Structure and RNA binding of the third KH domain of poly(C)-binding protein 1

M. Sidiqi¹, J. A. Wilce¹, J. P. Vivian¹,², C. J. Porter¹, A. Barker¹,², P. J. Leedman³,⁴ and M. C. J. Wilce¹,²,⁴,*

¹School of Biomedical and Chemical Sciences, ²School of Pharmacology and Medicine, ³Laboratory for Cancer Medicine, the UWA Centre for Medical Research and ⁴Western Australian Institute for Medical Research, The University of Western Australia, WA Australia 6009

Received November 18, 2004; Revised January 12, 2005; Accepted February 3, 2005

ABSTRACT

Poly(C)-binding proteins (CPs) are important regulators of mRNA stability and translational regulation. They recognize C-rich RNA through their triple KH (hnRNP K homology) domain structures and are thought to carry out their function though direct protection of mRNA sites as well as through interactions with other RNA-binding proteins. We report the crystallographically derived structure of the third domain of αCP1 to 2.1 Å resolution. αCP1-KH3 assumes a classical type I KH domain fold with a triple-stranded β-sheet held against a three-helix cluster in a βαββα configuration. Its binding affinity to an RNA sequence from the 3'-untranslated region (3'-UTR) of androgen receptor mRNA was determined using surface plasmon resonance, giving a Kd of 4.37 μM, which is indicative of intermediate binding. A model of αCP1-KH3 with poly(C)-RNA was generated by homology to a recently reported RNA-bound KH domain structure and suggests the molecular basis for oligonucleotide binding and poly(C)-RNA specificity.

INTRODUCTION

αCP1 is a member of the poly(C)-binding protein family of proteins that include αCP1, αCP2, αCP3, αCP4 (also known as PCBP and hnRNP E proteins) and the earliest member to be characterized, the heterogeneous ribonucleoprotein K, hnRNP K. These proteins contain a triplet K homology (KH) RNA-binding motif, as first identified in hnRNP K, which confer specificity for single-stranded poly(C) tracts of both RNA and DNA. The two N-terminal domains are closely spaced, whereas the C-terminal KH domain is separated by a linking segment of variable length. The αCP family members are well conserved, with the highest conservation observed between corresponding KH domains. The main differences between αCP family members and their isoforms occur in the regions between KH domains, which vary in both sequence and length [for excellent reviews see (1,2)]. The proteins exist in both the nucleus and cytoplasm of the cell and are involved in a diverse range of functions affecting the post-transcriptional regulation of specific genes. These include the shuttling of mRNA between the nucleus and the cytoplasm, the stabilization of specific mRNAs, translational silencing and translational enhancement. This range of functions, which confer seemingly opposed effects on gene expression, are likely to be modulated through variation in cellular conditions, specific RNA secondary structures and through interactions with other mRNA-binding proteins.

αCP1 has been studied in particular detail. It has been implicated in the stabilization of specific mRNAs, leading to the upregulation of their gene products. It has been shown to be sufficient for the formation of the ‘α-complex’ at a specific C-rich region of the 3'-untranslated region (3'-UTR) of α-globin mRNA, causing its accumulation during terminal erythroid differentiation (3,4). The mechanism of the mRNA stabilization is thought to be through the mutually cooperative binding of αCP1 (or other members of the αCP family) and poly(A)-binding protein at the poly(A) tail, resulting in both inhibition of deadenylation and protection of a specific endonuclelease site, adjacent to the αCP-binding site, from nucleases (5,6). Interactions between the αCP protein of the α-complex and the AU-rich element binding degradation factor AUF1 (hnRNP D) have also been detected (7).
Binding of αCP proteins to the 3′-UTR pyrimidine-rich motifs have also been implicated in the stabilization of tyrosine hydroxylase (8), erythropoietin (9) and β-globin (10) mRNAs. αCP2 binds with high affinity to a C-rich region within the 3′-UTR of collagen α1(I) mRNA resulting in its increased stability (11). αCP1 and αCP2 have also been shown to target a specific UC-rich region of the 3′-UTR of androgen receptor (AR) mRNA. Through cooperative binding with HuR, these proteins are thought to be a part of the post-transcriptional control mechanism for the AR expression (12).

αCP proteins have also been shown to affect translational control. Their binding to a CU-rich region of the 3′-UTR differentiatation control element of 15-lipoxgenase mRNA along with hnRNP K suppresses the translation in erythroid cells until the terminal stages of erythroid differentiation (13). This occurs through interference with the joining of the ribosomal 60S and 40S subunits at the initiation AUG codon (14). Similarly, human papillomavirus type 16 L2 mRNA appears to be silenced via binding to αCP proteins (15), though the L2 sequence is not so C-rich. In contrast, translational enhancement has been reported due to αCP binding to two sites of the S′-UTR of picornavirus RNA. αCP binds to the 5′-terminal cloverleaf structure of stem–loop I, and to a C-rich region of the stem–loop IV of the internal ribosome entry site resulting in the cap-independent translation of the gene (16,17). Further αCP-mediated effects on translation have also been reported for other viral systems (18,19). Thus, αCP binding to RNA can result in both silencing and enhancement of translation through a diverse array of pathways.

The specificity of αCP proteins for poly(C)-oligonucleotides is conferred through their KH domains. These domains were originally defined by the repeated 45 amino acid motif identified in hnRNP K (20). A more extensive 68–72 amino acid motif has since been defined (21). The 3D structures of several KH domains have been determined, both in the absence (21–24) and in the presence of RNA or ssDNA (25–27). Two structural subtypes have been identified, type I possesses the 45 residue core βαββ motif plus a C-terminal βα extension, whereas type II possesses the core and an N-terminal αββ extension (28). The αCP KH domains, like the hnRNP K, are type I and thus comprise a three-stranded anti-parallel β-sheet packed against three α-helices (βααβα). Oligonucleotide binding by KH domains occurs primarily via hydrophobic interactions through a groove bounded by two unstructured surface loops. The loop between α-helices 1 and 2 contains an invariant GXXG, crucial to oligonucleotide binding; that between β-strands 2 and 3 is of variable length and sequence and flanks the RNA-binding groove (25).

The relative contributions by the three KH domains in the αCP proteins to RNA-binding affinity and specificity are not yet clearly understood. While all three of the individual KH domains of hnRNP K have been shown to contribute significantly to poly(C) binding (29), only the first and third KH domains of αCP1 and αCP2 have been shown to independently bind poly(C)-RNA with high affinity and specificity (30). It may be, however, that the second KH domain also binds poly(C)-RNA when tethered by its neighbouring domains. The αCP-binding motif within the α-globin S′-UTR contains three C-rich patches. The disruption of any of these interferes with the α-complex formation and decreases α-globin stability in vivo (3,31). Likewise, the optimal target sequence of the closely related αCP-2KL isoform generated by in vitro SELEX contained three short C-patches within an exposed single-stranded conformation (32). This suggests that optimum binding by αCP proteins may be achieved via the interaction of all three KH domains with poly(C)-regions. It is not yet known, however, how these individual KH domains could be juxtaposed to facilitate such binding.

The current study describes the overexpression, crystallization and structure determination of the third KH domain of αCP1 solved to 2.1 Å resolution using X-ray crystallography. This represents the first structure of an αCP1 domain to be structurally analysed. We also verify its RNA-binding capacity to a target RNA sequence using surface plasmon resonance and present a model of its interaction with poly(C)-oligonucleotides, based on a homologous KH domain bound to RNA. This has provided insight into the basis for the poly(C)-binding specificity of αCP1-KH3 and is the first step towards the structural definition of the full-length protein.

MATERIALS AND METHODS

Protein expression and purification

The third KH domain of αCP1 (KH3) was expressed as a fusion protein with glutathione-S-transferase (GST). The DNA coding sequence comprising amino acids 279–356 of αCP1 were cloned into pGEX-6P2 plasmids and expressed by the Escherichia coli BL21 (Codon+) cell line in Luria broth at 37°C. Protein expression was induced with 0.02 mM isopropyl-β-d-thiogalactopyranoside (IPTG) (140 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4) containing 0.5% Triton X-100. They were then lysed by French Pressing (SLM Instruments, Inc.), supplemented with 0.5 mM phenylmethylsulfonyl fluoride. αCP1-KH3 was purified by affinity chromatography using glutathione agarose beads equilibrated with PBS buffer and the GST removed using 2 U Prescission protease (Amerham) in 50 mM Tris–HCl, pH 7.0, 150 mM NaCl, 1 mM EDTA and 1 mM DTT. Size exclusion chromatography using a Sephadex 75 column (Pharmacia) was used as the final purification step of αCP1-KH3 after dialysis into phosphate buffer, pH 6.0 (1 mM DTT, 25 mM phosphate, 150 mM NaCl and 1 mM EDTA). The purified protein was concentrated to 5 mg/ml with centrifugal concentrators of 3 K cutoff (Millipore) and quantified using a detergent compatible (Biorad) protein assay.

Crystallization of αCP1-KH3

Crystals of αCP1-KH3 were grown using vapour diffusion in 2 μl hanging drops containing 1:1 mixtures of protein and reservoir solutions. The protein solution contained 5.0 mg/ml of protein in 25 mM potassium phosphate, pH 6.0, 1 mM DTT, 1 mM EDTA and 150 mM NaCl, and the reservoir solution was composed of 0.1 M Na HEPES, pH 7.5 in 1.5 M lithium sulphate–Hampton Crystal Screen reagent formulation number 16 (Hampton Research, CA). Crystals typically grew in 2 days to dimensions of ~0.3 × 0.2 × 0.02 mm with the outline...
of a rugby football, and diffraction data were collected to 2.1 Å resolution.

X-ray data collection

Data were recorded with a Rigaku R-Axis V imaging plate detector mounted on a Rigaku RU-200 rotating anode generator with a Cu target and focusing mirror optics. Flash freezing was carried out in a stream of cold nitrogen gas. Prior to flash-freezing at 100 K, the crystals were passed through a solution of the reservoir solution modified to include 15% glycerol as a cryoprotectant. Data were integrated and scaled with DENZO and SCALEPACK (33). Structure factor amplitudes were calculated using TRUNCATE (34). The data collection statistics are given in Table 1. The data collection and refinement statistics

<table>
<thead>
<tr>
<th>Table 1. Data collection and refinement statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
</tr>
<tr>
<td>Symmetry</td>
</tr>
<tr>
<td>Unit cell (Å)</td>
</tr>
<tr>
<td>Measured reflections</td>
</tr>
<tr>
<td>Unique reflections</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;%</td>
</tr>
<tr>
<td>Wilson B (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Refinement</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
</tr>
<tr>
<td>R&lt;sub&gt;cryst&lt;/sub&gt; (%)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
</tr>
<tr>
<td>r.m.s. deviation from ideal values</td>
</tr>
<tr>
<td>Bond length (Å)</td>
</tr>
<tr>
<td>Bond angle (°)</td>
</tr>
<tr>
<td>Average temperature factor (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Number of water molecules</td>
</tr>
</tbody>
</table>

Values in parentheses are for the last resolution shell (2.16–2.1 Å). R<sub>merge</sub> = Σ |F<sub>j</sub> − c<sub>j</sub>|/Σ|c<sub>j</sub> where F<sub>j</sub> is the observed diffraction intensity and c<sub>j</sub> is the average diffraction intensity from several measurements of one reflection. R<sub>cryst</sub> = Σ |F<sub>j</sub>| − |F<sub>c</sub>|/Σ|F<sub>c</sub>| where |F<sub>j</sub>| and |F<sub>c</sub>| are the observed and calculated structure factors, respectively.

Modelling of αCP1-KH3 bound to poly(C)-oligonucleotide

The αCP1-KH3 structure was superposed with the structure of Nova2-KH3 bound to RNA (accession no. 1EC6) (25) using LSQMAN (36). In this way, the coordinates of oligonucleotides (9–16) could be extracted and used to generate an 8 nt poly(C)-RNA docked to the αCP1-KH3 structure (using the Insight II software package to change the bases to cytosine). The structure was subjected to molecular dynamics simulations using NAMD (37) in a fully solvated box, with overall neutral charge (through the addition of randomly placed sodium ions). The complex structure was allowed to equilibrate in 10<sup>8</sup> fs time steps using the CHARMM27 energy forcefield (38) at 310K and 1 atm using periodic boundary conditions. This ensured that there were no steric clashes in the final model and allowed a full set of possible intermolecular interactions to be viewed. The stereochemistry of the oligonucleotide and the intermolecular hydrogen bond formation during the simulation were recorded at picosecond intervals for analysis (Supplementary Material).

RESULTS

Structural overview

The αCP1-KH3 adopts a classic KH type I domain fold (28), with a triple-stranded β-sheet held against a three-helix cluster in a [βαβ]αβ configuration (Figure 1A). The β-sheet is antiparallel and displays the usual left-handed twist. From its inner surface emanate numerous hydrophobic residues which contribute both to the hydrophobic core and the oligonucleotide

RNA-binding studies

Surface plasmon resonance (using a BIAcore 2000 instrument) was employed to characterize the αCP1-KH3 interaction with RNA. A research grade chip coated with streptavidin was purchased from BIAcore. The target 5′-biotinylated mRNA (mRNA: 5′-CUCUCUUUUUUUCUUCUUCCCUC- CUA-3′) representing nt 3296–3325 of AR mRNA was obtained from Dharmacon and immobilized on the second flow cell as the captured molecule. The first flow cell coated with only streptavidin was used as the reference surface. The immobilization steps were carried out at a flow rate of 10 µl/min in running buffer, 10 mM Tris–HCl, pH 7.4, 150 mM NaCl, 0.5% Triton X and 2 mM DTT, 2 mM EDTA, 125 µg/ml tRNA and 62.5 µg/ml BSA. An average of 30 RU of RNA was immobilized on flow cell 2. αCP1-KH3 was injected over flow cells 1 and 2 with concentrations of 10, 5, 2.5, 1.25 and 0.625 µM for 2 min using a flow rate of 50 µl/min. All experiments were duplicated (or performed multiple times) to determine the reproducibility of the signal. Regeneration involved removal of the bound protein from the streptavidin chip with a 2 min wash at 20 µl/min with 2 M NaCl. Data were analysed with the BIAevaluation software to obtain a binding constant using a steady-state model.

MW 8525; expected MW 8552.72). The final model has been deposited with the Worldwide Protein Data Bank (accession no. 1WVN).
The crystal structure of αCP1-KH3 (residues 279–356) solved to 2.1 Å resolution depicted in (A) cartoon form and (B) as a molecular surface in the same orientation. The structure is shown from the beginning of β-strand 1 to the end of α-helix 3, since the regions outside these bounds were random coil or not visible in the density. The GXXG motif, common to this oligonucleotide-binding motif, is coloured blue. The ‘variable loop’ region between β-sheets 2 and 3 is coloured pink. These regions bound the hydrophobic oligonucleotide-binding cleft that accommodates C-rich RNA or ssDNA. (C) The electrostatic potential emanating from the αCP1-KH3 structure calculated using the APBS software package (http://agave.wustl.edu/apbs/) (39–43). Potential contours are shown at +1 kT/e (blue) and −1 kT/e (red) and obtained by solution of the linearized Poisson–Boltzmann equation at 150 mM ionic strength with a solute dielectric of 2 and a solvent dielectric of 78.5. The blue contour represents a striking positive potential directing oligonucleotides to the binding cleft.

**Comparison with other KH domain structures**

αCP1-KH3 shows high structural similarity to other type I KH domains. The seven most similar KH structures, including hnRNPK (24,26), Nova2-KH3 and Nova1-KH3 (23,25), FBP-KH3 and FBP-KH4 (27), vigilin-KH6 (21) and FMR-KH1 (22), are shown superimposed in Figure 2A (in the case of NMR-derived structures, the first chain in the PDB
file is depicted). Their backbone traces are highly convergent with pairwise root-mean-square deviation (RMSD) scores compared with αCP1-KH3 over the matched regions (according to LSQMAN) <1.8 Å. Vigilin-KH6 and FMR-KH1 show the greatest deviations, with several stretches of backbone fold unmatched to regions within αCP1-KH3 (>3.5 Å away). These include the variable loop and the region about the GXXG motif, which are also the regions that show the least definition in the NMR-derived structures.

Figure 2B shows the deviations numerically, with α-carbon distances from matched αCP1-KH3 residues plotted against the αCP1-KH3 residue number. The divergent regions are shown as off-scale in this plot. The KH structures are superimposed with most α-carbon atoms within 2 Å of the corresponding αCP1-KH3 atom. Apart from Vigilin-KH6 and FMR-KH1, greater deviations only occur at the termini and variable loop region between β-sheets 2 and 3. A subtle variation also occurs at the GXXG motif possibly reflecting the inherent flexibility of the glycines. It is remarkable that these KH domains retain such high structural similarity and yet possess distinct oligonucleotide-binding preferences.

αCP1-KH3 shows the greatest structural similarity to its fellow poly(C)-binding family member, hnRNP K, with an RMSD of 0.63 Å. A structure-based sequence alignment of these KH domains with the others serves to highlight the conservation of residues reportedly underlying oligonucleotide binding (Figure 3). In particular, residues about the GXXG motif as well as those in the β-strand 2 provide the main contact surface. Of these, Ile 20, Ile 21, Ile28 and Ile 41 are highly conserved as bulky hydrophobic residues, and Gly18, Gly22 and Gly25 are integral to the oligonucleotide-binding motif. Basic residues Arg23 and Arg51 have also been shown to be involved in the oligonucleotide-binding interaction and basic residues are retained at these positions except in Vigilin-KH6 and FMR-KH1.

RNA-binding studies

The binding of αCP1-KH3 to a 30 nt UC-rich sequence from the 3′-UTR of AR mRNA (nt 3296–3325) was examined using surface plasmon resonance. This sequence was derived from a 51 nt sequence (nt 3275–3325) shown to be an αCP1 target using αCP1 antibody supershift identification of the components of LNCaP cellular extract binding to the AR mRNA construct (12). Since the αCP1 binding was localized to the 3′-CCCUCCC sequence in this UC-rich stretch, we utilized the 3′ 30 nt sequence as our target RNA. The protein, as prepared using our reported method, was determined to be fully folded and capable of binding RNA completely using NMR titration methods (46). The protein was injected over a biosensor to which purified biotinylated RNA had been immobilized. The binding curves are illustrated in Figure 4, and are characteristic of a fast kinetics macromolecular interaction. A simple 1:1 Langmuir model could not be fitted to the curves due to the fast binding kinetics at the start and finish of the protein injection period, so reliable association and dissociation constants could not be determined. An equilibrium analysis of the data, however, yielded a $K_d$ value in the µM range (4.37 µM), which is indicative of intermediate binding. This is in contrast to the tight binding ($K_d \sim 28$ nM) determined for the full-length protein binding to the 51 nt AR mRNA construct using REMSA (12). While this may be an overestimate due to
the absence of non-specific binding competitors used in this study, the tighter binding of full-length protein suggests that multiple KH domains of αCP1 are likely to be participating synergistically in binding the target RNA (12).

Figure 3. Structure-based sequence alignment of seven KH domains of high structural similarity to αCP1-KH3. Each KH domain was structurally aligned using LSQMAN (36) against αCP1-KH3. Amino acid residues with α-carbon positions within 3.5 Å of a corresponding αCP1-KH3 residue are shown in black. Highlighted in purple are the amino acid residues that do not align well with residues of αCP1-KH3. Secondary structural elements, as defined in Lewis et al., (23) are shown above the corresponding sequence in cartoon form. Parenthesized numbers represent the amino acid numbers at the start and finish of the superimposed core region for each structure, and indicate the extent of the structure used to calculate sequence identity with αCP1-KH3 (final column). The GXXG motif and the variable loop regions are blocked with grey. Amino acid residues reported to make contact with the oligonucleotide [in the cases of structures determined in complex with either RNA or ssDNA (25–27)] are highlighted in red, and the αCP1-KH3 predicted to make contact with oligonucleotide in the current study are highlighted in tan. NMR structures were structurally aligned on the basis of the first chain in the deposited PDB coordinate file and all were deemed to be representative of the set of structures.

Figure 4. Interaction of the αCP1-KH3 with the 30 nt of the 3'-UTR of AR mRNA measured by surface plasmon resonance. (A) 30 RU RNA was immobilized on a streptavidin-coated chip. Binding interactions were measured for a series of dilutions of the αCP1-KH3 domain from 10 to 0.625 μM for 2 min using flow rate of 50 μl/min. (B) Steady-state analysis of the interaction yielded a K_d value of 4.37 μM.

Model of αCP1-KH3 bound to poly(C)-oligonucleotide
The high degree of similarity of αCP1-KH3 to Nova2-KH3 has permitted its interaction with poly(C)-RNA to be
modelled. Nova2-KH3 has been structurally characterized, complexed with a 20 base stem–loop RNA (25) as well as in its uncomplexed forms (23). Oligonucleotide binding incurred no significant structural differences in the backbone conformation, suggesting that the αCP1-KH3 structure may also represent a close approximation of its oligonucleotide bound form. Poly(C)-RNA was therefore positioned in the binding cleft of αCP1-KH3 by analogy to this structure to help predict interactions that may underlie its poly(C)-binding specificity. The poly(C)-RNA is positioned along the hydrophobic cleft and across the GXXG motif with four bases making most of the contacts with the binding site. The orientation of the oligonucleotide is with the sugar–phosphate backbone directed towards the helix edge of the cleft and the bases, planar to the protein surface and pointing towards the centre and β-sheet 2 (Figure 5A).

Figure 5B summarizes the possible electrostatic and hydrophobic contacts between αCP1-KH3 and RNA. These were determined with allowance for some molecular flexibility (as assessed using molecular dynamics simulations using the CHARMM27 energy forcefield). They include non-specific hydrophobic interactions with Ile17, Gly18, Cys19, Ile21, Ile 28 and Ile41, which form the surface of the binding cleft, as well as numerous electrostatic contacts to the sugar–phosphate backbone involving Gly22, Arg23, Gln24, Gly25 backbone atoms (the GXXG tetrad) and contact with the Cyt4 sugar hydroxyl by the Lys40 side chain amino group. Interactions that may help to favour pyrimidine binding include Arg 32 and Arg51 guanidino groups positioned in close proximiy to pyrimidine carbonyls (C2 carbonyls in Cyt3 and Cyt2, respectively; Figure 5C). Interactions that could underlie cytosine specificity include potential hydrogen bonds between Ile28 and Ile41 side chains and the central two cytosine bases (via their O2, N3 and N4 atoms). These iso-leucines are conserved in hnRNP K and form an extensive methyl–oxygen and methyl–nitrogen hydrogen bond network with the equivalent bases in ssDNA (26). In addition, several water-mediated hydrogen bonds between the protein and RNA occur fleetingly during the simulation. In particular, Ile41 carbonyl oxygen alternates between being hydrogen bonded to Cyt4 carbonyl and sugar hydroxyl groups, and thus contributes to the preference for ribopyrimidyl oligonucleotide. A summary of the occurrences of each hydrogen bond and water-mediated hydrogen bond during the 1 ns molecular dynamics simulation is provided as Supplementary Material.

Poly(C)-RNA structure may favour binding

Many of the αCP1-KH3-oligonucleotide contacts would be predicted to occur upon either RNA or ssDNA binding, such as the hydrophobic contacts listed above and electrostatic interactions with Gly25, Arg51 and Lys40. Other contacts are precluded from occurring in the case of ssDNA, due to the absence of sugar hydroxyl groups. These include potential hydrogen bonds between sugar hydroxyls and Gly25, Arg32, Arg51 and Lys40 as well as water-mediated hydrogen bonds as mentioned above.

Inter-nucleotide phosphate hydrogen bonds may also impact on the RNA structure and potential interactions with αCP1-KH3. Phosphates of nt 2 and 4 can hydrogen bond to sugar hydroxyls of nt 2 and 3, respectively. Phosphates of nt 1 and 3, on the other hand, may hydrogen bond to Cyt1 and Cyt4 amino groups. The former of these interactions are unique to RNA and the latter are also cytosine specific. Thus, it may be that the uniquely stable conformation of RNA in this binding.
cleft, and in particular that of poly(C)-RNA, favours binding to the KH domain.

**DISCUSSION**

αCP1-KH3 is reported to preferentially bind poly(C)-RNA over other bases and over ssDNA (30), though the ssDNA sequence is not clearly specified in this study. The crystal structure of this domain confirms its adoption of the classical type I KH fold and has allowed a precise model of its interactions with poly(C)-RNA to be examined. Specificity for pyrimidines can be understood in terms of its narrow binding cleft that would only readily accommodate the smaller bases. Specificity for cytosines over uracil or thymine can also be rationalized on the basis of specific hydrogen bond interactions to cytosine C2 carbonyl, N3 and C4 functionalities. Preferential binding to RNA over ssDNA would be explained in part by sugar hydroxyl intermolecular hydrogen bonding. It may also be that a poly(C)-RNA oligonucleotide is able to contour perfectly in the binding cleft, with inter-nucleotide hydrogen bonds from sugar hydroxyls stabilizing this conformation. On the other hand, C-rich ssDNA has been shown to adopt very similar interactions with hnRNPs K, and is reported to bind just as well, if not better, than RNA to this closely related KH domain (26).

This study has shown that oligonucleotide binding by αCP1-KH3 is likely to involve extensive interactions with only four bases. The question remains as to how adjacent KH domains are arranged when full-length αCP1-KH3 binds to RNA. It may be that the KH domains are able to bind in relatively close proximity. Indeed, the two adjacent KH domains (KH3 and KH4 of FBP) were shown to contact stretches of 6–7 bases, respectively, with only 5 bases in domains (KH3 and KH4 of FBP) were shown to contact stretches of 6–7 bases, respectively, with only 5 bases in between (27). In addition, the consensus binding sequence for the αCP-2KL isoform involves three C-rich stretches (of 3–5 bases) separated by 2–6 A/U stretches (32). Thus, αCP binding may well involve participation by all three KH domains.

This study represents the beginning of a structural and biochemical examination of all three KH domains of αCP1. The basis for RNA-binding affinity and specificity of the three KH domains will allow us to predict the occurrence of αCP1 interactions with mRNA and better understand the multi-KH domain binding complex. A comparison with the other αCP isoforms will also be of interest. To date, little is known about their oligonucleotide-binding preferences and differences in their binding affinities to RNA. Further analyses will help us to rationalize the role of the whole αCP family in mRNA stability and translational efficiency.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at NAR Online.

**ACKNOWLEDGEMENTS**

We would like to acknowledge the contribution of Aaron Oakley for helpful instruction in setting up the NAMD calculation. This work has been funded by an Australian Research Council Grant (M.C.J.W., J.A.W. and P.I.L.), an Australian Research Council Fellowship (J.A.W.), an Australian Postgraduate Award (M.S.) and a Small Grant awarded by the University of Western Australia (J.A.W.). Funding to pay the Open Access publication charges for this article was provided by Australian Research Council Grant.

**REFERENCES**


