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 1.6Å resolution


Presented by Yuefan Li
Outline

• Introduction
  ➢ poly(C)-binding proteins
  ➢ KH domain

• Experimental procedures
• Results
• Discussion
• Summary
Poly(C)-binding proteins (PCBPs)

- PCBPs comprised of five members in mammalian cells, namely hnRNPK and PCBP 1-4
- PCBPs has a function which involved in transcriptional and translational silencing and enhancement.
- PCBPs also play a role in mRNA stabilization and splicing.
Poly(C)-binding proteins (PCBPs)

A. mRNA Stabilization
   1) Nonviral (via 3'UTR)
      Ex: human α-globin
      [Diagram: αCP-1, αCP-2 binding to C-rich elements]
   2) Viral (via 5'UTR)
      Ex: polio virus
      [Diagram: αCP-1, αCP-2 binding to 3CD]

B. Translational silencing
   Ex: 15-lipoxygenase
   [Diagram: hnRNP K, αCP-1 binding to multiple DICE elements]

C. Translational enhancement
   Ex: polio virus
   [Diagram: αCP-2 binding to Domain IV]

D. Transcriptional activation
   Ex: SV40 and c-myc
   [Diagram: hnRNP K binding to cell cycle regulatory unit]

E. Transcriptional inhibition
   Ex: Thymidine kinase
   [Diagram: hnRNP K, hnRNP A1 inhibition]

F. Induction of programmed cell death
   [Diagram: hnRNP K, αCP-4 (MCG10) binding to p53, leading to death genes]
KH domain

- KH domains are presented in a variety of nucleic-binding proteins.
- KH domains have a conserved sequence of around 70 amino acids.
KH domain

• All PCBPs contain three different KH domains and recognize poly(C)-sequences with high affinity and specificity.
Materials and Methods

• Sample preparation
  N-terminal His-tagged PCBP2 KH3 was over expressed in the BL21(DE3) strain of E. coli. Purified protein with Ni-NTA resin and removed His-tag. Concentrated to a final concentration of 4 mg/ml.
  7-nt DNA sequence (5’-AACCCTA-3’) corresponding to one repeat of the C-rich strand of human telomeric DNA.

• Cryatallization
  protein : DNA ratio  1:1.2
  hanging-drop vapor diffusion against 2M ammonium sulfate, 80 mM lithium sulfate, 100 mM CAPS, pH 8.18 and 5% glycerol at 22 °C
  crystals grew to useful size within about 5 days with diffraction to 1.55 Å.
Crystallography refinement statistics

<table>
<thead>
<tr>
<th>Crystal data</th>
<th>KH3–DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>R32</td>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
<td>a = 81.07</td>
</tr>
<tr>
<td></td>
<td>b = 81.07</td>
</tr>
<tr>
<td></td>
<td>c = 87.82</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$z^a$</td>
<td></td>
</tr>
<tr>
<td>X-ray data collection statistics</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å) (SeMet SAD Peak)</td>
<td>0.979462</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>60.0 – 1.55</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>272.583</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>15 516</td>
</tr>
<tr>
<td>Completeness (last shell) (%)</td>
<td>98.5 (84.66)</td>
</tr>
<tr>
<td>$R_{merge}$ (%) ($^b$ (last shell)</td>
<td>3.3 (38.2)</td>
</tr>
<tr>
<td>$I/\sigma$ (last shell)</td>
<td>22.8 (7.3)</td>
</tr>
<tr>
<td>Phasing and refinement statistics</td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>40.0 – 1.6</td>
</tr>
<tr>
<td>Reflections in working set</td>
<td>13 842</td>
</tr>
<tr>
<td>Reflections in test set (5%)</td>
<td>723</td>
</tr>
<tr>
<td>$R_{cryst}$ (%) ($^c$</td>
<td>20.5</td>
</tr>
<tr>
<td>$R_{free}$ (%) ($^c$</td>
<td>24.4</td>
</tr>
<tr>
<td>RMSD bonds (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>RMSD angles (°)</td>
<td>1.334</td>
</tr>
<tr>
<td>Mean B-factors (Å²)</td>
<td>28.9</td>
</tr>
</tbody>
</table>

$^a_z$ is the number of equivalent structures per asymmetric unit.
$^b R_{merge} = \sum |I_{hkl} - \langle I_{hkl} \rangle|/\sum I_{hkl}$, where $I_{hkl}$ is the measured intensity of hkl reflection and $\langle I_{hkl} \rangle$ is the mean of all measured intensity of hkl reflection.
$^c R_{cryst} = \sum_{hkl} |F_{obs} - |F_{calc}|\sum_{hkl}|F_{obs}|$, where $F_{obs}$ is the observed structure factor amplitude and $F_{calc}$ is the structure factor calculated from model. $R_{free}$ is computed in the same manner as is $R_{cryst}$, with the test set of reflections (10%).
Overall protein-DNA complex structure

- The PCBP2 KH3 adopts a typical type-I KH domain fold of a $\beta_1$-$\alpha_1$-$\alpha_2$-$\beta_2$-$\beta_3$-$\alpha_3$ configuration.
- Antiparallel $\beta$-sheet with a left-handed twist and a spatial order of $\beta_1$-$\beta_3$-$\beta_2$
- Invariable GXXG loop is located between $\alpha_1$ and $\alpha_2$. Variable loop lies between $\beta_2$ and $\beta_3$
Crystal contacts and overall DNA structure

- The core of the protein forming the hydrophobic floor of a narrow groove for DNA binding.
- The first three residues from an adjacent DNA are stacking on the top of the complex DNA.
- A remarkable feature of this structure is the arrangement of Ade2 and Cys3, which is called AC platform.

(B) Surface representation of the protein showing the nucleic-acid-binding groove. Positively charged, negatively charged, uncharged hydrophilic and hydrophobic residues are colored blue, red, yellow and green, respectively. The floor of the nucleic acid-binding groove is mainly defined by hydrophobic residues. The DNA in the complex is colored orange. The first three residues from a symmetry-related DNA (labeled A1, A2 and C3) are shown in dark teal and are stacking on top of the complex DNA. See text for details.
Crystal contacts

Figure 3. (A) Comparison between the structure of Ade2 and Cyt3 of our crystallization DNA (top) and Ade 441 and Ade 442 from the crystal structure of the 50S ribosomal subunit of Haloarcula marismortui (bottom). Hydrogen bonds are depicted as yellow dashed bars, atoms are color coded light blue, red, grey, white and dark blue for nitrogen, oxygen, carbon, hydrogen and phosphorous, respectively. (B) Top: Overlay of the structure of Ade2 and Cyt3 of our crystallization DNA (red) and Ade 441 and Ade 442 from the crystal structure of the 50S ribosomal subunit of H. marismortui (blue). Bottom: Overlay of our DNA (red) and Ade2532 and Cyt 2533 from the crystal structure of the 50S ribosomal subunit of H. marismortui (blue).
Overview of the DNA binding

- The DNA orients in such a way that its 5’ and 3’ ends contact the C and the N-terminal regions of PCBP2 KH3 domain, respectively.

- The backbone is mainly contacted via a set of hydrogen bonds. Three phosphate backbones of the DNA are involved. These are: Cyt4, Cyt3 and Ade 7

- A cluster of 4 hydrophobic isoleucine residues defines the floor of the binding groove, which contacts the riboses and bases of Cyt4 and Cyt5

- The variable loop remains remarkably open upon DNA binding

Figure 4. Stereo view of the PCBP2 KH3–DNA complex structure. The dense network of hydrogen bonds (yellow dashed bars) involved in DNA–protein interaction is shown. Red spheres represent structured water molecules that participate in the hydrogen-bonding network. The DNA and protein are colored orange and deep blue, respectively. Protein residues that participate in hydrogen bond interaction with the DNA are shown in stick representation.
Specific recognition of the crystallization DNA

- Ade1 does not interact with the protein directly. The base of Ade1 is sandwiched between the base of Ade2 form the same DNA on one side and the base of Cyt3 from adjacent DNA on the other side.

- Both Ade2 and Cys3 lie on the top of helix α1 and involved in intra- or intermolecular hydrogen bonds.

- Cyt4 is the only nucleotide in the complex that has no stacking interaction with other bases. It occupies hydrophobic floor of the groove.

- Base-stacking interaction is observed among the last three residues (Cyt5, Thy6 and Ade7) of the crystallization DNA
Specific recognition of the crystallization DNA

Figure 5. Stereo views of detailed hydrogen-bonding interactions between the DNA and PCBP2 KH3.
Discussion

Previous and current crystal structures reveal some common structural features of KH domain-nucleic acid interactions.

- The structures of the KH domains are very similar to each other, with the possible exception of the variable loop.

- A core DNA/RNA recognition motif consisting of four nucleotides is observed in all structures.

- Each of the core recognition motifs recognized by the PCBP KH domains contains a triple-C sequence.
Discussion

Current structure reveals how the DNA is recognized by the KH domain in unprecedented details.

- Six out of the seven nucleotides of the crystallization DNA form hydrogen bonds, either directly or water-mediated with protein. For all four nucleotides in the core recognition motif involved in hydrogen bonds.

- No protein-protein interaction involved in the dimer formation.

- AC platform formation.
Summary

KH3 domain can bind the C-rich strand of human telomeric DNA repeat in vitro.

Hydrogen bonds and water molecules have a significant function in the binding between the KH3 domain and DNA.

The protein crystallize without protein-protein contacts, yielding new insights into the dimerization properties of different KH domains.
In the future....

Further studies are required to reveal whether the molecular interactions in this structure also occur in vivo and how PCBP proteins might be involved in telomere regulation.