

FIG. 2 A physical map of the GCPS/*GLI3* region. The translocation breakpoints of the three GCPS patients are shown on the top in relation to the *GLI3* cDNA sequence. The position of the PCR probes used is indicated by black boxes. A restriction map derived from genomic phage clones isolated by screening with the PCR probes or phage end fragments is drawn on a larger scale (E, *EcoRI*; S, *SalI*). The location of the RoH breakpoint spanned by

Gli_{dist11} is indicated by an asterisk.

METHODS. Overlapping phage clones were isolated from a genomic *Mbol* partial digest library in EMBL3 using the PCR probes described in Fig. 1. Clone *Gli_{dist11}* was isolated by chromosome walking with the 6-kp *EcoRI/SalI* end-fragment of phage *Gli_{dist3}* subcloned into pBluescript II SK-(Stratagene). All subcloning and hybridization was according to standard procedures¹⁸.

gene is expressed in tissues such as testes, myometrium, placenta, colon and lung as an 8.5-kilobase messenger RNA and the protein product (relative molecular mass, 190,000) shows sequence-specific DNA binding^{5,6}. On our panel this probe, pcrGli3, mapped distal to the Greig translocation breakpoints in patients 1863 (46, XY, t(3; 7) (p21.1; p13)) and IT (46, XX, t(6; 7)(q12; p13)), but proximal to the one in RoH (46, XX, t(6; 7)(q27; p13)) (Fig. 1d). Thus, *GLI3* maps very closely or may even be identical to the GCPS gene.

To define the position of the three translocation breakpoints along the *GLI3* gene we used pcrGli_{prox} and pcrGli_{dist}, sequences corresponding to proximal and distal segments of the *GLI3* complementary DNA sequence⁶ (Fig. 2), amplified by PCR from lung cDNA. PcrGli_{prox} maps proximal to all GCPS translocation breakpoints (Fig. 1c), indicating that two of the translocations split the *GLI3* gene within the first third of the coding region, thereby precluding the formation of a normal functioning protein. This also establishes the transcriptional orientation of *GLI3* from centromere to telomere. Probes from a 30-kb genomic phage contig selected by pcrGli_{prox} did not detect the translocation breakpoints, suggesting that the 5' part of this gene is spread over a large genomic distance (Fig. 2).

On the other hand, pcrGli_{dist}, like pcrGli3, maps proximal to the RoH breakpoint and distal to the 1863 and IT breakpoints (Fig. 1e). The third translocation breakpoint thus has to be downstream of the 3' end of *GLI3* (Fig. 2). Analysis of five independent cDNA clones isolated from a human fetal kidney library¹⁰ using pcrGli_{dist} as probe showed no evidence for alternatively spliced transcripts in this region, which might extend to the RoH translocation breakpoint. Thus, the translocation probably falls distal to the last exon of *GLI3*.

To map the distal translocation breakpoint, we isolated a 50-kb phage contig, including 15 kb of DNA downstream of the *GLI3* coding sequence. The RoH translocation breakpoint was found by hybrid mapping in a 3.5-kb *EcoRI/SalI* fragment about 10 kb downstream of *GLI3* (Figs 1f, and 2). In this case a *cis*-acting element might have been brought into the neighbourhood of *GLI3*, thereby modifying its expression. The influence of 3'-acting domains has been well documented in variant Burkitt

lymphomas, in which translocations of more than 10 kb downstream of the *c-myc* gene deregulate its expression¹¹.

Our results support the assumption that mutations affecting only one allele of the zinc finger gene *GLI3* cause the GCPS syndrome. With the potential function as a transcriptional regulator, *GLI3* could well control limb and craniofacial development, perhaps through involvement in the generation of a morphogen gradient, as proposed for vertebrate limb formation¹². The fact that *GLI3* is widely expressed in adult tissues, with only some of these being phenotypically altered by heterozygous mutation, implies that precise control of *GLI3* expression is crucial only at certain developmental stages. Studies of mouse mutant extra toes (Xt)¹³, and anterior digit deformity (add)¹⁴, which are presumed to have similarities with GCPS, may reveal more about the function of *GLI3*. Various finger proteins, among them those from the *GLI-Krüppel* gene family, have key roles in *Drosophila* and *Xenopus* development. Our results provide an example of a zinc-finger gene that is mutated in a hereditary human malformation. □

Received 23 April; accepted 21 June 1991.

- Greig, D. M. *Edin. med. J.* **33**, 189–218 (1926).
- Gollop, T. R. & Fontes, L. R. *Am. J. med. Genet.* **22**, 59–68 (1985).
- Tommerup, N. & Nielsen, F. *Am. J. med. Genet.* **16**, 313–321 (1983).
- Krüger, G. et al. *Am. J. med. Genet.* **32**, 411–416 (1989).
- Ruppert, J. M. et al. *Molec. cell. Biol.* **8**, 3104–3113 (1988).
- Ruppert, J. M., Vogelstein, B., Arheden, K. & Kinzler, K. W. *Molec. cell. Biol.* **10**, 5408–5415 (1990).
- Wagner, K., Kroisel, P. M. & Rosenkranz, W. *Genomics* **8**, 487–491 (1990).
- Brueton, L., Huson, S. M., Winter, R. M. & Williamson, R. *Am. J. med. Genet.* **31**, 799–804 (1988).
- Drabkin, H. et al. *Genomics* **4**, 518–529 (1989).
- Gessler, M. et al. *Nature* **343**, 774–778 (1990).
- Bornkamm, G. W., Pollack, A. & Eick, D. in *Cellular Oncogene Activation* (ed. Klein, G.) 223–273 (Dekker, New York, 1988).
- Eichele, G. *Trends Genet.* **5**, 246–251 (1989).
- Winter, R. & Huson, M. *Am. J. med. Genet.* **31**, 793–798 (1988).
- Pohl, T. M., Mattei, M.-G. & Rüther, U. *Development* **110**, 1153–1157 (1990).
- Rommens, J. M. et al. *Am. J. hum. Genet.* **43**, 645–633 (1988).
- Jobs, A. et al. *Hum. Genet.* **84**, 147–150 (1990).
- Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995 (1984).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Vortkamp, A. et al. *Genomics* (in the press).

ACKNOWLEDGEMENTS. We thank N. Tommerup and L. Pelz for fibroblast cell lines 1863 and RoH, C. Croce for hybrid 5387-3c10 and S. Orkin for an aliquot of the EMBL3 library. This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

Amino-terminal domains of *c-myc* and *N-myc* proteins mediate binding to the retinoblastoma gene product

Anil K. Rustgi[‡], Nicholas Dyson[†] & Rene Bernards^{*}

Divisions of^{*} Molecular Genetics and [†] Molecular Oncology, The Cancer Center of the Massachusetts General Hospital and Harvard Medical School, and [‡]Gastrointestinal Unit of Massachusetts General Hospital, 149 13th Street, Charlestown, Massachusetts, USA

THE proteins encoded by the *myc* gene family are involved in the control of cell proliferation and differentiation, and aberrant expression of *myc* proteins has been implicated in the genesis of a variety of neoplasms¹. In the carboxyl terminus, *myc* proteins have two domains that encode a basic domain/helix-loop-helix and a leucine zipper motif, respectively. These motifs are involved both in DNA binding and in protein dimerization²⁻⁵. In addition, *myc* protein family members share several regions of highly conserved amino acids in their amino termini that are essential for transformation^{6,7}. We report here that an N-terminal domain present in both the *c-myc* and *N-myc* proteins mediates binding to the retinoblastoma gene product, pRb. We show that the human papilloma virus E7 protein competes with *c-myc* for binding to pRb, indicating that these proteins share overlapping binding sites on pRb. Furthermore, a mutant Rb protein from a human tumour cell line that carried a 35-amino-acid deletion in its C terminus failed to bind to *c-myc*. Our results suggest that *c-myc* and pRb cooperate through direct binding to control cell proliferation.

The study of proteins that interact with *c-myc* has been hindered by the inherent insolubility of *myc* proteins. Conditions that allow extraction of *c-myc* from the nucleus are too stringent to preserve protein-protein interactions, whereas mild extraction conditions fail to solubilize *c-myc*⁸. To circumvent this problem, we used a bacterial expression vector in which the glutathione S-transferase (GST) gene is linked to a DNA fragment that encodes the first 204 amino acids of human *c-myc*. The encoded GSTΔ*myc* fusion protein is highly soluble and can be readily purified from bacterial lysates by binding to glutathione agarose beads (ref. 9, and D. Hill, unpublished observations).

To investigate whether cellular proteins form specific complexes with any of the conserved N-terminal domains of the *c-myc* protein, we labelled various human cell lines with ³⁵S-methionine and prepared mild-detergent lysates from labelled cells. These extracts were then incubated with either GST protein bound to agarose beads (GST beads) or with GST-*myc* chimaeric protein bound to agarose beads (GSTΔ*myc* beads) as described¹⁰. Cellular proteins bound by both types of beads were analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). These experiments indicated that a small number of cellular proteins were bound by GSTΔ*myc* beads, but not by control GST beads (A.K.R. and R.B., unpublished data). One of the cellular proteins that was specifically bound by GSTΔ*myc* beads had a relative molecular mass of about 105,000 (*M*_r, 105K).

As pRb has a similar *M*_r¹¹, we tested whether GSTΔ*myc* could form a specific complex with pRb. We incubated GSTΔ*myc* beads with unlabelled mild-detergent lysates from ML1 leukaemia cells (which express wild-type Rb protein) and separated bound proteins by SDS-PAGE. The presence of pRb was detected by western blot analysis using monoclonal antibodies to pRb. As a positive control we used GST-E1A beads, as

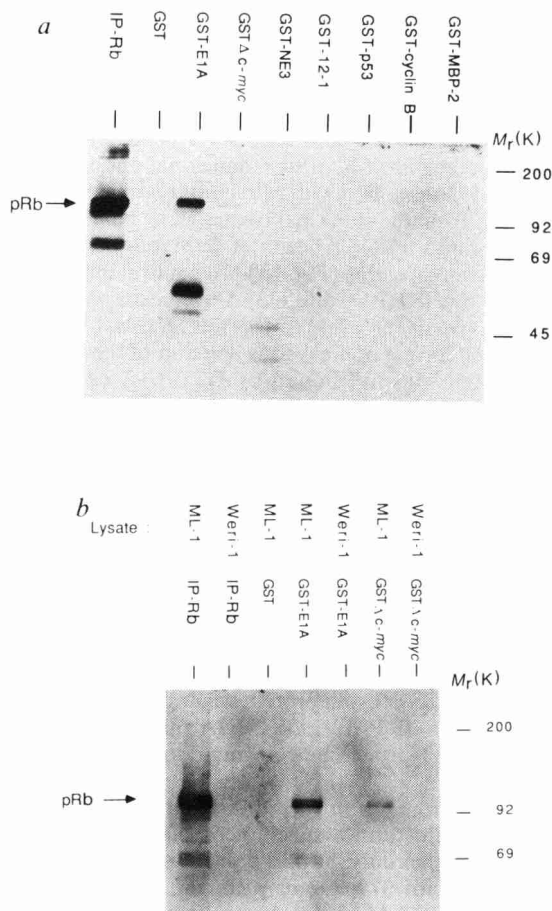


FIG. 1 Western blot analysis of Rb protein bound by GST fusion protein beads. Mild-detergent extracts were prepared from ML1 leukaemia cells or WERI-1 retinoblastoma cells as described¹² and incubated with agarose beads loaded with 20 µg GST fusion protein. As a control, pRb protein was immunoprecipitated from cell lysates for comparison (lanes labelled IP-Rb). Bound cellular proteins were released by boiling in SDS buffer and separated on a 6% acrylamide-SDS gel. Proteins were then transferred to nitrocellulose. The pRb protein was detected with monoclonal antibodies to pRb and an alkaline phosphatase-conjugated second antibody. *a*, Proteins bound using extracts from 2×10^7 ML1 leukaemia cells per lane. *b*, Proteins bound using both ML1 extracts and extracts derived from an equal number of WERI-1 retinoblastoma cells. The following GST fusion proteins were used: GST alone (lanes labelled GST), GST fused to the first 204 amino acids of human *c-myc* protein (GSTΔ*myc*), GST fused to the Ad5 243 amino acid E1A protein (GST-E1A), GST fused to the *N-myc* basic domain/helix-loop-helix region (GST-NE3), GST fused to the *N-myc* helix-loop-helix region (GST-12-1), GST fused to human p53 carrying a mutation in codon 273 (GST-p53), GST fused to cyclin B (GST-cyclin B) and GST fused to the MBP-2 transcription factor gene (ref. 26, GST-MBP-2). The lower-*M_r* protein species seen in the western blots are probably the result both of pRb protein degradation and of cross-reactivity of the pRb antisera with the bacterial GST fusion proteins.

METHODS. GST fusion proteins were prepared basically as described by Kaelin *et al.*¹⁰ with the modification that GST fusion proteins bound to glutathione-agarose beads were eluted with 5 mM glutathione. After elution, fusion proteins were dialysed overnight in 0.1 M sodium phosphate buffer, pH 7.8, and protein concentration was measured by modified Bradford assay. Dialysed proteins were checked for integrity by SDS-PAGE followed by Coomassie blue staining. This procedure of protein elution followed by rebinding to glutathione-agarose beads makes it possible to check protein integrity and quantify the amount of GST fusion protein loaded on the beads. To generate GST fusion protein-loaded beads, 50 µl glutathione-agarose beads (50% slurry, Sigma) was incubated with 20 µg GST fusion protein for 30 min at 4 °C. After this, beads were washed three times with 1 ml ELB buffer¹². Beads were then incubated with cell lysates for 2 h at 4 °C and washed five times 1 ml ELB buffer. Bound proteins were released by boiling in SDS buffer and separated on a polyacrylamide-SDS gel. Western blotting was as described²⁷.

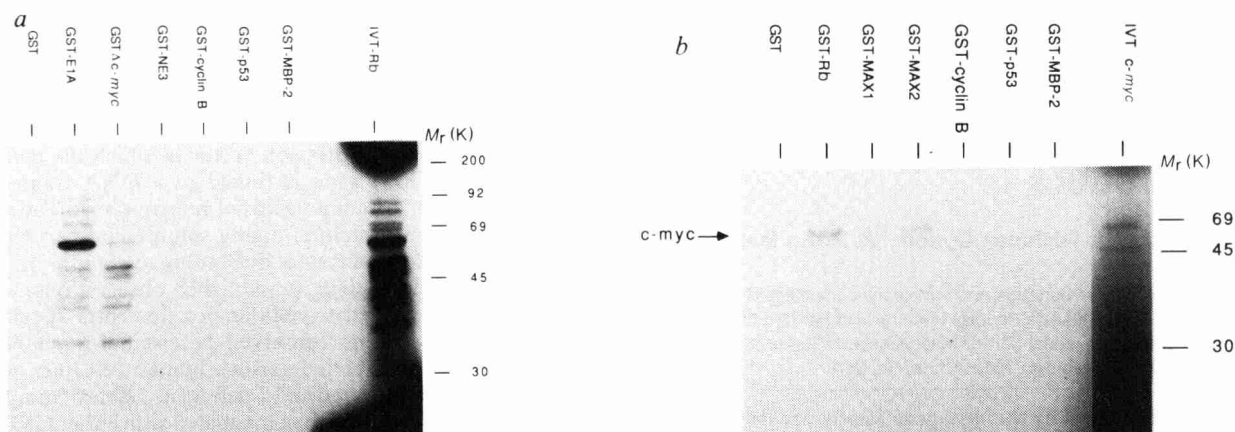


FIG. 2 Binding of *in vitro*-translated proteins to GST fusion protein beads. 35 S-labelled *in vitro*-translated human pRb and human c-myc proteins were incubated with agarose beads loaded with GST fusion proteins as indicated. For comparison, 1 μ l total *in vitro*-translated protein was loaded on the gel (lanes labelled IVT-Rb and IVT c-myc). The 45K protein species seen in the c-myc *in vitro* translation is a nonspecific protein also found in unprogrammed rabbit reticulocyte lysates. Bound proteins were separated on a 10% polyacrylamide gel and detected by fluorography. *a*, *In vitro*-translated pRb protein bound to GST fusion protein beads. *b*, *In vitro*-translated c-myc protein bound to GST fusion protein beads. *c*, *In vitro*-translated murine N-myc protein bound to GST fusion protein beads.

METHODS. GST fusion protein (1 μ g) was loaded on 20 μ l glutathione-agarose beads as described in the legend to Fig. 1. *In vitro*-translated proteins (10 μ l) were mixed with fusion protein beads in 0.5 ml NETN buffer¹⁰ and allowed to bind for 2 h at 4 °C. After this, beads were washed four times with 1 ml NETN. Bound proteins were released by boiling in SDS buffer and separated on a 10% acrylamide gel. The pRb and c-myc proteins were generated by sequential *in vitro* transcription and translation (Stratagene). GST fusion proteins used were as described in Fig. 1. GST-MAX1 and GST-MAX-2 represent GST fused to MAX, a c-myc binding protein. MAX-1 and MAX2 differ by nine amino acids as a result of alternative splicing of the MAX transcript³.

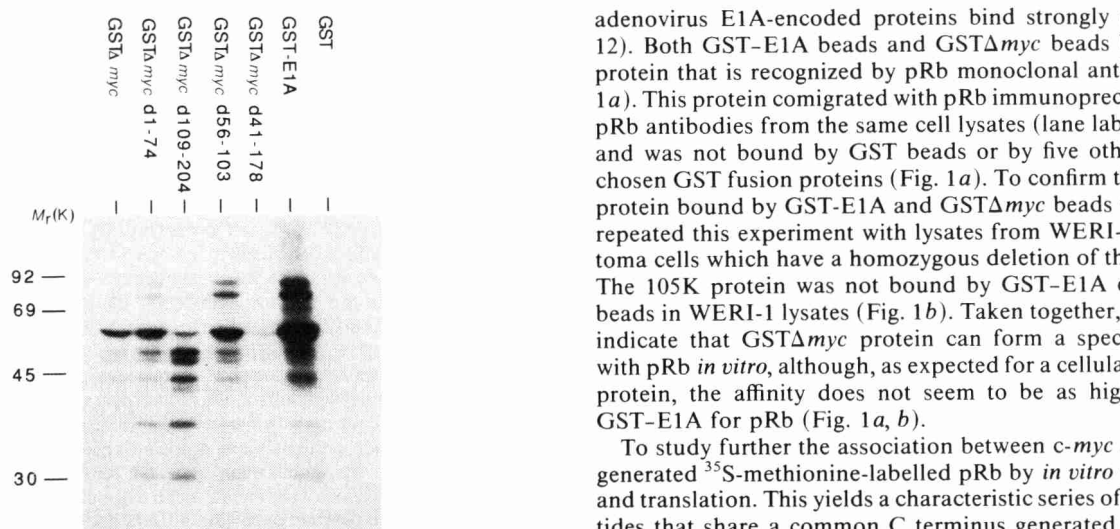


FIG. 3 Binding of pRb to mutant c-myc proteins. Agarose beads harbouring GSTΔc-myc protein or GSTΔc-myc mutant proteins were incubated with 35 S-methionine labelled *in vitro* translated pRb as described in the legend to Fig. 2. Bound pRb protein was separated on a 12% polyacrylamide gel and detected by fluorography. GSTΔc-mycd1-74 contains amino acids 75–204 of human c-myc linked to GST; GSTΔc-mycd109–204 contains amino acids 1–108 of c-myc; GSTΔc-mycd56–103 contains amino acids 1–55 and 103–204 of c-myc; GSTΔc-mycd41–178 has amino acids 1–41 and 178–204 of c-myc. The c-myc deletion mutants were generated by polymerase chain reaction amplification with specific c-myc primers using either wild-type c-myc plasmids or mutant c-myc plasmids⁶ as a template.

adenovirus E1A-encoded proteins bind strongly to pRb (ref. 12). Both GST-E1A beads and GSTΔc-myc beads bind a 105K protein that is recognized by pRb monoclonal antibodies (Fig. 1a). This protein comigrated with pRb immunoprecipitated with pRb antibodies from the same cell lysates (lane labelled IP-Rb) and was not bound by GST beads or by five other randomly chosen GST fusion proteins (Fig. 1a). To confirm that the 105K protein bound by GST-E1A and GSTΔc-myc beads was pRb, we repeated this experiment with lysates from WERI-1 retinoblastoma cells which have a homozygous deletion of the *Rb* gene¹³. The 105K protein was not bound by GST-E1A or GSTΔc-myc beads in WERI-1 lysates (Fig. 1b). Taken together, these results indicate that GSTΔc-myc protein can form a specific complex with pRb *in vitro*, although, as expected for a cellular Rb binding protein, the affinity does not seem to be as high as that of GST-E1A for pRb (Fig. 1a, b).

To study further the association between c-myc and pRb, we generated 35 S-methionine-labelled pRb by *in vitro* transcription and translation. This yields a characteristic series of Rb polypeptides that share a common C terminus generated by initiation of translation at internal start codons^{12,14}. *In vitro* synthesized pRb was incubated with GST-E1A beads, GSTΔc-myc beads or a series of control GST fusion protein beads and bound proteins were separated by SDS-PAGE. GST-E1A and GSTΔc-myc beads, but not control GST-protein beads, bound to *in vitro*-translated pRb (Fig. 2a).

We then tested whether *in vitro*-translated full-length human c-myc protein could form a specific complex with a GST-Rb fusion protein that has only the pRb domain required for binding to E1A (amino acids 379–792 of pRb; ref. 10). As a positive control, we used GST-MAX beads, as the protein MAX forms

heterodimers with c-myc through the helix-loop-helix/leucine zipper motifs³. Human c-myc translated *in vitro* bound to GST-Rb and GST-MAX beads with equal affinity (Fig. 2b). No binding of c-myc was seen to four other GST fusion proteins, including another tumour suppressor gene product, p53.

The N-terminal half of c-myc protein has several regions of high sequence homology to other myc protein family members⁷. We therefore tested whether *in vitro*-translated N-myc could bind to GST-Rb beads. Murine N-myc translated *in vitro* binds to GST-Rb beads and GST-MAX beads, but failed to bind to other GST fusion proteins (Fig. 2c). Our data seem to indicate that c-myc binds pRb better than N-myc (Fig. 2b, c). But because human c-myc and murine N-myc might differ in their affinity for MAX, this conclusion is not warranted by our data. We nonetheless conclude that the ability to bind pRb may be a feature common to myc protein family members. It is likely that pRb binding requires one or more of the highly conserved myc homology domains.

To map the domains in the c-myc protein involved in binding to pRb, we used a series of GST-myc mutant proteins. These mutant myc proteins were used in a binding assay with *in*

vitro-translated ³⁵S-methionine-labelled pRb protein. Both deletion of amino acids 1-74 and 109-204 of c-myc protein yielded proteins that bound to pRb (Fig. 3). Deletion of amino acids 56-103 also did not affect binding to pRb. But very weak binding was observed when amino acids 41-178 were deleted in c-myc. Thus at least two distinct regions in the c-myc protein mediate binding to pRb.

The Rb protein species generated by *in vitro* translation from internal AUG start codons represent a series of N-terminal truncations of pRb that can be used to map roughly the binding site of the c-myc protein on pRb¹⁴. High-affinity binding of E1A to pRb requires two domains of the Rb protein comprising amino acids 393-572 and 646-772, respectively¹⁴. The *in vitro* translation products of pRb with $M_r \geq 56$ K will therefore contain both E1A binding domains. Under the experimental conditions used, GST-E1A also shows a preference for ≥ 56 K pRb products, but additionally can bind to pRb *in vitro* translation products of ≥ 28 K with reduced affinity. GST-E1A protein did not bind to pRb *in vitro* translation products smaller than 28K, indicating that a region between amino acids 695 and 761 is essential for binding of GST-E1A to pRb. Similarly, GST Δ myc

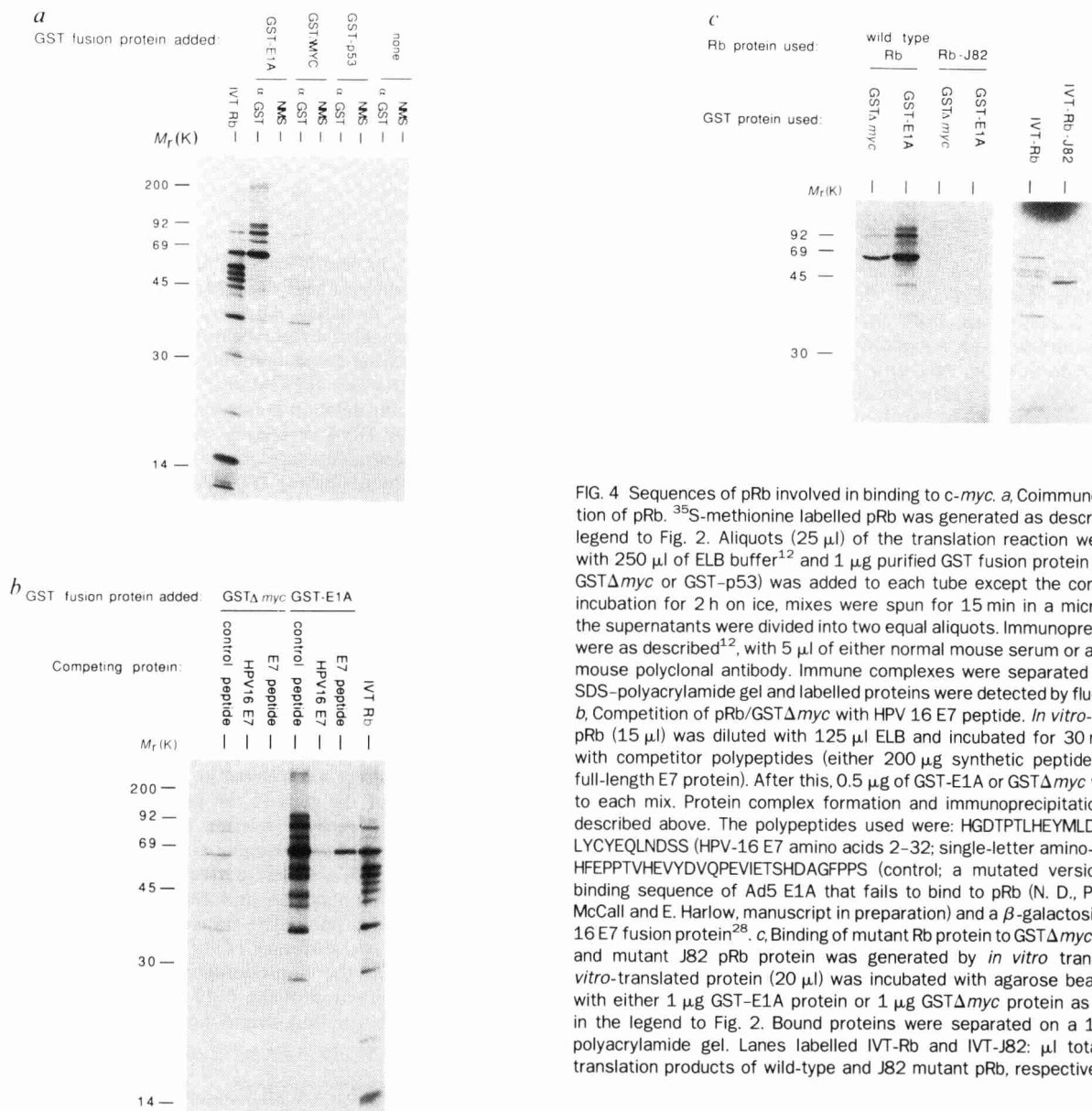


FIG. 4 Sequences of pRb involved in binding to c-myc. *a*, Coimmunoprecipitation of pRb. ³⁵S-methionine labelled pRb was generated as described in the legend to Fig. 2. Aliquots (25 μ l) of the translation reaction were diluted with 250 μ l of ELB buffer¹² and 1 μ g purified GST fusion protein (GST-E1A, GST Δ myc or GST-p53) was added to each tube except the control. After incubation for 2 h on ice, mixes were spun for 15 min in a microfuge and the supernatants were divided into two equal aliquots. Immunoprecipitations were as described¹², with 5 μ l of either normal mouse serum or an anti-GST mouse polyclonal antibody. Immune complexes were separated on a 12% SDS-polyacrylamide gel and labelled proteins were detected by fluorography. *b*, Competition of pRb/GST Δ myc with HPV 16 E7 peptide. *In vitro*-translated pRb (15 μ l) was diluted with 125 μ l ELB and incubated for 30 min on ice with competitor polypeptides (either 200 μ g synthetic peptide or 25 μ g full-length E7 protein). After this, 0.5 μ g of GST-E1A or GST Δ myc was added to each mix. Protein complex formation and immunoprecipitation was as described above. The polypeptides used were: HGDPTLHEYMLDLQPETTD-LYCQEQLNDSS (HPV-16 E7 amino acids 2-32; single-letter amino-acid code) HFEPTTVHEVYDVQPEVIETSHDAGFPSS (control; a mutated version of pRb binding sequence of Ad5 E1A that fails to bind to pRb (N. D., P. Guida, C. McCall and E. Harlow, manuscript in preparation) and a β -galactosidase-HPV 16 E7 fusion protein²⁸. *c*, Binding of mutant Rb protein to GST Δ myc. Wild-type and mutant J82 pRb protein was generated by *in vitro* translation. *In vitro*-translated protein (20 μ l) was incubated with agarose beads loaded with either 1 μ g GST-E1A protein or 1 μ g GST Δ myc protein as described in the legend to Fig. 2. Bound proteins were separated on a 12% SDS-polyacrylamide gel. Lanes labelled IVT-Rb and IVT-J82: μ l total *in vitro* translation products of wild-type and J82 mutant pRb, respectively.

protein bound to pRb translation products $\geq 28K$, but did not show increased affinity for translation products $\geq 56K$, as seen with GST-E1A (Fig. 4a). Furthermore, no binding to GST-E1A and GST Δmyc protein was observed with an *in vitro*-translated pRb polypeptide that contained a C terminus at amino acid 667 of the wild-type protein (data not shown). These data suggest that the binding sites for E1A and c-myc protein pRb may include similar regions in the C terminus of pRb.

To investigate further whether c-myc and DNA tumour virus-encoded proteins bind to similar sites on pRb, we did a competition experiment. We used human papilloma virus 16 (HPV-16) E7 protein for this experiment as this protein binds strongly to pRb¹⁵. Bacterially synthesized HPV-16 E7 protein or an HPV-16 E7 peptide that contained the pRb binding site on the E7 protein were incubated with *in vitro*-translated pRb. After this, GST Δmyc protein or GST-E1A protein was added and binding of pRb to GST fusion proteins was assayed. Both HPV-16 E7 protein and HPV-16 E7 peptide, but not control peptide, competed efficiently for binding of both GST Δmyc and GST-E1A protein to pRb (Fig. 4b). We conclude that c-myc, HPV-16 E7 and adenovirus E1A proteins have overlapping binding sites on pRb.

The C-terminal region of pRb to which c-myc binds is frequently deleted in human cancer¹⁶⁻¹⁸. This raised the possibility that in these tumour cells the interaction between c-myc protein and pRb is perturbed. To test this hypothesis, we used a cloned pRb cDNA from the J82 bladder carcinoma cell line which carries a 35-amino-acid C-terminal deletion in pRb (amino acids 697-731; ref. 16). *In vitro*-translated wild-type pRb and mutant J82 pRb was subsequently tested for binding to GST Δmyc . The mutant pRb from the J82 bladder carcinoma binds both GST-E1A and GST Δmyc with greatly reduced affinity (Fig. 4c). We conclude that the 35-amino-acid deletion of pRb in the J82 carcinoma cell line greatly alters the ability of pRb to bind to c-myc protein.

Interesting parallels exist between the c-myc and the adenovirus E1A genes. Both encode nuclear phosphoproteins that cooperate with an activated *ras* oncogene in transformation^{19,20}. Additionally, *myc* and E1A are structurally distantly related^{21,22}. We now extend these similarities by showing that c-myc protein, like E1A, can bind to pRb. Because full-length c-myc protein is highly insoluble, it was necessary to study the interaction between c-myc protein and pRb *in vitro* with a truncated soluble form of c-myc protein. Our findings are nevertheless intriguing because the region of pRb to which c-myc protein binds is frequently deleted in human tumours¹⁶⁻¹⁸. This suggests that c-myc protein interacts with a physiologically important site of pRb.

The ability of viral oncogenes to transform is intimately linked to their ability to bind pRb²³⁻²⁴. Our finding that the HPV-16 E7-encoded protein competes with c-myc protein for binding to pRb suggests that both proteins control cell proliferation by binding to overlapping sites on pRb. In this regard it is noteworthy that either antisense c-myc oligonucleotides or pRb overexpression results in a block in cell cycle progression from G1 to S phase (ref. 25, and S. Friend, personal communication). On the basis of our observations, it is tempting to speculate that dominant and recessive oncogenes can cooperate through direct binding to control progression through the cell cycle. □

Received 22 May; accepted 15 July 1991.

1. Lüscher, B. & Eisenman, R. N. *Genes Dev.* **4**, 2025-2035 (1990).
2. Murre, C. M. M., McCaw, P. S. & Baltimore, D. *Cell* **56**, 777-783 (1989).
3. Blackwood, E. M. & Eisenman, R. N. *Science* **251**, 1211-1217 (1991).
4. Blackwell, T. K. et al. *Science* **250**, 1149-1151 (1990).
5. Pendergast, G. C. & Ziff, E. B. *Science* **251**, 186-189 (1991).
6. Stone, J. et al. *Molec. cell. Biol.* **7**, 1697-1709 (1987).
7. Sugiyama, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9144-9148 (1989).
8. Moore, J. P., Hancock, D. C., Littlewood, T. D. & Evan, G. I. *Oncogene Res.* **2**, 65-80 (1987).
9. Smith, D. B. & Johnson, K. S. *Gene* **67**, 31-40 (1988).
10. Kaelin, W. G. et al. *Cell* **64**, 521-532 (1991).
11. Lee, W. H. et al. *Nature* **329**, 642-645 (1987).

12. Whyte, P. et al. *Nature* **334**, 124-129 (1988).
13. Friend, S. H. et al. *Nature* **323**, 643-646 (1986).
14. Hu, Q., Dyson, N. & Harlow, E. *EMBO J.* **9**, 1147-1155 (1990).
15. Dyson, N., Howley, P. M., Münger, K. & Harlow, E. *Science* **243**, 934-936 (1989).
16. Horowitz, J. M. et al. *Science* **243**, 937-940 (1989).
17. Yandell, D. W. et al. *New Engl. J. Med.* **321**, 1689-1695 (1989).
18. Dunn, J. M. et al. *Molec. cell. Biol.* **9**, 4596-4604 (1989).
19. Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596-602 (1983).
20. Ruley, H. E. *Nature* **304**, 602-606 (1983).
21. Ralston, R. & Bishop, J. M. *Nature* **306**, 803-806 (1983).
22. Figge, J., Wester, T., Smith, T. F. & Paucha, E. *J. Virol.* **62**, 1814-1818 (1988).
23. Whyte, P., Williamson, N. M. & Harlow, E. *Cell* **56**, 67-75 (1989).
24. De Caprio, J. A. et al. *Cell* **54**, 275-283 (1988).
25. Heikkilä, R. et al. *Nature* **328**, 445-449 (1987).
26. Rustgi, A. K., Van't Veer, L. J. & Bernards, R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8707-8710 (1990).
27. Windle, J. J. et al. *Nature* **343**, 665-669 (1990).
28. Jones, R. E. et al. *J. biol. Chem.* **265**, 12782-12785 (1990).

ACKNOWLEDGEMENTS. We thank M. Billaud, A. Fattaey, T. Frebourg, D. Hill, W. Kaelin, S. Shiff, L.-H. Tsai and L. van't Veer for GST fusion proteins, D. Patrick for HPV E7 protein, R. Kingston for the c-myc *in vitro* expression vector, J. Stone for mutant c-myc plasmids, J. Horowitz for the pRb-J82 plasmid, E. Harlow for discussions and critical reading of the manuscript, and M. Dembski for expert technical assistance. This work was supported by grants from the National Institutes of Health and the Searle Scholarship Foundation (R.B.), and an Institutional Grant from the Massachusetts General Hospital (N.D.). A.K.R. was supported by the American Cancer Society and by a Clinical Investigator Award.

Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*

Thuan Q. Trinh* & Richard R. Sinden†

Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati College of Medicine, 231 Bethesda Avenue, Cincinnati, Ohio 45267-0524, USA

WHEN present in single-stranded DNA, palindromic or quasi-palindromic sequences have the potential to form complex secondary structures, including hairpins, which may facilitate inter-strand misalignment of direct repeats and be responsible for diverse types of replication-based mutations, including deletions, additions, frameshifts and duplications¹⁻⁵. In regions of palindromic symmetry, specific deletion events may involve the formation of a hairpin or other DNA secondary structures which can stabilize the misalignment of direct repeats^{1,2}. One model suggests that these deletions occur during DNA replication by slippage of the template strand and misalignment with the progeny strand^{6,7}. The concurrent DNA replication model, involving an asymmetric dimeric DNA polymerase III complex which replicates the leading and lagging strands⁸, has significant implications for mutagenesis. The intermittent looping of the lagging strand template, and the fact that the lagging strand template may contain a region of single-stranded DNA the length of an Okazaki fragment, provides an opportunity for DNA secondary-structure formation and misalignment. Here we report our design of a palindromic fragment to create an 'asymmetric palindromic insert' in the chloramphenicol acetyltransferase gene of plasmid pBR325. The frequency with which the insert was deleted in *Escherichia coli* depends on the orientation of the gene in the plasmid. Our results suggest that replication-dependent deletion between direct repeats may occur preferentially in the lagging strand.

Plasmid pBR325, a ColE1-derived plasmid, has a unidirectional origin of replication and requires many host-encoded proteins for replication⁹. The transcribed strand of the chloramphenicol acetyltransferase (*CAT*) gene is the leading template strand and the non-coding strand the lagging strand. Reversing the direction of the *CAT* gene changes the relationship between the coding strand and the leading and lagging

† To whom correspondence should be addressed.

* Present address: Merck, Sharp, and Dohme Research Laboratories, Department of Membrane Biochemistry and Biophysics PO Box 2000, Rahway, New Jersey 07065, USA.

| Plasmids | Nucleotide sequence at the <i>Eco</i> RI site. |
|----------|--|
| pBR325p1 | <pre> <----- <----- <----- gctcatccggAATTCCTATAGGAATTCaattccgtaggca ----->----- </pre> |
| pBR523p1 | <pre> <----- <----- <----- tgccatacggAATTAGAATTCCTATAGGaattccggatgagc ----->----- </pre> |
| pBR325n1 | <pre> <----- <----- gctcatccggAATTCGTCTGATGCACGaattccgtaggca ----->----- </pre> |
| pBR523n1 | <pre> <----- <----- tgccatacggAATTCGTGCATCAGACGaattccggatgagc ----->----- </pre> |

FIG. 1 Nucleotide sequence of the plasmids around the *Eco*RI site in the *CAT* gene of pBR325 and pBR523. Shown are the sequences of the 'deletion insert' (capital letters) and the surrounding *CAT* gene sequence (lower-case). Inserts were chemically synthesized as *Eco*RI fragments and cloned into pBR325. To construct the pBR523 derivatives, the *CAT* gene was reversed by cutting with *Asu*II and religating. A difference in the *Pvu*II restriction patterns of pBR325 and pBR523 derivatives was used to screen potential pBR523 clones. The nucleotide sequence of all plasmids was confirmed by DNA sequencing. The various direct repeats are indicated by arrows above the sequence pointing in the direction of replication of the leading strand. The inverted repeat in the p1 insert is indicated by arrows below the sequence.

template strands of replication. Reversing the gene, rather than reversing the insert within the *CAT* gene, preserves the relationship between the insert and flanking DNA sequence that might influence the frequency of mutation.

The frequency of spontaneous deletion between direct repeats represents the sum of all possible events that result from homologous or illegitimate recombination, DNA repair activities or errors of DNA replication. If deletion between direct repeats occurs with equal frequency in the leading and lagging strands, there would be no effect of reversing the *CAT* gene. During replication, if deletion occurs preferentially in either the leading or lagging strand, deletion might occur at different frequencies. For a 'symmetric deletion insert', the mechanism of deletion or the DNA secondary structures that may be involved in deletion will be the same in either the leading or lagging

strand. Consequently, the frequency of deletion should be unchanged by reversing the *CAT* gene. But for an 'asymmetric deletion insert', the DNA secondary structures that may be involved in deletion will be different in the leading and lagging strands, and thus a different deletion frequency will be observed when the *CAT* gene is reversed.

We have designed several 17- and 18-base-pair symmetric and asymmetric deletion inserts and have shown that the frequency of deletion from the *Eco*RI site in the *CAT* gene in pBR325 can vary over more than a 100-fold range in *recA*⁻ *E. coli*. Reversion to chloramphenicol resistance (*Cm*^r) occurs primarily by deletion between the direct repeats, which include the *Eco*RI sites, resulting in a restoration of the *CAT* gene (T.Q.T. and R.R.S., manuscript in preparation). Here we examine the effect of reversing the *CAT* gene on the frequency of deletion of a

A pBR325n1, Leading Strand

a

```

          <----- <-----
5' gctcatccggAATTCGTCTGATGCACGaattccgtaggca 3'
3' cgagtaggccttaaGCAGACTACGTGCTTAaggcataccgt 5'

```

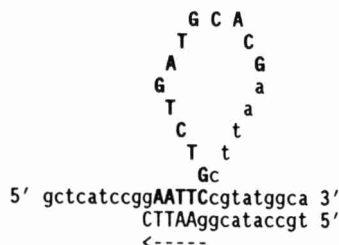
b

```

          <----- <-----
5' gctcatccggAATTCGTCTGATGCACGaattccgtaggca 3'
3' CTTAaggcataccgt 5'
          <-----

```

C



B pBR325n1, Lagging Strand

a

```

          -----> ----->
5' gctcatccggAATTCGTCTGATGCACGaattccgtaggca 3'
3' cgagtaggccttaaGCAGACTACGTGCTTAaggcataccgt 5'

```

b

```

          ----->
5' gctcatccggAATTC
3' cgagtaggccttaaGCAGACTACGTGCTTAaggcataccgt 5'

```

C

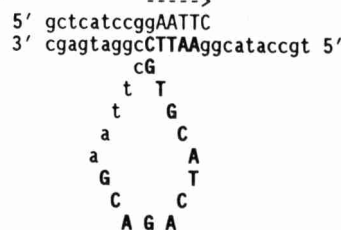


FIG. 2 Potential structural intermediates involved in deletion of a non-palindromic symmetric deletion insert. A, Potential DNA secondary structure in the leading strand of pBR325n1 containing a 'symmetric non-palindromic insert'. a, Nucleotide sequence of the 17-bp n1 insert and flanking DNA. The n1 insert is in upper-case bold characters. The direct repeats are represented by arrows above the sequence. b, DNA replication of the first copy of the

direct repeat 5'GAATTC3'. c, The template strand slips and the second (left-most) copy of the direct repeat in the template pairs with the direct repeat in the progeny strand. Replication continues resulting in deletion of the fragment between direct repeats in the progeny strand. B, Potential DNA secondary structure in the lagging strand of pBR325n1. Events shown are analogous to that described above for the leading strand.