Protein Science (1993), *2*, 155-164. Cambridge University Press. Printed in the USA.
Copyright © 1993 The Protein Society

The binding domain structure of retinoblastoma-binding proteins

JAMES FIGGE,[1,2] KIMBERLY BREESE,[1] SANDOR VAJDA,[4] QING-LIN ZHU,[4] LESLIE EISELE,[3] THOMAS T. ANDERSEN,[3] ROBERT MACCOLL,[5] THOMAS FRIEDRICH,[2] AND TEMPLE F. SMITH[4]

Departments of [1] Medicine, [2] Microbiology, Immunology and Molecular Genetics, and
[3] Biochemistry and Molecular Biology, Albany Medical College, Albany, New York 12208
[4] BioMolecular Engineering Research Center, Boston University, Boston, Massachusetts 02215
[5] New York State Wadsworth Center for Laboratories and Research, Albany, New York 12201

(RECEIVED July 8, 1992; REVISED MANUSCRIPT RECEIVED September 30, 1992)

Abstract

The retinoblastoma gene product (Rb), a cellular growth suppressor, complexes with viral and cellular proteins that contain a specific binding domain incorporating three invariant residues: Leu-X-Cys-X-Glu, where X denotes a nonconserved residue. Hydrophobic and electrostatic properties are strongly conserved in this segment even though the nonconserved amino acids vary considerably from one Rb-binding protein to another. In this report, we present a diagnostic computer pattern for a high-affinity Rb-binding domain featuring the three conserved residues as well as the conserved physico-chemical properties. Although the pattern encompasses only 10 residues (with only 4 of these explicitly defined), it exhibits 100% sensitivity and 99.95% specificity in database searches. This implies that a certain pattern of structural and physico-chemical properties encoded by this short sequence is sufficient to govern specific Rb binding. We also present evidence that the secondary structural conformation through this region is important for effective Rb binding.

Keywords: CD spectroscopy; conformational filtering; papilloma virus E7 protein; peptide conformation; protein-protein binding interactions; retinoblastoma protein; SV40 T antigen

The human retinoblastoma locus encodes a 928-residue nuclear product (Rb) that functions as a cellular growth suppressor (see Weinberg, 1991, for review), possibly by governing passage through a checkpoint in G1 of the cell cycle (Goodrich et al., 1991). This activity may result from complex formation between Rb and other nuclear proteins (Huang et al., 1991; Kaelin et al., 1991) including the transcription factor, E2F (Bandara et al., 1991; Chellappan et al., 1991; Chittenden et al., 1991; Helin et al., 1992; Kaelin et al., 1992), the c-myc proto-oncogene product (Rustgi et al., 1991), and RBP-1 and RBP-2 (Defeo-Jones et al., 1991). Rb also complexes specifically with transforming proteins of several oncogenic DNA viruses, including the simian virus 40 (SV40) and polyomavirus large T antigens (DeCaprio et al., 1988; Dyson et al., 1990), the adenovirus E1a protein (Whyte et al., 1988), and the papilloma virus E7 protein (Dyson et al., 1989). The ability to complex Rb may contribute to the transforming properties of these viral products (Weinberg, 1991).

The aforementioned viral proteins as well as two cellular proteins, RBP-1 and RBP-2, share a structurally similar segment (Stabel et al., 1985; Figge et al., 1988; Moran, 1988; Defeo-Jones et al., 1991) containing three invariant residues: Leu-X-Cys-X-Glu, where X represents a nonconserved residue. Genetic mapping (Munger et al., 1989; Whyte et al., 1989; Barbosa et al., 1990; Firzlaff et al., 1991; Heck et al., 1992) and peptide-binding competition experiments (Jones et al., 1990; Kaelin et al., 1990) have demonstrated that this segment encodes a high-affinity Rb-binding domain. Short peptides (9-14 residues) containing the Rb-binding domain from SV40 T or papilloma virus E7 can compete against their full-length counterparts for specific Rb-binding (Jones et al., 1990; Kaelin et al., 1990), implying that a short segment contains the necessary structural information to govern

Reprint requests to: Dr. James Figge, Department of Medicine, Albany Medical College, A44, 47 New Scotland Ave., Albany, New York 12208.

specific binding. In this report, we show that hydrophobic and electrostatic properties are strongly conserved in this segment even though the nonconserved amino acids vary considerably from one Rb-binding protein to another. We present a diagnostic computer pattern for a high-affinity Rb-binding domain containing the three conserved residues as well as the common hydrophobicity profile and specific charged-group requirements. Although the pattern encompasses only 10 residues (with only 4 of these explicitly defined) it exhibits 100% sensitivity and 99.95% specificity in identifying high-affinity Rb-binding proteins in database searches. In keeping with the peptide competition experiments, this implies that a certain pattern of structural and physico-chemical properties encoded by this short sequence is sufficient to govern specific Rb-binding. We also hypothesize that the secondary structural conformation of the binding domain is important for effective Rb binding, and we present evidence in support of this idea.

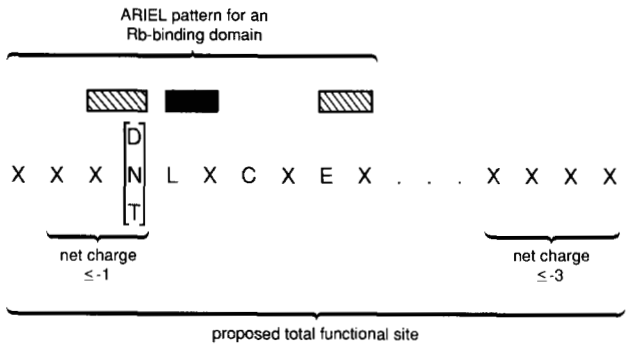

Fig. 1. Diagnostic pattern of a high-affinity retinoblastoma (Rb)-binding domain. Invariant amino acids are shown in single letter code. Alternate residues at a given position are grouped within brackets. The X's represent any residue except lysine or arginine. A net charge of ≤ -1 occurs within three residues amino-terminal to the conserved leucine, indicating that at least one glutamic or aspartic acid residue is found in that region. The cross-hatched bars represent regions where minima occur in the hydrophobicity profile, and the solid bar shows the region where a hydrophobicity maximum occurs. In most Rb-binding proteins there is an acidic domain carboxy-terminal to the binding domain, separated by a spacer of variable length (denoted by . . .). Both the Rb-binding domain and the acidic domain are required for the full transforming function of the viral proteins (Barbosa et al., 1990; Firzlaff et al., 1991).

Results

Diagnostic pattern for a high-affinity Rb-binding domain

A diagnostic pattern for a high-affinity Rb-binding domain incorporating elements of primary structure, hydrophobicity, and electrostatic charge is shown in Figure 1. This pattern was developed by analyzing sequences of 21 known high-affinity Rb-binding proteins (Table 1A). A control set of non-Rb-binding proteins was used to maximize the pattern's specificity and included the entire Swissprot-22 database and a set of nine mutant viral proteins (Table 1B). The pattern was refined with the assistance of the artificial intelligence pattern-induction program, ARIEL (Lathrop et al., 1990; Zhu et al., 1990), running on a massively parallel supercomputer, the Connection Machine 2. The program refines the pattern by simultaneously optimizing both the pattern's sensitivity and specificity. Local hydrophobicity profiles for each of the sequences were generated using Eisenberg's hydrophobicity scale (Rees et al., 1989) with a smoothing algorithm. A simple weighted average (2:11:2) was found to be optimal for smoothing, with the value assigned to the middle residue. The pattern's sensitivity was 100% and specificity 99.95%. None of the nine known nonbinding mutants in the control set were detected by the pattern.

The pattern (Fig. 1) matches to the cellular Rb-binding proteins RBP-1 and RBP-2 (Defeo-Jones et al., 1991) but does not match to c-myc or E2F. Although c-myc binds Rb, the interaction is weak (Rustgi et al., 1991), and the sequence similarity between c-myc and the other Rb-binding proteins is distant (Figge et al., 1988). The gene encoding E2F, a cellular transcription factor, has recently been cloned (Helin et al., 1992; Kaelin et al., 1992), and its product also binds Rb. The Rb-binding domain of E2F was mapped to an 18-amino acid segment (Helin et al., 1992) that does not contain a match to our pattern. Because E1A and E7 sequences can displace E2F from Rb (Bandara & La Thangue, 1991; Helin et al., 1992; Kaelin et al., 1992), these results suggest three possible models: (1) E1A/E7 and E2F may bind to separate sterically interfering sites on Rb; (2) E1A/E7 may bind to and stabilize Rb in a particular conformation that excludes E2F from binding at another site; or (3) E1A/E7 may compete with E2F for binding to the same site within Rb via partially related or entirely different sequence motifs (see Discussion). Similar models apply to the c-myc interaction with Rb.

Of note, the pattern (Fig. 1) matches to the E7 proteins from 12 of 14 human papillomavirus strains in the Swissprot-22 database (strains 1A, 2A, 5, 8, 16, 18, 31, 33, 39, 45, 47, and 57). A number of these (strains 16, 18, 31, 33, 39, and 45) are known to be associated with carcinoma of the cervix in humans (Gage et al., 1990; Howley, 1991). The pattern does not match two strains (6b and 11) that are associated with benign lesions that rarely progress to malignancy. The oncogenic strains contain the sequence DLXCXE, whereas the nononcogenic types encode the sequence GLXCXE. In keeping with our pattern, Heck et al. (1992) have shown that the interchange of G for D in the indicated position decreases the affinity of the E7 protein for Rb.

Even though the pattern appears to have a low information content, it defines a highly specific, conserved

Table 1. *Binding domains and flanking sequences*[a]

Protein	Sequence	Reference
A. Wild-type (wt) and mutant high-affinity retinoblastoma (Rb)-binding proteins		
SV40 T antigen (wt)	-NEENLFCS**E**EMPSSDDEATA-	DeCaprio et al., 1988
E1A/T hybrid	-VIENLFCS**E**EMPSSDDEATA-	Moran, 1988
Mouse Py (wt)	-EQPDLFCY**E**EPLLSPNPSSP-	Dyson et al., 1990
Monkey LPV (wt)	-CCDDLFCS**E**TMSSSSDEDTP-	Dyson et al., 1990
Baboon SA12 (wt)	-WDEDLFCH**E**DMFQSDE-	Dyson et al., 1990
JCV (wt)	-WDEDLFCH**E**EMFASDDENTG-	Dyson et al., 1990
BKV (wt)	-WDEDLFCH**E**DMFASDEEATA-	Dyson et al., 1990
HPV E7 (strain 16, wt)	-ETTDLYCY**E**QLNDSSEEEDE-	Munger et al., 1989
HPV E7 (strain 18, wt)	-IPVDLLCH**E**QLSDSEEENDE-	Munger et al., 1989
E7-Ala	-ETTDLYCY**E**QLNDAAEEEDE-	Firzlaff et al., 1991
E7-Asp	-ETTDLYCY**E**QLNDDDEEEDE-	Firzlaff et al., 1991
E7-Gln	-ETTDLYCY**E**QLNQSSQQQQQ-	Firzlaff et al., 1991
Delta-EDE	-ETTDLYCY**E**QLNDSSEEIDG-	Munger et al., 1989
S31-Arg	-ETTDLYCY**E**QLNDRSEEEDE-	Barbosa et al., 1990
S32-Trp	-ETTDLYCY**E**QLNDSWEEEDE-	Barbosa et al., 1990
S31,32-Arg/Pro	-ETTDLYCY**E**QLNDRPEEEDE-	Barbosa et al., 1990
E35,D36-Asp/His	-ETTDLYCY**E**QLNDSSEEDHE-	Barbosa et al., 1990
E1A (strain 2,5, wt)	-EVIDLTCH**E**AGFPPSDDEDE-	Whyte et al., 1989
NCdl	-FPPDLTCH**E**AGFPPSDDEDE-	Whyte et al., 1989
CTdl 940	-EVIDLTCH**E**AL-[COOH]	Whyte et al., 1989
RBP-1 (wt)	-GPETLVCH**E**VDLDDLDEKDK-	Defeo-Jones et al., 1991
B. Mutant non-Rb-binding sequences		
T antigen mutants		
K1	-NEENLFCSKEMPSSDDEATA-	DeCaprio et al., 1988
PVUI	-NEENLFCYDEATADSQHSTP-	DeCaprio et al., 1988
C105D114	-NEENLFGS**E**EMPSSDKEATA-	DeCaprio et al., 1988
E7 mutants		
E26-Q	-ETTDLYCYQQLNDSSEEEDE-	Munger et al., 1989
C24-Gly	-ETTDLYGY**E**QLNDSSEEEDE-	Barbosa et al., 1990
E26-Gly	-ETTDLYCYGQLNDSSEEEDE-	Barbosa et al., 1990
Delta-DLYC	-ETTYEQLNDSSEEEDEIDGP-	Munger et al., 1989
E1A mutants		
928	-EVIDLTGH**E**AGFPPSDDEDE-	Moran, 1988
CTdl 934	-EVIDLTCHL-[COOH]	Whyte, 1989

[a] Bold letters denote conserved signature sequences.

structure with a characteristic charged-group profile. First, a Glu or Asp appears one to three residues before the conserved Leu. Second, a conserved Glu is found four residues C-terminal to the Leu. Third, no Lys or Arg residues are found. These conserved features suggest that electrostatic interactions are required for binding to Rb (Jones et al., 1990; Breese et al., 1991). The hydrophobicity profile of this region demonstrates two local minima flanking a region of maximum hydrophobicity, suggesting that a hydrophobic interaction may also be involved in Rb binding. Only three residues are absolutely conserved, and a fourth residue is characterized by a two-carbon hydrophilic side chain (Asp, Asn, or Thr). In addition to these criteria, it has been suggested that the secondary structural conformation of the binding domain may be a condition of the interaction with Rb (Jones et al., 1990; Breese et al., 1991).

Binding studies on a proline-substituted SV40 T antigen mutant

To determine whether there are conformational constraints that govern the ability of a peptide to competitively bind Rb, we designed a peptide derivative of SV40 T conforming to the pattern in Figure 1 but containing prolines in two of the nonconserved positions (Table 2, peptide P2). It was expected that the prolines, due to their unique conformational properties, might prevent the peptide from assuming its active binding conformation. This peptide was compared with a wild-type SV40 T-derived peptide (wt), a nonbinding mutant analog (K1), an unrelated peptide (X), and bovine serum albumin (BSA) in a competitive binding assay (Table 2). The assay employs a recombinant Rb/glutathione *S*-transferase (Rb/GST) hybrid protein (Kaelin et al., 1991), encoded by the plasmid

Table 2. *Competitive Rb-binding assay*[a]

Competitor peptide/protein		Peptide sequence	Relative amount of bound ^{35}S T
wt	(0.70 mM)	ENLFCSEEMPSSDDE	0
	(0.35 mM)		12
K1	(1.0 mM)	ENLFCSKEMPSSDDE	1,522
P2	(1.0 mM)	ENLPCPEEMPSSDDE	880
X	(1.5 mM)	SRALTPSPVMVLENIEPE	853
BSA	(1.0 mM)		1,000

[a] The peptides were synthesized as described in Materials and methods. The primary peptide sequences were confirmed by automated gas-phase sequencing. Competitive Rb-binding assays against radiolabeled full-length SV40 T antigen were carried out as described in Materials and methods. The table gives the relative amount of radiolabeled T bound to Rb/glutathione *S*-transferase in the presence of the indicated competitor. The wild-type (wt) peptide is a known strong competitor and serves as a positive control. K1, bovine serum albumin (BSA), and an unrelated peptide (X) are weak competitors and serve as negative controls. The results have been normalized such that BSA was set to 1,000.

pGEX2T-RB(379-792). The Rb/GST hybrid protein retains the ability to specifically bind SV40 T antigen, E1A, and E7 proteins. The hybrid protein also can be immobilized by binding to glutathione/Sepharose 4B beads, making it a convenient reagent to use in a competitive binding assay (Kaelin et al., 1991). A mutant Rb/GST sequence, encoded by the plasmid pGEX2T-RB(379–792; DL 573–645), was used as a negative control and failed to bind radiolabeled T antigen as expected (Kaelin et al., 1991; data not shown).

The first peptide (Table 2, peptide wt), a 15-residue wild-type segment of SV40 T antigen containing the Rb-binding domain, competes strongly with intact, full-length radiolabeled T antigen for binding to the hybrid Rb/GST protein. The second peptide (Table 2, peptide K1) contains a single point mutation (Glu to Lys) known to inactivate Rb binding (DeCaprio et al., 1988) and is the weakest competitor in the series. Likewise, the other negative controls (Table 2, peptide X and BSA) show only weak competitor activity. The P2 peptide also fails to compete effectively for Rb binding (Table 2). Because the P2 peptide conforms to the pattern in Figure 1 and has appropriate hydrophobicity and charged-group profiles, these results imply that the prolines prevent the peptide from assuming the optimal Rb-binding conformation. Thus, the secondary structure propensity of a peptide may be a factor in determining its ability to bind Rb.

Model of secondary structure conformational propensities

If the secondary structure propensity of a peptide is an important factor that contributes to its Rb-binding specificity, then we would expect that secondary structural profiles for diverse Rb-binding sequences would exhibit common elements. To test this hypothesis, we used a new conformational filtering algorithm (Vajda, 1992) to generate profiles of statistically probable conformations for peptide sequences in our function set (Table 1A), as well as for other test sequences such as the P2 peptide (Table 2). In this method, conformational constraints are established based upon a statistical analysis of local interactions among first, second, and third nearest-neighbor residues in known protein structures (Vajda, 1992; see Materials and methods). This analysis results in the assignment of one or more allowed conformational states to each residue in the sequence. These assignments are given in terms of the 16 regions in the (ϕ, ψ) map of Zimmerman et al. (1977). The resulting conformational profile gives information about structures that are stabilized by local interactions, and reflects the intrinsic propensity of the sequence to form secondary structure. Thus, the method is particularly suitable for the analysis of short linear peptides that usually exist in solution as ensembles of rapidly interconverting conformers (Dyson & Wright, 1991), and it has been successfully employed to predict the NMR structures of a number of peptides (Vajda, 1992).

Secondary structure models were developed for all 21 high-affinity Rb-binding peptide sequences (Table 1A). The Rb-binding sequences show flexibility in conformation through the Rb-binding domain, but alpha-helix is the most probable conformation in this region for all 21. This result is illustrated in Figure 2, which gives upper and lower boundaries on the position-specific probability of alpha-helix (Zimmerman region A) for the group of 21 Rb-binding sequences. These boundaries are delineated by the solid black lines (results are given for the first 14 residues only). In the region of the conserved signature sequence (L X C X E) there is a window with at least 40% alpha-helical probability predicted for all 21 binding sequences. In contrast, the helical probability exhibits very large variability in the rest of the chain, changing from 0% to 100% depending upon the particular amino acid sequence. Thus, although these peptide sequences would all be expected to demonstrate flexibility in solution, our model predicts that they share a common intrinsic propensity to form alpha-helical structure in the signature sequence region.

In the next section (see below), we present experimental evidence that our conformational filtering technique gives reliable predictions of alpha-helical content when applied to peptides containing the Rb-binding signature sequence. Thus, we expect that the alpha-helical probability boundaries (solid black lines) in Figure 2 could be used as a discriminator for determining whether candidate sequences might bind Rb. As an initial test of this, the conformational profile of a proline-substituted T antigen sequence (NEENLPCPEEMPSSDDEATA) was also plotted on the graph (Fig. 2, dotted line). The alpha-helical profile for the proline-substituted sequence falls far below the 40% probability level in the signature sequence

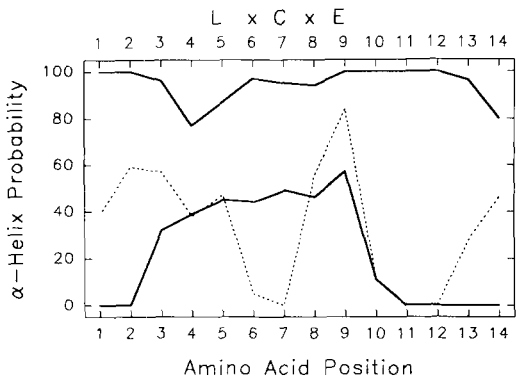

Fig. 2. Alpha-helical probability profile of 21 Rb-binding sequences and one nonbinding mutant. Position-specific probabilities for Zimmerman region A (alpha-helical) were calculated according to the conformational filter of Vajda (1992) (see Materials and methods) for each of the 21 high-affinity Rb-binding peptide sequences shown in Table 1A. The resulting probability profiles were plotted against amino acid position, and the lower and upper boundaries of the plots were determined for each position. Figure 2 depicts these lower and upper probability boundaries as solid black lines (results for only the first 14 residues are shown). These lines define a window of alpha-helical probabilities that are common to all of the binding sequences. The position of the conserved signature sequence in the binding domain (L X C X E) is given above the plot for orientation. In the region of the signature sequence, all 21 binding sequences are calculated to have at least 40% alpha-helical probability. The profile of a proline-substituted SV40 T antigen derivative sequence (NEENLPCPEEMPSSDDEATA) is plotted as a dotted line and falls significantly below the window in the region of the conserved signature sequence.

region. This difference in the conformational profile in the signature sequence region might, at least in part, be related to the reduced Rb-binding affinity of the P2 peptide (Table 2). The conformational properties of the prolines may reduce the probability that the peptide will assume an optimal conformation that is needed to interact with the Rb pocket.

CD spectroscopy of peptides containing Rb-binding domains: Comparison with conformational filtering results

To test the validity of our secondary structure filtering technique as applied to peptides containing an Rb-binding domain, we have modeled the secondary structure of three such peptides (Table 3) and have compared the predictions with the results of CD spectroscopy. We asked whether the filtering technique could be used to predict the overall alpha-helical content of a peptide as documented by its CD spectrum in an appropriate chemical environment. As noted above, the conformational filtering technique generates position-specific predictions of alpha-helical probability. To translate these data into the predicted net alpha-helical content of the peptide, we perform a Monte Carlo simulation in which 1 million possible conformations are randomly generated using the position-specific probability values for each of the allowed Zimmerman regions. Each generated conformation is analyzed for fractional alpha-helical content (an alpha-helix is defined as at least three contiguous residues in the Zimmerman region A state). The net alpha-helical character is then calculated by averaging the fractional alpha-helical content of all 1 million generated conformations. This method assumes that the peptide exists as an ensemble of rapidly interconverting conformers. The Monte Carlo method is a statistical approach to simulate such a population of conformers. The method further assumes that an alpha-helix is stabilized by intramolecular forces (hydrogen bonding and charged-group effects). The results of this analysis are then compared to the CD spectrum of the peptide in an appropriate chemical environment. The CD spectrum can be analyzed to determine the net alpha-helical content of the peptide (Manavalan & Johnson, 1987; Johnson, 1990).

Table 3. *Predicted alpha-helical content of three peptides containing Rb-binding domains: comparison with CD spectral analyses*[a]

Peptide sequence	Alpha-helical content (%)	
	Predicted	CD result
NLFCSEEMPSSDDE-amide	20.6	18
NAFNEENLFCSEEMPSSDDEATA-amide	40.8	40
acetyl-AETAAADLLCHEQLS-amide	89.1	82

[a] Probable alpha-helical regions and net alpha-helical content of three peptides containing an Rb-binding domain. The first two peptides are segments from SV40 T antigen. The third peptide contains an alpha-helical seed (AETAAA) followed by a segment from human papilloma virus strain 18 E7. The analysis is based upon a statistical treatment of local interactions among first, second, and third nearest-neighbor residues in known protein structures (Vajda, 1992; see Materials and methods). Residues calculated to have at least a 40% probability of being in an alpha-helical conformation (Zimmerman region A) are underlined. The overall net alpha-helical character of the peptide was predicted with a Monte Carlo simulation (see text). The CD analyses were performed in 90% 2,2,2-trifluoroethanol (see Breese et al., 1991; Figs. 3, 4). The first two peptides were studied at 4 degrees C and the third was studied at room temperature. Alpha-helical content was derived from an analysis of each CD spectrum (Manavalan & Johnson, 1987; Johnson, 1990).

For CD spectroscopy, the peptides are studied in 90% 2,2,2-trifluoroethanol (TFE), usually at 4 °C. These conditions are selected to maximally stabilize intramolecular hydrogen bonding and provide experimental conditions that are appropriate for testing our model, which is based on the assumption that short-range intramolecular forces will govern the peptide's structure. Low temperature and TFE stabilize intramolecular hydrogen bonding and promote secondary structure formation in short peptides (Brown & Klee, 1971; Nelson & Kallenbach, 1986, 1989; Dyson et al., 1988; Zhong & Johnson, 1992). Both alpha-helices (Brown & Klee, 1971; Nelson & Kallenbach, 1986) and beta structure (Lintner et al., 1977; Barrow & Zagorski, 1991) can be stabilized in peptides in aqueous TFE solutions. TFE appears to stabilize the structure for which

a peptide has intrinsic propensity (Zhong & Johnson, 1992).

The first peptide (Table 3) studied by the above approach is a 14-amino acid peptide amide containing the Rb-binding domain of SV40 T antigen (NLFCSEEMPS SDDE-amide). As shown in Table 3, this peptide is predicted to be 20.6% alpha-helical when stabilized by local intramolecular forces. The peptide was previously studied in 90% TFE at 4 °C (Breese et al., 1991) and was found to be 18% alpha-helical, in good agreement with the predicted value.

The second peptide (Table 3) is a 23-amino acid segment of SV40 T antigen that encompasses the first peptide sequence (NAFNEENLFCSEEMPSSDDEATA-amide). It was expected that the 23-residue peptide would exhibit increased alpha-helical character for two reasons. First, there are now two glutamic acid residues in the amino-terminal region that might stabilize the helical macrodipole (Collawn & Paterson, 1990) of a potential alpha-helical region encompassing the signature sequence LXCXE. Second, there are two alanine residues in the carboxy-terminus that might stabilize an alpha-helical structure in that region. The Monte Carlo simulation for this peptide also predicts an increased alpha-helical content of 40.8%.

To test the above theoretical considerations, the 23-amino acid peptide was synthesized as a carboxy-terminal amide and was then purified to homogeneity by preparative reverse-phase high performance liquid chromatography (HPLC). The primary structure was confirmed by automated gas-phase sequencing on a Porton 2090E system, and by plasma desorption mass spectrometry. The latter revealed a single positive quasimolecular ion at m/z 2,551.7 ± 1.0. The calculated isotopically averaged mass of the $[M + H]^+$ species for this peptide is 2,550.7.

The CD spectrum was collected at 4 °C in 90% TFE from 290.0 to 182.4 nm at 0.2-nm decrements and is shown in Figure 3. The spectrum was analyzed with the variable selection method of Manavalan and Johnson (1987) to obtain estimates of each type of secondary structure present in the sample. The analysis yielded: 40% alpha-helix, 15% anti-parallel beta-sheet, 26% turn, and 20% other. Thus, the alpha-helical content is in excellent agreement with that predicted by our model.

As a further test of our conformational filter model, we sought to design a peptide containing an Rb-binding domain that would exhibit a very high alpha-helical content even at room temperature. Our modeling results with the sequences in Table 1A suggested that the papilloma virus strain 18 E7 sequence had the strongest alpha-helical propensity through the Rb-binding domain. Accordingly, we designed a sequence with a potent helical seed, AETAAA (Mitchinson & Baldwin, 1986), linked to a segment of the papilloma virus 18 E7 protein sequence. This yielded the following 15-residue peptide design (Table 3, peptide 3): acetyl-AETAAADLLCHEQLS-amide. The

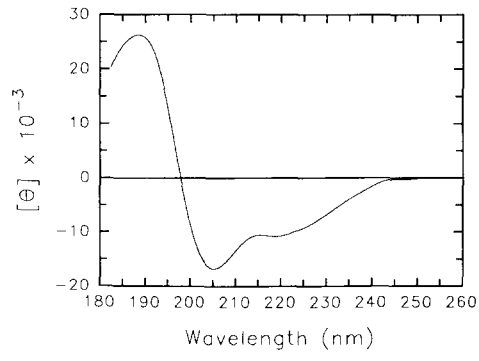

Fig. 3. CD spectrum of a peptide segment of T antigen. The CD spectrum of the peptide NAFNEENLFCSEEMPSSDDEATA-amide (corresponding to simian virus 40 [SV40] large T antigen residues 96–118) was recorded at 4 °C in aqueous 90% (v/v) 2,2,2-trifluoroethanol (TFE). The peptide concentration was 0.083 mg/mL. Data are displayed from 260.0 nm to 182.4 nm and are expressed as mean residue ellipticity, $[\theta]$ (deg cm^2 dmol^{-1}), calculated using 22 peptide amide bonds. The secondary structure composition of the peptide under these conditions was estimated with the variable selection program of Manavalan and Johnson (1987), which gave the following results (±SEM): 40 ± 5% alpha-helix, 15 ± 4% antiparallel beta-sheet, 26 ± 0.4% turn and 20 ± 0.8% other.

net helical content of this peptide was predicted to be 89.1% using the Monte Carlo technique.

To determine whether the designed 15-mer would have a high helical propensity as predicted, the peptide was synthesized and purified to homogeneity. The primary structure was confirmed with quantitative amino acid analysis and with plasma desorption mass spectrometry (sequencing was not attempted because of the blocked N-terminus). The mass spectrometric analysis revealed positive quasimolecular ions at m/z 1,614.3 ± 1.0 and 1,636.6 ± 1.0. These correspond to the $[M + H]^+$ and $[M + Na]^+$ species, which have expected isotopically averaged masses of 1,613.8 and 1,635.8, respectively.

The CD spectrum of the 15-mer was obtained at room temperature in 90% TFE from 290.0 nm to 186.0 nm and is shown in Figure 4. The spectrum shows diagnostic features of an alpha-helix including double minima near 208 nm and 222 nm, and a maximum near 191 nm. Analysis of the spectrum with the variable selection program of Manavalan and Johnson (1987) yielded an alpha-helical content of 82%, in close agreement with the predicted value. To determine whether peptide aggregation might be a factor in the stabilization of the alpha-helix, the CD spectrum was recorded with peptide concentrations of 0.0046 mg/mL (2.8 μM), 0.046 mg/mL (28 μM), and 0.46 mg/mL (280 μM) in 90% TFE. The mean residue ellipticity was determined at 222 nm for each concentration (−28.4, −30.6, and −30.0 [10^3 deg cm^2 dmol^{-1}], respectively) and was plotted as a function of the logarithm of the micromolar concentration. The slope (±SE) of the regression line (−0.80 ± 0.81) is not significantly different

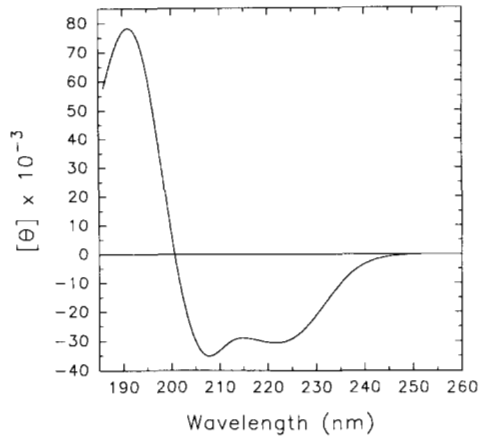

Fig. 4. CD spectrum of a helical seed sequence linked to the Rb-binding domain of the human papilloma virus strain 18 E7 protein. The CD spectrum of the peptide acetyl-AETAAADLLCHEQLS-amide was recorded at room temperature in aqueous 90% (v/v) TFE. The peptide concentration was 0.046 mg/mL. Data are displayed from 260.0 nm to 186.0 nm and are expressed as mean residue ellipticity, $[\theta]$ (deg cm^2 dmol^{-1}), calculated using 14 peptide amide bonds. The peptide contains the helical seed sequence, AETAAA (Mitchinson & Baldwin, 1986), linked to the Rb-binding domain of the human papilloma virus strain 18 E7 protein. The alpha-helical composition of the peptide under these conditions was estimated with the variable selection program of Manavalan and Johnson (1987), which gave a value (±SEM) of 81.5 ± 0.7% alpha-helix.

from zero. Thus, there is no evidence for a systematic concentration-dependent aggregation effect.

Discussion

In this report we have developed a diagnostic pattern for a high-affinity Rb-binding domain. This pattern has both high sensitivity and specificity. The pattern can effectively identify cellular Rb-binding proteins such as RBP-1 and RBP-2, which are known to bind with high affinity (Defeo-Jones et al., 1991). Thus, the pattern should serve as a useful tool for the identification of other high-affinity binding proteins (e.g., by screening predicted protein sequences from cloned genomic DNA segments). In addition, the pattern can effectively discriminate between human papilloma virus strains with high and low Rb-binding affinity. This has significant implications in terms of the oncogenicity of those viruses (Heck et al., 1992). The pattern can also be used to identify weak Rb-binding proteins if one allows partial matches. The nononcogenic E7 proteins from papilloma virus strains 6b and 11 have one mismatch. A mutant RBP-1 protein with a Cys to Ala mutation in the binding domain still binds Rb, although weakly (Defeo-Jones et al., 1991). The product of the c-myc gene also binds Rb weakly through an amino-terminal domain (Rustgi et al., 1991). The pattern has partial matches to regions in the amino-terminus of c-myc, although the exact sequences responsible for binding await experimental determination. The pattern also has a partial match to the Rb-binding sequence of E2F (L X E X E), another Rb-binding protein, but the significance of this partial match also awaits further experimental investigation.

When one reviews Table 1A and the RBP-2 sequence, it can be seen that there is no obvious conservation of residues in the positions flanking the conserved Cys. For example, the residue just amino-terminal to the Cys can be Phe, Tyr, Leu, Thr, or Val. The residue following the Cys can be Ser, Tyr, His, or Asp (the latter is in RBP-2). There are no obvious patterns that would explain these allowed substitutions. However, in this report, we propose that the common feature is that these residues all result in a similar alpha-helical propensity profile (Fig. 2), while at the same time they maintain the proper hydrophobic and electrostatic environment (Fig. 1). Our data derived from binding studies with a proline-substituted peptide (Table 2) support the idea that secondary structural propensity is important for the ability of a peptide to bind Rb. For example, the ability to assume a certain conformation might be important for a peptide's initial interaction with the Rb-binding pocket. The prolines, due to their unique conformational properties, might prevent the peptide from assuming its active binding conformation.

To model the secondary structure propensity in the binding domain, we have devised a new approach that incorporates a statistical filtering tool (Vajda, 1992). The statistical filtering algorithm is based on nearest-neighbor interactions extracted from 43 small proteins with known X-ray structures. In short peptides, such as the ones employed here, structural conformations are largely dictated by short-range interactions as well as by interactions with the environment (e.g., the solvent) (Dyson & Wright, 1991; Zhong & Johnson, 1992). Thus, our statistical filtering approach is particularly suitable for the analysis of short peptides. The method correctly identifies the state of 79% of residues in the set of 43 small proteins when all conformational states of 5% or higher probability are included. In addition, it correctly predicts the conformation of a set of peptides with known NMR structures (Vajda, 1992). We have presented data in this report (Table 3) showing that the method is also useful in predicting the alpha-helical content of Rb-binding peptide sequences in an appropriate chemical environment, and the results compare favorably with those determined by CD spectroscopy (see below).

The result of the statistical filtering algorithm sets boundaries on the alpha-helical probabilities for residues in the Rb-binding domain (Fig. 2). The profile of the proline-substituted peptide falls significantly outside of the boundaries, as expected. This suggests that the conformational properties of the prolines prevent that peptide from effectively interacting with Rb.

We have further tested the validity of the statistical filter by synthesizing a panel of three peptide amides con-

taining the Rb-binding domain of SV40 T antigen or of the papilloma virus E7 protein. Using CD spectroscopy, we have shown that these peptides can assume different levels of alpha-helical structure in TFE, ranging from approximately 20% to 80%. These results are in accord with the prediction of our model (see Table 3). The model predictions are based on a Monte Carlo method that translates position-specific probabilities into net alpha-helical content by generating a large number of probable conformations. TFE has been used to stabilize other alpha-helix forming peptides (Brown & Klee, 1971; Dyson et al., 1988; Nelson & Kallenbach, 1986, 1989). TFE is a hydrophilic solvent that tends to stabilize intramolecular hydrogen bonding and stabilizes the secondary structure for which a sequence has propensity (Zhong & Johnson, 1992).

It is of course not known whether an Rb-binding peptide's active structure is induced upon interaction with Rb, is a precondition, or is a combination. We propose that the peptide's Rb-binding domain must be in a certain conformation during effective interaction with Rb. This conformation could serve to orient the conserved Leu, Cys, and Glu residues into the proper three-dimensional alignment that is required to bind Rb. Prolines may render it improbable to achieve the appropriate conformation. Our statistical conformational profile suggests that the active binding conformation has an element of alpha-helix. We plan to test this model in future studies by using the theoretical modeling approach presented in this paper to design a larger series of peptides with differing alpha-helical propensities. These can be studied to determine their equilibrium binding constants.

Materials and methods

Peptide synthesis

A peptide amide (NAFNEENLFCSEEMPSSDDEATA-amide) corresponding to SV40 large T antigen residues 96–118 was synthesized with solid-phase methods and purified to homogeneity by preparative reverse-phase (RP)-HPLC. The synthesis was performed by Multiple Peptide Systems (San Diego, California). Homogeneity was verified by analytical RP-HPLC using a linear gradient of 5–60% acetonitrile in water containing 0.05% trifluoroacetic acid on a Vydac C-18 column. Absorbance was monitored at 215 nm. The primary structure was confirmed by automated gas-phase sequencing on a Porton 2090E system and by plasma desorption mass spectrometry (BIOION 20 Analyzer) using a 252-Cf source (Sundqvist et al., 1984). Peptides used for competitive binding assays were synthesized in a similar fashion by Research Genetics, Huntsville, Alabama, and were confirmed by automated gas-phase sequencing. A hybrid peptide containing an alpha-helical seed sequence linked to the Rb-binding domain of the human papilloma virus strain 18 E7 protein (acetyl-AETAAADLLCHEQLS-amide) was synthesized and purified to homogeneity by Dr. Richard J. Ridge at Woods Hole, Massachusetts. This peptide was confirmed by quantitative amino acid analysis and plasma desorption mass spectrometry. All peptides were stored under argon prior to use to avoid oxidation.

CD spectroscopy

CD spectra were recorded on a Jasco J-720 spectropolarimeter using 0.20-mm to 2.0-mm pathlength cells. The cell was mounted on a temperature-controlled block. Peptides were dissolved in 90% (v/v) TFE (NMR grade; Aldrich) and 10% water. Peptide concentrations were determined by quantitative amino acid analyses. The instrument was calibrated with D-10 camphorsulfonic acid. CD data were recorded within the wavelength range of 290.0 nm to 180.0 nm at 0.2-nm decrements with a scan speed of 20 nm/min, a 1.0-nm band width, and an averaging time of 0.5 s. Scans were terminated when the voltage exceeded 550 V. Five scans of each sample and solvent were taken at a specified temperature, and the averaged, baseline-corrected spectra were smoothed with a Savitzky-Golay (Savitzky & Golay, 1964) filter. Data are expressed as mean residue ellipticity, $[\theta]$ (deg cm^2 dmol^{-1}), calculated using the number of peptide amide bonds. Spectra were analyzed with the variable selection program of Manavalan and Johnson (1987) as described by Johnson (1990).

Competitive Rb-binding assay

The hybrid Rb/GST protein, encoded by the plasmid pGEX2T-RB(379–792) (Kaelin et al., 1991), was produced in bacteria and harvested with glutathione/Sepharose 4B beads as previously described (Kaelin et al., 1991). The Rb/GST bead complexes were then incubated with ^{35}S-labeled extracts of SV40-transformed human lung fibroblast cells (AG2804C) that express full-length SV40 T antigen. The incubation was carried out either in the presence or absence of cold competing peptides or BSA at 4 °C for 1 h in EBC buffer (50 mM Tris-HCl [pH 8.0], 120 mM NaCl, 0.5% NonidetP-40, aprotinin [10 μg/mL], leupeptin [10 μg/mL], phenylmethylsulfonylfluoride [10 μg/mL]). The binding of radiolabeled T antigen to the immobilized Rb/GST protein was monitored by eluting bound proteins off the beads, followed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and autoradiography. The relative amount of labeled T antigen bound was determined by densitometry and was standardized against the level bound in the presence of 1.0 mM BSA, which was set to a level of 1,000.

Conformational filtering

Conformational filtering is based on dividing the (ϕ, ψ) map of each residue into the 16 regions (Zimmerman

et al., 1977) and using these regions as conformational states. The number of candidate states for each residue of a sequence is reduced by a series of conformational filters. In each filter we use a well-defined set of assumptions to estimate the probabilities of conformational states and eliminate a state if its estimated probability is below an acceptance probability p_α. The resulting state assignment is not necessarily unique but provides information that can be further exploited in searches for the detailed tertiary structure. The method is particularly well suited for the analysis of short linear peptides. The structure of such peptides is predominantly determined by local interactions, and they usually exist in solution as ensembles of rapidly interconverting conformers (Dyson & Wright, 1991). Retaining all states with nonnegligible conformational probabilities, filtering generates an entire ensemble of conformers and provides information on the expected structural variability of peptide segments. If there are dominant conformational states for all residues of a segment, then the structure is well defined by local interactions alone and is likely to be preserved when the segment is isolated from the rest of the chain. Conversely, a segment with large structural variability may adopt a variety of conformations in solution or in a protein environment.

Let p_i^m, $i = 1, \ldots, 16$ denote the probabilities of the 16 conformational states for the mth residue. The simplest filters are based on the use of forward and backward Markov models in which the probability of a state for a residue depends only on the state of the nearest neighbors of that residue. Extracting doublet data from the X-ray structures of 43 small proteins (Lambert & Scheraga, 1989), we established transitional probabilities for the 400 residue pairs (Vajda, 1992). For example, $p(\mathrm{Ala}_i^{m+1} | \mathrm{Gly}_j^m)$ denotes the normalized frequency of an Ala residue in state i when it is preceded by a Gly residue in state j. More generally, we compile 400, 16×16 forward Markov matrices with elements of the form $T_{ij}^m = p(R_i^{m+1} | R_j^m)$. The same information is arranged in form of backward Markov matrices $\bar{T}_{ij}^m = p(R_i^{m-1} | R_j^m)$. These matrices are the input data for predicting the probability distributions for the residues of a particular sequence by repeated application of the forward and backward transition formulas $\mathbf{p}^{m+1} = \mathbf{T}^m \mathbf{p}^m$ and $\mathbf{p}^{m-1} = \bar{\mathbf{T}}^m \mathbf{p}^m$ to obtain forward and backward probabilities for each residue. The starting probability $\mathbf{p}^1$ is an eigenvector of the matrix $\bar{\mathbf{T}}_2 \bar{\mathbf{T}}_2 \ldots \bar{\mathbf{T}}_n \mathbf{T}_{n-1} \ldots \mathbf{T}_2 \mathbf{T}_1$. Notice that in spite of using only nearest-neighbor interactions, changes in probabilities of any residue progress along the entire chain. Further filters were implemented by generating and selecting acceptable conformations for overlapping tripeptide and tetrapeptide fragments.

Conformational constraints for each residue are formed by rejecting any state if its estimated probability is below the acceptance level $p_\alpha = 5\%$ in any of the filters. While the prediction is rarely unique (i.e., several states are retained), the conformational degrees of freedom are reduced by a factor of two on the average for each residue. The method has been tested in a cross-validation study on the 43 proteins, in each step deleting the protein to be predicted from the learning set. At $p_\alpha = 5\%$ the prediction is correct for 79% of all residues.

Acknowledgments

J.F. is a Lucille P. Markey Scholar. This research was supported in part by a grant from the Lucille P. Markey Charitable Trust to J.F. S.V. was supported in part by NIAID grant R01-AI30535 (awarded to C. DeLisi). The pGEX2T-RB(379-792) and pGEX2T-RB(379-792; DL 573-645) constructs were a gift from W.G. Kaelin, Jr. (Harvard University) to T.F. We thank L. Hof (Albany) for use of a densitometer and David S. Eisenberg (UCLA) for reading the manuscript prior to submission.

References

Bandara, L.R., Adamczewski, J.P., Hunt, T., & La Thangue, N.B. (1991). Cyclin A and the retinoblastoma gene product complex with a common transcription factor. *Nature 352*, 249-251.

Bandara, L.R. & La Thangue, N.B. (1991). Adenovirus E1a prevents the retinoblastoma gene product from complexing with a cellular transcription factor. *Nature 351*, 494-497.

Barbosa, M.S., Edmonds, C., Fisher, C., Schiller, J.T., Lowy, D.R., & Vousden, K.H. (1990). The region of the HPV E7 oncoprotein homologous to adenovirus E1a and SV40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. *EMBO J. 9*, 153-160.

Barrow, C.J. & Zagorski, M.G. (1991). Solution structures of beta-peptide and its constituent fragments: Relation to amyloid deposition. *Science 253*, 179-182.

Breese, K., Friedrich, T., Andersen, T.T., Smith, T.F., & Figge, J. (1991). Structural characterization of a 14-residue peptide ligand of the retinoblastoma protein: Comparison with a nonbinding analog. *Peptide Res. 4*, 220-226.

Brown, J.E. & Klee, W.A. (1971). Helix-coil transition of the isolated amino terminus of ribonuclease. *Biochemistry 10*, 470-476.

Chellappan, S.P., Hiebert, S., Mudryj, M., Horowitz, J.M., & Nevins, J.R. (1991). The E2F transcription factor is a cellular target for the RB protein. *Cell 65*, 1053-1061.

Chittenden, T., Livingston, D.M., & Kaelin, W.G., Jr. (1991). The T/E1A-binding domain of the retinoblastoma product can interact selectively with a sequence-specific DNA-binding protein. *Cell 65*, 1073-1082.

Collawn, J.F. & Paterson, Y. (1990). Stabilization of helical structure in two 17-residue amphipathic analogues of the C-terminal peptide of cytochrome C. *Biopolymers 29*, 1289-1296.

DeCaprio, J.A., Ludlow, J.W., Figge, J., Shew, J.-Y., Huang, C.-M., Lee, W.-H., Marsilio, E., Paucha, E., & Livingston, D.M. (1988). SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene. *Cell 54*, 275-283.

Defeo-Jones, D., Huang, P.S., Jones, R.E., Haskell, K.M., Vuocolo, G.A., Hanobik, M.G., Huber, H.E., & Oliff, A. (1991). Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene product. *Nature 352*, 251-254.

Dyson, H.J., Rance, M., Houghten, R.A., Wright, P.E., & Lerner, R.A. (1988). Folding of immunogenic peptide fragments of proteins in water solution. II. The nascent helix. *J. Mol. Biol. 201*, 201-217.

Dyson, H.J. & Wright, P.E. (1991). Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Biophys. Chem. 20*, 519-538.

Dyson, N., Bernards, R., Friend, S.H., Gooding, L.R., Hassell, J.A., Major, E.O., Pipas, J.M., Vandyke, T., & Harlow, E. (1990). Large T antigens of many polyomaviruses are able to form complexes with the retinoblastoma protein. *J. Virol. 64*, 1353-1356.

Dyson, N., Howley, P.M., Munger, K., & Harlow, E. (1989). The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science 243*, 934-937.

Figge, J., Webster, T., Smith, T.F., & Paucha, E. (1988). Prediction of similar transforming regions in simian virus 40 large T, adenovirus E1A, and myc oncoproteins. *J. Virol. 62*, 1814–1818.

Firzlaff, J.M., Luscher, B., & Eisenman, R.N. (1991). Negative charge at the casein kinase II phosphorylation site is important for transformation but not for Rb binding by the E7 protein of human papillomavirus type 16. *Proc. Natl. Acad. Sci. USA 88*, 5187–5191.

Gage, J.R., Meyers, C., & Wettstein, F.O. (1990). The E7 proteins of the nononcogenic human papillomavirus type 6b (HPV-6b) and of the oncogenic HPV-16 differ in retinoblastoma protein binding and other properties. *J. Virol. 64*, 723–730.

Goodrich, D.W., Wang, N.P., Qian, Y.-W., Lee, E.Y.-H.P., & Lee, W.-H. (1991). The retinoblastoma gene product regulates progression through the G1 phase of the cell cycle. *Cell 67*, 293–302.

Heck, D.V., Yee, C.L., Howley, P.M., & Munger, K. (1992). Efficiency of binding the retinoblastoma protein correlates with the transforming capacity of the E7 oncoproteins of the human papillomaviruses. *Proc. Natl. Acad. Sci. USA 89*, 4442–4446.

Helin, K., Lees, J.A., Vidal, M., Dyson, N., Harlow, E., & Fattaey, A. (1992). A cDNA encoding a pRB-binding protein with properties of the transcription factor E2F. *Cell 70*, 337–350.

Howley, P.M. (1991). Role of the human papillomaviruses in human cancer. *Cancer Res. (Suppl.) 51*, 5019s–5022s.

Huang, S., Lee, W.-H., & Lee, E.Y.-H.P. (1991). A cellular protein that competes with SV40 T antigen for binding to the retinoblastoma gene product. *Nature 350*, 160–162.

Johnson, W.C., Jr. (1990). Protein secondary structure and circular dichroism: A practical guide. *Proteins Struct. Funct. Genet. 7*, 205–214.

Jones, R.E., Wegrzyn, R.J., Patrick, D.R., Balishin, N.L., Vuocolo, G.A., Riemen, M.W., Defeo-Jones, D., Garsky, V.M., Heimbrook, D.C., & Oliff, A. (1990). Identification of HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein. *J. Biol. Chem. 265*, 12782–12785.

Kaelin, W.G., Jr., Ewen, M.E., & Livingston, D.M. (1990). Definition of the minimal simian virus 40 large T antigen- and adenovirus E1A-binding domain in the retinoblastoma gene product. *Mol. Cell. Biol. 10*, 3761–3769.

Kaelin, W.G., Jr., Krek, W., Sellers, W.R., DeCaprio, J.A., Ajchenbaum, F., Fuchs, C.S., Chittenden, T., Li, Y., Farnham, P.J., Blanar, M.A., Livingston, D.M., & Flemington, E.K. (1992). Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties. *Cell 70*, 351–364.

Kaelin, W.G., Jr., Pallas, D.C., DeCaprio, J.A., Kaye, F.J., & Livingston, D.M. (1991). Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product. *Cell 64*, 521–532.

Lambert, M.H. & Scheraga, H.A. (1989). Pattern recognition in the prediction of protein structure. I. Tripeptide conformational probabilities calculated from the amino acid sequence. *J. Comp. Chem. 10*, 770–797.

Lathrop, R.H., Webster, T.A., Smith, T.F., & Winston, P.H. (1990). ARIEL: A massively parallel symbolic learning assistant for protein structure/function. In *Artificial Intelligence at MIT: Expanding Frontiers* (Winston, P.H. & Shellard, S.A., Eds.), pp. 70–103. MIT Press, Cambridge, Massachusetts.

Lintner, K., Fermandjian, S., Fromageot, P., Khosla, M.C., Smeby, R.R., & Bumpus, F.M. (1977). Circular dichroism studies of angiotensin II and analogues: Effects of primary sequence, solvent, and pH on the side-chain conformation. *Biochemistry 16*, 806–812.

Manavalan, P. & Johnson, W.C., Jr. (1987). Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem. 167*, 76–85.

Mitchinson, C. & Baldwin, R.L. (1986). The design and production of semisynthetic ribonucleases with increased thermostability by incorporation of S-peptide analogues with enhanced helical stability. *Proteins Struct. Funct. Genet. 1*, 23–33.

Moran, E. (1988). A region of SV40 large T can substitute for a transforming domain of the adenovirus E1A products. *Nature 334*, 168–170.

Munger, K., Werness B.A., Dyson, N., Phelps, W.C., Harlow, E., & Howley, P.M. (1989). Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *EMBO J. 8*, 4099–4105.

Nelson, J.W. & Kallenbach, N.R. (1986). Stabilization of the ribonuclease S-peptide alpha-helix by trifluoroethanol. *Proteins Struct. Funct. Genet. 1*, 211–217.

Nelson, J.W. & Kallenbach, N.R. (1989). Persistence of the alpha-helix stop signal in the S-peptide in trifluoroethanol solutions. *Biochemistry 28*, 5256–5261.

Rees, D.C., DeAntonio, L., & Eisenberg, D. (1989) Hydrophobic organization of membrane proteins. *Science 245*, 510–513.

Rustgi, A.K., Dyson, N., & Bernards, R. (1991). Amino-terminal domains of c-myc and N-myc proteins mediate binding to the retinoblastoma gene product. *Nature 352*, 541–544.

Savitzky, A. & Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem. 36*, 1627–1639.

Stabel, S., Argos, P., & Philipson L. (1985). The release of growth arrest by microinjection of adenovirus E1A DNA. *EMBO J. 4*, 2329–2336.

Sundqvist, B., Roepstorff, P., Fohlman, J., Hedin, A., Hakansson, P., Kamensky, I., Lindberg, M., Salehpour, M., & Sawe, G. (1984). Molecular weight determinations of proteins by californium plasma desorption mass spectrometry. *Science 226*, 696–698.

Vajda, S. (1992). Conformational filtering in polypeptides and proteins. *J. Mol. Biol.*, in press.

Weinberg, R.A. (1991). Tumor suppressor genes. *Science 254*, 1138–1146.

Whyte, P., Buchkovich, K.J., Horowitz, J.M., Friend, S.H., Raybuck, M., Weinberg, R.A., & Harlow, E. (1988). Association between an oncogene and an anti-oncogene: The adenovirus E1A proteins bind to the retinoblastoma gene product. *Nature 334*, 124–129.

Whyte, P., Williamson, N.M., & Harlow, E. (1989). Cellular targets for transformation by the adenovirus E1A proteins. *Cell 56*, 67–75.

Zhong, L. & Johnson, W.C., Jr. (1992). Environment affects amino acid preference for secondary structure. *Proc. Natl. Acad. Sci. USA 89*, 4462–4465.

Zimmerman, S.S., Pottle, M.S., Nemethy, G., & Scheraga, H.A. (1977). Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP. *Macromolecules 10*, 1–9.

Zhu, Q., Smith, T.F., Lathrop, R.H., & Figge, J. (1990). Acid helix-turn activator motif. *Proteins Struct. Funct. Genet. 8*, 156–163.