

Contribution of the first K-homology domain of poly(C)-binding protein 1 to its affinity and specificity for C-rich oligonucleotides

Yano M. K. Yoga¹, Daouda A. K. Traore¹, Mahjooba Sidiqi^{2,3,4}, Chris Szeto¹, Nicole R. Pardini¹, Andrew Barker⁴, Peter J. Leedman⁴, Jacqueline A. Wilce^{1,*} and Matthew C. J. Wilce^{1,*}

¹Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC Australia 3800,

²School of Biomedical, Biomolecular and Chemical Sciences, The University of Western Australia, WA Australia

6009, ³School of Medicine and Pharmacology, The University of Western Australia, WA Australia 6009, and

⁴Laboratory for Cancer Medicine, Centre for Medical Research and Western Australian Institute for Medical Research, The University of Western Australia, WA Australia 6009

Received November 24, 2011; Revised and Accepted January 17, 2012

ABSTRACT

Poly-C-binding proteins are triple KH (hnRNP K homology) domain proteins with specificity for single stranded C-rich RNA and DNA. They play diverse roles in the regulation of protein expression at both transcriptional and translational levels. Here, we analyse the contributions of individual α CP1 KH domains to binding C-rich oligonucleotides using biophysical and structural methods. Using surface plasmon resonance (SPR), we demonstrate that KH1 makes the most stable interactions with both RNA and DNA, KH3 binds with intermediate affinity and KH2 only interacts detectibly with DNA. The crystal structure of KH1 bound to a 5'-CCCTCCCT-3' DNA sequence shows a 2:1 protein:DNA stoichiometry and demonstrates a molecular arrangement of KH domains bound to immediately adjacent oligonucleotide target sites. SPR experiments, with a series of poly-C-sequences reveals that cytosine is preferred at all four positions in the oligonucleotide binding cleft and that a C-tetrad binds KH1 with 10 times higher affinity than a C-triplet. The basis for this high affinity interaction is finally detailed with the structure determination of a KH1.W.C54S mutant bound to 5'-ACCCCA-3' DNA sequence. Together, these data establish the lead role of KH1 in oligonucleotide binding by α CP1 and reveal the molecular basis of its specificity for a C-rich tetrad.

INTRODUCTION

Poly-C-binding proteins (PCBP) are ubiquitous oligonucleotide-binding proteins in eukaryotic cells that play a fundamental role in the regulation of gene expression via interaction with C-rich oligonucleotides. The family consists of the archetypal hnRNP K (heterogeneous nuclear ribonucleoprotein K) and isoforms of α CP1 (also known as PCBP and hnRNP E) including α CP1-4 and α CP-KL (1,2). The effects of binding by α CPs vary—and are thought to be dependent upon the ternary complex within which the α CPs are involved. α CPs are involved at several levels of post-transcriptional gene regulation. Within the nucleus, α CP binding at the 3'-UTR or intron 1 of α -globin mRNA impacts upon its splicing and enhances its cleavage and polyadenylation (3,4). Outside the nucleus, α CPs are implicated in the stabilization of specific mRNAs, leading to the up-regulation of their gene products. They have been shown to be sufficient for formation of the ' α -complex' at a specific C-rich region of the 3'-UTR of α -globin mRNA, causing its accumulation during terminal erythroid differentiation (5,6). Binding of α CPs to 3'-UTR mRNA have been implicated in the stabilization of tyrosine hydroxylase (7), erythropoietin (8), β -globin (9) and collagen α 1(I) (10) mRNAs. α CP proteins have also been shown to effect translational control. Their binding to a CU-rich region of the 3'-UTR differentiation control element (DICE) of 15-lipoxygenase mRNA along with hnRNP K suppresses translation through interference with the joining of the ribosomal 60S and 40S subunits at the initiation AUG codon (11,12). Similarly, human papillomavirus type 16 L2 mRNA appears to be silenced

*To whom correspondence should be addressed. Tel: +1 613 99 02 92 44; Fax: +1 613 99 02 95 00; Email: matthew.wilce@monash.edu
Correspondence may also be addressed to Jacqueline A. Wilce. Tel: +1 613 99 02 92 26; Fax: +1 613 99 02 95 00; Email: Jackie.Wilce@monash.edu

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

© The Author(s) 2012. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

via binding to α CPs (13). In contrast, translational enhancement has been reported due to α CP binding to the 5'-UTR of the folate receptor mRNA (14), the 3'-UTR of phosphatase 2A mRNA (15) and the 5'-UTR of picornavirus mRNA (16,17). Thus, α CP binding to RNA can result in both silencing and enhancement of translation through a diverse set of mechanisms.

In addition to their more recognized ability to bind RNA, α CPs have also been shown to bind single stranded DNA (ssDNA). Such interactions play a role in transcriptional regulation with α CP identified as the ssDNA binding protein underlying proximal promoter activity of mouse μ -opioid receptor (18). The closely related hnRNP K is established as a transcription factor, binding to the CT element in the promoter region of c-myc (19), and also to specific ssDNA elements within the promoter region of a neuronal nicotinic acetylcholine receptor gene (20). The α CPs have also been found to recognize the C-rich strand of human telomeric DNA with high affinity (21). Of all the α CPs, α CP1 in particular, showed remarkable specificity for the telomeric (CCCTAA)_n repeat motif (22).

The structural basis for α CP interactions with oligonucleotide is thus of interest, including the basis for their affinity and specificity for RNA and ssDNA. Oligonucleotide binding by α CPs is via their triple K homology (KH) domain structure, as first identified in hnRNP K (23). These type I KH domains are 68–72 amino acid structures involving a three-stranded anti-parallel β -sheet packed against three α -helices ($\beta\alpha\alpha\beta\alpha$) (24). The two N-terminal KH domains of α CPs are closely spaced, whereas the C-terminal KH domain is separated by a linker of variable length. Nuclear localization sequences in the linker regions between KH2 and KH3 and within KH3 of the α CPs have been shown to dictate their differential subcellular localization (25). α CP1 and α CP2 are predominantly nuclear, whereas α CP3 and α CP4 are restricted to the cytoplasm and the splice variant, α CP2-KL, is present in significant amount in both the nucleus and cytoplasm. It may be that this differential localization dictates the involvement of the α CP family members in their various functions.

The three-dimensional structures of several α CP-KH domains in apo- and oligonucleotide-bound form have been reported including: α CP1-KH3, KH domains 1, 2 and 3 of α CP2 and hnRNP-KH3 (26–31), providing considerable insight into their basis of binding and specificity. Oligonucleotide binding occurs at a cleft formed across α -helix 1 and bounded by two unstructured surface loops. The first loop, between α -helices 1 and 2, contains an invariant GXXG sequence, crucial to oligonucleotide binding and forming contacts with the sugar-phosphate backbone. The second loop, between β -strands 2 and 3, is of variable length and sequence, and flanks the edge of the RNA binding cleft that the bases are directed towards. The oligonucleotide binding cleft is narrow and reported to more readily accommodate pyrimidine over purine bases, and four bases are accommodated within the oligonucleotide binding site (26,27,31). Here, surprisingly, specific contacts to cytosines at the two central positions are consistently observed, but the existence of specificity at the first and fourth position is still unclear.

Another feature of the α CP/oligonucleotide interaction that is not completely understood concerns the relative roles of the three KH domains and their overall topological arrangement when bound to target oligonucleotide (32). The optimal target sequence of the α CP-2KL isoform generated by *in vitro* SELEX contained three short C-rich patches within an exposed single stranded conformation spaced 2–6 nt apart (33). This suggests a close spatial arrangement of all three α CP KH domains. It is also notable that the disruption of any of the three C-rich patches within the α CP target site of α -globin mRNA interferes with α -complex formation and decreases α -globin stability *in vivo* (5,34). Interestingly, X-ray crystallographic and nuclear magnetic resonance (NMR) spectroscopic analysis of α CP KH domains reveal both hetero- and homo-dimer formation between KH domains (29–31). These protein–protein interactions can occur simultaneously to oligonucleotide binding. Whether these interactions occur within the context of the full-length protein bound to its target oligonucleotide is not yet known.

We have previously reported α CP binding to a specific UC-rich region of the 3'-UTR of androgen receptor mRNA (35). Through mutual binding with HuR, α CP1 and α CP2 are thought to be a part of the post-transcriptional control mechanism for androgen receptor expression. They have been shown to bind to a specific 5'-CCCUCCC-3' motif (AR mRNA nucleotides 3317–3323) immediately adjacent to a U-rich sequence (AR mRNA 3275–3316) which is the target for HuR binding. It is speculated that the combination of α CP and HuR binding plays a role in the regulation of AR mRNA gene expression. Here, we report a binding analysis of full-length α CP1, as well as the three separate α CP1-KH domains to their target sequence in the 3'-UTR of androgen receptor mRNA using electrophoretic mobility shift assays (EMSA) and surface plasmon resonance (SPR). A binding comparison is made with the analogous DNA sequence. We also report the crystallization and structure determination of α CP1-KH1 bound to a target DNA 11-mer analogous to the androgen receptor mRNA target site. This structure reveals the mode of interaction with the oligonucleotide, as well as the topological arrangement of two α CP1-KH domains bound at adjacent target sites. Further, SPR experiments are used to investigate the preferential binding of cytosine at each of the four nucleotide binding positions. This is further explored using a structural approach in which we determine the structure of a α CP1-KH1 mutant bound to a cytosine tetrad. Together, our data establish the lead role of KH1 in oligonucleotide binding by α CP1 and reveal the molecular basis of its specificity to a C-rich tetrad.

MATERIALS AND METHODS

Protein expression and purification

Full-length α CP1 or individual KH domains were expressed as fusion proteins with glutathione-S-transferase (GST). The DNA coding sequences comprising amino acids 1–356 (α CP1), 14–86 (α CP1-KH1), 97–150 (α CP1-KH2) and 279–356 (α CP1-KH3), were cloned into pGEX-6P2

plasmids and expressed by *Escherichia coli* BL21 (Codon⁺) in Luria Broth at 37°C. Protein expression was induced with 0.02 mM IPTG (isopropyl-D-thiogalactoside) at an optical density of 0.8 at 595 nm. The cells were harvested after 3 h of further growth by centrifugation, resuspended in phosphate buffered saline (PBS) (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4) containing 0.5% Triton X-100 and was then lysed by French Pressing (SLM Instruments, Inc.), supplemented with 0.5 mM phenylmethanesulphonyl fluoride (PMSF) and, in the case of full-length α CP1, a cocktail of protease inhibitors were used including aprotinin (2 μ g/ml), leupeptin (2 μ g/ml) and pepstatin (1 μ g/ml). The GST-fusion proteins were purified by affinity chromatography using glutathione agarose beads equilibrated with PBS buffer at 4°C and the GST was cleaved using 2U Prescission protease (GE Healthcare) in 50 mM Tris-HCl, pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT) at 4°C. α CP1 was finally purified using anion exchange chromatography (Mono-Q Pharmacia) and α CP1-KH1 and α CP1-KH2 purified by size exclusion chromatography using a Sephadex 75 column (Pharmacia). The purified proteins were dialysed into phosphate buffer pH 6.0 (1 mM DTT, 25 mM NaH₂PO₄, 150 mM NaCl, 1 mM EDTA), concentrated with centrifugal concentrators of 3 K cut-off (Millipore) and quantified using a detergent compatible (Biorad) protein assay. The cloning, overexpression and purification of the α CP1-KH1.W.C54S mutant protein is as described previously (36). In brief, this protein was prepared as a His-tagged protein, and the His-tag was not removed prior to crystallization trials.

Preparation of oligonucleotides for EMSA

Long RNA transcript. A 107 nt RNA containing vector sequences and nucleotides 3275–3325 of the androgen receptor mRNA 3'UTR was generated from the relevant DNA and cloned into pBLUESCRIPT II KS+ using Ambion *in vitro* transcription kits according to manufacturers specifications. After purification of full-length transcripts on denaturing polyacrylamide gels, RNAs were end-labelled with γ [³²P]ATP (Perkin Elmer). Radiolabelled RNAs were subjected to a second round of denaturing gel purification, then eluted, precipitated and resuspended in water. Prior to use, radiolabelled RNAs were heated to 70°C for 10 min, then quenched on ice and diluted to 2 \times the final concentration.

Short oligonucleotides. Synthetic RNA sequence 5'-UUC CCUCCCUA-3' corresponding to nucleotides 3315–3325 of AR mRNA, was purchased from Dharmacon (Dallas, Tx). The analogous synthetic DNA, of sequence 5'-TTCC CTCCTA-3', was purchased from Sigma Genosys (Castle Hill, NSW, Australia). Both were gel purified and 5'-end labelled as described above.

RNA and DNA EMSA

The EMSA was utilized for examining the binding of full-length α CP1 and α CP1-KH1 to target RNA and DNA. Purified full-length α CP1 and α CP1-KH1 were thawed and diluted to 2 \times the final concentration in

binding buffer [10 mM HEPES pH 7.5, 3 mM MgCl₂, 14 mM KCl, 5% (v/v) glycerol, 0.2% (v/v) Nonidet[®] P-40, 1 mM DTT]. A quantity of 5 μ l of diluted protein was mixed with 1 μ l of tRNA (1 μ g/ μ l) and incubated at room temperature for 10 min to prevent non-specific binding, before addition of an equal volume of labelled probe, also diluted to 2 \times the final concentration in binding buffer. After 30 min incubation at room temperature, 4 μ l of loading dye [binding buffer also containing 50% (v/v) glycerol, 0.1% (w/v) bromophenol blue and 0.1% (w/v) xylene cyanol] was added and the mixture loaded immediately on to a running non-denaturing polyacrylamide/0.5 \times tris/borate/ethylenediaminetetraacetic acid (TBE) gel and run at 10 V/cm. After electrophoresis, the gel was dried and then exposed to either a phosphorimager plate or X-ray film overnight for detection of protein and oligonucleotide complexes.

SPR studies

SPR (using a BIAcore 2000 or T100 instrument) was employed to characterize the α CP1-KH domain interactions with target RNA and DNA. A research grade CM5 chip coated with streptavidin was purchased from BIAcore. For comparison of KH1, KH2 and KH3 binding, 5'-biotinylated mRNA (5'-CUCUCCUUUCUUUUUCUUCUUCUCCUCCCUA-3') representing nucleotides 3296–3325 of androgen receptor mRNA was obtained from Dharmacon and immobilized on flow cell 2 as the captured molecule. 5'-biotinylated DNA (5'-CTC TCCTTCTTTTCTTCTTCCCTCCCTA-3') analogous to the above RNA sequence was also obtained from Dharmacon and immobilized on flow cell 3 as the captured molecule. For comparison of α CP1-KH1 binding to a series of tetrad DNA target sites, the following 5'-biotinylated sequences were purchased from Geneworks Australia: (5'-AAAAAATCCA-3'; 5'-AAAA AACTCA-3'; 5'-AAAAAACCTA-3'; 5'-AAAAAACCC A-3'; 5'-AAAAACCCCA-3').

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 μ l/min in immobilization buffer 10 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5% Triton X (or 0.025% P20) and 1–2 mM dithiothreitol, 2 mM EDTA and finally blocked with 1 mM Biotin in HEPES buffered saline (HBS) at 10 μ l/min for 3 min. An average of 30 RU of oligonucleotide was immobilized. α CP-KH domains were injected over flow cells 1–3 at a range of concentrations from 50–0.3 μ M using a flow rate of 50 μ l/min and at room temperature in running buffer 10 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5% Triton X (or 0.025% P20) and 1–2 mM dithiothreitol, 2 mM EDTA, 66–125 μ g/ml tRNA and 62.5 mg/ml bovine serum albumin. All experiments were replicated (duplicate or triplicate) to confirm the reproducibility of the signal. Regeneration involved removal of the bound protein from the streptavidin chip with a 1–2 min wash at 20–50 μ l/min with 2 M NaCl. Data were analysed with the BIAevaluation software to obtain a binding constant using a steady state model.

Preparation of DNA targets for co-crystallization

The DNA '11-mer' sequence 5'-TTCCCTCCCTA-3' analogous to nucleotides 3315–3325 of AR mRNA, was purchased from Dharmacon in crude form and further purified by denaturing PAGE. After the separation of the sample by 20% PAGE, the band was visualized by UV shadowing when the DNA had run approximately half way down the gel. The sample was recovered by excising the band, which was then crushed and eluted overnight in 0.3 M sterile sodium acetate at 37°C. The eluent was filtered and desalted using a reverse-phase solid-extraction cartridge (C18 Sepak cartridge, Waters). The eluted fractions were lyophilized and dissolved in distilled water for quantification using UV spectroscopy. The DNA '6-mer' sequence 5'-ACCCCA-3' was purchased from Geneworks Australia in purified form and resolubilized in water without further purification steps. DNA concentrations were determined by measuring the absorbance at 260 nm and assuming one absorbance unit to be equivalent to 34 µg/ml.

Preparation of α CP1–KH1/DNA complexes for co-crystallization

DNA and protein complexes were prepared by dissolving the lyophilized DNA with the protein solution to a final molar ratio of 1:1. The mixture was left on ice for 30 min prior to setting crystal drops. Crystals were grown using vapour diffusion in 1 µl hanging drops containing 1:1 mixtures of protein and reservoir solutions. The α CP1–KH1/11-mer complex solution contained 309 µM of protein and DNA in 50 mM Tris–HCl pH 8.0, 1 mM DTT, 1 mM EDTA, 150 mM NaCl, and the reservoir solution was composed of 0.1 M Na cacodylate pH 6.5 in 0.2 M magnesium acetate, 30% MPD (2-methyl-2,4-pentanediol) Sigma Crystal Screen reagent formulation number 21. Crystals typically grew in 2 months to dimensions of $\sim 0.2 \times 0.2 \times 0.04$ mm with the outline of a diamond. The α CP1–KH1.W.C54S/6-mer were produced using 0.1 M phosphate–citrate pH 4.2, 40% (v/v) PEG 300 as the precipitant as reported (36).

X-Ray data collection, structure determination and refinement of α CP1–KH1/DNA complexes

Data for the α CP1–KH1/11-mer complex were recorded in-house using a Rigaku RUH2R X-ray source with a rotating copper anode equipped with Osmic confocal optics, a MAR Research MAR345 detector and an Oxford cryosystems 600 series cryostream from a single crystal. The diffraction measurements indicated that the space group was $P2_1$ with unit cell constants $a = 45.6$, $b = 76.8$, $c = 61.4$ Å and $\beta = 111.7^\circ$ and diffraction was recorded to 3.0 Å resolution. Data were integrated and scaled with DENZO and SCALEPACK (37). Structure factor amplitudes were calculated using TRUNCATE (38). The data collection statistics are given in Table 1.

Initial phases were derived by molecular replacement using the protein coordinates from the Nova2-KH3/RNA structure (pdb accession code: 1EC6) as a search

model. The search model included residues 4–76 with the oligonucleotide and solvent molecules removed. The programme PHASER (39) was used to locate four molecules in the asymmetric unit. Electron density derived from the initial set of phases revealed the positions of the two molecules of DNA in addition to the four KH domains in the asymmetric unit. The protein and oligonucleotide were built using COOT (40). To aid in interpretation of the initial electron density maps, 4-fold non-crystallographic symmetry averaging was applied by the programme DM (38). Tight non-crystallographic symmetry constraints were maintained throughout the refinement. Translation libration screw (TLS) refinement was also utilized. The refinement statistics are reported in Table 1 and the atomic coordinates are available from the Protein data bank (PDB code: 1ZTG).

The data collection for the α CP1–KH1.W.C54S/6-mer complex is described previously (36). The structure was solved by molecular replacement using PHASER (39) included in the CCP4 suite (38). A single copy of the polypeptide chain was extracted from the coordinate of α CP1–KH1/11-mer structure and used as search model for both rotation and translation functions. Four copies of the protein were identified in the asymmetric unit and the oligonucleotide was manually built in COOT (40). Several cycles of refinement were carried out using non-crystallographic symmetry constraints in REFMAC5 (41) and PHENIX (42). Water molecules were added by ARP/WARP (43) and amino acids side chains were manually checked and rebuilt in COOT. TLS refinement was also utilized. Data collection statistics are previously reported (36) and the refinement statistics are reported in Table 1. The structure factors and the coordinate of the final model are available in the protein data bank (PDB code: 3VKE).

RESULTS

Multimeric RNA complexes are formed by full-length α CP1, but not α CP1–KH1 alone

α CP1 has been shown to bind in a sequence-specific manner to a UC-rich region, located in the 3'-UTR of androgen receptor mRNA (nucleotides 3275–3325), that is implicated in the regulation of AR mRNA stability and translational efficiency (35). We have previously shown that α CP1–KH3 independently binds this sequence and that α CP1–KH2 does not using EMSA (44). In the current study, we examine the binding of this AR mRNA sequence by full-length α CP1 and α CP1–KH1.

The binding of full-length α CP1 to the target 51-nt AR mRNA sequence in an EMSA experiment is shown in Figure 1A. The probe (10 nM RNA) is shifted upon the addition of increasing concentrations of α CP1 (10 nM–1 µM), with half of the probe shifted when between 20 and 50 nM, α CP1 is added. This is consistent with a previous report of a K_D in the nM range (35). Notably, full-length α CP1 shifts the probe to positions that increase in size with α CP1 concentration revealing the formation of higher molecular weight complexes. Binding by

Table 1. Statistics for X-ray crystallographic data collection and refinement

	α CP1-KH1/11-mer	α CP1-KH1.W.C54S/6-mer
Data collection ^a		
Space group	$P2_1$	$P2_1$
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	45.6, 76.8, 61.4	38.6, 114.9, 43.4
<i>abg</i> (°)	90.0, 111.7, 90.0	90.0, 93.4, 90.0
Resolution (Å)	30.0 – 3.0 (3.11 – 3.0)	30.0 – 1.77 (1.87 – 1.77)
R_{merge} (%)	5.7 (65.8)	9.0 (60.2)
$I/\sigma(I)$	22.7 (2.1)	7.9 (2.2)
Reflections measured (total)	35 603 (394)	113 451 (16 460)
Completeness (%)	99.8 (99.9)	99.7 (100)
Redundancy	4.4 (4.0)	3.2 (3.1)
Refinement		
Resolution (Å)	3.0	1.77
No. of reflections used (unique)	7994	34 271
$R_{\text{work}}/R_{\text{free}}$	23.0 / 27.9	16.4 / 21.3
No. of atoms		
Protein	2193	2281
Oligonucleotide	308	392
Water		71
<i>B</i> -factors (Å ²)		
Protein	79.2	30.4
Oligonucleotide	91.5	34.3
Water		43.1
RMS deviations		
Bond lengths (Å)	0.007	0.027
Bond angles (°)	1.164	2.243
Ramachandran plot ^c (%)		
Favoured regions	95.3	98.9

The statistics for both the α CP1-KH1/11-mer and α CP1-KH1.W.C54S/6-mer structures are shown.

^aValues in parentheses refer to the highest resolution shell.

^b $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$ where *I* is the intensity of individual reflections.

^cDetermined using the program MOLPROBITY(45).

α CP1-KH1 to the target RNA is also shown in Figure 1A. The single KH domain shifts the target RNA (likely via a 2:1 protein:RNA binding stoichiometry) to a discrete position showing that full-length α CP1, but not α CP1-KH1 alone, forms large multimeric complexes with RNA.

Full-length α CP1 and α CP1-KH1 bind to target RNA, as well as DNA

The ‘CCCUECC’ motif at the 3′-end of the target 51-nt AR mRNA has been shown to be the binding site of α CP proteins through mutational analysis of the two poly(C) triads (35). In order to verify whether full-length α CP1 and α CP1-KH1 binding occurs to this motif *in vitro*, we conducted gel shift assays using an 11-nt probe corresponding to nucleotides 3315–3325 of AR mRNA (5′-UU CCCUECCUA-3′). Figure 1B shows that the probe is shifted by full-length protein to a constant position, indicating good binding to the probe via a single binding interaction. α CP1-KH1 also binds, but only marginally shifts the probe under the conditions of this experiment, indicating a weaker interaction. Interestingly, the binding profiles of α CP1 and α CP1-KH1 to an 11-nt DNA probe analogous to the AR target sequence above (DNA: 5′-TT CCCTCCCTA-3′) are very similar to the binding observed to RNA (Figure 1B). Full-length α CP1 and α CP1-KH1 bind to the DNA with good and weak binding respectively. Notably, the binding of full-length α CP1 appears to

shift more of the DNA probe than the RNA probe, suggesting a marginally higher affinity for DNA over RNA.

α CP1-KH1 forms the most stable interactions to a C-rich RNA and DNA

Having shown that α CP1-KH1 binds to both DNA and RNA using EMSA, we utilized SPR for a more sensitive comparison of KH domain binding affinities and binding kinetics to target RNA and the analogous DNA sequence. SPR binding curves showing a series of α CP1-KH interactions (at concentrations of between 0.312 μ M and 10 μ M) with a 30-nt oligonucleotide representing AR mRNA (nucleotides 3296–3325) are shown in Figure 2. This RNA sequence includes 19-nt of U-rich sequence preceding the 11-nt C-rich target sequence (as used in the EMSA above) as a spacer between the Biacore chip and the binding site. Alongside these curves are the sensorgrams obtained simultaneously using the analogous sequence of DNA. It is to be noted that the binding curves exhibited complex binding kinetics, most likely due to the lengthy oligonucleotides immobilized on the chip and the presence of two cytosine triplets at the 3′-ends that constitute two α CP1-KH domain target sites. Where binding was observed, an approximated steady state binding analysis is presented.

α CP1-KH1 was found to readily associate with the RNA and a slow dissociation phase was observed,

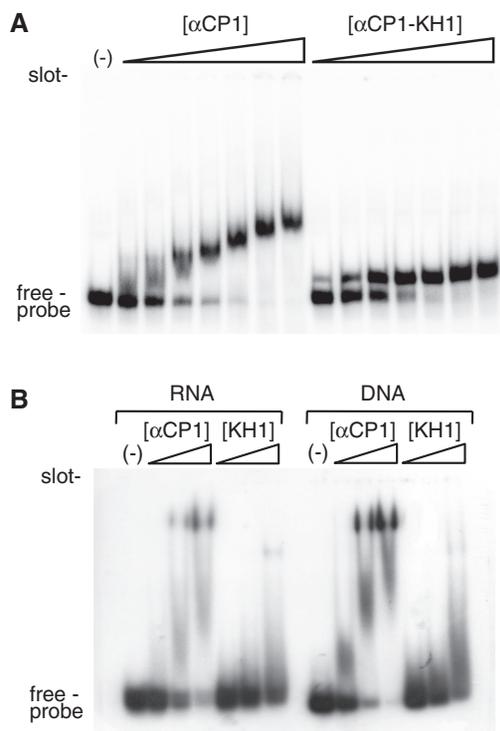


Figure 1. Binding of α CP1 and α CP1-KH1 to oligonucleotides representing the UC-rich region of androgen receptor 3'-UTR. (A) A mobility shift assay, run on a 5% PAA/0.5 \times TBE gel is shown. 5'-end labelled RNA was incubated either alone (-) or in the presence of purified α CP1 or α CP1-KH1 as indicated above the gel. The probe comprises a total of 107-nt RNA containing vector sequences and nucleotides 3275–3325 of the androgen receptor mRNA. Increasing protein concentrations are indicated by a wedge and are, from the left: 1×10^{-8} M, 2×10^{-8} M, 5×10^{-8} M, 1×10^{-7} M, 2×10^{-7} M, 5×10^{-7} M or 1×10^{-6} M for both proteins. In all cases, RNA concentration is 1×10^{-8} M. All lanes also contained 1 μ g yeast tRNA added as a non-specific competitor. (B) A mobility shift assay, run on a 10% PAA/0.5 \times TBE gel is shown. 5' -end labelled 11-mer RNA or DNA was incubated either alone (-) or in the presence of purified α CP1 or α CP1-KH1 as indicated above the gel. Probe sequences are: RNA: 5'-UUCCCUCCUA; DNA: 5'-TCCC TCCCTA. Increasing protein concentrations are indicated by a wedge and are, from the left: for α CP 1, 1×10^{-7} M, 3×10^{-7} M and 1×10^{-6} M; for α CP1-KH1, 1×10^{-6} M, 3×10^{-6} M and 1×10^{-5} M. In all cases, RNA or DNA concentration is 1×10^{-7} M. All lanes also contained 1 μ g yeast tRNA added as a non-specific competitor.

indicating a relatively stable interaction. A dissociation equilibrium constant (K_D) calculated from a steady state analysis was estimated to be $3.6 \pm 0.6 \mu\text{M}$. In the case of α CP-KH1 binding to DNA, a relatively slow dissociation was also observed and the K_D was determined to be $0.33 \pm 0.04 \mu\text{M}$, representing ~ 10 -fold increase in affinity of α CP1-KH1 for DNA over RNA.

The interactions of α CP1-KH2 with target RNA and DNA were also examined for comparison. The α CP1-KH2 domain has not previously been shown to be able to independently bind to RNA (32,44). Consistent with these reports, no binding interaction was detectable between α CP1-KH2 and RNA even at the highest protein concentrations. However, we detected binding to DNA, albeit with a low response. A steady state analysis

of the α CP1-KH2 interaction with DNA yielded a K_D of $13 \pm 4 \mu\text{M}$. This indicates that a functional α CP1-KH2 was formed, but that interactions with RNA could not be detected.

In contrast, α CP1-KH3 bound to both RNA and DNA with similar affinities to α CP1-KH1. Dissociation equilibrium constants calculated from a steady state analysis were $K_D = 3.5 \pm 0.8 \mu\text{M}$ and $K_D = 1.5 \pm 0.2 \mu\text{M}$, respectively. The complex half-life of α CP1-KH3 with DNA and RNA, however, were shorter than that for α CP1-KH1, with response curves returning to the baseline promptly at the end of the protein injection phase.

Thus, oligonucleotide binding could be measured for all three individual α CP1-KH domains with the exception of KH2 binding to RNA. KH1 engages in the most stable interaction with both RNA and DNA of the three KH domains. KH2 and KH3 are also both capable of binding DNA.

Structural overview of α CP1-KH1 bound to adjacent CCCT binding sites

In order to examine the structural basis for the differences in KH domain interactions with target oligonucleotide, crystal trials were undertaken with each of the three α CP1-KH domains in complex with the DNA sequence 5'-TTCCCTCCCTA-3' (analogous to nucleotides 3315–3325 of AR mRNA). Only α CP1-KH1 yielded crystals, which was not unexpected considering that KH1 forms the most stable complexes with DNA. α CP1-KH1 (residues 14–86, numbered as in Swiss-Prot entry Q15365, preceded by the sequence GPLGSPGI present due to cloning procedures) yielded crystals containing two crystallographically independent copies of a 2:1 protein–DNA complex in the asymmetric unit. Equivalent crystallization experiments utilizing RNA did not produce crystals suitable for structure determination. Experimental phases were obtained by molecular replacement using coordinates from the Nova2-KH3 structure (PDB ID:1EC6) with the oligonucleotide removed. The current refinement model has a working R -factor of 23.0% and a free R -value of 27.9% at 3.0 Å resolution (Table 1), with good stereochemistry 95% in the favoured regions of a Ramachandran plot (100% in the allowed regions). The coordinates are deposited in the protein structure database (pdb id:1ZTG).

The α CP1-KH1/11-mer structure reveals two α CP1-KH1 domains bound at adjacent CCCT sequences (Figure 3A). The final model defined the position of residues 13–83 of α CP1-KH1 and the four bound bases in each of the binding clefts. Although the two KH domains bound to the same oligonucleotide are held very closely, they do not make contact with one another (Figure 3A). Interestingly, in this crystal form, each KH1 domain is covalently linked to an adjacent KH1 domain by a disulphide bond formed through their C54 residues. Furthermore, each KH domain exists as a dimer with a KH domain from an adjacent unit as previously observed for other KH domain structures (26,27,30,31).

The protein conforms to the classical type I KH domain structure, with a three-stranded anti-parallel β -sheet

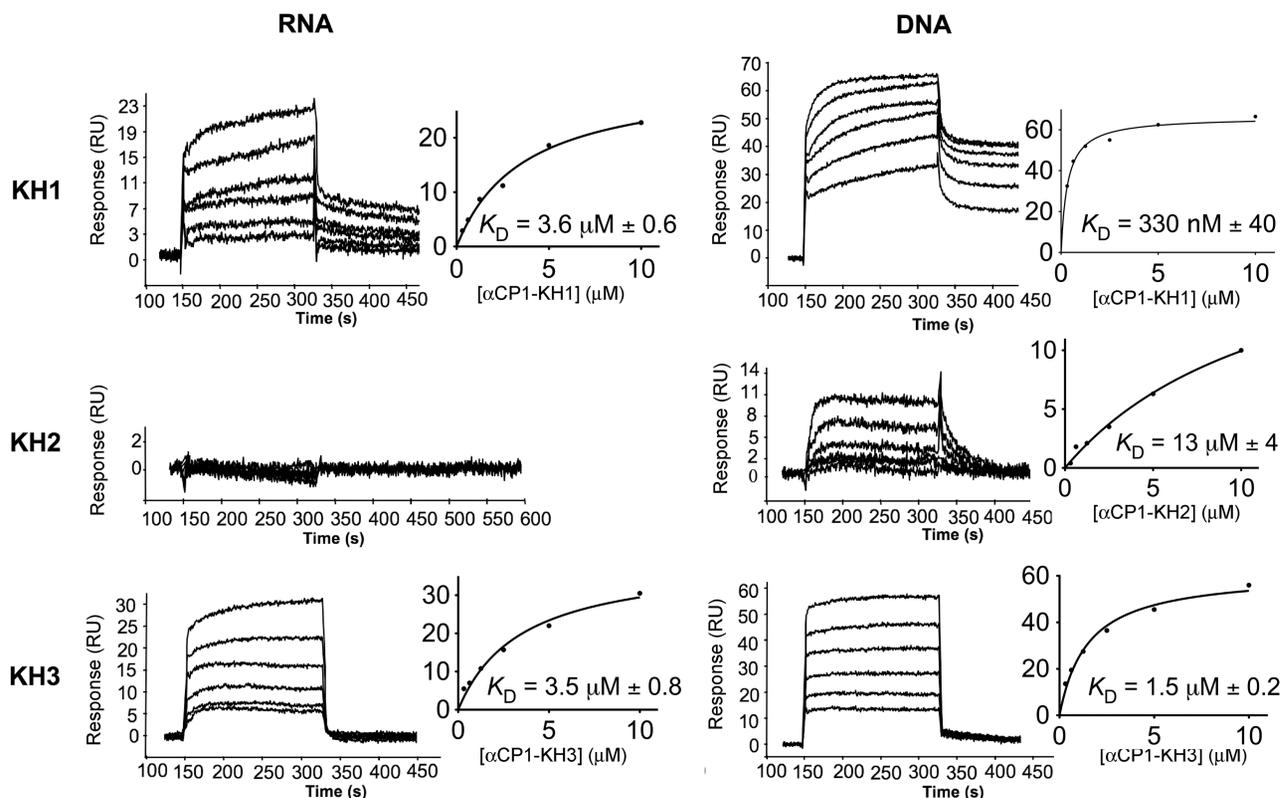


Figure 2. Binding analysis of separate KH domains of α CP1 to target RNA and DNA using SPR. Sensorgrams of α CP1 KH1, KH2 and KH3 binding to biotinylated mRNA (5'-CUCUCCUUUCUUUUUCUUCUCCUCCCUA-3') representing nucleotides 3296–3325 of androgen receptor mRNA (flow cell 2) and biotinylated DNA (5'-CTCTCCTTTCTTTTCTTCTTCCCTCCCTA-3') analogous to the above RNA sequence (flow cell 3) captured on SA-coated sensor chips at a range of protein concentrations are shown. Binding curves, derived from the approximated steady state binding of the proteins, were used to determine equilibrium dissociation constants (K_D s). Errors are standard errors arising from fits.

packed against three α -helices in a $\beta\alpha\alpha\beta\beta\alpha$ topological arrangement (Figure 3B). The structure is consistent with that reported for the α CP2-KH1 homologue bound to a telomeric DNA sequence solved to 1.7 Å resolution [pdb id: 2AXY; identity 97% and $C\alpha$ pair-wise root mean square deviation (RMSD) 1.1 Å] (31). A hydrophobic core provides the structure's stability, with hydrophobic residues emanating from the inner face of the β -sheet (including Leu14, Ile16, Leu18, Met20, Ala45, Ile47, Ile49, Ile59 and Leu61) and all three helices (including Glu24, Val25, Ile28, Val36, Ile39, Arg40, Ala67, Ile68, Ala71, Ile75, Lys78 and Leu79). This core is partly exposed to create the base of the hydrophobic oligonucleotide binding cleft.

The dimer interface is formed via an anti-parallel interaction between the third α -helices of paired α CP1-KH1 domains via hydrophobic interactions including a ring stacking arrangement of Phe72 side chain aromatic rings. This interaction also brings together the first β -strand of each α CP1-KH1 domain to effectively form a continuous six-stranded anti-parallel β -sheet, though the two strands are only connected via backbone hydrogen bonds between Arg17 NH and Leu19 CO of respective KH domains. Additional inter-molecular hydrogen bond interactions are made from the Arg17 side chain guanidinium to both the Leu19 backbone carbonyl and

Glu24 side chain carboxylic acid groups of the opposing α CP1-KH1 domain and from the Glu80 side chain to the Thr65 side chain hydroxyl of the opposing α CP1-KH1 domains.

The oligonucleotide is accommodated in a hydrophobic cleft formed across the top of α -helix 1 and bounded by the 'GXXG' motif and variable loop (Figure 3B). The DNA binding cleft is on the opposite side of the molecule to the dimerization interface so that both interactions can occur simultaneously. Basic residues surrounding the binding site, including Lys 23, Lysines 31 and 32 at the 'XX' positions, Lys37 and Arginines 40, 46 and 57, create a positive potential along the length of the cleft (Figure 3C). Such a potential, similarly observed for α CP1-KH3 (28), likely provides a driving force for the docking of the oligonucleotide to the site, and also provides specific electrostatic contacts to the bound oligonucleotide. The two α CP1-KH1 monomers bound to the 11-base oligonucleotide make equivalent contacts to sequential 'CCCT' tetrads, so that the two monomers are arranged in a tail to head arrangement on the DNA. The oligonucleotide is arranged in a linear fashion with bases arranged planar to the surface of the protein except for Thy-4 of each tetrad that partially base stacks over Cyt-3. Only a single intra-molecular hydrogen bond between Cyt-2 N4 of the first (5') tetrad and the phosphate

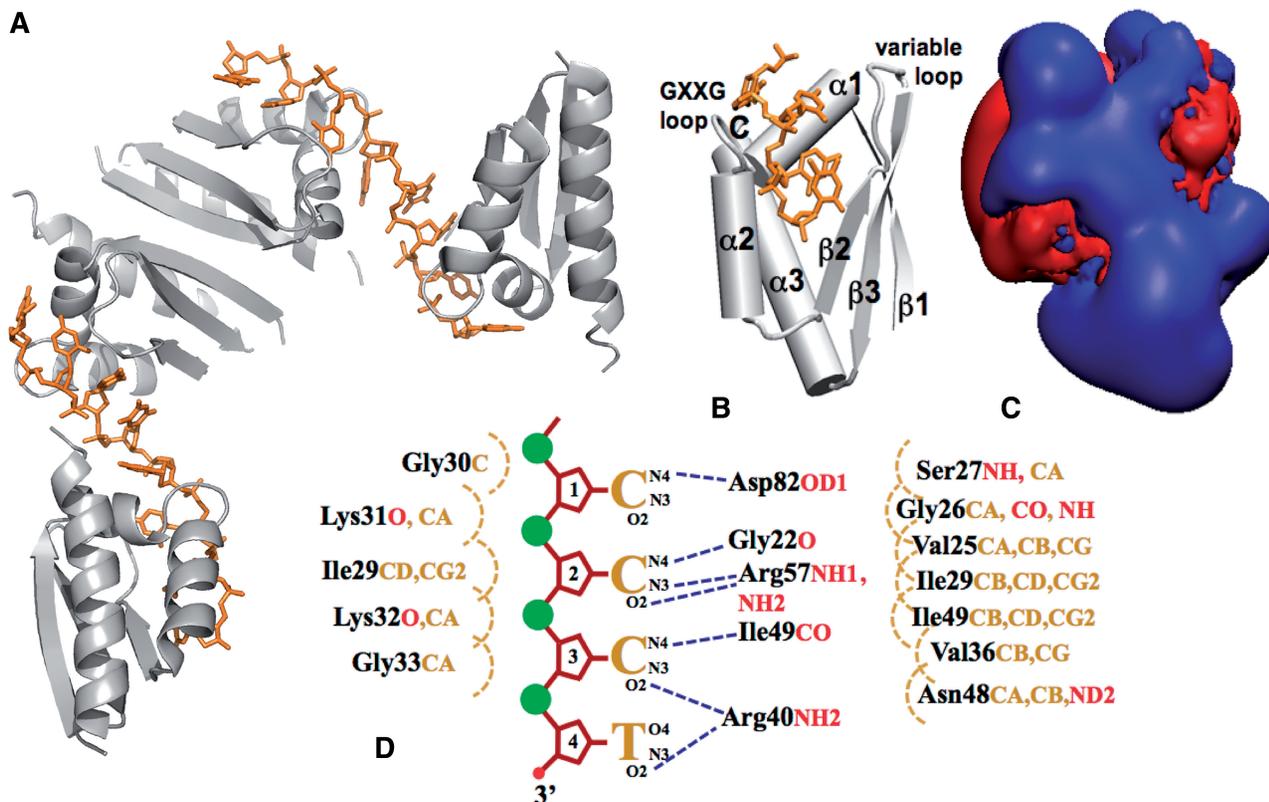


Figure 3. Schematic representations of the α CP1-KH1/11-mer DNA complex. (A) α CP1-KH1 formed as a dimer with two protein molecules bound to a single 11-nt strand of DNA, resulting in a continuous network throughout the crystal. (B) A cartoon representation of the KH domain bound to the target DNA. The 5'-tetrad of the target DNA that form contacts with the first KH domain is shown, illustrating the positioning of the critical bases about α -helix 1 and between the GXXG and variable loops. (C) The electrostatic potential emanating from the α CP1-KH1 (in the same orientation as cartoon alongside) structure calculated using the APBS software package (46). Potential contours are shown at +1 kT/e (blue) and -1 kT/e (red) and were 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 positive potential directing oligonucleotides to the binding cleft. (D) Summary of the contacts between α CP1-KH1 and bound DNA tetrad of sequence 5'-CCCT-3'. Van der Waals contacts are coloured orange, and hydrogen bond interactions are coloured blue. The residues making important contacts with the oligonucleotide sugar-phosphate backbone are listed on the left, and the residues making contacts with the pyrimidine rings, and thus underlying base specificity, are listed on the right.

preceding Cyt-1 is formed. The oligonucleotide twists about the phosphate bond of the fourth base to allow the second KH1 monomer to bind downstream of the first site $\sim 180^\circ$ about the oligonucleotide axis (Figure 3A). Similarly to reports of other KH domains bound to DNA with multiple binding sites (26,30), this shows the way in which two KH domains may be closely juxtaposed when bound at adjacent C-rich binding sites.

Binding specificity of α CP1-KH1 for cytosines in the CCCT tetrad

Of particular interest in this study, was the determination of the inter-molecular contacts underlying cytosine specificity. The α CP1-KH1/11-mer interactions are summarized in Figure 3D. As seen for isoform α CP2-KH1 (30), the central two cytosines of the bound tetrad (Cyt-2 and Cyt-3) engage in hydrogen bond interactions that form the basis for cytosine specificity at these positions. The cytosine in the second position forms van der Waals contacts with Val25, Gly26 and Ile29. The basis for cytosine specificity in this position is dominated by

interactions with Arg57 which projects from the variable loop region and forms bipartite hydrogen bonds to Cyt-2 O2 and N3 atoms. In addition, the backbone carbonyl of Gly22 is able to form a hydrogen bond with Cyt-2 N4. The cytosine in the third position (Cyt-3) is positioned over hydrophobic residues Ile29, Ile49 and Val36. Binding specificity for Cyt-3 is conferred by Ile49 and Arg40 which are conservatively substituted and conserved respectively in poly(C)-binding proteins. The backbone carbonyl of Ile49 is positioned to form a hydrogen bond with Cyt-3 N4. Arg40 extends from the C-terminal end of α -helix 2 and is able to make hydrogen-bond contact with the Cyt-3 O2 atom.

Unlike previous studies, this structure also reveals possible interactions underlying base specificity at positions 1 and 4. Cyt-1, in the first position, forms a hydrogen bond with the side chain of Asp82 via its N4 group. In addition, non-specific interactions between Cyt-1 and Gly26, Ser27, Gly30 and Lys31 backbone atoms, that form part of the oligonucleotide binding cleft, are also observed. Thy-4 is not extensively contacted by the KH domain. It forms base stacking interactions

with Cyt-3 rather than forming many protein contacts. However, Arg40 is positioned such that its guanidinium nitrogens are positioned to form hydrogen bonds with both Thy-4 and Cyt-3 O2 groups. Such an interaction could favour either thymine or cytosine at position 4.

Cytosine is preferentially bound by α CP1-KH1 at all four base positions

In previous studies of α CP-KH1 domains, it has been shown that the first position in the oligonucleotide binding cleft can accommodate an A, T or C, with little to indicate a preferred base at this position (30,31). The fourth position has been shown to accommodate T or C, but not a purine base. Yet, systematic evolution of ligand by exponential enrichment (SELEX) studies originally performed to characterize the α CP target sequence clearly identified sequences containing C-triplets and tetrads (33).

To investigate the position-specific binding preferences of α CP1-KH1, we used SPR to monitor binding to a series of oligonucleotides in which cytosines were systematically replaced by adenine or thymine. Figure 4 shows sensorgrams of α CP1-KH1 binding to different 10-mer DNA containing oligonucleotide target tetrads and the steady state binding analysis in the cases where quantifiable binding was measured. A 5-nt A-rich spacer was included at the 5'-end of the oligonucleotide to distance the binding site from the matrix surface and another adenine was added at the 3'-end to restrict the possible binding mode to the target tetrad since purines have not been observed to bind at the fourth position. By this reasoning, we could assume that the target tetrad (underlined in each of the sequences) would be bound in the same register at the α CP1-KH1 binding site.

The first experiment was conducted with the sequence 5'-AAAAAACCCA-3'. The binding to this short single binding site DNA sequence occurred with a fast off-rate compared with the slow off-rate observed in the previous experiments, and readily reached equilibrium. Binding by α CP1-KH1 to this oligonucleotide was found to occur with micromolar affinity ($K_D = 33 \mu\text{M}$). The second experiment, in comparison with the first, shows that cytosine in position 2 is important for binding. The sequence 5'-AAAAAATCCA-3' (in which Cyt-2 is replaced by thymine) showed no binding by α CP1-KH1. This experiment also confirmed the assumption that adenine cannot be accommodated in position 4 (since, if it could, there would be no impediment to TCC binding in positions 1, 2 and 3). The non-binding of the ATCC sequence then, demonstrates the critical role of cytosine at position 2. Likewise, in the third experiment, the sequence 5'-AAAAAACTCA-3' (in which Cyt-3 is replaced) showed no binding by α CP1-KH1. By the same logic, this confirms the critical role of cytosine in position 3. In contrast, the fourth experiment shows that cytosine is preferred, but not essential, in the fourth position. The sequence 5'-AAAAAACCTA-3' (in which the Cyt-4 is replaced), showed some binding (though not quantifiable). The binding was reduced compared with the oligonucleotide containing a C-triplet, and therefore,

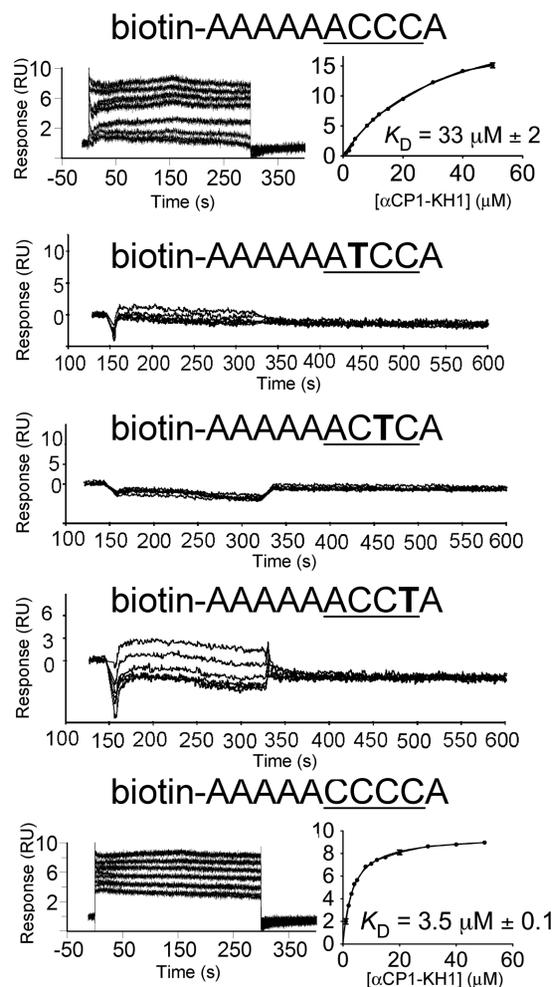


Figure 4. Binding analysis of α CP1-KH domains of α CP1 to target RNA and DNA using SPR. Sensorgrams of α CP1-KH1 binding to a series of biotinylated DNA sequences, designed to test the preferential binding of cytosine at each of the four nucleotide binding positions. The sequences are shown above each set of sensorgrams with the binding tetrad underlined. Five adenines were used as a spacer between the biotin and the oligonucleotide binding site. Oligonucleotides were captured on SA-coated sensor chips at a range of protein concentrations. Binding curves, derived from the approximated steady state binding of the proteins, were used to determine equilibrium dissociation constants (K_D s). Errors are standard errors arising from fits.

shows that a thymine is tolerated in position 4, though cytosine is preferred. Last, the sequence 5'-AAAAACCCCA-3' containing a C-rich tetrad was bound by α CP1-KH1 with 10-fold higher affinity ($K_D = 3.5 \mu\text{M}$) than the ACCC sequence. This enhanced binding suggests that, in fact, a cytosine is preferred over adenine in the first position. From this series of experiments, it was established that in fact, there is a binding preference for cytosine by α CP1-KH1 at positions 1 and 4 (as well as 2 and 3).

High resolution structure α CP1-KH1.W.C54S bound to a CCCC tetrad

In order to obtain higher resolution structural data for the α CP1-KH1/DNA complex, we strategically designed

protein and DNA constructs for co-crystal studies. Having recognized the tendency of C54 to form inter-molecular disulphide bonds upon crystallization, we prepared a C54S α CP-KH1 mutant and also added a tryptophan to the C-terminus to facilitate detection at 280 nm that would assist with purification. This construct, referred to as α CP1-KH1.W.C54S formed high quality crystals in complex with a C-tetrad DNA sequence 5'-A CCCC-3' established by our binding studies as a high affinity ligand. The crystals were grown as reported and diffracted to 1.77 Å resolution (36). The structure was solved by molecular replacement using the coordinates of a monomer from the α CP1-KH1/11-mer DNA complex as starting model (pdb id: 1ZTG). The oligonucleotide was removed from the coordinate prior to calculation. The refined model has a working *R*-factor of 16.35% and a free *R*-value of 21.34% at 1.77 Å resolution with an excellent stereochemistry including 99% of all amino acids in the favoured regions of a Ramachandran plot (100% in the allowed regions) (Table 1). As indicated in the coordinate file, the final model includes residues 14–86 of α CP1 and supplementary residues due to cloning procedures at N- and C-terminus. The coordinates are deposited in the protein structure data base (pdb id:3VKE).

The α CP1-KH1.W.C54S/6-mer structure reveals the same organization as that described above for the α CP1-KH1/11-mer structure (*C α* pair-wise RMSD 0.395 Å), establishing that the W and C54S mutations do not affect the structure or binding of oligonucleotide. The asymmetric unit contains four monomers organized as a pair of dimers. Each monomer binds a single copy of the 6-mer DNA. The position of 6 nt out of 6 was clearly identified in two chains in the electron density map (Figure 5A). In the two remaining chains, only the central four nucleotides are present. Absence of the extreme nucleotides can be attributed to a crystallization artefact, since those positions coincide with the crystallographic 2-fold screw axis. The oligonucleotide is arranged in an analogous fashion to that observed for the α CP1-KH1/11-mer DNA. In this case though, as well as partial base stacking of Cyt-4 across Cyt-3, base stacking is also observed for the terminal adenosine bases. Interestingly, an intra-molecular hydrogen bond is formed between Cyt-2 N4 and the phosphate preceding Cyt-1, as also seen for the α CP1-KH1/11-mer DNA. In addition, an equivalent hydrogen bond is formed between Cyt-4 N4 and the phosphate preceding Cyt-3. Such hydrogen bonds, that are uniquely formed by cytosine, may help to stabilize the bound oligonucleotide conformation.

Cytosine recognition at all four nucleotide binding positions

The four cytosines are positioned in the hydrophobic binding cleft in similar positions as seen for the CCCT sequence in the α CP1-KH1/11-mer structure (Figure 5B). The α CP1-KH1.W.C54S/6-mer interactions are summarized in Figure 5C and D. The central two cytosines of the tetrad (Cyt-2 and Cyt-3) are positioned analogously to that previously described. Specific contacts to

Cyt-2 are made via hydrogen bond interactions with the Arg57 guanidinium side chain and Gly22 backbone oxygen groups. Specific contacts to Cyt-3 are made via hydrogen bonds with the Arg40 guanidinium side chain and the backbone carbonyl of Ile49.

In addition to these contacts underlying specificity at positions 2 and 3, the α CP1-KH1.W.C54S/6-mer structure revealed further contacts that underlie a preference for cytosine in position 1, as well as the hydrogen bond between the side chain of Asp82 and the Cyt-1 N4 group, a hydrogen bond is also seen between the Lys31 backbone NH and the Cyt-1 O2 group. In the case of Cyt-4, a clear difference is seen from the α CP1-KH1/11-mer structure in which thymine was present at this position. The side chain of Glu51, that was not positioned towards the oligonucleotide in the α CP1-KH1/11-mer structure is clearly defined in the density of the α CP1-KH1.W.C54S/6-mer structure and forms a hydrogen bond to the Cyt-4 N4 group. Together, these favourable and specific interactions likely contribute towards the higher affinity of a C-tetrad at the binding site of α CP1-KH1.

Interestingly, the 5'-Ade also makes contact with the α CP1-KH1 binding cleft in the α CP1-KH1.W.C54S/6-mer structure. It forms base stacking interactions with Cyt-1 and also is able to form hydrogen bonds with the Ser85 side chain hydroxyl. Ser85 was not positioned to interact with the oligonucleotide in the α CP1-KH1/11-mer structure, suggesting that an additional interaction with adenine, not previously detected, to our knowledge, in the study of KH domains, may play a role in the binding of α CP1-KH1 to target oligonucleotides. In contrast, the 3'-Ade forms base stacking interactions with Cyt-4 and is held away from the surface of the protein where it does not make direct protein interactions.

DISCUSSION

PCBP are triple KH domain proteins that exert an extraordinarily diverse array of functions through their ability to interact with C-rich single stranded oligonucleotides. This includes the stabilization of mRNA, translational repression, translational enhancement, transcriptional enhancement and possible involvement in telomere functioning. This functional diversity suggests that PCBP do not confer a specific functional outcome, but may play a general scaffolding role and help to form a variety of functional oligonucleotide structures, depending upon the target C-rich sequence. Such a role could be achieved via the interactions of the individual KH domains of PCBP with adjacent C-rich regions of the target oligonucleotide, and also through intra- and/or inter-molecular interactions between PCBP and other RNA-binding proteins.

In the current study, we focus on the PCBP α CP1 (also known as PCBP1 and hnRNP E1). In the previous work, we identified α CP1 as one of the proteins that targets the androgen receptor (AR) mRNA at a specific 3'-UTR site, and plays a role in the regulation of its translation (35). In an effort to better understand the mode of interaction of α CP1 with this mRNA target, we first undertook EMSA experiments with the 51-nt target (representing AR

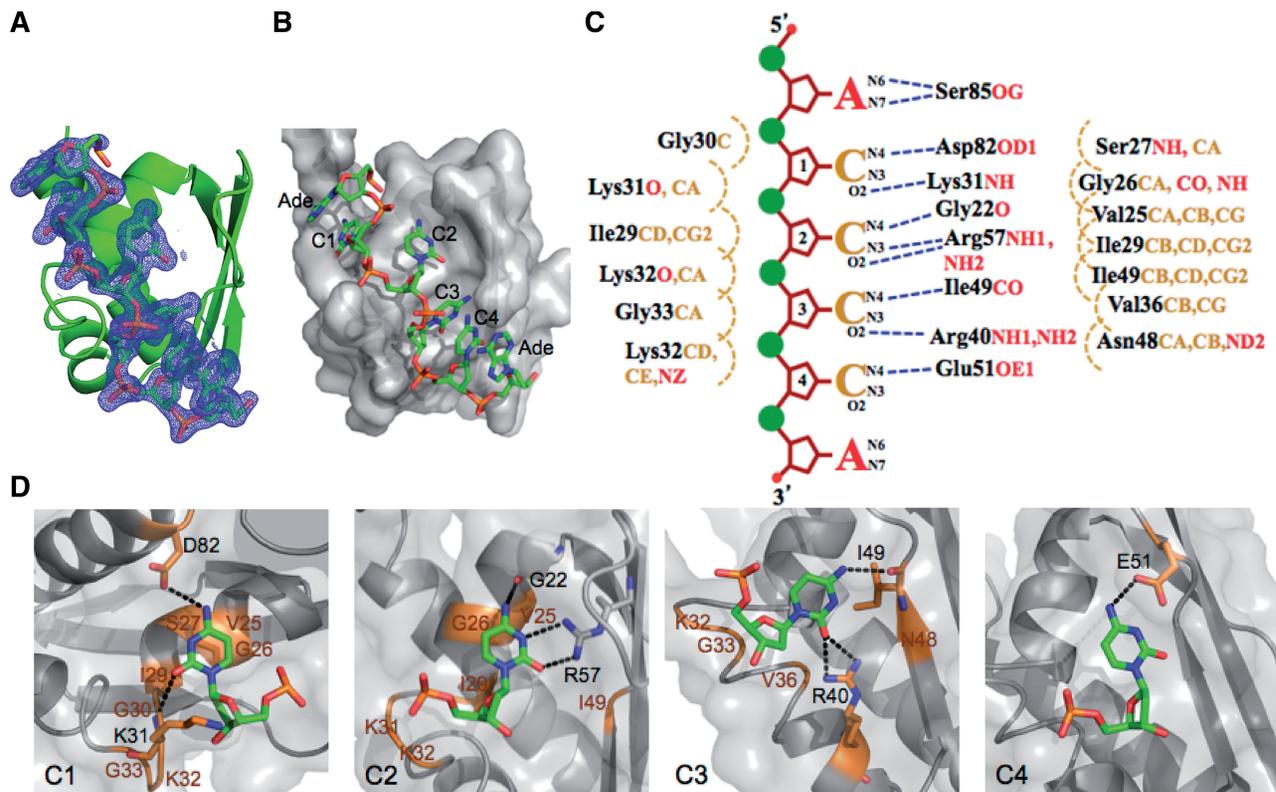


Figure 5. Schematic representations of the α CP1-KH1/6-mer DNA complex. (A) α CP1-KH1.W.C54S is shown in cartoon representation with the electron density (blue mesh representation) surrounding the bound oligonucleotide (stick representation) showing the well resolved positioning of the oligonucleotide. (B) The surface of the α CP1-KH1.W.C54S structure is shown in grey illustrating the binding cleft that accommodates the oligonucleotide (stick representation). (C) Summary of the contacts between α CP1-KH1.W.C54S and bound DNA of sequence 5'-ACCCCA-3'. Van der Waals contacts are coloured orange, and hydrogen bond interactions are coloured blue. The residues making important contacts with the oligonucleotide sugar-phosphate backbone are listed on the left, and the residues making contacts with the pyrimidine rings, and thus underlying base specificity, are listed on the right. (D) Cartoon view of each of the four cytosines of the C-tetrad (stick representations) within the binding cleft of α CP1-KH1.W.C54S (cartoon representation and transparent surface). Key interacting residues are coloured in orange and hydrogen bonds underlying specificity are shown in black.

mRNA nucleotides 3275–3325). This showed not only an interaction, but the formation of increasingly large multimeric structures as higher proportions of α CP1 were added. This suggests that either α CP1 is capable of binding the target RNA at multiple sites at high concentrations (i.e. not only at the C-rich site) or that α CP1 can initiate the formation of multi-protein/RNA complexes. Since α CP1-KH1 alone was also able to shift the probe, but did not cause increasingly large structures to be formed, it suggests that the triple KH motif structure of full-length α CP1 facilitates the formation of multi-protein/RNA complexes. This is most likely to represent a function of the dimerizing ability of KH domains 1 and 2 that have been identified as forming intra-molecular homodimer, but could play a role in inter-molecular heterodimerization (30).

Through further EMSA studies focusing on the specific 11-nt C-rich site of AR mRNA and its DNA equivalent sequence, we confirmed that α CP1 could interact with both oligonucleotides. α CP1-KH1 alone could also shift these probes, though did not bind with as high affinity as the full-length protein. Thus, the interaction of full-length α CP1 with RNA and DNA is supported by its multiple KH domain interactions. Furthermore, it appeared from

these experiments that DNA was more readily shifted than RNA, implying a higher affinity interaction with DNA. This would be in contrast to a previous report of higher affinity binding by α CP proteins to RNA over DNA (32). In their report, however, binding to poly-C-RNA was being compared with binding to a ssDNA corresponding to the sequence from the V8 domain from human ribosomal RNA, i.e. a long GC-rich sequence that may not have availed the equivalent density of C-rich target sites. This study, in contrast, utilized directly equivalent RNA and DNA. The moderately higher binding affinity to DNA over RNA *in vitro* likely reflects the greater structural accessibility of the DNA (due to the absence of the sugar 2'-OH that assists RNA secondary structure formation) for forming productive binding interactions. Within the cell, however, the actual target of α CPs will be dictated by the localization of the protein and the accessibility of single-stranded C-rich regions of the oligonucleotides, rather than be a function of the difference in affinity.

In order to more quantitatively define the roles of the three individual α CP1 KH domains in interacting with RNA and DNA, SPR studies were carried out with each of the individual KH domains. The binding affinities of the individual KH domains were in the micro-molar

range, in contrast to the nano-molar affinity determined for full-length α CP-KH1 to target AR mRNA (35). The KH domains thus act synergistically to achieve tighter binding by several orders of magnitude over their individual binding. A comparison of KH domain binding interactions revealed that α CP1-KH1 and KH3 contribute the most to the binding affinity of α CP1 to the C-rich oligonucleotide, with a slow dissociation phase observed for KH1 indicating the most stable interaction. Interestingly, whereas interactions between KH2 and RNA could not be detected, KH2 binding to DNA was observed. Furthermore, for all three α CP1 KH domains, higher affinities were again observed for DNA over RNA. This was not a reflection of different protein preparations, as experiments were conducted simultaneously (with RNA and DNA on individual flow cells of the same chip). Overall, it is evident that KH1 and KH3 are likely to make the primary interactions with target oligonucleotide, with KH1 conferring the greatest stability to the interaction. Although RNA binding by KH2 was not observed, in the context of the full-length α CP1 where it is covalently tethered close to the RNA via KH1 and KH3, it would have a greatly enhanced chance of interaction.

The structure of α CP1-KH1 bound to an 11-mer with two immediately adjacent C-rich sites provides further insight into the topological arrangement between poly-C-binding proteins and their target RNA or DNA. The α CP1-KH1 domains were arranged without the formation of contacts between α CP1 and KH1 domains bound at adjacent DNA binding sites. The way in which the α CP2-KH domains can bind in close proximity without forming contacts has previously been shown crystallographically at C-rich sites separated by 2 nt (30). In contrast, the structure of KH3 of hnRNPk bound at immediately adjacent sites places the DNA tightly sandwiched between the two KH domains (26). Their close proximity potentially allows interactions between the GXXG region of one KH domain with the variable loop region of the next. The current structure shows, however, that it is possible for KH domains to be bound arranged head-to-tail at immediately adjacent oligonucleotide sites without forming contacts with one another. This may reflect the way in which covalently attached KH domains within full-length α CP1 interact with target oligonucleotide.

Furthermore, the current structure demonstrates that α CP1-KH1 binding to two immediately adjacent C-rich sites can occur simultaneously with dimer formation. Throughout the crystal lattice of the α CP1-KH1/11-mer DNA complex was a network of protein/DNA complexes interconnected via homo-dimer KH domain interactions. Both homo- and heterodimer formation have previously been observed for KH domains (26,27,30,31) and, especially within the context of the higher affinity multivalent full-length α CP1, this type of network could explain the observation of increasing molecular weight species observed with increasing α CP1 added to target oligonucleotide.

In the current study, we were particularly interested to better understand the basis for α CP1-KH1 specificity for

C-rich oligonucleotides. In previous studies of related KH domains, it has been established that the oligonucleotide binding cleft accommodates four nucleotide binding positions and that it is the central two positions that confer specificity to cytosines (26–31). The structural studies showed, however, A, T or C bound in the first position and C or T, but not a purine, bound in the fourth position. In our own α CP1-KH1/11-mer structure, we observed CCCT bound in the oligonucleotide binding site in the same way as reported for α CP2-KH1 bound to a telomeric sequence containing CCCT (31). A subsequent study by the same group however, showed that this telomeric sequence could also be bound by α CP2-KH1 in a different register, with ACCC placed in the oligonucleotide binding site (30). This raises the question of whether there is, in fact, a binding preference by α CP-KH1 domains for specific bases in the first and fourth positions.

We therefore used SPR to measure the affinity of a series of oligonucleotides in which cytosines within a target tetrad were systematically replaced by thymine or adenine. These experiments confirmed that the presence of cytosine at the central positions (positions 2 and 3) is essential for binding by α CP1-KH1. The experiments also showed that cytosine was preferentially bound at the fourth position, though thymine was tolerated and interestingly, that a cytosine tetrad bound with the highest affinity interactions of all. Thus for the first time, it is demonstrated that, in fact, α CP1-KH1 domain preferentially binds cytosine at all four positions of its oligonucleotide binding cleft.

The structural basis for this preferential binding is revealed in the high resolution structure of a α CP1-KH1 mutant (α CP1-KH1.W.C54S) in complex with a cytosine tetrad DNA (5'-ACCCCA-3'). At position 1, the same protein-oligonucleotide interactions as observed for α CP2-KH1 bound to a pair of tandem 'CCCT' motifs were observed (30). In the current structure, however, we observed an interaction between Cyt-1 and the Asp82 side chain not observed for α CP2-KH1 since the α CP2-KH1 construct stopped short of residue 82. We also observed an interaction between Cyt-1 O2 with the backbone NH of Lys31 that was present, but not reported for α CP2-KH1. Both of these interactions would favour the binding of cytosine at position 1. A further feature of this structure is the positioning of the nucleotide that precedes the bound tetrad. A nucleotide in this position has not been previously reported to form significant interactions with a KH domain, but owing to the few extra residues at the C-terminal end of our α CP1-KH1.W.C54S construct we were able to observe hydrogen bonds to the adenine base in this position. Together, these interactions may contribute to the higher affinity of the C-tetrad observed and reveal the molecular basis for the preferential binding of cytosine in the first nucleotide binding position.

With respect to the fourth position, we did not observe any base-specific interactions to Thy-4 in the α CP1-KH1/11-mer structure. A hydrogen bond from the Arg40 guanidinium group to the Thy-4 O2 was observed, but this could equally well support binding to cytosine in this position. The only reported base-specific interaction

to a base in the fourth position has been a hydrogen bond between the Glu51 side chain and the cytosine-N4 group (31). Indeed, this was also observed in the current α CP1–KH1.W.C54S/6-mer structure and clearly supports specific binding to cytosine. A further feature of cytosine in this position, however, is the intra-molecular hydrogen bond that is formed between the Cyt-4 N4 group and the phosphate linking Cyt-3 and Cyt-4. This interaction, that thymine does not form, may contribute to the overall preferential binding of cytosine though the stabilization of the bound oligonucleotide conformation.

Thus, together these data demonstrate that α CP1–KH1 preferentially binds a C-rich tetrad and reveal the molecular basis for this specificity. In the context of the full-length protein binding to target oligonucleotide, KH1, by virtue of its highest binding affinity, may form the initial interaction. We speculate that this would enable binding by KH2 through its restricted proximity to an adjacent C-rich site on the RNA. KH3, also able to bind independently to RNA, may bind at a nearby C-rich site. Due to the capacity of α CP1 KH1 to form homodimers or heterodimers with KH2, interactions with other RNA bound α CPs may underlie the formation of higher order multi-protein/RNA complexes that direct the outcome for the RNA.

ACCESSION NUMBERS

1ZTG, 3VKE.

ACKNOWLEDGEMENTS

We would like to thank the staff at the Micro Crystallography Beamline (MX2) of the Australian Synchrotron for their help during the diffraction data collection, and acknowledge the use of the Australian Synchrotron facilities.

FUNDING

Australian Research Council Project Grant funding (awarded to M.C.J.W. and J.A.W.); National Health and Medical Research Senior Research Fellowship (awarded to M.C.J.W.). Funding for the open access charge: Australian Research Council.

Conflict of interest statement. None declared.

REFERENCES

- Ostareck-Lederer, A., Ostareck, D. and Hentze, M. (1998) Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2. *Trends Biochem. Sci.*, **23**, 409–411.
- Makeyev, A. and Liebhaber, S. (2002) The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA*, **8**, 265–278.
- Ji, X., Kong, J., Carstens, R.P. and Liebhaber, S.A. (2007) The 3' untranslated region complex involved in stabilization of human alpha-globin mRNA assembles in the nucleus and serves an independent role as a splice enhancer. *Mol. Cell Biol.*, **27**, 3290–3302.
- Ji, X., Kong, J. and Liebhaber, S.A. (2011) An RNA-protein complex links enhanced nuclear 3' processing with cytoplasmic mRNA stabilization. *EMBO J.*, **30**, 2622–2633.
- Wang, X., Kiledjian, M., Weiss, I. and Liebhaber, S. (1995) Detection and characterization of a 3' untranslated region ribonucleoprotein complex associated with human alpha-globin mRNA stability. *Mol. Cell Biol.*, **15**, 1769–1777.
- Chkheidze, A., Lyakhov, D., Makeyev, A., Morales, J., Kong, J. and Liebhaber, S. (1999) Assembly of the alpha-globin mRNA stability complex reflects binary interaction between the pyrimidine-rich 3' untranslated region determinant and poly(C) binding protein alphaCP. *Mol. Cell Biol.*, **19**, 4572–4581.
- Paulding, W. and Czyzyk-Krzeska, M. (1999) Regulation of tyrosine hydroxylase mRNA stability by protein-binding, pyrimidine-rich sequence in the 3' untranslated region. *J. Biol. Chem.*, **274**, 2532–2538.
- Czyzyk-Krzeska, M. and Bendixen, A. (1999) Identification of the poly(C) binding protein in the complex associated with the 3' untranslated region of erythropoietin messenger RNA. *Blood*, **93**, 2111–2120.
- Yu, J. and Russell, J. (2001) Structural and functional analysis of an mrnp complex that mediates the high stability of human beta-globin mRNA. *Mol. Cell Biol.*, **21**, 5879–5888.
- Stefanovic, B., Hellerbrand, C., Holcik, M., Briendl, M., Liebhaber, S. and Brenner, D. (1997) Posttranscriptional regulation of collagen alpha1(I) mRNA in hepatic stellate cells. *Mol. Cell Biol.*, **17**, 5201–5209.
- Ostareck, D., Ostareck-Lederer, A., Shatsky, I. and Hentze, M. (2001) Lipoxygenase mRNA silencing in erythroid differentiation: The 3'UTR regulatory complex controls 60S ribosomal subunit joining. *Cell*, **104**, 281–290.
- Ostareck, D., Ostareck-Lederer, A., Wilm, M., Theile, B.J., Mann, M. and Hentze, M. (1997) mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. *Cell*, **89**, 597–606.
- Collier, B., Goobar-Larsson, L., Sokolowski, M. and Schwartz, S. (1998) Translational inhibition in vitro of human papillomavirus type 16 L2 mRNA mediated through interaction with heterogeneous ribonucleoprotein K and poly(rC)-binding proteins 1 and 2. *J. Biol. Chem.*, **273**, 22648–22656.
- Xiao, X., Tang, Y., Mackins, J., Sun, X., Jayaram, H., Hansen, D. and Antony, A.C. (2001) Isolation and characterization of a folate receptor mRNA-binding trans-factor from human placenta. Evidence favoring identity with heterogeneous nuclear ribonucleoprotein E1. *J. Biol. Chem.*, **276**, 41510–41517.
- Pailard, L., Maniey, D., Lachaume, P., Legagneux, V. and Osborne, H. (2000) Identification of a C-rich element as a novel cytoplasmic polyadenylation element in Xenopus embryos. *Mech. Dev.*, **93**, 117–125.
- Blyn, L., Swiderek, K., Richards, O., Stahl, D., Semler, B.L. and Ehrenfeld, E. (1997) Poly(rC) binding protein 2 binds to stem-loop IV of the poliovirus RNA 5' noncoding region: Identification by automated liquid chromatography-tandem mass spectrometry. *Proc. Natl Acad. Sci. USA*, **93**, 11115–11120.
- Blyn, L., Towner, J., Semler, B.L. and Ehrenfeld, E. (1997) Requirement of poly(rC) binding protein 2 for translation of poliovirus RNA. *J. Virol.*, **71**, 6243–6246.
- Kim, S., Pandey, K., Choi, H., Kim, S., Law, P., Wei, L. and Loh, H. (2005) Poly(C) binding protein family is a transcription factor in mu-opioid receptor gene expression. *Mol. Pharmacol.*, **68**, 729–736.
- Michelotti, E., Michelotti, G., Aronsohn, A. and Levens, D. (1996) Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol. Cell Biol.*, **16**, 2350–2360.
- Du, Q., Melnikova, I. and Gardner, P. (1998) Differential effects of heterogeneous nuclear ribonucleoprotein K on Sp1- and Sp3-mediated transcriptional activation of a neuronal nicotinic acetylcholine receptor promoter. *J. Biol. Chem.*, **273**, 19877–19883.
- Lacroix, L., Liénard, H., Labourier, E., Djavaheri-Mergny, M., Lacoste, J., Leffers, H., Tazi, J., Hélène, C. and Mergny, J. (2000) Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres in vitro. *Nucleic Acids Res.*, **28**, 1564–1575.

22. Bandiera, A., Tell, G., Marsich, E., Scaloni, A., Pocsfalvi, G., Akintunde Akindahunsi, A., Cesaratto, L. and Manzini, G. (2003) Cytosine-block telomeric type DNA-binding activity of hnRNP proteins from human cell lines. *Arch. Biochem. Biophys.*, **409**, 305–314.
23. Siomi, H., Matunis, M., Michael, W.M. and Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.*, **21**, 1193–1198.
24. Grishin, N. (2001) KH domain: one motif, two folds. *Nucleic Acids Res.*, **29**, 638–643.
25. Chkheidze, A. and Liebhaber, S. (2003) A novel set of nuclear localization signals determine distributions of the alphaCP RNA-binding proteins. *Mol. Cell Biol.*, **23**, 8405–8415.
26. Backe, P., Messias, A., Ravelli, R., Sattler, M. and Cusack, S. (2005) X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure*, **13**, 1055–1067.
27. Fenn, S., Du, Z., Lee, J., Tjhen, R., Stroud, R. and James, T. (2007) Crystal structure of the third KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.6 Å resolution. *Nucleic Acids Res.*, **35**, 2651–2660.
28. Sidiqi, M., Wilce, J., Vivian, J., Porter, C., Barker, A., Leedman, P.J. and Wilce, M.C.J. (2005) Structure and RNA binding of the third KH domain of poly(C)-binding protein 1. *Nucleic Acids Res.*, **33**, 1213–1221.
29. Du, Z., Fenn, S., Tjhen, R. and James, T. (2008) Structure of a construct of a Human Poly(C)-binding Protein containing the first and second KH domains reveals Insights into its regulatory mechanisms. *J. Biol. Chem.*, **283**, 28757–28766.
30. Du, Z., Lee, J., Fenn, S., Tjhen, R., Stroud, R. and James, T. (2007) X-ray crystallographic and NMR studies of protein-protein and protein-nucleic acid interactions involving the KH domains from human poly(C)-binding protein-2. *RNA*, **13**, 1043–1051.
31. Du, Z., Lee, J., Tjhen, R., Li, S., Pan, H., Stroud, R. and James, T. (2005) 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. *J. Biol. Chem.*, **280**, 38823–38830.
32. Deigaard, K. and Leffers, H. (1996) Characterisation of the nucleic-acid-binding activity of KH domains. Different properties of different domains. *Eur. J. Biochem.*, **241**, 425–431.
33. Thisted, T., Lyakhov, D. and Liebhaber, S. (2001) Optimized RNA targets of two closely related triple KH domain proteins, heterogeneous nuclear ribonucleoprotein K and α CP-2KL, suggest Distinct modes of RNA recognition. *J. Biol. Chem.*, **276**, 17484–17496.
34. Weiss, I. and Liebhaber, S. (1995) Erythroid cell-specific mRNA stability elements in the alpha 2-globin 3' nontranslated region. *Mol. Cell Biol.*, **15**, 2457–2465.
35. Yeap, B., Voon, D., Vivian, J., McCulloch, R., Thomson, A.M., Giles, K.M., Czyzyk-Krzeska, M., Furneaux, H., Wilce, M.C., Wilce, J. *et al.* (2002) Novel binding of HuR and poly(C)-binding protein to a conserved UC-rich motif within the 3'-untranslated region of the androgen receptor messenger RNA. *J. Biol. Chem.*, **277**, 27183–27192.
36. Yoga, Y.M., Traore, D.A., Wilce, J.A. and Wilce, M.C.J. (2011) Mutation and crystallisation of the first KH domain of human Polycytosine-binding protein 1 (PCBP1) in complex with DNA. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **67**, 1257–1261.
37. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
38. Collaborative Computational Project. (1997) (1997) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
39. Storoni, L., McCoy, A. and Read, R. (2004) Likelihood-enhanced fast rotation functions. *Acta. Crystallogr. D Biol. Crystallogr.*, **60**, 432–438.
40. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr. D*, **60**, 2126–2132.
41. Murshudov, G., Vagin, A. and Dodson, E. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
42. Adams, P., Afonine, P., Bunkóczi, G., Chen, V., Davis, I., Echols, N., Headd, J., Hung, L.W., Kapral, G., Grosse-Kunstleve, R. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
43. Langer, G., Cohen, S., Lamzin, V. and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.*, **3**, 1171–1179.
44. Sidiqi, M., Wilce, J.A., Porter, C.J., Barker, A., Leedman, P.J. and Wilce, M.C. (2005) Formation of an alphaCP1-KH3 complex with UC-rich RNA. *Eur. Biophys. J.*, **34**, 423–429.
45. Chen, V., Arendall, W.B. 3rd, Headd, J., Keedy, D., Immormino, R., Kapral, G., Murray, L., Richardson, J. and Richardson, D. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
46. Baker, N., Sept, D., Joseph, S., Holst, M. and McCammon, J. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.