Solution Structure of the Human Oncogenic Protein Gankyrin Containing Seven Ankyrin Repeats and Analysis of Its Structure—Function Relationship†‡

Chunhua Yuan,*§,|| Junan Li,†,‡,† Anjali Mahajan,,@ Ming Jye Poi,+ In-Ja L. Byeon,‖ and Ming-Daw Tsai*,§,||,‡,+C

Department of Chemistry, Department of Biochemistry, Biophysics Program, Ohio State Biochemistry Program, and Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, and Genomics Research Center, Academia Sinica, Taipei, Taiwan

Received April 30, 2004; Revised Manuscript Received July 8, 2004

ABSTRACT: Human gankyrin (226 residues, 24.4 kDa) is a liver oncoprotein that plays an important role in the development of human hepatocellular carcinomas. In this paper, its solution structure is reported, which is the largest ankyrin protein ever determined by NMR. The highly degenerate primary sequences of the seven ankyrin repeats presented a major challenge, which was overcome by combined use of TROSY experiments, perdeuterated samples, isotope-filtered NMR experiments, and residual dipolar couplings. The final structure was of high quality, with atomic rmsds for the backbone (N, Cα, and C′) and all heavy atoms (residues 4–224) of 0.69 ± 0.09 and 1.04 ± 0.09 Å, respectively. Detailed analyses of NMR data suggested that the conserved TPLH motifs play important structural roles in stabilizing the repeating ankyrin scaffold. Gankyrin is conformationally more stable than the tumor suppressor p16INK4A, possibly due to the structural roles of conserved residues evidenced by slowly exchanging backbone amides as well as hydrogen bonding networks involving labile side chain protons. Structural comparison with p16INK4A identified several residues of gankyrin that are potentially important for CDK4 binding, whereas observation of the thiol proton of C180 indicated a well-structured Rb-binding site in the helical region of the sixth ankyrin repeat. Interestingly, the CDK4-binding site and Rb-binding site located in N- and C-terminal regions, respectively, are separated by comparatively more stable ankyrin repeats and highly condensed positive surface charge. These results and analyses will shed light on the structural basis of the function of human gankyrin.

Human gankyrin is a liver oncoprotein that plays an important role in the development of hepatocellular carcinomas (1). Overexpression of gankyrin has been observed in all studied human and rodent hepatocellular carcinomas, and has been further correlated with increases in the levels of both phosphorylation and degradation of tumor suppressor retinoblastoma protein (Rb) (J–J). In relation to these effects, gankyrin has been shown to interact with three components in the Rb pathway, namely, Rb, the S6 ATPase subunit of the 26S proteasome, and cyclin-dependent kinase 4 (CDK4) (2). From in vitro studies, dual functional roles have been proposed for gankyrin in the expanded INK4-CDK4/6-Rb pathway (5). On one hand, gankyrin binds to both tumor suppressor Rb and the S6 ATPase subunit of the 26S proteasome, which in turn facilitates CDK4/6-mediated Rb phosphorylation and drives Rb into ubiquitin-mediated degradation. On the other hand, gankyrin directly interacts with CDK4 and counteracts the inhibitory function of p16INK4A and p18INK4C. Moreover, it has been shown that the oncogenic activity of gankyrin can be suppressed by binding to melanoma antigen (MAGE)-A4, a tumor specific antigen with potential in antitumor immunotherapy (6). These findings taken together strongly suggest that gankyrin is a promising target for developing novel therapeutic strategies against liver cancers.

† This work was supported by NIH Grant CA69472 (to M.-D.T.). The Bruker DRX-800 NMR spectrometer was funded by Ohio Board of Regents, and the Bruker DMX-600 spectrometer was funded by NIH Grant RR08299. Structural calculation performed in the Ohio Supercomputer Center was supported by the BALE Cluster supercomputer resource grant.
†‡ PDB entry 1TR4. Backbone chemical shift assignments (15N, 1H N, 13Cα, 13C′, and 1H′) accession code BMRB-5898.
+ To whom correspondence should be addressed: Department of Chemistry, The Ohio State University, 100 West 18th Ave., Columbus, OH 43210. E-mail: cyuan@ccic.ohio-state.edu and tsai.7@osu.edu.
§ Department of Chemistry, The Ohio State University.
|| Campus Chemical Instrument Center, The Ohio State University.
| These authors contributed equally to this work.
‖ Department of Biochemistry, The Ohio State University.
* Current address: Department of Surgery, College of Medicine, The Ohio State University, Columbus, OH 43210.
† Biophysics Program, The Ohio State University.
‡ Ohio State Biochemistry Program, The Ohio State University.
| Current address: Building 5, Room B2-41, National Institutes of Health, Bethesda, MD 20892.
C Academia Sinica.
NMR Structure of Human Gankyrin Oncoprotein

biochemistry, Vol. 43, No. 38, 2004 12153

For CDK4 binding (e.g., p16INK4A and its intricate relationship with other ankyrin proteins that have only one functional domain, binds to multiple partners to promote cell cycle control as briefly above. To understand its functional role and its intricate relationship with other ankyrin proteins for CDK4 binding (e.g., p16INK4A and IkBa) from structural perspectives (5, 9), it is important to determine the tertiary structure of gankyrin under physiological conditions.

We here report the tertiary structure of gankyrin determined by multidimensional NMR spectroscopy, which reveals seven ankyrin repeats (numbered AR1–AR7 starting from N-terminus) with some structural variance in the capping units. The results are significant. (i) Hydrogen bonds involving the side chains of Thr and His residues in a TPLH motif were deduced from resonance and NOE assignments, underscores the role of highly conserved residues in the structural integrity of an ankyrin repeat. (ii) Stabilizing features such as amino–aromatic interactions were uncovered in the capping units, which could be significant for ankyrin repeat protein design. (iii) Structural comparison with p16INK4A has identified several putative residues in AR1–AR3 of gankyrin that are potentially important for CDK4 binding. (iv) The Rb-binding motif LXCXE was identified in the first helix of AR6, and C180 was found in a reduced form with its thiol proton observed. (v) Finally, between the CDK4- and Rb-binding sites, a more stable loop conformation as well as more condensed positive surface charge was observed in AR4 and AR5.

Materials and Methods

Sample Preparation. Full-length human gankyrin (residues 1–226) was expressed as a glutathione S-transferase (GST) fusion protein, and the GST tag was removed, leaving GPLGS residues at the N-terminus of gankyrin (5), [U-15N], [U-13C,15N], and [U-13C,15N,70% 2H]-labeled gankyrin were expressed and purified essentially as described previously (5) except that the bacteria harboring the expression vector were grown at 37 °C in M9 minimal medium containing 1.0 g/L [13NH4]ammonium sulfate with or without 1.5 g/L [13C6]glucose and 80% (v/v) 2H2O. The protein concentrations were 0.3–0.6 mM in 5 mM HEPES, 1 μM EDTA, and 1 mM DTT (pH 7.4). A higher concentration results in visible aggregation.

NMR Spectroscopy. The lyophilized gankyrin samples were dissolved in either 90% H2O and 10% 2H2O or 100% 2H2O. The NMR spectra were collected at 300 K on a Bruker DMX-600 spectrometer as well as on a Bruker DRX-800 spectrometer, each of which is equipped with a 5 mm triple-resonance probe with three-axis gradients. The TROSY-based triple-resonance experiments, including TROSY-HNCA, TROSY-HN(CO)CA, TROSY-HNCO, TRSY-HN(CA)CB, and TROSY-HN(CO)CAB, were carried out with the [U-15N,70% 2H]-labeled sample for sequential assignment (10). 3D 15N-edited TOCSY and 3D HCCH-TOCSY were carried on [U-15N]- and [U-13C,15N]-labeled samples, respectively, for side chain chemical shift assignment. 3D 15N-edited NOESY data sets were collected on a [U-15N]-labeled sample with two mixing times of 100 and 200 ms, and on a [U-13C,15N,70% 2H]-labeled gankyrin with a mixing time of 150 ms. 3D 13C-edited NOESY data sets were recorded with a [U-13C,15N]-labeled sample dissolved in 90% H2O and 10% 2H2O, each of which has a mixing time of 100 ms. 2D NOEY spectra were collected in 100% 2H2O at 90% 2H2O and in 100% 2H2O as well as in 100% 2H2O, each of which has a mixing time of 100 ms. 2D NOEY spectra were collected in 100% 2H2O or in 90% H2O and 10% 2H2O. Slow 1H−H exchanging backbone amides were identified by recording 1H−15N HSQC spectra after dissolving the lyophilized protein in 100% 2H2O buffer for a period of time. NMR data were processed with XWINNMR version 3.1 (Bruker, Inc.).

Residual Dipolar Coupling Experiment. Backbone amide residual dipolar coupling (RDC) experiments were performed on the Bruker DRX-800 spectrometer on a [U-15N]-labeled sample in the presence of filamentosus P11 phase (ASLA Ltd., Riga, Latvia) (11). The concentration of phosphate was ca. 3.5 mg/mL, resulting in a 5 Hz splitting of the 2H2O signal that ensures a weak alignment (11, 12). Doublet-separated sensitivity-enhanced HSQC spectra were collected for the
determination of scalar and dipolar one-bond $J$ coupling (13). The data were processed with NMRPipe (14) and analyzed with NMRView (15). Approximate values of the axial and rhombic components of the molecular alignment tensor were determined by the distribution of NH residual dipolar couplings and a grid search (16, 17). The accuracy between experimental RDC data and gankyrin NMR structure was evaluated using the program Module (18), and the dipolar $R$-factor that measures the agreement between observed and calculated dipolar couplings was calculated (19).

Structure Calculation. The solution structures were determined on the basis of distance restraints derived from the identified NOEs and H-bonds, the backbone torsion angle restraints derived from chemical shifts using TALOS (20), and residual dipolar coupling constants (12, 16). NOE-derived distance constraints were classified as 1.8–2.7, 1.8–3.7, 1.8–5.0, and 1.8–6.0 Å according to the relative NOE cross-peak intensities. Upper limits for distances involving methyl protons and non-stereospecifically assigned protons were corrected appropriately as described previously (21). The structures were calculated utilizing a dynamical annealing protocol (22) implemented in CNS (23) and analyzed with PROCHECK (24) and MOLMOL (25). The structure figures were produced using MOLMOL (25), and the NMR spectra were generated with XWINNMR (Bruker, Inc.).

RESULTS AND DISCUSSION

A High Level of Intramolecular Sequence Homology Presents a Challenge for NMR Assignments. To date, structures of only five ankyrin repeat proteins, four INK4 members (26–29) and myotrophin (30), have been determined by NMR. Human gankyrin is not only larger in size (24.4 kDa and seven ankyrin repeats) but also higher in its level of internal sequence homology. With the exclusion of (24.4 kDa and seven ankyrin repeats) but also higher in its

The chemical shift index (CSI) method (32) and TALOS (20) enabled us to derive secondary structures from backbone chemical shift data that are found to be exquisitely sensitive to local conformation (Figure 2). The CSI method, which is based on chemical shift deviations of $^{13}$C$\alpha$, $^{13}$C$\beta$, $^{13}$C, and $^1$H$\alpha$ resonances from respective random coil values, predicted 13 long $\alpha$-helices (7–10 residues), one short $\alpha$-helix (∼4 residues), and three intercalated short $\beta$-strands. Similar results were obtained with TALOS, which predicts backbone dihedral angle values by searching a database for chemical shift ($^{13}$C$\alpha$, $^{13}$C$\beta$, $^{13}$C, $^1$H$\alpha$, and $^{15}$N) and sequence homology. The helices derived from chemical shift data were generally supported by short-range NOE assignments such as $^1$H$(i)$–$^1$H$(i + 3)$ and $^1$H$(i)$–$^1$H$(i + 3)$, although not all of the expected NOE correlations can be assigned, largely due to overlap of cross-peaks. Nevertheless, the following helices were identified with NOE assignments: V10–S16, L19–L28, A42–S49, T53–Q61, P76–S82, D86–G94, P109–S115, H119–E126, A142–A148, L152–Y160, P175–D181, V185–S193, L209–A212, and G217–M223. These in turn form seven helix–turn–helix motifs (the core of seven ankyrin repeats).

Investigation of Histidine and Threonine Residues in the TPLH Tetrapeptide. A salient feature of an ankyrin repeat is the highly conserved TPLH motif. Crystal structures of other ankyrin proteins showed that this motif makes a tight turn to initiate a helix–turn–helix motif, and Thr coupled with His serves as the $\alpha$-helix N-cap (7). However, this motif happens to be absent in the INK4 proteins and myotrophin, the AR proteins whose structures have been determined by
NMR Structure of Human Gankyrin Oncoprotein

Figure 3: Schematic diagram showing the side chain of an N\(^{\alpha2}\)-protonated histidine residue with the dotted arrows indicating NOEs and two-bond \(^2\)J\(\text{HN}\) coupling cross-peaks observed in Figure 4.

NMR. The NMR study presented here provides additional insight into this motif, that the histidine assumes an N\(^{\alpha2}\)-H tautomeric form (Figure 3) and the threonine receives structural protection for its labile hydroxyl proton.

Figure 4a shows the \(^1\)N downfield region of a \(^1\)H-\(^{15}\)N heteronuclear multiple-bond correlation (HMBC) spectrum optimized for observation of two-bond correlation. Five histidine residues displayed a pattern of the N\(^{\alpha2}\)-H form, which is characterized by a set of two-bond coupling peaks of H\(^1\)-N\(^{\alpha\alpha}\), H\(^{\alpha1}\)-N\(^{\alpha2}\), and H\(^{\alpha2}\)-N\(^{\alpha2}\). The corresponding H\(^2\) protons were observed in the proton downfield region (11−12 ppm) with strong NOEs to both neighboring H\(^1\) and H\(^{\alpha2}\) protons (Figure 4b). Further NOE assignments led to their identification all in a TPLH motif or a variation (Table 1). One of those NOEs was assigned to the interaction between His H\(^2\) and the Thr H\(^a\) in the following TPLH motif, which supports the formation of a hydrogen bond between His H\(^2\) and a distant carboxyl oxygen at the His+29 position, the residue just preceding the TPLH peptide in the next \(\beta\)-hairpin loop (e.g., H177 H\(^{\alpha2}\)-K206 O in Figure 5).

Histidine is also engaged in interactions with threonine in a TPLH motif. An example shown in Figure 5 indicates two reciprocal side chain−backbone hydrogen bonds (Thr H\(^{\beta-}\) His N\(^{\alpha\alpha}\) and Thr O\(^{\beta-}\)His H\(^{\beta}\)) and another potential side chain−side chain hydrogen bond (Thr H\(^{\beta-}\)His N\(^{\alpha\alpha}\)) deduced from NMR assignments and generated structures. The last hydrogen bond, which was not addressed to the best of our knowledge in any crystal structures of ankyrin protein, is credited to the ability of NMR to detect this protection for the labile hydroxyl proton and the associated NOEs such as the one between T174 H\(^{\alpha1}\) and H177 H\(^{\alpha1}\). It is important to note that these NMR assignments were challenging due to the nature of oxygen-attached protons. In an attempt to resolve ambiguities, several isotope-filtered NMR experiments have been performed, including 2D \(^{13}\)C- and \(^{15}\)N-filtered (f1) and f2 unfiltered NOESY on a [U-\(^{13}\)C,\(^{15}\)N]-labeled sample and a 2D double half-\(^{15}\)N-filtered NOESY on a [U-\(^{15}\)N]-labeled sample (34, 35). This body of data taken together with other NOE data sets allowed us to make unambiguous NMR assignments (Table 1). The observation of S75 H\(^{\beta}\) in \(^{75}\)SPLH\(^{78}\) suggests that a serine residue can also adopt a similar hydrogen bonding network with its side chain. However, threonine is clearly preferred because its buried bulky methyl group could further contribute to stability via hydrophobic interactions with the residue at the Thr+26 position, which is Ala in the ankyrin repeat consensus sequence.

It has been well documented that an amino acid at the N-cap has a significant impact on the stability of \(\alpha\)-helical structure (36, 37). In nature, a high frequency of Ser at the N-cap coupled with Glu at the Ser+3 position is found which could provide reciprocal side chain−backbone interactions in stabilizing a helical conformation (37). In gankyrin with predominant \(\alpha\)-helical elements, the three aforementioned potential hydrogen bonds between Thr and His residues together with one more involving backbone atoms only (Thr O and His H\(^{\beta}\)) (38, 39) should contribute significantly to secondary structure stability. Moreover, the inter-ankyrin repeat hydrogen bond involving His H\(^2\) and the hydrophobic interactions involving the methyl group of threonine further contribute to the stabilization of the elongated tertiary fold. Therefore, the NMR data presented here could well explain why the TPLH motif, the Thr and His pair in particular, is highly conserved among ankyrin repeats. The data also nicely complement the results from crystal structures of ankyrin proteins, which can only elucidate information about heavy atoms.

**Tertiary Structure of Seven Ankyrin Repeats.** The NMR constraints were largely collected from 2D and 3D NOESY spectra recorded in either 100% \(^2\)H\(_2\)O or 90% H\(_2\)O and 10% \(^2\)H\(_2\)O. Effort was made to resolve ambiguous NOEs by using every piece of information and an iterative strategy of structure-assisted assignment. Figure 6 shows an example how an NOE between T108 H\(^{\beta}\) and H111 H\(^{\beta}\) is assigned, which is comparable to the NOE between T174 H\(^{\beta}\) and H177 H\(^{\beta}\) indicated in Figure 5. Hydrogen bonds and dihedral angle constraints in the identified secondary structure as well as RDC data of backbone amides were also incorporated in structural calculation. The RDC experiment conducted in a slightly anisotropic environment assesses the one-bond internuclear dipolar interactions which can be correlated with the average orientation of the corresponding vectors relative to the magnetic field (12, 40).

In the first stage, a preliminary low-resolution structure was built with \(~1000\) unambiguous NOEs mostly attributed to unique resonances. The structure revealed an elongated globular shape that was used to aid in more NOE assignments; in other words, the candidate for an NOE assignment could be significantly narrowed down to the neighboring repeats. More structure-assisted NOEs were assigned, and the next-generation structure was calculated. When the NOE number reached \(~1400\), an ensemble of 10 calculated structures revealed a backbone (residues 20−200) atomic rmsd value of 2.3 Å. As many as 20 rounds of structure calculations have been performed, during which NOE assignments were added and cross-checked. The final ensemble of 20 structures was selected from a total of 80 calculations based on 3185 constraints, and the structural statistics are summarized in Table 2. The structure has revealed seven ankyrin repeats that linearly stack on each other, with a global architecture reminiscent of p16\(^{NK4A}\) (Figure 7). Interactions between short \(\beta\)-strands were observed, but the evidence was not sufficient to propose a continuous \(\beta\)-sheet in hairpin loops. While the internal five ankyrin repeats show a high level of structural similarity and superimpose with an average pairwise root-mean-square deviation (rmsd) of 0.88 Å for backbone atoms, several points about the capping repeats are worth mentioning. (1) The N-terminus is well-folded and appears to consist of a distorted \(\beta\)-strand (Figure 7b). The well-ordered nature is evidenced from resonance and NOE assignments such as
ankyrin proteins, a slight bending of the repeat stack toward of G226 displays several short-range NOEs.

The first helix in the second helix motif, a four-residue stretch (K30- \( > \)-hairpin loop can be clearly discerned for gankyrin, the short helix in the loop following the first helix motif, a four-residue stretch (K30-A33) appears to adopt a helical conformation. As a result, the \( \beta \)-hairpin loop is one residue longer than that observed in a consensus AR. These residues indeed display an average 2.3 ppm upfield shift of 13C resonance from respective random coil values (Figure 2b). However, the NOE data are insufficient for firmly establishing the helical conformation or let alone for differentiating between an \( \alpha \)-helix and a 3_10-helix. (3) The first \( \alpha \)-helix in AR7 is atypically short, leading to a distorted ankyrin repeat. Like the N-terminus, the C-terminal region is also well-folded. Even the backbone amide proton distortion ankyrin repeat. Like the N-terminus, the C-terminal region is also well-folded. Even the backbone amide proton distortion ankyrin repeat.

Table 1: Chemical Shift Assignments of Hydroxyl Protons and Imidazole Ring Protons in the TPLH Motif or Its Variants

<table>
<thead>
<tr>
<th></th>
<th>Thr H_1 or Ser H_1 (ppm)</th>
<th>His H_{R2} (ppm)</th>
<th>His H_{R1} (ppm)</th>
<th>His H_{R2} (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>42TALH</td>
<td>6.12</td>
<td>6.68</td>
<td>7.81</td>
<td>11.24</td>
</tr>
<tr>
<td>75SPLH</td>
<td>7.05</td>
<td>6.15</td>
<td>7.27</td>
<td>11.94</td>
</tr>
<tr>
<td>106TPLH</td>
<td>?</td>
<td>6.51</td>
<td>7.36</td>
<td>11.20</td>
</tr>
<tr>
<td>145TAMH</td>
<td>6.40</td>
<td>6.77</td>
<td>7.60</td>
<td>11.86</td>
</tr>
<tr>
<td>176TPLH</td>
<td>6.24</td>
<td>6.62</td>
<td>7.58</td>
<td>11.79</td>
</tr>
<tr>
<td>20TPLQ</td>
<td>5.78</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The uncertainty is estimated to be 0.03 ppm for 1H resonances.

short-range NOEs assigned to E2 and the ring current shift experienced by V5 H_{R2} (at \(-0.26\) ppm in the vicinity of Y15 and W46. (2) In the loop following the first helix-\( \beta \)-turn-\( \beta \)-helix motif, a four-residue stretch (K30-A33) appears to adopt a helical conformation. As a result, the \( \beta \)-hairpin loop is one residue longer than that observed in a consensus AR. These residues indeed display an average 2.3 ppm upfield shift of 13C resonance from respective random coil values (Figure 2b). However, the NOE data are insufficient for firmly establishing the helical conformation or let alone for differentiating between an \( \alpha \)-helix and a 3_{10}-helix. (3) The first \( \alpha \)-helix in AR7 is atypically short, leading to a distorted ankyrin repeat. Like the N-terminus, the C-terminal region is also well-folded. Even the backbone amide proton distortion ankyrin repeat. Like the N-terminus, the C-terminal region is also well-folded. Even the backbone amide proton distortion ankyrin repeat. Like the N-terminus, the C-terminal region is also well-folded. Even the backbone amide proton distortion ankyrin repeat. Like the N-terminus, the C-terminal region is also well-folded. Even the backbone amide proton distortion ankyrin repeat. Like the N-terminus, the C-terminal region is also well-folded. Even the backbone amide proton distortion ankyrin repeat.

While it is generally described as “linearly stacking” in ankyrin proteins, a slight bending of the repeat stack toward the \( \beta \)-hairpin loop can be clearly discerned for gankyrin, the largest AR protein structure ever determined by NMR. This observation is impressive considering that as an elongated protein close contacts are made by residues close in primary sequence, and that long-range NOEs can only be observed between adjacent repeats. However, the RDC data in this work have provided an extremely valuates source of orientation constraints to the overall folding that delineates the slight curvature of the elongated shape (12, 16, 40). The 2-3_10 twisting of each ankyrin repeat with respect to the preceding repeat has also been revealed in crystal structures of larger proteins such as human ankyrinR with 12 ankyrin repeats (41). When the internal five ankyrin repeats (residues 37-201) of gankyrin are overlaid with human ankyrinR (residues 533-697), the backbone rmsd is only 1.58 Å and the twists along the long molecular axis agree with each other.

Structure-Function Relationship of Gankyrin. Previous \textit{in vitro} studies have narrowed down the CDK4 binding site of gankyrin to the first four ankyrin repeats (gankyrin residues 1-138), and the binding affinity for CDK4 was found to decrease in the following order: p16\textsubscript{INK4A} > p18\textsubscript{INK4C} > gankyrin (5). Although they share a low degree of homology in sequence, gankyrin and p16\textsubscript{INK4A} share a high degree of homology in structure. However, aside from different numbers of ankyrin repeats, several major differences are worth noting. (1) p16\textsubscript{INK4A} has disordered N- and C-termini (27, 29, 42). (2) The sequence of p16\textsubscript{INK4A} is less similar to the consensus sequence of an ankyrin repeat, and a repeat unit could be varied by one residue in length. (3) The first helix in the second helix-\( \beta \)-turn-\( \beta \)-helix motif of p16\textsubscript{INK4A} is atypically short. (4) p16\textsubscript{INK4A} does not have a short helix in the loop following the first helix-\( \beta \)-turn-\( \beta \)-helix motif.

On the basis of the p16\textsubscript{INK4A}-CDK6 crystal structure (43), the important residues for mediating CDK binding were observed mostly in the first three ankyrin repeats. Those residues using side chains for hydrogen bond formation (R22, E27, N42, D74, D84, R87, and D92 of p16\textsubscript{INK4A}) should contribute significantly to binding specificity as well as binding affinity. It would be instructive to compare gankyrin with p16\textsubscript{INK4A} to identify prime candidates in gankyrin for...
CDK4 binding. Figure 1b shows the structure-based sequence alignment of p16\textsuperscript{INK4A}, p18\textsuperscript{INK4C}, and gankyrin, while the structural superposition of p16\textsuperscript{INK4A} and gankyrin is shown in Figure 8a. It needs to be pointed out that using AR1–AR3 rather than AR2–AR4 of gankyrin to align with AR1–AR3 of p16\textsuperscript{INK4A} for the CDK4 binding (Figure 1b) leads to a more meaningful alignment and shows that E21, D37, D70, and E87 of gankyrin play roles equivalent to those of E27, N42, D74, and D92, respectively, in p16\textsuperscript{INK4A}. Interestingly, E21 and E87 in gankyrin are among the least conserved residues, whereas D37 and D70 are among the highly conserved ones. These putative binding residues will be subjected to further examination by site-directed mutagenesis.

The Rb-binding site of gankyrin is located in the last two ankyrin repeats (5) (Figure 8b). More specifically, a peptide fragment of gankyrin (L\textsuperscript{70}LHLCDEERV\textsuperscript{185}), in which LXCXE is a Rb-binding motif, was found to be responsible for Rb binding (5). Point mutation of L178, C180, or E182 to Ala has been shown to abolish or significantly reduce the Rb binding (5). Two lines of evidence in this NMR work showed that C180 exists in a reduced form and its sulfhydryl group is well protected from solvent exposure (Figure 8b).

Table 2: Structural Statistics of 20 Gankyrin Structures

<table>
<thead>
<tr>
<th>restraint</th>
<th>mean</th>
<th>stdev</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>long-range NOE ((</td>
<td>i-j</td>
<td>\geq 5))</td>
<td>703</td>
</tr>
<tr>
<td>short-range NOE (1 &lt;</td>
<td>i-j</td>
<td>&lt; 5)</td>
<td>756</td>
</tr>
<tr>
<td>sequential NOE (</td>
<td>i-j</td>
<td>= 1)</td>
<td>644</td>
</tr>
<tr>
<td>intraresidue NOE</td>
<td>561</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hydrogen bond constraints</td>
<td>77 x 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dihedral angles(a)</td>
<td>1.72</td>
<td>0.53</td>
<td>4.07</td>
</tr>
<tr>
<td>backbone amide RDC</td>
<td>140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rmsd for distance restraints ((\AA))</td>
<td>0.029 ± 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-factor for RDC (%(b))</td>
<td>4.1 ± 0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rmsd from idealized covalent geometry bonds ((\AA))</td>
<td>0.0034 ± 0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>angles (deg)</td>
<td>0.50 ± 0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>improper (deg)</td>
<td>0.47 ± 0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROCHECK (Ramachandran plot)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>most favored regions (%)</td>
<td>76.9 ± 1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>additionally allowed region (%)</td>
<td>18.5 ± 1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>generously allowed region (%)</td>
<td>3.6 ± 0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>disallowed region (%)</td>
<td>1.0 ± 0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rmsd with respect to mean structure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>backbone (residues 4–224)</td>
<td>0.69 ± 0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>heavy atom (residues 4–224)</td>
<td>1.04 ± 0.09</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(a\) The dihedral angle constraints include 111 \(\phi\) and 111 \(\psi\) torsion angle restraints derived by using TALOS (20) and five \(\gamma\) restraints derived from NOE assignments. \(b\) \(R_{\text{dip}} = ([D_{\text{obs}} - D_{\text{calc}}] / [2(D_{\text{obs}}^2)])^{1/2}\), where \(D_{\text{obs}}\) and \(D_{\text{calc}}\) are the observed and calculated RDC values, respectively (19).

Conserved residues were identified by comparing the structure of human gankyrin with those of Cin89 and Cin55, using the program Dali (21).
indication of a well-structured binding site in a helical region and also suggests that in terms of structure the reduced cysteine may be substituted with a structurally equivalent serine residue. Interestingly, the LXSXE motif has been observed in protein phosphatase 1, a Rb-binding protein, and the mutation from LXSXE to LXCXE only increases the binding affinity slightly (45). It remains to be established whether the cysteine in the Rb-binding motif can be generally replaced with a serine residue for its functional role.

Stability of Human Gankyrin and Ankyrin Protein Design. In contrast to p16 INK4A which displays very low stability and very high conformational flexibility (46, 47), gankyrin is comparatively stable which is evident from the observation of slowly exchanging amide protons. After $^1$H→$^2$H exchange had proceeded at 300 K for 3 h (the lyophilized protein was dissolved in 100% $^2$H$_2$O buffer), approximately 60 slowly exchanging amide protons could be observed (Figure 9). Most of them are localized in $\alpha$-helices (Figure 10), and none was found in the vicinity of Gly-induced turn, including two residues N-terminal and four residues C-terminal to the helix-breaking Gly residue. The two projected binding residues, E21 and E87, are both three residues C-terminal to Gly.

Figure 7: (a) Stereoview showing backbone N, C′, and Cω atoms of 20 gankyrin structures (residues 1–226). The backbone atoms of residues 4–224 were used for superposition. (b) Stereoview of the ribbon diagram of a representative gankyrin structure.

Figure 8: (a) Superposition of the backbone atoms of gankyrin (blue, residues 1–226) and p16$^{INK4A}$ (red, residues 14–134) NMR structures (27). Residues 8–31, 34–49, 51–70, and 73–106 of gankyrin are superimposed with residues 14–37, 38–54, 55–74, and 78–111 of p16$^{INK4A}$. The backbone rmsd is ca. 2.0 Å for the 94 residues in a span of three ankyrin repeats. The residues potentially important for CDK4 binding are highlighted in green (E21, D37, D70, and E87) and the corresponding residues in p16$^{INK4A}$ in magenta. (b) Rb-binding motif $^{179}$LACDE$^{182}$ (highlighted in red) located in the first helix of AR6. C180 Sα is highlighted in blue, and residues E21, D37, D70, and E87 are shown in green. (c) Surface charge distribution. Positive and negative charges are colored blue and red, respectively. (d) Amino–aromatic interactions (K23–F58 and Y199–K221) identified in the capping units. The figures are generated in the same orientation except that panel d is rotated 180° along the long molecular axis.
The backbone amide peaks of L59, L60, L92, L125, L126, L158, and L159 are the only ones that could be detected even after 2 weeks. These Leu residues belong to a homologous pair and are highly conserved in the ankyrin repeats. Together with other conserved residues, particularly the TPLH peptide at the beginning of a helix-turn-helix motif, they could play an important structural role in stabilizing the turn of the L-shaped AR cross section formed by antiparallel \( \alpha \)-helices and the extended \( \beta \)-hairpin loop at an approximately right angle.

Though the two capping ankyrin repeats of gankyrin appear to be more flexible from \( ^1 \)H--\( ^2 \)H exchange data (Figure 10), they are better structured than the two capping repeats of p16INK4A. Several features could have been introduced during the evolution to help to stabilize the two termini. (1) Substitution of hydrophobic residues with polar residues in both capping units could avoid the exposure of the hydrophobic surface to solvent. (2) The change from histidine to glutamine in the C-terminal TPLH motif could avoid a destabilizing effect, solvent exposure of the imidazole ring. A similar phenomenon was observed in the last ankyrin repeat of IκBε and other proteins (49). (3) Several hydrogen bonds, involving side chains between AR1 and AR2 and between AR6 and AR7, were identified. (4) Two amino-aromatic interactions (50) are present between K23 and F58 and between Y199 and K221 in N- and C-termini, respectively (Figure 8d). In favor of this type of interaction, F58 is mostly buried (~4% surface exposure), whereas Y199 has ~42% surface exposure. Furthermore, the exposed amino groups of K23 and K221 could also increase the protein solubility at physiological pH (51).

There has been intense interest in designing repeat proteins with novel functionality (52). The ankyrin repeat, one of the major modular repeat proteins, has been one of these subjects. Recent work has revealed consensus sequences of building blocks (38, 53), which has significantly advanced our knowledge of ankyrin proteins. Apparently, the design of the capping unit is more challenging since the sequence should be varied to take stability, solubility, and function into consideration (38, 39, 51, 53). This could explain why a terminal capping unit is so divergent in the primary sequences of natural proteins. The features uncovered in gankyrin and other ankyrin proteins would be valuable in guiding the design of capping units.

**Comparison with X-ray Crystal Structures.** During the structural refinement and preparation of the manuscript, two crystal structures of human gankyrin (54, 55) and another crystal structure of a gankyrin homologue (Nas6p) (56) were reported. These developments support the significance of gankyrin and provide a good opportunity to compare crystal and solution structures from independent research. As shown in Figure 11, the overall folds of gankyrin structures determined by NMR and X-ray crystallography appear to be very similar with an rmsd value of 1.58 Å for backbone
CDK4 binding and Rb binding functions of gankyrin, and their relationship with gankyrin (15). The packing force. For example, an NOE was observed between chain atoms, which can be partially attributed to the crystal packing force. For example, an NOE was observed between V5 Cα and W46 H1 at 10.96 ppm, whereas the distance between V5 Cα and W46 H1 was measured to be 7.7 Å in the crystal structure (54), which is somewhat larger than the NOE limit after correcting for the Cα–H bond length. Most significant backbone deviations are observed in the C-terminal ankyrin repeat, particularly at G214 and G215, which could be attributed to the paucity of NOE related to these Gly residues (the HN and Hα protons of G215 were not assigned).

SUMMARY AND CONCLUSIONS

We have successfully determined the human gankyrin structure under physiological conditions. While the structural studies by NMR and X-ray crystallography complemented each other nicely, the NMR study was clearly a challenge due to the significant homology of the seven ankyrin repeats. Once this difficulty had been overcome, we were able to perform detailed analyses of the structure–function relationship of gankyrin. The NMR studies on H–D exchange provided further insight into the conformational stability of different ankyrin repeats. Detailed NMR analyses of the conserved TPLH motif also led to insight into its structural roles. Building on our previous studies of the structure–function relationship of the INK4 proteins (28, 46, 57) and their relationship with gankyrin (5), this work has also set a good stage for uncovering the structural basis of the CDK4 binding and Rb binding functions of gankyrin.

SUPPORTING INFORMATION AVAILABLE

2D 1H–15N HSQC spectrum with assignments labeled. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

NMR Structure of Human Gankyrin Oncoprotein

Biochemistry, Vol. 43, No. 38, 2004


