Human programmed cell death 5 protein has a helical-core and two dissociated structural regions

Dongsheng Liu, Yingang Feng, Yuan Cheng, and Jinfeng Wang*

National Laboratory of Biomacromolecules, Center for Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences,
15 Datun Road, Beijing 100101, China

Received 1 April 2004
Available online 16 April 2004

Abstract

Programmed cell death 5 (PDCD5) protein is phylogenetically conserved in both the nucleus and cytoplasm. The human PDCD5 protein is expressed in tumor cells during apoptosis independent of the apoptosis-inducing stimuli, and recently it was found that PDCD5 is an important regulator in both apoptotic and non-apoptotic programmed cell death. In this study, human PDCD5 was expressed in Escherichia coli cell and studied using heteronuclear NMR method. The NMR results indicate that PDCD5 protein can be divided into three structural regions, a core region and two dissociated terminal regions. The core region (41–101) represents a rigid sub-domain consisting mainly of a triple-helix bundle. The N-terminal 38 residues (3–40) are ordered, but not a rigid structural region which contains abundant secondary structure, and packs very loosely against the core. The C-terminal 17 residues (102–118) represent a mobile unstructured region, which may be capable of interaction with nucleic acid.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Programmed cell death 5; Heteronuclear NMR method; Triple-helix bundle; Dynamic behavior

A TF-1 cell apoptosis related gene-19 (TFAR19) has been cloned from human premyeloid cell line established from a patient with erythroleukemia [1], and designated by the name of PDCD5 (programmed cell death 5). Human PDCD5 gene encodes a protein expressed in tumor cells, and translocated rapidly from the cytoplasm into the nuclei of cells during apoptosis. The expression of PDCD5 protein in the tumor cell during apoptosis is independent of the apoptosis-inducing stimuli. The expression level of PDCD5 protein is significantly increased in cells undergoing apoptosis. Using antibody electroporation, PDCD5 protein has been proved as an apoptosis accelerating protein and may play an important role in the apoptotic process of cells [2,3]. Recently, Wang et al. [4] showed that PDCD5 protein overexpression enhanced TAJ/TROY-induced parapoptotic cell death, a death pathway distinct from apoptosis, and suggested that PDCD5 protein is an important regulator in both apoptotic and non-apoptotic programmed cell death. Xu et al. [5] demonstrated that the 5'-upstream region of human PDCD5 gene contains a highly active TATA-less promoter that is up-regulated by etoposide. As was indicated [1], human PDCD5 protein is a well-conserved protein, sharing significant homology to the corresponding proteins of species ranging from yeast to mice (Fig. 1). The PDCD5 homolog in archaear is usually shorter than those in eukaryota for about 10 residues, which are usually high negative charged. One of the proteins in this family, named MTH1615 from the archaea, is 15 residues shorter than human PDCD5 protein. The 3D solution structure of the protease resistant domain of MTH1615 protein was determined using NMR spectroscopy, however, the N-terminal 31 residues of this protein were found to be unstructured in solution [6].

In this study, the heteronuclear NMR method was adopted for understanding the solution conformation of human PDCD5 protein. The 3D solution structure of PDCD5 protein is supposed to be divided into three structural regions, a rigid core region and two dissociated terminal regions. The core region (N41–Q101)
consists mainly of a triple-helix bundle. The N-terminal region (D3–R40) is an ordered, but not a rigid, structural region which contains abundant secondary structures, and packs very loosely against the core. The C-terminal extension represents a mobile and unstructured region (Q102–D118) with a dynamically frayed tail (S119–Y125) in the protein that may be capable of interactions with nucleic acid.

Materials and methods

Expression and purification of the human PDCD5. For expression of the human PDCD5 protein, the fusion expression system with a small fusion partner, His6-tagged N-terminal fragment of staphylococcal nuclease R (HR52), was used [7]. In the expression plasmid pET-3D-HR52-PDCD5, there is a thrombin cleavage site (LVPR/GIS) between the fusion partner HR52 and target protein PDCD5 for removing the fusion partner. For obtaining the highly purified human PDCD5 protein, the purification procedure [7] was modified. A 10 ml of buffer A (50 mM Tris–HCl, pH 8.0, 250 mM NaCl) followed by a 100 ml of buffer B (50 mM Tris–HCl, pH 8.0, 250 mM NaCl) and 10 mM imidazole) and a 100 ml of buffer C (50 mM Tris–HCl, pH 8.0, 140 mM NaCl, and 1 mM CaCl2) was charged with Ni2+, was used for purification of HR52-PDCD5. The column, on which the recovered supernatant containing the overexpressed HR52-PDCD5 was loaded, was washed first by a 100 ml of buffer C and thrombin (Sigma T4648) at a flow rate of 0.2 ml/min for 10 h. The column was washed by a 20 ml of buffer C, and the eluate containing PDCD5 and thrombin was collected and condensed to 5 ml immediately on the column at 25°C by cycling a buffer containing of 100 mM of buffer B (50 mM Tris–HCl, pH 8.0, 250 mM NaCl, and 1 mM CaCl2). The digestion was carried out immediately on the column at 25°C by cycling a buffer containing of 100 mM NH4HCO3. The major peak was identified as target protein PDCD5. The collected protein was lyophilized and stored at −20°C. The sequence of the purified protein, the purification procedure [7] was modified. A 10 ml of buffer A (50 mM Tris–HCl, pH 8.0, 250 mM NaCl) followed by a 100 ml of buffer B (50 mM Tris–HCl, pH 8.0, 250 mM NaCl) and 10 mM imidazole) and a 100 ml of buffer C (50 mM Tris–HCl, pH 8.0, 140 mM NaCl, and 1 mM CaCl2) was charged with Ni2+, was used for purification of HR52-PDCD5. The column, on which the recovered supernatant containing the overexpressed HR52-PDCD5 was loaded, was washed first by a 100 ml of buffer C and thrombin (Sigma T4648) at a flow rate of 0.2 ml/min for 10 h. The column was washed by a 20 ml of buffer C, and the eluate containing PDCD5 and thrombin was collected and condensed to 5 ml immediately on the column at 25°C by cycling a buffer containing of 100 mM NH4HCO3. The major peak was identified as target protein PDCD5. The collected protein was lyophilized and stored at −20°C. The sequence of the purified PDCD5 consists of 126 amino acid residues resulting from thrombin cleavage, replacing residue Met1 in gene sequence (Fig. 1) by Gly–Ser.

NMR sample preparation. The uniformly 13N- and 15C-labeled PDCD5, and unlabeled PDCD5 as well were used in preparation of the NMR sample for different experiments. 1.0–2.0 mM PDCD5 protein was dissolved in 90% H2O/10% D2O containing 50 mM phosphate buffer (pH 6.5), 200 mM NaCl, and 0.01 NaN3. The protein solubility was highly dependent on pH and salt concentration. Low pH, low salt concentration, and high temperature were found to enhance the protein precipitation during experiments. Therefore, both pH and ionic strength, and the experimental temperature were considered seriously for obtaining the proper solubility of PDCD5 protein.

NMR spectroscopy. The heteronuclear NMR experiments [8], 1H–13C–15N HNCACB, CBCA(CO)NH, HN(CO)CA, HNCO, 13C–15N HNCACB, CBCA(CO)NH, HN(CO)CA, HNCO, and 13C–15N NOESY-HSQC, TOCSY-HSQC, were performed with isotope-labeled PDCD5 at 298 K on Bruker DMX 600 MHz spectrometer. All NMR data were processed and analyzed using FE-LIX98 (Msi/Accelrys Inc.). The data points in each indirect dimension were usually doubly by linear prediction [9] before zero filling to the appropriate size. A 90–70° shifted square sine bell apodizations were used for all three dimensions prior to Fourier transformation. 1H chemical shifts were referenced to internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). 13C and 15N chemical shifts were referenced indirectly [8].

Heteronuclear 1H–15N nuclear Overhauser enhancement (NOE), 15N-transverse relaxation rate (R1), and 15N-longitudinal relaxation rate (R2) were used as indicators for the 1H–15N NOE measurements, whereas the saturation period was replaced by a delay of equivalent duration in the control experiment. For R1 measurements, a series of experiments with different relaxation delays (17.0, 33.9, 55.9, 67.8, 84.8, 101.8, 118.7, 135.7, 169.6, and 203.5 ms) using the 1H Carr–Purcell–Meiboom–Gill pulse train [10]. For R2 measurements, the relaxation delay times were set to 0.020, 0.040, 0.140, 0.240, 0.360, 0.530, 0.760, 1.150, and 1.500 s. The NOE values were calculated from the ratios of peak intensities in the NOE spectrum (lunsat) to those in the control spectrum (l0). The standard deviations in NOE values (σNOE) were estimated from the equation, σNOE/NOE = ([σsat/lsat]2 + [σunsat/lunsat])1/2, where σsat and σunsat are the standard deviation values of peak intensities with and without saturation, respectively. R1 and R2 relaxation times were obtained by non-linear least square fits of signal decays to an exponential decay function, 1/l = exp(−R1·t) + exp(−R2·t), using the Gnuplot software (ftp://ftp.gnuplot.info/pub/gnuplot/l). The uncertainties of the R1 and R2 values were obtained directly from Gnuplot.
Results

Assignment of backbone resonances was obtained with triple-resonance NMR spectra recorded on PDCD5 sample at pH 6.5 using $^1$H–$^{13}$C–$^{15}$N resonance experiment. The assigned backbone resonances include 97.5%, 92.0%, 97.6%, 92.8%, and 97.5% of $^1$HN, $^1$Ha, $^{13}$Ca, $^{13}$Co, and $^{15}$N atoms, respectively. The assignments of amide resonances ($^1$HN and $^{15}$N) are represented in the 2D $^1$H–$^{15}$N HSQC spectrum of PDCD5 (Fig. 2). Dispersion of the $^1$H–$^{15}$N cross-peaks in Fig. 2 characterizes a tertiary structure of PDCD5 which consists mainly of $\alpha$-helices.

The secondary chemical shifts [11,12] of $^{13}$Ca, $^{13}$C, and $^1$Ha($\Delta$Ca, $\Delta$C, and $\Delta$Ha) were used for determining the secondary structural elements of PDCD5. Six $\alpha$-helices, consisting of residues D3–A19, D26–K33, N41–Q46, Q50–L61, E65–R79, and G91–Q101, were determined by continuous large positive $\Delta$Ca and $\Delta$C and large negative $\Delta$Ha (Fig. 3). The observed medium to strong dNN NOE signals and weak $d_{NN}(i, i + 3)$, $d_{NN}(i, i + 4)$ NOE signals for residues in these segments (data not shown) provided the conclusive evidence of formation of the six $\alpha$-helices in PDCD5 molecule. The amide resonances from residues in the C-terminal region Q102–Y125, generating less NOE signal, show very strong signals in a narrow spectral region (8.0–8.5 ppm) of HSQC spectrum (Fig. 2). This suggests that PDCD5 has a flexible unstructured C-terminal segment.

$R_1$, $R_2$, and $^1$H–$^{15}$N NOE data are shown as a function of amino acid sequence of PDCD5 in Fig. 4. Sequence variations can be observed for these relaxation data. The relaxation data of PDCD5 fall into three classes. In the first class, a broader sequence region (N41–Q101) containing helices $\alpha$3, $\alpha$4, $\alpha$5, and $\alpha$6, the relaxation rate constants $R_1$, and $R_2$, and NOEs are fairly constant within the $\alpha$-helix regions. The averaged values of $R_1$, $R_2$, and NOE are 1.14 ± 0.22, 20.8 ± 3.0 Hz, and 0.66 ± 0.10, respectively. This indicates that the helices $\alpha$3–$\alpha$6, having similar dynamic behavior, exhibit low internal mobility. Presumably, they form a more

Fig. 2. 2D $^1$H–$^{15}$N HSQC spectrum of 2.0 mM uniformly $^{15}$N-enriched human PDCD5 protein in 50 mM phosphate buffer, 200 mM NaCl, pH 6.5, 298 K. Resonance assignments for backbone amide groups are indicated in the figure. The side-chain amide resonances of asparagines and glutamine residues are connected by horizontal lines.
rigid core region in the protein. In the second class, C-terminal region Q102–Y125, the relaxation data are classified into two classes again. For residues Q102–D118, the NOE and $R_2$ values are lower than those in the core, whereas, $R_1$ is larger. This indicates that the backbone of segment Q102–D118 is less ordered and exhibits large amplitude motion on the sub-nanosecond timescale. For residues in the tail region S119–Y125, the $R_1$ and $R_2$ values decrease and NOE values are negative, thus human PDCD5 has a very flexible tail. In the third class, N-terminal region (A2-R40) containing helices $\alpha_1$ and $\alpha_2$, the NOEs for residues in helix $\alpha_1$ have similar values to those in the core, whereas the $^15$N NOEs of helix $\alpha_2$ decrease gradually from 0.50 at the C-termini to 0.34 at the N-termini of the helix. However, the $R_1$ and $R_2$ values for residues in the helix $\alpha_1$ indicate slow sub-nanosecond motion of the helix. The gradually decreasing $R_1$ and increasing $R_2$ values from the C-termini to N-termini of helix $\alpha_2$ reveal that helix $\alpha_2$ exhibits a heterogeneous dynamic behavior. Thus, low internal mobility in ns–ps timescale, and large amplitude motion on sub-nanosecond timescale were observed for the C-terminal and N-terminal extensions of helix $\alpha_2$, respectively. The loop linking helices $\alpha_1$ and $\alpha_2$ (L-$\alpha_1\alpha_2$: K21-G25) is less ordered and flexible.

The above experimental data provide the structural and dynamical features of the PDCD5 proteins. Human PDCD5 has a rigid core region consisting of four $\alpha$-helices, an ordered N-terminal, and an unstructured C-terminal region. Both terminal regions exhibit slow motions in sub-nanosecond timescale except the dynamically frayed tail.

**Discussion**

Human PDCD5 protein shares 32.4% sequence identity with a 110-residue archaeon protein, MTH1615, from *Methanobacterium thermoautotrophicum* [6] at residues 33–102 (Fig. 1). The 3D solution structure of 39-residue deletion MTH1615 (deletion of 1–31 and 103–110 residues) consists of four helices which has a three right-handed twisted fold. Four helices including helices $\alpha_3$–$\alpha_6$ in PDCD5 have a good sequence alignment with four $\alpha$-helices in MTH1615 (Fig. 1). The Kyte–Doolittle hydrophobic plot [13], obtained with a window size of 7 amino acids, reveals a hydrophobic core consisting of helices $\alpha_3$–$\alpha_6$ in PDCD5 (Fig. 4). Therefore, four helices in the core region of PDCD5 are supposed to construct a triple-helix bundle. The dynamic behavior of the helices in the core region is different explicitly from that of N- and C-terminal extensions (Fig. 4). Four helices in the core region, exhibiting low internal mobility, represent an individual structural domain of the protein, which is flanked by the unstructured C-terminal and ordered N-terminal extensions.
The N-terminal extension of PDCD5 includes three structural regions, two estimated α-helices and a short loop. Helix α1 is about 17-residue long, consisting of residues D3-A19. Helix α2 consists of 8 residues (D26–D33). Residues K20–G25 are a flexible loop connecting helix α1 to helix α2 (L<sub>α1α2</sub>). Helix α1 shows very different dynamic behavior from that of four helices in the core, exhibiting slow sub-nanosecond motions. This indicates there is no compact packing of helix α1 with the helices in the core and helix α1 remains as an unpaired helix in PDCD5. The estimated helix α2 connects to helix α3 in the core through an 8-residue loop (L<sub>α2α3</sub>) consisting mainly of hydrophilic amino acid residues, and links to helix α1 through a flexible L<sub>α1α2</sub> which undergo a cis–trans isomerization as it contains an X-prolyl bond of peptide D23–P24. Therefore, the less ordered helix α2 exhibits heterogeneous backbone dynamics. The internal mobility of helix α2 increases from the C-terminal backwards to the N-terminal of the helix. These dynamic properties of helix α2 may cause the present dynamic behavior of the ordered helix α1. The determination of helix α1 as an ordered structural region

Fig. 4. 15N relaxation experimental data and hydrophobicity of human PDCD5 protein as a function of sequence (A) 1H–15N heteronuclear NOE values. (B) 15N R<sub>2</sub> values. (C) 15N R<sub>1</sub> values. (D) Kyte–Doolittle hydrophathy plot with a window size of seven amino acids [13].
uncorrelating to the core is a significant finding for human PDCD5. The N-terminal 31 residues of MTH1615 were identified as an unstructured region, although sequence alignment shows seven highly conserved orthologs from MTH1615 to human PDCD5 [6].

The unstructured C-terminal extension of PDCD5, undergoing large amplitude motion on sub-nanosecond timescale, consists of 19 hydrophilic and 5 hydrophobic residues. A great number of positive charged and polar residues having high propensities for nucleic acid binding can be found in segment Q102–K115, such as Q102, T103, K105, T106, T107, T108, K119, F111, R113, R114, and K115, according to the proposed residue interface propensity for protein-nucleic acid binding [14]. Since the human PDCD5 protein has been shown to localize mainly in the nucleus, and MTH1615 protein can interact nonspecifically with a randomly chosen 20-mer of double stranded DNA, as demonstrated by electrophoretic mobility shift assay (EMSA) [6], thus PDCD5, presumably, can be involved in nucleic acid binding. Many studies were related to protein-nucleic acid binding, and the binding motifs, such as helix–turn–helix, were proposed. Interest in disordered-order transitions of protein involved in binding with nucleic acid or other proteins has been raised recently. The disordered or random-coil structural regions of protein may be stabilized to take an ordered conformation by binding of nucleic acid or other protein [15–17]. A commonly accepted viewpoint that binding between proteins requires certain flexibility of the building motifs that facilitate the induced fit binding process is acceptable for the binding of protein with nucleic acid. Therefore, the unstructured C-terminal extension could be capable of interactions with nucleic acid.

Above analysis indicates that the 3D solution structure of PDCD5 protein can be divided into three structural regions: a rigid core region and two dissociated terminal regions. The core region of the protein consists mainly of a triple-helix bundle. The N-terminal region is an ordered, but not a rigid, structural region which contains abundant secondary structures, and packs very loosely against the core. The unstructured C-terminal is a mobile structural region in the protein and may be capable of interactions with nucleic acid.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (NNSFC 30170201).

References