a slurry of pre-equilibrated Tycon resin (Clontech) and incubated at 4°C on a nutater for 2 hours. The protein-bound resin was washed three times with a buffer containing 50 mM sodium phosphate (pH 8.0) and 300 mM sodium chloride. Elution of the His-tagged protein was achieved by incubating the resin in a buffer containing 25 mM sodium phosphate (pH5.5), 25 mM sodium chloride and 0.5 M imidazole for 10 minutes at 4°C. The protein solution was dialyzed extensively against NMR buffer containing 25 mM sodium phosphate (pH5.5) and 25 mM sodium chloride, and it was concentrated by centrifugation on Centricon 3 (Amicon) to 1-2 mM. For experiments in D2O, the proteins were lyophilized and redissolved in 99.996% D2O. Loop B and stem-loop B RNA samples with sequences illustrated in Figure3c were prepared by in vitro transcription using T7
RNA polymerase and a synthetic DNA template (26,27). For annealing, the RNA solutions were heated to 90°C and slowly cooled to room temperature. DNA and RNA oligonucleotides were ordered from Integrated DNA Technology (IDT).

Samples of the KH1–nucleic acid complexes for NMR studies were prepared by titrating solutions of the nucleic acids into solutions of the KH1 protein domain and monitoring the NMR signals of the protein in a 15N-HSQC spectrum.

**NMR Spectroscopy.** All NMR experiments were performed on Varian Inova spectrometers operating at 600 MHz for protons. Spectra were processed with NMRPipe/NMRDraw (28) and analyzed with SPARKY (29). In the 2D and 3D homonuclear and heteronuclear experiments, the States-TPPI method (30) was used to achieve quadrature detection in the indirect dimensions. All NMR experiments except those involving the anisotropic (liquid crystal) sample were carried out at 30°C. 1H, 15N, and 13C resonance assignments were achieved by using the following 3D heteronuclear NMR experiments: HNCO, HNCA, HN(CO)CA (31), CBCANH (32), CBCA(CO)NH (33), HBHA(CO)NH, C(CO)NH, H(CCO)NH, 15N TOCSY-HSQC and HCCH-TOCSY (34). Backbone φ and ψ torsion angle restraints were derived from 3JHNHA values measured in a quantitative J correlation HNHA experiment (35), and backbone chemical shifts were derived using the program TALOS (36). Interproton distance restraints were derived from 2D homonuclear NOESY spectra (60 ms and 150 ms mixing time), 3D 15N-separated NOESY (80 ms mixing time) and 13C-separated NOESY (150 ms mixing time) experiments. Measurements of amide one bond N-H residual dipolar couplings were carried out at 25°C by using a 2D IPAP 15N-1H HSQC experiment. Liquid crystalline media for measurements of residual dipolar coupling constants were prepared by step-wise addition with vigorous vortexing of microliter aliquots of hexanol to the protein solution containing 2.5% (w/v) of C12E6 until the
solution became translucent. Dipolar coupling values were calculated from the difference in corresponding $J_{\text{NH}}$ splittings measured in the anisotropic (liquid crystalline) phase and in the isotropic phase.

To monitor complex formation between the PCBP2 KH1 domain and various DNA/RNA molecules, we compared $^{15}$N-HSQC or TROSY-$^{15}$N-HSQC spectra of the $^{15}$N-labeled protein before and after nucleic acids titration. It should be noted that although TROSY-type experiments are usually used in systems suffered from line-broadening as a result of increased molecular weight or exchanges and therefore suggestive of such problems, the choice of TROSY experiment for some of the complexes in this study is largely a matter of personal flavor, rather than necessity. TROSY experiment is less sensitive but has superior line-shape because no decoupling is required during data acquisition. Since solubility of the KH1 domain and its nucleic acid complexes is generally quite good, we were able to perform our experiments at very high protein concentration. Regular and TROSY-type $^{15}$N-HSQC spectra were used quite randomly in our studies without noticeably difference; with the only exception being the complex with a 34-nt RNA bulged stem-loop for which reasonable spectra could only be obtained using TROSY-$^{15}$N-HSQC.
*Structure Calculations.* In addition to the NOE-derived distance restraints, residual dipolar coupling restraints, and TALOS/HNHA-derived torsion angle restraints, generic hydrogen bond distance restraints were utilized for regions of regular secondary structure that were based on characteristic NOE patterns and secondary chemical shifts. Structure refinement was carried out by simulated annealing in torsion angle space using established procedures implemented in the program XPLOR-NIH (37). NMR restraint violations and structure quality were analyzed via the programs AQUA and PROCHECK-NMR (38).

Structure figures were generated using MidasPlus (39) or Chimera (40) from the Computer Graphics Laboratory, University of California, San Francisco.

**RESULTS**

*Structure of the PCBP2 KH1 Domain.* The secondary structure of the PCBP2 KH1 domain was defined by established NMR methods based on chemical shift indexing (CSI), backbone torsion angles derived from the HNHA experiment and TALOS prediction, as well as characteristic NOE patterns. Figure 1c shows the NMR-derived secondary structure, with a summary of the sequential and medium range NOEs characteristic of α-helices or β-strands, and chemical shift indexing (CSI) data taking into account of chemical shifts of amide proton, amide nitrogen, α and β carbons. The secondary structure consists of three α helices and three β strands arranged in the order β1-α1-α2-β2-β3-α3. The evolutionarily conserved invariable GXXG loop is located between α1 and α2; the variable loop is between β2 and β3. As shown in Figure 1c, measured one-bond amide H-N RDC values show different signs for the α helices and β strands, indicting the roughly parallel orientation of these secondary structure elements.
The three-dimensional structure of the PCBP2 KH1 domain was calculated using residual dipolar couplings (RDC) restraints in addition to distance restraints derived from NOE measurements and torsion angle restraints derived from scalar coupling experiments. Structure calculation statistics are summarized in Table 1. Figure 2a shows the backbone superimposition of the twenty final structures with the lowest energy penalty; Figure 2b is a ribbon diagram of the structure with lowest overall energy. The spatial arrangement of the structural elements of the PCBP2 KH1 domain is that the three β strands form an antiparallel β-sheet, with a spatial order of β1-β3-β2; the three α helices are packed against one side of the β-sheet. The two loops, which are less ordered, point upward in the orientation shown and help to define the edges of the nucleic acid binding platform (details in later sections). The three α helices and the three β strands are all amphipathic in nature, with their hydrophobic residues pointing inward to form a hydrophobic core. The residues in the core (L14, I16, L18, I47, I49, I59, and L61 from the β strands; V25, I28, I29, V36, M39, I68, A71, F72, I75, I76, and L79 from the α helices) are exclusively hydrophobic; most of these residues are isoleucines, leucines, or valines, which are the most hydrophobic amino acids (highest positive values for the Kyte-Doolittle hydropathy index). It therefore seems like a very high degree of hydrophobicity is sought in the core. This also provides a good explanation for the high degree of conservation for these residues among the KH domains (Figure 1b). There are several sharp turns of the backbone in the molecule, at junctions between α1 and β1, β2 and α2, α3 and β3 (Figures 2a and 2b); all of these sharp turns are mediated by a glycine (G22, G44, and G63, respectively, at the three junctions). G44 and G63 are highly conserved among the KH domains from PCBPs. G22 is conserved among the KH1 domains except that from hnRNP K. For those KH domains in which a glycine is not present at position 22, a proline is most often used as a substitute at position 21 (Figure 1b). The
absolute conservation of G30 and G33 of the invariant GXXG loop (Figure 1b) is also most likely due to steric consideration; the two helices (α1 and α2) connected by the GXXG loop assume a tight interhelical angle of ~110-120°.

**PCBP2 KH1 Has Poly(C)-Binding Specificity for Both DNA and RNA Sequences.** To test the nucleic acid binding specificity of the PCBP2 KH1 domain, we used $^{15}$N-HSQC spectra to monitor the chemical shift changes upon titration of the protein solution with various nucleic acid probes. We first tested whether PCBP2 KH1 binds to a six-nucleotide RNA oligomer, 5’-UCCCCA-3’, which corresponds to one of the three C-rich patches comprising the RNA targets for PCBP2 within the 3’-UTR of the mRNAs of α-globin, collagen-α1, tyrosine hydroxylase, and 15-lipoxygenase. This sequence also conforms to the SELEX consensus of PCBP2-binding RNAs. As shown in Figure 3a, addition of this RNA oligonucleotide to PCBP2 KH1 causes very dramatic chemical shift changes in the $^{15}$N-HSQC spectrum; these include a couple of substantial downfield shifts unambiguously indicating the formation of a protein-RNA complex. To test whether this interaction is specific, we also titrated the protein with three other RNA oligonucleotides with the sequences 5’-UUUUUA-3’, 5’-UAAAAA-3’, and 5’-UGGGGA-3’, respectively. None of these oligonucleotides produced significant changes in the $^{15}$N-HSQC spectra that would suggest specific RNA-protein interaction (data not shown; an example of one such $^{15}$N-HSQC spectrum is included in the Supplementary Materials).

To test the ability of the PCBP2 KH1 domain to interact with DNA sequences, four DNA oligonucleotides, which correspond in sequence to the four RNA oligonucleotides, were used as probes. Similar to the RNA sequences, only the DNA oligonucleotide 5’-TCCCCA-3’ could cause significant chemical shift changes in the $^{15}$N-HSQC spectrum (Figure 3a). Comparing the $^{15}$N-HSQC spectra of the RNA and DNA complexes, it is clear that most of the chemical shift...
changes induced by RNA and DNA binding are identical or very close; only three cross-peaks exhibit relatively large differences between RNA and DNA complexes (the cross-peaks from the RNA complex spectrum that have no apparent corresponding cross-peaks from the DNA complex in close proximity are boxed). In addition to the backbone amides, two out of the four arginine side-chain epsilon H-N groups also showed drastic chemical shift changes upon titration with poly(C) RNA or DNA oligonucleotides (Figure 3a), indicating that these arginine side-chains are directly involved in complex formation.

Our NMR binding assays on the eight RNA and DNA oligonucleotides prove that PCBP2 KH1 domain recognizes poly(C) RNA and DNA sequences specifically. Moreover, our data suggest that the RNA-and DNA-binding modes should be very similar although minor differences may exist.

**PCBP2 KH1 Binds to the C-rich Strand of Human Telomeric DNA and the Template Sequence of the RNA Component of Human Telomerase.** Inspired by the recent suggestion that PCBP2 might interact with the C-rich strand of human telomeric DNA (Bandiera et al., 2003), we titrated PCBP2 KH1 domain with a 10-nucleotide DNA sequence, 5’-CTAACCCTAA-3’, which corresponds to a C-rich strand telomeric repeat. As shown in Figure 3b, this DNA oligonucleotide causes very significant chemical shift changes in the $^{15}$N-HSQC spectrum of PCBP2 KH1, indicating formation of a specific DNA-protein complex. Since the C-rich telomeric repeat is virtually a DNA counterpart of the RNA template sequence of human telomerase, we went on to test whether PCBP2 KH1 could also interact with the RNA template sequence of human telomerase, 5’-CUAACCCUAA-3’, in a specific manner. As shown in Figure 3b, the RNA oligonucleotide also causes dramatic chemical shift changes in the $^{15}$N-HSQC spectrum of PCBP2 KH1. Similar to observations from Figure 3a, the spectra in Figure
3b from the DNA and RNA complexes show identical or quite similar chemical shift changes for most of the residues, with only a couple NMR signals manifesting a relatively larger difference between DNA and RNA complexes (some such cross-peaks from the RNA complex spectrum are boxed).

Since we had established that a DNA oligonucleotide with the sequence 5’-TAAAAA-3’ did not interact with PCBP2 KH1, we hypothesized that the three C residues within the telomeric DNA repeat or the telomerase RNA template provided the determinant for specific interaction with PCBP2 KH1. Indeed, when we titrated PCBP2 KH1 with a 6-nucleotide DNA 5’-AACCCT-3’, which is a shorter version of the C-rich human telomeric DNA, we obtained a virtually identical 15N-HSQC spectrum as that with the 10-nucleotide DNA (Figure 3d); G30 and I49 cross-peaks are not observed in the 10-nucleotide DNA complex most likely due to different dynamics; all other cross-peaks could be superimposed with corresponding cross-peaks in the 10-nucleotide DNA-complex spectrum. Based on these data, we conclude that specific recognition of the human telomeric DNA sequence by PCBP2 KH1 is mainly mediated by a motif centered on the three C residues. Although we did not perform a similar experiment on the telomerase RNA template, we believe this conclusion should also be valid for the RNA case, based on our results for the other RNA oligonucleotides and the similarity between DNA and RNA binding.

**PCBP2 KH1 Specifically Recognizes the Stem-Loop B and Loop B RNAs from the 5’UTR of Poliovirus RNA.** Biochemical data have shown that PCBP2 KH1 is the major determinant for specific recognition by PCBP2 of the RNA structures constituting both the cloverleaf stem-loop B and loop B from IRES stem-loop IV of RNA of the 5’UTR Poliovirus (41). By monitoring chemical shift changes in 15N-HSQC spectra, we showed that both RNAs (sequences of the
NMR constructs are shown in the upper right and lower left corners in Figure 3c) induced dramatic chemical shift changes in the PCBP2 KH1 spectra (Figure 3c), proving that the PCBP2 KH1 domain, as an independent entity, can achieve specific recognition of the stem-loop B and loop B RNAs. Some similarities between the spectra of the viral RNA complexes and those of the short RNA oligonucleotide complexes are observed; most obviously for the characteristic peaks, down-field shifted around 10 ppm on the proton axis, and the two glycine cross-peaks from G30 and G33 of the GXXG loop. These similarities suggest that recognition of the viral structured RNAs and RNAs within a “non-structured” context by PCBP2 KH1 should to a certain extent be similar to each other. Of course, some differences are fully expected and can only be revealed by solving the structures of the respective complexes.

_Nucleic Acid Binding Does Not Cause Substantial Structural Changes in PCBP2 KH1 Domain._ We have made almost complete backbone assignments for PCBP2 KH1 domain in complex with the 6-nucleotide (5’-AACCCT-3’) human telomeric DNA. We have also assigned a significant number of NOEs that permit a meaningful analysis of the NOE patterns. Based on identification of characteristic short- and medium-range NOEs and chemical shift indexing, it is clear that the DNA-bound form of PCBP2 KH1 domain assumes the same secondary structure as in the free form; DNA binding does not disrupt any existing secondary structure nor induce any new secondary structure. Moreover, DNA-binding seems to only affect residues in and around the nucleic-acid-binding pocket, which is presented on one surface of the protein (_vide infra_). Structures of free and RNA- or DNA-bound forms of two KH domains; namely, the KH3 domains from hnRNP K and Nova2, respectively, have been solved by NMR and X-ray crystallography (42); (43); (44,45). In both cases, the overall structure of the KH domain was not changed significantly by nucleic acid binding. Based on the data we have at this point, this
scenario should also be true for the PCBP2 KH1 domain interacting with the human telomeric DNA.

**Mapping the Telomeric DNA Binding Interface of PCBP2 KH1 by Chemical Shift Perturbations.** With substantial assignments available for the PCBP2 KH1 domain in both the free and telomeric DNA bound forms, we were able to map unambiguously the DNA binding interface by chemical shift perturbations. Chemical shift changes for the backbone amide protons and nitrogens of PCBP2 KH1 are plotted against residue number in the top two panels of Figure 4. Except for T60, all of the residues experiencing significant chemical shift changes are located in the first two α helices (α1 and α2), the second β strand (β2), and the two loops (the GXXG loop and the variable loop). Significant chemical shift changes for other backbone nuclei (Hα, Cα and C') are also confined to the α1-α2-GXXG loop-β2-variable loop region. These data indicate that the nucleic acid binding site of PCBP2 KH1 is an α/β platform defined by the juxtaposition of two α helices and one β strand, with further contribution from the two flanking loops.

A composite chemical shift change index was calculated based on the observed HN and NH chemical shift differences. This index is plotted against the residue number in the bottom panel of Figure 4. To map the composite chemical shift differences onto the NMR-derived PCBP2-KH1 structure, the residues were classified into three categories based on the magnitude of their composite shift values and colored accordingly in a surface representation of the PCBP2 KH1 structure. Residues with no changes or small changes (composite shift values < 0.15) are colored dark grey; residues with medium changes (composite shift values between 0.15 and 0.30) are colored orange; residues experiencing the biggest changes (composite shift values > 0.3) are colored red. As shown in Figures 5a and 5b, residues with red or orange colors are clustered on one surface of the structure and clearly define a narrow and elongated groove for nucleic acid
binding. The narrow width of the groove may impose a steric impediment for bulky purine residues, explaining the preference for pyrimidine-rich sequences; the limited length of the groove may dictate that only a limited number of residues are allowed in the core nucleic acid recognition sequences.

Properties of the Nucleic Acid Binding Groove of PCBP2 KH1. Now that we have mapped the nucleic acid binding groove of PCBP2 KH1, we can analyze this in some detail to gain insights into its attributes that may dictate specificity and affinity for binding nucleic acids. To assist this analysis, we generated another surface representation of the nucleic acid binding groove, which has an identical viewing angle as in Figure 5a but with a different coloring scheme based on residue properties instead of on chemical shift perturbation. In Figure 6, hydrophobic residues are colored green, positively charged residues blue, negatively charged residues red, and uncharged hydrophilic residues yellow. It is seen clearly that the floor of the nucleic acid binding groove consists of hydrophobic residues exclusively. Isoleucine and valine, two of the most hydrophobic residues, dominate the floor, with a string of four glycines lying along the groove ridge on the upper left corner of the floor. The absolute conservation of hydrophobicity on the floor strongly indicates its functional or structural importance. A single point mutation, Ile to Asn, in KH2 of FMR1 (corresponding to V36 in PCBP2 KH1) results in fragile X mental retardation syndrome. Introduction of the asparagines changes the electrostatic properties of the nucleic acid binding groove, making it significantly less hydrophobic. At the ridge on the left-hand side of the groove, all residues but one are positively charged (either lysine or arginine). Residues at the right-hand side ridge of the groove seem more diverse in terms of electrostatic property, but they are common in that they are all hydrophilic residues and possess side-chain functional groups that can act as hydrogen bond donors; the later attribute is probably
more critical in defining nucleic acid binding specificity and affinity. No aromatic residues are present in the binding groove, indicating that base-stacking interaction between protein side chains and nucleic acid bases does not play any role in nucleic acid binding by the PCBP2 KH1 domain.

DISCUSSION

The structure of the PCBP2 KH1 domain we report here is similar to previously reported structures of other KH domains (46); (43); (44); (42); (47); (48), especially the third KH domain from hnRNP K. hnRNP K and PCBP2 both belong to the PCBP subfamily of KH-domain containing proteins. PCBP2 KH1 is ~31% identical to KH3 of hnRNP K. Pairwise backbone RMSDs between the structures of PCBP2 KH1 we report here and the free KH3 of hnRNP K reported by Baber and coworkers (42) is only about 1.4 Å (excluding the variable loop which has a different number of residues for the two domains (Supplementary Materials). KH3 is the only KH domain within hnRNP K that is capable of binding poly(C) nucleic acid sequences as an independent motif (25). This domain specifically recognizes a DNA tetrad of 5’-TCCC-3’, whose RNA equivalent conforms to the optimal consensus SELEX RNA sequence motif for hnRNP K, UC\textsubscript{3}W\textsubscript{2}, W is either A or U (24). Molecular details about specific recognition of the DNA tetrad by KH3 of hnRNP K have been recently revealed by a NMR structure of the complex (45). In the complex, specific recognition of the bases of the three C residues is mediated by an extensive network of hydrogen bonds involving residues Ile29, Ile36, Ile49, and Arg59 (for example, the backbone amide of Ile49 forms a hydrogen bond with the O2 group of the third C residue in the triple C sequence); electrostatic interaction with the RNA backbone phosphate groups comes from the backbone amide of Gly32 and the side chains of Lys31, Lys37 and Arg40. In the PCBP2 KH1 domain, corresponding residues at these key positions are Ile29,
Val36, Ile49, Arg57, Lys32, Lys31, Lys37 and Arg40, respectively (Figure 1b). In the structure of the free PCBP2 KH1 domain (Figure 5), these residues are positioned at roughly the right places (as compared to those in the hnRNP K KH3-DNA complex structure) to effect a similar set of interactions with a triple C sequence; only relatively minor rearrangements of these amino acids would be required to effect complex formation.

Although both hnRNP K KH3 and PCBP2 KH1 domains exhibit specificity towards poly(C) sequences, they might differ in their recognition of the identity of the first nucleotide of the core recognition sequence. As indicated by SELEX experiments (24), the first nucleotide of the optimal RNA sequences for hnRNP K was always a U, while it could be U or A for PCBP2. Corresponding to this SELEX study, we show that PCBP2 KH1 can recognize RNA and DNA sequences with an A or U/T preceding the poly(C) patch (Figure 3). Superposition of the $^{15}$N-HSQC spectra of PCBP2 KH1 in complex with 5’-AACCCT-3’ and 5’-TCCCCA-3’ showed that they were virtually identical, indicating that the binding of these two DNAs should be rather similar. In the hnRNP K KH3-DNA complex, the first T residue participated exclusively in hydrophobic interactions with the α-methylene groups of Gly26 and Gly30, the α-methylene group of Ser27, and the aromatic ring of Tyr84 (45). In PCBP2 KH1 domain, these residues correspond to Gly26, Gly30, Ser27, and Asp82 (Figure 1b), with Asp82 versus Tyr84 as the only difference. This difference might be significant due to the very different nature of tyrosine and aspartic acid. We also notice that a stretch of residues immediately N-terminal to Gly26 shows the least conservation between the two domains (Figure 1b; HGKEV in PCBP2 KH1 versus PKDLA in hnRNP K KH3). This may also contribute to a possible difference in local structures, resulting in the more generous binding propensity of the PCBP2 KH1 domain.
Prior to this report, studies on the biological functions of PCBP2 have heavily focused on its RNA-binding ability, with only one report (to the best of our knowledge) suggesting its possible involvement in mechanisms associated with DNA-recognition. In the present study, we have established the ability of PCBP2 KH1 domain to recognize a couple of poly(C) DNA sequences specifically, including the C-rich strand of human telomeric repeats. Comparison with binding to the RNA counterpart sequences unequivocally demonstrates that PCBP2 KH1 domain utilizes not only the same nucleic acid binding platform but also very similar interactions to recognize both the RNA and DNA targets. This is in contrast to some other dual-specificity nucleic acid binding proteins such as the prototypic *Xenopus* TFIIIA, which recognizes RNA and DNA targets with different motifs; the structures of the RNA and DNA targets being recognized are also different (49).

So far, structures of several KH domains in complex with various DNA or RNA constructs have been solved by NMR or X-ray crystallography. Besides hnRNP K KH3 (Braddock et al., 2002a), these include FBP KH3 and KH4 recognizing 5’-TTTT-3’ and 5’-TATTC-3’, respectively (48), nova2 KH3 recognizing 5’-UCAY-3’ (Y=pyrimidine) presented within the loop region of an RNA stem-loop (44), and Splicing factor 1 (SF1) KH domain recognizing 5’-UAAC-3’ within a single-stranded intron branch point sequence (47). Although the nucleic acid targets and the detailed interactions dictating specificities are varied, a number of common features emerge. First, all of the KH domains adopt the same overall topology and use a common groove formed by the first two α helices, the second β strand, and the two flanking loops for nucleic acid binding; the floor of the groove is hydrophobic, and the ridges on both sides are hydrophilic and charged. Secondly, the core recognition sequences are single-stranded and extended, with the 5’-end positioned at the top end of the binding groove (referenced to the
orientation shown in Figure 6); the sugar-phosphate backbone is located on the left side of the groove, forming neutralizing electrostatic interactions with a stretch of positively charged residues on the left-side ridge; the bases of the nucleic acids point towards the center and right side of the groove. While hydrophobic interaction between the nucleic acid bases and the hydrophobic floor of the groove is of general importance for binding, specificity is defined by case-specific hydrogen bonding interactions between the functional groups of the bases and the protein side chain and backbone. Thirdly, in all complexes, the core recognition sequence is 4-5 residues in length, which is dictated by the length of the binding groove. Finally, in both cases of hNRNP K KH3 and nova2 KH3, no significant difference is observed between the structures of the protein domain in its free form and DNA/RNA bound form (backbone RMSD ~1.0 Å and ~0.8 Å, respectively). Based on our detailed analysis of the nucleic acid binding groove of PCBP2 KH1 domain and characterization of its interactions with various DNA and RNA targets, we have either confirmed or gained strong evidence to support most of these common features in our system of PCBP2 KH1–nucleic acid interactions. Of course, confirmation of the molecular details that define specificity await determination of several representative complex structures.

In essence, the interactions between PCBP2 KH1 domain and human telomere/telomerase DNA/RNA we established in the present study imply that proteins of the PCBP family may participate in regulation of telomere and telomerase activities through specific binding to the C-rich sequences of telomeric DNA and telomerase RNA. Of course, we could speculate that since the PCBPs have three KH domains, different KH domains in the same molecule may possibly bind to telomeric DNA and telomerase RNA template simultaneously. In fact, should any poly(C)-specific binding activity be required in functions related to telomere and telomerase biology, the ubiquitous PCBP proteins should come to mind as the best candidates, because these
proteins are simply the major poly(C)-specific binding proteins inside mammalian cells. Adding to the credibility of involvement of PCBPs in telomere/telomerase regulation, note that hnRNPs (the PCBPs are also members of the hnRNP family) are known to play roles in telomere biology (50). hnRNPs A1, C1/C2, and D are capable of interacting with the human telomerase holoenzyme; hnRNPs A1, A2-B1, D, and E and hnRNP homologous proteins from other organisms can associate with the single-stranded telomeric sequence (G-rich overhang). hnRNPs are integral components of the nuclear matrix. The nuclear matrix is a putative attachment site for telomeres, so hnRNPs and telomeres are in close proximity if not directly associated. hnRNP C1/C2 binds directly and specifically to a six-nucleotide U-rich tract just 5’ to the hTR template region. Since hnRNPs are often found in complex with each other, it is possible that other hnRNPs would bind to the hTR template region.

Acknowledgements — Partial support for this work was provided by National Institutes of Health grant AI46967 (T.L.J.). We acknowledge use of the UCSF Computer Graphics Laboratory, which is supported by NIH grant P41 RR-01081.
REFERENCES

1. Makeyev, A. V., and Liebhaber, S. A. (2002) RNA 8, 265-278

2. Gamarnik, A. V., and Andino, R. (2000) J Virol 74, 2219-2226

3. Weiss, I., and Liebhaber, S. (1995) Mol. Cell. Biol. 15, 2457-2465

4. Chkheidze, A. N., Lyakhov, D. L., Makeyev, A. V., Morales, J., Kong, J., and Liebhaber, S. A. (1999) Mol. Cell. Biol. 19, 4572-4581

5. Stefanovic, B., Hellerbrand, C., Holcik, M., Briendl, M., Aliebhaber, S., and Brenner, D. (1997) Mol. Cell. Biol. 17, 5201-5209

6. Lindquist, J. N., Kauschke, S. G., Stefanovic, B., Burchardt, E. R., and Brenner, D. A. (2000) Nucl. Acids. Res. 28, 4306-4316

7. Paulding, W. R., and Czyzyk-Krzeska, M. F. (1999) J. Biol. Chem. 274, 2532-2538

8. Czyzyk-Krzeska, M. F., and Bendixen, A. C. (1999) Blood 93, 2111-2120

9. Ostareck-Lederer, A., Ostareck, D. H., Standart, N., and Thiele, B. J. (1994) EMBO J. 13, 1476-1481

10. Ostareck, D. H., Ostareck-Lederer, A., Wilm, M., Thiele, B. J., Mann, M., and Hentze, M. W. (1997) Cell. 89, 597-606

11. Ostareck, D. H., Ostareck-Lederer, A., Shatsky, I. N., and Hentze, M. W. (2001) Cell 104, 281-290

12. Blyn, L. B., Swiderek, K. M., Richards, O., Stahl, D. C., Semler, B. L., and Ehrenfeld, E. (1996) Proc Nat Acad Sci USA 93, 11115-11120

13. Blyn, L. B., Towner, J. S., Semler, B. L., and Ehrenfeld, E. (1997) J Virol 71, 6243-6246

14. Gamarnik, A. V., and Andino, R. (1997) RNA 3, 882-892

15. Gamarnik, A. V., and Andino, R. (1998) Gene Dev 12, 2293-2304

16. Parsley, T. B., Towner, J. S., Blyn, L. B., Ehrenfeld, E., and Semler, B. L. (1997) RNA 3, 1124-1134
17. Andino, R., Rieckhof, G. E., and Baltimore, D. (1990) *Cell* 63, 369-380

18. Du, Z., Ulyanov, N. B., Yu, J., Andino, R., and James, T. L. (2004) *Biochemistry* 43, 5757-5771

19. Waggoner, S. A., and Liebhaber, S. A. (2003) *Mol Cell Biol.* 23, 7055-7067

20. Zhu, J., and Chen, X. (2000) *Mol. Cell. Biol.* 20, 5602-5618

21. Tomonaga, T., and Levens, D. (1996) *Proc. Nat. Acad. Sci. USA* 93, 5830-5835

22. Lacroix, L., Lienard, H., Labourier, E., Djavaheri-Mergny, M., Lacoste, J., Leffers, H., Tazi, J., Helene, C., and Mergny, J.-L. (2000) *Nucl. Acids. Res.* 28, 1564-1575

23. Bandiera, A., Tell, G., Marsich, E., Scaloni, A., Pocsfalvi, G., Akindahunsi, A. A., Cesaratto, L., and Manzini, G. (2003) *Arch Biochem Biophys.* 409, 305-314

24. Thisted, T., Lyakhov, D. L., and Liebhaber, S. A. (2001) *J. Biol. Chem.* 276, 17484-17496

25. Dejgaard, K., and Leffers, H. (1996) *Eur J Biochem* 241, 425-431

26. Milligan, J. F., and Uhlenbeck, O. C. (1989) in *Meth. Enzym.* (Dahlberg, J. E., and Abelson, J. N., eds) Vol. 180, pp. 51-62, Academic Press, New York

27. Du, Z., Yu, J., Andino, R., and James, T. L. (2003) *Biochemistry* 42, 4373-4383

28. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomolec. NMR* 6, 277-293

29. Goddard, T. D., and Kneller, D. G. (1998), 3.0 Ed., University of California, San Francisco

30. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson.* 85, 393-399

31. Bax, A., and Ikura, M. (1991) *J. Biomolec. NMR* 1, 99-104

32. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc.* 114, 6291-6293

33. Grzesiek, S., and Bax, A. (1992) *J. Magn. Res.* 96, 432-440

34. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) *J. Magn. Reson.* 88, 425-431
35. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc.* **115**, 7772-7777
36. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13**, 289-302
37. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) *J. Magn. Res.* **160**, 66-74
38. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomolec. NMR* **8**, 477-486
39. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) *Mol Graphics* **6**, 13-27
40. Huang, C. C., Couch, G. S., Pettersen, E. F., and Ferrin, T. E. (1996) *Pacific Symposium on Biocomputing* **1**, 724
41. Silvera, D., Gamarnik, A. V., and Andino, R. (1999) *J Biol Chem* **274**, 38163-38170
42. Baber, J. L., Libutti, D., Levens, D., and Tjandra, N. (1999) *J Mol Biol* **289**, 949-962
43. Lewis, H. A., Chen, H., Edo, C., Buckanovich, R. J., Yang, Y. Y., Musunuru, K., Zhong, R., Darnell, R. B., and Burley, S. K. (1999) *Structure Fold Des* **7**, 191-203
44. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. (2000) *Cell* **100**, 323-332
45. Braddock, D. T., Baber, J. L., Levens, D., and Clore, G. M. (2002) *EMBO J.* **21**, 3476-3485
46. Musco, G., Stier, G., Joseph, C., Morelli, M. A. C., Nilges, M., Gibson, T. J., and Pastore, A. (1996) *Cell* **85**, 237-245
47. Liu, Z. H., Luyten, I., Bottomley, M. J., Messias, A. C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., and Sattler, M. (2001) *Science* **294**, 1098-1102
48. Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D., and Clore, G. M. (2002) *Nature* **415**, 1051-1056
49. Cassiday, L. A., and Maher III, L. J. (2002) *Nucl. Acids. Res.* **30**, 4118-4126
50. Ford, L. P., Wright, W. E., and Shay, J. W. (2002) *Oncogene* **21**, 580-583
Table 1. NMR and structure determination statistics for PCBP2 KH1

| Restraint Type                        | Count  |
|---------------------------------------|--------|
| NOE-derived distance restraints       | 780    |
| Intra-residue                         | 136    |
| Sequential                            | 267    |
| Medium range (i+2 to i+5)             | 224    |
| Long range (>i+5)                     | 153    |
| Total hydrogen bond restraints        | 70     |
| Total torsion angle restraints        | 115    |
| Dipolar coupling restraints (\(1^\text{D}_{\text{NH}}\)) | 66     |
| R.m.s. deviation from experimental restraints \(^b\) | 0.098 ± 0.004 |
| Distance restraints (Å) \(^c\)        | 0.098 ± 0.004 |
| Dihedral restraints (°)               | 0.000 ± 0.000 |
| Dipolar coupling restraints (Hz)      | 0.59 ± 0.04 |
| Deviations from idealized covalent geometry |        |
| Bonds (Å)                             | 0.0026 ± 0.0002 |
| Angles (°)                            | 0.205 ± 0.001  |
| Impropers (°)                         | 0.74 ± 0.06   |
| RMSD from mean structure (Å) \(^b\)  | 0.47    |
| Backbone                              | 0.47    |
| All non-hydrogen                      | 1.08    |
| Ramachandran plot analysis \(^d\)    |        |
| Residues in most favored regions (%)  | 84.8%   |
| Residues in additional allowed regions (%) | 11.3% |
| Residues in generously favored regions (%) | 3.2%   |
| Residues in disallowed regions (%)    | 0.8%    |

\(^a\)Includes two restraints per hydrogen bond.

\(^b\)Averaged over the 20 accepted structures.

\(^c\)No violation >0.6 Å.

\(^d\)Checked by PROCHECK_NMR (38)
FIGURE CAPTIONS

**FIG. 1.** (a) Schematic diagram of the domain structure of human poly(C) binding protein-2 (PCBP2). Similar domain structures are observed in other members of the PCBP family. (b) Sequence alignments of the KH domains from the known PCBP proteins PCBP1-4, and hnRNP-K. Alignments were carried out using the program ClustalX. The KH domains are listed in descending order of homology to the first KH domain of PCBP2 (PCBP2-KH1). The sequence shown for PCBP2-KH1 corresponds to residues 11-82 in the full-length protein. The actual NMR construct has two extra amino acids (MK) at the N-terminus and a (His)x6 tag at the C-terminus. (c) Summary of sequential and medium range NOEs and chemical shift indexing (CSI) data for the PCBP2-KH1 domain. In CSI, chemical shifts of amide proton, amide nitrogen, \( \alpha \) and \( \beta \) carbons were analyzed by the program PsiCSI (http://protinfo.compbio.washington.edu/psicsi) to predict the secondary structures; H-helix, E-extended (\( \beta \)-sheet), C-coil. Residue dipolar coupling constants (RDCs) for the amide H-N vectors were also shown. The different signs of the RDCs from residues within the \( \alpha \)-helices and \( \beta \)-strands reflect the relative orientation of these secondary structure elements.

**FIG. 2.** Structure of the PCBP2-KH1 domain determined using NOE and residual dipolar coupling restraints. (a) A view of the backbone superimposition of the 20 lowest energy structures. The three antiparallel \( \beta \)-sheets and the N-terminal sequence are colored magenta; the first, second, and third \( \alpha \)-helices are colored blue, cyan, and yellow respectively; the highly conserved GxxG loop and the variable loop are colored red and forest green respectively. (b) A ribbon representation of the lowest energy structure, with a similar viewing angle and identical coloring scheme as in (a).
**Fig. 3.** Complex formation between PCBP2 KH1 and various oligonucleotides as monitored by chemical shifts changes in $^{15}$N-HSQC spectra of PCBP2 KH1. (a) Overlay of three $^{15}$N-HSQC spectra of KH1: free (red), in complex with a 6-nt RNA 5’-UCCCCA-3’ (green), and in complex with a 6-nt DNA 5’-TCCCCCA-3’ (blue). The RNA sequence corresponds to one of the C-patches found in PCBP2 recognition motifs within 3’-UTRs of several mRNAs. The boxes highlight some noticeable differences in chemical shifts between RNA- and DNA-bound forms. Cross-peaks labeled with stars within the dashed rectangle near the upper-right corner are aliased cross-peaks from the arginine side-chain NH groups. Spectra showing the real chemical shifts of these peaks are shown within the rectangle at the upper-left corner. (b) Overlay of three TROSY-$^{15}$N-HSQC spectra of KH1: free (red), in complex with a 10-nt human telomeric DNA, 5’-CTAACCCTAA-3’ (blue), and its 10-nt RNA counterpart, 5’-CUAACCCUA-3’, which comprises the template region of human telomerase (green). The boxes highlight some noticeable differences in chemical shifts between RNA- and DNA-bound forms. Cross-peaks labeled with stars within the dashed rectangle near the upper-right corner are aliased cross-peaks from the arginine side-chain NH groups. The real chemical shifts of these peaks are similar to those shown in Figure 3a. Note that the RNA and DNA complexes show similar chemical shift changes although the aliased peaks do not superimpose due to different spectral widths used in acquiring the HSQC spectra. (c) Overlay of three TROSY-$^{15}$N-HSQC spectra of KH1: free (red), in complex with a 19-nt cloverleaf stem-loop B RNA (green) and a 34-nt IRES loop B RNA (blue) from the 5’UTR of Poliovirus-1. (d) Overlay of the TROSY-$^{15}$N-HSQC spectra of free KH1 (red) and KH1 in complex with a 6-nt C-rich human telomeric DNA, 5’-AACCCT-3’ (blue). Assignments for some of the cross-peaks experiencing significant chemical shift changes are labeled.
**Fig. 4.** Chemical shift differences between the free form and human telomeric DNA (6 nt, 5’-AACCCT-3’)-bound form of PCBP2-KH1. Chemical shift differences are calculated as: $\Delta \delta = \delta_{\text{bound}} - \delta_{\text{free}}$, where $\delta$ is the chemical shift value in ppm. The top and middle panels are for amide proton and nitrogen, respectively. The bottom panel shows the composite difference calculated as $((\Delta \delta_H)^2 + (\Delta \delta_N/5)^2)/2)^{1/2}$.

**Fig. 5.** Mapping the composite chemical shift differences as shown in the bottom panel of Fig. 4 onto a surface representation of the lowest energy PCBP2 KH1 structure. Different colorings of the protein residues are employed to indicate the magnitude of the composite differences. Residues with no changes or small changes (composite shift values < 0.15) are colored dark grey; residues with medium changes (composite shift values between 0.15 and 0.30) are colored orange; residues experiencing the biggest changes (composite shift values > 0.3) are colored red. (A) Viewed from a similar angle as in Figures 2a and 2b; (B) viewed from the back of Fig. (A). Note that the red spot is seen through a deep and narrow hole, which should not be accessible by the nucleic acids from this side of the structure.

**Fig. 6.** A view of the nucleic acid binding groove in a surface representation of PCBP2 KH1. Hydrophobic residues are colored green; positively and negatively charged residues are colored blue and red, respectively; hydrophilic noncharged residues are colored yellow. Note that the residues on the floor of the binding groove are exclusively hydrophobic.
Figure 1
Figure 3
Figure 4
Figure 5
Figure 6
Figure S1 2D $^{1}$H-$^{15}$N HSQC spectrum of a $^{13}$C/$^{15}$N doubly labeled PCBP2 KH1 sample. Note that residue numbering starts from the first residue of the NMR construct as residue 1, therefore all residue numbers need to add a value of 8 to agree to the numbering used in the main text, which corresponds to residue numbering in the full-length PCBP2 protein.
Figure S2: Example of a backbone sequential walk from S27 to C46 (S35 to C54 in the numbering system used in the main text) in a 3D HNCA experiment of a $^{13}\text{C}^{15}\text{N}$ doubly labeled PCBP2 KH1 sample.
**Figure S3**: Sequence alignments generated by the program CLUSTAL X of the two KH domains (KH1 and KH2) of MCG10 (also known as PCBP4b, MCG10 is used here to avoid confusion with the major isoform of PCBP4), to the three KH domains of PCBP4 (KH1, KH2, and KH3), the third KH domain of hnRNP K, and PCBP2 KH1. Residue numbering in the “ruler” field is for PCBP2 KH1 as in the main text.
Figure S4: Superimposition of the lowest energy structures of PCBP2 KH1 (in red) and hnRNP K KH3 (in blue). Pair-wise RMSD on backbone heavy atoms (excluding residues in the variable loops which have different numbers of residues) between the two structure is ~1.4 Å.
Ramachandran Plot
kh1 (20 models)

Plot statistics

- Residues in most favoured regions [A,B,L] 1068 84.8%
- Residues in additional allowed regions [a,b,l,p] 142 11.3%
- Residues in generously allowed regions [-a,-b,-l,-p] 40 3.2%
- Residues in disallowed regions 10 0.8%
- Number of non-glycine and non-proline residues 1260 100.0%
- Number of end-residues (excl. Gly and Pro) 40
- Number of glycine residues (shown as triangles) 140
- Number of proline residues 40
- Total number of residues 1480

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.
Model numbers shown inside each data point.
Figure S5: Ramachandran plot of the NMR ensemble of PCBP2 KH1 produced by the quality checking program PROCHECK-NMR.

| Table: Backbone and Sidechain Assignment of KH1 |
|------------------------------------------------|
| No. | HN | HA1 | HA2 | CA | CB | CD | CE | CGI | CG2 | HB | HD1 | HD2 | HE | HE1 | HG2 |
|-----|----|-----|-----|----|----|----|----|-----|-----|----|-----|-----|----|-----|-----|
| K2  | 56.5 | 32.5 | 29.2 | 42.0 | 25.0 |
| N4  | 6.64 | 119.0 | | 53.4 | 28.8 |
| V4  | 4.00 | 119.7 | | 63.0 | 33.3 | 31.7 | 26.7 |
| T5  | 8.15 | 118.7 | | 61.6 | 70.5 | 21.3 |
| L6  | 5.63 | 126.4 | | 54.2 | 45.4 | 26.9 |
| T7  | 6.03 | 115.3 | 3.14 | 62.1 | 21.5 | 3.89 |
| I8  | 9.42 | 127.1 | | 57.4 | 46.7 | 37.3 | 17.5 |
| E9  | 9.24 | 125.8 | | 54.9 | | |
| L10 | 9.38 | 122.5 | 5.26 | 53.5 | 44.2 | 25.3 | 25.9 |
| L11 | 6.66 | 124.4 | | 58.9 | 44.4 |
| M12 | 8.99 | 124.5 | | 58.3 | 33.2 |
| H13 | 9.04 | 119.9 | | 56.1 | 36.5 |
| G14 | 8.92 | 111.8 | | 48.3 |
| R15 | 4.19 | 59.1 | 32.2 | 39.3 | 25.1 | 1.87 | 1.68 | 2.97 | 1.48 |
| E16 | 7.61 | 118.8 | 4.19 | 58.6 | 36.5 | 37.4 | 2.36 | 2.46 |
| V17 | 7.63 | 116.9 | 3.61 | 66.3 | 23.7 | 22.1 | 2.12 | 0.58 |
| G18 | 6.13 | 124.4 | 3.83 | 27.8 | 47.8 |
| S19 | 7.66 | 116.9 | 4.19 | 61.2 | 63.1 | 3.94 |
| E20 | 7.16 | 119.4 | | 63.7 | 38.2 | 29.1 | 17.2 |
| L21 | 7.09 | 115.3 | 3.82 | 64.9 | 27.9 | 17.2 | 1.99 | 0.84 | 1.28 | 1.07 |
| G22 | 7.66 | 126.1 | 3.96 | 4.31 | 44.3 |
| R23 | 6.02 | 122.7 | 4.28 | 56.6 | 32.2 | 29.0 | 2.51 | 1.86 | 1.75 | 3.02 | 1.46 |
| K24 | 9.36 | 120.9 | 4.28 | 56.3 | 31.6 | 29.2 | 2.84 | 1.78 | 1.71 | 3.02 | 1.48 |
| G25 | 7.67 | 124.5 | | 46.4 |
| F26 | 4.14 | 59.1 | 26.4 | 36.0 | 2.18 | 2.41 |
| S27 | 8.03 | 115.8 | 4.53 | 61.2 | 62.3 | 3.89 |
| V28 | 7.73 | 122.1 | 4.06 | 65.1 | 31.2 | 22.4 | 19.9 | 2.20 | 1.36 | 0.97 |
| R29 | 7.66 | 122.6 | 4.07 | 60.6 | 32.8 | 29.9 | 33.2 | 1.95 | 1.73 | 3.03 | 1.96 |
| E30 | 7.49 | 119.2 | 4.32 | 59.2 | 32.4 | 29.4 | 25.0 | 2.03 | 1.75 | 3.02 | 1.90 |
| M31 | 6.13 | 117.0 | 4.23 | 59.6 | 32.4 | 2.95 |
| R32 | 6.63 | 122.6 | 4.06 | 60.3 | 31.1 | 38.9 | 29.3 | 3.29 |
| E33 | 8.05 | 119.1 | 4.02 | 59.3 | 36.3 | 23.3 |
| E34 | 6.64 | 115.4 | 4.24 | 50.4 | 36.3 | 3.52 | 2.32 | 2.46 |
| S35 | 6.10 | 111.0 | 4.31 | 60.7 | 65.0 | 3.91 |
| G36 | 7.76 | 109.5 | 3.87 | 4.37 | 45.9 |
| A37 | 8.53 | 122.0 | | 52.8 | 18.5 | 1.19 |
| R38 | 8.86 | 120.7 | | 56.3 | 36.4 | 41.0 | 27.6 |
| D39 | 8.11 | 123.4 | 5.03 | 60.6 |
| N40 | 8.84 | 135.7 | 5.17 | 52.4 | 42.5 |
Table S1: Chemical shift assignment table PCBP2 KH1. Note that residue numbering starts from the first residue of the NMR construct as residue 1, therefore all residue numbers need to add a value of 8 to agree to the numbering used in the main text, which corresponds to residue numbering in the full-length PCBP2 protein.

| No. | H   | HN1 | HN2 | CA  | CB  | CD  | CE  | CG1 | CG2 | HB  | HD1 | HD2 | HF  | HG1 | HG2 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| E1  | 5.86| 123.5| 4.85| 60.4| 26.5| 17.9| 1.73| 0.80| 0.97|
| E2  | 8.71| 124.4| 4.37| 59.5| 63.6| 3.69|
| E4  | 4.65| 120.7| 4.39| 36.1| 36.2| 2.29| 2.36|
| E4  | 4.47| 109.9| 3.93| -0.20| 45.1|
| N4  | 5.54| 190.0| 57.9| 37.9|
| C6  | 7.68| 118.8| 58.1|
| C7  | 6.51| 124.4| 32.4| 27.7|
| E40 | 7.51| 106.1| 55.3| 31.2| 36.7| 1.82|
| E40 | 4.75| 120.7| 38.6|
| E5  | 8.53| 120.7| 59.8| 28.4| 17.8|
| E1  | 9.49| 130.7| 5.00| 60.1| 28.0| 17.1|
| F32 | 9.37| 124.7| 3.03| 62.8| 21.9| 3.93|
| L35 | 9.53| 127.0| 3.19| 52.8| 45.2| 25.6|
| A54 | 4.86| 123.5| 5.73| 50.8| 23.3| 1.32|
| G55 | 4.16| 104.2| 44.7|
| F56 | 6.25| 32.4| 27.4|
| T57 | 8.83| 135.5| 3.53| 67.3| 68.1| 23.5| 3.46| 1.41|
| N58 | 8.70| 106.3| 4.54| 56.8| 37.4| 2.69|
| A59 | 7.26| 123.5| 4.49| 54.8| 35.5| 1.43|
| B60 | 7.89| 117.7| 65.0| 36.4| 17.6|
| R61 | 4.66| 118.6| 42.9| 38.8| 3.29| 3.42|
| K62 | 7.79| 118.2| 59.7| 22.8| 29.6| 26.0|
| A63 | 8.46| 120.8| 55.5| 16.5|
| F64 | 8.95| 117.7| 62.4| 46.1|
| A65 | 5.32| 139.5| 55.5| 17.6|
| M66 | 7.74| 116.4| 59.7| 56.2| 35.5|
| E67 | 5.93| 122.4| 66.3| 29.0| 15.9|
| N68 | 6.83| 138.2| 63.9| 26.8| 18.0|
| D69 | 8.00| 199.5| 57.8| 46.5|
| K70 | 7.64| 118.6| 57.6| 31.5| -0.5|
| L71 | 8.16| 120.4| 4.11| 57.8| 41.8| 22.3| 25.8| 1.86| 0.77| 0.84| 1.46|
| L72 | 8.43| 120.2| 4.39| 59.7| 26.5| 36.4|
| E73 | 7.93| 120.8| 4.01| 59.2| 29.8| 36.2| 2.23| 2.40|
| D74 | 7.84| 118.5| 4.54| 56.6| 41.3| 2.77|
| H75 | 7.76| 120.4| 57.3| 42.5| 23.7| 25.6|
| H76 | 8.64| 118.5| 57.8|
Specific recognition of the C-rich strand of human telomeric DNA and the RNA template of human telomerase by the first KH domain of human poly(C)-binding Protein-2

Zhihua Du, Jinghua Yu, Yihong Chen, Raul Andino and Thomas L. James

J. Biol. Chem. published online August 24, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M405371200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2004/09/01/M405371200.DC1