

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

M. Sidiqi¹, J. A. Wilce¹, J. P. Vivian^{1,2}, C. J. Porter¹, A. Barker^{1,2}, P. J. Leedman^{3,4} and M. C. J. Wilce^{1,2,4,*}

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 through 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 α CP1 to 2.1 Å resolution. α CP1-KH3 assumes a classical type I KH domain fold with a triple-stranded β -sheet held against a three-helix cluster in a $\beta\alpha\beta\beta\alpha$ 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 K_d of 4.37 μ M, which is indicative of intermediate binding. A model of α CP1-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

α CP1 is a member of the poly(C)-binding protein family of proteins that include α CP1, α CP2, α CP3, α CP4 (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 α CP family members are well conserved, with the highest conservation observed between corresponding KH domains. The main differences between α CP 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.

α CP1 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 α CP1 (or other members of the α CP family) and poly(A)-binding protein at the poly(A) tail, resulting in both inhibition of deadenylation and protection of a specific endoribonuclease site, adjacent to the α CP-binding site, from nucleases (5,6). Interactions between the α CP protein of the α -complex and the AU-rich element binding degradation factor AUF1 (hnRNP D) have also been detected (7).

*To whom correspondence should be addressed at School of Pharmacology and Medicine and School of Biomedical and Chemical Sciences, University of Western Australia, 35 Stirling Highway, Crawley, Perth, Western Australia, 6009, Australia. Tel: +61 8 9346 2981; Fax: +61 8 9346 3469; Email: mwilce@receptor.pharm.uwa.edu.au

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

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oupjournals.org

Binding of α CP 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. α CP2 binds with high affinity to a C-rich region within the 3'-UTR of collagen α 1(I) mRNA resulting in its increased stability (11). α CP1 and α CP2 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).

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 of 15-lipoxygenase 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 α CP proteins (15), though the L2 sequence is not so C-rich. In contrast, translational enhancement has been reported due to α CP binding to two sites of the 5'-UTR of picornavirus RNA. α CP 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 α CP-mediated effects on translation have also been reported for other viral systems (18,19). Thus, α CP binding to RNA can result in both silencing and enhancement of translation through a diverse array of pathways.

The specificity of α CP 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 $\beta\alpha\alpha\beta$ motif plus a C-terminal $\beta\alpha$ extension, whereas type II possesses the core and an N-terminal $\alpha\beta$ extension (28). The α CP KH domains, like the hnRNP K, are type I and thus comprise a three-stranded anti-parallel β -sheet packed against three α -helices ($\beta\alpha\alpha\beta\beta\alpha$). 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 α CP 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 α CP1 and α CP2 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 α CP-binding motif within the α -globin 3'-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 α CP-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 α CP 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 α CP1 solved to 2.1 Å resolution using X-ray crystallography. This represents the first structure of an α CP1 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 α CP1-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 α CP1 (KH3) was expressed as a fusion protein with glutathione-S-transferase (GST). The DNA coding sequence comprising amino acids 279–356 of α CP1 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 at an optical density of 0.8 at 595 nm. The cells were harvested after 3 h further growth by centrifugation and resuspension 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. They were then lysed by French Pressing (SLM Instruments, Inc.), supplemented with 0.5 mM phenylmethylsulfonyl fluoride. α CP1-KH3 was purified by affinity chromatography using glutathione agarose beads equilibrated with PBS buffer and the GST removed using 2 U PreScission protease (Amersham) 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 α CP1-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 α CP1-KH3

Crystals of α CP1-KH3 were grown using vapour diffusion in 2 μ l 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 $\sim 0.3 \times 0.2 \times 0.02$ mm with the outline

Table 1. Data collection and refinement statistics

Data collection	
Symmetry	$P2_12_12$
Unit cell (Å)	$a = 33.4$ $b = 71.0$ $c = 29.1$
Measured reflections	14 721
Unique reflections	4310
Completeness (%)	98.1 (87.2)
R_{merge} (%)	5.7 (25.8)
Wilson B (Å ²)	28.2
Refinement	
Resolution range (Å)	35.0–2.1
R_{cryst} (%)	21.4
R_{free} (%)	25.4
r.m.s. deviation from ideal values	
Bond length (Å)	0.012
Bond angle (°)	1.3
Average temperature factor (Å ²)	26.1
Number of water molecules	55

Values in parentheses are for the last resolution shell (2.16–2.1 Å).

$R_{\text{merge}} = \sum |I - \langle I \rangle| / \langle I \rangle$ where I is the observed diffraction intensity and $\langle I \rangle$ is the average diffraction intensity from several measurements of one reflection.

$R_{\text{cryst}} = \sum |F_o| - |F_c| / \sum |F_o|$ where $|F_o|$ and $|F_c|$ are the observed and calculated structure factors, respectively.

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.

Structure solution and refinement

The structure of α CP1-KH3 was solved by molecular replacement using the coordinates of the Nova-2 KH3 RNA-binding domain (accession no. 1EC6) as the search model as implemented in AMORE (34). With one molecule of α CP1-KH3 per asymmetric unit, the estimated solvent content of the crystals is 39%. Matthew's Coefficient was calculated as 2.0, which is within the normal range of proteins (35). Success with molecular replacement was achieved using space group $P2_12_12$, which was also consistent with observed systematic absences. Molecular replacement with other primitive orthorhombic space groups was not successful. Cycles of manual model building and refinement were carried out using REFMAC (34). A total of 10% of the reflections were used for R_{free} calculations. The final model, containing 74 amino acid residues of the α CP1-KH3 construct and 55 water molecules, has a crystallographic R_{cryst} of 21.4% ($R_{\text{free}} = 25.4\%$) with 96% of all amino acids within the most favourable region of a Ramachandran plot. All residues were visible in the electron density map except the N-terminal glycine and the C-terminal eight residues (SEKGMGCS) present in the construct, as confirmed by mass spectrometry (measured

MW 8525; expected MW 8552.72). The final model has been deposited with the Worldwide Protein Data Bank (accession no. 1WVN).

RNA-binding studies

Surface plasmon resonance (using a BIAcore 2000 instrument) was employed to characterize the α CP1-KH3 interaction with RNA. A research grade chip coated with streptavidin was purchased from BIAcore. The target 5'-biotinylated mRNA (mRNA: 5'-CUCUCCUUUCUUUUUCUUCUCCCCUCC-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 μ l/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 μ g/ml tRNA and 62.5 μ g/ml BSA. An average of 30 RU of RNA was immobilized on flow cell 2. α CP1-KH3 was injected over flow cells 1 and 2 with concentrations of 10, 5, 2.5, 1.25 and 0.625 μ M for 2 min using a flow rate of 50 μ l/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 μ l/min with 2 M NaCl. Data were analysed with the BIAevaluation software to obtain a binding constant using a steady-state model.

Modelling of α CP1-KH3 bound to poly(C)-oligonucleotide

The α CP1-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 α CP1-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 10⁶ 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 (see Supplementary Material).

RESULTS

Structural overview

The α CP1-KH3 adopts a classic KH type I domain fold (28), with a triple-stranded β -sheet held against a three-helix cluster in a $\beta\alpha\beta\beta\alpha$ configuration (Figure 1A). The β -sheet is anti-parallel 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

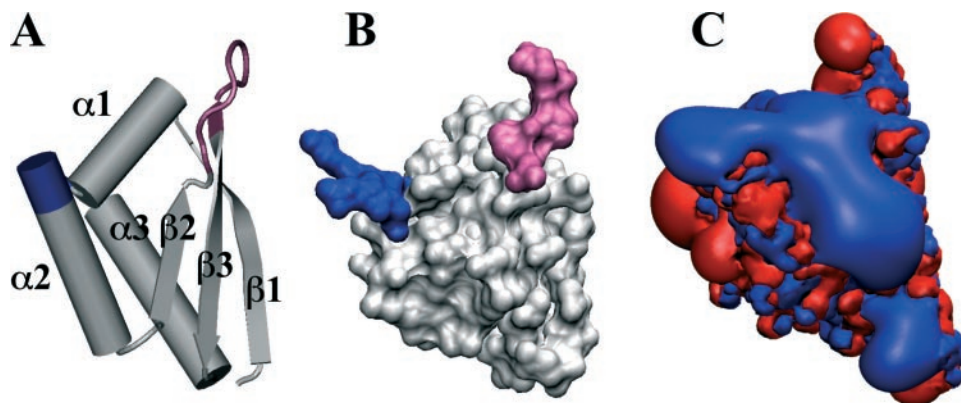


Figure 1. The crystal structure of α CP1-KH3 (residues 279–356) solved to 2.1 Å resolution depicted in (A) cartoon form and (B) as a molecular surface in the same orientation. The structure is shown from the beginning of β -strand 1 to the end of α -helix 3, since the regions outside these bounds were 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 α CP1-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.

binding cleft. The bundle of three amphipathic helices provides the complementary hydrophobic surface within this compact motif. The N-terminal four residues in the model (PLGS) are not shown. These residues, which are not part of the α CP1 sequence but present due to cloning procedures, adopt a random coil structure.

β -strand 1 commences with the first native amino acid residue, Gln5. This strand extends the length of the molecule and projects residues Leu10 and Ile12 into the hydrophobic core, before breaking into a turn at Pro13. α -helix 1 is held in position through its hydrophobic face (including Leu16, Ile17, Ile20 and Ile21) before its structure is interrupted by the invariant GXXG sequence (Figure 1B, blue) that is essential to the KH domain oligonucleotide-binding site. In the case of α CP1-KH3, where RQ fills the XX positions, these side chains are projected outward and provide a hydrophobic edge of the oligonucleotide-binding cleft. Numerous hydrophobic side chains also emanate from α -helix 2 to form contacts with the inner face of the β -sheet, and to provide a hydrophobic environment for oligonucleotide binding (Ile 28, Ile31 and the aliphatic chain of Arg32). Gly36 facilitates a break from helical secondary structure and the remaining two strands of the β -sheet follow. They provide hydrophobic core residues Ile39, Ile41, Arg51, Val53 and Ile55.

β -strands 2 and 3 are separated by the ‘variable loop’ (Figure 1B, pink) which bulges slightly away from the β -sheet and forms the opposing edge of the narrow oligonucleotide-binding cleft. This is the region of the greatest sequence variability between KH domains. The C-terminal helix extends the length of the main body of the molecule with residues Ile62, Leu68, Ile69, Arg72 and Leu73 projected into the hydrophobic core or towards adjacent α -helix 2. α -helix 3 is not visible over the last six residues, due to high mobility.

The oligonucleotide-binding cleft

The oligonucleotide-binding site has long been supposed to involve the GXXG motif. This has been confirmed through

the recent structural analysis of four KH domains in the presence of oligonucleotide. These include Nova2-KH3 in the presence of a 20 base loop of RNA (25), hnRNP-KH3 solved with a 10 base stretch of ssDNA (26) and KH3/4 domains of FBP solved in the presence of a 29 base ssDNA (27). In each of these cases, the main oligonucleotide contacts have been made with the narrow hydrophobic cleft that runs between α -helix 2 and β -sheet 2 and across the GXXG motif. It is thought that the narrowness of the cleft confers the specificity of these KH domains for pyrimidines. Likewise, α CP1-KH3 possesses a narrow hydrophobic cleft that would be expected to accommodate pyrimidine-rich RNA or ssDNA.

The edges of the cleft are polar and charged with basic side chains (Arg23, Arg 32, Lys40 and Arg51) providing attractive electrostatic forces for both the docking of the oligonucleotide as well as making specific contacts with the oligonucleotide (see further discussion below). The electrostatic potential emanating from α CP1-KH3 was calculated using the Adaptive Poisson–Boltzmann Solution (APBS) software package (<http://agave.wustl.edu/apbs/>) (39–43) and is shown in Figure 1C. The contours represent a numerical solution to the Poisson–Boltzmann equation (44,45) and simulate the sum total of the electrostatic potential of the molecule in salty aqueous media. The outstanding feature of the calculation is the positive potential arising precisely from the oligonucleotide-binding cleft (blue contour). This positive potential would provide an attractive force for the approach of the oligonucleotide since its potential is dominated by the electronegative phosphate backbone.

Comparison with other KH domain structures

α CP1-KH3 shows high structural similarity to other type I KH domains. The seven most similar KH structures, including hnRNP K (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

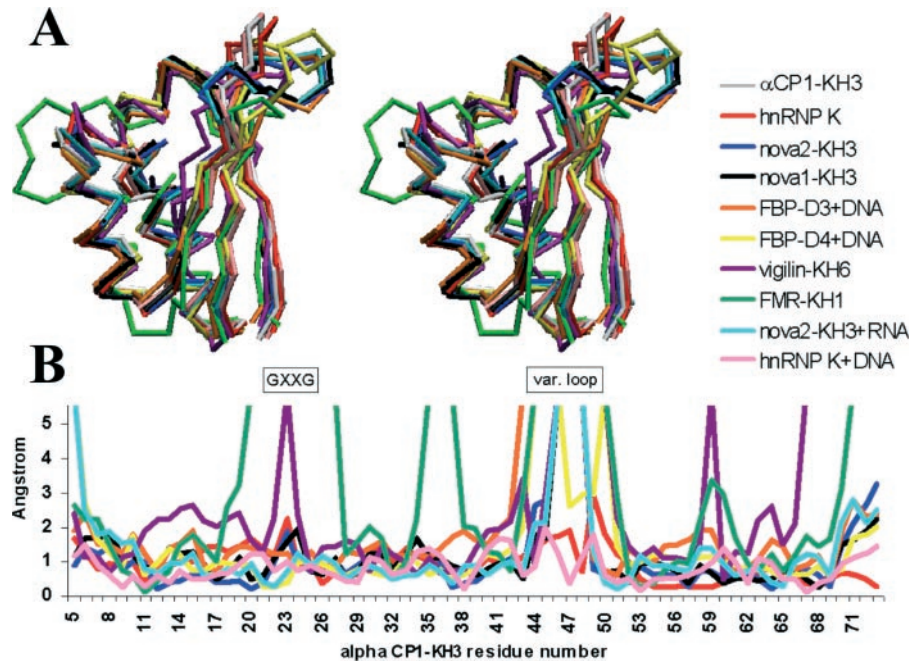


Figure 2. Comparison of KH domain structures. (A) Backbone trace of α CP1-KH3 (grey) shown in stereo superimposed with those of other KH domain structures as listed. These include KH domain structures both in the absence and presence of bound oligonucleotide. (B) The α -carbon deviation for each KH domain residue from the corresponding aligned residue of α CP1-KH3 is plotted versus the α CP1-KH3 residue number. Amino acids >3.5 Å, or with no corresponding aligned residue, are indicated with an off-scale score (>5).

file is depicted). Their backbone traces are highly convergent with pairwise root-mean-square deviation (RMSD) scores compared with α CP1-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 α CP1-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 α CP1-KH3 residues plotted against the α CP1-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 α CP1-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.

α CP1-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 α CP1-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 α CP1 target using α CP1 antibody supershift identification of the components of LNCaP cellular extract binding to the AR mRNA construct (12). Since the α CP1 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 μ M range (4.37 μ M), 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

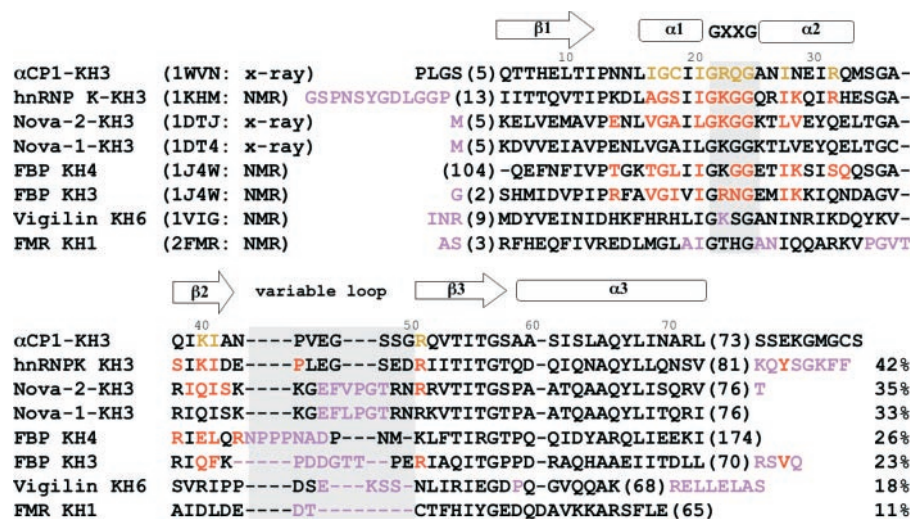


Figure 3. Structure-based sequence alignment of seven KH domains of high structural similarity to α CP1-KH3. Each KH domain was structurally aligned using LSQMAN (36) against α CP1-KH3. Amino acid residues with α -carbon positions within 3.5 Å of a corresponding α CP1-KH3 residue are shown in black. Highlighted in purple are the amino acid residues that do not align well with residues of α CP1-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 α CP1-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 α CP1-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.

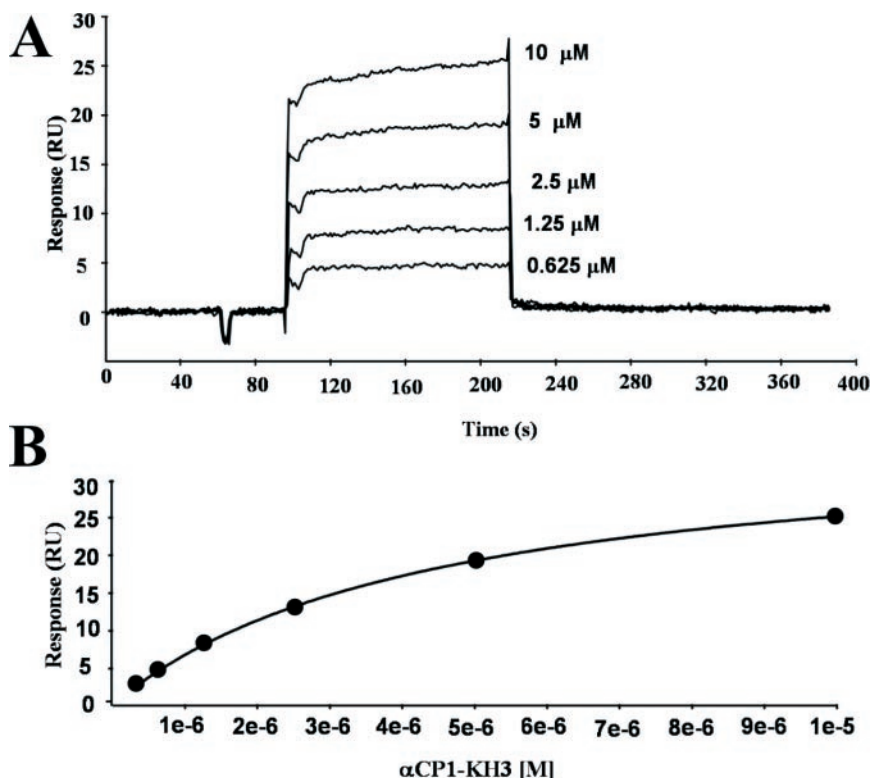


Figure 4. Interaction of the α CP1-KH3 with the 30 nt of the 3'-UTR of AR mRNA measured by surface plasmon resonance. (A) 30 RU RNA was immobilized on a streptavidin-coated chip. Binding interactions were measured for a series of dilutions of the α CP1-KH3 domain from 10 to 0.625 μ M for 2 min using flow rate of 50 μ l/min. (B) Steady-state analysis of the interaction yielded a K_d value of 4.37 μ M.

the absence of non-specific binding competitors used in this study, the tighter binding of full-length protein suggests that multiple KH domains of α CP1 are likely to be participating synergistically in binding the target RNA (12).

Model of α CP1-KH3 bound to poly(C)-oligonucleotide

The high degree of similarity of α CP1-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 α CP1-KH3 structure may also represent a close approximation of its oligonucleotide bound form. Poly(C)-RNA was therefore positioned in the binding cleft of α CP1-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 α CP1-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 proximity 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 isoleucines 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 α CP1-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 α CP1-KH3. Phosphates of nt 2 and 4 can hydrogen bond to sugar hydroxyls of nt 2 and 3, respectively. Phosphates of nt 1

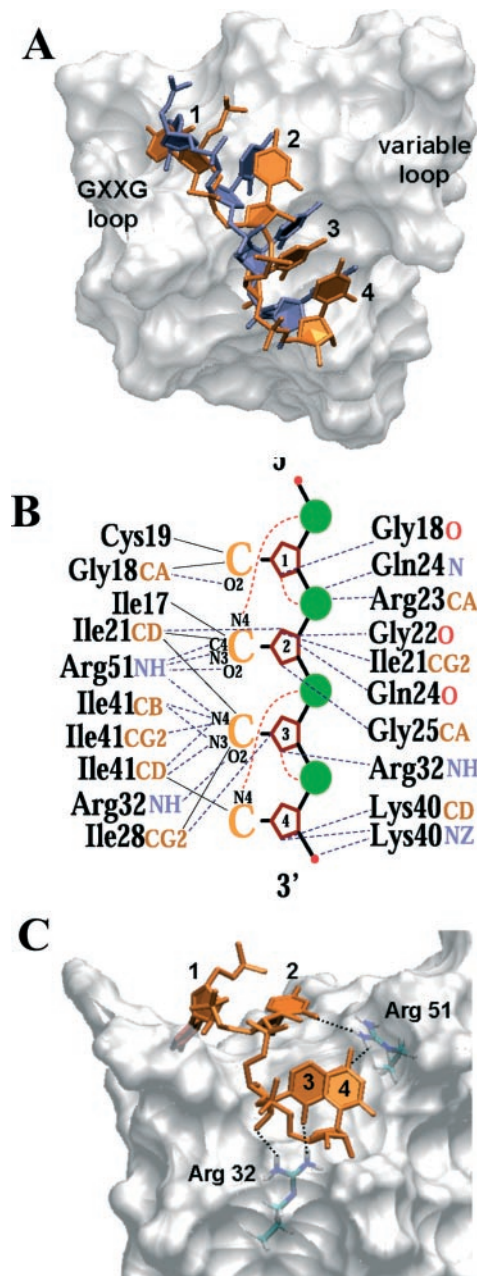


Figure 5. (A) Molecular surface of α CP1-KH3 showing modelled position of poly(C)-RNA (orange) based on the Nova2-KH3-RNA structure (accession no. 1EC6). The poly(C)-tetrad is viewed from above the GXXG and variable loops, highlighting their position either side of the hydrophobic binding cleft. (B) Summary of potential interactions occurring between the modelled α CP1-KH3 and poly(C)-RNA. A poly(C)-RNA-tetrad is represented schematically. Potential hydrogen bond interactions are indicated by dotted lines. Those from specific residue atoms to the RNA backbone are listed on the right, and those to the cytosine bases are listed on the left. The red dotted lines represent intramolecular hydrogen bonds that may stabilize the RNA in its binding mode to the KH domain. Hydrophobic or van der Waals contacts to the cytosine bases are indicated by solid lines. (C) The positions of Arg 32 and Arg 51 side chains are highlighted beneath the molecular surface of α CP1-KH3. Potential hydrogen bonds to the poly(C)-RNA are shown as dotted lines.

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

α CP1-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 hnRNP 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 α CP1-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 α CP1-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 between (27). In addition, the consensus binding sequence for the α CP-2KL isoform involves three C-rich stretches (of 3–5 bases) separated by 2–6 A/U stretches (32). Thus, α CP binding may well involve participation by all three KH domains.

This study represents the beginning of a structural and biophysical examination of all three KH domains of α CP1. The basis for RNA-binding affinity and specificity of the three KH domains will allow us to predict the occurrence of α CP1 interactions with mRNA and better understand the multi-KH domain binding complex. A comparison with the other α CP 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 α CP 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.J.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

- Ostareck-Lederer, A., Ostareck, D.H. and Hentze, M.W. (1998) Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2. *Trends Biochem. Sci.*, **23**, 409–411.
- Makeyev, A.V. and Liebhaber, S.A. (2002) The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA*, **8**, 265–278.
- Wang, X., Kiledjian, M., Weiss, I.M. and Liebhaber, S.A. (1995) Detection and characterization of a 3'-untranslated region ribonucleoprotein complex associated with human α -globin mRNA stability. *Mol. Cell. Biol.*, **15**, 1769–1777.
- Chkheidze, A.N., Lyakhov, D.L., Makeyev, A.V., Morales, J., Kong, J. and Liebhaber, S.A. (1999) Assembly of the α -globin mRNA stability complex reflects binary interaction between the pyrimidine-rich 3'-untranslated region determinant and poly(C) binding protein α CP. *Mol. Cell. Biol.*, **19**, 4572–4581.
- Wang, Z., Day, N., Trifillius, P. and Kiledjian, M. (1999) An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA *in vitro*. *Mol. Cell. Biol.*, **19**, 4552–4560.
- Wang, Z. and Kiledjian, M. (2000) The poly(A)-binding protein and an mRNA stability protein jointly regulate an endoribonuclease activity. *Mol. Cell. Biol.*, **20**, 6334–6341.
- Kiledjian, M., Wang, X. and Liebhaber, S.A. (1995) Identification of two KH domain proteins in the α -globin stability complex. *EMBO J.*, **14**, 4357–4364.
- Paulding, W.R. and Czyzyk-Krzeska, M.F. (1999) Regulation of tyrosine hydroxylase mRNA stability by the protein-binding, pyrimidine-rich sequence in the 3'-untranslated region. *J. Biol. Chem.*, **274**, 2532–2538.
- Czyzyk-Krzeska, M.F. and Bendixen, A.C. (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.E. (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.A. and Brenner, D.A. (1997) Posttranscriptional regulation of collagen alpha1(I) mRNA in hepatic stellate cells. *Mol. Cell. Biol.*, **17**, 5201–5209.
- Yeap, B.B., Voon, D.C., Vivian, J.P., McCulloch, R.K., Thomson, A.M., Giles, K.M., Czyzyk-Krzeska, M.F., Furneaux, H., Wilce, M.C., Wilce, J.A. and Leedman, P.J. (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.
- Ostareck, D.H., Ostareck-Lederer, A., Wilm, M., Theile, B.J., Mann, M. and Hentze, M.W. (1997) mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. *Cell*, **89**, 597–606.
- Ostareck, D.H., Ostareck-Lederer, A., Shatsky, I.N. and Hentze, M.W. (2001) Lipoxygenase mRNA silencing in erythroid differentiation. The 3'-UTR regulatory complex controls 60S ribosomal subunit joining. *Cell*, **104**, 281–290.
- 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.
- Blyn, L.B., Swiderek, K.M., Richards, O., Stahl, D.C., Semier, B.L. and Ehrenfeld, E. (1996) 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.B., Towner, J.S., Semier, B.L. and Ehrenfeld, E. (1997) Requirement of poly(rC) binding protein 2 for translation of poliovirus RNA. *J. Virol.*, **71**, 6243–6246.

18. Graff, J., Cha, J., Blyn, L.B. and Ehrenfeld, E. (1998) Interaction of poly(rC) binding protein 2 with the 5' noncoding region of hepatitis A virus RNA and its effects on translation. *J. Virol.*, **72**, 9668–9675.
19. Spangberg, K. and Schwartz, S. (1999) Poly(C)-binding protein interacts with the hepatitis C virus 5' untranslated region. *J. Gen. Virol.*, **80**, 1371–1376.
20. Siomi, H., Matunis, M.J., Michael, M.W. and Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionary conserved motif. *Nucleic Acids Res.*, **21**, 1193–1198.
21. Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M.A., Nilges, M., Gibson, T.J. and Pastore, A. (1996) Three-dimensional structure and stability of the KH domain: Molecular insights into the fragile X syndrome. *Cell*, **85**, 237–245.
22. Musco, G., Kharrat, A., Stier, G., Fraternali, F., Gibson, T.J., Nilges, M. and Pastore, A. (1997) The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome [Erratum (1997) *Nature Struct. Biol.*, **4**, 840.]. *Nature Struct. Biol.*, **4**, 712–716.
23. Lewis, H.A., Chen, H., Edo, C., Buckanovich, R.J., Yang, Y.Y., Musunuru, K., Zhong, R., Darnell, R.B. and Burley, S.K. (1999) Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure Fold. Des.*, **15**, 191–203.
24. Baber, J.L., Libutti, D., Levens, D. and Tjandra, N. (1999) High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. *J. Mol. Biol.*, **289**, 949–262.
25. Lewis, H.A., Musunuru, K., Jensen, K.B., Edo, C., Chen, H., Darnell, R.B. and Burley, S.K. (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell*, **100**, 323–332.
26. Braddock, D.T., Baber, J.L., Levens, D. and Clore, G.M. (2002) Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J.*, **21**, 3476–3485.
27. Braddock, D.T., Louis, J.M., Baber, J.L., Levens, D. and Clore, G.M. (2002) Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature*, **415**, 1051–1056.
28. Grishin, N.V. (2001) KH domain: one motif, two folds. *Nucleic Acids Res.*, **29**, 638–643.
29. Siomi, H., Choi, M., Siomi, M.C., Nussbaum, R.L. and Dreyfuss, G. (1994) Essential role for KH domains in RNA binding: Impaired RNA binding by a mutation in the KH domain in FMR1 that causes fragile X syndrome. *Cell*, **77**, 33–39.
30. Dejgaard, 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.
31. Weiss, I.M. and Liebhaber, S.A. (1995) Erythroid cell-specific mRNA stability elements in the alpha 2-globin 3' non-translated region. *Mol. Cell. Biol.*, **15**, 2456–2465.
32. Thisted, T., Lyakhov, D.L. and Liebhaber, S.A. (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.
33. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
34. Collaborative Computational Project Number 4. (1994), The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
35. Matthews, B.W. (1977) X ray Structure of Proteins. In Neurath, H. and Hill, R.L. (eds), vol 3, Academic Press, NY, pp. 468–477.
36. Kleywegt, G.J. (1996) Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 842–857.
37. Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K. and Schulten, K. (1999) NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.*, **151**, 283–312.
38. Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.
39. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
40. Holst, M. and Saied, F. (1993) Multigrid solution of the Poisson–Boltzmann equation. *J. Comput. Chem.*, **14**, 105–113.
41. Holst, M. and Saied, F. (1995) Numerical solution of the nonlinear Poisson–Boltzmann equation: developing more robust and efficient methods. *J. Comput. Chem.*, **16**, 337–364.
42. Holst, M. (2001) Adaptive numerical treatment of elliptic systems on manifolds. *Adv. Comput. Math.*, **15**, 139–191.
43. Bank, R. and Holst, M. (2003) A new paradigm for parallel adaptive meshing algorithms. *SIAM Rev.*, **45**, 291–323.
44. Davis, M.E. and McCammon, J.A. (1990) Electrostatics in biomolecular structure and dynamics. *Chem. Rev.*, **94**, 7684–7692.
45. Honig, B. and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.
46. Sidiqi, M., Wilce, J.A., Porter, C.J., Barker, A., Leedman, P.J. and Wilce, M.C.J. (2005) Formation of CP1-KH3 complexed with UC-rich RNA. *Eur. Biophys. J.*, in press.