Structural Consequences of Tumor-Derived Mutations in p16^{INK4a} Probed by Limited Proteolysis†

Bin Zhang‡ and Zheng-yu Peng*

Department of Biochemistry, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06032

Received August 22, 2001; Revised Manuscript Received March 13, 2002

ABSTRACT: The cyclin-dependent kinase inhibitor p16^{INK4a} (hereafter p16) functions as a multiple tumor suppressor. Mutations in p16, which are distributed throughout the entire protein, have been identified in a variety of human cancers and cancer-derived cell lines. It is unclear how tumor-derived mutations disrupt the structure and function of p16, especially since many of these mutations are located far away from the cyclin-dependent kinase binding site. In this study, we investigated the effect of two tumor-derived mutations, P81L and V126D, on the structure of p16 by limited proteolysis. The proteolytic products were characterized by gel electrophoresis, HPLC, and mass spectrometry. Our results show that the N-terminal half of p16 is significantly more sensitive to proteolysis in both tumor-derived mutant proteins than in the wild type, suggesting that this region is particularly unstable. Interestingly, the N-terminal half of p16 contains many residues that are important for cyclin-dependent kinase binding. Thus, our results provide a structural mechanism by which tumor-derived mutations inactivate the function of p16 and suggest that stabilization of the N-terminal region could be a useful strategy for future therapeutic development.

Cancer is characterized by uncontrolled cell proliferation. The *INK4a*/*ARF* locus is particularly important for cell cycle regulation because it encodes two unrelated tumor suppressor proteins, using different open reading frames (for recent reviews, see refs *1*−*3*). The protein p16^{INK4a} (hereafter p16)[1] specifically inhibits the G1 phase cyclin-dependent kinases, cdk4 and cdk6, which phosphorylate the retinoblastoma protein, pRb, thereby regulating the commitment to cell cycle progression (*4*−*6*). The other protein p14^{ARF} sequesters mdm2 and stabilizes the tumor suppressor p53 (*7*, *8*). A large number of mutations and deletions in p16 have been identified in a variety of human malignant diseases, including melanoma, pancreatic cancer, esophageal cancer, nonsmall cell lung cancer, and certain types of leukemia (*9*, *10*; for a recent review, see ref *11*). The strongest evidence for p16's involvement in tumorigenesis comes from human genetic studies. Over a dozen point mutations have been found to cosegregate with a genetic predisposition to familial melanoma (*12*, *13*), and biochemical analyses show that these tumor-derived mutant proteins are generally inactive (*14*−*18*).

The three-dimensional structure of p16 has been determined by X-ray crystallography and heteronuclear NMR (*19*, *20*). The protein consists of a linear array of four ankyrin repeats, a 33-residue sequence motif often found in proteins mediating specific protein−protein interactions (*21*, *22*). Each ankyrin repeat folds into two antiparallel α-helices, with a connecting *β*-hairpin loop protruding out at a perpendicular direction (*23*). p16 is a relatively unstable protein; the ΔG of unfolding is only 3.1 kcal/mol (*24*). In addition, the tumor-derived mutations are distributed throughout the entire molecule, without any significant clustering in a particular region (*11*). Taken together, these observations suggest that the majority of the tumor-derived mutations may inhibit the proper folding or destabilize the native structure of p16 rather than locally disturbing the p16−cdk4/6 binding site. Indeed, several spectroscopic studies suggested that the tumor-derived p16 mutant proteins are generally structurally compromised, and some of these proteins tend to aggregate in solution (*19*, *24*−*26*).

The mechanisms by which proteins respond to disruptive mutations can vary extensively. In some cases, the mutation may cause local unfolding of a specific secondary structural element (for example, see ref *27*). In other cases, the mutation may convert the entire protein into a partially folded, molten globule-like species, which often results in the loss of function (*28*−*32*). Early spectroscopic studies indicate that many tumor-derived p16 mutant proteins have a partially folded structure. These proteins are not suitable for high-resolution X-ray or NMR studies (for example, see refs *19* and *25*). Thus, from a mechanistic point of view, how the effect of tumor-derived mutations at a site far away from the binding interface between p16 and cdk4/6 propagates to the functionally important regions of the protein is still unclear.

In this study, we investigated the structural consequences of tumor-derived mutations in p16 by limited proteolysis.

† This work was supported by grants from the Patrick and Catherine Weldon Donaghue Medical Research Foundation and the Robert Leet and Clara Guthrie Patterson Trust.

* Corresponding author. Tel: (860) 679-2885. Fax: (860) 679-3408. E-mail: peng@sun.uchc.edu.

‡ Current address: EngeneOS, Inc., 40 Bear Hill Road, Waltham, MA 02451.

[1] Abbreviations: p16, the tumor suppressor protein p16^{INK4a}; cdk4 and cdk6, cyclin-dependent kinase 4 and cyclin-dependent kinase 6; HMPK, heart muscle protein kinase; CD, circular dichroism; HPLC, high-performance liquid chromatography; NMR, nuclear magnetic resonance; LC-MS, liquid chromatography−mass spectrometry.

10.1021/bi0117100 CCC: $22.00
Published on Web 04/26/2002

The proteolysis method has been used extensively to study partially folded proteins, since it does not require the molecule to have a well-defined tertiary structure (*33*). If a tumor-derived p16 mutant protein contains a locally unfolded region, this region should be cleaved at a much higher rate than regions that are buried and involved in regular secondary structures (*33*, *34*). Our studies were carried out under "single-hit" conditions using relatively nonspecific proteolytic enzymes, so the result will reflect the intrinsic properties of the target molecule. To determine the location of the proteolytic cleavage sites, we used two different techniques, gel electrophoresis with N-terminal radioactive labeling and reverse-phase HPLC with mass spectrometry.

Two tumor-derived mutations were selected for detailed analysis. The mutation P81L was originally identified in sporadic melanoma and melanoma cell lines (*35*, *36*), and the mutation V126D was found in familial melanoma (*12*, *13*). The choice of these two mutations was based on the following considerations. First, there is a plethora of structural and functional data regarding these mutant proteins. The mutant P81L has a high level of secondary structure (*24*, *26*). However, both biochemical and yeast two-hybrid studies show that it does not bind cdk4 or cdk6, nor can it inhibit the kinase reaction (*15*, *16*). The mutant V126D has approximately half of the amount of secondary structure compared to the wild type (*24*, *26*). The functional properties of V126D are similar to that of P81L, except that V126D has been shown to have a temperature-sensitive phenotype (*15*−*17*, *37*). Second, both P81L and V126D do not change the amino acid sequence of $p14^{ARF}$. Therefore, the effects of these mutations are solely due to the loss of p16 function. Finally, in the X-ray crystal structure of the p16−cdk6 complex, neither P81 nor V126 is directly in contact with the cyclin-dependent kinase. P81 is located near the center of p16, at the beginning of the third ankyrin repeat, while V126 is located in the second α-helix of the fourth ankyrin repeat. The side chains of P81 and V126 are both buried in the interior of p16; therefore, substitutions of these residues by larger aliphatic residues or charged residues likely will disrupt the hydrophobic packing and/or structural integrity of the protein.

Our results indicate that the N-terminal half of p16 comprising the first two ankyrin repeats is particularly unstable and sensitive to proteolysis in tumor-derived mutant proteins than in the wild type. Interestingly, the proteolytic sensitive region contains a number of residues that is important for cyclin-dependent kinase binding. Thus, structural perturbation in this region could easily cause the loss of function. This study provides a better understanding of the structure−function relationship of p16 and useful information for the design of future therapeutic agents.

MATERIALS AND METHODS

Protein Expression and Purification. The cDNA clone of p16 was a gift from Dr. David Beach at Cold Spring Harbor Laboratory. The p16 gene was amplified by PCR with an N-terminal HMPK recognition sequence (RRASVA, single letter code) and a C-terminal 6×His tag encoded in the primers and cloned into the T7 polymerase-based expression vector, pAED4 (*38*). The tumor-derived mutations were transferred from the truncated p16 proteins to the current construct by exchange of a *Bss*HII−*Eag*I restriction fragment. The sequences of all clones were confirmed by automated DNA sequencing reactions. The proteins were expressed in *Escherichia coli* BL21 (Novagen) growing in terrific media and induced with 0.4 mM IPTG.

After lysis of the bacteria by sonication in a buffer containing 10 mM Tris (pH 8.0), 200 mM NaH_2PO_4, and 8 M urea, the insoluble material was removed by centrifugation at 27000*g* for 30 min. The protein was first purified by metal chelating chromatography using a 4 mL Ni-NTA−agarose column (Qiagen) according to the manufacturer's protocol. The protein was eluted with a buffer at pH 4.5, and fractions containing p16 were combined and dialyzed against 5% acetic acid. The protein was further purified by reverse-phase HPLC using a Vydac C_{18} preparative column, eluted with a linear gradient of acetonitrile in 0.1% TFA. The purified protein was lyophilized and stored at −80 °C. The identity of all proteins was confirmed by electrospray mass spectrometry.

Circular Dichroism (CD). CD studies were performed on a JASCO J-715 spectropolarimeter equipped with a thermoelectric temperature controller using a 1 mm path-length cuvette for wavelength scans and a 1 cm path-length cuvette for thermal denaturation studies. The samples consisted of 2−10 μM protein in 10 mM Tris (pH 8.5), 1 mM EDTA, and 1 mM DTT. The protein concentrations were determined by the absorbance at 280 nm in 6 M guanidine hydrochloride (*39*).

Limited Proteolysis. (a) Determining the Overall Rate of Proteolysis. To determine the overall rate of proteolytic digestion, a 1 mL, 10 μM solution of wild-type p16 or its mutant proteins in 10 mM Tris (pH 8.5), 1 mM EDTA, and 1 mM DTT or 10 mM Tris (pH 8.5), 1 mM EDTA, 1 mM DTT, and 6 M urea was treated with 3×10^{-4} mg of proteinase K at 0 °C. A 200 μL aliquot of the reaction mixture was taken out at 5, 15, 45, and 120 min and mixed with 3 μL of 0.1 M PMSF (in 100% 2-propanol) and 400 μL of 5% acetic acid. The resulting solution was analyzed using a Waters reverse-phase HPLC system and a Vydac C_{18} analytical column. The elution profiles were monitored at 229 and 280 nm using a four-channel absorbance detector. The peak areas were quantified using Waters Millennium software and normalized by the peak area of an equal amount of undigested protein.

(b) Identification of Proteolytic Cleavage Sites Using N-Terminal Radioactive Labeled Protein. To label p16 at the N-terminus, approximately 0.3 mg of purified protein was dissolved in 300 μL of labeling buffer containing 10 mM Tris (pH 7.4), 150 mM KCl, and 10 mM DTT. To this solution were added 32 nmol of [γ-^{32}P]ATP and 500 units of HMPK (P2645, Sigma), and the reaction mixture was incubated at room temperature for 30−45 min. The labeled p16 and free radioactive labels were separated by passing through a disposable gel filtration column (PD-10, Amersham Pharmacia Biotech). The column was equilibrated and eluted with the storage buffer containing 10 mM Tris (pH 8.5) and 1 mM DTT. The labeled p16 proteins can be stored at 0 °C for up to 1 week without significant degradation as detected by SDS gel electrophoresis.

The labeled p16 proteins were diluted in the storage buffer to approximately the same concentration, based on the amount of radioactivity determined by using a scintillation

counter. The proteases were diluted from a 1 mg/mL stock solution to a final concentration of 10^{-3} or 10^{-4} mg/mL in 100 mM Tris (pH 8.5) and 10 mM $CaCl_2$. Typically, 1 μL of diluted protease, 4 μL of labeled protein, and 5 μL of storage buffer were mixed together to make a 10 μL reaction, which was allowed to proceed for 5−10 min at 0 °C. In the final reaction mixture, the protein concentration was on the order of 0.1 mg/mL, and the protease concentration was between 10^{-4} and 10^{-5} mg/mL. The conditions were carefully adjusted such that, even for the least stable mutant (V126D), at least 85% of the starting material remained uncut at the end of the reaction, and the same amount of each protease was always used to treat both the wild-type and mutant proteins. The cleavage reaction was stopped by addition of 2 μL of 0.1 M PMSF and 2 μL of 0.1 M EDTA with 6 μL of 4× gel loading buffer. The proteolytic products were separated by Tris−tricine gel electrophoresis using a large format apparatus (Hoffer Scientific SE420, 30 cm long). No significant difference was observed between samples that were analyzed immediately and samples that had been frozen overnight. The gels were dried and scanned using a phosphorimager (Storm 860, Molecular Dynamics). Data were processed using the program ImageQuant from Molecular Dynamics.

A set of molecular mass standards was generated by digestion of the same N-terminally labeled p16 proteins with four specific proteases (trypsin, chymotrypsin, endoproteinase Asp-N, and V8 protease) under denaturing conditions. The buffers contained 10 mM Tris (pH 8.5), 1 mM DTT, and 8 M urea for trypsin, chymotrypsin, and V8 protease and 10 mM Tris, 2 mM $ZnCl_2$, and 4 M urea for endoproteinase Asp-N. The reactions were carried out at room temperature for 30 min (3 h for endoproteinase Asp-N) with the final protease concentration at 0.01 mg/mL. These conditions were chosen to maximize the number of fragments that can be seen on the gel. The proteolytic reactions were stopped by addition of PMSF and EDTA. The products were separated on the same gel with the nonspecific proteolytic fragments.

We can routinely identify about two-thirds of all proteolytic fragments predicted by the known protease specificity (typically, we can identify 10−12 of 17 trypsin sites, 7−8 of 11 endoproteinase Asp-N sites, and 3−4 of 5 chymotrypsin sites; V8 protease tends to generate small molecular mass fragments, which were useful for assigning the proteolytic cleavage sites near the N-terminus). The theoretical molecular masses of these fragments were fit to a function of the electrophoretic mobility using

$$M_W = a + b(d - d_0) \quad (1)$$

where a and b were fitting parameters and $d - d_0$ was the migration distance of a proteolytic product relative to the uncut material. The intercept values (a) determined from these fits were in excellent agreement with the molecular mass of the intact protein. The slopes (b) depended on how long the gel was run and, therefore, exhibited some variability between different experiments. The root mean square deviations of these fits were between 152 and 227 Da, suggesting that the accuracy of our molecular mass determination was on the order of ±2 residues.

The patterns of limited proteolysis shown in Figure 4 were generated by first normalizing the radioactive intensity measured for each gel lane to the maximum intensity of the uncut protein (the lane in the middle on each gel). The electrophoretic mobility was converted into the molecular mass of the fragment using eq 1 and then to the residue number in p16 using a mean residue molecular mass of 107 Da, calculated on the basis of the amino acid composition of the protein. Fragments with molecular masses less than 2 kDa had an abnormal electrophoretic mobility; therefore, the number of residues in these fragments cannot be determined accurately.

(c) Identification of Proteolytic Cleavage Sites Using HPLC/Mass Spectrometry. To purify each of the individual proteolytic fragments by HPLC, a 1 mL solution of 1 mg/mL unlabeled protein in 10 mM Tris (pH 8.5), 1 mM EDTA, and 1 mM DTT was treated with 6×10^{-4} mg of proteinase K at 0 °C for 10 min. After the reaction was terminated, the digested protein was loaded onto a Vydac C_{18} analytical column and eluted with a linear gradient of 10−50% acetonitrile over 100 min. Only peaks that appeared early (≤5 min) in the analytical studies and that had a higher intensity in the chromatogram generated from a mutant protein than from the wild type were collected and lyophilized for further analysis.

The molecular mass of each proteolytic fragment was determined by electrospray mass spectrometry using a Bruker Esquire ion trap mass spectrometer located at the University of Massachusetts Mass Spectrometry Center. The samples were dissolved in 50% methanol and 3% acetic acid and injected into the electrospray interface at a rate of 1 μL/min. Data were collected and processed using the software provided by Bruker. The assignment was made on the basis of the assumption that the molecule had been cleaved only once; thus, each fragment must contain either the N- or C-terminus, and the predicted molecular mass must be within ±2 Da from the measured value. The N-terminal sequencing analyses were performed at the Tufts University Core Facility.

NMR. The NMR spectra were acquired using a Varian INOVA 500 MHz spectrometer at 20 °C. The samples contained approximately 200 μM protein in 10 mM Tris (pH 7.5), 1 mM EDTA, and 1 mM DTT. The spectra were processed using the software NMRpipe and NMRdraw.

RESULTS

Several compilations of p16 mutations found in various types of human cancers have been published (*40*, *41*). Figure 1 shows the location of tumor-derived mutations mapped onto the three-dimensional structure of p16. Less than one-third of these mutations are located on or near the cdk4/6 binding site. Thus, for the majority of tumor-derived mutations, the effect must be indirect.

Ideally, one would like to solve the high-resolution structure of tumor-derived p16 mutant proteins. However, this is often impossible because the protein is partially unfolded and exhibits high levels of conformational dynamics. Our initial NMR studies show that, in the $^{15}N-^{1}H$ HSQC spectrum of P81L, the number of well-resolved peaks is much less than the number of residues in the protein (Figure S1, Supporting Information). Most of these peaks have "random coil" chemical shift values, suggesting that they arise from unfolded regions of the molecule. The mutant

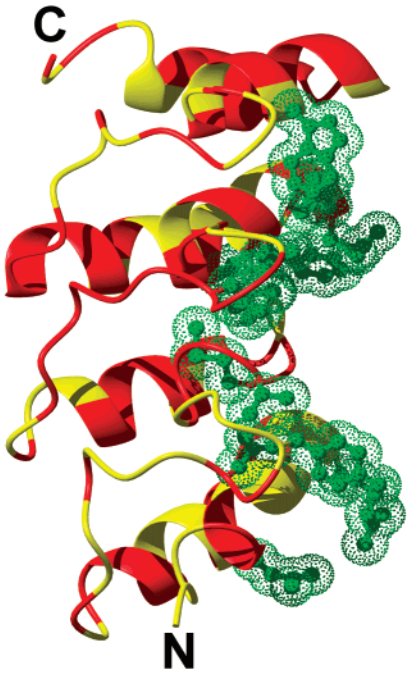

FIGURE 1: Distribution of tumor-derived mutations mapped onto the three-dimensional structure of p16. The backbone of p16 is colored in yellow with residues involved in tumor-derived mutations colored in red. Side chains of p16 in contact with cdk6 in the X-ray crystal structure of the p16−cdk6 complex are shown using the ball-and-stick model and colored in green. The molecular graphics in this paper were generated using the program MOLMOL (*58*) and the PDB coordinates 1BI7.

V126D tends to aggregate at high concentrations; therefore, it is not suitable for NMR studies. Taken together with the previously published data (*19*, *25*), we have decided to use a different technique, namely, limited proteolysis, to probe the structural integrity of p16 mutant proteins.

In the following study, we used p16 proteins with an N-terminal heart muscle protein kinase (HMPK) site and a C-terminal 6×His tag. The N- and C-terminal regions of p16 are unstructured and not required for its function (*19*, *20*); therefore, we do not anticipate that these extensions will significantly change the properties of the molecule. Our control studies using circular dichroism showed that the proteins with and without the N- and C-terminal extensions have essentially the same level of secondary structure and thermal denaturation profiles (data not shown).

Figure 2 shows the overall proteolysis rate for wild-type p16 and its mutant proteins as determined by digesting the protein with proteinase K, a low-specificity protease, and following the reaction by reverse-phase HPLC. The digestion was carried out at 0 °C in order to maximize the stability of partially folded structures. Under native conditions, upon a limited proteinase K digestion, wild-type p16 was quickly converted into several species lacking 5−15 C-terminal residues (*26*). These species eluted closely with the full-length protein on the HPLC chromatogram, and on the basis of previous studies, the C-terminal residues are dispensable for the structure and function of p16 (*18*−*20*, *37*). Therefore, we considered these species as intact molecules. In contrast to the wild type, the mutant proteins were degraded at a much higher rate and often directly into small molecular mass fragments (Figure 2A and data not shown). The accumulation of the C-terminally truncated species was observed at a much

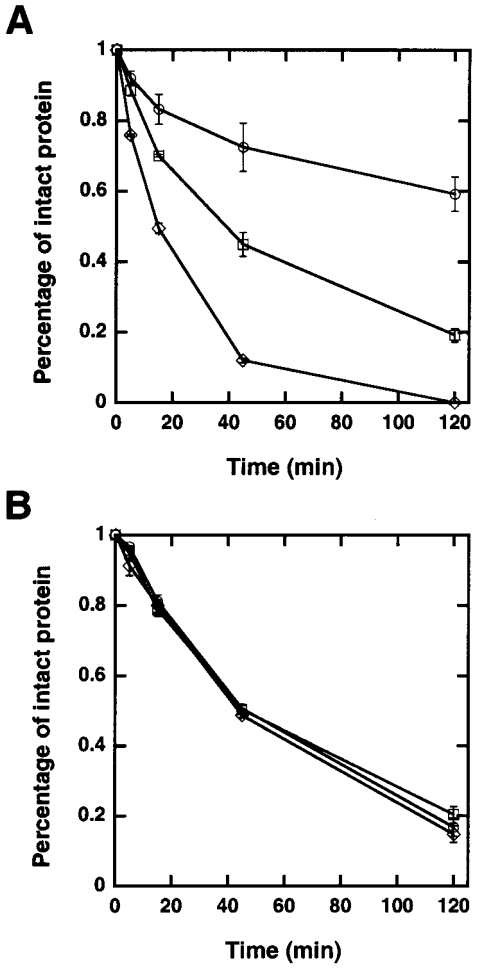

FIGURE 2: (A) Percentage of intact protein remaining in the proteolytic mixture for wild-type p16 (○), the mutant P81L (□), and the mutant V126D (◇) upon a mild proteinase K digestion under native conditions. (B) The same experiments were repeated under denaturing conditions.

lower level. As a control, we also carried out the same experiments under denaturing conditions. In 6 M urea, all three proteins were degraded at the same rate within experimental error, resulting in simultaneous formation of small molecular mass fragments (Figure 2B and data not shown). Taken together, these data indicate that the differences in the rate of proteolysis between wild-type p16 and its mutant proteins are due to the intrinsic structural differences between these proteins rather than the change in amino acid sequences and these differences disappear when the molecule is completely unfolded.

Two independent methods were used to identify the regions with enhanced proteolytic sensitivity in p16 mutant proteins. First, we labeled the proteins with a radioactive label at the N-terminus. The proteolytic fragments were separated by gel electrophoresis, and the gels were exposed to a phosphorimager (Figure 3, lanes 6−9 on each gel). Since only fragments that carry the intact N-terminal radioactive label are visible on these gels, we can infer the proteolytic cleavage sites by the molecular mass of the fragments. An advantage of this method is that it can be used to simultaneously analyze the cleavage pattern of several proteases. This will help to reduce the problem of residual protease specificity. In this study, we used four commercially available, relatively nonspecific proteases, proteinase K, Pronase,

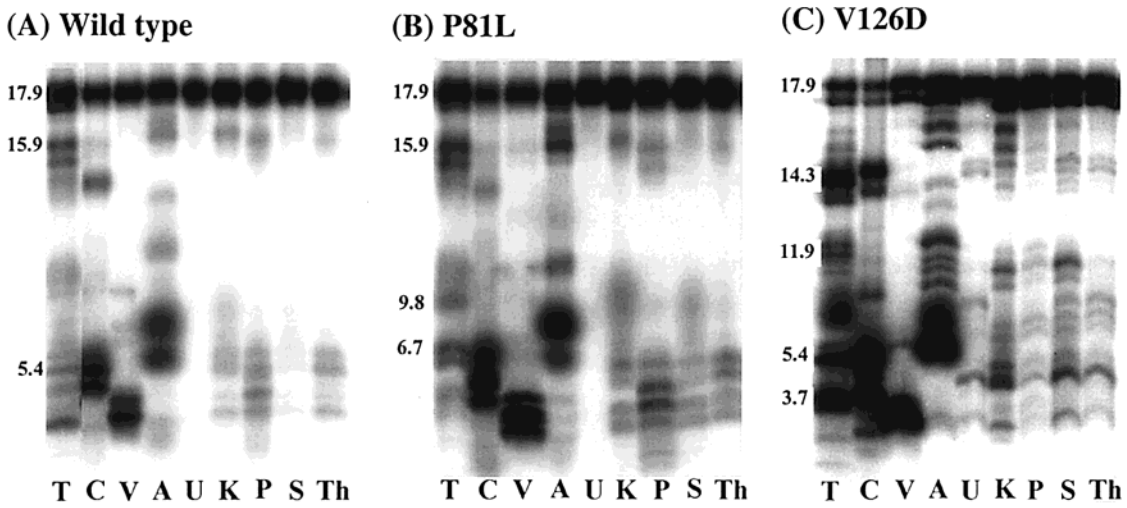

FIGURE 3: Patterns of limited proteolysis for (A) wild-type p16, (B) the mutant P81L, and (C) the mutant V126D as detected by gel electrophoresis. On the left four lanes, p16 proteins with a N-terminal radioactive label were digested with specific proteases under denaturing conditions and used as molecular mass standards (T, trypsin; C, chymotrypsin; V, V8 protease; A, endoproteinase Asp-N). On the right four lanes, the same N-terminally labeled p16 proteins were digested with nonspecific proteases under native and single-hit conditions. (K, proteinase K; P, Pronase; S, subtilisin; Th, thermolysin). In the middle was undigested protein (U), which serves as a standard to normalize the radioactive intensity.

subtilisin, and thermolysin. The digestion time and the protease-to-substrate ratio were optimized according to the single-hit conditions (*42*, *43*). Exactly the same amount of each protease was used to treat both the wild-type and mutant proteins such that the rates of proteolysis were directly comparable. To determine the molecular mass accurately, we generated a set of molecular mass standards from the same N-terminally labeled p16 by specific protease cleavage under denaturing conditions. These standards were run on the same gel with the nonspecific proteolytic products such that their electrophoretic mobility can be compared side by side (Figure 3, lanes 1−4 on each gel). Within the range of our interest (from 2 to 20 kDa), the molecular mass of the standard can be fit to a straight line with respect to the electrophoretic mobility, as determined by the migration distance of a proteolytic product relative to the uncut material (Figure 4, panels D−F). Using these fits as calibration curves, we can convert the electrophoretic mobility to the molecular mass of the fragment and then to the residue number in p16. The final results, given as the percentage of proteolytic cleavage as a function of amino acid residues in p16, are presented in Figure 4 (panels A−C).

The method described above cannot provide single residue resolution. To determine the exact location of the proteolytic cleavage sites, we analyzed the proteolytic mixtures generated by proteinase K digestion using reverse-phase HPLC, which allow us to separate most, if not all, individual proteolytic fragments (Figure 5). By examining the proteolytic profile as a function of time, we had identified a number of peaks that appeared early in the proteolytic time course when the majority of the substrate (≥85%) remained intact. A subset of these peaks had a higher intensity in the chromatogram generated from the mutant proteins than that from the wild type. These peaks were purified and analyzed by electrospray mass spectrometry in order to determine the accurate molecular mass of each fragment (Table 1). A computer program was used to automatically scan the entire amino acid sequence of p16 and identify all potential fragments with a theoretical molecular mass less than ±2 Da from the measured value. In all cases, there was always one and only one candidate that contained either the N- or C-terminus. This candidate was taken as the assignment of our proteolytic fragment on the basis of the assumption that the polypeptide chain had been cleaved only once. The robustness of this procedure has been demonstrated by direct N-terminal sequencing of more than half of all peptide fragments. In all cases, our assignment was correct.

Table 1: Identity of Proteolytic Fragments of Mutant p16 Generated by Proteinase K Digestion and Purified by Reverse-Phase HPLC

HPLC peaks[a]	mass (Da) obsd	mass (Da) predicted	assignment[c]	N-terminal amino acid sequence
P81L-1[b]	2499.5	2499.8	M^{1}RRASVAE^{2}−A^{19}	MRRAS
P81L-1a	2861.2	2861.2	M^{1}RRASVAE^{2}−R^{22}	
P81L-2[b]	3871.6	3871.4	M^{1}RRASVAE^{2}−L^{31}	MRRAS
P81L-3	6108.3	6108.0	M^{1}RRASVAE^{2}−M^{52}	
P81L-4[b]	6912.5	6912.0	M^{1}RRASVAE^{2}−A^{60}	MRRAS
P81L-5[b]	6586.4	6585.6	M^{1}RRASVAE^{2}−A^{57}	MRRAS
P81L-6[b]	6427.9	6427.4	M^{1}RRASVAE^{2}−G^{55}	MRRAS
P81L-7	11446.0	11444.8	R^{58}−D^{156}HHHHHH	
P81L-8	11923.1	11922.4	M^{53}−D^{156}HHHHHH	
V126D-1[b]	3086.4	3087.2	A^{134}−D^{156}HHHHHH	AGGTR
V126D-2	3870.0	3871.4	M^{1}RRASVAE^{2}−L^{31}	
V126D-3[b]	7477.6	7479.2	V^{95}−D^{156}HHHHHH	VVLHR
	7166.3	7167.7	H^{98}−D^{156}HHHHHH	HRAGA
V126D-3a	6404.7	6405.9	D^{105}−D^{156}HHHHHH	
V126D-4	6106.4	6108.0	M^{1}RRASVAE^{2}−M^{52}	
V126D-5	6910.0	6912.0	M^{1}RRASVAE^{2}−A^{60}	
V126D-6[b]	6583.9	6585.6	M^{1}RRASVAE^{2}−A^{57}	MRRAS
V126D-7	6425.5	6427.4	M^{1}RRASVAE^{2}−G^{55}	
V126D-8[b]	11442.1	11444.6	R^{58}−D^{156}HHHHHH	RVAEL
	9641.8	9641.6	P^{75}−D^{156}HHHHHH	

[a] Only peaks that appeared early (in the first chromatogram taken at 5 min after the reaction was started) and have a higher intensity in the chromatogram generated from a mutant protein than that from the wild type were collected and analyzed. [b] The assignment of these peaks was confirmed by N-terminal sequencing. [c] The residues in the N-terminal kinase site and the C-terminal 6×His tag are listed explicitly. Regions before residue 10 or after residue 134 are unfolded in the X-ray crystal and NMR structure of p16. Fragments that are entirely located in these regions are not shown.

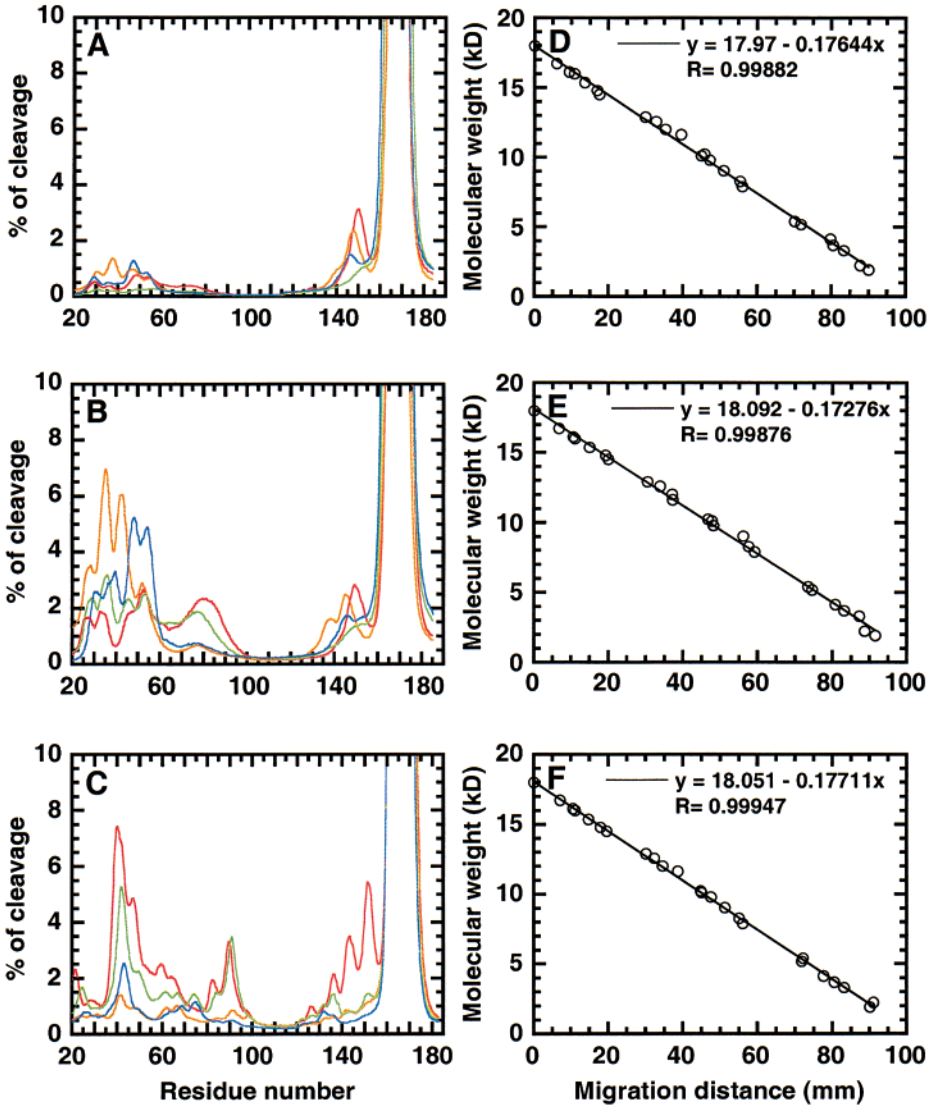

FIGURE 4: On the left, the percentage of proteolytic cleavage as a function of the residue number in p16 derived from the phosphorimager scans of the proteolytic gel shown in Figure 3. On the right, the molecular mass calibration curves used to convert the migration distances to the molecular mass of the fragments. The migration distance of the uncut protein was defined as zero. The data obtained for different proteases are color-coded: green, proteinase K; red, Pronase; blue, subtilisin; orange, thermolysin. Panels A and D are for wild-type p16, panels B and E are for the mutant P81L, and panels C and F are for the mutant V126D.

Three general observations can be made from these data. First, different proteases generally give rise to similar, although not always identical, cleavage patterns. This observation suggests that the cleavage pattern is primarily determined by the intrinsic property of the target proteins rather than the specificity of proteases, which is an important criterion for our study. Second, both the rate and pattern of limited proteolysis in wild-type p16 and its mutant proteins are significantly different. For wild-type p16, the predominant effect of proteolytic digestion is to remove the C-terminal tail, which is in the unfolded regions of the protein (*18*−*20*, *37*). The remaining part of the molecule, which consists of the four ankyrin repeats, is relatively resistant to proteolysis, although there is a small amount of cleavage that can be observed in the region corresponding to the first and second ankyrin repeat. In the mutant proteins P81L and V126D, this region becomes significantly more sensitive to proteolysis than in the wild type, suggesting that it is partially unfolded. Finally, there are some detailed differences between the limited proteolysis patterns obtained for the two mutant proteins. For example, the proteolytic cleavage sites between residues A19 and A20 and between residues R22 and G23 in P81L were not observed in V126D, and the proteolytic sensitive regions in V126D extend more toward the C-terminus than in P81L.

Figure 6 shows the regions with enhanced proteolytic sensitivity in mutant p16 proteins mapped onto the three-dimensional structure of wild-type p16. To identify these regions from the gel electrophoresis data, we averaged the percentage of proteolytic cleavage made by each of the four proteases. For the wild-type protein, the maximum value was 0.79%. Thus, we assign any region with an averaged percentage of cleavage greater than 0.8% as potentially unfolded and sensitive to proteolysis. These regions are colored in red in Figure 6. In the same figure, we also indicated the proteolytic cleavage sites identified by using the HPLC−mass spectrometry (LC-MS) method. In general, the two sets of data agree reasonably well. For P81L, all of the cleavage sites identified by using the LC-MS method are located within the proteolytic sensitive regions determined

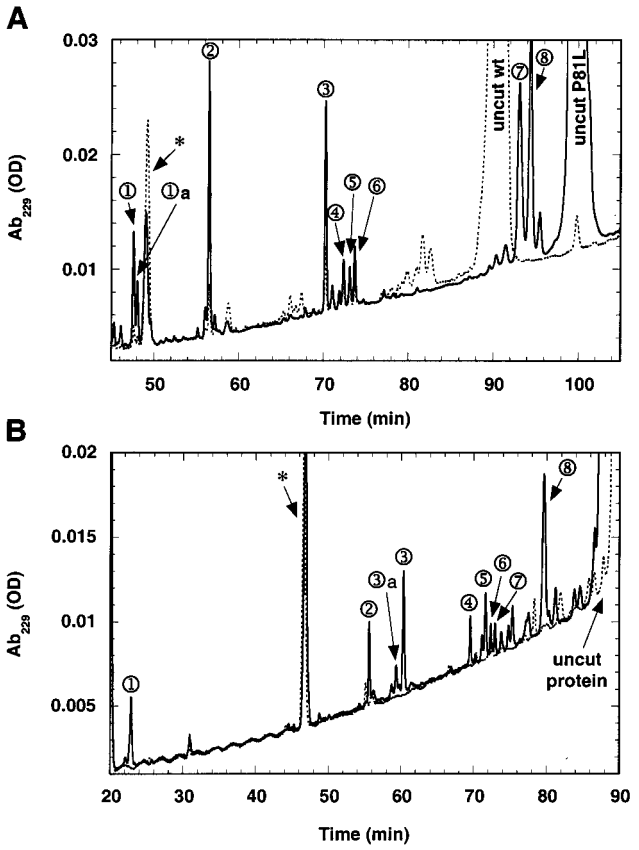

FIGURE 5: HPLC analysis of the proteolytic mixtures generated by proteinase K digestion. (A) Chromatogram of P81L (black, solid line) versus that of the wild type (gray, dashed line). (B) Chromatogram of V126D (black, solid line) versus that of the wild type (gray, dashed line). The peaks that were collected are indicated in numerical order. The asterisk indicates a small molecule species probably corresponding to the protease inhibitor.

by using the gel electrophoresis method. For V126D, this is true for the majority of the proteolytic cleavage sites identified by using the LC-MS method; however, there are some discrepancies. The cleavage sites located between residues V94 and V95 and between residues L97 and A98 are next to but outside of the proteolytic sensitive regions identified by using the gel electrophoresis method. In addition, there is a small amount of cleavage at the peptide bond after residue L104, which may be below the threshold that can be detected by gel electrophoresis. Despite these discrepancies, both methods show that the proteolytic sensitive regions in V126D are shifted somewhat toward the C-terminus with respect to those in P81L.

DISCUSSION

The goal of this study is to identify the location of structural defects introduced by tumor-derived mutations in the cyclin-dependent kinase inhibitor p16 and to understand how these mutations disrupt the structure and function of the protein. With the completion of human genome project and the effort to screen for all single nucleotide polymorphisms, a large number of disease-related mutations will be discovered in the next few years. Many of these mutations may exert their effect by modulating the structure and stability of proteins. Therefore, understanding the structural consequences of these mutations will become increasingly important.

Limited proteolysis has long been used to identify domains and flexible regions in proteins (*44*, *45*). It has been shown that the rate of proteolysis in native proteins is primarily determined by the local flexibility of the polypeptide chain, as measured by the crystallographic *B* factors (*33*, *34*, *46*, *47*). Both modeling and experimental studies suggest that, in order for the polypeptide chain to fit into the active site

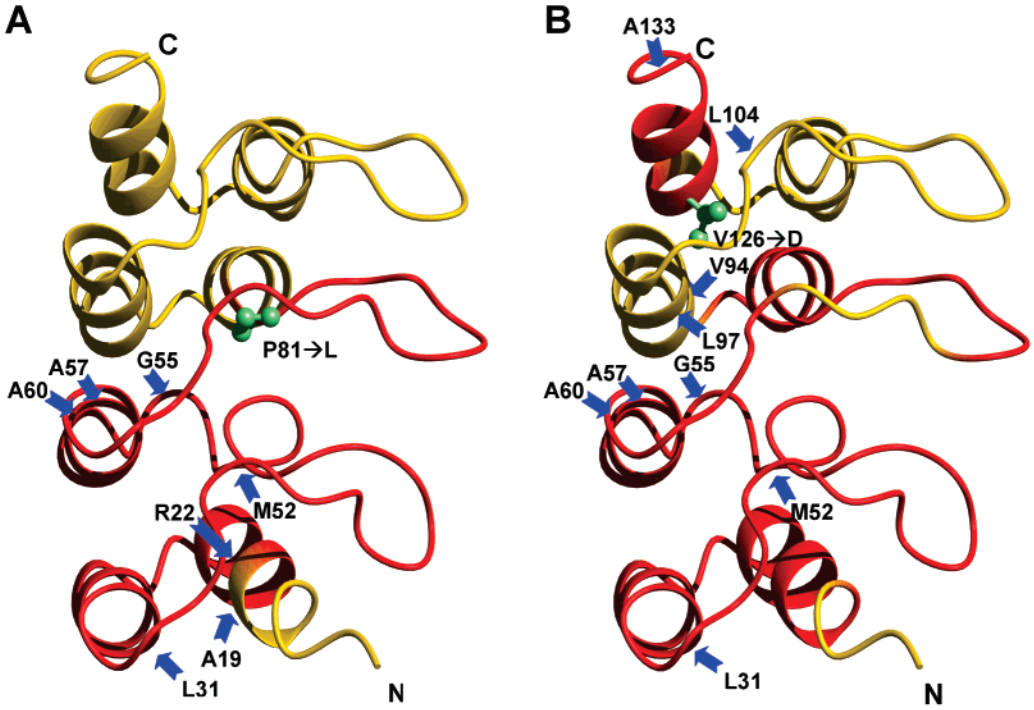

FIGURE 6: Regions with enhanced proteolytic sensitivity (red) in (A) P81L and (B) V126D mapped onto the three-dimensional structure of wild-type p16. The gel electrophoretic method cannot detect unfolded regions in the first α-helix. The arrows indicate the approximate position of the proteolytic cleavage sites identified by using the LC-MS method with the name of the residue just preceding the cleavage site. The side chains of the mutated residues are colored in green and shown using the ball-and-stick model.

of a protease, the region must be sufficiently mobile and be able to change its conformations (*48*, *49*). Regions that are involved in regular secondary structures are usually protected from limited proteolysis (*50*−*52*). Thus, limited proteolysis can be used to identify locally unfolded regions. This approach is particularly useful for studying partially folded proteins, such as the tumor-derived mutants of p16 described in this paper. The main advantage of limited proteolysis is that it does not require the molecule to have a fixed tertiary structure, although it cannot be used to provide direct three-dimensional information.

The use of limited proteolysis for protein structural analysis requires several stringent conditions. First, the proteolytic cleavage rates must be primarily determined by the intrinsic structure and stability of the substrate rather than the specificity of proteases. To satisfy this criterion, we used four relatively nonspecific proteases and verified that the cleavage pattern is largely independent of the choice of enzymes. Second, the experiment must be carried out under single-hit conditions, such that the molecules are not cleaved more than once. This is because a molecule that has been cleaved already may lose its structure, allowing additional cleavage to occur at a site that is normally buried in the intact protein. In previous DNA footprinting studies, it has been estimated that when 85% of the starting material remains intact, the amount of second and multiple cleavages is less than 1% (*42*). This criterion was used to optimize our proteolytic conditions. Since the unfolding of the protein may expose additional cleavage sites, we expect that, for protease digestions, the amount of multiple cleavages may be slightly higher. In any case, all of our fragments isolated by using the LC-MS method contain either the N- or C-terminus, which provides direct evidence that our experimental design is consistent with the single-hit condition. Finally, because the intrinsic proteolytic rates may vary from residue to residue, it is necessary to compare the proteolytic pattern of a mutant protein with that of the wild type.

In this study, we used two different methods to identify proteolytic cleavage sites in p16 proteins. Each of these methods has its intrinsic advantages and disadvantages. The gel electrophoresis method is particularly useful for screening a large number of proteases and conditions. However, the result is more difficult to interpret. The 0.8% cutoff value was chosen somewhat arbitrarily. A more relevant measure would be the relative proteolytic sensitivity observed in a mutant protein versus that in the wild type. This was not implemented due to technical difficulties. The HPLC method is more reproducible and quantitative. However, it can only resolve a finite number of proteolytic fragments, making it less useful in providing a global picture.

The two mutations investigated here are unrelated to each other and located in different regions of the molecule. P81 and V126 are likely to play different roles in stabilizing the three-dimensional structure of p16. Residue P81 is located at the beginning of a α-helix, serving as a N-cap residue. The peptide bond of P81 is in a *trans* conformation; thus, we do not anticipate that the mutation will significantly alter the backbone structure. However, this substitution may destabilize the α-helix and the hydrophobic packing surrounding this residue. Interestingly, substitutions P48L and P114L, which are located at equivalent positions in the second and fourth ankyrin repeats, have also been documented as tumor-derived mutations (*11*). It has been shown that both P81L and P114L exhibit a molten globule-like conformation (*26*). Residue V126 is a buried hydrophobic residue located at the interface between two ankyrin repeats. Substitution of V126 by a charged residue will likely disrupt the assembly of this multiple repeat protein. V126D has been reported to be a temperature-sensitive mutation, although previous CD studies show that the protein has a lower level of α-helical secondary structure than the wild type and the level of secondary structure does not strongly depend on temperature (*24*, *26*; Zhang and Peng, unpublished results).

FIGURE 7: Commonly unfolded regions in mutant p16 proteins (red) partially overlap with the p16−cdk6 binding site. In this figure, the p16 molecule is colored in yellow, and the cdk6 molecule is colored in blue. The side chains of p16 in contact with cdk6 are colored in green.

Perhaps the most interesting observation of this study is that there is a large overlap between the proteolytic sensitive regions in P81L and V126D. The unfold regions in mutant P81L are located almost entirely in the N-terminal two ankyrin repeats. Interestingly, although the mutation V126D is located near the C-terminus, the unfolded regions in this mutant protein also include the N-terminal two ankyrin repeats, as well as the ankyrin repeats closer to the C-terminus (i.e., the mutant V126D is a more globally unfolded protein). In the LC-MS study, approximately half of the proteolytic cleavage sites (i.e., the cleavage sites after residues L31, M52, G55, A57, and A60) occurred in both mutant proteins. The commonly unfolded regions in P81L and V126D include the second α-helix of the first ankyrin repeat, the entire second ankyrin repeat, and the *β*-hairpin loops connecting the first to the second ankyrin repeat and the second to the third ankyrin repeat. This region contains many residues that are important for cyclin-dependent kinase binding (Figure 7). The *β*-hairpin loops in ankyrin repeat proteins are often used as a recognition motif; therefore, it is not surprising that the loss of structural integrity in these regions may result in functional impairment. In addition, the second ankyrin repeat region of p16 and other *INK4* inhibitor

proteins all contain a distorted α-helical loop in place of a regular α-helix. The conservation of this feature suggests that it is important for cyclin-dependent kinase binding. According to the X-ray crystal structure of the p16−cdk6 complex, approximately 50% of all residues that are in direct contact with cdk6 are located in the second ankyrin repeat region (including the adjacent β-hairpin loops). These residues contribute significantly to the extensive hydrogen bond network between the two molecules. Thus, our results strongly suggest that the unfolding of the second ankyrin repeat region could be a general mechanism by which tumor-derived mutations inactivate p16.

The unfolded regions in mutant p16 proteins are not necessarily located in the vicinity of, or symmetric with respect to, the site of mutations, suggesting that the energetic consequence of mutations may propagate through a long distance in protein structure (*53*). A careful inspection of Figure 3 suggests that the proteolytic digestion pattern of wild-type p16 bears certain similarity with the digestion pattern of P81L (this will become clear if the wild-type protein is treated with a higher concentration of proteases; data not shown). This observation suggests that the N-terminal half of p16 may be intrinsically unstable and the differences in intrinsic stability may determine which part of the molecule becomes unfolded first in response to disruptive mutations. Among the four ankyrin repeat motifs, the second ankyrin repeat is likely to have the lowest stability due to structural distortion. Consistent with this view, our early studies indicate that the C-terminal half of p16, containing the third and fourth ankyrin repeat, can fold independently, whereas the N-terminal half of p16 cannot be expressed in *E. coli* (*54*). Previous NMR studies also show that the resonances corresponding to residues in the first β-hairpin loop between the first and second ankyrin repeats and residues in the second ankyrin repeat in $p19^{INK4d}$ exhibit exchange broadening (*55*) and that the amide protons in the N-terminal half of $p15^{INK4b}$, $p16^{INK4a}$, and $p18^{INK4c}$ are poorly protected (*56*). Both of these studies suggest that the second ankyrin repeat region may be involved in local unfolding and has high levels of conformational dynamics. Thus, it is possible that the stability of this region in wild-type p16 is already at a critical threshold. Consequently, a small perturbation in the rest of the protein may induce a large conformational rearrangement.

Both the CD and limited proteolysis studies suggest that the tumor-derived mutant proteins are not entirely unfolded. To determine the residual structure in these mutant proteins will be difficult since none of the existing high-resolution techniques is designed to work with partially unfolded proteins. However, this observation suggests that it may be possible to restore the function of tumor-derived mutant proteins by stabilizing its native state.

Understanding the structural defects caused by tumor-related mutations may eventually lead to the discovery of novel therapeutics. One potential strategy is to stabilize the native structure of p16 by interaction with small molecules. In this regard, it is important to note that a small molecule ligand of tumor suppressor p53 identified by high throughput screening has been shown recently to exhibit the ability to rescue tumor-derived mutations (*57*). A small molecule that binds to the distorted second ankyrin repeat region may provide both the necessary stabilization energy and specificity for p16, such that it will not interfere with other ankyrin repeat proteins in a crowded cellular environment.

ACKNOWLEDGMENT

We thank David Beach for the cDNA clone of p16 and Nikola Pavletich for the X-ray coordinates of the p16−cdk6 complex. The DNA sequencing was performed by Ming Li and John Glynn at the Molecular Core Facility at the University of Connecticut Health Center. The phosphorimager was maintained by the Center of Biomedical Imaging Technology. We are grateful to Igor Kaltashov and Steve Eyles for assistance on mass spectrometry studies, Michael Bourne for N-terminal sequencing, Mark Maciejewski for help on NMR and figure preparation, and Leila Mosavi, Tobin Cammett, and Peter Setlow for critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Figure S1 showing the $^{15}N-^{1}H$ HSQC spectra of wild-type p16 and its mutant P81L, with the protein containing an N-terminal 6×His tag in front of full-length human p16 (total 162 residues). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Chin, L., Pomerantz, J., and DePinho, R. A. (1998) *Trends Biochem. Sci. 23*, 291−296.
2. Sharpless, N. E., and DePinho, R. A. (1999) *Curr. Opin. Genet. Dev. 9*, 22−30.
3. Sherr, C. J. (2000) *Cancer Res. 60*, 3689−3695.
4. Serrano, M., Hannon, G. J., and Beach, D. (1993) *Nature 366*, 704−707.
5. Sherr, C. J., and Roberts, J. M. (1995) *Genes Dev. 9*, 1149−1163.
6. Sherr, C. J. (1996) *Science 274*, 1672−1677.
7. Pomerantz, J., Schreiber-Agus, N., Liegeois, N. J., Silverman, A., Alland, L., Chin, L., Potes, J., Chen, K., Orlow, I., Lee, H.-W., Cordon-Cardo, C., and DePinho, R. A. (1998) *Cell 92*, 713−723.
8. Zhang, Y., Xiong, Y., and Yarbrough, W. G. (1998) *Cell 92*, 725−734.
9. Kamb, A., Gruis, N. A., Weaver-Feldhaus, J., Liu, Q., Harshman, K., Tavtigian, S. V., Stockert, E., Day, R. S. R., Johnson, B. E., and Skolnick, M. H. (1994) *Science 264*, 436−440.
10. Nobori, T., Miura, K., Wu, D. J., Lois, A., Takabayashi, K., and Carson, D. A. (1994) *Nature 368*, 753−756.
11. Ruas, M., and Peters, G. (1998) *Biochim. Biophys. Acta 1378*, F115−F177.
12. Hussussian, C. J., Struewing, J. P., Goldstein, A. M., Higgans, P. A. T., Ally, D. S., Sheahan, M. D., Clark, W. H., Tucker, M. A., and Dracopoli, N. C. (1994) *Nat. Genet. 8*, 15−21.
13. Kamb, A., Shattuck-Eidens, D., Eeles, R., Liu, Q., Gruis, N. A., Ding, W., Hussey, C., Tran, T., Miki, Y., Weaver-Feldhaus, J., McClure, M., Aitken, J. F., Anderson, D. E., Bergman, W., Frants, R., Goldger, D. E., Green, A., MacLennan, R., Martin, N. G., Meyer, L. J., Youl, P., Zone, J. J., Skolnick, M. H., and Cannon-Albright, L. A. (1994) *Nat. Genet. 8*, 22−26.
14. Koh, J., Enders, G. H., Dynlacht, B. D., and Harlow, E. (1995) *Nature 375*, 506−510.
15. Ranade, K., Hussussian, C. J., Sikorski, R. S., Varmus, H. E., Goldstein, A. M., Tucker, M. A., Serrano, M., Hannon, G. J., Beach, D., and Dracopoli, N. C. (1995) *Nat. Genet. 10*, 114−116.
16. Reymond, A., and Brent, R. (1995) *Oncogene 11*, 1173−1178.
17. Wick, S. T., Dubay, M. M., Imanil, I., and Brizuela, L. (1995) *Oncogene 11*, 2013−2019.

18. Yang, R., Gombart, A. F., Serrano, M., and Koeffler, H. P. (1995) *Cancer Res. 55*, 2503−2506.
19. Byeon, I. J., Li, J., Ericson, K., Selby, T. L., Tevelev, A., Kim, H. J., O'Maille, P., and Tsai, M. D. (1998) *Mol. Cell 1*, 421−431.
20. Russo, A. A., Tong, L., Lee, J.-O., Jeffrey, P. D., and Pavletich, N. P. (1998) *Nature 395*, 237−243.
21. Lux, S. E., John, K. M., and Bennett, V. (1990) *Nature 344*, 36−42.
22. Bork, P. (1993) *Proteins: Struct., Funct., Genet. 17*, 363−374.
23. Sedgwick, S. G., and Smerdon, S. J. (1999) *Trends Biochem. Sci. 24*, 311−316.
24. Tang, K. S., Guralnick, B. J., Wang, W. K., Fersht, A. R., and Itzhaki, L. S. (1999) *J. Mol. Biol. 285*, 1869−1886.
25. Tevelev, A., Byeon, I.-J., Selby, T., Ericson, K., Kim, H.-J., Kraynov, V., and Tsai, M.-D. (1996) *Biochemistry 35*, 9475−9487.
26. Zhang, B., and Peng, Z.-Y. (1996) *J. Biol. Chem. 271*, 28734−28737.
27. Eberstadt, M., Huang, B., Olejniczak, E. T., and Fesik, S. W. (1997) *Nat. Struct. Biol. 4*, 983−985.
28. Lim, W. A., Farruggio, D. C., and Sauer, R. T. (1992) *Biochemistry 31*, 4324−4333.
29. Uversky, V. N., Leontiev, V. V., and Gudkov, A. T. (1992) *Protein Eng. 5*, 781−783.
30. Lascu, I., Schaertl, S., Wang, C., Sarger, C., Giartosio, A., Briand, G., Lacombe, M. L., and Konrad, M. (1997) *J. Biol. Chem. 272*, 15599−15602.
31. Yuan, C., Byeon, I. J., Poi, M. J., and Tsai, M. D. (1999) *Biochemistry 38*, 2919−2929.
32. Matthews, J. M., Norton, R. S., Hammacher, A., and Simpson, R. J. (2000) *Biochemistry 39*, 1942−1950.
33. Fontana, A., de Laureto, P. P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997) *Folding Des. 2*, R17−R26.
34. Hubbard, S. J. (1998) *Biochim. Biophys. Acta 1382*, 191−206.
35. Castellano, M., Pollock, P. M., Walters, M. K., Sparrow, L. E., Down, L. M., Gabrielli, B. G., Parsons, P. G., and Hayward, N. K. (1997) *Cancer Res. 57*, 4868−4875.
36. Flores, J. F., Walker, G. J., Glendening, J. M., Haluska, F. G., Castresana, J. S., Rubio, M. P., Pastorfide, G. C., Boyer, L. A., Kao, W. H., Bulyk, M. L., Barnhill, R. L., Hayward, N. K., Housman, D. E., and Fountain, J. W. (1996) *Cancer Res. 56*, 5023−5032.
37. Parry, D., and Peters, G. (1996) *Mol. Cell. Biol. 16*, 3844−3852.
38. Doering, D. S. (1992), Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
39. Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.
40. Pollock, P. M., Pearson, J. V., and Hayward, N. K. (1996) *Genes, Chromosomes Cancer 15*, 77−88.
41. Smith-Sorensen, B., and Hovig, E. (1996) *Hum. Mutat. 7*, 294−303.
42. Brenowitz, M., Senear, D. F., Shea, M. A., and Ackers, G. K. (1986) *Method Enzymol. 130*, 132−181.
43. Heyduk, E., and Heyduk, T. (1994) *Biochemistry 33*, 9643−9650.
44. Porter, R. R. (1950) *Biochem. J. 46*, 479−484.
45. Porter, R. R. (1959) *Biochem. J. 73*, 119−126.
46. Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., and Zambonin, M. (1986) *Biochemistry 25*, 1847−1851.
47. Novotny, J., and Bruccoleri, R. E. (1987) *FEBS Lett. 211*, 185−189.
48. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.
49. Hubbard, S. J., Eisenmenger, F., and Thornton, J. M. (1994) *Protein Sci. 3*, 754−768.
50. Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., and Scaramella, E. (1997) *J. Mol. Biol. 266*, 223−230.
51. Polverino, de Laureto, P., De Filippis, V., Di Bello, M., Zambonin, M., and Fontana, A. (1995) *Biochemistry 34*, 12596−12604.
52. Wu, L. C., and Kim, P. S. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 14314−14319.
53. Lockless, S. W., and Ranganathan, R. (1999) *Science 286*, 295−299.
54. Zhang, B., and Peng, Z.-Y. (2000) *J. Mol. Biol. 299*, 1121−1132.
55. Renner, C., Baumgartner, R., Noegel, A. A., and Holak, T. A. (1998) *J. Mol. Biol. 283*, 221−229.
56. Yuan, C., Li, J., Selby, T. L., Byeon, I.-J. L., and Tsai, M.-D. (1999) *J. Mol. Biol. 294*, 201−211.
57. Foster, B. A., Coffey, H. A., Morin, M. J., and Rastinejad, F. (1999) *Science 286*, 2507−2510.
58. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.

BI0117100