Crystal Structure of the First KH Domain of Human Poly(C)-binding Protein-2 in Complex with a C-rich Strand of Human Telomeric DNA at 1.7 Å*

Received for publication, July 27, 2005, and in revised form, September 14, 2005. Published, JBC Papers in Press, September 25, 2005, DOI 10.1074/jbc.M508183200

Zhihua Du†, John K. Lee‡, Richard Tjhen§, Shang Li§, Hu Pan†, Robert M. Stroud†, and Thomas L. James††

From the Departments of †Pharmaceutical Chemistry, and ‡Biochemistry and Biophysics, University of California, San Francisco, California 94143-2280

Recognition of poly(C) DNA and RNA sequences in mammalian cells is achieved by a subfamily of the KH (hnRNP K homology) domain-containing proteins known as poly(C)-binding proteins (PCBPs). To reveal the molecular basis of poly(C) sequence recognition, we have determined the crystal structure, at 1.7-Å resolution, of PCBP2 KH1 in complex with a 7-nucleotide DNA sequence (5'-AACCTA-3') corresponding to one repeat of the human C-rich strand telomeric DNA. The protein-DNA interaction is mediated by the combination of several stabilizing forces including hydrogen bonding, electrostatic interactions, van der Waals contacts, and shape complementarities. Specific recognition of the three cytosine residues is realized by a dense network of hydrogen bonds involving the side chains of two conserved lysines and one glutamic acid. The co-crystal structure also reveals a protein-protein dimerization interface of PCBP2 KH1 located on the opposite side of the protein from the DNA binding groove. Numerous stabilizing protein-protein interactions, including hydrophobic contacts, stacking of aromatic side chains, and a large number of hydrogen bonds, indicate that the protein-protein interaction interface is most likely genuine. Interaction of PCBP2 KH1 with the C-rich strand of human telomeric DNA suggests that PCBPs may participate in mechanisms involved in the regulation of telomere/telomerase functions.

K-homology domain (KH2 domain, originally identified in hnRNP-K) is one of the most frequently occurring and conserved nucleic acid-binding protein motifs. So far, a large number of KH domain-containing proteins have been identified in a wide variety of species ranging from bacteria to human. These KH domain proteins assume a wide spectrum of biological functions, including transcriptional and translational controls, mRNA stabilization, and mRNA splicing, among others. Different KH domains possess quite different nucleic acid binding specificities. Unveiling how different KH domains interact specifically with their nucleic acid targets and how these interactions contribute to the complex regulatory processes is of central importance to a better understanding of how KH domain proteins function.

One of the most distinctive nucleic acid binding specificities achieved by the KH domains is manifested by a subfamily of KH domain-containing proteins known as poly(C)-binding proteins (PCBPs). As implied by the family name, PCBPs recognize poly(C) RNA or DNA sequences with high affinity and specificity (for reviews, see Refs. 1 and 2). To date, five evolutionarily related PCBPs have been identified in mammalian cells: PCBP1–4 (also known as hCP1–4; PCBP1 and -2 are also known as hnRNP E1 and E2) and hnRNP K. Each PCBP contains three KH domains: 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 KH domains (Fig. 1A). In general, corresponding KH domains share a higher degree of homology than KH domains within each protein (Fig. 1B). No other discernable nucleic acid binding motif is present in PCBPs; the KH domains are responsible for the ability of the PCBPs to interact with poly(C) sequences.

The established examples of functional roles carried out by PCBPs indicate that these proteins are key mediators in a number of important cellular processes (1). Binding of PCBP1 or PCBP2 to target RNA sequences harboring tandem poly(C) stretches 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. In the case of α-globin mRNA, it was established that the stoichiometry of the RNA-protein complex is 1:1, and a minimum RNA sequence of 20 nt (5'-CCCACGGGGCCUCUCCCC-3') was able to form the complex (9). Interaction of two PCBPs, hnRNP K and PCBP1 or -2, with a C-rich sequence within the 3'-UTR of some mRNAs can also result in translational silencing, as seen in 15-lipoxygenase (LOX) mRNA (10–12). A recent study identified 160 mRNA species that associate in vivo with PCBP2 from a human hematopoietic cell line (13), suggesting that the contribution of PCBPs in post-transcriptional regulation of cellular genes may be far more profound than currently known.

Besides cellular RNAs, PCBPs also participate in regulating critical viral RNA functions. Binding of PCBP1 or -2 to two cis-acting C-rich sequence containing RNA elements within the 5'-UTR of Poliovirus mRNA (which is also the genomic RNA) is critical for regulation of cap-independent translation and replication of the viral RNA (14–18).

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 C-rich sequence in the promoter of the human c-myec gene activates transcription (19). It was also shown that hnRNP K and PCBP1 could recognize the C-rich strand of human telomeric DNA with high affinity in vitro (20, 21); whether such an interaction is functionally significant is a subject for further biochemical/biological investigations.
It should be noted that although PCBP functions are diverse, they are nearly all dependent on the ability of the KH domains to recognize single-stranded C-rich RNA or DNA sequences with high specificity and affinity. To reveal the molecular basis of KH domain-poly(C) DNA/RNA interaction, we have previously used NMR to determine the solution structure of the first KH domain (KH1) from human PCBP2, and characterize its interaction with various DNA/RNA molecules (22). In this study, we report the 1.7-Å resolution crystal structure of the PCBP2 KH1 domain in complex with a seven-nucleotide single-stranded DNA sequence (5′/H11032-AACCCTA-3′/H11032) corresponding to one repeat of the C-rich strand of human telomeric DNA (htDNA). The structure shows that PCBP2 KH1 makes substantial contacts with four nucleotides (5′-ACCC-3′, the core recognition sequence). Interestingly, each of the three cystosines is specifically recognized by a network of strong hydrogen bonds to the Watson-Crick positions involving the side chain functional group of a particular amino acid. Another outstanding feature of the PCBP KH1 domain is the presence of a protein-protein dimerization interface located on the other side of KH1 domain from the DNA binding interface. A large number of protein-protein interactions, including hydrophobic contacts, stacking of aromatic side chains, and numerous hydrogen bonds, stabilize the dimerization interface, a feature not observed in any other KH domain structure previously characterized. Insights into mechanisms of PCBP functions, in the context of comparison with other available KH domain-nucleic acids complex structures, are discussed.

**EXPERIMENTAL PROCEDURES**

**Sample Preparation and Crystallization—**N-terminal His-tagged PCBP2 KH1 was overexpressed in BL21(DE3) strain of *Escherichia coli* (Stratagene). For Se-Met-labeled protein, the bacteria were grown in M9 minimal medium until they reached an *A*<sub>600</sub> of 0.6–0.8, whereupon leucine, isoleucine, lysine, phenylalanine, threonine, and valine were added to the culture to inhibit methionine biosynthesis. After 15 min, L-seleno-methionine (50 mg/liter) was added, followed by isopropyl-D-thiogalactopyranoside to a final concentration of 0.4 mM to induce expression of the Se-Met-labeled protein. After purification by Ni-nitritotriacetic acid resin (Qiagen), the His tag was removed by the TAGzyme system from Qiagen. Crystals of the PCBP2 KH1-DNA (5′-AACCTA-3′) complex were obtained by hanging drop vapor diffusion against 25% polyethylene glycol 8000, 100 mM sodium acetate, 100 mM sodium cacodylate, pH 6.1, at 22 °C. The protein concentration was about 250 μM with a 1:1.2 protein:DNA ratio. Orthorhombic crystals grew to useful size within 1 day with diffraction to 1.7 Å. The crystals are in space group P2₁2₁2₁(*a* = 65.60 Å, *b* = 115.18 Å, *c* = 45.53 Å), with four protein-DNA complexes in one asymmetric unit.

**Data Collection, Structure Determination, and Refinement—**A single SAD data set was collected at the peak wavelength of the selenium K absorption edge from a single frozen selenomethionine-containing crystal using Beamline 8.3.1 of the Advanced Light Source (ALS) at Berkeley National Laboratory. Diffraction intensities were integrated and reduced by using the program DENZO and were scaled by using SCALEPACK (23) (TABLE ONE). All twelve selenium atoms from the four protein molecules in an asymmetric unit were located by CNS (24). An interpretable electron density map was obtained after solvent flattening. The model was built by MOLOC (25) and refined in CNS to an *R*<sub>f</sub> of 21.5% (*R<sub>f</sub>free = 23.8%). The final model includes all of the protein residues 11–82 (the PCBP2 numbering is used), four DNA molecules (two molecules have all of the seven nucleotides built; the other...
Structure of KH Domain in Complex with Human Telomeric DNA

| Crystallographic refinement statistics |
|---------------------------------------|
| Crystal data                          |
| Space group                           | P2₁,2,2 |
| Unit cell dimensions (Å)              | a = 66.60, b = 115.18, c = 45.53 |
| x²                                   | 4 |
| X-ray data collection statistics      |
| Wavelength (Å) (SeMet SAD peak)       | 0.979594 |
| Resolution (Å)                        | 60.0–1.70 |
| Observed reflections                  | 314,041 |
| Unique reflections                     | 39,246 |
| Completeness (last shell) (%)         | 99.6 (97.6) |
| R_merge (%) (last shell)              | 9.2 (88.7) |
| F/σ (last shell)                     | 15.3 (2.1) |
| Crystallographic refinement statistics |
| Resolution (Å)                        | 45.5–1.70 |
| Reflections in working set            | 39,092 |
| Reflections in test set (10.0%)       | 1,935 |
| Rmerge (%)                            | 23.5 |
| Rfree (%)                             | 25.6 |
| Phasing power                         | 2.20 |
| Figure of merit (after solvent flattening) | 0.41 (0.92) |
| RMSD bonds (Å)                        | 0.0046 |
| RMSD angles (°)                       | 1.16 |
| Mean B-factors (Å²)                   | 22.7 |

* x is the number of equivalent structures per asymmetric unit.

** Rmerge = ΣᵢΣⱼΣk(Fobsᵢⱼₖ) − [ΣᵢΣⱼΣk(Fobsᵢⱼₖ)]/ΣᵢΣⱼΣk(Fobsᵢⱼₖ), where Fobsᵢⱼₖ is the measured intensity of hkl reflection, and ⟨Fobs⟩ is the mean of all measured intensity of hkl reflection.

† Rmerge = ΣᵢΣⱼΣk(Fcalcᵢⱼₖ) − [ΣᵢΣⱼΣk(Fcalcᵢⱼₖ)]/ΣᵢΣⱼΣk(Fcalcᵢⱼₖ), where Fcalcᵢⱼₖ is the observed structure factor amplitude, and Fcalcᵢⱼₖ is the structure factor calculated from the model. Rmerge is computed in the same manner as is Rmerge with the test set of reflections (10%).

RESULTS

Overall Structure of the PCBP2 KH1-htDNA Complex—There are four PCBP2 KH1-htDNA complexes in one asymmetric unit, labeled as complex A, B, C, and D in Fig. 2; the two complexes A and B form a dimer; the two complexes C and D form another dimer. Electron density is clearly present for every protein residue in all four complexes. All seven DNA residues were built in complexes A and C; five DNA residues were built for complexes B and D, but the two terminal residues were not built because of a lack of electron density. Structures of the four complexes are otherwise very similar (non-crystalllographic symmetry averaging was not applied; average pair-wise RMSD is ~0.22 Å), including details in molecular recognition.

The overall structure of the complex (using complex A as a representative) is depicted in Fig. 3. A and B. The DNA-bound form of PCBP2 KH1 is very similar to the NMR structure of the free protein we previously determined (22). The structure consists of three α-helices and three β-strands arranged in the order β1-α1-α2-β2-β3-α3. The evolutionarily conserved invariable Gly³⁰-Lys³¹-Lys³²-Gly³³ loop is located between α1 and α2; the variable loop (Ser⁵⁰-Pro⁵⁵) is between β2 and β3. 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. Hydrophobic interactions seem to play an important role in the packing of the structural elements to form a compact global fold; the residues in the core are exclusively hydrophobic.

The C-rich strand of human telomeric DNA binds to PCBP2 KH1 in a groove defined by the juxtaposition of two α-helices (α1 and α2), two β-strands (β2 and β3; only one residue from β3, Arg⁵⁷, participates in direct contact with the DNA), and two connecting loops (the GKKG loop and the variable loop). This binding groove is consistent with what we previously determined by NMR chemical shift perturbations (22); note that chemical shifts are reported in the Supplementary Materials in Ref. 22. The limited length of the groove only allows the accommodation of four DNA residues (Adenine-2 to Cytosine-5) as the core recognition motif. Other flanking DNA residues are not involved in direct contact with the protein (Fig. 3A).

3 W. L. DeLano (2002) The PyMOL Molecular Graphics System on the World Wide Web, www.pymol.org.
Structure of KH Domain in Complex with Human Telomeric DNA

**Specific Recognition of the Core Sequence**—Recognition of the core motif, Ade-2 to Cyt-5, is achieved by the combination of several forces including hydrogen bonding, electrostatic interactions, van der Waals contacts, and shape complementarities. The human telomeric DNA binds in a groove running from the C-terminal region to the N-terminal region of PCBPD2 KH1 domain (Fig. 3A). The overall orientation of the DNA backbone is similar to that previously observed in other DNA/RNA-KH domain complexes (26–30), with the 5′- and 3′-ends interacting with the C- and N-terminal regions of the KH domain, respectively.

Two phosphate groups of the DNA participate in intermolecular hydrogen bonds (Fig. 4A). O1P of Cyt-3 accepts a hydrogen bond from the backbone amide of Lys32 within the conserved GKKG loop. This hydrogen bond formation explains the large NMR chemical shift of the Lys32 amide proton (downfield, 0.86 ppm) we previously observed (22). A water molecule forms a bridge between the phosphate group of Thy-6 and the backbone carbonyl of Ile57.

The DNA backbone roughly runs along the left ridge of the binding groove, where four positively charged residues (Lys32 and Gly36 from the invariable GKKG loop, Lys57 and Arg40 from helix α2) make close contacts with the phosphate groups of Cyt-3, Cyt-4, Cyt-5, and Thy-6 (Fig. 4A). Conservation of these positively charged residues suggests that electrostatic interactions may play a role in interaction with the DNA backbone.

The ribose and base moieties of the core recognition sequence fit nicely inside the binding groove (Fig. 4A). For Ade-2, no specific interactions are observed, but van der Waals contacts are mainly provided by Gly26, Ser27, and Lys31. Base-stacking with Ade-1 is observed. Binding at this position does not seem to be sequence-specific. For the three cytosines, most of their Watson-Crick functional groups of the bases point to the right, forming specific hydrogen bonds with protein residues from the β-strands (β2 and β3) and variable loop. The size of the groove may dictate the preference for pyrimidine bases at these positions. There are extensive hydrophobic contacts between the riboses/hydrophobic faces of the cytosine bases and the aliphatic side chains that dominate the floor of the binding groove (Fig. 4A). For Cyt-3, van der Waals contacts are provided by Val25, Gly26, and Ile29; for Cyt-4, the hydrophobic environment is created by Ile20, Val25, and Ile49. For Cyt-5, Ile47 is involved. These conserved hydrophobic residues are presumably important for the function of the KH domain. A single mutation of an isoleucine to an asparagine in the second KH domain of FMR1 (corresponding to Val36 in PCBPD2 KH1) is known to cause a particularly severe presentation of the mental retardation syndrome (31). In our hand, mutation of Val36 to an asparagine caused PCBPD2 KH1 domain to aggregate in the inclusion body; we were not able to refold the protein despite substantial efforts. It seems as though the conserved hydrophobic residues are somehow important for proper folding and/or solubility of PCBPD2 KH1 domain.

Specific recognition of the poly(C) sequence is achieved by an extensive network of hydrogen bonding interactions (Fig. 4, B and C). For Cyt-3, the side chain of Arg57 provides two hydrogen bond donors to the O2 group of Cyt-4. The O2 group of Cyt-4 forms hydrogen bonds to the backbone carbonyl oxygen of Ser50. Hydrogen bonds involving the N4 amino group of Cyt-3 entail an intrastrand hydrogen bond to the O1P phosphate group of Ade-2 and a water bridge to the backbone carbonyl oxygen atoms of Gly22 and Cys44. This network of hydrogen bonds provides a molecular mimicry of a guanine to form Watson-Crick-like interactions with Cyt-3, therefore defining the specificity for a cytosine at this position of the DNA sequence.

For Cyt-4, the N4 amino group of Cyt-4 forms two hydrogen bonds: one with the backbone carbonyl oxygen of Ile49, the other with a water which bridges to the amide group of Gly32. The N3 group of Cyt-4 forms a hydrogen bond with another water, which bridges to the amide of Ile49 and the O2 group of Cyt-5. Involvement of the amide group of Ile49 in a hydrogen bond explains the big downfield NMR chemical shift change (1.31 ppm) we previously observed for Ile49 amide proton upon complex formation with the htDNA (22). The O2 group of Cyt-4 forms hydrogen bonds with the side chain of Arg40 from helix α2. The extended conformation of the Arg40 side chain is stabilized by hydrogen bonds to the backbone carbonyl of Ile57 from strand β2. Intrastrand base stacking with Cyt-5 also contributes to defining the binding environment for Cyt-4.

For Cyt-5, the side chain of Glu51 from the variable loop forms two hydrogen bonds to the N4 and N3 groups. The conformation of the Glu51 side chain is further stabilized by a water bridge to the side chain carbonyl oxygen of Asn46. The base of Cyt-5 engages in stacking interactions with the base of Cyt-4 on one face and the base of an adenosine.
from a symmetry-related DNA strand (labeled as sym_Ade-7 in Fig. 4C) from an adjacent complex in the crystal lattice on the other face.

**Dimerization of the PCBP2-KH1 Domain**—Numerous interactions stabilize the formation of the PCBP2 KH1 homodimer (Fig. 5, A, B, and C). The two protein domains in the dimer are arranged in a head-to-toe manner, with the dimerization interface located on the opposite side of the protein from the DNA binding groove. The dimerization interface is defined by the anti-parallel positioning of the longest α-helix (α3) and β-strand (β1) in the protein domain. The molecular interactions mediating the dimerization of the KH domains are truly remarkable in that they are the kinds of interactions normally encountered in the folding of a compact, integral protein motif. There is a hydrophobic interior core defined by the hydrophobic side chains of Leu18, Ile16, Leu19 from the two β-strands (β1s), and Ile66, Phe72, Ile76, Leu79 from the α-helices (α3s). Stacking of the two Phe72 aromatic rings is also observed. Dimerization orients the two three-stranded antiparallel β-sheets of the monomers in such a way that a six-stranded antiparallel β-sheet is formed. Four generic (backbones only) hydrogen bonds stabilize the antiparallel arrangement of the two β1-strands. These include: two from Leu19 amide to Arg17 carbonyl oxygen, and two water bridges from Arg17 amide to Leu19 carbonyl oxygen. There are also quite a number of non-generic (involving side chain functional groups) intermolecular hydrogen bonds formed at the dimerization interface, including two from the side chain of Arg17 in molecule A to the side chain of Glu56 in molecule B, one from the side chain of Arg17 in molecule B to the side chain of Glu56 in molecule A, one from Val12 backbone in molecule A to the side chain of Glu56 in molecule B, one from the side chain hydroxyl of Thr65 in molecule B to Glu60 side chain in molecule A, and one water bridge from the amide of Thr65 in molecule A to the side chain of Glu60 in molecule B.

Formation of the dimer buries 1188 Å² of solvent-accessible surface area in each monomer. The molecular surface forming the dimerization interface is rather hydrophobic in nature (Fig. 5B). Such a big area of hydrophobic surface should provide a significant driving force for formation of the protein-protein interface. Whether the dimerization of the PCBP2 KH1 domain depicted in Fig. 5, A and C is biologically significant is not clear at the present time. However, the details about the driving forces for dimer formation as revealed by our crystal structure strongly suggest that the dimer should be very stable. KH domain dimers were also observed in the crystal structures of free and RNA-bound forms of NOVA2 KH3 (32), and DNA-bound form of hnRNP K KH3 (32). Only the dimerization interface of free NOVA2 KH3 is similar to that of PCBP2 KH1 in terms of the involvement of both strand β1 and helix α3 in an antiparallel arrangement. Other interfaces do not possess such a full set of "native-like" stabilizing interactions as we observe in the PCBP2 KH1 interface. The same PCBP2 KH1 dimer is also present in the crystal structures of PCBP2 KH1 in complex with a 12-nt DNA and RNA with different crystal packings.4 Given the remarkable features and reoccurrence of the PCBP2 KH1 dimer, it is tempting to speculate that some KH domains may have a natural propensity to dimerize (self-association, forming heterodimers with other KH domains, or interacting with other proteins) through the β1/α3 interface; depending on the properties of the residues present at the interface. For such KH domains, they cannot only play a role in interaction with nucleic acids, but also in protein-protein interactions. Moreover, because there is no overlap between the nucleic acids and protein interaction, binding to nucleic acids and protein partners can happen simultaneously, which may be functionally important in certain scenarios.

**DISCUSSION**

The DNA-bound form crystal structure of the PCBP2 KH1 is very similar to the unliganded structure we previously determined by NMR (22). The most noticeable structural change upon DNA binding is seen in the variable loop region. In the free form structure, the variable loop is pointing outward, resulting in a more open binding groove; in the DNA-bound form, the variable loop wraps back toward the groove to achieve close contacts with the DNA (Fig. 3B). In several other KH domain-nucleic acid interactions (26–30), the nucleic acid binding groove is also largely preformed. This appears to be a common property among KH domains.

Prior to this study, there were two KH domain-nucleic acid co-crystal structures available in the literature: the 2.4-Å structure of NOVA2 KH3 in complex with a SELEX RNA stem loop (28), and the 1.8-Å structure of hnRNP K KH1 in complex with a 6-nt DNA (30). There are also several complex structures determined by NMR: splicing factor 1 (SF1) KH with single-stranded RNA (29), FUSE-binding protein (FBP) KH3 and KH4 with ssDNA (27), and hnRNP K KH3 with ssDNA (26). A previous analysis of these structures (30) revealed that the NMR structures of the KH domains of hnRNP K and FBP with ssDNA differed from other KH domain-nucleic acid complexes in that they employed weak methyl-mediated hydrogen bonds to achieve specific recognition of the Watson-Crick positions of the DNA bases, with different relative positioning of the DNA bases as a possible result. A comparison of the other structures led the authors to the following conclusions: each KH domain recognized a core motif of four nucleotides; only pyrimidines

---

4 Z. Du, J. K. Lee, R. Tjhen, L. Shang, R. M. Stroud, and T. L. James, unpublished results.
Figure 6. Comparison of the PCBP2 KH1-htDNA and the NOVA2 KH3-RNA co-crystal structures. The sequences of PCBP2 KH1 and NOVA2 KH3 are ~23% identical and ~47% similar. A, structure of the PCBP2 KH1-htDNA complex. This rendering is virtually identical to that in Fig. 4C except that the viewing window includes the whole structure, and the residues defining the hydrophobic floor of the binding groove are shown as sticks (labeled in green). See Fig. 4C for a detailed annotation. B, crystal structure of the NOVA2 KH3-RNA complex (28). For clarity, only five residues out of the 20-nt crystallization stem-loop/RNA are shown. Coloring and annotation are comparable to the PCBP2 KH1-htDNA complex. Stick representations are shown for residues involved in hydrogen bonds with the bases of the RNA and defining the hydrophobic floor of the binding groove, colored in blue and green, respectively.

The proteins of the PCBP family contain three copies of the KH domain, with the two central copies having similar structures and the third copy differing. A comparison of these domains in the PCBP2 KH1 complex shows that the residues involved in base recognition are highly conserved. For example, the backbone carbonyl oxygen of a conserved isoleucine is involved in hydrogen bonding to the nucleic acid base. The specific hydrogen bonds involved in recognition are not always the same, but they are generally conserved within the family. For instance, recognition of the cytosine at the fourth position in the PCBP2 KH1 complex involves only water-mediated hydrogen bonds, whereas in the NOVA2 KH3 complex, a water molecule is involved in the hydrogen bond to the cytosine. Intriguingly, a guanidino group of a conserved arginine is also involved in hydrogen bonding to the nucleic acid base.

Recognition of the second nucleotide in the core sequence (cytosine) is very similar for both KH domains. The base of the first nucleotide is positioned on top of helix $\alpha_1$; the backbone carbonyl oxygen of a conserved isoleucine is involved in hydrogen bonding to the nucleic acid base. The specific hydrogen bonds involved in recognition are not the same, but they are generally conserved within the family. For instance, recognition of the cytosine at the fourth position in the PCBP2 KH1 complex involves only water-mediated hydrogen bonds, whereas in the NOVA2 KH3 complex, a water molecule is involved in the hydrogen bond to the cytosine. Intriguingly, a guanidino group of a conserved arginine is also involved in hydrogen bonding to the nucleic acid base.

Recognition of the second nucleotide in the core sequence (cytosine) is very similar for both KH domains. The base of the first nucleotide is positioned on top of helix $\alpha_1$; the backbone carbonyl oxygen of a conserved isoleucine is involved in hydrogen bonding to the nucleic acid base. The specific hydrogen bonds involved in recognition are not the same, but they are generally conserved within the family. For instance, recognition of the cytosine at the fourth position in the PCBP2 KH1 complex involves only water-mediated hydrogen bonds, whereas in the NOVA2 KH3 complex, a water molecule is involved in the hydrogen bond to the cytosine. Intriguingly, a guanidino group of a conserved arginine is also involved in hydrogen bonding to the nucleic acid base.

Recognition of the second nucleotide in the core sequence (cytosine) is very similar for both KH domains. The base of the first nucleotide is positioned on top of helix $\alpha_1$; the backbone carbonyl oxygen of a conserved isoleucine is involved in hydrogen bonding to the nucleic acid base. The specific hydrogen bonds involved in recognition are not the same, but they are generally conserved within the family. For instance, recognition of the cytosine at the fourth position in the PCBP2 KH1 complex involves only water-mediated hydrogen bonds, whereas in the NOVA2 KH3 complex, a water molecule is involved in the hydrogen bond to the cytosine. Intriguingly, a guanidino group of a conserved arginine is also involved in hydrogen bonding to the nucleic acid base.

Recognition of the second nucleotide in the core sequence (cytosine) is very similar for both KH domains. The base of the first nucleotide is positioned on top of helix $\alpha_1$; the backbone carbonyl oxygen of a conserved isoleucine is involved in hydrogen bonding to the nucleic acid base. The specific hydrogen bonds involved in recognition are not the same, but they are generally conserved within the family. For instance, recognition of the cytosine at the fourth position in the PCBP2 KH1 complex involves only water-mediated hydrogen bonds, whereas in the NOVA2 KH3 complex, a water molecule is involved in the hydrogen bond to the cytosine. Intriguingly, a guanidino group of a conserved arginine is also involved in hydrogen bonding to the nucleic acid base.
is now clear that PCBP2 KH1 and hnRNP K KH3 can both do so as an isolated domain. Based on knowledge about the critical residues responsible for specific recognition gained by the crystal structures and sequence alignments (Fig. 1B), all KH1 domains (probably except hnRNP K KH1, which has an aspartic acid instead of glutamic acid at position 51) and KH3 domains should be able to bind poly(C) sequences in a way similar (if not identical) to PCBP2 KH1 and hnRNP K KH3, respectively. We have prepared an 15N-labeled sample of the PCBP2 KH3 domain (which has an Asp at the position corresponding to Glu51 in PCBP2 KH1) and observed significant chemical shift changes in the fingerprint amide resonances upon addition of the same 7-nt htDNA (Supplementary Materials, Ref. 22), indicating a specific binding event. Because of conservation of the two arginines at positions 40 and 57 (PCBP2 KH1 numbering), all KH2 domains should also be able to specifically recognize at least two cytosines at the second and third positions of the tetrancleotide core motif. It is likely that all three of the KH domains within each PCBP are capable of poly(C) binding. Interdomain interactions as well as protein-protein interactions could mediate binding of particular nucleic acids to the proteins of this three KH domain family.

In the PCBP1/2-α-globin mRNA complex, it was established that a minimum RNA construct containing three stretches of poly(C) sequences formed a 1:1 complex with PCBP2 or PCBP1 (9), consistent with each KH domain recognizing one stretch of the poly(C) sequence. This kind of nucleic acid–PCBP interaction would increase the sequence specificity and affinity of interaction on the one hand, and might help to constrain one or both of the interacting partners (the nucleic acid and the PCBP) in certain biologically significant conformational changes on the other hand. Of course, other ways of interaction exist. One established example is the interaction of PCBP1 or 2 with two C-rich sequence-harboring RNA structures within the 5′-UTR of poliovirus type-1. The KH1 domain is the major determinant for interaction with both RNAs. Therefore, although PCBP2 KH1 and KH3 (very likely also KH2) domains can all bind single-stranded DNA/RNA, they are clearly not functionally equivalent to one another in these cases. The crystal structures of the PCBP2 KH1 and hnRNP K KH3 complexes reveal some different structural features in binding to poly(C) sequences. These may correlate to distinguishing differences in binding affinity and specificity (regarding the identity of the first nucleotide in the core motif). It is also possible that some KH domains, such as PCBP2 KH1, have evolved some special features to cope with recognition of poly(C) sequences presented in more constrained conformations within highly structured RNAs.

Another important insight into PCBP function gained from our crystal structure is the revelation that PCBP2 KH1 domain has a protein-protein interaction interface (the dimerization interface), suggesting that given an appropriate composition of amino acids residues on the α3 and β1 surface, some KH domains are capable of assuming non-excluding dual functional roles in nucleic acid and protein interactions. Although the nucleic acid binding interface and the protein interaction interface are located on opposite sites of the domain, a study on the Nova2 KH3 domain (33) suggested that the processes of nucleic acid binding and protein interaction might be correlated. Binding to one interface induced stiffening in other regions of the protein and therefore reduced the entropic cost of binding to the other interface.

Most of the published studies on PCBP2 have been directed to functions dependent on RNA binding events, with only a few reports (21, 22) suggesting its possible involvement in mechanisms associated with DNA recognition. Our co-crystal structure of the PCBP2 KH1-htDNA complex, in conjunction with our previous NMR study of the complex in solution (21, 22), proves that PCBP2 has the ability to bind to poly(C) DNA sequences specifically; poly(G), poly(A), poly(T), and poly(U) sequences did not yield chemical shift changes. This feature should enable PCBP2 to assume functional roles in mechanisms dependent on DNA binding, such as transcriptional regulation and telomere maintenance.

The tandem arrangement of poly(C) stretches on the human C-rich strand telomeric DNA is very similar to some of the known RNA targets for PCBP1 and PCBP2 (such as the 3′-UTR of α-globin mRNA and some other ultrastable mRNAs). It is fully possible that PCBP1 or -2 would bind to the C-rich strand of htDNA in a manner somewhat similar to the α-globin complex. Recent progress on telomere/telomerase biology has shown that the telomere/telomerase complex can exist in different stages (34). Whereas the C-rich strand may be present in a double-stranded form with the complementary G-rich strand, at certain stages the DNA telomere or telomerase RNA may have the C-rich htDNA or the RNA template in a single-stranded conformation. Such a scenario would permit the involvement of proteins of the PCBP family in regulation of telomere and telomerase activities through specific binding to the exposed C-rich strands. To this end, we have confirmed, through antibody pull-down experiments and mass spectroscopy, that PCBP1 is one of the nucleic acid-binding proteins present in the human telomere/telomerase complex.2 Further structural and biological studies are being carried out to increase our knowledge of the exact involvement of PCBPs in telomere/telomerase regulation.

Acknowledgment—We thank Chris Wadding for managing the UCSF X-ray Crystalization Laboratory.

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. Chikheidez, 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., Bellerbrand, C., Holcz, M., Beindl, 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) Nucleic 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. Waggoner, S. A., and Liebhaber, S. A. (2000) Exp. Biol. Med. 225, 387–395
10. Ostareck-Lederer, A., Ostareck, D. H., Standart, N., and Thiele, B. J. (1994) EMBO J. 13, 1476–1481
11. Ostareck, D. H., Ostareck-Lederer, A., Wilm, M., Thiele, B. J., Mann, H., and Hentze, M. W. (1997) Cell 89, 597–606
12. Ostareck, D. H., Ostareck-Lederer, A., Shatsky, I. N., and Hentze, M. W. (2001) Cell 104, 281–290
13. Waggoner, S. A., and Liebhaber, S. A. (2000) Mol. Cell. Biol. 23, 7055–7067
14. Gamarnik, A. V., and Andino, R. (1997) RNA 3, 882–892
15. Gamarnik, A. V., and Andino, R. (1998) Genes 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. Blyn, L. B., Swiderek, K. M., Richards, O., Stahl, D. C., Semler, B. L., and Ehrenfeld, E. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 11115–11120
18. Blyn, L. B., Towner, J. S., Semler, B. L., and Ehrenfeld, E. (1997) J. Virol. 71, 6243–6246
19. Tomonaga, T., and Levens, D. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 5830–5835
20. Lacroix, L., Cesaratto, L., and Manzini, G. (2003) Mol. Cell. Biol. 23, 5201–5209
21. Bandiera, A., Tell, G., Marsich, E., Scaloni, A., Poesfeldt, G., Akindahunsi, A. A., Cesartano, L., and Manzini, G. (2003) Arch. Biochem. Biophys. 409, 305–314
22. Du, Z., Yu, J., Chen, Y., Andino, R., and James, T. L. (2004) J. Biol. Chem. 279, 48126–48134
23. Ostwinoewski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326
24. Brünger, A. T. (1996) X-PLOR version 3.843, Yale University, New Haven, CT
25. Gerber, P. R. A. M., K. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 5830–5835
26. Bradlick, D. T., Baber, J. L., Levens, D., and Clare, G. M. (2002) EMBO J. 21, 48126–48134
Structure of KH Domain in Complex with Human Telomeric DNA

3476–3485
27. Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D., and Clore, G. M. (2002) Nature 415, 1051–1056
28. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. (2000) Cell 100, 323–332
29. Liu, A. Z., Riek, R., Wider, G., Von Schroetter, C., Zahn, R., and Wuthrich, K. (2000) J. Biomol. NMR 16, 127–138
30. Backe, P. H., Messias, A. C., Ravelli, R. B., Sattler, M., and Cusack, S. (2005) Structure (Camb.) 13, 1055–1067
31. De Boulle, K., Verkerk, A. J., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., Van Den Bos, F., De Graaff, E., Oostra, B. A., and Willems, P. J. (1993) Nat. Genet. 3, 31–35
32. 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
33. Ramos, A., Hollingsworth, D., Major, S. A., Adinolfi, S., Kelly, G., Muskett, F. W., and Pastore, A. (2002) Biochemistry 41, 4193–4201
34. Blackburn, E. (2005) FEBS Letters 579, 859–862