(12) PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. AU 199523845 B2
(10) Patent No. 703908

(54) Title

Cell-cycle regulatory proteins, and uses related thereto

(51)[6] International Patent Classification(s)

C12N 015/12 C12N 015/11
C07K 014/47 C12Q 001/68
C07K 016/18 G01N 033/53

(21) Application No: 199523845 (22) Application Date: 1995.04.14

(87) WIPO No: WO95/28483

(30) Priority Data

(31) Number	(32) Date	(33) Country
08/227371	1994.04.14	US
08/248812	1994.05.25	US
08/306511	1994.09.14	US
08/346147	1994.11.29	US

(43) Publication Date : 1995.11.10
(43) Publication Journal Date : 1995.12.21
(44) Accepted Journal Date : 1999.04.01

(71) Applicant(s)

Cold Spring Harbor Laboratory

(72) Inventor(s)

David H Beach; Douglas J Demetrick; Manuel Serrano; Gregory J Hannon

(74) Agent/Attorney

GRIFFITH HACK,GPO Box 1285K,MELBOURNE VIC 3001

(56) Related Art

NATURE, VOL 366, 16 DEC 1993, M. SERRANO ET AL, PP.704-707
NATURE, VOL 368, 21 APRIL 1994, T. NOBORI ET AL, PP.753-756
AU 54440/94

OPI DATE 10/11/95 APPLN. ID 23845/95
AOJP DATE 21/12/95 PCT NUMBER PCT/US95/04636
AU9523845

(51) International Patent Classification 6 :
C12N 15/12, 15/11, C07K 14/47, 16/18, C12Q 1/68, G01N 33/53

A1

(11) International Publication Number: WO 95/28483

(43) International Publication Date: 26 October 1995 (26.10.95)

(21) International Application Number: PCT/US95/04636

(22) International Filing Date: 14 April 1995 (14.04.95)

(30) Priority Data:

08/227,371	14 April 1994 (14.04.94)	US
08/248,812	25 May 1994 (25.05.94)	US
08/306,511	14 September 1994 (14.09.94)	US
08/346,147	29 November 1994 (29.11.94)	US

(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(71) Applicant: COLD SPRING HARBOR LABORATORY [US/US]; One Bungtown Road, Cold Spring Harbor, NY 11724 (US).

(72) Inventors: BEACH, David, H.; 19 Woodland Drive, Huntington Bay, NY 11743 (US). DEMETRICK, Douglas, J.; 17 Monmouth Drive, East Northport, NY 11731 (US). SERRANO, Manuel; 297 Mill Hill Road, Mill Neck, NY 11765 (US). HANNON, Gregory, J.; 92 Sammis Street, Huntington, NY 11743 (US).

(74) Agents: VINCENT, Matthew, P. et al.; Lahive & Cockfield, 60 State Street, Boston, MA 02109 (US).

(54) Title: CELL-CYCLE REGULATORY PROTEINS, AND USES RELATED THERETO

H9	U18	CCL119	MCF-7	HTB 125	SaOs2	A431	normal #1	normal #2	cell
absent	absent	absent	absent	absent	altered	altered	norm	norm	exon 1
absent	absent	absent	absent	absent	altered	absent	norm	norm	exon 2

WI38	CCL120	HeLa	HTB100	ZRB75	GM130	Tera2	HTB172	HTB173	cell
norm	norm	norm	norm	norm	norm	norm	norm	norm	exon 1
norm	norm	norm	norm	norm	norm	norm	norm	norm	exon 2

(57) Abstract

The present invention relates to the discovery in eukaryotic cells, particularly mammalian cells, of a novel family of cell-cycle regulatory proteins ("CCR-proteins"). As described herein, this family of proteins includes a polypeptide having an apparent molecular weight of 16 kDa, and a polypeptide having an apparent molecular weight of approximately 15 kDa, each of which can function as an inhibitor of cell-cyle progression, and therefore ultimately of cell growth. Thus, similar to the role of p21 to the p53 checkpoint, the subject CCR-proteins may function coordinately with the cell-cycle regulatory protein, retinoblastoma (RB). Furthermore, the CCR-protein family includes a protein having an apparent molecular weight of 13.5 kDa (hereinafter "p13.5"). The presumptive role of p13.5, like p16 and p15, is in the regulation of the cell-cycle.

CELL-CYCLE REGULATORY PROTEINS, AND USES RELATED THERETO

Background of the Invention

Neoplasia is characterized by deregulated cell growth and division. Inevitably, molecular pathways controlling cell growth must interact with those regulating cell division. It was not until very recently, however, that experimental evidence became available to bring such connection to light. Cyclin A was found in association with the adenovirus oncoprotein E1A in virally transformed cells (Giordona et al. *Cell* 58:981 (1989); and Pines et al. *Nature* 346:760 (1990)). In an early hepatocellular carcinoma, the human cyclin A gene was found to be the integration site of a fragment of the hepatitis B virus, which leads to activation of cyclin A transcription and a chimeric viral cyclin A protein that is not degradable *in vitro* (Wang et al. *Nature* 343:555 (1990)). The cell-cycle gene implicated most strongly in oncogenesis thus far is the human cyclin D1. It was originally isolated through genetic complementation of yeast G_1 cyclin deficient strains (Xiong et al. *Cell* 65:691(1991); and Lew et al. *Cell* 66:1197 (1991)), as cellular genes whose transcription is stimulated by CSF-1 in murine macrophages (Matsushine et al. *Cell* 65:701 (1991)) and in the putative oncogene *PRAD1* rearranged in parathyroid tumors (Montokura et al. *Nature* 350:512 (1991). Two additional human D-type cyclins, cyclins D2 and D3, were subsequently identified using PCR and low-stringency hybridiazation techniques (Inaba et al. *Genomics* 13:565 (1992); and Xiong et al. *Genomics* 13:575 (1992)). Cyclin D1 is genetically linked to the *bcl-1* oncogene, a locus activated by translocation to an immunoglobulin gene enhancer in some B-cell lymphomas and leukemias, and located at a site of gene amplification in 15-20% of human breast cancers and 25-48% of squamous cell cancers of head and neck origin.

However, the creation of a mutant onocogene is only one of the requirements needed for tumor formation; tumorigenesis appears to also require the additional inactivation of a second class of critical genes: the "anti-oncogenes" or "tumor-suppressing genes." In their natural state these genes act to suppress cell proliferation. Damage to such genes leads to a loss of this suppression, and thereby results in tumorigenesis. Thus, the deregulation of cell growth may be mediated by either the activation of oncogenes or the inactivation of tumor-suppressing genes (Weinberg, R.A., (Sept 1988) *Scientific Amer.*pp 44-51).

Oncogenes and tumor-suppressing genes have a basic distinguished feature. The oncogenes identified thus far have arisen only in somatic cells, and thus have been incapable of transmitting their effects to the germ line of the host animal. In contrast, mutations in tumor-suppressing genes can be identified in germ line cells, and are thus transmissible to an animal's progeny.

The classic example of a hereditary cancer is retinoblastomas in children. The incidence of the retinoblastomas is determined by a tumor suppressor gene, the retinoblastoma (RB) gene (Weinberg, R.A., (Sept 1988) *Scientific Amer.*pp 44-51; Hansen et al. (1988) *Trends Genet* 4:125-128). Individuals born with a lesion in one of the RB alleles are predisposed to early childhood development of retinoblastomas. Inactivation or mutation of the second RB allele in one of the somatic cells of these susceptible individuals appears to be the molecular event that leads to tumor formation (Caveneee et al. (1983) *Nature* 305:799-784; Friend et al. (1987) *PNAS* 84:9059-9063).

The RB tumor-suppressing gene has been localized onto human chromosome 13. The mutation may be readily transmitted through the germ line of afflicted individuals (Cavenee, et al. (1986) *New Engl. J. Med* 314:1201-1207). Individuals who have mutations in only one of the two naturally present alleles of this tumor-suppressing gene are predisposed to retinoblastoma. Inactivation of the second of the two alleles is, however, required for tumorigenesis (Knudson (1971) *PNAS* 68:820-823).

A second tumor-suppressing gene is the p53 gene (Green (1989) *Cell* 56:1-3; Mowat et al (1985 *Nature* 314:633-636). The protein encoded by the p53 gene is a nuclear protein that forms a stable complex with both the SV40 large T antigen and the adenovirus E1B 55 kd protein. The p53 gene product may be inactivated by binding to these proteins.

Based on cause and effect analysis of p53 mutants, the functional role of p53 as a "cell-cycle checkpoint", particularly with respect to controlling progression of a cell from G1 phase into S phase, has implicated p53 as able to directly or indirectly affect cycle cycle machinery. The first firm evidence for a specific biochemical link between p53 and the cell-cycle comes a finding that p53 apparently regulates expression of a second protein, p21, which inhibits cyclin-dependent kinases (CDKs) needed to drive cells through the cell-cycle, e.g. from G1 into S phase. For example, it has been demonstrated that non-viral transformation, such as resulting at least in part from a mutation of deletion of of the p53 tumor suppressor, can result in loss of p21 from cyclin/CDK complexes. As described Xiong et al. (1993) *Nature* 366:701-704, induction of p21 in response to p53 represents a plausible mechanism for effecting cell-cycle arrest in response to DNA damage, and loss of p53 may deregulate growth by loss of the p21 cell-cycle inhibitor.

The role of RB as a tumor-suppressor protein in cell-cycle control is believed to be similiar to that of p53. However, whereas p53 is generally believed to be responsive to such indigenous environmental cues as DNA damage, the RB protein is apparently involved in coordinating cell growth with exogenous stimulus that normally persuade a cell to cease proliferating, such as diffusible growth inhibitors. In normal cells, RB is expressed throughout the cell cycle but exists in multiple phosphorylated forms that are specfic for certain phases of the cycle. The more highly phosphorylated forms are found during S and G_2/M, whereas the underphosphorylated forms are the primary species seen in G_1 and in the

growth arrested state. Based on these observations, it has been argued that if RB is to have a regulatory (suppressive) activity in the cell-cycle, this activity must be regulated at the post-translational level. Accordingly, underphosphorylated RB would be the form with growth-suppressive activity, since this form is prevalent in G1 and growth arrested cells.

To this end, it is noted that various paracrine growth inhibitors, such as members of the TGF-β family, prevent phosphorylation of RB and arrest cells in late G_1. Current models suggest that during G1, cyclin dependent kinases and particularly cyclin D-associated kinases, CDK4 and CDK6, phosphorylate the product of the retinoblastoma susceptibility gene, RB, and thus release cells from its growth inhibitory effects. TGF-β treatment causes accumulation of RB in the under-phosphorylated state and expression of RB-inactivating viral oncoproteins prevent TGF-β induced cell cycle arrest. In similar fashion, other related differentiation factors, such as activin, induce accumulation of unphosphorylated RB that is correlated with arrest in G_1 phase.

Recently, it has been demonstrated that the RB protein is a phosphorylation substrate for both CDK4 and CDK6 (Serano et al. (1993) *Nature* 366:704-707; Kato et al. (1993) *Genes Dev* 7:331-342; and Meyerson et al. (1994) *Mol Cell Biol* 14:2077-2086). However, prior to the present discovery, there was little information concerning the manner by which CDK phosphorylation of RB was negatively regulated.

For the purposes of this specification it will be clearly understood that the word "comprising" means "including but not limited to", and that the word "comprises" has a corresponding meaning.

AUSTRALIAN PATENT OFFICE

Summary of the Invention

The present invention relates to the discovery in eukaryotic cells, particularly mammalian cells, of a novel family of cell-cycle regulatory proteins ("CCR-proteins"). As described herein, this family of proteins includes a polypeptide having an apparent molecular weight of 16 kDa, and a polypeptide having an apparent molecular weight of approximately 15 kDa, each of which can function as an inhibitor of cell-cycle progression, and therefore ultimately of cell growth. Thus, similar to the role of p21 to the p53 checkpoint, the subject CCR-proteins may function coordinately with the cell-cycle regulatory protein, retinoblastoma (RB). Furthermore, the CCR-protein family includes a protein having an apparent molecular weight of 13.5 kDa (hereinafter "p13.5"). The presumptive role of p13.5, like p16 and p15, is in the regulation of the cell-cycle.

One aspect of the invention features a substantially pure preparation of a cell cycle regulatory (CCR) protein, or a fragment thereof, the full-length form of the CCR-protein having an approximate molecular weight in the range of 12 to 20kD, preferably 13kD to 18kD. In preferred embodiments, the full length form of the CCR-protein has an apparent molecular weight of approximately 13.5kD, 15kD, 15.5kD or 16kD. In a preferred embodiment: the polypeptide has an amino acid sequence at least 60% homologous to the amino acid sequence represented in one of SEQ ID No. 2, 4, 6 or 8; the polypeptide has an

AUSTRALIAN PATENT OFFICE

amino acid sequence at least 80% homologous to the amino acid sequence represented in one of SEQ ID No. 2, 4, 6 or 8; the polypeptide has an amino acid sequence at least 90% homologous to the amino acid sequence represented in one of SEQ ID No. 2, 4, 6 or 8; the polypeptide has an amino acid sequence identical to the amino acid sequence represented in one of SEQ ID No. 2, 4, 6 or 8. In a preferred embodiment: the fragment comprises at least 5 contiguous amino acid residues of SEQ ID No. 2, 4, 6 or 8; the fragment comprises at least 20 contiguous amino acid residues of SEQ ID No. 2, 4, 6 or 8; the fragment comprises at least 50 contiguous amino acid residues of SEQ ID No. 2, 4, 6 or 8. For instance, the CCR-protein can comprise an amino acid sequence represented by the general formula:

Met-Met-Met-Gly-Xaa-Xaa-Xaa-Val-Ala-Xaa-Leu-Leu-Leu-Xaa-Xaa-Gly-Ala-Xaa-Xaa-Asn-Cys-Xaa-Asp-Pro-Xaa-Thr-Xaa-Xaa-Xaa-Arg-Pro-Val-His-Asp-Ala-Ala-Arg-Glu-Gly-Phe-Leu-Asp-Thr-Leu-Val-Val-Leu-His-Xaa-Xaa-Gly-Ala-Arg-Leu-Asp-Val-Arg-Asp-Ala-Trp-Gly-Arg-Leu-Pro-Xaa-Asp-Leu-Ala-Xaa-Glu-Xaa-Gly-His-Xaa-Asp-Xaa-Xaa-Xaa-Tyr-Leu-Arg-Xaa-Ala-Xaa-Gly

For example, the CCR-protein can be represented by the sequence:

Met-Asp-Pro-Ala-Ala-Gly-Ser-Ser-Met-Glu-Pro-Ser-Ala-Asp-Trp-Leu-Ala-Thr-Ala-Ala-Ala-Arg-Gly-Arg-Val-Glu-Glu-Val-Arg-Ala-Leu-Leu-Glu-Ala-Val-Ala-Leu-Pro-Asn-Ala-Pro-Asn-Ser-Tyr-Gly-Arg-Arg-Pro-Ile-Gln-Val-Met-Met-Met-Gly-Xaa-Xaa-Xaa-Val-Ala-Xaa-Leu-Leu-Leu-Xaa-Xaa-Gly-Ala-Xaa-Xaa-Asn-Cys-Xaa-Asp-Pro-Xaa-Thr-Xaa-Xaa-Xaa-Arg-Pro-Val-His-Asp-Ala-Ala-Arg-Glu-Gly-Phe-Leu-Asp-Thr-Leu-Val-Val-Leu-His-Xaa-Xaa-Gly-Ala-Arg-Leu-Asp-Val-Arg-Asp-Ala-Trp-Gly-Arg-Leu-Pro-Xaa-Asp-Leu-Ala-Xaa-Glu-Xaa-Gly-His-Xaa-Asp-Xaa-Xaa-Xaa-Tyr-Leu-Arg-Xaa-Ala-Xaa-Gly-Gly-Thr-Arg-Gly-Ser-Asn-His-Ala-Arg-Ile-Asp-Ala-Ala-Glu-Gly-Pro-Ser-Asp-Ile-Pro-Asp;

or alternatively, by the sequence:

Met-Arg-Glu-Glu-Asn-Lys-Gly-Met-Pro-Ser-Gly-Gly-Gly-Ser-Asp-Glu-Gly-Leu-Ala-Thr-Pro-Ala-Arg-Gly-Leu-Val-Glu-Lys-Val-Arg-His-Ser-Trp-Glu-Ala-Gly-Ala-Asp-Pro-Asn-Gly-Val-Asn-Arg-Phe-Gly-Arg-Arg-Ala-Ile-Gln-Val-Met-Met-Met-Gly-Xaa-Xaa-Xaa-Val-Ala-Xaa-Leu-Leu-Leu-Xaa-Xaa-Gly-Ala-Xaa-Xaa-Asn-Cys-Xaa-Asp-Pro-Xaa-Thr-Xaa-Xaa-Xaa-Arg-Pro-Val-His-Asp-Ala-Ala-Arg-Glu-Gly-Phe-Leu-Asp-Thr-Leu-Val-Val-Leu-His-Xaa-Xaa-Gly-Ala-Arg-Leu-Asp-Val-Arg-Asp-Ala-Trp-Gly-Arg-Leu-Pro-Xaa-Asp-Leu-Ala-Xaa-Glu-Xaa-Gly-His-Xaa-Asp-Xaa-Xaa-Xaa-Tyr-Leu-Arg-Xaa-Ala-Xaa-Gly-Asp,

or yet in another embodiment, by the sequence:

Met-Met-Met-Gly-Xaa-Xaa-Xaa-Val-Ala-Xaa-Leu-Leu-Leu-Xaa-Xaa-Gly-Ala-Xaa-Xaa-Asn-Cys-Xaa-Asp-Pro-Xaa-Thr-Xaa-Xaa-Xaa-Arg-Pro-Val-His-Asp-Ala-Ala-Arg-Glu-Gly-Phe-Leu-Asp-Thr-Leu-Val-Val-Leu-His-

Xaa-Xaa-Gly-Ala-Arg-Leu-Asp-Val-Arg-Asp-Ala-Trp-Gly-Arg-Leu-Pro-
Xaa-Asp-Leu-Ala-Xaa-Glu-Xaa-Gly-His-Xaa-Asp-Xaa-Xaa-Xaa-Tyr-Leu-
Arg-Xaa-Ala-Xaa-Gly-Cys-Ser-Leu-Cys-Ser-Ala-Gly-Trp-Ser-Leu-Cys-
Thr-Ala-Gly-Asn-Val-Ala-Gln-Thr-Asp-Gly-His-Ser-Phe-Ser-Ser-Ser-
Thr-Pro-Arg-Ala-Leu-Glu-Leu-Arg-Gly-Gln-Ser-Gln-Glu-Gln-Ser.

In preferred embodiments, the CCR-protein specifically binds a CDK, e.g. a G_1 phase CDK, e.g. CDK4 and/or CDK6. The CCR-protein can be cloned from a mammalian cell, e.g. a human cell, e.g. a mouse cell.

Another aspect of the present invention features a polypeptide, of the CCR-protein family, which functions in one of either role of an agonist of cell-cycle regulation or an antagonist of cell-cycle regulation. In a preferred embodiment: the subject CCR-protein specifically binds a cyclin dependent kinase (CDK), e.g. specifically binds CDK4; e.g. specifically binds CDK6; e.g. inhibits a kinase activity of CDK4; inhibits a kinase activity of CDK6; e.g. inhibits phosphorylation of an RB protein by CDK4. In a more preferred embodiment: the CCR-protein regulates a eukaryotic cell-cycle, e.g. a mammalian cell-cycle, e.g., a human cell-cycle; the CCR-protein inhibits proliferation/cell growth of a eukaryotic cell, e.g., a human cell; the CCR-protein inhibits progression of a eukaryotic cell from G1 phase into S phase, e.g., inhibits progression of a mammalian cell from G1 phase into S phase, e.g., inhibits progression of a human cell from G1 phase into S phase; the CCR-protein inhibits the kinase activity of a cyclin dependent kinase (CDK), e.g. a CDK active in G1 phase, e.g. CDK 4; the CCR-protein suppresses tumor growth, e.g. in a tumor cell, e.g. in a tumor cell having an unimpaired RB or RB-like protein checkpoint. Moreover, CCR-proteins of the present invention may also have biological activities which include: an ability to regulate cell-cycle progression in response to extracellular factors and cytokines, e.g. functional in paracrine or autocrine regulation of cell growth and/or differentiation, e.g. inhibit CDK activation in response to transforming growth factor-β (TGF-β) or related growth, differentiation or morphogenesis factor.

Still another aspect of the invention concerns a polypeptide having an amino acid sequence homologous to an amino acid sequence represented by one of SEQ ID Nos: 2, 4, 6 or 8

Yet another aspect of the present invention concerns an immunogen comprising a CCR-protein of the present invention, or a fragment thereof, in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for the CCR-protein; e.g. a humoral response, e.g. an antibody response; e.g. a cellular response.

Another aspect of the present invention features recombinant CCR-protein, or a fragment thereof, having an amino acid sequence preferably: at least 60% homologous to the amino acid sequence represented in one of SEQ ID No. 2, 4, 6 or 8; at least 80% homologous to the amino acid sequence represented in one of SEQ ID No. 2, 4, 6 or 8; at least 90% homologous to the amino acid sequence represented in one of SEQ ID No. 2, 4, 6 or 8; identical to the amino acid sequence represented in one of SEQ ID No. 2, 4, 6 or 8. In a

AUSTRALIAN PATENT OFFICE

preferred embodiment: the fragment comprises at least 5 contiguous amino acid residues of SEQ ID No. 2, 4, 6 or 8; the fragment comprises at least 20 contiguous amino acid residues of SEQ ID No. 2, 4, 6 or 8; the fragment comprises at least 50 contiguous amino acid residues of SEQ ID No. 2, 4, 6 or 8. In a preferred embodiment, the recombinant CCR-protein functions in one of either role of an agonist of cell-cycle regulation or an antagonist of cell-cycle regulation. In a more preferred embodiment: the CCR-protein specifically binds a cyclin dependent kinase (CDK), e.g. specifically binds CDK4; e.g. specifically binds CDK6; e.g. inhibits a kinase activity of CDK4; inhibits a kinase activity of CDK6; e.g. inhibits phosphorylation of an RB protein by CDK4. In a more preferred embodiment: the CCR-protein regulates a eukaryotic cell-cycle, e.g. a mammalian cell-cycle, e.g., a human cell-cycle; the CCR-protein inhibits proliferation/cell growth of a eukaryotic cell, e.g., a human cell; the CCR-protein inhibits progression of a eukaryotic cell from G1 phase into S phase, e.g., inhibits progression of a mammalian cell from G1 phase into S phase, e.g., inhibits progression of a human cell from G1 phase into S phase; the CCR-protein inhibits the kinase activity of a cyclin dependent kinase (CDK), e.g. a CDK active in G1 phase, e.g. CDK 4; the CCR-protein suppresses tumor growth, e.g. in a tumor cell, e.g. in a tumor cell having an unimpaired RB or RB-like protein checkpoint.

In yet other preferred embodiments, the recombinant CCR-protein is a fusion protein further comprising a second polypeptide portion having an amino acid sequence from a protein unrelated the protein of SEQ ID No. 2, 4, 6 or 8. Such fusion proteins can be functional in a two-hybrid assay.

Another aspect of the present invention provides a substantially pure nucleic acid having a nucleotide sequence which encodes a CCR-protein, or a fragment thereof, having an amino acid sequence at least 60% homologous to one of SEQ ID Nos. 2, 4, 6 or 8. In a more preferred embodiment: the nucleic acid encodes a protein having an amino acid sequence at least 80% homologous to SEQ ID No. 2, more preferably at least 90% homologous to SEQ ID No. 2, and most preferably at least 95% homologous to SEQ ID No. 2; the nucleic acid encodes a protein having an amino acid sequence at least 80% homologous to SEQ ID No. 6, more preferably at least 90% homologous to SEQ ID No. 6, and most preferably at least 95% homologous to SEQ ID No. 6. The nucleic preferably encodes a CCR-protein which specifically binds a cyclin dependent kinase (CDK); e.g. specifically binds CDK4; e.g. specifically binds CDK6; e.g. which inhibits a kinase activity of CDK4; e.g. which inhibits phosphorylation of an RB protein by CDK4.

In another embodiment, the nucleic acid hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides of SEQ ID No. 1; more preferably to at least 20 consecutive nucleotides of SEQ ID No. 1; more preferably to at least 40 consecutive nucleotides of SEQ ID No. 1.

In a further embodiment, the nucleic acid hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides of SEQ ID No. 3; more preferably to at least 20 consecutive nucleotides of SEQ ID No. 3; more preferably to at least 40 consecutive nucleotides of SEQ ID No. 3.

In yet a further embodiment, the nucleic acid hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides of SEQ ID No. 5; more preferably to at least 20 consecutive nucleotides of SEQ ID No. 5; more preferably to at least 40 consecutive nucleotides of SEQ ID No. 5.

In yet a further embodiment, the nucleic acid hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides of SEQ ID No. 7; more preferably to at least 20 consecutive nucleotides of SEQ ID No. 7; more preferably to at least 40 consecutive nucleotides of SEQ ID No. 7.

Furthermore, in certain embodiments, the CCR nucleic acid will comprise a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, operably linked to the CCR-gene sequence so as to render the recombinant CCR-gene sequence suitable for use as an expression vector.

The present invention also features transgenic non-human animals, e.g. mice, which either express a heterologous CCR-gene, e.g. derived from humans, or which mis-express their own CCR-gene, e.g. p16, p15 or p13.5 expression is disrupted. Such a transgenic animal can serve as an animal model for studying cellular disorders comprising mutated or mis-expressed CCR alleles.

The present invention also provides a probe/primer comprising a substantially purified oligonucleotide, wherein the oligonucleotide comprises a region of nucleotide sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence of one of SEQ ID No. 1, 3, 5 or 7, or naturally occurring mutants thereof. In preferred embodiments, the probe/primer further comprises a label group attached thereto and able to be detected, e.g. the label group is selected from a group consisting of radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors. Such probes can be used as a part of a diagnostic test kit for identifying transformed cells, such as for measuring a level of a p16, p15 or p13.5 encoding nucleic acid in a sample of cells isolated from a patient; e.g. for measuring the mRNA level in a cell or determining whether the genomic CCR gene has been mutated or deleted.

The present invention also provides a method for treating an animal having unwanted cell growth characterized by a loss of wild-type CCR-protein function, comprising administering a therapeutically effective amount of an agent able to inhibit a kinase activity of a CDK, e.g. CDK4. In one embodiment, the method comprises administering a nucleic acid construct encoding a CCR protein, e.g. p16, p15 or p13.5, e.g. a polypeptide represented in one of SEQ ID Nos. 2, 4, 6 or 8, under conditions wherein the construct is incorporated by

CCR-deficient cells and the polypeptide is expressed, e.g. by gene therapy techniques. In another embodiment, the method comprises administering a CCR mimetic, e.g. a peptidomimetic, which binds to and inhibits the CDK.

Another aspect of the present invention provides a method of determining if a subject, e.g. a human patient, is at risk for a disorder characterized by unwanted cell proliferation, comprising detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a protein represented by one of SEQ ID Nos. 2, 4, 6 or 8, or a homolog thereof; or (ii) the mis-expression of the CCR-gene, e.g. the p16, p15 or p13.5 gene. In preferred embodiments: detecting the genetic lesion comprises ascertaining the existence of at least one of a deletion of one or more nucleotides from said gene, an addition of one or more nucleotides to said gene, an substitution of one or more nucleotides of said gene, a gross chromosomal rearrangement of said gene, a gross alteration in the level of a messenger RNA transcript of said gene, the presence of a non-wild type splicing pattern of a messenger RNA transcript of said gene, or a non-wild type level of said protein. For example, detecting the genetic lesion can comprise (i) providing a probe/primer comprising an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of one of SEQ ID Nos. 1, 3, 5 or 7, or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the CCR-gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and (iii) detecting, by hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion; e.g. wherein detecting the lesion comprises utilizing the probe/primer to determine the nucleotide sequence of the CCR-gene and, optionally, of the flanking nucleic acid sequences; e.g. wherein detecting the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR); e.g. wherein detecting the lesion comprises utilizing the probe/primer in a ligation chain reaction (LCR). In alternate embodiments, the level of said protein is detected in an immunoassay.

Yet another aspect of the invention pertains to a peptidomimetic which binds to a CCR-protein, e.g. p15 or p16, and inhibits its binding to a CDK, e.g. CDK4 or CDK6. For example, a preferred peptidomimetic is an analog of a peptide having the sequence VAEIG(V/E)GAYG(T/K)-V(F/Y)KARD, though more preferably the peptidomimetic is an analog of the hexa-peptide V(F/Y)KARD, and even more preferably of the tetrapeptide KARD. Likewise, another preferred peptide analog of CDK4/CDK6 comprises a peptide or peptidomimetic corresponding to the sequence SRTDRE(I/T)K(V/L)TLVFEHVDQDL(R/T)-TYLDK(A/V)PPPG(L/V), though more preferably the peptidomimetic is an analog of the hexa-peptides FEHVDQ or EHVDQD, and even more preferably of the tetrapeptides HVDQ or EHVD. Non-hydrolyzable peptide analogs of such residues can be generated using, for example, benzodiazepine, azepine, substituted gama lactam rings, keto-methylene pseudopeptides, β-turn dipeptide cores, or β-aminoalcohols.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims. The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press:1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No. 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Description of the Drawings

Figures 1A-1B are a schematic representation of the p16 cDNA, indicating the location of exon boundaries and PCR primers used in the present invention.

Figures 2A and 2B are the genomic nucleic acid sequence for exon 1 and the non-coding sequences directly flanking exon 1. The sequence is a composite sequence from several primers. Figure 2C is the genomic sequence about exon 2.

Figures 3A-3D are the genomic nucleic acid sequence for exon 3 and the non-coding sequences directly flanking exon 3. The sequence is a composite sequence from several primers.

Figure 4 illustrates the loss of p16 sequences from the genomes of several human tumor cells, as compared to normal human controls and other human tumors.

Figure 5 illustrates the restriction map for the mouse p16 gene.

Figure 6 is a sequence alignment of a highly conserved portion of the CCR-proteins p16, p15 and p13.5.

Figure 7 is a sequence alignment of portions of cyclin-dependent kinases.

Detailed Description of the Invention

Progression through the cell-cycle is marked by a series of irreversible transitions that separate discrete tasks necessary for faithful cell duplication. These transitions are negatively regulated by signals that constrain the cell-cycle until specific conditions are fulfilled. Entry in to mitosis, for example, is inhibited by incompletely replicated DNA or DNA damage. These restriction on cell-cycle progression are essential for preserving the fidelity of the genetic information during cell division. The transition from G_1 to S phase, on the other hand, coordinates cell proliferation with environmental cues, after which the checks on the cell-cycle progression tend to be cell autonomous. Among the signals that restrict cell-cycle progression during G_1 are extracellular proteins which inhibit cell proliferation, growth factor or amino acid depletion, and cell-cell contact. Disruption of these signaling pathways uncouples cellular responses from environmental controls and may lead to unrestrained cell proliferation.

Eukaryotic cells, in general, require cyclin-dependent kinases (CDKs) for progression through G_1 and entry into S phase. For instance, in mammalian cells, both D- and E-type cyclins are rate limiting for the G_1 to S transition, and both reduce, but do not eliminate, the cell's requirement for mitogenic growth factors. However, prior to the present discovery, there was little information concerning the manner by which these cyclins and CDKs are negatively regulated by either intracellular or extracellular signals that inhibit cell proliferation.

The present invention is directed to the discovery of a family of related cell-cycle regulatory proteins (herein refered to as "CCR-proteins") which function typically to restrict progression of a cell through mitosis, and are likely to be involved in controlling progression through meiosis. Members of this family are evolutionarily related to the approximately 16kd protein represented in SEQ ID No. 2, designated "p16" (also "$p16^{INK4a}$"). This family includes, for example, a polypeptide (termed "p15") having an approximate molecular weight of 15Kd, and a 13.5 Kd polypeptide (termed "p13.5"). The nucleotide sequences for the human p16, the human p15, the mouse p13.5, and the mouse p15 coding sequences are provided in SEQ ID Nos. 1, 3, 5 and 7, respectively. The corresponding amino acid sequences are represented in SEQ ID Nos. 2, 4, 6 and 8. Moreover, data from hybridization and immunoprecipitation experiments indicates still other members of the CCR-protein family exist, comprising proteins representing both evolutionarily divergent sequences as well as differentially spliced variants.

One function of members of this family of proteins in cell-cycle regulation is in modulating the activity of cyclin/CDK complexes during various stages of the cell-cycle, particularly those which include CDKs active in G_1 phase, such as CDK4 or CDK6. To illustrate, both p16 and p15 are demonstrated below to exert an inhibitory effect on the

activity of cyclin/CDK complexes, particularly those which include CDK4 or CDK6. For instance, each protein is able to inhibit the activity of cyclin D1/CDK complexes *in vivo*. As is generally known, cyclin D1 has been associated with a wide variety of proliferative diseases. Consequently, the present invention identifies a potential inhibitor of cell proliferation resulting from oncogenic expression of cyclin D1. Moreover, the diversity of members of the CCR-protein family, like the diversity of CDKs, is suggestive of individualistic roles of each member of this family, which may be tissue-type of cell-type specific, occur at different points in the cell-cycle, occur as part of different extracellular or intracellular signaling pathways, or a combination thereof.

As described in the examples below, certain of the CCR-proteins have been shown to be deleted or mutated at high frequency in tumors, such as derived from lung, breast, brain, bone, skin, bladder, kidney, ovary, or lymphocytes. Consequently, as set forth in the present application, replacement of CCR protein function by gene therapy or by CCR mimetics, or by direct inhibition of CDK4 or CDK6 activity, is therefore a potential therapy for treating such proliferative disorders. Moreover, the present data demonstrates that p15 expression is regulated by treatment with transforming growth factor-beta (TGF-β), suggesting that p15 may function as an effector of TGF-β mediated cell cycle arrest via inhibition of CDK4 or CDK6 kinases. Considered in light of recent findings demonstrating that reduced responsiveness to TGF-β may be an important event in the loss of growth control in many proliferative disorders, an approach to modulate CDK4/6 activity by CCR mimetics or CCR-gene therapy, or by mechanism based inhibitors of the kinases themselves, is even more attractive for treating such proliferative disorders.

Accordingly, the present invention makes available diagnostic and therapeutic assays and reagents for detecting and treating proliferative disorders arising from, for example, tumorigenic transformation of cells, or other hyperplastic or neoplastic transformation processes, as well as differentiative disorders, such as degeneration of tissue, e.g. neurodegeneration. For example, the present invention makes available reagents, such as antibodies and nucleic acid probes, for detecting altered complex formation, and/or altered levels of CCR-protein expression, and/or CCR-gene deletion or mutation, in order to identify transformed cells. Moreover, the present invention provides a method of treating a wide variety of pathological cell proliferative conditions, such as by gene therapy utilizing recombinant gene constructs encoding the subject CCR-proteins, or by providing CCR-mimetics, with the general strategy being the inhibition of aberrantly proliferating cells.

The subject proteins can also be used in assay systems to identify agents which either decrease the ability of the CCR-protein to bind a CDK (e.g. CDK4 or CDK6) and thereby relieve inhibition of cyclin/CDK complexes, or alternatively, which agonize or mimic the CCR-mediated inhibition of CDK activation. In the latter, e.g. CCR mimetics, the consequence of inhibiting activation of a cyclin/CDK complex, e.g. cyclin D/CDK4, is the

failure of the cell to advance through the cell-cycle, which inhibition can lead ultimately to cell death. Reactivation of the CDK/cyclin complexes, on the other hand, can disrupt or otherwise unbalance the cellular events occurring in a transformed cell. Such agents can be of use therapeutically to activate CDK complexes in cells transformed, for example, by tumor viruses. Treatment of such cells can cause premature progression through a checkpoint, e.g. the retinoblastoma (RB) checkpoint, and result in mitotic catastrophe (cell death) or induction of apoptosis.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

As used herein, the terms "gene", "recombinant gene" and "gene construct" refer to a nucleic acid comprising an open reading frame encoding a cell-cycle regulatory of the present invention, including both exon and (optionally) intron sequences. In preferred embodiments, the nucleic acid is DNA or RNA. Exemplary recombinant genes include nucleic acids which encode all or a CDK-binding portion of the p16 protein represented in SEQ ID No. 2, the p15 proteins represented in SEQ ID Nos. 4 and 8, or the p13.5 protein represented in SEQ ID No. 6. The term "intron" refers to a DNA sequence present in a given CCR-gene which is not translated into protein and is generally found between exons.

"Homology" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences.

The term "transfection" refers to the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of one of the subject cell-cycle regulatory proteins, e.g. p16, p15 or p13.5.

"Cells" or "cell cultures" or "recombinant host cells" or "host cells" are often used interchangeably as will be clear from the context. These terms include the immediate subject cell which expresses the cell-cycle regulatory protein of the present invention, and, of course, the progeny thereof. It is understood that not all progeny are exactly identical to the parental cell, due to chance mutations or difference in environment. However, such altered progeny are included in these terms, so long as the progeny retain the characteristics relevant to those

conferred on the originally transformed cell. In the present case, such a characteristic might be the ability to produce a recombinant CCR-protein.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. The term "expression vector" includes plasmids, cosmids or phages capable of synthesizing the subject CCR-protein encoded by the respective recombinant gene carried by the vector. Preferred vectors are those capable of autonomous replication and/expression of nucleic acids to which they are linked. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. Moreover, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, as well as polyadenylation sites, which induce or control transcription of protein coding sequences with which they are operably linked. In preferred embodiments, transcription of a recombinant CCR-gene is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are different from those sequences which control transcription of the naturally-occurring form of the regulatory protein.

The term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as cells of a neuronal lineage, e.g. glial cells, or alternatively, in epithelial cells, e.g. melanocytes. In an illustrative embodiment, gene constructs utilizing glial-specific promoters can be used as a part of gene therapy to cause expression of recombinant forms of one of the subject CCR-proteins in glioma cells with a feature of the gene construct being a tissue-specific promoter for directing expression of the subject protein in only glial tissue. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

As used herein, a "transgenic animal" is any animal, preferably a non-human mammal in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by trangenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a

recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form of the subject CCR proteins, e.g. either agonistic or antagonistic forms, or in which an endogenous CCR-gene has been disrupted. However, transgenic animals in which the recombinant CCR-gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. The "non-human animals" of the invention include vertebrates such as rodents, non-human primates, sheep, dog, cow, amphibians, reptiles, etc. Preferred non-human animals are selected from the rodent family including rat and mouse, most preferably mouse. The term "chimeric animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that the recombinant CCR-gene is present and/or expressed, or disrupted, in some tissues but not others.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., a CCR-polypeptide), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

As used herein, the terms "transforming growth factor-beta" and "TGF-β" denote a family of structurally related paracrine polypeptides found ubiquitously in vertebrates, and prototypic of a large family of metazoan growth, differentiation, and morphogenesis factors (see, for review, Massaque et al. (1990) *Ann Rev Cell Biol* 6:597-641; and Sporn et al. (1992) *J Cell Biol* 119:1017-1021).

The term "evolutionarily related to", with respect to nucleic acid sequences encoding CCR-proteins, refers to nucleic acid sequences which have arisen naturally in an organism, including naturally occurring mutants. In general, genes which encode proteins that are evolutionarily related to the p16 protein can be recognized by the general features of ankyrin-like repeats and the ability to bind to a cyclin dependent kinase. Moreover, the term "evolutionarily related to" also refers to nucleic acid sequences which, while derived from a naturally occurring CCR-proteins, have been altered by mutagenesis, as for example, the combinatorial mutagenesis techniques described below, yet still encode polypeptides which have at least one activity of a CCR-protein. For instance, the sequence of p16 can be altered by mutagenesis based on amino acid substitutions derived from alignment with the p15 and/or p13.5 sequences.

One aspect of the present invention pertains to an isolated nucleic acid comprising the nucleotide sequence encoding a CCR-protein, fragments thereof encoding polypeptides having at least one biological activity of a CCR-protein, and/or equivalents of such nucleic acids. The term nucleic acid as used herein is intended to include such fragments and equivalents. The term equivalent is understood to include nucleotide sequences encoding functionally equivalent CCR-proteins or functionally equivalent peptides having an activity of a CCR-protein such as described herein. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will also include sequences that differ from the nucleotide sequence encoding the subject p16 protein represented by SEQ ID No. 2, or the p15 proteins represented by SEQ ID Nos. 4 and 8, or the p13.5 protein represented by SEQ ID No. 6 due to the degeneracy of the genetic code. Equivalents will also include nucleotide sequences that hybridize under stringent conditions (i.e., equivalent to about 20-27°C below the melting temperature (T_m) of the DNA duplex formed in about 1M salt) to the nucleotide sequence of a CCR-gene shown in SEQ ID No. 1, 3, 5 or 7. In one embodiment, equivalents will further include nucleic acid sequences derived from and evolutionarily related to the nucleotide sequence shown in one of SEQ ID No. 1, 3, 5 or 7.

The term "isolated" as also used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. For example, an isolated nucleic acid encoding on of the subject CCR-proteins preferably includes no more than 10 kilobases (kb) of nucleic acid sequence which naturally immediately flanks the CCR-gene in genomic DNA, more preferably no more than 5kb of such naturally occurring flanking sequences, and most preferably less than 1.5kb of such naturally occurring flanking sequence. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state.

Polypeptides referred to herein as having an activity of a cell-cycle regulatory protein preferably have an amino acid sequence homologous to and corresponding to all or a portion of the amino acid sequence of the p16 protein shown in SEQ ID No. 2. Exemplary polypeptides include the p15 protein shown in either SEQ ID No. 4 or 8, or of the p13.5 protein shown in SEQ ID No. 6, or isoforms of one of these proteins (including differential splicing variants). In preferred embodiments, the biological activity of a CCR-protein includes one or more of the following: (i) an ability to regulate a eukaryotic cell-cycle, e.g. a mammalian cell-cycle, e.g., a human cell-cycle; (ii) an ability to inhibit proliferation/cell growth of a eukaryotic cell, e.g. a mammalian cell, e.g., a human cell; (iii) an ability to

inhibit progression of a eukaryotic cell from G1 phase into S phase, e.g., inhibit progression of a mammalian cell from G1 phase into S phase, e.g., inhibit progression of a human cell from G1 phase into S phase; and/or (iv) an ability to inhibit the kinase activity of a cyclin dependent kinase (CDK), e.g. a CDK active in G1 phase, e.g. CDK 4, e.g. CDK6; e.g. an ability to inhibit phosphorylation of a retinoblastoma (RB) or retinoblastoma-like protein by a cyclin dependent kinase. Moreover, CCR-proteins of the present invention may also have biological activities which include: an ability to suppress tumor growth, e.g. in a tumor having an unimpaired RB protein; an ability to regulate cell-cycle progression in response to extracellular factors and cytokines, e.g. functional in paracrine or autocrine regulation of cell growth and/or differentiation, e.g. inhibit CDK activation in response to transforming growth factor-β (TGF-β) or related growth, differentiation or morphogenesis factor. In this respect, the CCR-proteins of the present invention may also function to prevent de-differentiation of cells/tissue. Other biological activities of the subject CCR-proteins are described herein or will be reasonably apparent to those skilled in the art in light of the present disclosure.

Moreover, it will be generally appreciated that, under certain circumstances, it will be advantageous to provide homologs of naturally-occurring forms of particular CCR-proteins which are either agonists or antagonists of only a subset of that protein's biological activities. Thus, specific biological effects can be elicited by treatment with a homolog of limited function, and with fewer side effects relative to treatment with agonists or antagonists which are directed to all of the biological activities of that protein. For example, p16 homologs can be generated which bind to and inhibit activation of CDK4 without substantially interfering with the activation of CDK6.

In one embodiment, the nucleic acid of the invention encodes a peptide which is an agonist or antagonist of the p16 protein and comprises an amino acid sequence shown in SEQ ID No. 2. Preferred nucleic acids encode a peptide having a p16 protein activity and being at least 60% homologous, more preferably 70% homologous and most preferably 80% homologous with an amino acid sequence shown in SEQ ID No. 2. Nucleic acids which encode peptides having an activity of a p16 protein and having at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homology with a sequence shown in SEQ ID No. 2 are also within the scope of the invention.

In another embodiment, the nucleic acid of the invention encodes a peptide which is an agonist or antagonist of the p15 protein and comprises an amino acid sequence shown in SEQ ID No. 4. Preferred nucleic acids encode a peptide having a p15 protein activity and being at least 60% homologous, more preferably 70% homologous and most preferably 80% homologous with an amino acid sequence shown in one or both of SEQ ID Nos. 4 and 8. Nucleic acids which encode peptides having an activity of a p15 protein and having at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homology with a sequence shown in SEQ ID No. 4 or 8 are also within the scope of the

invention. In a representative embodiment, the nucleic acid is a cDNA molecule comprising at least a portion of the nucleotide sequence encoding a p15 protein shown in SEQ ID No. 3 or 7. A preferred portion of the cDNA molecule shown in SEQ ID No. 3 or 7 includes the coding region of the molecule.

In yet another embodiment, the nucleic acid of the invention encodes a peptide having an activity of a p13.5 protein and comprising an amino acid sequence shown in SEQ ID No. 6. Preferred nucleic acids encode a peptide having a p13.5 protein activity and being at least 60% homologous, more preferably 70% homologous and most preferably 80% homologous with an amino acid sequence shown in SEQ ID No. 6. Nucleic acids which encode peptides having an activity of a p13.5 protein, such as the ability to bind a CDK, and having at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homology with a sequence shown in SEQ ID No. 6 are also within the scope of the invention. Preferably, the nucleic acid is a cDNA molecule comprising at least a portion of the nucleotide sequence encoding a p13.5 protein shown in SEQ ID No. 5. A preferred portion of the cDNA molecule shown in SEQ ID No. 5 includes the coding region of the molecule.

Another aspect of the invention provides a nucleic acid which hybridizes under high or low stringency conditions to a nucleic acid which encodes a CCR polypeptide having all or a portion of an amino acid sequence shown in one of SEQ ID Nos. 2, 4, 6 or 8. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C.

Isolated nucleic acids which differ from the nucleotide sequences shown in one of SEQ ID Nos. 1, 3, 5 or 7 due to degeneracy in the genetic code are also within the scope of the invention. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC are synonyms for histidine) may result in "silent" mutations which do not affect the amino acid sequence of the protein. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject CCR-proteins will exist among eukaryotic cells. One skilled in the art will appreciate that these variations in one or more nucleotides (up to about 3-4% of the nucleotides) of the nucleic acids encoding a particular member of the CCR-protein family may exist among individuals of a given species due to natural allelic variation. Any and all such nucleotide variations and resulting amino acid polymorphisms are within the scope of this invention.

Fragments of the nucleic acid encoding a biologically active portion of the subject CCR-proteins are also within the scope of the invention. As used herein, a fragment of the nucleic acid encoding an active portion of a CCR-protein refers to a nucleotide sequence having fewer nucleotides than the nucleotide sequence encoding the full length amino acid sequence of, for example, the CCR-proteins represented in SEQ ID Nos. 2, 4, 6 or 8, and which encodes a peptide which retains at least a portion of the biological activity of the full-length protein (i.e., a peptide capable of binding a CDK) as defined herein, or alternatively, which is functional as an antagonist of the biological activity of the full-length protein. Nucleic acid fragments within the scope of the invention include those capable of hybridizing under high or low stringency conditions with nucleic acids from other species, e.g. for use in screening protocols to detect homologs of the subject CCR-proteins. Nucleic acids within the scope of the invention may also contain linker sequences, modified restriction endonuclease sites and other sequences useful for molecular cloning, expression or purification of such recombinant peptides.

As indicated by the examples set out below, a nucleic acid encoding a peptide having an activity of a CCR-protein may be obtained from mRNA or genomic DNA present in any of a number of eukaryotic cells in accordance with protocols described herein, as well as those generally known to those skilled in the art. A cDNA encoding a CCR-protein, for example, can be obtained by isolating total mRNA from a cell, e.g. a mammalian cell, e.g. a human cell. Double stranded cDNAs can then be prepared from the total mRNA, and subsequently inserted into a suitable plasmid or bacteriophage vector using any one of a number of known techniques. A gene encoding a CCR-protein can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. Alternatively, the two hybrid assay described in the appended examples can be used to identify other p16 homologs, e.g. which have ankyrin-like repeats and bind to a cyclin dependent kinase, such as CDK6.

A preferred nucleic acid is a cDNA homologous to a sequence shown in SEQ ID No. 1. Another preferred nucleic acid is a cDNA including a sequence shown in SEQ ID No. 3. Yet another preferred nucleic acid is a cDNA including a sequence shown in SEQ ID No. 5.

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridizes (e.g. binds) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding a CCR-protein so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques generally

employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a CCR-protein. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences encoding one of the subject CCR proteins. Such oligonucleotide probes are preferably modified oligonucleotide which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and is therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by van der krol et al. (1988) *Biotechniques* 6:958-976; and Stein et al. (1988) *Cancer Res* 48:2659-2668.

Accordingly, the modified oligomers of the invention are useful in therapeutic, diagnostic, and research contexts. In therapeutic applications, the oligomers are utilized in a manner appropriate for antisense therapy in general. For such therapy, the oligomers of the invention can be formulated for a variety of modes of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remmington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneuos for injection, the oligomers of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by transmucosal or transdermal means, or the compounds can be administered orally. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligomers are formulated into conventional oral administration forms such as capsules, tablets, and tonics. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

Preferred antisense oligonucleotides contain all or a unique fragment of a CCR gene. The fragment can be, for example, 10, 20, 50, 75 or 100 nucleotides in length. The uniqueness of the sequence selected can be relative to all other non-CCR genes, e.g., the antisense nucleic acid will hybridizes under physiological conditions to several CCR genes/messages, or the construct can be selected to uniquely hybridizes to only one or a specific subset of CCR genes, e.g., by chosing a sequence unique even among the various CCR genes.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target DNA or RNA sequences to which they specifically bind. Such diagnostic tests are described in further detail below.

This invention also provides expression vectors comprising a nucleotide sequence encoding a subject cell-cycle regulatory protein and operably linked to at least one regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of the peptide having an activity of a CCR-protein. Accordingly, the term regulatory sequence includes promoters, enhancers and other expression control elements. Exemplary regulatory sequences are described in Goeddel; *Gene Expression Technology*: *Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences-sequences that control the expression of a DNA sequence when operatively linked to it may be used in these vectors to express DNA sequences encoding the CCR-proteins of this invention. Such useful expression control sequences, include, for example, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage lambda , the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α-mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed. Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered.

As will be apparent, the subject gene constructs can be used to cause expression of the subject CCR-proteins in cells propagated in culture, e.g. to produce proteins or peptides, including fusion proteins or peptides, for purification. In addition, recombinant expression of the subject CCR-proteins, or agonist forms thereof, in cultured cells can be useful for

preventing de-differentiation of cells *in vitro*. To illustrate, *in vitro* neuronal culture systems have proved to be fundamental and indispensable tools for the study of neural development, as well as the identification of neurotrophic factors. Once a neuronal cell has become terminally-differentiated, it typically will not change to another terminally differentiated cell-type. However, neuronal cells can nevertheless readily lose their differentiated state. This is commonly observed when they are grown in culture from adult tissue, and when they form a blastema during regeneration.

Agonizing the function of the subject CCR-proteins, such as by maintaining expression (or overexpression) of p16 or p15, provides a means for ensuring an adequately restrictive environment in order to maintain neuronal cells at various stages of differentiation, and can be employed, for instance, in cell cultures designed to test the specific activities of trophic factors. Other tissue culture systems which require maintenance of differentiation will be readily apparent to those skilled in the art. In this respect, each of the agonist and antagonist of CCR inhibition of CDK4 and CDK6 can be used for *ex vivo* tissue generation, as for example, to enhance the generation of prosthetic cartilage devices for implantation.

Conversely, by antagonizing the activity of the wild-type CCR-proteins, such as by expression of antagonistic homologs, antisense constructs, or treatment with agents able to disrupt binding of a CCR-protein with a CDK, the cultured cells can be prevented from following certain differentiative pathways, and, importantly, can cause transformation of cells in culture. In similar fashion, the dominant negative CDK4 and CDK6 mutants described below can be used to cause cellular transformation, as each of these CDK mutants is insensitive to p16 inhibition and is effectively an antagonist of p16. The ability of CCR antagonists to promote cell growth is particularly significant in light of the observation that human cells are notoriously difficult to grow *in vitro*. Accordingly, such reagents are therefore useful for transforming, and in certain instances, immortalizing, cells from primary cell cultures.

Moreover, the subject gene constructs can also be utilized in diagnostic assays to determine if a cell's growth is no longer dependent on the regulatory function of a CCR-protein, e.g. in determining the phenotype of a transformed cell. To illustrate, a sample of cells from the tissue can be obtained from a patient and dispersed in appropriate cell culture media, a portion of the cells in the sample can be caused to express a recombinant CCR-protein, e.g. by transfection with a p16, p15 or p13.5 expression vector, and subsequent growth of the cells assessed. The ability of cells to proliferate despite expression of the CCR-protein is indicative of a lack of dependence on cell regulatory pathways which include the CCR-protein, e.g. RB-mediated checkpoints. Depending on the nature of the tissue of interest, the sample can be in the form of cells isolated from, for example, a blood sample, an exfoliated cell sample, a fine needle aspirant sample, or a biopsied tissue sample. Where the initial sample is a solid mass, the tissue sample can be minced or otherwise dispersed so that

cells can be cultured, as is known in the art. Such knowledge can have both prognostic and therapeutic benefits.

This invention also pertains to a host cell transfected with a recombinant CCR-gene in order to express a polypeptide having an activity of a CCR-protein. The host cell may be any prokaryotic or eukaryotic cell. For example, a CCR-protein of the present invention may be expressed in bacterial cells such as *E. coli*, insect cells (baculovirus), yeast, or mammalian cells. Other suitable host cells are known to those skilled in the art.

Another aspect of the present invention concerns recombinant CCR-proteins which are encoded by genes derived from eukaryotic organisms, e.g. mammals, e.g. humans, and which have at least one biological activity of a CCR-protein, or which are naturally occurring mutants thereof. The term "recombinant protein" refers to a protein of the present invention which is produced by recombinant DNA techniques, wherein generally DNA encoding the CCR-protein is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. Moreover, the phrase "derived from", with respect to a recombinant gene encoding the recombinant CCR-protein, is meant to include within the meaning of "recombinant protein" those proteins having an amino acid sequence of a native CCR-protein, or an amino acid sequence similar thereto which is generated by mutations including substitutions and deletions of a naturally occurring CCR-protein of a organism. To illustrate, recombinant proteins preferred by the present invention, in addition to native p15 or p13.5 proteins, are those recombinantly produced proteins which are at least 60% homologous, more preferably 70% homologous and most preferably 80% homologous with an amino acid sequence shown in SEQ ID No. 2, 4, 6 or 8. Polypeptides having an activity of a CCR- protein, such as CDK-binding, and having at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homology with a sequence shown in SEQ ID No. 2, 4, 6 or 8 are also within the scope of the invention. In other preferred embodiments, the recombinant CCR-protein can be characterized by ankyrin-like repeats and the ability to bind to a cyclin dependent kinase, such as CDK4 or CDK6. Thus, the present invention pertains to recombinant CCR-proteins which are encoded by genes derived from a organism and which have amino acid sequences evolutionarily related to a CCR-protein represented by one of ID No. 2, 4, 6 or 8, wherein "evolutionarily related to", refers to CCR-proteins having amino acid sequences which have arisen naturally (e.g. by allelic variance or by differential splicing), as well as mutational variants of CCR-proteins which are derived, for example, by combinatorial mutagenesis.

The present invention further pertains to methods of producing the subject CCR-proteins. For example, a host cell transfected with expression vector encoding one of the subject CCR-protein can be cultured under appropriate conditions to allow expression of the peptide to occur. The peptide may be secreted and isolated from a mixture of cells and medium containing the peptide. Alternatively, the peptide may be retained cytoplasmically

and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. The peptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for particular epitopes of the subject CCR-proteins. In a preferred embodiment, the CCR-protein is a fusion protein containing a domain which facilitates its purification, such as p16-GST, p15-GST, or a p13.5-GST fusion proteins.

Thus, a nucleotide sequence derived from the cloning of a CCR-protein of the present invention, encoding all or a selected portion of the protein, can be used to produce a recombinant form of the protein via microbial or eukaryotic cellular processes. Ligating the polynucleotide sequence into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells), are standard procedures used in producing other well-known proteins, e.g. insulin, interferons, human growth hormone, IL-1, IL-2, and the like. Similar procedures, or modifications thereof, can be employed to prepare recombinant CCR-proteins, or portions thereof, by microbial means or tissue-culture technology in accord with the subject invention.

The recombinant CCR-protein can be produced by ligating the cloned gene, or a portion thereof, into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vehicles for production of a recombinant CCR-protein include plasmids and other vectors. For instance, suitable vectors for the expression of the subject polypeptides include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as *E. coli*.

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into *S. cerevisiae* (see, for example, Broach *et al.* (1983) in *Experimental Manipulation of Gene Expression*, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in *E. coli* due the presence of the pBR322 ori, and in *S. cerevisiae* due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used.

The preferred mammalian expression vectors contain both prokaryotic sequences to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pcDNAI/amp, pcDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic

cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papilloma virus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. Examples of other viral (including retroviral) expression systems can be found below in the description of gene therapy delivery systems. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press:1989) Chapters 16 and 17. In some instances, it may be desirable to express the recombinant CCR-protein by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the ß-gal containing pBlueBac III).

When expression of a carboxy terminal fragment of the full-length CCR-protein is desired, i.e. a truncation mutant, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from *E. coli* (Ben-Bassat et al. (1987) *J. Bacteriol.* 169:751-757) and *Salmonella typhimurium* and its *in vitro* activity has been demonstrated on recombinant proteins (Miller et al. (1987) *PNAS 84*:2718-1722). Therefore, removal of an N-terminal methionine, if desired, can be achieved either *in vivo* by expressing such recombinant polypeptides in a host which produces MAP (e.g., *E. coli* or CM89 or *S. cerevisiae*), or *in vitro* by use of purified MAP (e.g., procedure of Miller et al.).

Alternatively, the coding sequences for the polypeptide can be incorporated as a part of a fusion gene including a nucleotide sequence encoding a different polypeptide. This type of expression system can be useful under conditions where it is desirable to produce an immunogenic fragment of one of the subject CCR-proteins. For example, the VP6 capsid protein of rotavirus can be used as an immunologic carrier protein for portions of polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequences corresponding to the portion of the CCR protein to which antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising a portion of the protein as part of the virion. The Hepatitis B surface antigen can also be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of a CCR-protein and the poliovirus capsid protein can be created to

enhance immunogenicity (see, for example, EP Publication No. 0259149; and Evans et al. (1989) *Nature* 339:385; Huang et al. (1988) *J. Virol.* 62:3855; and Schlienger et al. (1992) *J. Virol.* 66:2).

The Multiple Antigen Peptide system for peptide-based immunization can be utilized, wherein a desired portion of a CCR-protein is obtained directly from organo-chemical synthesis of the peptide onto an oligomeric branching lysine core (see, for example, Posnett et al. (1988) *JBC* 263:1719 and Nardelli et al. (1992) *J. Immunol.* 148:914). Antigenic determinants of the CCR-proteins can also be expressed and presented by bacterial cells.

In addition to utilizing fusion proteins to enhance immunogenicity, it is widely appreciated that fusion proteins can also facilitate the expression of proteins. For example, the CCR-protein of the present invention can be generated as a glutathione-S-transferase (GST) fusion proteins. Such GST fusion proteins can be used to simply purification of the CCR-protein, such as through the use of glutathione-derivatized matrices (see, for example, *Current Protocols in Molecular Biology*, eds. Ausabel et al. (N.Y.: John Wiley & Sons, 1991)).

In another embodiment, a fusion gene coding for a purification leader sequence, such as a poly-(His)/enterokinase cleavage site sequence at the N-terminus of the desired portion of the recombinant protein, can allow purification of the expressed fusion protein by affinity chromatography using a Ni^{2+} metal resin. The purification leader sequence can then be subsequently removed by treatment with enterokinase to provide the purified CCR-protein (e.g., see Hochuli et al. (1987) *J. Chromatography* 411:177; and Janknecht et al. *PNAS* 88:8972).

Techniques for making fusion genes are well known. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992).

The present invention also makes available isolated and/or purified forms of the subject CCR-proteins, which are isolated from, or otherwise substantially free of other extracellular proteins, especially cell-cycle proteins, e.g. CDKs, cyclins, p21, p19, or PCNA, normally associated with the CCR-protein. The term "substantially free of other cellular proteins" (also referred to herein as "contaminating proteins") is defined as encompassing, for

example, p16, p15 or p13.5 preparations comprising less than 20% (by dry weight) contaminating protein, and preferably comprises less than 5% contaminating protein. Functional forms of the CCR-proteins can be prepared, for the first time, as purified preparations by using a cloned gene as described herein. By "purified", it is meant, when referring to a polypeptide, that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins (particularly other cell-cycle proteins such as CDK4 or CDK6, as well as other contaminating proteins). The term "purified" as used herein preferably means at least 80% by dry weight, more preferably in the range of 95-99% by weight, and most preferably at least 99.8% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (e.g., in an acrylamide gel) but not obtained either as pure (e.g. lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, e.g. acrylamide or agarose) substances or solutions.

However, the subject polypeptides can also be provided in pharmaceutically acceptable carriers for formulated for a variety of modes of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remmington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. In an exemplary embodiment, the CCR-polypeptide is provided for transmucosal or transdermal delivery. For such administration, penetrants appropriate to the barrier to be permeated are used in the formulation with the CCR polypeptide. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

Another aspect of the invention related to peptides derived from the full-length CCR proteins. Isolated peptidyl portions of the subject CCR-proteins can be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a CCR-protein of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of, for example, CDK4 activation, such as by microinjection

assays. In an illustrative embodiment, peptidyl portions of the subject CCR proteins can tested for CDK-binding activity, as well as inhibitory ability, by expression as, for example, thioredoxin fusion proteins each of which contains a discrete fragment of the CCR-protein (see, for example, U.S. Patents 5,270,181 and 5,292,646; and PCT publication WO94/ 02502).

It is also possible to modify the structure of the subject CCR-proteins for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*). Such modified polypeptides, when designed to retain at least one activity of the naturally-occurring form of the protein, are considered functional equivalents of the cell-cycle regulatory proteins described in more detail herein. Such modified polypeptides can be produced, for instance, by amino acid substitution, deletion, or addition.

Moreover, it is reasonable to expect, for example, that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. conservative mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur -containing = cysteine and methionine. (see, for example, *Biochemistry*, 2nd ed, Ed. by L. Stryer, WH Freeman and Co.:1981). Whether a change in the amino acid sequence of a peptide results in a functional homolog can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein. For instance, such variant forms of p16 can be assessed for their ability to complement a p16-deficient cell. Peptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method of generating sets of combinatorial mutants of the present CCR-proteins, as well as truncation mutants, and is especially useful for identifying potential variant sequences (e.g. homologs) that are functional in binding to a CDK, especially CDK4 or CDK6. The purpose of screening such combinatorial libraries is to generate, for example, novel p16, p15 or p13.5 homologs which can act as either agonists

or antagonist, or alternatively, possess novel activities all together. To illustrate, p16 and/or p15 homologs can be engineered by the present method to provide more efficient binding to CDK4, yet have a significantly reduced binding affinity for CDK6 relative to the naturally-occurring form of the protein. Thus, combinatorially-derived homologs can be generated which have a selective potency relative to a naturally occurring CCR-protein. Such proteins, when expressed from recombinant DNA constructs, can be used in gene therapy protocols.

Likewise, mutagenesis can give rise to CCR homologs which have intracellular half-lives dramatically different than the corresponding wild-type protein. For example, the altered protein can be rendered either more stable or less stable to proteolytic degradation or other cellular process which result in destruction of, or otherwise inactivation of, the CCR-protein. Such homologs, and the genes which encode them, can be utilized to alter the envelope of p16, p15 or p13.5 expression by modulating the half-life of the protein. For instance, a short half-life can give rise to more transient biological effects and, when part of an inducible expression system, can allow tighter control of recombinant CCR-protein levels within the cell. As above, such proteins, and particularly their recombinant nucleic acid constructs, can be used in gene therapy protocols.

In similar fashion, CCR homologs can be generated by the present combinatorial approach to act as antagonists, in that they are able to interfere with the ability of the corresponding wild-type protein to regulate cell proliferation.

In a representative embodiment of this method, the amino acid sequences for a population of CCR-protein homologs are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, p16 and/or p15 homologs from one or more species, or homologs from the same species but which differ due to mutation. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. The presence or absence of amino acids from an aligned sequence of a particular variant is relative to a chosen consensus length of a reference sequence, which can be real or artificial. In order to maintain the highest homology in alignment of sequences, deletions in the sequence of a variant relative to the reference sequence can be represented by an amino acid space (*), while insertional mutations in the variant relative to the reference sequence can be disregarded and left out of the sequence of the variant when aligned.

To illustrate, upon inspection of the p16, p15 and p13.5 sequences (see Figure 6), it was noted that an internal fragment of the three CCR-proteins was highly conserved. Based on these alignments, combinatorial libraries can be generated from this portion, the members of which can be expressed in the absence of other portions of a CCR-protein, or as a part of a CCR-protein in which other portions of the protein are static (e.g. a p16 or p15 protein in which only residues Met-52 to Gly-135 or Met-54 to Gly-137, respectively, are varied by combinatorial mutagenesis). For instance, a library of CCR homologs can be generated based

on the sequence of the human p16 and p15 proteins so as to have an amino acid sequence represented by the degenerate formula:

Met-Met-Met-Gly-Xaa(1)-Xaa(2)-Xaa(3)-Val-Ala-Xaa(4)-Leu-Leu-Leu-Xaa(5)-Xaa(6)-Gly-Ala-Xaa(7)-Xaa(8)-Asn-Cys-Xaa(9)-Asp-Pro-Xaa(10)-Thr-Xaa(11)-Xaa(12)-Xaa(13)-Arg-Pro-Val-His-Asp-Ala-Ala-Arg-Glu-Gly-Phe-Leu-Asp-Thr-Leu-Val-Val-Leu-His-Xaa(14)-Xaa(15)-Gly-Ala-Arg-Leu-Asp-Val-Arg-Asp-Ala-Trp-Gly-Arg-Leu-Pro-Xaa(16)-Asp-Leu-Ala-Xaa(17)-Glu-Xaa(18)-Gly-His-Xaa(19)-Asp-Xaa(20)-Xaa(21)-Xaa(22)-Tyr-Leu-Arg-Xaa(23)-Ala-Xaa(24)-Gly

wherein each of Xaa(1)-Xaa(24) is selected from one of the amino acid residues of the same position in SEQ ID No. 2, 4, 6 or 8.

Further expansion of the combinatorial library can be made by, for example, by including amino acids which would represent conservative mutations at one or more of the degenerate positions. Inclusion of such conservative mutations can give rise to a library of potential cell-cycle regulatory sequences represented by the above formula, but wherein Xaa(1) represents Ser, Thr, Asn or Gln; Xaa(2) represents Gly, Ala, Val, Leu, or Ile; Xaa(3) represents Arg, Lys or His; Xaa(4) represents Gly, Ala, Val, Leu, Ile, Asp or Glu; Xaa(5) represents Gly, Ala, Val, Leu, Ile, Asn or Gln; Xaa(6) represents Arg, Lys, His, Tyr or Phe; Xaa(7) represents Asp or Glu; Xaa(8) represents Pro, Gly, Ser or Thr; Xaa(9) represents Gly, Ala, Val, Leu, Ile, Asp or Glu; Xaa(10) represents Gly, Ala, Val, Leu, Ile, or an amino acid gap; Xaa(11) represents Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(12) represents Phe, Tyr, Trp or an amino acid gap; Xaa(13) represents Ser or Thr; Xaa(14) represents Gly, Ala, Val, Leu, Ile, Arg, Lys or His; Xaa(15) represents Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(16) represents Gly, Ala, Val, Leu or Ile; Xaa(17) represents Glx; Xaa(18) represents Gly, Ala, Val, Leu, Ile, Lys, His or Arg; Xaa(19) represents Arg or Gln; Xaa(20) represents Gly, Ala, Val, Leu or Ile; Xaa(21) represents Gly, Ala, Val, Leu or Ile; Xaa(22) represents Gly, Ala, Val, Leu, Ile, Lys, His or Arg; Xaa(23) represents Gly, Ala, Val, Leu, Ile, Thr or Ser; Xaa(24) represents Gly, Ala, Val, Leu, Ile, Ser, Thr or an amino acid gap, where in this context, an amino acid gap is understood to mean the deletion of that amino acid position from the polypeptide. Alternatively, amino acid replacement at degenerate positions can be based on steric criteria, e.g. isosteric replacement, without regard for polarity or charge of amino acid sidechains. Similarly, completely random mutagenesis of one or more of the variant positions (Xaa) can be carried out.

In a preferred embodiment, the combinatorial CCR library is produced by way of a degenerate library of genes encoding a library of polypeptides which each include at least a portion of potential CCR-protein sequences. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential CCR nucleotide sequences are expressible as individual polypeptides, or

alternatively, as a set of larger fusion proteins (e.g. for phage display) containing the set of CCR protein sequences therein.

There are many ways by which the library of potential CCR homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then be ligated into an appropriate gene for expression. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential CCR sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations, and, for that matter, for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of CCR homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the illustrative assays described below are amenable to high through-put analysis as necessary to screen large numbers of degenerate sequences created by combinatorial mutagenesis techniques.

In an illustrative embodiment of a screening assay, the candidate combinatorial gene products are displayed on the surface of a cell, and the ability of particular cells or viral particles to bind a CDK, such as CDK4 or CDK6, via this gene product is detected in a "panning assay". For instance, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140), and the resulting fusion protein detected by panning, e.g. using a fluorescently labeled molecule which binds the CCR-protein, e.g. FITC-CDK4, to score for potentially functional CCR homologs. Cells can be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, separated by a fluorescence-activated cell sorter.

In similar fashion, the gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd, and f1 are most often used in phage display libraries, as either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

In an illustrative embodiment, the recombinant phage antibody system (RPAS, Pharmacia Catalog number 27-9400-01) can be easily modified for use in expressing and screening CCR combinatorial libraries of the present invention. For instance, the pCANTAB 5 phagemid of the RPAS kit contains the gene which encodes the phage gIII coat protein. The CCR combinatorial gene library can be cloned into the phagemid adjacent to the gIII signal sequence such that it will be expressed as a gIII fusion protein. After ligation, the phagemid is used to transform competent *E. coli* TG1 cells. Transformed cells are subsequently infected with M13KO7 helper phage to rescue the phagemid and its candidate CCR-gene insert. The resulting recombinant phage contain phagemid DNA encoding a specific candidate CCR-protein, and display one or more copies of the corresponding fusion coat protein. The phage-displayed candidate proteins which are capable of, for example, binding a CDK, are selected or enriched by panning. For instance, the phage library can be panned on glutathione immobilized CDK-GST fusion proteins, and unbound phage washed away from the cells. The bound phage is then isolated, and if the recombinant phage express at least one copy of the wild type gIII coat protein, they will retain their ability to infect *E. coli*. Thus, successive rounds of reinfection of *E. coli*, and panning will greatly enrich for CCR homologs, e.g. p16, p15 or p13.5 homologs, which can then be screened for further biological activities in order to differentiate agonists and antagonists. Subsequent selection, e.g. of a reduced set of variants from the library, may then be based upon more meaningful criteria rather than simple CDK-binding ability. For instance, intracellular half-life or inhibitory potency can become selection criteria in secondary screens.

In light of the present disclosure, other forms of mutagenesis generally applicable will be apparent to those skilled in the art in addition to the aforementioned combinatorial mutagenesis. For example, p16, p15 or p13.5 homologs (both agonist and antagonist forms,

and including fragments) can be generated and screened using, for example, alanine scanning mutagenesis and the like (Ruf et al. (1994) *Biochemistry* 33:1565-1572; Wang et al. (1994) *J. Biol. Chem.* 269:3095-3099; Balint et al. (1993) *Gene* 137:109-118; Grodberg et al. (1993) *Eur. J. Biochem.* 218:597-601; Nagashima et al. (1993) *J. Biol. Chem.* 268:2888-2892; Lowman et al. (1991) *Biochemistry* 30:10832-10838; and Cunningham et al. (1989) *Science* 244:1081-1085), by linker scanning mutagenesis (Gustin et al. (1993) *Virology* 193:653-660; Brown et al. (1992) *Mol. Cell Biol.* 12:2644-2652; McKnight et al. (1982) *Science* 232:316); by saturation mutagenesis (Meyers et al. (1986) *Science* 232:613); by PCR mutagenesis (Leung et al. (1989) *Method Cell Mol Biol* 1:11-19); or by random mutagenesis (Miller et al. (1992) *A Short Course in Bacterial Genetics*, CSHL Press, Cold Spring Harbor, NY; and Greener et al. (1994) *Strategies in Mol Biol* 7:32-34).

Consequently, the invention also provides for reduction of the subject CCR-proteins to generate mimetics, e.g. peptide or non-peptide agents, which are able to mimic binding of the authentic CCR protein to a cyclin dependent kinase, e.g. CDK4 and/or CDK6. Such mutagenic techniques as described above, as well as the thioredoxin system, are also particularly useful for mapping the determinants of a CCR-protein which participate in protein-protein interactions involved in, for example, binding of the subject CCR-protein to a CDK. To illustrate, the critical residues of a subject CCR-protein which are involved in molecular recognition of CDK4 can be determined and used to generate CCR-derived peptidomimetics which bind to CDK4 or CDK6 and, like the authentic CCR-protein, inhibit activation of the kinase. By employing, for example, scanning mutagenesis to map the amino acid residues of a particular CCR-protein involved in binding a cyclin dependent kinase, peptidomimetic compounds (e.g. diazepine or isoquinoline derivatives) can be generated which mimic those residues in binding to the kinase. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gama lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), β-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231), and β-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

In similar fashion, identification of mutations in CDK4 and/or CDK6 which effect binding to a CCR-protein can be used to identify potential peptidyl fragments of

CDK4/CDK6 which can competitively bind a CCR-protein and interfere with its ability to inhibit the CDK. As describe below, we have characterized mutations to CDK4 which abrogate binding by p15 and p16, and consequently provide kinases that are insensitive to inhibition by those CCR-proteins. These mutations, in fact, occur in stretches of amino acid residues which are conserved between CDK4 and CDK6. Accordingly, peptidomimetics based on these portions of CDK4/CDK6 might be useful as antagonists of CCR-proteins, in that they are expected to compete with CDK4 or CDK6 for binding to the CCR-protein. In a preferred embodiment, the CCR antagonist is a peptide or non-peptide analog of the amino acid sequence VAEIG(V/E)GAYG(T/K)-V(F/Y)KARD, though more preferably a peptidomimetic of the amino acid sequence V(F/Y)KARD, and even more preferably of the tetrapeptide KARD. Likewise, another preferred peptide analog of CDK4/CDK6 comprises a peptide or peptidomimetic corresponding to the sequence SRTDRE(I/T)K(V/L)TLVFEHVD-QDL(R/T)TYLDK(A/V)PPPG(L/V), though more preferably the peptidomimetic is an analog of the hexa-peptides FEHVDQ or EHVDQD, and even more preferably of the tetrapeptides HVDQ or EHVD. Non-hydrolyzable peptide analogs of such residues can be generated using, for example, benzodiazepine, azepine, substituted gama lactam rings, keto-methylene pseudopeptides, β-turn dipeptide cores, or β-aminoalcohols. These and other peptidyl portions of CDKs can be tested for binding to CCR-proteins such as p15 or p16 using, for example, the thioredoxin fusion proteins constructs mention above.

Another aspect of the invention pertains to an antibody specifically reactive with one of the subject CCR-proteins. For example, by using peptides based on the cDNA sequence of the subject p16 protein, anti-p16 antisera or anti-p16 monoclonal antibodies can be made using standard methods. Likewise, anti-p13.5 and anti-p15 antibodies can be generated. A mammal such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide (e.g., an antigenic fragment which is capable of eliciting an antibody response). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. For instance, a peptidyl portion of the protein represented by one of SEQ ID No. 2, 4, 6 or 8 can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies.

Following immunization, anti-CCR antisera can be obtained and, if desired, polyclonal anti-CCR antibodies isolated from the serum. To produce monoclonal antibodies, antibody producing cells (lymphocytes) can be harvested from an immunized animal and fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, an include, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) *Nature*, 256: 495-497), as the human B cell hybridoma technique (Kozbar et al., (1983)

Immunology Today, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., (1985) *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. pp. 77-96). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with the CCR-protein of interest and the monoclonal antibodies isolated.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with a CCR-protein, e.g. anti-p16, anti-p15 or anti-p13.5 antibodies. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab')_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab')_2$ fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the present invention is further intended to include bispecific and chimeric molecules.

Both monoclonal and polyclonal antibodies (Ab) directed against the subject CCR-proteins, and antibody fragments such as Fab` and $F(ab`)_2$, can be used to block the action of particular CCR and allow the study of the cell-cycle or cell proliferation.

One application of anti-CCR antibodies is in the immunological screening of cDNA libraries constructed in expression vectors, such as λgt11, λgt18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of ß-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a CCR-protein, such as proteins antigenically related to p16, p15 or p13.5, can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with an anti-CCR antibody. Phage, scored by this assay, can then be isolated from the infected plate. Thus, the presence of CCR homologs, such as p16, p15 or p13.5 homologs, can be detected and cloned from other sources.

Antibodies which are specifically immunoreactive with one or more CCR-proteins of the present invention can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of the CCR-protein family, or particular members thereof. Anti-CCR antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate levels of one or more CCR-proteins in tissue or cells isolated from a bodily fluid as part of a clinical testing procedure. For instance, such measurements can be useful in predictive valuations of the onset or progression of tumors. Likewise, the ability to monitor certain CCR-protein levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. Diagnostic assays using anti-CCR antibodies, such as anti-p16 or anti-p15 antibodies, can include, for example, immunoassays designed to aid

in early diagnosis of a neoplastic or hyperplastic disorder, e.g. the presence of cancerous cells in the sample, e.g. to detect cells in which a lesion of a CCR-gene has occurred.

In addition, nucleotide probes can be generated from the cloned sequence of the subject CCR-proteins, which allow for histological screening of intact tissue and tissue samples for the presence of a CCR-protein encoding mRNA. Similar to the diagnostic uses of anti-CCR-protein antibodies, the use of probes directed to CCR-protein encoding mRNAs, or to genomic CCR-gene sequences, can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, neoplastic or hyperplastic disorders (e.g. unwanted cell growth). Used in conjunction with anti-CCR protein antibody immunoassays, the nucleotide probes can help facilitate the determination of the molecular basis for a developmental disorder which may involve some abnormality associated with expression (or lack thereof) of a CCR-protein. For instance, variation in CCR-protein synthesis can be differentiated from a mutation in the coding sequence.

Accordingly, the present method provides a method for determining if a subject is at risk for a disorder characterized by unwanted cell proliferation. In preferred embodiments, method can be generally characterized as comprising detecting, in a tissue of said subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a CCR-protein, such as p16, p15 or p13.5 or (ii) the mis-expression of the CCR-gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a CCR-gene, (ii) an addition of one or more nucleotides to a CCR-gene, (iii) a substitution of one or more nucleotides of a CCR-gene, (iv) a gross chromosomal rearrangement of a CCR-gene, (v) a gross alteration in the level of a messenger RNA transcript of a CCR-gene, (vi) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a CCR-gene, and (vii) a non-wild type level of a CCR-protein. In one aspect of the invention, there is provided a probe/primer comprising an oligonucleotide containing a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of any of SEQ ID Nos: 1, 3, 5 or 7 or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject CCR-genes or naturally occurring mutants thereof. The probe is exposed to nucleic acid of a tissue sample; and the hybridization of the probe to the sample nucleic acid is detected. In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1944) *PNAS* 91:360-364), the later of which can be particularly useful for detecting point mutations in the CCR-gene. Alternatively, the level of CCR-protein can detected in an immunoassay.

Moreover, the use of anti-sense techniques (e.g. microinjection of antisense molecules, or transfection with plasmids whose transcripts are anti-sense with regard to a

certain CCR mRNA) can be used to investigate role of particular CCR-proteins, e.g. p16, p15 or p13.5, in the cell-cycle and cell proliferation, in a controlled environment, by inhibiting endogenous production of the protein. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals.

Another aspect of the invention features transgenic non-human animals which express a heterologous CCR-gene of the present invention, or which have had one or more genomic CCR-gene(s), disrupted in at least one of the tissue or cell-types of the animal. For instance, transgenic mice that are disrupted at their p13.5 gene locus are described in Example 5.

In another aspect, the invention features an animal model for developmental diseases, which has a CCR allele which is mis-expressed. For example, a mouse can be bred which has a p16 or p15 allele deleted, or in which all or part of one or more p16 exons are deleted. Such a mouse model can then be used to study disorders arising from mis-expressed p16 genes.

Accordingly, the present invention concerns transgenic animals which are comprised of cells (of that animal) which contain a transgene of the present invention and which preferably (though optionally) express an exogenous CCR-protein in one or more cells in the animal. The CCR-transgene can encode the wild-type form of the protein, or can encode homologs thereof, including both agonists and antagonists, as well as antisense constructs. In preferred embodiments, the expression of the transgene is restricted to specific subsets of cells, tissues or developmental stages utilizing, for example, cis-acting sequences that control expression in the desired pattern. In the present invention, such mosaic expression of the subject proteins can be essential for many forms of lineage analysis and can additionally provide a means to assess the effects of, for example, lack of p16 and p15 inhibition of CDKs which might grossly alter development in small patches of tissue within an otherwise normal embryo. Toward this and, tissue-specific regulatory sequences and conditional regulatory sequences can be used to control expression of the transgene in certain spatial patterns. Moreover, temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences.

Genetic techniques which allow for the expression of transgenes can be regulated via site-specific genetic manipulation *in vivo* are known to those skilled in the art. For instance, genetic systems are available which allow for the regulated expression of a recombinase that catalyzes the genetic recombination a target sequence. As used herein, the phrase "target sequence" refers to a nucleotide sequence that is genetically recombined by a recombinase. The target sequence is flanked by recombinase recognition sequences and is generally either excised or inverted in cells expressing recombinase activity. Recombinase catalyzed recombination events can be designed such that recombination of the target sequence results in either the activation or repression of expression of the subject CCR polypeptide. For example, excision of a target sequence which interferes with the expression of a recombinant

CCR-gene can be designed to activate expression of that gene. This interference with expression of the protein can result from a variety of mechanisms, such as spatial separation of the CCR-gene from the promoter element or an internal stop codon. Moreover, the transgene can be made wherein the coding sequence of the gene is flanked recombinanse recognition sequences and is initially transfected into cells in a 3' to 5' orientation with respect to the promoter element. In such an instance, inversion of the target sequence will reorient the subject gene by placing the 5' end of the coding sequence in an orientation with respect to the promoter element which allow for promoter driven transcriptional activation.

In an illustrative embodiment, either the *cre/loxP* recombinase system of bacteriophage P1 (Lakso et al. (1992) *PNAS* 89:6232-6236; Orban et al. (1992) *PNAS* 89:6861-6865) or the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355; PCT publication WO 92/15694) can be used to generate *in vivo* site-specific genetic recombination systems. Cre recombinase catalyzes the site-specific recombination of an intervening target sequence located between *loxP* sequences. *loxP* sequences are 34 base pair nucleotide repeat sequences to which the Cre recombinase binds and are required for Cre recombinase mediated genetic recombination. The orientation of *loxP* sequences determines whether the intervening target sequence is excised or inverted when Cre recombinase is present (Abremski et al. (1984) *J. Biol. Chem.* 259:1509-1514); catalyzing the excision of the target sequence when the *loxP* sequences are oriented as direct repeats and catalyzes inversion of the target sequence when *loxP* sequences are oriented as inverted repeats.

Accordingly, genetic recombination of the target sequence is dependent on expression of the Cre recombinase. Expression of the recombinase can be regulated by promoter elements which are subject to regulatory control, e.g., tissue-specific, developmental stage-specific, inducible or repressible by externally added agents. This regulated control will result in genetic recombination of the target sequence only in cells where recombinase expression is mediated by the promoter element. Thus, the activation expression of the CCR protein can be regulated via regulation of recombinase expression.

Use of the *cre/loxP* recombinase system to regulate expression of a recombinant CCR protein, such as p15 or p16, requires the construction of a transgenic animal containing transgenes encoding both the Cre recombinase and the subject protein. Animals containing both the Cre recombinase and the recombinant CCR genes can be provided through the construction of "double" transgenic animals. A convenient method for providing such animals is to mate two transgenic animals each containing a transgene, e.g., the CCR-gene and recombinase gene.

One advantage derived from initially constructing transgenic animals containing a CCR-transgene in a recombinase-mediated expressible format derives from the likelihood that the subject protein will be deleterious upon expression in the transgenic animal. In such

an instance, a founder population, in which the subject transgene is silent in all tissues, can be propagated and maintained. Individuals of this founder population can be crossed with animals expressing the recombinase in, for example, one or more tissues. Thus, the creation of a founder population in which, for example, an antagonistic p15 transgene is silent will allow the study of progeny from that founder in which disruption of cell cycle regulation by p15 in a particular tissue or at developmental stages would result in, for example, a lethal phenotype.

Similar conditional transgenes can be provided using prokaryotic promoter sequences which require prokaryotic proteins to be simultaneous expressed in order to facilitate expression of the transgene. Exemplary promoters and the corresponding trans-activating prokaryotic proteins are given in U.S. Patent No. 4,833,080. Moreover, expression of the conditional transgenes can be induced by gene therapy-like methods wherein a gene encoding the trans-activating protein, e.g. a recombinase or a prokaryotic protein, is delivered to the tissue and caused to be expressed, such as in a cell-type specific manner. By this method, the CCR-transgene could remain silent into adulthood until "turned on" by the introduction of the trans-activator.

In an exemplary embodiment, the "transgenic non-human animals" of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonal target cell. The zygote is the best target for micro-injection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter which allows reproducible injection of 1-2pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster et al. (1985) *PNAS* 82:4438-4442). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene. Microinjection of zygotes is the preferred method for incorporating transgenes in practicing the invention.

Retroviral infection can also be used to introduce transgene into a non-human animal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, R. (1976) *PNAS* 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (*Manipulating the Mouse Embryo*, Hogan eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al. (1985) *PNAS* 82:6927-6931; Van der Putten et al. (1985) *PNAS* 82:6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of

virus-producing cells (Van der Putten, *supra*; Stewart et al. (1987) *EMBO J.* 6:383-388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al. (1982) *Nature* 298:623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner et al. (1982) *supra*).

A third type of target cell for transgene introduction is the embryonal stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (Evans et al. (1981) *Nature* 292:154-156; Bradley et al. (1984) *Nature* 309:255-258; Gossler et al. (1986) *PNAS* 83: 9065-9069; and Robertson et al. (1986) *Nature* 322:445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) *Science* 240:1468-1474.

Methods of making knock-out or disruption transgenic animals are also generally known. See, for example, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Recombinase dependent knockouts can also be generated, e.g. by homologous recombination to insert target sequences, such that tissue specific and/or temporal control of inactivation of a CCR-gene can be controlled as above.

Yet another aspect of the invention pertains to methods of treating proliferative and/or differentiative disorders which arise from cells which, despite aberrant growth control, still require CDK4 or CDK6 for cell growth. There are a wide variety of pathological cell proliferative conditions for which the CCR-gene constructs and CCR-mimetics of the present invention can provide therapeutic benefits, with the general strategy being the inhibition of an anomalous cell proliferation. For instance, the gene constructs of the present invention can be used as a part of a gene therapy protocol, such as to reconstitute the function of a CCR-protein, e.g. p16, p15 or p13.5, in a cell in which the protein is misexpressed or in which signal transduction pathways upstream of the CCR-protein are dysfunctional. To illustrate, cell types which exhibit pathological or abnormal growth presumably dependent at least in part on a function of a CCR-protein include various cancers and leukemias, psoriasis, bone diseases, fibroproliferative disorders such as involving connective tissues, atherosclerosis and other smooth muscle proliferative disorders, as well as chronic inflammation. In addition to proliferative disorders, the treatment of differentiative disorders which result from, for example, de-differentiation of tissue which may (optionally) be accompanied by abortive

reentry into mitosis. Such degenerative disorders include chronic neurodegenerative diseases of the nervous system, including Alzheimer's disease, Parkinson's disease, Huntington's chorea, amylotrophic lateral sclerosis and the like, as well as spinocerebellar degenerations. Other differentiative disorders include, for example, disorders associated with connective tissue, such as may occur due to de-differentiation of chondrocytes or osteocytes, as well as vascular disorders which involve de-differentiation of endothelial tissue and smooth muscle cells, gastric ulcers characterized by degenerative changes in glandular cells, and renal conditions marked by failure to differentiate, e.g. Wilm's tumors. It will also be apparent that, by transient use of gene therapy constructs of the subject CCR proteins (e.g. agonist and antagonist forms), *in vivo* reformation of tissue can be accomplished, e.g. in the development and maintenance of organs. By controlling the proliferative and differentiative potential for different cells, the subject gene constructs can be used to reform injured tissue, or to improve grafting and morphology of transplanted tissue. For instance, CCR agonists and antagonists can be employed therapeutically to regulate organs after physical, chemical or pathological insult. For example, gene therapy can be utilized in liver repair subsequent to a partial hepatectomy, or to promote regeneration of lung tissue in the treatment of emphysema.

For instance, as described in the Examples below, transformation of a cell can be due in part to a loss-of-function mutation to a particular CCR-gene, e.g., ranging from a point mutation to gross deletion of the gene. Additionally, other data provided in the appended examples suggests that disorders susceptible to treatment with CCR agonists include those arising from cells which have lost the ability to induce CCR-protein expression. Normal cell proliferation, for instance, is generally marked by responsiveness to negative autocrine or paracrine growth regulators, such as members of the TGF-β family, e.g. TGF-β1, TGF-β2 or TGF-β3, and related polypeptide growth inhibitors, e.g. activins, inhibins, Müllerian inhibiting substance, decapentaplegic, bone morphogenic factors, and vg1 (e.g. terminal differentiation inducers). Ordinarily, control of cellular proliferation by such growth regulators, particularly in epithelial and hemopoietic cells, is in the form of growth inhibition. This is generally accompanied by differentiation of the cell to a post-mitotic phenotype. However, it has been observed that a significant percentage of human cancers derived from these cells types display a reduced responsiveness to growth regulators such as TGF-β. For instance, some tumors of colorectal, liver epithelial, and epidermal origin show reduced sensitivity and resistance to the growth-inhibitory effects of TGF-β as compared to their normal counterparts. In this context, a noteworthy characteristic of several retinoblastoma cell lines is the absence of detectable TGF-β receptors. Treatment of such tumors with CCR agonists (e.g. CCR-proteins delivered by gene therapy or CCR-mimetics) provides an opportunity to restore the function of the RB-mediated checkpoint.

However, it will be appreciated that the subject method can be used to inhibit proliferation of cells which, in general, are still reliant on cyclin dependent kinases active in G_1, e.g. CDK4 or CDK6, irrespective of involvement of RB or RB-like proteins.

In accordance with the subject method, expression constructs of the subject CCR-proteins may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively transfecting cells *in vivo* with a recombinant CCR-gene. Approaches include insertion of the subject gene in viral vectors including recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors can be used to transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramacidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or $CaPO_4$ precipitation carried out *in vivo*. It will be appreciated that because transduction of appropriate target cells represents the critical first step in gene therapy, choice of the particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g. locally or systemically.

A preferred approach for *in vivo* introduction of nucleic acid encoding one of the subject proteins into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA, encoding the gene product. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded within the viral vector, e.g., by a cDNA contained in the viral vector, are expressed efficiently in cells which have taken up viral vector nucleic acid.

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous genes *in vivo*, particularly into humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host. A major prerequisite for the use of retroviruses is to ensure the safety of their use, particularly with regard to the possibility of the spread of wild-type virus in the cell population. The development of specialized cell lines (termed "packaging cells") which produce only replication-defective retroviruses has increased the utility of retroviruses for gene therapy, and defective retroviruses are well characterized for use in gene transfer for gene therapy purposes (for a review see Miller, A.D. (1990) *Blood* 76:271). Thus, recombinant retrovirus can be constructed in which part of the retroviral coding sequence (*gag, pol, env*) has been replaced by nucleic acid encoding one of the subject CCR-proteins, rendering the retrovirus replication defective. The replication defective retrovirus is then packaged into virions which can be used to infect a target cell through the use of a helper virus by standard techniques. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* or *in vivo* with such viruses can be found in Current Protocols in Molecular Biology, Ausubel, F.M. et

al. (eds.) Greene Publishing Associates, (1989), Sections 9.10-9.14 and other standard laboratory manuals. Examples of suitable retroviruses include pLJ, pZIP, pWE and pEM which are well known to those skilled in the art. Examples of suitable packaging virus lines for preparing both ecotropic and amphotropic retroviral systems include ψCrip, ψCre, ψ2 and ψAm. Retroviruses have been used to introduce a variety of genes into many different cell types, including neural cells, epithelial cells, endothelial cells, lymphocytes, myoblasts, hepatocytes, bone marrow cells, *in vitro* and/or *in vivo* (see for example Eglitis, et al. (1985) *Science* 230:1395-1398; Danos and Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464; Wilson et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:3014-3018; Armentano et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6141-6145; Huber et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8039-8043; Ferry et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8377-8381; Chowdhury et al. (1991) *Science* 254:1802-1805; van Beusechem et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:7640-7644; Kay et al. (1992) *Human Gene Therapy* 3:641-647; Dai et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:10892-10895; Hwu et al. (1993) *J. Immunol.* 150:4104-4115; U.S. Patent No. 4,868,116; U.S. Patent No. 4,980,286; PCT Application WO 89/07136; PCT Application WO 89/02468; PCT Application WO 89/05345; and PCT Application WO 92/07573).

In choosing retroviral vectors as a gene delivery system for the subject CCR-genes, it is important to note that a prerequisite for the successful infection of target cells by most retroviruses, and therefore of stable introduction of the recombinant CCR-gene, is that the target cells must be dividing. In general, this requirement will not be a hindrance to use of retroviral vectors to deliver CCR-gene constructs. In fact, such limitation on infection can be beneficial in circumstances wherein the tissue (e.g. nontransformed cells) surrounding the target cells does not undergo extensive cell division and is therefore refractory to infection with retroviral vectors.

Furthermore, it has been shown that it is possible to limit the infection spectrum of retroviruses and consequently of retroviral-based vectors, by modifying the viral packaging proteins on the surface of the viral particle (see, for example PCT publications WO93/25234, WO94/06920, and WO94/11524). For instance, strategies for the modification of the infection spectrum of retroviral vectors include: coupling antibodies specific for cell surface antigens to the viral *env* protein (Roux et al. (1989) *PNAS* 86:9079-9083; Julan et al. (1992) *J. Gen Virol* 73:3251-3255; and Goud et al. (1983) *Virology* 163:251-254); or coupling cell surface ligands to the viral *env* proteins (Neda et al. (1991) *J Biol Chem* 266:14143-14146). Coupling can be in the form of the chemical cross-linking with a protein or other variety (e.g. lactose to convert the *env* protein to an asialoglycoprotein), as well as by generating fusion proteins (e.g. single-chain antibody/*env* fusion proteins). This technique, while useful to limit or otherwise direct the infection to certain tissue types, and can also be used to convert an ecotropic vector in to an amphotropic vector.

Moreover, use of retroviral gene delivery can be further enhanced by the use of tissue- or cell-specific transcriptional regulatory sequences which control expression of the CCR-gene of the retroviral vector.

Another viral gene delivery system useful in the present invention utilizes adenovirus-derived vectors. The genome of an adenovirus can be manipulated such that it encodes a gene product of interest, but is inactivate in terms of its ability to replicate in a normal lytic viral life cycle (see, for example, Berkner et al. (1988) *BioTechniques* 6:616; Rosenfeld et al. (1991) *Science* 252:431-434; and Rosenfeld et al. (1992) *Cell* 68:143-155). Suitable adenoviral vectors derived from the adenovirus strain Ad type 5 dl324 or other strains of adenovirus (e.g., Ad2, Ad3, Ad7 etc.) are well known to those skilled in the art. Recombinant adenoviruses can be advantageous in certain circumstances in that they are not capable of infecting nondividing cells and can be used to infect a wide variety of cell types, including airway epithelium (Rosenfeld et al. (1992) cited *supra*), endothelial cells (Lemarchand et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6482-6486), hepatocytes (Herz and Gerard (1993) *Proc. Natl. Acad. Sci. USA* 90:2812-2816) and muscle cells (Quantin et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:2581-2584). Furthermore, the virus particle is relatively stable and amenable to purification and concentration, and as above, can be modified so as to affect the spectrum of infectivity. Additionally, introduced adenoviral DNA (and foreign DNA contained therein) is not integrated into the genome of a host cell but remains episomal, thereby avoiding potential problems that can occur as a result of insertional mutagenesis in situations where introduced DNA becomes integrated into the host genome (e.g., retroviral DNA). Moreover, the carrying capacity of the adenoviral genome for foreign DNA is large (up to 8 kilobases) relative to other gene delivery vectors (Berkner et al., *supra*; Haj-Ahmand and Graham (1986) *J. Virol.* 57:267). Most replication-defective adenoviral vectors currently in use and therefore favored by the present invention are deleted for all or parts of the viral E1 and E3 genes but retain as much as 80% of the adenoviral genetic material (see, e.g., Jones et al. (1979) *Cell* 16:683; Berkner et al., *supra*; and Graham et al. in Methods in Molecular Biology, E.J. Murray, Ed. (Humana, Clifton, NJ, 1991) vol. 7. pp. 109-127). Expression of the inserted CCR-gene can be under control of, for example, the E1A promoter, the major late promoter (MLP) and associated leader sequences, the E3 promoter, or exogenously added promoter sequences.

Yet another viral vector system useful for delivery of the subject CCR-gene is the adeno-associated virus (AAV). Adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle. (For a review see Muzyczka et al. *Curr. Topics in Micro. and Immunol.* (1992) 158:97-129). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (see for example Flotte et al. (1992) *Am. J. Respir. Cell. Mol. Biol.* 7:349-356; Samulski et al.

(1989) *J. Virol.* 63:3822-3828; and McLaughlin et al. (1989) *J. Virol.* 62:1963-1973). Vectors containing as little as 300 base pairs of AAV can be packaged and can integrate. Space for exogenous DNA is limited to about 4.5 kb. An AAV vector such as that described in Tratschin et al. (1985) *Mol. Cell. Biol.* 5:3251-3260 can be used to introduce DNA into cells. A variety of nucleic acids have been introduced into different cell types using AAV vectors (see for example Hermonat et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6466-6470; Tratschin et al. (1985) *Mol. Cell. Biol.* 4:2072-2081; Wondisford et al. (1988) *Mol. Endocrinol.* 2:32-39; Tratschin et al. (1984) *J. Virol.* 51:611-619; and Flotte et al. (1993) *J. Biol. Chem.* 268:3781-3790).

Other viral vector systems that may have application in gene therapy have been derived from herpes virus, vaccinia virus, and several RNA viruses. In particular, herpes virus vectors may provide a unique strategy for persistence of the recombinant CCR-gene in cells of the central nervous system and ocular tissue (Pepose et al. (1994) Invest Ophthalmol Vis Sci 35:2662-2666)

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of a CCR-protein in the tissue of an animal. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the subject CCR-gene by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In a representative embodiment, a gene encoding one of the subject CCR-proteins can be entrapped in liposomes bearing positive charges on their surface (e.g., lipofectins) and (optionally) which are tagged with antibodies against cell surface antigens of the target tissue (Mizuno et al. (1992) *No Shinkei Geka* 20:547-551; PCT publication WO91/06309; Japanese patent application 1047381; and European patent publication EP-A-43075). For example, lipofection of neuroglioma cells can be carried out using liposomes tagged with monoclonal antibodies against glioma-associated antigen (Mizuno et al. (1992) *Neurol. Med. Chir.* 32:873-876).

In yet another illustrative embodiment, the gene delivery system comprises an antibody or cell surface ligand which is cross-linked with a gene binding agent such as poly-lysine (see, for example, PCT publications WO93/04701, WO92/22635, WO92/20316, WO92/19749, and WO92/06180). For example, the subject CCR-gene construct can be used to transfect hepatocytic cells *in vivo* using a soluble polynucleotide carrier comprising an asialoglycoprotein conjugated to a polycation, e.g. poly-lysine (see U.S. Patent 5,166,320). It will also be appreciated that effective delivery of the subject nucleic acid constructs via -mediated endocytosis can be improved using agents which enhance escape of the gene from

the endosomal structures. For instance, whole adenovirus or fusogenic peptides of the influenza HA gene product can be used as part of the delivery system to induce efficient disruption of DNA-containing endosomes (Mulligan et al. (1993) *Science* 260-926; Wagner et al. (1992) *PNAS* 89:7934; and Christiano et al. (1993) PNAS 90:2122).

In clinical settings, the gene delivery systems can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Patent 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) *PNAS* 91: 3054-3057).

Moreover, the pharmaceutical preparation can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced in tact from recombinant cells, e.g. retroviral packages, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system. In the case of the latter, methods of introducing the viral packaging cells may be provided by, for example, rechargeable or biodegradable devices. Various slow release polymeric devices have been developed and tested *in vivo* in recent years for the controlled delivery of drugs, including proteinacious biopharmaceuticals, and can be adapted for release of viral particles through the manipulation of the polymer composition and form. A variety of biocompatible polymers (including hydrogels), including both biodegradable and non-degradable polymers, can be used to form an implant for the sustained release of an the viral particles by cells implanted at a particular target site. Such embodiments of the present invention can be used for the delivery of an exogenously purified virus, which has been incorporated in the polymeric device, or for the delivery of viral particles produced by a cell encapsulated in the polymeric device.

By choice of monomer composition or polymerization technique, the amount of water, porosity and consequent permeability characteristics can be controlled. The selection of the shape, size, polymer, and method for implantation can be determined on an individual basis according to the disorder to be treated and the individual patient response. The generation of such implants is generally known in the art. See, for example, *Concise Encyclopedia of Medical & Dental Materials*, ed. by David Williams (MIT Press: Cambridge, MA, 1990); and the Sabel et al. U.S. Patent No. 4,883,666. In another embodiment of an implant, a source of cells producing a the recombinant virus is

encapsulated in implantable hollow fibers. Such fibers can be pre-spun and subsequently loaded with the viral source (Aebischer et al. U.S. Patent No. 4,892,538; Aebischer et al. U.S. Patent No. 5,106,627; Hoffman et al. (1990) *Expt. Neurobiol.* 110:39-44; Jaeger et al. (1990) *Prog. Brain Res.* 82:41-46; and Aebischer et al. (1991) *J. Biomech. Eng.* 113:178-183), or can be co-extruded with a polymer which acts to form a polymeric coat about the viral packaging cells (Lim U.S. Patent No. 4,391,909; Sefton U.S. Patent No. 4,353,888; Sugamori et al. (1989) *Trans. Am. Artif. Intern. Organs* 35:791-799; Sefton et al. (1987) *Biotechnol. Bioeng.* 29:1135-1143; and Aebischer et al. (1991) *Biomaterials* 12:50-55). Again, manipulation of the polymer can be carried out to provide for optimal release of viral particles.

To further illustrate the use of the subject method, the therapeutic application of a CCR-gene, e.g., by gene therapy, can be used in the treatment of a neuroglioma. Gliomas account for 40-50% of intracranial tumors at all ages of life. Despite the increasing use of radiotherapy, chemotherapy, and sometimes immunotherapy after surgery for malignant glioma, the mortality and morbidity rates have not substantially improved. However, there is increasing experimental and clinical evidence that for a significant number of gliomas, loss of TGF-β responsiveness is an important event in the loss of growth control. Irrespective of the cause of decreased responsiveness, e.g. the loss of function of p15 or the loss of other TGF-β signal transduction proteins, exogenous expression of p15 (or other CCR-protein) in the cell can used effectively to inhibit cell proliferation.

It has been demonstrated that gene therapy can be used to target glioma cells for expression of recombinant proteins (Miyao et al. (1993) *J. Neurosci. Res.* 36:472-479; Chen et al. (1994) PNAS 91:3054-3057; and Takamiya et al. (1993) *J. Neurosurg.* 79:104-110). Thus, a gene construct for expressing a CCR-protein can be delivered to the tumor, preferably by stereotactic-dependent means. In preferred embodiments, the gene delivery system is a retroviral vector. Since rapidly growing normal cells are rare in the adult CNS, glioma cells can be specifically transduced with a recombinant retrovirus. For example, the retroviral particle can be delivered into the tumor cavity through an Ommaya tube after surgery, or alternatively, packaging fibroblasts encapsulated in retrievable immunoisolatory vehicles can be introduced into the tumor cavity. In order to increase the effectiveness and decrease the side effects of the retrovirus-mediated gene therapy, glioma-specific promoters can be used to regulate expression of the CCR-gene. For example, the promoter regions of glial fibrillary acidic protein (GFAP) and myelin basis protein (MBP) can operably linked to the recombinant gene in order to direct glial cell-specific expression of the recombinant gene construct.

In another embodiment, gene therapy can be used in conjunction with the subject CCR-protein in the treatment of various carcinomas. Among the carcinomas, a strong correlation has been observed between TGF-β responsiveness and both increasing malignant

tumor grade and increasing proliferating cell nuclear antigen index (Coffery et al. (1988) *Cancer Res* 48:1596-1602; Masui et al. (1986) *PNAS* 83:2438-2442; and Knabble et al. (1987) *Cell* 48:417-428). In a representative embodiment, a gene therapy system comprising the subject CCR-gene in a retroviral vector is used to treat certain breast cancers. In preferred embodiments, expression of the subject CCR-protein- gene is controlled at least in part by a mammary-specific promoter, a number of which are available (for review, see Hennighausen (1990) *Protein Expression and Purification* 1:3-8; and Günzberg et al. (1992) *Biochem J* 283:625-632).

In similar fashion, gene therapy protocols involving delivery of the subject CCR-gene can be used in the treatment of malignant melanoma, which also serves as a model for progressive TGF-β resistance in transformation. In preferred embodiments, gene therapy protocols for treatment of melanomas include, in addition to the delivery of the CCR-gene construct by retroviral delivery, the delivery of a pharmaceutical preparation of the gene by direct injection. For instance, U.S. patent no. 5,318,514 describes an applicator for the electroporation of genes into epidermal cells and can be used in accordance with the present invention.

The subject CCR-gene constructs can be used in the treatment of hyperproliferative vascular disorders, e.g. smooth muscle hyperplasia (such as atherosclerosis) or restinosis, as well as other disorders characterized by fibrosis, e.g. rheumatoid arthritis, insulin dependent diabetes mellitus, glomerulonephritis, cirrhosis, and scleroderma, particularly proliferative disorders in which loss of TGF-β autocrine or paracrine signaling is implicated.

For example, restinosis continues to limit the efficacy of coronary angioplasty despite various mechanical and pharmaceutical interventions that have been employed. An important mechanism involved in normal control of intimal proliferation of smooth muscle cells appears to be the induction of autocrine and paracrine TGF-β inhibitory loops in the smooth muscle cells (Scott-Burden et al. (1994) *Tex Heart Inst J* 21:91-97; Graiger et al. (1993) *Cardiovasc Res* 27:2238-2247; and Grainger et al. (1993) *Biochem J* 294:109-112). Loss of sensitivity to TGF-β, or alternatively, the overriding of this inhibitory stimulus such as by PDGF autostimulation, can be a contributory factor to abnormal smooth muscle proliferation in restinosis. It may therefore be possible to treat or prevent restinosis by the use of gene therapy with CCR-gene constructs of the present invention. The CCR-gene construct can be delivered, for example, by percutaneous transluminal gene transfer (Mazur et al. (1994) *Tex Heart Inst J* 21:104-111) using viral or liposomal delivery compositions. An exemplary adenovirus-mediated gene transfer technique and compositions for treatment of cardiac or vascular smooth muscle is provided in PCT publication WO 94/11506.

Transforming growth factor-β is also understood to play a significant role in local glomerular and interstitial sites in human kidney development and disease. Consequently, the subject method provides a method of treating or inhibiting glomerulopathies and other

renal proliferative disorders comprising the *in vivo* delivery and recombinant expression of the subject CCR-protein in kidney tissue.

The subject method can also be used to treat retinoblastomas in which the retinoblastoma gene (RB) is not itself impaired, e.g. the effective impairment of the RB checkpoint is the result of a failure to control CDK4 phosphorylation of RB. Thus, an exogenous CCR-gene can be expressed in a retinoblastoma cell, thereby causing inhibition of CDK4 activation and down-regulating RB phosphorylation. To illustrate, a recombinant retrovirus can be constructed to facilitate expression of p16 or p15, and infectivity of retinoblastoma cells enhanced by derivatizing the *env* protein with antibodies specific for retinoblastoma cells, e.g. antibodies to retinal S-antigen (Doroso et al. (1985) *Invest Opthalmol Vis Sci* 26:560-572; see also Liao et al. (1981) *Eur J Immunol* 11:450-454; and U.S. Patent No. 4,444,744).

In yet another embodiment, the subject CCR-gene is delivered to a sarcoma, e.g. an osteosarcoma or Kaposi's sarcoma. In a representative embodiment, the gene is provided in a viral vector and delivered by way of a viral particle which has been derivatized with antibodies immunoselective for an osteosarcoma cell (see, for example, U.S. Patents 4,564,517 and 4,444,744; and Singh et al. (1976) *Cancer Res* 36:4130-4136).

In a still further embodiment, the subject CCR-gene is recombinantly expressed in tissue which is characterized by unwanted de-differentiation and which may also be undergoing unwanted apoptosis. For instance, many neurological disorders are associated with degeneration of discrete populations of neuronal elements. For example, Alzheimer's disease is associated with deficits in several neurotransmitter systems, both those that project to the neocortex and those that reside with the cortex. For instance, the nucleus basalis in patients with Alzheimer's disease were observed to have a profound (75%) loss of neurons compared to age-matched controls. Although Alzheimer's disease is by far the most common form of dementia, several other disorders can produce dementia. Many are age-related, occurring in far greater incidence in older people than in younger. Several of these are degenerative diseases characterized by the death of neurons in various parts of the central nervous system, especially the cerebral cortex. However, some forms of dementia are associated with degeneration of the thalmus or the white matter underlying the cerebral cortex. Here, the cognitive dysfunction results from the isolation of cortical areas by the degeneration of efferents and afferents. Huntington's disease involves the degeneration of intrastraital and cortical cholinergic neurons and GABAergic neurons. Pick's disease is a severe neuronal degeneration in the neocortex of the frontal and anterior temporal lobes, sometimes accompanied by death of neurons in the striatum. Accordingly, the subject CCR-gene can be delivered to the effected tissue by gene therapy techniques. It is noted that numerous advances have been made in the construction of expression vectors, cellular and viral transgene carriers, and the characterization of target cells for neuronal gene therapy, and

can be readily adapted for delivery of the subject CCR-genes (see, for example, Suhr et al. (1993) *Arch Neurol* 50:1252-1268; Jiao et al. (1993) *Nature* 362:450-453; Friedmann (1992) *Ann Med* 24:411-417; and Freese et al. (1991) *Nuc Acid Res* 19:7219-7223)

In addition to degenerative-induced dementias, the subject gene therapy systems can be applied opportunely in the treatment of neurodegenerative disorders which have manifestations of tremors and involuntary movements. Parkinson's disease, for example, primarily affects subcortical structures and is characterized by degeneration of the nigrostriatal pathway, raphe nuclei, locus cereleus, and the motor nucleus of vagus. Ballism is typically associated with damage to the subthalmic nucleus, often due to acute vascular accident. Also included are neurogenic and myopathic diseases which ultimately affect the somatic division of the peripheral nervous system and are manifest as neuromuscular disorders. Examples include chronic atrophies such as amyotrophic lateral sclerosis, Guillain-Barre syndrome and chronic peripheral neuropathy, as well as other diseases which can be manifest as progressive bulbar palsies or spinal muscular atrophies. Moreover, the use of CCR gene therapy constructs is amenable to the treatment of disorders of the cerebellum which result in hypotonia or ataxia, such as those lesions in the cerebellum which produce disorders in the limbs ipsilateral to the lesion. For instance, p16 or p15 gene constructs can used to treat a restricted form of cerebellar cortical degeneration involving the anterior lobes (vermis and leg areas) such as is common in alcoholic patients.

Furthermore, the subject CCR-genes can also be used in the treatment of autonomic disorders of the peripheral nervous system, which include disorders affecting the innervation of smooth muscle and endocrine tissue (such as glandular tissue). For instance, recombinant p16 or p15 expression by gene therapy can be used to treat tachycardia or atrial cardiac arrythmias which may arise from a degenerative condition of the nerves innervating the striated muscle of the heart.

Yet another aspect of the invention pertains to CDK4 and CDK6 mutants which are insensitive to the inhibitory control of one or more of the subject CCR proteins. With respect to cell cycle control, these mutants antagonize the action of CCR-proteins, and in preferred embodiments, are dominant negative mutants in that they remove from the control of mitosis certain cell-cycle checkpoints, such as checkpoints mediated by RB or RB-like proteins. Exemplary CDK4 and CDK6 mutants are represented in SEQ ID Nos. 9 (CDK4) and SEQ ID Nos. 10 (CDK6), wherein Xaa at position 24 and 31 respectively, which is an Arg in the wild-type enzyme, is mutated to render the enzyme insensitive to inhibition by CCR proteins. In preferred embodiments, Xaa is an amino acid other than arginine, though more preferably an amino acid which does not have a basic sidechain (e.g. not Arg, Lys or His). In a preferred embodiment, the Arginine is changed to a Cysteine.

Other CDK4 mutants which have lost the ability to bind p16 are described in Example 10 below. Accordingly, the Xaa at position 22 of CDK4 (residue 29 of CDK6) can

be an amino acid other than lysine, preferably an amino acid with a basic sidechain or a non-polar sidechain; the Xaa at position 95 of CDK4 (residue 100 of CDK6) can be an amino acid other than a histidine, and is preferably an amino acid with an acidic or non-polar or hydrophobic sidechain; and the Xaa position 97 (residue 102 of CDK6) can be an amino acid other than and aspartic acid, and is preferably an amino acid with a basic sidechain or a non-polar sidechain. It will be appreciated that such CDK4/CDK6 mutants can comprise one or more of any of the above mutations.

The mutant CDK4 and CDK6 proteins, genes encoding such proteins, and antibodies specific for such proteins, can be provided in a manner analogous to any of the embodiments described above for each of the CCR-proteins. For instance, nucleic acids encoding the mutant kinases can be provided in any of a number of forms, e.g. as recombinant expression vectors as well as probes. Both isolated and recombinant forms of the enzymes are specifically contemplated, as well as antibodies which specifically bind the mutant forms of each of the kinases. Moreover, it will be appreciated that, wherever the present application mentions the utility for CCR antagonists, the subject CDK mutants can be utilized.

Moreover, while the Arg→Cys mutant of CDK4 was first discovered in cells from a melanoma patient, and the subsequent CDK4/CDK6 mutants isolated from analysis of molecular models of CDK4 based on this mutation (see Example 10), mutagenic techniques such as those described above can be utilized to isolate yet other CDK4 and CDK6 mutants which are resistant to CCR protein inhibition. For instance, scanning mutagenesis or random mutagenesis can be used to generate a library of CDK mutants which can be screened for dominant negative mutations employing assays sensitive to CDK4 or CDK6 inhibition. For example, a library of CDK4 mutants can be transfected into cells which overexpress p16 and which are quiescent as a result. In the absence of mutations which uncouple the CDK4 mutant from p16/p15, the cell will remain quiescent. In contrast, mutations to the kinase which uncouple its activity from p16 will result in proliferation of cells expressing that mutant of CDK4. In a similar embodiment, the level of phosphorylated RB can be used to detect dominant negative mutants of CDK4, such as by immunodetection of phosphorylated RB, or the use of RB-sensitive reporter constructs the expression of which is dependent on the phosphorylation state of RB (e.g., see Park et al. (1994) *J Biol Chem* 269:6083-6088; Ouellette et al. (1992) *Oncogene* 7:1075-1081; and Slack et al. (1993) *Oncogene* 8:1585-1591). Mutants which uncouple p16/p15 inhibition of CDK4 will cause higher levels of phosphorylated RB protein to accumulate in the cell.

Another aspect of the invention relates to the discovery of the CCR-binding pocket of CDK4/CDK6. As described in Example 10 below, we have ascertained from point mutations and molecular modelling that a surface determined at least in part by residues K22, R24, H95 and D97 of CDK4 (K29, R31, H100 and D102 of CDK6) defines a binding pocket for CCR proteins such as p16. This binding pocket, which is within the small lobe of CDK4, is in

close proximity to the ATP binding site for the kinase, the binding to which by p16 likely blocks or distorts the ATP binding site of CDK4/CDK6. Accordingly, the present invention concerns three-dimensional molecular models of CDK proteins, and their use as templates for the design of agents able to inhibit at least one biological activity of that kinase. An integral step to an approach to designing CCR-like inhibitors of CDK4 or CDK6 involves construction of computer graphics models of the kinase which can be used to design pharmacophores by rational drug design based on mimicking the binding of CCR-proteins to the p16-binding site. For instance, for an inhibitor to interact optimally with the kinase, it will generally be desirable that it have a shape which is at least partly complimentary to that of the p16-binding site of the enzyme. Additionally, other factors, including electrostatic interactions, hydrogen bonding, hydrophobic interactions, desolvation effects, and cooperative motions of ligand and enzyme, all influence the binding effect and should be taken into account in attempts to design bioactive inhibitors.

A computer-generated molecular model of the subject enzyme can be created. In preferred embodiments, at least the Cα-carbon positions of CDK4 or CDK6 are mapped to a particular coordinate pattern, such as the coordinates of CDK2 (DeBondt et al. (1993) *Nature* 363:595-602; Endicott et al. (1994) *Prot Eng* 7:243-253; and Morgan et al. (1994) *Curr Opin Cell Biol* 6:239-246), by homology modeling, and the structure of the kinase and velocities of each atom are calculated at a simulation temperature (T_o) at which the docking simulation is to be determined. Typically, such a protocol involves primarily the prediction of side-chain conformations in the modeled kinase, while assuming a main-chain trace taken from a tertiary structure such as provided by the CDK2 structure. Computer programs for performing energy minimization routines are commonly used to generate molecular models. For example, both the CHARMM (Brooks et al. (1983) *J Comput Chem* 4:187-217) and AMBER (Weiner et al (1981) *J. Comput. Chem.* 106: 765) algorithms handle all of the molecular system setup, force field calculation, and analysis (see also, Eisenfield et al. (1991) *Am J Physiol* 261:C376-386; Lybrand (1991) *J Pharm Belg* 46:49-54; Froimowitz (1990) *Biotechniques* 8:640-644; Burbam et al. (1990) *Proteins* 7:99-111; Pedersen (1985) *Environ Health Perspect* 61:185-190; and Kini et al. (1991) *J Biomol Struct Dyn* 9:475-488). At the heart of these programs is a set of subroutines that, given the position of every atom in the model, calculate the total potential energy of the system and the force on each atom. These programs may utilize a starting set of atomic coordinates, such as the model coordinates for CDK2, the parameters for the various terms of the potential energy function, and a description of the molecular topology (the covalent structure). Common features of such molecular modeling methods include: provisions for handling hydrogen bonds and other constraint forces; the use of periodic boundary conditions; and provisions for occasionally adjusting positions, velocities, or other parameters in order to maintain or change temperature, pressure, volume, forces of constraint, or other externally controlled conditions.

Most conventional energy minimization methods use the input data described above and the fact that the potential energy function is an explicit, differentiable function of Cartesian coordinates, to calculate the potential energy and its gradient (which gives the force on each atom) for any set of atomic positions. This information can be used to generate a new set of coordinates in an effort to reduce the total potential energy and, by repeating this process over and over, to optimize the molecular structure under a given set of external conditions. These energy minimization methods are routinely applied to molecules similar to CDK4 and CDK6 as well as nucleic acids, polymers and zeolites.

In general, energy minimization methods can be carried out for a given temperature, T_i, which may be different than the docking simulation temperature, T_o. Upon energy minimization of the molecule at T_i, coordinates and velocities of all the atoms in the system are computed. Additionally, the normal modes of the system are calculated. It will be appreciated by those skilled in the art that each normal mode is a collective, periodic motion, with all parts of the system moving in phase with each other, and that the motion of the molecule is the superposition of all normal modes. For a given temperature, the mean square amplitude of motion in a particular mode is inversely proportional to the effective force constant for that mode, so that the motion of the molecule will often be dominated by the low frequency vibrations.

After the molecular model has been energy minimized at T_i, the system is "heated" or "cooled" to the simulation temperature, T_o, by carrying out an equilibration run where the velocities of the atoms are scaled in a step-wise manner until the desired temperature, T_O, is reached. The system is further equilibrated for a specified period of time until certain properties of the system, such as average kinetic energy, remain constant. The coordinates and velocities of each atom are then obtained from the equilibrated system.

Further energy minimization routines can also be carried out. For example, a second class of methods involves calculating approximate solutions to the constrained EOM for the protein. These methods use an iterative approach to solve for the Lagrange multipliers and, typically, only need a few iterations if the corrections required are small. The most popular method of this type, SHAKE (Ryckaert et al. (1977) *J Comput Phys* 23:327; and Van Gunsteren et al. (1977) *Mol Phys* 34:1311) is easy to implement and scales as O(N) as the number of constraints increases. Therefore, the method is applicable to macromolecules such as the CDK proteins of interest. An alternative method, RATTLE (Anderson (1983) *J Comput Phys* 52:24) is based on the velocity version of the Verlet algorithm. Like SHAKE, RATTLE is an iterative algorithm and can be used to energy minimize the model of the subject protein.

The increasing availability of biomacromolecule structures of potential pharmacophoric molecules that have been solved crystallographically has prompted the development of a variety of direct computational methods for molecular design, in which the

steric and electronic properties of substrate binding sites are use to guide the design of potential inhibitors of CDK4 and/or CDK6 (Cohen et al. (1990) *J. Med. Cam.* 33: 883-894; Kuntz et al. (1982) *J. Mol. Biol* 161: 269-288; DesJarlais (1988) *J. Med. Cam.* 31: 722-729; Bartlett et al. (1989) (*Spec. Publ., Roy. Soc. Chem.*) 78: 182-196; Goodford et al. (1985) *J. Med. Cam.* 28: 849-857; DesJarlais et al. *J. Med. Cam.* 29: 2149-2153). Directed methods generally fall into two categories: (1) design by analogy in which 3-D structures of known molecules (such as from a crystallographic database) are docked to the enzyme structure and scored for goodness-of-fit; and (2) *de novo* design, in which the ligand model is constructed piece-wise in the enzyme. The latter approach, in particular, can facilitate the development of novel molecules, uniquely designed to bind to the kinase CCR-binding site.

In an illustrative embodiment, the design of potential CDK4/CDK6 inhibitors begins from the general perspective of shape complimentary for the p16-binding site, and a search algorithm is employed which is capable of scanning a database of small molecules of known three-dimensional structure for candidates which fit geometrically into the target binding site. It is not expected that the molecules found in the shape search will necessarily be leads themselves, since no evaluation of chemical interaction necessarily be made during the initial search. Rather, it is anticipated that such candidates might act as the framework for further design, providing molecular skeletons to which appropriate atomic replacements can be made. Of course, the chemical complimentary of these molecules can be evaluated, but it is expected that atom types will be changed to maximize the electrostatic, hydrogen bonding, and hydrophobic interactions with the enzyme. Most algorithms of this type provide a method for finding a wide assortment of chemical structures that are complementary to the shape of the p16-binding site of the subject kinases. Each of a set of small molecules from a particular data-base, such as the Cambridge Crystallographic Data Bank (CCDB) (Allen et al. (1973) *J. Chem. Doc.* 13: 119), is individually docked to the p16-binding site of the kinase in a number of geometrically permissible orientations with use of a docking algorithm. In a preferred embodiment, a set of computer algorithms called DOCK, can be used to characterize the shape of invaginations and grooves that form the p16 recognition surfaces of the CDK protein (Kuntz et al. (1982) *J. Mol. Biol* 161: 269-288). The program can also search a database of small molecules for templates whose shapes are complementary to the p16-binding sites of the enzyme (DesJarlais et al. (1988) *J Med Chem* 31: 722-729). These templates normally require modification to achieve good chemical and electrostatic interactions (DesJarlais et al. (1989) *ACS Symp Ser* 413: 60-69). However, the program has been shown to position accurately known cofactors for inhibitors based on shape constraints alone.

The orientations are evaluated for goodness-of-fit and the best are kept for further examination using molecular mechanics programs, such as AMBER or CHARMM. Such algorithms have previously proven successful in finding a variety of molecules that are

complementary in shape to a given binding site of a receptor-enzyme, and have been shown to have several attractive features. First, such algorithms can retrieve a remarkable diversity of molecular architectures. Second, the best structures have, in previous applications to other proteins, demonstrated impressive shape complementarity over an extended surface area. Third, the overall approach appears to be quite robust with respect to small uncertainties in positioning of the candidate atoms.

Goodford (1985, *J Med Chem* 28:849-857) and Boobbyer et al. (1989, *J Med Chem* 32:1083-1094) have produced a computer program (GRID) which seeks to determine regions of high affinity for different chemical groups (termed probes) on the molecular surface of the binding site. GRID hence provides a tool for suggesting modifications to known ligands, such as p16 and the like, that might enhance binding. It may be anticipated that some of the sites discerned by GRID as regions of high affinity correspond to "pharmacophoric patterns" determined inferentially from a series of known ligands. As used herein, a pharmacophoric pattern is a geometric arrangement of features of the anticipated ligand that is believed to be important for binding. Accordingly, short overlapping peptidyl portions of the subject CCR-proteins can be scanned for ability to bind the p16-binding site defined for CDK4 or CDK6.

Yet a further embodiment of the present invention utilizes a computer algorithm such as CLIX which searches such databases as CCDB for small molecules which can be oriented in the receptor binding site in a way that is both sterically acceptable and has a high likelihood of achieving favorable chemical interactions between the candidate molecule and the surrounding amino acid residues. The method is based on characterizing the p16-binding site of the kinase in terms of an ensemble of favorable binding positions for different chemical groups and then searching for orientations of the candidate molecules that cause maximum spatial coincidence of individual candidate chemical groups with members of the ensemble. The current availability of computer power dictates that a computer-based search for novel ligands follows a breadth-first strategy. A breadth-first strategy aims to reduce progressively the size of the potential candidate search space by the application of increasingly stringent criteria, as opposed to a depth-first strategy wherein a maximally detailed analysis of one candidate is performed before proceeding to the next. CLIX conforms to this strategy in that its analysis of binding is rudimentary -it seeks to satisfy the necessary conditions of steric fit and of having individual groups in "correct" places for bonding, without imposing the sufficient condition that favorable bonding interactions actually occur. A ranked "shortlist" of molecules, in their favored orientations, is produced which can then be examined on a molecule-by-molecule basis, using computer graphics and more sophisticated molecular modeling techniques. CLIX is also capable of suggesting changes to the substituent chemical groups of the candidate molecules that might enhance binding to the p16-binding pocket of the kinase.

The algorithmic details of CLIX is described in Lawerence et al. (1992) *Proteins* 12:31-41, and the CLIX algorithm can be summarized as follows. The GRID program is used to determine discrete favorable interaction positions (termed target sites) in the p16-binding site of the kinase for a wide variety of representative chemical groups. For each candidate ligand in the CCDB an exhaustive attempt is made to make coincident, in a spatial sense in the binding site of the protein, a pair of the candidate's substituent chemical groups with a pair of corresponding favorable interaction sites proposed by GRID. All possible combinations of pairs of ligand groups with pairs of GRID sites are considered during this procedure. Upon locating such coincidence, the program rotates the candidate ligand about the two pairs of groups and checks for steric hindrance and coincidence of other candidate atomic groups with appropriate target sites. Particular candidate/orientation combinations that are good geometric fits in the binding site and show sufficient coincidence of atomic groups with GRID sites are retained.

Consistent with the breadth-first strategy, this approach involves simplifying assumptions. Rigid protein and small molecule geometry is maintained throughout. As a first approximation, rigid geometry is acceptable as the energy minimized coordinates of the CDK4 or CDK6 deduced structure describe an energy minimum for the molecule, albeit a local one. A further assumption implicit in CLIX is that the potential ligand, when introduced into the p16-binding site of the kinase, does not induce change in the protein's stereochemistry or partial charge distribution and so alter the basis on which the GRID interaction energy maps were computed. It must also be stressed that the interaction sites predicted by GRID are used in a positional and type sense only, i.e., when a candidate atomic group is placed at a site predicted as favorable by GRID, no check is made to ensure that the bond geometry, the state of protonation, or the partial charge distribution favors a strong interaction between the protein and that group. Such detailed analysis should form part of more advanced modeling of candidates identified in the CLIX shortlist.

Yet another embodiment of a computer-assisted molecular design method for identifying specific inhibitors of CDK4 or CDK6 comprises the *de novo* synthesis of potential inhibitors by algorithmic connection of small molecular fragments that will exhibit the desired structural and electrostatic complementarity with the p16-binding site of the enzyme. The methodology employs a large template set of small molecules with are iteratively pieced together in a model of the CDK4/CDK6 p16-binding site. Each stage of ligand growth is evaluated according to a molecular mechanics-based energy function, which considers van der Waals and coulombic interactions, internal strain energy of the lengthening ligand, and desolvation of both ligand and enzyme. The search space can be managed by use of a data tree which is kept under control by pruning according to the binding criteria.

In an illustrative embodiment, the search space is limited to consider only amino acids and amino acid analogs as the molecular building blocks. Such a methodology generally

employs a large template set of amino acid conformations, though need not be restricted to just the 20 natural amino acids, as it can easily be extended to include other related fragments of interest to the medicinal chemist, e.g. amino acid analogs. The putative ligands that result from this construction method are peptides and peptide-like compounds rather than the small organic molecules that are typically the goal of drug design research. The appeal of the peptide building approach is not that peptides are preferable to organics as potential pharmaceutical agents, but rather that: (1) they can be generated relatively rapidly *de novo*; (2) their energetics can be studied by well-parameterized force field methods; (3) they are much easier to synthesize than are most organics; and (4) they can be used in a variety of ways, for peptidomimetic inhibitor design, protein-protein binding studies, and even as shape templates in the more commonly used 3D organic database search approach described above. Moreover, peptidyl fragments such as may be virtually derived from the CCR-protein sequences described herein, can be employed.

Such a *de novo* peptide design method has been incorporated in a software package called GROW (Moon et al. (1991) *Proteins* 11:314-328). In a typical design session, standard interactive graphical modeling methods are employed to define the structural environment in which GROW is to operate. For instance, the environment could be the p16-binding site of CDK4, or it could be a set of features on the protein's surface to which the user wishes to bind a peptide-like molecule, e.g. residues K22, R24, H95 and D97 of CDK4. The GROW program then operates to generate a set of potential ligand molecules. Interactive modeling methods then come into play again, for examination of the resulting molecules, and for selection of one or more of them for further refinement.

To illustrate, GROW operates on an atomic coordinate file generated by the user in the interactive modeling session, such as the coordinates provided by homology modelling of CDK4, or the simply the coordinates of the p16-binding site, plus a small fragment (e.g., an acetyl group) positioned in the p16-binding site to provide a starting point for peptide growth. These are referred to as "site" atoms and "seed" atoms, respectively. A second file provided by the user contains a number of control parameters to guide the peptide growth (Moon et al. (1991) *Proteins* 11:314-328).

The operation of the GROW algorithm is conceptually fairly simple. GROW proceeds in an iterative fashion, to systematically attach to the seed fragment each amino acid template in a large preconstructed library of amino acid conformations. When a template has been attached, it is scored for goodness-of-fit to the receptor site, and then the next template in the library is attached to the seed. After all the templates have been tested, only the highest scoring ones are retained for the next level of growth. This procedure is repeated for the second growth level; each library template is attached in turn to each of the bonded seed/amino acid molecules that were retained from the first step, and is then scored. Again, only the best of the bonded seed/dipeptide molecules that result are retained for the third level

of growth. The growth of peptides can proceed in the N-to-C direction only, the reverse direction only, or in alternating directions, depending on the initial control specifications supplied by the user. Successive growth levels therefore generate peptides that are lengthened by one residue. The procedure terminates when the user-defined peptide length has been reached, at which point the user can select from the constructed peptides those to be studied further. The resulting data provided by the GROW procedure include not only residue sequences and scores, but also atomic coordinates of the peptides, related directly to the coordinate system of the receptor site atoms.

In yet another embodiment, potential pharmacophoric compounds can be determined using a method based on an energy minimization-quenched molecular dynamics algorithm for determining energetically favorable positions of functional groups in the p16-binding sie of a CDK. The method can aid in the design of molecules that incorporate such functional groups by modification of known ligands or *de novo* construction.

For example, the multiple copy simultaneous search method (MCSS) described by Miranker et al. (1991) *Proteins* 11: 29-34 can be employed. To determine and characterize a local minima of a functional group in the forcefield of the protein, multiple copies of selected functional groups are first distributed in the p16-binding site of the CDK protein of interest. Energy minimization of these copies by molecular mechanics or quenched dynamics yields the distinct local minima. The neighborhood of these minima can then be explored by a grid search or by constrained minimization. In one embodiment, the MCSS method uses the classical time dependent Hartee (TDH) approximation to simultaneously minimize or quench many identical groups in the forcefield of the protein.

Implementation of the MCSS algorithm requires a choice of functional groups and a molecular mechanics model for each of them. Groups must be simple enough to be easily characterized and manipulated (3-6 atoms, few or no dihedral degrees of freedom), yet complex enough to approximate the steric and electrostatic interactions that the functional group would have in binding to the p16-binding site of the kinase. A preferred set is, for example, one in which most organic molecules can be described as a collection of such groups (*Patai's Guide to the Chemistry of Functional Groups*, ed. S. Patai (New York: John Wiley, and Sons, (1989)). This includes fragments such as acetonitrile, methanol, acetate, methyl ammonium, dimethyl ether, methane, and acetaldehyde.

Determination of the local energy minima in the binding site requires that many starting positions be sampled. This can be achieved by distributing, for example, 1,000-5,000 groups at random inside a sphere centered on the binding site; only the space not occupied by the protein needs to be considered. If the interaction energy of a particular group at a certain location with the protein is more positive than a given cut-off (e.g. 5.0 kcal/mole) the group is discarded from that site. Given the set of starting positions, all the fragments are minimized simultaneously by use of the TDH approximation (Elber et al. (1990) *J Am Chem Soc* 112:

9161-9175). In this method, the forces on each fragment consist of its internal forces and those due to the protein. The essential element of this method is that the interactions between the fragments are omitted and the forces on the protein are normalized to those due to a single fragment. In this way simultaneous minimization or dynamics of any number of functional groups in the field of a single protein can be performed.

Minimization is performed successively on subsets of, e.g. 100, of the randomly placed groups. After a certain number of step intervals, such as 1,000 intervals, the results can be examined to eliminate groups converging to the same minimum. This process is repeated until minimization is complete (e.g. RMS gradient of 0.01 kcal/mole/Å). Thus the resulting energy minimized set of molecules comprises what amounts to a set of disconnected fragments in three dimensions representing potential pharmacophores.

The next step then is to connect the pharmacophoric pieces with spacers assembled from small chemical entities (atoms, chains, or ring moieties). In a preferred embodiment, each of the disconnected can be linked in space to generate a single molecule using such computer programs as, for example, NEWLEAD (Tschinke et al. (1993) *J Med Chem* 36: 3863,3870). The procedure adopted by NEWLEAD executes the following sequence of commands (1) connect two isolated moieties, (2) retain the intermediate solutions for further processing, (3) repeat the above steps for each of the intermediate solutions until no disconnected units are found, and (4) output the final solutions, each of which is single molecule. Such a program can use for example, three types of spacers: library spacers, single-atom spacers, and fuse-ring spacers. The library spacers are optimized structures of small molecules such as ethylene, benzene and methylamide. The output produced by programs such as NEWLEAD consist of a set of molecules containing the original fragments now connected by spacers. The atoms belonging to the input fragments maintain their original orientations in space. The molecules are chemically plausible because of the simple makeup of the spacers and functional groups, and energetically acceptable because of the rejection of solutions with van-der Waals radii violations.

Still another aspect of the invention relates to CDK mutants which have been rendered sensitive to CCR proteins such as p16 or p15. For instance, CDK2 or CDC2, which do not normally bind p16, and be rendered sensitive to p16 inhibition by replacing certain residues of the kinase with the corresponding residues from CDK4 which are involved in p16 binding. For example, as Figure 7 illustrates, portions of CDK4 corresponding to β1, β2, β3, β4, β5, or L7 can be interchanged into CDK2 or CDC2 in order to yield a p16-sensitive CDK. Such mutants can be used, for example, to generate recombinant cells either *in vitro* or *in vivo* which can be inhibited at other points of the cell-cycle by the expression of a CCR-protein.

Exemplification

The invention now being generally described, it will be more readily understood by reference to the following examples which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

As described previously (see U.S. Patent Applications Serial Nos. 08/154,915, 07/991,997 and 07/963,308, as well as Xiong et al. (1993) *Nature* 366:701; Xiong et al. (1993) *Genes Dev* 7:1572; Xiong et al. (1992) *Cell* 71:505; and Zhang et al. (1993) *Mol Cell Biol* 4:897), immunological procedures have been used to establish that cyclins associate, in eukaryotic cells, with a variety of potential catalytic subunits (e.g., CDKs, such as CDK2, CDK4 and CDK5). To illustrate, human cyclin D1 has been associated with a wide variety of proliferative diseases. In human diploid cells, specifically human diploid fibroblasts, cyclin D1 is complexed with a number of other cellular proteins. Among them are the catalytic subunits CDK2, CDK4 (previously called PSK-J3), CDK5 (also called PSSALRE), and CDK6 (PLSTIRE). In addition, polypeptides of 21 kDa and 36 kDa were identified in association with cyclin D1. It was shown that the 36 kDa protein is the Proliferating Cell Nuclear Antigen, PCNA. PCNA has been described as an essential accessory factor to the delta polymerase, which is required for leading-strand DNA replication and DNA repair. Cyclin D3 was also found to associate with multiple protein kinases, p21 and PCNA. It was therefore proposed that there exists a quaternary complex of D type cyclins, CDK, PCNA and p21, and that many combinatorial variations (cyclin D1, D3 with CDK2, 4, 5 and 6) may assemble *in vivo*.

The importance of the quaternary complex is emphasized by the discovery that cellular transformation by DNA tumor viruses is associated with selective subunit rearrangement of the cyclin D complexes, as well as other cell-cycle complexes, including cyclin A, CDC2, CDK2, CDK4 and CDK5 complexes. In particular, introduction of SV40 DNA tumor virus or its oncogenic gene product large T antigen into normal human diploid fibroblasts (HDF) causes disruption of the association between cyclin D and PCNA, CDKs (such as CDK2, CDK4, CDK5 and CDK6) and p21. For example, after dissociation from cyclin D and p21, CDK4 kinase becomes associated with a 16 kDa polypeptide (p16). Similarly, SV40 transformation causes a decrease of association of p21 with cyclin A in HDF; and adenovirus-(293 cell line) or human papilloma virus- (HeLa cell line) transformed cells, p21 is completely disassociated from cyclin A. A 19 kDa peptide, p19, then appears in a complex with cyclin A.

Thus, in many transformed cells, cyclin and CDK's associate in binary complexes which form the core of the cell-cycle regulatory machinery. In normal cells, a major fraction of the cyclin kinases acquire two additional subunits (p21 and PCNA) and thereby form quaternary complexes. Reconstitution of quaternary complexes in insect cells revealed that

p21 is a universal inhibitor of cyclin kinases. As such, p21 inhibits cell-cycle progression and cell proliferation upon overexpression in mammalian cells. Taken in conjunction with the previously demonstrated absence of p21 protein in the cell-cycle kinase complexes of cells with deficient p53, these results suggest that p21 is a transcriptional target of the tumor suppressor protein, p53. One function of p53 is to act in a cellular signaling pathway which causes cell-cycle arrest following DNA damage (see for example, Kastan et al. *Cell* 71:587-5971993). It has therefore been suggested that p21 forms a critical link between p53 and the cell-cycle control machinery.

Cyclin D/CDK4 kinase differs from the others in its inability to utilize histone H1 as a substrate. To date, the only substrates known for cyclin D/CDK4 kinases are the members of the RB family of "pocket" proteins (Matsushime et al., *Cell* 71:323-334 (1992)). Therefore, the effect of p21 was tested on the ability of cyclin D/CDK4 to phosphorylated RB. Insect cell lysates containing cyclin D or CDK4 alone showed little activity toward GST-RB. However, cyclin D/CDK4 binary complexes catalyzed substantial RB phosphorylation. Addition of increasing amounts of p21 resulted in the accumulation of cyclin D/CDK4/p21 ternary complexes with a corresponding inhibition of RB phosphorylation. Inclusion of PCNA was essentially without effect. However, cells lacking functional p53 may nevertheless retain a functional RB checkpoint which undergoes differential phosphorylation despite lack of endogenous p21.

The two-hybrid screening system (Fields et al. *Nature* 340:254 (1989)) was utilized to search for proteins that could interact with human CDK4, and more specifically, to isolate a cDNA encoding p16. Two-hybrid screening relies on reconstituting a functional GAL4 activator from two separated fusion proteins: the GAL4 DNA-binding domain fused to CDK4, GAL4db-CDK4; and the GAL4 activation domain fused to the proteins encoded by HeLa cDNAs, GAL4ad-cDNA. YPB2 was used as the recipient yeast as it is a strain that contains two chromosomal genes under the control of two different GAL4-dependent promoters: HIS3 and LacZ. YPB2 was transformed with a mixture of two plasmids encoding, respectively, the GAL4db-CDK4 and the GAL4ad-cDNA fusions; several clones were obtained that grew in the absence of histidine and that turned blue in the presence of β-gal. From DNA sequencing data it was determined that each of the positive clones derived from the same gene, although one group represented mRNAs with a shorter 3' end. The sequence of these cDNAs contained, in-phase with the GAL4ad, an open reading frame encoding a protein of 148 amino acids with a predicted molecular weight of 15,845 daltons (see SEQ ID Nos. 1 and 2). The sequence of p16 was compared by standard methods with those present in the currently available data banks and no significant homologies were found.

To test if p16 would specifically bind CDK4, YPB2 were cotransformed with the GAL4ad-p16 fusion as well as with several target GAL4db fusion constructs containing, respectively, cdc2, CDK2, CDK4, CDK5, PCNA and Snfl (a fission yeast kinase).

Transformed cells were plated with and without histidine. Only the GAL4db-CDK4 fusion interacted with GAL4ad-p16 to an extent which allowed growth in the absence of histidine, indicating that this pair of fusion proteins specifically reconstituted a functional GAL4 activator able to enhance the expression of the HIS3 gene. The same result was obtained when the ability to transactivate the expression of the β-galactosidase gene was assayed.

The specificity of this interaction was further demonstrated in a cell-free system, by mixing *in vitro* translated (^{35}S)-labeled CDKs with a purified bacterially-produced fusion protein consisting of glutathione-S transferase (GST) linked to p16 (17). The GST-p16 fusion was recovered by binding to glutathione-sepharose beads and the association of each CDK was analyzed by gel electrophoresis. Consistent with the previous observations, GST-p16 bound much more efficiently to CDK4 than to cdc2, CDK2 or CDK5.

Since the predicted molecular weight of p16 is close to 16 Kd, the identity of p16 as the CDK4-associated p16 protein found in transformed cell lines (see above) was determined. Two *in vitro* translation products of 15 KD and 17 KD were obtained from the p16 cDNA. These products, as well as the CDK4-associated p16 protein from HeLa cells were treated with N-chlorosuccinimide. The partial NCS-proteolytic pattern of the 17 KD cDNA-derived product was very similar to the pattern obtained with the CDK4-associated p16 protein from HeLa cells, strongly suggesting that the p16 cDNA actually corresponds to p16. Partial digestion with V8 protease of the 17 KD cDNA-derived product and p16 also yielded similar patterns. It is interesting to note that the p16 protein overexpressed in insect cells has an electrophoretic mobility of 15 KD, and its NCS proteolytic map is identical to that obtained with the 15 KD cDNA derived product. This suggests that the actual p16 found in human cells and the 17 KD *in vitro* translation product correspond to posttranslationally modified proteins. The fact that the p16 protein overexpressed in insect cells interacts with CDK4 suggests that this modification is not essential for the interaction (see below).

The identity between p16 and the CDK4-associated protein p16 was further confirmed using antibodies raised against the purified GST-p16 fusion protein. Several human cell lines were used for this experiment: a normal cell line WI38, derived from normal lung fibroblasts; the VA13 cell line derived from WI38 by transformation with the SV40 T-antigen; and HeLa cells. As set out above, anti-CDK4 immunoprecipitates of WI38 revealed the association of CDK4 with cyclin D1, PCNA, p21 and p16. In contrast, in VA13 and HeLa cells CDK4 is only associated with p16. Anti-p16 immunoprecipitates contained a protein with an apparent molecular weight of 16 KD which was readily detectable in the two transformed cell lines, VA13 and HeLa but to a lesser extent in the normal cell line WI38. This protein not only had the same electrophoretic mobility as the p16 protein coimmunoprecipitated with anti-CDK4 serum, but also had an identical NCS partial proteolytic pattern. In addition to p16 a protein of 33Kd was observed in anti-p16 coimmunoprecipitates that was shown to be identical to CDK4 by V8-proteolytic mapping.

Northern analysis of the transcripts present in WI38 and VA13 cells indicated that the p16 mRNA was around many times less abundant in WI38 cells compared to VA13 cells. This difference approximately corresponded to the observed difference in the amount of p16 protein between the two cell lines, suggesting the possibility that p16 expression might be regulated at a transcriptional or post-transcriptional level. Indeed, in three non-virally transformed cell lines the expression of p16 could not be detected even after overexposure of the gel.

To study the biochemical consequences of the interaction of p16 with CDK4, active CDK4-cyclin D complexes have been reconstituted *in vitro* by standard protocols (Kato et al. *Genes Der* 7:331 (1993); and Ewen et al. *Cell* 73:487 (1993)). The three relevant components, CDK4, p16 and cyclin D1, were expressed in baculovirus-infected insect cells. Extracts were prepared from metabolically (^{35}S)-labeled insect cells that separately overexpressed p16, CDK4 or cyclin D1, as well as from cells overexpressing both CDK4 and cyclin D1. In response to increasing amounts of p16, corresponding decreases in the ability of CDK4 to phosphorylate RB was observed. This inhibition correlated with the association between p16 and CDK4 as detected by immunoprecipitation. No inhibition was observed when CDK2-cyclin D2 complexes were used in a similar assay. To confirm that the inhibition of CDK4 was due to p16, a His-tagged p16 fusion protein (His-p16) was created to have an amino terminal extension of 20 amino acids containing a tract of 6 histidine residues. This fusion protein was overexpressed in baculovirus-infected insect cells, and was purified by virtue of the high-affinity association of the histidine tract to nickel-agarose beads. The His-p16 protein preparation was shown to be >90% pure, and inhibited the activity of the CDK4-cyclin D1 complex under conditions similar to those used for inhibition by the whole lysates.

The role of the retinoblastoma gene product (RB), appears to be as a cell-cycle checkpoint which appears to at least act be sequestering transcription factors responsible for the proteins of phase. In many carcinomas, p53 function is lost by mutation or deletion. RB, on the other hand, is not apparently altered as often. However, because p16 down-regulates the CDK4/cyclin D complex, which acts to phosphorylate RB, it is proposed herein that p16 loss in certain carcinomas can alleviate the effects of the RB checkpoint and, in some manner of speaking, represent a checkpoint deficiency analogous to p53 loss. The loss of p16 would result in more effective phosphorylation of RB and hence would remove the RB-mediated inhibition of the cell-cycle. Consistent with this notion, it is described below that in a variety of human tumor cells, such as cells which over-express a D-type cyclin, e.g. cyclin D1 or D2, the p16 gene is lost from the cell, e.g. homozygously deleted.

Moreover, as described in the examples below, the p16 gene was found to map to the human region 9p21-22, a known melanoma locus (Walker et al. (1994) *Oncogene* 9:819; Coleman et al. (1994) *Cancer Res* 54:344; Cheng et al. (1993) *Cancer Res* 53:4761; and

Cannon-Albright et al. (1992) *Science* 258:1148). The chromosomal mapping was further confirmed by analysis of somatic cell hybrids through PCR amplification (using primers ex1A and ex13 of Figure 2A). Somatic hybrids containing human chromosome 9 resulted in positive PCR products being amplified.

Utilizing primers generated from the cDNA sequence of human p16 (SEQ ID No. 1) which are shown in Figure 1, the genomic p16 gene was partially sequenced to determine intron/exon boundaries. The approximate sequences of the nucleic acid flanking Exon 1 and Exon 2 (see Figure 1) are shown in Figures 2A and 2B and 3A-D respectively.

Genomic DNA was isolated from a variety of human tumor lines (Sambrook et al. *Molecular Cloning: A Laboratory Manual*, CSHL Press, Cold Spring Harbor, NY (1989)) and was probed by PCR reactions for the presence or absence of p16 sequences. In particular, primers ex1A and ex13 (Figure 2A) were used to score for exon 1 of p16, and primers ex14 and ex15 were likewise used to detect exon 2 of p16. As shown in Figure 4, the p16 gene is disrupted in several tumor cell lines, confirming that p16 is indeed likely to be critical in cell transformation in certain cancerous cells. Moreover, probing of these cell lines with full length p16 cDNA (SEQ ID No. 1) demonstrated that in at least 3 of those cells apparently missing a portion of the p16 gene, the entire gene was in fact absent.

Based on immunoprecipitation experiments with anti-p16 antibodies, as well as oligonucleotide hybridization assays, it became apparent that the p16 protein represented by SEQ ID No. 2 is merely one member of a larger family of related cell-cycle regulatory proteins. For instance, even under high stringency conditions, Southern hybridization experiments of mRNA from different tissue types has indicated that approximately 4 closely related transcripts are produced. These p16 homologs, members of the CCR-protein family, may have arisen by gene duplication (e.g. each CCR-protein arises from a distinct gene) or from alternate exon splicing at the mRNA level, or a combination thereof.

Utilizing a probe consisting of the coding region of the human p16 gene, we have screened a mouse embryonal stem cell library and have isolated a genomic clone containing the coding region for a mouse homolog of the human p16 gene described above. This clone was isolated under low to moderate stringency conditions (1 x SSC at 50°C). This DNA (14kB) has been cloned in two independent pieces and the restriction map for nine restriction endonucleases has been performed (see Figure 5). The mouse CCR-gene has been completely sequenced and the coding region is apparently made up of only two exons that have been located in the restriction map by Southern hybridization. The apparent molecular weight of the mouse CCR-protein is 13.5kDa, and the nucleic acid and amino acid sequence of the mouse CCR-protein, termed herein "p13.5", is given in SEQ ID No. 5 and 6 respectively.

Moreover, utilizing degenerate probes based on the most highly conserved sequences between the human p16 clone and mouse p13.5 clone (e.g. between residues Met-52 through Gly-135 of human p16 and Met-1 through Gly-83 of mouse p13.5), we have isolated a

number of human and mouse p16 homologs, such as the mouse p15 homolog (SEQ ID Nos. 7 and 8).

In addition, it has been noted that TGF-β treatment causes accumulation of RB in the under-phosphorylated state and expression of RB-inactivating viral oncoproteins prevents TGF-β induced cell cycle arrest (Laiho et. al. (1990) *Cell* 62:175-185; and Pietenpol et al. (1990) *Cell* 61:777-785). While prior publications have suggested that TGF-β treatment results in down-regulation of CDK4 expression (Ewen et al. (1993) *Cell* 74:1009-1020), the data suggested to us that TGF-β might function through suppression of RB phosphorylation and pointed to the possibility that one result of TGF-β treatment might be inhibition of cyclin dependent kinases.

Accordingly, to investigate the mechanism by which TGF-β inhibits cell proliferation, we examined anti-CDK immunoprecipitates from human keratinocytes which had been arrested in G1 by exposure to TGF-β. Notably, immunoprecipitates of two G1-specific cyclin kinases, CDK4 and CDK6, contained several low molecular weight, associated proteins. These included p16 and two additional proteins of approx. 15 and 15.5 kDa. These proteins were not recovered in parallel CDC2 or CDK2 immunoprecipitates but were recovered in anti-p16 immunoprecipitates, suggesting that p15, p15.5 and p16 might be related. This was confirmed by western blotting of CDK4 and CDK6 immunoprecipitates which demonstrated that p15 and p15.5 were weakly cross-reactive with the p16 antiserum.

To isolate clones encoding putative p16 relatives, we constructed a cDNA library from TGF-β arrested HaCaT cells (Boukamp et al. (1988) *J Cell Biol* 106:761-771) and probed this library at low-stringency with the p16 coding sequence (SEQ ID No. 1). One clone obtained in this screen encoded a 137 amino acid protein (predicted M.W. 14.7 kDa.) with homology to p16. Based upon this homology and upon biochemical properties described below, we have named this protein p15. The first 50 amino acids of p15 and p16 share approx. 44% identity. This is followed by an 81 amino acid region of approx. 97% identity (see Figure 6) after which p15 and p16 sequences diverge. The sequence of p15 can be divided into four ankyrin repeats suggesting that this structural motif is conserved in the CCR-protein family. *In vitro* translation of the p15 cDNA produced a protein which precisely comigrated with the p15 band present in CDK4, CDK6 and p16 immunoprecipitates from TGF-β arrested HaCaT cells. Identity of these proteins was confirmed by protease and chemical cleavage mapping.

To investigate the functional similarity of p15 and p16, we expressed p15 as a fusion protein in bacteria and tested its ability to bind and inhibit cyclin dependent kinases. p15 specifically bound CDK4 and CDK6 but did not appreciably bind CDC2, CDK2 or CDK5. To assess the consequences of binding, p15 was added to active cyclin/CDK complexes expressed in insect cells. p15 specifically inhibited the cyclin D/CDK4 and cyclin D/CDK6 enzymes but had no effect on CDK2/cyclin A kinase. Thus p15 is a functional member of the

CCR-protein family. Moreover, FISH mapping of the p15 gene demonstrated that this gene lies adjacent to the p16 gene at 9p21.

While we first noted p15 in immunoprecipitates from HaCaT cells which had been arrested in G1 by serum starvation and re-stimulation in the presence of TGF-β, by comparison, we found that asynchronous, rapidly proliferating HaCaT cells contained considerably lower levels of p15 in CDK4 and CDK6 immunoprecipitates. To separate effects of TGF-β treatment from effects of Gl arrest, asynchronous cultures were treated with TGF-β for various periods, after which patterns of CDK4 and CDK6 associated proteins were examined. In as little as four hours following TGF-β addition, p15 levels rose in CDK4 and CDK6 immunoprecipitates, reaching peak levels after 6-8 hours. In contrast, CDK-associated p16 levels were unaffected by TGG-β. Northern blotting of RNA from cultures treated in parallel revealed that increased CDK4-associated p15 reflected increased abundance of p15 mRNA. In 2 hours following TGF-β treatment, p15 mRNA began to rise and reached a peak induction of approx. 30-fold after 6-8 hours. In contrast, p16 mRNA levels did not vary.

Two other mechanisms for TGF-β mediated cell cycle arrest have been previously proposed. In Mv1Lu cells, TGF-β treatment suppressed CDK4 synthesis. This was deemed causal since cells could be rendered resistant to TGF-β by constitutive overexpression of CDK4. In HaCaT cells, TGF-β treatment had no effect on CDK4 protein or mRNA levels. Based upon the properties of p15, we would predict that CDK4 overexpression could also render HaCaT cells TGF-β resistant by titrating the p15 CDK4/CDK6 inhibitor. p27, a CDK inhibitor which was purified from TGF-β arrested cells, has also been proposed as a link between TGF-β and cell cycle control. However, in HaCaT cells, TGF-β treatment had no effect on p27 mRNA levels. Thus any contribution that p27 may make to TGF-β mediated cell cycle arrest in these cells must occur by regulation at the post-translational level.

Considered together, our data suggest that p15 may function as an effector of TGF-β mediated cell cycle arrest via inhibition of CDK4 and CDK6 kinases. p15 may be the sole mediator of TGF-β induced arrest in some cells, or may cooperate with other TGF-β responsive pathways. TGF-β can regulate differentiation in some cell types, and the ability of TGF-β to affect cell cycle progression through p15 may also contribute to these processes.

Moreover, cytogenetic abnormalities at 9p21 are common in many types of human tumors suggesting the presence of a tumor suppressor gene at this locus. An inherited cancer syndrome which causes predisposition to melanoma also maps to 9p21. In addition to our data presented herein, and in WO 95/28483, p16 was initially proposed as a candidate for both of these activities based upon analysis of p16 deletions and point mutations in cell lines. However, the presence of a second functional member of the p16 family at 9p21 raises the possibility that loss of tumor suppression may involve inactivation of either or both genes. The response of p16 to viral oncoproteins indicates that it may function in intracellular growth regulatory pathways, while results presented here

AUSTRALIAN PATENT OFFICE

suggest that p15 may transduce extacellular growth inhibitory signals. Thus deletions of 9p21 which remove both genes (or other mutations that might inactivate both) could simultaneously negate two major proliferative control pathways. In this regard, the ability of TGF-β to induce growth arrest is reduced or lost in many neoplastically transformed cell lines. In particular, melanocytes are sensitive to growth inhibition by TGF-β, but many metastatic melanoma cells are TGF-β resistant.

Example 1

Demonstration of Selective Subunit Rearrangement of Cell-cycle Complexes In Association With Cellular Transformation by a DNA Tumor Virus or Its Oncogenic Product

(i) Cellular Transformation With DNA Tumor Virus SV40 Is Associated With Subunits Rearrangement of Cell-cycle Complexes

Preparation of [^{35}S] methionine-labelled cell lysates and polyacrylamide gel electrophoresis were as described above, as well as described in PCT Publication No. WO92/20796. Cell lysates were prepared from either human normal diploid fibroblast cells WI38 or DNA tumor virus SV40 transformed WI38 cells, VA13. Cell lysates were immunoprecipitated with antibodies against each cell-cycle gene products.

(ii) *Subunit Rearrangements of Cell-cycle Complexes In Two Different Pair Cell Lines*

Methods for preparation of cell lysates are the same as described above. Two different pair cell lines were used in these experiments. HSF43 is a normal human diploid fibroblast cell line and CT10 (full name CT10-2C-T1) is a derivative of HSF43 transformed by SV40 large tumor antigen. CV-1 is an African green monkey kidney cell line and COS-1 is a derivative of CV-1 transformed by SV40.

(iii) Cellular Transformation by DNA Tumor Virus SV40 Is Associated With Rearrangement of PCNA Subunit of Cell-cycle Complexes

Preparation of cell lysate, electrophoresis, and Western blotting conditions are the same as described above. Normal human diploid fibroblast cell lines and their SV40 transformed cell lines are described above. Immunoprecipitates derived from each antibody were separated on polyacrylamide gels and blotted with anti-PCNA antibody.

(iv) Cellular Transformation by DNA Tumor Virus SV40 Is Associated With Rearrangement of CDK4 Subunit of Cell-cycle Complexes

Preparation of cell lysate, electrophoresis, and Western blotting conditions are the same as previously described. Normal human diploid fibroblast cell lines and their SV40

transformed cell lines are described above. Immunoprecipitates derived from each antibody were separated on polyacrylamide gels and blotted with anti-CDK4 antibody.

Example 2
Cloning of p16, an inhibitor of CDK4 activity

(i) Cloning of p16 using the two hybrid assay

Saccharomyces cerevisiae YPB2 cells were transformed simultaneously with a plasmid containing a GAL4db-p16 fusion and with a plasmid containing, respectively, the GAL4ad fused to cdc2 (CDK1), CDK2, CDK4, CDK5, PCNA (proliferating cell nuclear antigen), and the fission yeast kinase Snf 1. After growing cells in medium selective for both plasmids (minus tryptophan and minus leucine), two colonies were picked randomly and were streaked in plates that either contained or lacked histidine. The ability to grow in the absence of histidine depends on the expression of the HIS3 gene that is under a GAL4-responsive promoter and, therefore, indicates that a functional GAL4 activator has been reconstituted through the interaction of p16 with the corresponding target protein.

(ii) Interaction of p16 CDKs

Purified bacterially-produced GSTp16 fusion protein was mixed with (^{35}S)-labeled *in vitro* translated cdc2, CDK2, CDK4 and CDK5. Mixtures contained 0.5 µg of purified GST-p16 and an equivalent amount of *in vitro* translated protein (between 0.5 to 5 ,µl; TNT Promega) in a final volume of 200 µl of a buffer containing 50 mM Tris-HCl pH 8, 120 mM NaCl and 0.5% Nonidet P-40. After 1 h at 4°C, 15 µl of glutathione-agarose beads were added and incubation was resumed for an additional hour. Beads were recovered by centrifugation, washed 4 times with the incubation buffer, and mixed with standard protein-gel loading buffer. Samples were loaded into a 15% poly-acryllamide gel and (^{35}S)-labeled proteins were detected by fluorography. The GSTp16 fusion protein was overexpressed in the pGEX-KG vector and purified by standard techniques. The *in vitro* translation templates were derived from the pBluescript vector (Stratagene).

(iii) Proteolytic mapping of p16

The *in vitro* translated (^{35}S)-labeled p16 (TNT Promega) was obtained using the p16 cDNA cloned into pBluescript vector (Stratagene) as a template, and the CDK4-associated p16 protein was co-immunoprecipitated with an anti-CDK4 serum from metabolically (^{35}S)-labeled HeLa cells lysates. Partial proteolysis was done over the corresponding gel slices after extensive equilibration in a buffer and digestion was accomplished by addition of NCS at different concentrations. The products were run in a 17.5% polyacrylamide gel and detected in a phosphoimager Fujix 2000.

(iv) Detecting the effects of p16 on CDK4-cyclin D complexes

Baculovirus-infected insect cells overexpressing p16, CDK4, cyclin D1, or both CDK4 and cyclin D1 together were metabolically (^{35}S)-labeled. The different incubation mixtures were composed by extracts containing p16, CDK4, cyclin D1 and both CDK4 and cyclin D1, and were immunoprecipitated with anti-p16 serum, anti-CDK4 serum without any previous preincubation, and anti-CDK serum preincubated with the peptide originally used to raise the antiserum and anti-cyclin D1 serum. Immunoprecipitates were then analyzed by SDS-PAGE.

Example 3
Chromosomal Mapping of p16

Genomic clones of the human p16 gene were isolated by stringency screening (68°C with 0.1xSSC wash) of a λFIXII human genomic library (Strategene) with cDNA probes. Isolated phage clones were confirmed by high stringency Southern hybridization and/or partial sequence analysis. Purified whole phage DNA was labelled for fluorescent *in situ* hybridization (FISH) analysis.

FISH analysis was performed using established methods (Demetrick et al. (1994) *Cytogenet Cell Genet* 66:72-74; Demetrick et al. (1993) *Genomics* 18:144-147; and DeMarini et al. (1991) *Environ Mol Mutagen* 18:222-223) on methotrexate or thymidine synchronized, phytohemagglutinin stimulated, normal peripheral blood lymphocytes. Suppression with a mixture of sonicated human DNA and cot1 DNA was required to reduce the background. The stained slides were counterstained with propidiem iodide (for an R banding pattern) or DAPI and actinomycin D (for a DA-DAPI banding pattern), mounted in antifade medium and visualized utilizing a Zeiss Axiphot microscope. Between 30 and 100 mitotises were examined for each gene location. Photographs were taken using a cooled CCD camera. Alignment of three color fluorescence was done under direct visualization through a triple bandpass filter (FITC/Texas Red/DAPI). The p16 gene was visualized to map to 9p21-22.

Example 4
Cloning Of Mouse CCR-proteins

Utilizing a probe consisting of the coding region of the human p16 gene, we have screened a mouse embryonal stem cell library and have isolated a genomic clone containing

the coding region for a mouse homolog of the human p16 gene described above. This clone was isolated under low to moderate stringency conditions (1 x SSC at 50°C). This DNA (14kB) has been cloned in two independent pieces and the restriction map for nine restriction endonucleases has been performed (see Figure 5). The mouse CCR-gene has been completely sequenced and the coding region is apparently made up of only two exons that have been located in the restriction map by Southern hybridization. The apparent molecular weight of the mouse CCR-protein is 13.5kDa, and the nucleic acid and amino acid sequence of the mouse p16 homolog, termed herein "p13.5", is given in SEQ ID No. 5 and 6 respectively.

Utilizing a probe based on the nucleotide sequence of the conserved region of Met1-Arg80 (see Figure 6) of the mouse p13.5 clone, a p19 embryonal carcinoma library (Stratagene) was probed under high stringency conditions. Several clones were isolated from the library which were significantly homologous to the p13.5 clone. Sequence comparison of one of the murine clone, the sequence for which is provided in SEQ ID Nos. 7 and 8, indicated that it was the murine homolog of p15 (described in Example 5 below).

Example 5
Cloning of p15 from human cells

The sequence of the p15 cDNA is shown along with the deduced amino acid sequence of the protein in SEQ ID Nos. 3 and 4 respectively. The deduced amino acid sequence of p15 was compared to that of p16 (e.g. see Figure 6), and areas of homology were identified using the BLAST program.

(i) Cell Culture.

HaCaT cells were routinely maintained in DMEM containing 10% fetal bovine serum (FBS) (Boukamp et al. (1988) *J Cell Biol.* 106:761-771). For TGF-β arrest, HaCaT cells were grown to confluence in DMEM containing 10% FBS and then serum starved for 3 days in DMEM containing 0.1% FBS. Cells were re-stimulated by addition of new media containing 10% FBS and 2 ng/ml TGF-β (purified from human platelets, Calbiochem). For library construction, RNA was prepared 22 hours after re-stimulation. For immunoprecipitations experiments, cells were labelled with ^{35}S-methionine in the presence of TGF-β for four hours beginning at 19 hours after re-stimulation.

(ii) Library Construction and Screening.

RNA was prepared from TGF-β treated cells and from cells that were serum starved and then re-stimulated in the absence of TGF-β using RNAZOL B according to the manufacturer's instructions. The cDNA library which was used to isolate the p15 clone was constructed from a mixture of RNA derived from treated and untreated cells. Messenger

RNA preparation and cDNA synthesis were exactly as previously described (Hannon et al. (1993) *Genes Dev.* 7:2378-2391). Double-stranded cDNA was ligated into λ-ZapII arms according to the manufacturers instructions. Low stringency hybridization was performed at 50°C in 500 mM $NaPO_4$, pH 7.0, 1 mM EDTA, 15% Formamide, 7% SDS, 0.1% bovine serum albumin (wash: 1 x SSC, 50°C). The p15 cDNA was also isolated from a human mammary epithelial cell library. Cell lysis, immunoprecipitation and protease mapping were performed exactly as previously described (Xiong et al. (1993) *Nature* 366:701-704) For chymotrypsin digestion, either 7μg or 1.5μg of enzyme was used. Gels for chymotrypsin mapping contained 0.1% SDS.

Example 6
p15 is a specific inhibitor of CDK4 and CDK6

Preparation of GST-p16 fusion protein from bacteria is described above. GST-p15 was prepared identically.

(i) p15 binds CDK4 and CDK6

In vitro translated CDKs were prepared using the TNT-lysate *in vitro* translation kit (Promega) according to the manufacturer's instructions. For the binding assay, GST-p15 or GST-p16 (250 ng) was incubated with *in vitro* translated CDKs, e.g. ^{35}S-labelled CDC2, CDK2, CDK4, CDK5 or CDK6, for 30 minutes at 30°C in 30 μl containing 20 mM Tris, pH 8.0, 10 mM $MgCl_2$, 1 mM EGTA. Following incubation, mixtures were diluted to 250 μl in IP buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 0.5% NP-40) and incubated for 1 hour with 12.5μl of glutathione sepharose. Bound proteins were recovered on glutathione sepharose, washed four times with 1 ml of IP buffer, and then released by boiling in SDS gel sample buffer. Bound proteins were analyzed by electrophoresis in a 17.5% PAGE gel. For comparison, a similar experiment using GST-p16 was carried out.

(ii) p15 binding inhibits cyclin D/CDK4 kinase

Active cyclin D/CDK4 kinase, present in lysates of baculovirus infected insect cells, was incubated with increasing quantities of GST-p15 for 30 minutes at 30°C. Preparation of baculovirus lysates containing active cyclin D/CDK4 kinase is described above. Lysates containing active cyclin D/CDK6 were prepared identically. For kinase assays 10μl of baculoviral lysates were mixed with approximately 0, 10ng, 20ng, 50ng, or 100ng of GST-p15 or GST-p16 and incubated for 30 minutes at 30°C in a total volume of 30μl containing 20 mM Tris, pH 8.0, 10 mM $MgCl_2$, 1mM EGTA. Following incubation, 0.5μg GST-RB and 0.5 μl of γ-^{32}P-ATP (5 μlCi, 3000 Ci/mMol) were added for 10 minutes at 30°C as substrates. Reactions were stopped by the addition of 250μl of IP buffer and 15μl of glutathione sepharose. After a further 1 hour of incubation at 4°C, beads were washed four

times with IP buffer before release of bound proteins and subsequent electrophoresis. Like p16, the p15-GST fusion protein was able to inhibit the kinase activity of CDK4.

(iii) p15 inhibits cyclin D/CDK6 kinase

Cyclin D/CDK6 complexes were produced in baculovirus infected insect cells and incubated with increasing quantities of GST-p15 or GST-p16 as described above. Following incubation, the ability of these complexes to catalyze GST-RB phosphorylation was determined. Both p15 and p16 demonstrated an ability to inhibit the CDK6 kinase activity.

Example 7

TGF-β treatment increases association of p15 with CDK4 and CDK6

HaCaT cells were cultured as described above. TGF-β treatment was initiated by adding TGF-β to existing media to 2ng/ml. For the last two hours of TGF-β treatment, cells were labelled with ^{35}S-methionine. For cell labelling, media was changed to DMEM minus methionine (Gibco-BRL) containing 2ng/ml of TGF-β, 10% dialyzed FBS and 0.5 mCi/ml of ^{35}S-methionine/cysteine (trans-label, NEN). Cells were lysed and proteins were immunoprecipitated using either anti-CDK4 or anti-CDK6 antibodies. Cell lysis and immunoprecipitation were as described above. HaCaT cultures treated with TGF-β in parallel were stained with DAPI and analyzed by FACS. For 8 hours following TGF-β addition, there was no appreciable change in the percentage of G_1 cells. After fourteen hours, the G_1 population began to increase. Arrest of HaCaT cells in G_1 (approx. 85% G_1 cells) required at least 24 hours after treatment of an asynchronous culture.

Example 8

TGF-β treatment induces p15 mRNA

Asynchronous HaCaT cultures were treated with TGF-β for the indicated times. Total RNA was prepared from treated cells and used for Northern blotting with a probe specific for p15 (e.g. probes consisted of the first coding exons of this genes and was prepared by PCR). Hybridization used for northern blots consisted of 200 mM $NaPO_4$, pH 7.0, 1 mM EDTA, 15% Formamide, 7% SDS, 0.1% bovine serum albumin at 65°C; wash: 0.2xSSC, 65°C. The p15 probe recognized three p15 mRNAs of approximately 0.8, 2.2 and 3.2 Kb. Northern signals were quantitated on a Fuji BAS2000 phosphorimager and plotted to give a graphical representation of the results. RNA amounts were normalized by mass, and over-probing of the blot with a human actin probe suggested no more than 10% variance in RNA amounts between lanes.

RNAs identical to those used in p15 probe panel were also probed with a fragment specific for p16. Furthermore, RNA from TGF-β treated HaCaT cells was probed

oligonucleotides derived from either a fragment of human CDK4 coding sequence or a fragment of the p27 cDNA.

Example 9
Generating a Transgenic Mouse p13.5 Knockout

The disruptive construct is formed by the two DNA regions of approx. 3-4Kb flanking the p13.5 gene. These DNA pieces are cloned at both sides of a gene marker that will be used to select the mouse embryonal stem (ES) cells that have incorporated this DNA after transfection. Regions which are homologous to the p16 locus are in turn flanked by another marker which allows selection against cells which have incorporated the disruption vector by non-homologous recombination (e.g. at a locus other than that of the mouse p13.5 gene). Those cells where insertion has occurred in the appropriate position are injected into mouse blastocytes and implanted into the appropriate female mice following standard protocols (*Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986; and Jaenisch (1988) *Science* 240:1468-1474.)

Chimeric pups resulting from the engineered blastocysts can be identified by a coat color marker specific to the transfected ES cells (agouti). Mice with high degrees of chimerism are crossed to identify those with chimeric germ lines and to generate non-chimeric heterozygous disruptants. Homozygous disruptants are derived by breeding the non-chimeric heterozygotes.

Example 10
Characterization of CCR-insensitive CDK4 Mutants

A mutant of CDK4 was discovered in cells from a melanoma patient, and the sequence of the mutant enzyme was determined. From the sequence, it was determined that that the gene contained a Arg→Cys at position 24 (see SEQ ID No. 9).

We reisolated the mutant by cloning CDK4 (see Matsushime et al. (1992) *Cell* 71:323-334) and generated a the cysteine mutation by oligonucleotide primer mutagenesis. To characterize the effect of the mutation, we compared the mutant and wild-type enzyme based on a number of different criteria, including intrinsic activity (e.g. did the mutant constitutively activate CDK4), as well as the ability of other regulatory proteins to control CDK4 activation. Briefly, we generated a series of baculovirus expression systems for over-expressing various proteins. In particular, Sf9 cell lysates (Desai et al. (1992) *Mol Cell Biol* 3:571-582, and Example 6 above) were obtained for mutant and wild-type CDK4, cyclin D1, p16, p15, p21 and p27 (see Polyak et al. (1994) *Genes Dev* 8:9-22; and Toyoshima et al. (1994) *Cell* 78:67-74). Using a GST-RB fusion protein (Example 6) as a substrate for

detecting CDK4 kinase activity, various combinations of lysate were lixed and tested for CDK4 activation/inhibition.

When the mutant CDK4 was expressed alone in Sf9 cells, no appreciable phosphorylation of the RB substrate was detected, as is also the case with the wild-type enzyme, indicating that the mutation did cause constitutive acitvation of CDK4. Overexpression of a CDK4 and cyclin D1 in an Sf9 lysate was also identical for both mutant and wild-type kinase, as each was shown to be activated in the presence of cyclin D1. However, upon addition of increasing amounts of either p16- or p15-containing lysate to the CDK4/cyclin D mixture, the wild-type CDK4 was inhibited yet the mutant CDK4 was relatively unaffected, indicating that the mutation gave rise to kinase whose activity is insensitive to either p15 or p16. Furthermore, immunoprecipitation demonstrated that neither p15 or p16 were capable of binding the mutant, as they were apparently lost from the complex which is ordinarily seen with the wild-type CDK4. Finally, similar experiments carried out with p21 and p27 indicated that the particular mutation, Arg24-Cys, did not effect the binding or inhibitory ability of either of those proteins. An analogous mutation to Arg31 of CDK6 (SEQ ID No. 10; and Bates et al. (1994) *Oncogene* 9:71-79 for the wild-type gene) is expected to have the same effect.

Utilizing the Arg-24 residue as a reference point, we have further identified by molecular modeling other residues which may also be involved in the recognition of p16/p15. Utilizing the coordinates for CDK2 (DeBondt et al. (1993) *Nature* 363:595-602; Endicott et al. (1994) *Prot Eng* 7:243-253; and Morgan et al. (1994) *Curr Opin Cell Biol* 6:239-246)) we have constructed a model for CDK4. Focusing our attention on residues in the spatial vicinity of Arg-24 and that are conserved between CDK4 and CDK6 (but different from CDK2 or CDC2, we have recombinantly generated and analyzed a number of new CDK4 mutants for their ability to bind p16. These mutants and their p16-binding abilities are summarized in Table 1 below (see also Figure 7).

Three changes abolished the interaction with p16. When these changes were visualized onto the 3-dimensional structure, it was apparent that these residues form a cluster of four amino acid residues accessible to solvent. These residues, K22, R24, H95 and D97 define a surface in the small lobe of CDK4, in very close proximity to the ATP binding site, but far away from the cyclin binding site or the substrate binding site. This surface likely represents at least a portion of the p16/p15-recognition surface present in CDK4 (and homologously in CDK6). Accordingly, an attractive model for p16/p15 inhibition of CDK4/CDK6 provides an occlusion or distorting effect to the ATP-binding site upon binding of the CCR protein such that ATP either does not bind to CDK4 or is not properly positioned to be used as a phosphate donor.

Table 1
p16 binding to CDK4 mutants

residue #	conservation	mutation	p16 binding
7	CDK4/CDK6 specific	E→Q	no effect
10	CDK4/CDK6 specific	A→E	no effect
11	CDK4/CDK6 specific	E→K	no effect
22	conserved in a CDKs	K→A	no binding to p16
24	conserved in a CDKs	R→S	no binding to p16
25	CDK4/CDK6 specific	D→N	no effect
31	CDK4/CDK6 specific	F→V	no effect
78	CDK4/CDK6 specific	C→I	no effect
81	CDK4/CDK6 specific	S→E	no effect
82 to 86	CDK4/CDK6 specific	RTDRE→N	no effect
95 to 97	CDK4/CDK6 specific	HVD→FLH	no binding to p16

All of the above-cited references and publications are hereby incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Cold Spring Harbor Laboratory
(B) STREET: One Bungtown Road
(C) CITY: Cold Spring Harbor
(D) STATE: NY
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 11724

(ii) TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses Related Thereto

(iii) NUMBER OF SEQUENCES: 10

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: AscII(text)

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/346,147
(B) FILING DATE: 29-NOV-1994

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/306,511
(B) FILING DATE: 14-SEP-1994

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/248,812
(B) FILING DATE: 25-MAY-1994

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/227,371
(B) FILING DATE: 14-APR-1994

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 994 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 41..508

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CGGAGAGGGG GAGAACAGAC AACGGGCGGC GGGGAGCAGC ATG GAT CCG GCG GCG          55
                                            Met Asp Pro Ala Ala
                                              1               5

GGG AGC AGC ATG GAG CCT TCG GCT GAC TGG CTG GCC ACG GCC GCG GCC          103
Gly Ser Ser Met Glu Pro Ser Ala Asp Trp Leu Ala Thr Ala Ala Ala
                  10                  15                  20

CGG GGT CGG GTA GAG GAG GTG CGG GCG CTG CTG GAG GCG GTG GCG CTG          151
Arg Gly Arg Val Glu Glu Val Arg Ala Leu Leu Glu Ala Val Ala Leu
              25                  30                  35

CCC AAC GCA CCG AAT AGT TAC GGT CGG AGG CCG ATC CAG GTC ATG ATG          199
Pro Asn Ala Pro Asn Ser Tyr Gly Arg Arg Pro Ile Gln Val Met Met
          40                  45                  50

ATG GGC AGC GCC CGA GTG GCG GAG CTG CTG CTG CTC CAC GGC GCG GAG          247
Met Gly Ser Ala Arg Val Ala Glu Leu Leu Leu Leu His Gly Ala Glu
      55                  60                  65

CCC AAC TGC GCC GAC CCC GCC ACT CTC ACC CGA CCC GTG CAC GAC GCT          295
Pro Asn Cys Ala Asp Pro Ala Thr Leu Thr Arg Pro Val His Asp Ala
  70                  75                  80                  85

GCC CGG GAG GGC TTC CTG GAC ACG CTG GTG GTG CTG CAC CGG GCC GGG          343
Ala Arg Glu Gly Phe Leu Asp Thr Leu Val Val Leu His Arg Ala Gly
                  90                  95                  100

GCG CGG CTG GAC GTG CGC GAT GCC TGG GGC CGT CTG CCC GTG GAC CTG          391
Ala Arg Leu Asp Val Arg Asp Ala Trp Gly Arg Leu Pro Val Asp Leu
              105                 110                 115

GCT GAG GAG CTG GGC CAT CGC GAT GTC GCA CGG TAC CTG CGC GCG GCT          439
Ala Glu Glu Leu Gly His Arg Asp Val Ala Arg Tyr Leu Arg Ala Ala
          120                 125                 130

GCG GGG GGC ACC AGA GGC AGT AAC CAT GCC CGC ATA GAT GCC GCG GAA          487
Ala Gly Gly Thr Arg Gly Ser Asn His Ala Arg Ile Asp Ala Ala Glu
      135                 140                 145

GGT CCC TCA GAC ATC CCC GAT TGAAAGAACC AGAGAGGCTC TGAGAAACCT             538
Gly Pro Ser Asp Ile Pro Asp
150                 155

CGGGAAACTT AGATCATCAG TCACCGAAGG TCCTACAGGG CCACAACTGC CCCCGCCACA        598

ACCCACCCCG CTTTCGTAGT TTTCATTTAG AAAATAGAGC TTTTAAAAAT GTCCTGCCTT        658

TTAACGTAGA TATAAGCCTT CCCCCACTAC CGTAAATGTC CATTTATATC ATTTTTTATA        718

TATTCTTATA AAAATGTAAA AAAGAAAAAC ACCGCTTCTG CCTTTTCACT GTGTTGGAGT        778

TTTCTGGAGT GAGCACTCAC GCCCTAAGCG CACATTCATG TGGGCATTTC TTGCGAGCCT        838

CGCAGCCTCC GGAAGCTGTC GACTTCATGA CAAGCATTTT GTGAACTAGG GAAGCTCAGG        898

GGGGTTACTG GCTTCTCTTG AGTCACACTG CTAGCAAATG GCAGAACCAA AGCTCAAATA        958
```

AAAATAAAAT TATTTTCATT CATTCACTCA AAAAAA 994

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 156 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asp Pro Ala Ala Gly Ser Ser Met Glu Pro Ser Ala Asp Trp Leu
1 5 10 15

Ala Thr Ala Ala Ala Arg Gly Arg Val Glu Glu Val Arg Ala Leu Leu
20 25 30

Glu Ala Val Ala Leu Pro Asn Ala Pro Asn Ser Tyr Gly Arg Arg Pro
35 40 45

Ile Gln Val Met Met Met Gly Ser Ala Arg Val Ala Glu Leu Leu Leu
50 55 60

Leu His Gly Ala Glu Pro Asn Cys Ala Asp Pro Ala Thr Leu Thr Arg
65 70 75 80

Pro Val His Asp Ala Ala Arg Glu Gly Phe Leu Asp Thr Leu Val Val
85 90 95

Leu His Arg Ala Gly Ala Arg Leu Asp Val Arg Asp Ala Trp Gly Arg
100 105 110

Leu Pro Val Asp Leu Ala Glu Glu Leu Gly His Arg Asp Val Ala Arg
115 120 125

Tyr Leu Arg Ala Ala Ala Gly Gly Thr Arg Gly Ser Asn His Ala Arg
130 135 140

Ile Asp Ala Ala Glu Gly Pro Ser Asp Ile Pro Asp
145 150 155

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 850 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 338..751

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GCGGGGCAGT GAGGACTCCG CGACGGTCCG CACCCTGCGG CCAGAGCGGC TTTGAGCTCG      60

GCTGCTTCCG CGCTAGGCGC TTTTTCCCAG AAGCAATCCA GGCGCGCCCG CTGGTTCTTG     120

AGCGCCAGGA AAAGCCCGGA GCTAACGACC GGCCGCTCGG CACTGCACGG GGCCCCAAGC     180

CGCAGAAGAA GGACGACGGG AGGGTAATGA AGCTGAGCCC AGGTCTCCTA GGAAGGAGAG     240

AGTGCGCCGG AGCAGCGTGG GAAAGAAGGG AAGAGTGTCG TTAAGTTTAC GGCCAACGGT     300

GGATTATCCG GGCCGCTGCG CGTCTGGGGG CTGCGGA ATG CGC GAG GAG AAC AAG       355
                                          Met Arg Glu Glu Asn Lys
                                           1               5

GGC ATG CCC AGT GGG GGC GGC AGC GAT GAG GGT CTG GCC AGC GCC GCG         403
Gly Met Pro Ser Gly Gly Gly Ser Asp Glu Gly Leu Ala Ser Ala Ala
            10                  15                  20

GCG CGG GGA CTA GTG GAG AAG GTG CGA CAG CTC CTG GAA GCC GGC GCG         451
Ala Arg Gly Leu Val Glu Lys Val Arg Gln Leu Leu Glu Ala Gly Ala
        25                  30                  35

GAT CCC AAC GGA GTC AAC CGT TTC GGG AGG CGC GCG ATC CAG GTC ATG         499
Asp Pro Asn Gly Val Asn Arg Phe Gly Arg Arg Ala Ile Gln Val Met
    40                  45                  50

ATG ATG GGC AGC GCC CGC GTG GCG GAG CTG CTG CTG CTC CAC GGC GCG         547
Met Met Gly Ser Ala Arg Val Ala Glu Leu Leu Leu Leu His Gly Ala
55                  60                  65                  70

GAG CCC AAC TGC GCA GAC CCT GCC ACT CTC ACC CGA CCG GTG CAT GAT         595
Glu Pro Asn Cys Ala Asp Pro Ala Thr Leu Thr Arg Pro Val His Asp
                75                  80                  85

GCT GCC CGG GAG GGC TTC CTG GAC ACG CTG GTG GTG CTG CAC CGG GCC         643
Ala Ala Arg Glu Gly Phe Leu Asp Thr Leu Val Val Leu His Arg Ala
            90                  95                  100

GGG GCG CGG CTG GAC GTG CGC GAT GCC TGG GGT CGT CTG CCC GTG GAC         691
Gly Ala Arg Leu Asp Val Arg Asp Ala Trp Gly Arg Leu Pro Val Asp
        105                 110                 115

TTG GCC GAG GAG CGG GGC CAC CGC GAC GTT GCA GGG TAC CTG CGC ACA         739
Leu Ala Glu Glu Arg Gly His Arg Asp Val Ala Gly Tyr Leu Arg Thr
    120                 125                 130

GCC ACG GGG GAC TGACGCCAGG TTCCCCAGCC GCCCACAACG ACTTTATTTT             791
Ala Thr Gly Asp
135

CTTACCCAAT TTCCCACCCC CACCCACCTA ATTCGATGAA GGCTGCCAAC GGGGAGCGG        850
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 138 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Arg Glu Glu Asn Lys Gly Met Pro Ser Gly Gly Gly Ser Asp Glu
1 5 10 15

Gly Leu Ala Ser Ala Ala Ala Arg Gly Leu Val Glu Lys Val Arg Gln
20 25 30

Leu Leu Glu Ala Gly Ala Asp Pro Asn Gly Val Asn Arg Phe Gly Arg
35 40 45

Arg Ala Ile Gln Val Met Met Met Gly Ser Ala Arg Val Ala Glu Leu
50 55 60

Leu Leu Leu His Gly Ala Glu Pro Asn Cys Ala Asp Pro Ala Thr Leu
65 70 75 80

Thr Arg Pro Val His Asp Ala Ala Arg Glu Gly Phe Leu Asp Thr Leu
85 90 95

Val Val Leu His Arg Ala Gly Ala Arg Leu Asp Val Arg Asp Ala Trp
100 105 110

Gly Arg Leu Pro Val Asp Leu Ala Glu Glu Arg Gly His Arg Asp Val
115 120 125

Ala Gly Tyr Leu Arg Thr Ala Thr Gly Asp
130 135

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 853 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 213..587

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
GGAGTACAGC AGCGGGAGCA TGGGTCGCAG GTTCTTGGTC ACTGTAAGGA TTCAGCGCGC      60

GGGCCGCCCA CTCCAAGAGA GGGTTTTCTT GGTGAAGTTC GTGCGATCCC GGAGACCCAG     120

GACAGCGAGC TGCGCTCTGG CTTTCGTGAA CATGTTGTTG AGGCTAGAGA GGATCTTGAG     180

AAGAGGGCCG CACCGGAATC CTGGACCAGG TG ATG ATG ATG GGC AAC GTT CAC        233
                                   Met Met Met Gly Asn Val His
                                     1               5

GTA GCA GCT CTT CTG CTC AAC TAC GGT GCA GAT TCG AAC TGC GAG GAC        281
Val Ala Ala Leu Leu Leu Asn Tyr Gly Ala Asp Ser Asn Cys Glu Asp
        10                  15                  20

CCC ACT ACC TTC TCC CGC CCG GTG CAC GAC GCA GCG CGG GAA GGC TTC        329
Pro Thr Thr Phe Ser Arg Pro Val His Asp Ala Ala Arg Glu Gly Phe
    25                  30                  35

CTG GAC ACG CTG GTG GTG CTG CAC GGG TCA GGG GCT CGG CTG GAT GTG        377
Leu Asp Thr Leu Val Val Leu His Gly Ser Gly Ala Arg Leu Asp Val
 40                  45                  50                  55

CGC GAT GCC TGG GGT CGC CTG CCG CTC GAC TTG GCC CAA GAG CGG GGA        425
Arg Asp Ala Trp Gly Arg Leu Pro Leu Asp Leu Ala Gln Glu Arg Gly
                60                  65                  70

CAT CAA GAC ATC GTG CGA TAT TTG CGT TCC GCT GGG TGC TCT TTG TGT        473
His Gln Asp Ile Val Arg Tyr Leu Arg Ser Ala Gly Cys Ser Leu Cys
            75                  80                  85

TCC GCT GGG TGG TCT TTG TGT ACC GCT GGG AAC GTC GCC CAG ACC GAC        521
Ser Ala Gly Trp Ser Leu Cys Thr Ala Gly Asn Val Ala Gln Thr Asp
        90                  95                  100

GGG CAT AGC TTC AGC TCA AGC ACG CCC AGG GCC CTG GAA CTT CGC GGC        569
Gly His Ser Phe Ser Ser Ser Thr Pro Arg Ala Leu Glu Leu Arg Gly
    105                 110                 115

CAA TCC CAA GAG CAG AGC TAAATCCGCC TCAGCCCGCC TTTTTCTTCT               617
Gln Ser Gln Glu Gln Ser
120                 125

TAGCTTCACT TCTAGCGATG CTAGCGTGTC TAGCATGTGG CTTTAAAAAA TACATAATAA      677

TGCTTTTTTT GCAATCACGG GAGGGAGCAG AGGGAGGGAG CAGAAGGAGG GAGGGAGGGA      737

GGGAGGGACC TGGACAGGAA AGGAATGGCA TGAGAAACTG AGCGAAGGCG GCCGCGAAGG      797

GAATAATGGC TGGATTGTTT AAAAAAATAA AATAAAGATA CTTTTTAAAA TGTCAA          853
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 125 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Met Met Gly Asn Val His Val Ala Ala Leu Leu Leu Asn Tyr Gly
 1               5                  10                  15

Ala Asp Ser Asn Cys Glu Asp Pro Thr Thr Phe Ser Arg Pro Val His
            20                  25                  30

Asp Ala Ala Arg Glu Gly Phe Leu Asp Thr Leu Val Val Leu His Gly
        35                  40                  45

Ser Gly Ala Arg Leu Asp Val Arg Asp Ala Trp Gly Arg Leu Pro Leu
    50                  55                  60

Asp Leu Ala Gln Glu Arg Gly His Gln Asp Ile Val Arg Tyr Leu Arg
65                  70                  75                  80

Ser Ala Gly Cys Ser Leu Cys Ser Ala Gly Trp Ser Leu Cys Thr Ala
                85                  90                  95

Gly Asn Val Ala Gln Thr Asp Gly His Ser Phe Ser Ser Ser Thr Pro
            100                 105                 110

Arg Ala Leu Glu Leu Arg Gly Gln Ser Gln Glu Gln Ser
        115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 580 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 91..480

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
CTCCAGAGGG GAGGAGCCGC TCAGAGACCA GGCTGTAGCA ATCTCACGGC CGGCGAAGGA      60

CCATTTCTGC CACAGACCGG GGACAAGGGC ATG TTG GGC GGC AGC AGT GAC GCG      114
                                 Met Leu Gly Gly Ser Ser Asp Ala
                                   1               5

GGC CTG GCC ACC GCC GCG GCG CGG GGG CAA GTG GAG ACG GTG CGG CAG      162
Gly Leu Ala Thr Ala Ala Ala Arg Gly Gln Val Glu Thr Val Arg Gln
     10                  15                  20
```

```
CTC CTG GAA GCC GGC GCA GAT CCC AAC GCC CTG AAC CGC TTC GGG AGG          210
Leu Leu Glu Ala Gly Ala Asp Pro Asn Ala Leu Asn Arg Phe Gly Arg
 25                  30                  35                  40

CGC CCA ATC CAG GTC ATG ATG ATG GGC AGC GCC AGG GTG GCA GAG CTG          258
Arg Pro Ile Gln Val Met Met Met Gly Ser Ala Arg Val Ala Glu Leu
                 45                  50                  55

CTG CTG CTC CAC GGA GCA GAA CCC AAC TGC GCC GAC CCT GCC ACC CTT          306
Leu Leu Leu His Gly Ala Glu Pro Asn Cys Ala Asp Pro Ala Thr Leu
             60                  65                  70

ACC AGA CCT GTG CAC GAC GCA GCT CGG GAA GGC TTC CTG GAC ACG CTT          354
Thr Arg Pro Val His Asp Ala Ala Arg Glu Gly Phe Leu Asp Thr Leu
         75                  80                  85

GTC GTG CTG CAC CGG GCA GGG GCG CGG TTG GAT GTG TGT GAC GCC TGG          402
Val Val Leu His Arg Ala Gly Ala Arg Leu Asp Val Cys Asp Ala Trp
     90                  95                 100

GGC CGC CTG CCG GTA GAC TTG GCT GAA GAG CAG GGC CAC CGT GAC ATT          450
Gly Arg Leu Pro Val Asp Leu Ala Glu Glu Gln Gly His Arg Asp Ile
105                 110                 115                 120

GCG AGG TAT CTG CAC GCT GCC ACT GGA GAT TGACTGCGGG TTCCCTCCGC            500
Ala Arg Tyr Leu His Ala Ala Thr Gly Asp
                125                 130

CTTCCGCAAG GACTTCTTTC TCCCCAGCCC CATCTAGGAA GACTGTAAGC ACGAAGAGGC        560

CACCAGCGCC CAGCCTGCAG                                                    580
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 130 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Leu Gly Gly Ser Ser Asp Ala Gly Leu Ala Thr Ala Ala Ala Arg
 1               5                  10                  15

Gly Gln Val Glu Thr Val Arg Gln Leu Leu Glu Ala Gly Ala Asp Pro
            20                  25                  30

Asn Ala Leu Asn Arg Phe Gly Arg Arg Pro Ile Gln Val Met Met Met
        35                  40                  45

Gly Ser Ala Arg Val Ala Glu Leu Leu Leu Leu His Gly Ala Glu Pro
    50                  55                  60

Asn Cys Ala Asp Pro Ala Thr Leu Thr Arg Pro Val His Asp Ala Ala
65                  70                  75                  80
```

Arg Glu Gly Phe Leu Asp Thr Leu Val Val Leu His Arg Ala Gly Ala
85 90 95

Arg Leu Asp Val Cys Asp Ala Trp Gly Arg Leu Pro Val Asp Leu Ala
100 105 110

Glu Glu Gln Gly His Arg Asp Ile Ala Arg Tyr Leu His Ala Ala Thr
115 120 125

Gly Asp
130

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Ala Thr Ser Arg Tyr Glu Pro Val Ala Glu Ile Gly Val Gly Ala
1 5 10 15

Tyr Gly Thr Val Tyr Xaa Ala Xaa Asp Pro His Ser Gly His Phe Val
20 25 30

Ala Leu Lys Ser Val Arg Val Pro Asn Gly Gly Gly Gly Gly Gly Gly
35 40 45

Leu Pro Ile Ser Thr Val Arg Glu Val Ala Leu Leu Arg Arg Leu Glu
50 55 60

Ala Phe Glu His Pro Asn Val Val Arg Leu Met Asp Val Cys Ala Thr
65 70 75 80

Ser Arg Thr Asp Arg Glu Ile Lys Val Thr Leu Val Phe Glu Xaa Val
85 90 95

Xaa Gln Asp Leu Arg Thr Tyr Leu Asp Lys Ala Pro Pro Pro Gly Leu
100 105 110

Pro Ala Glu Thr Ile Lys Asp Leu Met Arg Gln Phe Leu Arg Gly Leu
115 120 125

Asp Phe Leu His Ala Asn Cys Ile Val His Arg Asp Leu Lys Pro Glu
130 135 140

Asn Ile Leu Val Thr Ser Gly Gly Thr Val Lys Leu Ala Asp Phe Gly
145 150 155 160

Leu Ala Arg Ile Tyr Ser Tyr Gln Met Ala Leu Thr Pro Val Val Val
165 170 175

Thr Leu Trp Tyr Arg Ala Pro Glu Val Leu Leu Gln Ser Thr Tyr Ala
180 185 190

Thr Pro Val Asp Met Trp Ser Val Gly Cys Ile Phe Ala Glu Met Phe
195 200 205

Arg Arg Lys Pro Leu Phe Cys Gly Asn Ser Glu Ala Asp Gln Leu Gly
210 215 220

Lys Ile Phe Asp Leu Ile Gly Leu Pro Pro Glu Asp Asp Trp Pro Arg
225 230 235 240

Asp Val Ser Leu Pro Arg Gly Ala Phe Pro Pro Arg Gly Pro Arg Pro
245 250 255

Val Gln Ser Val Val Pro Glu Met Glu Glu Ser Gly Ala Gln Leu Leu
260 265 270

Leu Glu Met Leu Thr Phe Asn Pro His Lys Arg Ile Ser Ala Phe Arg
275 280 285

Ala Leu Gln His Ser Tyr Leu His Lys Asp Glu Gly Asn Pro Glu
290 295 300

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 326 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Glu Lys Asp Gly Leu Cys Arg Ala Asp Gln Gln Tyr Glu Cys Val
1 5 10 15

Ala Glu Ile Gly Glu Gly Ala Tyr Gly Lys Val Phe Xaa Ala Xaa Asp
20 25 30

Leu Lys Asn Gly Gly Arg Phe Val Ala Leu Lys Arg Val Arg Val Gln
35 40 45

Thr Gly Glu Glu Gly Met Pro Leu Ser Thr Ile Arg Glu Val Ala Val
50 55 60

Leu Arg His Leu Glu Thr Phe Glu His Pro Asn Val Val Arg Leu Phe
65 70 75 80

Asp Val Cys Thr Val Ser Arg Thr Asp Arg Glu Thr Lys Leu Thr Leu
85 90 95

Val Phe Glu Xaa Val Xaa Gln Asp Leu Thr Thr Tyr Leu Asp Lys Val
100 105 110

Pro Glu Pro Gly Val Pro Thr Glu Thr Ile Lys Asp Met Met Phe Gln
115 120 125

Leu Leu Arg Gly Leu Asp Phe Leu His Ser His Arg Val Val His Arg
130 135 140

Asp Leu Lys Pro Gln Asn Ile Leu Val Thr Ser Ser Gly Gln Ile Lys
145 150 155 160

Leu Ala Asp Phe Gly Leu Ala Arg Ile Tyr Ser Phe Gln Met Ala Leu
165 170 175

Thr Ser Val Val Val Thr Leu Trp Tyr Arg Ala Pro Glu Val Leu Leu
180 185 190

Gln Ser Ser Tyr Ala Thr Pro Val Asp Leu Trp Ser Val Gly Cys Ile
195 200 205

Phe Ala Glu Met Phe Arg Arg Lys Pro Leu Phe Arg Gly Ser Ser Asp
210 215 220

Val Asp Gln Leu Gly Lys Ile Leu Asp Val Ile Gly Leu Pro Gly Glu
225 230 235 240

Glu Asp Trp Pro Arg Asp Val Ala Leu Pro Arg Gln Ala Phe His Ser
245 250 255

Lys Ser Ala Gln Pro Ile Glu Lys Phe Val Thr Asp Ile Asp Glu Leu
260 265 270

Gly Lys Asp Leu Leu Leu Lys Cys Leu Thr Phe Asn Pro Ala Lys Arg
275 280 285

Ile Ser Ala Tyr Ser Ala Leu Ser His Pro Tyr Phe Gln Asp Leu Glu
290 295 300

Arg Cys Lys Glu Asn Leu Asp Ser His Leu Pro Pro Ser Gln Asn Thr
305 310 315 320

Ser Glu Leu Asn Thr Ala
325

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Asp Pro Ala Ala Gly Ser Ser Met Glu Pro Ser Ala Asp Trp Leu
1 5 10 15

Ala Thr Ala Ala Ala Arg Gly Arg Val Glu Glu Val Arg Ala Leu Leu
20 25 30

Glu Ala Val Ala Leu Pro Asn Ala Pro Asn Ser Tyr Gly Arg Arg Pro
35 40 45

Ile Gln Val Met Met Met Gly Xaa Xaa Xaa Val Ala Xaa Leu Leu Leu
50 55 60

Xaa Xaa Gly Ala Xaa Xaa Asn Cys Xaa Asp Pro Xaa Thr Xaa Xaa Xaa
65 70 75 80

Arg Pro Val His Asp Ala Ala Arg Glu Gly Phe Leu Asp Thr Leu Val
85 90 95

Val Leu His Xaa Xaa Gly Ala Arg Leu Asp Val Arg Asp Ala Trp Gly
100 105 110

Arg Leu Pro Xaa Asp Leu Ala Xaa Glu Xaa Gly His Xaa Asp Xaa Xaa
115 120 125

Xaa Tyr Leu Arg Xaa Ala Xaa Gly Gly Thr Arg Gly Ser Asn His Ala
130 135 140

Arg Ile Asp Ala Ala Glu Gly Pro Ser Asp Ile Pro Asp
145 150 155

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 138 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Arg Glu Glu Asn Lys Gly Met Pro Ser Gly Gly Gly Ser Asp Glu
1 5 10 15

Gly Leu Ala Thr Pro Ala Arg Gly Leu Val Glu Lys Val Arg His Ser
20 25 30

Trp Glu Ala Gly Ala Asp Pro Asn Gly Val Asn Arg Phe Gly Arg Arg
35 40 45

PATENT OFFICE

Ala Ile Gln Val Met Met Met Gly Xaa Xaa Xaa Val Ala Xaa Leu Leu
50 55 60

Leu Xaa Xaa Gly Ala Xaa Xaa Asn Cys Xaa Asp Pro Xaa Thr Xaa Xaa
65 70 75 80

Xaa Arg Pro Val His Asp Ala Ala Arg Glu Gly Phe Leu Asp Thr Leu
85 90 95

Val Val Leu His Xaa Xaa Gly Ala Arg Leu Asp Val Arg Asp Ala Trp
100 105 110

Gly Arg Leu Pro Xaa Asp Leu Ala Xaa Glu Xaa Gly His Xaa Asp Xaa
115 120 125

Xaa Xaa Tyr Leu Arg Xaa Ala Xaa Gly Asp
130 135

AUSTRALIAN PATENT OFFICE

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. An isolated or recombinant cell-cycle regulatory (CCR) polypeptide including an amino acid sequence characterized by ankyrin-like repeats, which polypeptide binds to a cyclin-dependent kinase (CDK) and is at least 90% homologous to a polypeptide selected from the group consisting of SEQ ID No. 4 and SEQ ID No. 8.

2. The polypeptide of claim 1, wherein said CDK is a G1 phase CDK.

3. The polypeptide of claim 1, wherein said CDK is a G1 phase CDK.

4. The CCR polypeptide of claim 1, which polypeptide includes an amino acid sequence at least 90% homologous to an amino acid sequence represented by the general formula:

Met-Met-Met-Gly Xaa-Xaa-Xaa-Val-Ala-Xaa Leu-Leu-Leu-Xaa-Xaa-Gly Ala-Xaa-Xaa-Asn-Cys-Xaa-Asp-Pro Xaa Thr-Xaa-Xaa-Xaa-Arg Pro-Val-His-Asp-Ala-Ala Arg-Glu-Gly-Phe-Leu-Asp-Thr-Leu-Val-Val-Leu-His-Xaa-Xaa-Gly-Ala-Arg-Leu-Asp-Val-Arg-Asp-Ala-Trp-Gly Arg-Leu-Pro-Xaa-Asp-Leu-Ala-Xaa-Glu-Xaa-Gly-His-Xaa-Asp-Xaa-Xaa-Xaa Tyr-Leu-Arg-Xaa-Ala-Xaa-Gly.

5. The CCR polypeptide of claim 1, which polypeptide functions in one of either role of an agonist of cell-cycle regulation or an antagonist of cell-cycle regulation.

6. A substantially pure preparation of a p15 polypeptide including an amino acid sequence at least 90% homologous to SEQ ID NO:4, or a fragment thereof, which polypeptide or fragment binds to CDK.

AUSTRALIAN PATENT OFFICE

7. The polypeptide of claim 6, including an amino acid sequence represented by the general formula:

Met-Arg-Glu-Glu-Asn-Lys-Gly-Met-Pro-Ser-Gly-Gly-Gly-Ser-
Asp-Glu-Gly-Leu-Ala-Thr-Pro-Ala-Arg-Gly-Leu-Val-Glu-Lys-
Val-Arg-His-Ser-Trp-Glu-Ala-Gly-Ala-Asp-Pro-Asn-Gly-Val-
Asn-Arg-Phe-Gly-Arg-Arg-Ala-Ile-Gln-Val-Met-Met-Met-Gly-
Xaa-Xaa-Xaa-Val-Ala-Xaa-Leu-Leu-Leu-Xaa-Xaa-Gly-Ala-Xaa-
Xaa-Asn-Cys-Xaa-Asp-Pro-Xaa-Thr-Xaa-Xaa-Xaa-Arg-Pro-Val-
His-Asp-Ala-Ala-Arg-Glu-Gly-Phe-Leu-Asp-Thr-Leu-Val-Val-
Leu-His-Xaa-Xaa-Gly-Ala-Arg-Leu-Asp-Val-Arg-Asp-Ala-Trp-
Gly-Arg-Leu-Pro-Xaa-Asp-Leu-Ala-Xaa-Glu-Xaa-Gly-His-Xaa-
Asp-Xaa-Xaa-Xaa-Tyr-Leu-Arg-Xaa-Ala-Xaa-Gly-Asp.

8. The polypeptide of claim 6, wherein said polypeptide functions in one of either role of an agonist of cell cycle regulation or an antagonist of cell cycle regulation.

9. An isolated or recombinant polypeptide comprising an amino acid sequence homologous to an amino acid sequence represented in one or more of SEQ ID Nos. 4 or 8, which polypeptide specifically interacts with cyclin-dependent kinase 4 (CDK4), which specific interaction is affected by residues K22, R24, H95 and D97 of CDK4.

10. An antibody preparation specifically immunoreactive with a polypeptide of claim 1.

11. A fusion protein, which protein binds to a cyclin-dependent kinase (CDK), including a first amino acid sequence at least 60% homologous to an amino acid sequence represented in at least one of SEQ ID Nos. 4 or 8, and a second polypeptide having an amino acid sequence unrelated to the polypeptide of SEQ ID Nos. 2, 4, or 8, wherein said fusion protein is functional in a two-hybrid assay.

AUSTRALIAN PATENT OFFICE

12. A probe/primer including a substantially purified oligonucleotide, wherein said oligonucleotide hybridises under stringent conditions to a nucleic acid sequence corresponding to at least 12 consecutive nucleotides of a nucleic acid selected from the group consisting of SEQ ID NO:3 and SEQ ID NO:7, in which said nucleic acid sequence encodes a polypeptide that binds to a CDK.

13. The probe/primer of claim 12, further including a label group attached thereto and able to be detected.

14. The probe/primer of claim 13, wherein said label group is selected from a group consisting of radioisotopes, fluorescent compounds, enzymes, and enzyme cofactors.

15. A diagnostic test kit for identifying transformed cells, including the probe/primer of claim 12, for measuring, in a sample of cells isolated from a patient, a level of a nucleic acid encoding a cell-cycle regulatory protein.

16. An assay for screening test compounds for an inhibitor of an interaction of a cell-cycle regulatory (CCR) polypeptide selected from the group consisting of SEQ ID Nos. 2, 4, and 8 with a cyclin-dependent kinase (CDK) including

i. combining a CDK, and a CCR polypeptide including an amino acid sequence characterised by ankyrin-like repeats, which polypeptide binds to said CDK, under conditions wherein said CDK and said CCR polypeptide are able to interact;

ii. contacting said combination with a test compound; and

iii. detecting the formation of a complex including said CDK and said CCR polypeptide, wherein a decrease in the formation of said complex in the presence of said test compound is indicative of an inhibitor of the interaction between a CDK and a CCR polypeptide.

AUSTRALIAN PATENT OFFICE

17. A method of identifying an agent which disrupts the ability of a CCR protein selected from the group consisting of SEQ ID Nos. 2, 4, and 8 to regulate a eukaryotic cell cycle, including:

i. providing a two-hybrid assay system including

(a) a first fusion protein including a cyclin-dependent kinase (CDK), and

(b) a second fusion protein including a CCR polypeptide including an amino acid sequence characterized by ankyrin-like repeats, which polypeptide binds to said CDK, under conditions wherein said two-hybrid assay is sensitive to interactions between the CDK of said first fusion protein and said CCR polypeptide of said second polypeptide;

ii. contacting said two-hybrid assay system with a candidate agent;

iii. measuring a level of interactions between said fusion proteins in the presence of said candidate agent; and

iv. comparing the level of interaction of said fusion proteins in the presence of said candidate agent to a level of interaction of said fusion proteins in the absence of the candidate agent, wherein a decrease in the level of interaction in the presence of said candidate agent is indicative of inhibition of an interaction of a CCR protein with a cyclin-dependent kinase.

18. A diagnostic assay for identifying a cell or cells at risk for a disorder characterized by unwanted cell proliferation or differentiation, including detecting, in a cell sample, the presence or absence of a genetic lesion characterized by at least one of (i) aberrant modification or mutation of a gene encoding a cell-cycle regulatory (CCR) protein selected from the group consisting of SEQ ID Nos:2, 4, and 8, and (ii) misexpression of said gene; wherein a wild-type form of said gene encodes a CCR protein

AUSTRALIAN PATENT OFFICE

characterized by an ability to bind to a cyclin dependent kinase and an amino acid sequence including a series of ankyrin-like repeats.

19. The assay of claim 18 herein detecting said lesion includes ascertaining, relative to a wild-type CCR protein, the ability of said CCR protein from said cell sample to bind a cyclin-dependent kinase (CDK).

20. The assay of claim 19, wherein detecting said genetic lesion includes

i. providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of SEQ ID No:1 or 3, or naturally occurring mutants thereof, or 5′ or 3′ flanking or intronic sequences naturally associated with said gene;

ii. exposing said probe/primer to nucleic acid of said tissue; and

iii. detecting, by hybridization of said probe/primer to said nucleic acid, the presence or absence of said genetic lesion.

21. The method of claim 19 wherein the level of said protein is detected in an immunoassay.

22. The assay of claim 19, wherein the ability of said CCR protein to bind a CDK is ascertained by:

i. providing a two-hybrid assay system including a first fusion protein including a CDK protein portion, and a second fusion protein including a CCR protein portion cloned from said cell sample, under conditions wherein said two hybrid assay is sensitive to interactions between the CDK portion of said first fusion protein and said CCR protein portion of said second polypeptide;

ii. measuring a level of interactions between

AUSTRALIAN PATENT OFFICE

said fusion proteins; and

iii. comparing the level of interaction of said fusion proteins to a level of interaction of said first fusion protein with a second fusion protein including a wild-type CCR protein instead of the cloned CCR protein, wherein a decrease in the level of interaction is indicative of a lesion to said gene which disrupts the normal cellular function of said gene.

23. A diagnostic method for detecting mutation of a CCR polypeptide including

i. isolating a gene encoding a sample protein including an amino acid sequence characterised by ankyrin-like repeats, which polypeptide binds to a CDK and is at least 90% homologous to a polypeptide selected from the group consisting of SEQ ID NO:4 and SEQ ID NO:8 or a biologically active fragment thereof;

ii. cloning said gene into a two hybrid assay to produce a host cell including

(a) a reporter gene operably linked to a transcriptional regulatory sequence,

(b) a first chimeric gene which encodes a first fusion protein, said first fusion protein including a target protein to which a wild-type form of said sample protein binds and to which binding of the cloned sample protein is to be assessed,

(c) a second chimeric gene which encodes a second fusion protein including said cloned sample protein, wherein interaction of said first and second fusion proteins, if at all, is mediated by binding of said target protein and sample protein portions, and expression of said reporter gene is sensitive to interactions between said transcriptional regulatory sequence and a complex of said first and second fusion proteins;

iii. measuring expression of said reporter gene; and

iv. comparing the level of expression of said reporter gene to a level of expression in a two hybrid

AUSTRALIAN PATENT OFFICE

assay having a second fusion protein including a wild-type form of said sample protein instead of the cloned sample protein, wherein a decrease in the level of expression is indicative of a mutation to said gene which mutation disrupts the ability of said sample protein to bind to another cellular protein.

24. A diagnostic assay for identifying a cell or cells at risk for a disorder characterised by unwanted cell proliferation or differentiation, including detecting, in a cell sample, by determining the phenotype of a cell including detecting the presence or absence of a genetic lesion characterised by at least one of (i) aberrant modification or mutation of a gene encoding a CDK-inhibitory protein, and (ii) mis-expression of said gene; wherein a wild-type form of said gene encodes a CDK-inhibitory protein and hybridizes under stringent conditions to a nucleic acid represented in SEQ ID No. 3 characterised by an ability to bind to a cyclin-dependent kinase and an amino acid sequence including a series of ankyrin-like repeats.

25. The assay of claim 24, wherein detecting said lesion includes:

i. providing a diagnostic probe including a nucleic acid including a region of nucleotide sequence which hybridizes to a sense or antisense sequence of said gene, or naturally occurring mutants thereof, or 5′ or 3′ flanking sequences naturally associated with said gene;

ii. combining said probe with nucleic acid of said cell sample; and

iii. detecting, by hybridization of said probe to said cellular nucleic acid, the existence of at least one of a deletion of one or more nucleotides from said gene, an addition of one or more nucleotides to said gene, a substitution of one or more nucleotides of said gene, a gross chromosomal rearrangement of all or a portion of said

AUSTRALIAN PATENT OFFICE

gene, a gross alteration in the level of an mRNA transcript of said gene, or a non-wild type splicing pattern of an mRNA transcript of said gene.

26. A transgenic non-human animal having germline or somatic cells including a chromosomally incorporated transgene that alters the biological activity of a genomic CCR gene, wherein said genomic CCR gene is selected from the group consisting of SEQ ID Nos:4 and 8 and encodes an CCR protein including ankyrin repeats and which CCR protein specifically bind to CDK4 or CDK6.

27. Cells isolated from the transgenic animal of claim 26, which cells include the chromosomally incorporated transgene.

Dated this 3rd day of February 1999
COLD SPRING HARBOR LABORATORY
By their Patent Attorneys
GRIFFITH HACK
Fellows Institute of Patent and
Trade Mark Attorneys of Australia

AUSTRALIAN PATENT OFFICE

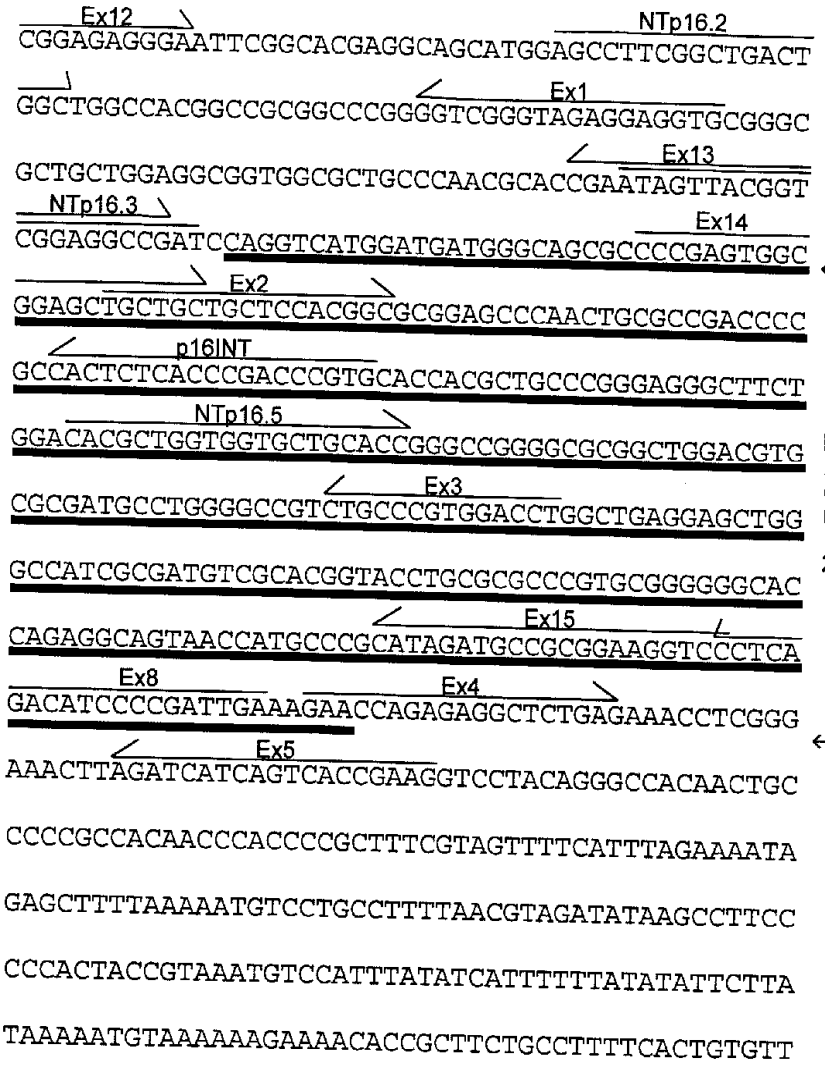
Ex12
NTp16.2
CGGAGAGGGAATTCGGCACGAGGCAGCATGGAGCCTTCGGCTGACT
Ex1
GGCTGGCCACGGCCGCGGCCCGGGGTCGGGTAGAGGAGGTGCGGGC
Ex13
GCTGCTGGAGGCGGTGGCGCTGCCCAACGCACCGAATAGTTACGGT
NTp16.3
Ex14
CGGAGGCCGATCCAGGTCATGGATGATGGGCAGCGCCCCGAGTGGC
Ex2
GGAGCTGCTGCTGCTCCACGGCGCGGAGCCCAACTGCGCCGACCCC
p16INT
GCCACTCTCACCCGACCCGTGCACCACGCTGCCCGGGAGGGCTTCT
NTp16.5
GGACACGCTGGTGGTGCTGCACCGGGCCGGGGCGCGGCTGGACGTG
Ex3
CGCGATGCCTGGGGCCGTCTGCCCGTGGACCTGGCTGAGGAGCTGG
GCCATCGCGATGTCGCACGGTACCTGCGCGCCCGTGCGGGGGGCAC
Ex15
CAGAGGCAGTAACCATGCCCGCATAGATGCCGCGGAAGGTCCCTCA
Ex8
Ex4
GACATCCCCGATTGAAAGAACCAGAGAGGCTCTGAGAAACCTCGGG
Ex5
AAACTTAGATCATCAGTCACCGAAGGTCCTACAGGGCCACAACTGC
CCCCGCCACAACCCACCCCGCTTTCGTAGTTTTCATTTAGAAAATA
GAGCTTTTAAAAATGTCCTGCCTTTTAACGTAGATATAAGCCTTCC
CCCACTACCGTAAATGTCCATTTATATCATTTTTTATATATTCTTA
TAAAAATGTAAAAAAGAAAACACCGCTTCTGCCTTTTCACTGTGTT
Exon 2

Figure 1A

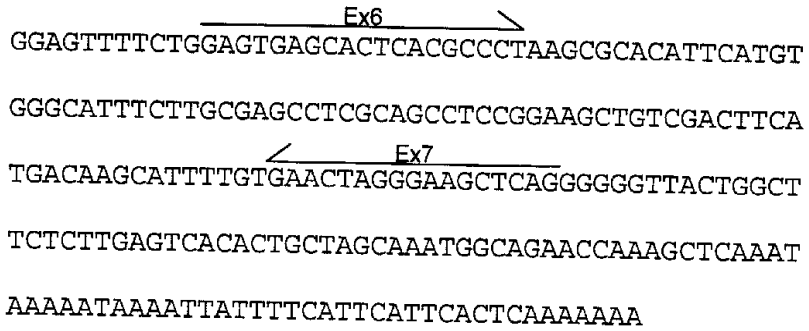
Ex6
GGAGTTTTCTGGAGTGAGCACTCACGCCCTAAGCGCACATTCATGT
GGGCATTTCTTGCGAGCCTCGCAGCCTCCGGAAGCTGTCGACTTCA
Ex7
TGACAAGCATTTTGTGAACTAGGGAAGCTCAGGGGGGTTACTGGCT
TCTCTTGAGTCACACTGCTAGCAAATGGCAGAACCAAAGCTCAAAT
AAAAATAAAATTATTTTCATTCATTCACTCAAAAAAA

Figure 1B

```
p16ex1    < GGNGGNAAGNTGTGGGGGAAAGTTTGGGGATGGAANACCAANCCCTCCTTTCNTTACCAA

          .........+.........+.........+.........+.........+.........+

p16ex1    < ACNCTGGCTCTGNCGAGGCTNCNTCCGANTGGTNCCCCCGGGGGAGACCCAACCTGGGNC
p16ex1    < GACTTCAGGGNTGCNACATTCACTAAGTGCTNGGAGNTAATANCACCTCCTCCGAGCANx
p16ex13                          TCNCTTATTGNTAGGANATAATAACACCTCCACCGATAACT

          .........+.........+.........+.........+.........+.........+

p16ex1    < TCGCTCACAGCGTCCCCTTACCTNGANAGATACCNCGXGXTCCCTCCAGAGGATTTGAGG
p16ex13   < TCACTTACAACGTCCCNNTTCCTGGAAAGATACACAGCGTTCCCTCCAGAGGATTTGTGG

          .........+.........+.........+.........+.........+.........+

p16ex1    < GACAGGNTCGGAGGGGGCTCTTCCCCCANCACCGGAGGAAGAAAGAGGAGGGNCTGACTG
p16ex13   < GACAGGGTNGGAGNGGTCTCTTCCNCCACCACCGGAGGAAGAAAGAGGAGGGGCTGNCTG

          .........+.........+.........+.........+.........+.........+

                                                          -----Ex1A--(12)-->
p16ex1    < GTCACCAGAGGGTGGGACGGACCGCGTGCGCTCGGCGNCTNCGGAGAGGGGGAGAACAGA
p16ex13   < TTCACCAGAGGGTGGGACGGACCNCGTACGCTCGNCGNCTNCGGAGAGGGGGAGAGCAGT

          .........+.........+.........+.........+.........+.........+

p16ex1    < CAACGGGCGGCGGGGAGCAGCATGGATCCGGCGGCGGGGAGCAGCATGGANCCTTCGACT
p16ex13   < CANCGGNCGNCGGGGAGCAACATGGAACCGNCGGCGGGGAGCAGCATGGANCCTTCGGCT

          .........+.........+.........+.........+.........+.........+

p16NT2    <             GACNNNCTCCGGCCGGNGTCGGGTAGAGGAGGTGCGGGCGCTGCTGGAG
p16ex1    < GACTGACTGCCTCGC
p16ex13   < GACTGGCTGNCCACGNCCACGNCCCGGGGTCGGGTAGAGGAGGTGCGGNCGCTNCTGGAG

          .........+.........+.........+.........+.........+.........+

                                    <---------Ex13----------
p16nt3    >                                                 CTCTNANCCCGGGTA
p16nt2    < GCGGGGGCGCTGCCCAACGCACCGAATAGTTACGGTCGGAGGCCGATCCAGGTXXGGGTA
p16ex13   < GCGGGGNCTCTGNCCAACNCGCTAAAAN

          .........+.........+.........+.........+.........+.........+

p16nt3    > GAGGGTCTGCAGCGGGAGCAGNGGATGGCGGGCGACTCTGGAGGACGAAGTTGGCAGGGG
p16nt2    < GAGGGTCTGCAGCGGGAGCAGGGGATGGCGGGCGACTCTGGAGGACGAAGTTTGCAGGGG

          .........+.........+.........+.........+.........+.........+

                                    <----------Ex1B------
p16nt3    > AATTGGAATCAGGTAGCGCTTCGANTCTCCGGAAAAAGGGGAGGCTTCCTGGGGAGTTNN
p16nt2    < AATTGGAATCAGGTAGCGCTTCGATTCTCCNGAAAAAGGGGAGGCTTCCTGGGGAGTTTT
```

Figure 2A

```
          .........+.........+.........+.........+.........+.........+
p16nt3   > CAGAAGGGGTTTGTAATCACAGNCCTCCNCCTGGCGACGCCCTGGGGGGTTGGGAAGCCA
p16nt2   < CAGAAGGGGTTTGTAATCACAGACCTCCTCCTGGCGACGTCCTGGGGGCTTGGGAAGCCA

          .........+.........+.........+.........+.........+.........+
p16nt3   > AGGAAGAGGAATGAGGAGNCACGCGCNTACAGNTCTCTCGAATNCTGANAAGATCTGAAG
p16nt2   < AGGAAGAGGAATNAGGAGCCACGCGCGTACAGATCTCTCGAATGCTGAGAAGATCTNAAG

          .........+.........+.........+.........+.........+.........+
p16nt3   > GGGGGAACATATTTGTATTAGxATNNAAGTATGCTCTTTATCAGATACAAAATTCACGAA
p16nt2   < GGGGGAACATATTTGTATTAGCNTCCAAGTNTNCTCTNTATCANATACAAANTxC

          .........+.........+.........+.........+.........+.........+
p16nt3   > CGTGTGGNATAAAAAGGGAGTCTTAAAGAAATNTAAGATGTGCTGGGACTACTTAGCCTC
p16nt3   > CAANACACAGATNCCTGGATGGAGCT
```

Figure 2B

```
p16int  > AAAANNAAAAAAAATCTCCCAGGCCTAACATAATTNTCAGGAAAGAAATTTCAGTAGTTG

          .........+.........+.........+.........+.........+.........+

p16int  > NATCTCAGGGGAAATACAGGAAGTTAGCCTGGAGTAAAAGTCAGTCTGTCCCTGCCCCTT

          .........+.........+.........+.........+.........+.........+

p16int  > TGCTANATTGCCCGTGCCTCACAGTGCTCTCTGCCTGTGACGACAGCTCCNCAGAAGTTC

          .........+.........+.........+.........+.........+.........+

p16int  > GGAGGATATAATGGAATTCATTGTGTACTGAAGAATGGATAGAGAACTCAAGAAGGAAAT

          .........+.........+... .....+.........+.........+.........+

p16int  > TGGAAACTGGAAGCAAATGTAGGG GTAATTAGACACCTGGGGCTTGTGTGGGGGTCTGCT
p16ex15 <                                    AANAAAAAAGAAATNGATAANATAGAGGAT

          .........+.........+.........+.........+.........+.........+

                          ----------EX2A-------->
p16int  > TGGCGGTGAGGGGGCTCTACACAAGCTTCCTTTCCGTCATGCCGNCCCCCACCCTGGCTC
p16ex15 < GAACANATTAAAATCAAAAAACANAACANAGACATAATAAAAAACGAGAATGTTCTAGAC

          .........+.........+.........+.........+.........+.........+

                                                          ------EX14-------
p16int  > TGACCATTCTGTTCTCTCTGGCAGGTCATGATGATGGGCAGCGCCCGAGTGGCGGAGCTG
p16ex15 < NTAATCATAATTATAAAGGTCAAGACTCATTGATATnAAGGAAATTGAAGGGAAATCTTA

          .........+.........+.........+.........+.........+.........+

          ->
p16int  > CTGCTGCTCCACGGCGCGGAGCCCAACTGCTCCGACGCCG
p16ex2  >                                CCTGCNACGACCCCGCCACTCTCACCCGACCCGTG
p16ex14 >     NCTCTCACGGTGGGGAGCCC ACTGCGCCGAACCCGCCACTCTCACCCGACCCGCG
p16ex15 < ACTAGCACAANNGNATNAAAAAAN ATTCCCACGACACCGCCACTCTCAACGCATCCGTG

          .........+.........+.........+.........+.........+.........+

p16ex2  > CACGACGCTGTCCGGGAGGGTTTCCTGGACACGCTGGTGGTGCTGCACCGGGCCGGGGNG
p16ex14 > CACGACGGTGCCCGGGAGGGGTTCCTGGACACGCTGGTGGTGCTGCACCGGGCCGGGGCG
p16ex15 < CTCGACACTGCCCGGGAGGTCNTCCTGGACACGCTGGTGGTNCTCCACCGGNCCGGGGCA

          .........+.........+.........+.........+.........+.........+

p16ex2  > CGGTTGGACGTGCGCGATGCCTGGGGCCGCCTNCCCGTGGXACCTGGTTGAGGAGCTGGG
p16ex14 > CGGCTGGACGTTCGNGATGCCTGGGGGCNTCTNTCCGTNGXACCTGGCTGAAGAGCTGGN
p16ex15 < CGTCTGGACGTGCGCGATGCCTGGGNCCGNCTACCCGTGGTACCTGACTGAGGACCTGGG

          .........+.........+.........+.........+.........+.........+

p16ex2  > NCATCGCGATGTCGCACGGTACCTGCGCGCGGTTGCGGGGGGCACCAGAGGXNAGTNACC
p16ex14 > NCATCGNGATGTCGCACGGCCNCTGTGTGNGGNTGCGGGGGGCACCATAGGTCAGTNTCC
p16ex15 < CCATCCCGATTTCGCNGGGTANCTGNGNGNGGCTGNGGGGGCCAANAGAGGXCANTACCC
```

Figure 2C

```
P16EX5  < XAAGTATGAGCGAAACNAATTGTGGTTTGAGAANAGGNAATCGTAGGGAACTTCGGGATC

          .........+.........+.........+.........+.........+.........+

P16EX5  < CCNCNGGGANCNCCAGAACCTGAGNCGCCNATTGGAAATNACAAACTGNCTGNATCACTC

          .........+.........+.........+.........+.........+.........+

P16EX5  < CGNACCAGGTNCAAAAGATACCTGGGGANGCGGGAAGGGAAAGACNACATCNAGACCGCC
P16EX9  <                                                          CCCC

          .........+.........+.........+.........+.........+.........+

P16EX5  < TTCGCNCCTXGGNATTGTGAGCAGCCTCTGAGACTCATTXATATNACACTCGTXTTTCTT
P16EX9  < ATCGCGCCTTGGGANTGTGAGCNACCATTGAGACTCATNAATATAGCACTCGTTTTTCTT

          .........+.........+.........+.........+.........+.........+

P16EX5  < CTTACAACCCTGCGGNCCGCGCGGTCGCGCTTTCTCTGCCCTCCGCCGGGTGGACCTGGA
P16EX9  < CTTGCAACCCTGCGGNCCGCGCGGTCGCGCTNTCTCTGCCCTCCGCNGGGTGGACCTGGA

          .........+.........+.........+.........+.........+.........+

P16EX5  < GCGCTTGAGCGGTCGGCGCGCCTGGAGCAGCCAGGCGGNCAGTGGACTAGCTGCTGGACC
P16EX9  < GCGCTTGAGCGGTCGGCGCNCCTGGANCAGCCAGGCGGGCAGTGGACTACCTNCTGGACC

          .........+.........+.........+.........+.........+.........+

P16EX5  < AGGGAGGTGTGGGAGAGCGGTGGCGGCGGGTACATGCACGTGAAGCCATTGCGAGAACTT
P16EX9  < AGGGAGGTGTGGGAGAGCGGTGNCGGCGGGTACATGCACGTGAAGCCATTGCGAGAACTT

          .........+.........+.........+.........+.........+.........+

P16EX5  < TATCCATAAGTATTTCAATACCGGTAGGGACGGCAAGAGAGGAGGGCGGGATGTGCCACA
P16EX5  < TATCCATAAGTATTTCAATGCCGGTAGGGACGGCAAGAGAGGAGGGCGGGATGTNCCACA

          .........+.........+.........+.........+.........+.........+

P16EX5  < CATCTTTGACCTCAGGTTTCTAACGCCTGTTTTCTTTCTGCCCTCTGCAGACAACCCCGA
P16EX9  < CATCTTTGACCTCAGGTTTCTAACGCCTGTTTTCTTTCTGCCCTCTGCAGACATCCCCGA

          .........+.........+.........+.........+.........+.........+

P16EX4  >                                          AGAAATTAGATCATCAGTCACCGATG
P16EX5  < TTGAAAGAACCAGAGAGGCTCTGAGAAACC
P16EX9  < TTGAAAGAACCAGAGAGGCTCTGAGAAACCTCCGGAAACTTAGXTCATCAXTCGCCGNAA

          .........+.........+.........+.........+.........+.........+

P16EX4  > GTCCTACAGGGNCACAACTGNCCCCGCCACAACCCACCCCGNTTTCGTAGTTTTCATTTA
P16EX9  < AA
```

Figure 3A

```
P16EX4  > GAAAATAGAGCTTTTAAAAATGTCCTGCCTTTTAACGTAGATATATGCCTTCCCCCACTA

          .........+.........+.........+.........+.........+.........+

P16EX4  > CCGNAAATGTCCATTTATATCATNTTTTATATATTCTTATAAAAATGTAAAAAAGAAAAA

          .........+.........+.........+.........+.........+.........+

P16EX4  > CACCGCTTCTGCCTTTTCACTGTGTTGGAGTTTTCTGGAGTGAGCACTCACGCCCTAAGC

          .........+.........+.........+.........+.........+.........+

P16EX6  >   CANCNATNTNCGGCATTTCTNGNGAGCCTCGTAGTCTCCGGATGNTGTCGACCTCGAG
P16EX6A >   CANCNATNTNCGGCATTTCTNGNGAGCCTCGTAGTCTCCGGATGNTGTCGACCTCGAG
P16EX4  > GCACATTCATGTGGGCATTTCTTGCGAGCCTCGCAGNCTCCGGAAGCTGTCGACCTCGAG

          .........+.........+.........+.........+.........+.........+

P16EX6  > GGGGGGNCCNGTACCCAATTCGNCCTATNGTGAGTCGTNTTACAATTCACTGGCCGCCGT
P16EX6A > GGGGGGNCCNGTACCCAATTCGNCCTATNGTGAGTCGTNTTACAATTCACTGGCCGCCGT
P16EX4  > GGGGGGNCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGNCGNCGN

          .........+.........+.........+.........+.........+.........+

P16EX6  > TTTXACAACGTCGXTGXACTGGGAAAACCCTGGTGTTACCCAACTTXAATCGCCTTGNAG
P16EX6A > TTTXACAACGTCGXTGXACTGGGAAAACCCTGGTGTTACCCAACTTXAATCGCCTTGNAG
P16EX4  > TTTTACAACGTCGGTGGACTGGGAAAACCCCGGNGTTACCCAACTTTAATCGNCTTGGAG

          .........+.........+.........+.........+.........+.........+

P16EX6  > NACATCCCCCTTTXCGCCAGCTGGTGTAATAGCGANGAGGCCCGCACCGATCGCCCTTCC
P16EX6A > NACATCCCCCTTTXCGCCAGCTGGTGTAATAGCGANGAGGCCCGCACCGATCGCCCTTCC
P16EX4  > GACATCCCCCTTTTCGCCAGNTGGGGTTATAGNGAAGAGGGCCNCACCNNTCGCCC

          .........+.........+.........+.........+.........+.........+

P16EX6  > CAACAGTTGNGCAGCCTGAATGGCGAATGGAAATTGTAAGCGTTAATATTTTGTTAAAAT
P16EX6A > CAACAGTTGNGCAGCCTGAATGGCGAATGGAAATTGTAAGCGTTAATATTTTGTTAAAAT

          .........+.........+.........+.........+.........+.........+

P16EX6  > TCGCGTTANATCNTCGGTTAANTCAGCTCATNTTTTATCCAATAGGCCGANATCGGCANA
P16EX6A > TCGCGTTANATCNTCGGTTAANTCAGCTCATNTTTTATCCAATAGGCCGANATCGGCANA

          .........+.........+.........+.........+.........+.........+

P16EX6  > ATCCCCAATAAATCAANAGAATAGACCGAGATAGGGTTGAGTGTCGTTCCAGTTNGGGAA
P16EX6A > ATCCCCAATAAATCAANAGAATAGACCGAGATAGGGTTGAGTGTCGTTCCAGTTNGGGAA

          .........+.........+.........+.........+.........+.........+

P16EX6  > CANGAGTCCACTATTAAAGANCGTAGNCTCNAACGTCANAGGGCGAAAAACCNTNTTTCA
P16EX6A > CANGAGTCCACTATTAAAGANCGTAGNCTCNAACGTCANAGGGCGAAAAACCNTNTTTCA
```

Figure 3B

```
          .........+.........+.........+.........+.........+.........+
P16EX6   > GNGGATTGGNCCACTACGCNTANCC
P16EX6A  > GNGGATTGGNCCACTACGCNTANCCATCACCCTATTC
```

Figure 3C

Figure 4

H9	U18	CCL119	MCF-7	HTB 125	SaOs2	A431	normal #1	normal #2	cell
absent	absent	absent	absent	absent	altered	altered	norm	norm	exon 1
absent	absent	absent	absent	absent	altered	absent	norm	norm	exon 2

WI38	CCL120	HeLa	HTB100	ZRB75	GM130	Tera2	HTB172	HTB173	cell
norm	norm	norm	norm	norm	norm	norm	norm	norm	exon 1
norm	norm	norm	norm	norm	norm	norm	norm	norm	exon 2

Figure 5

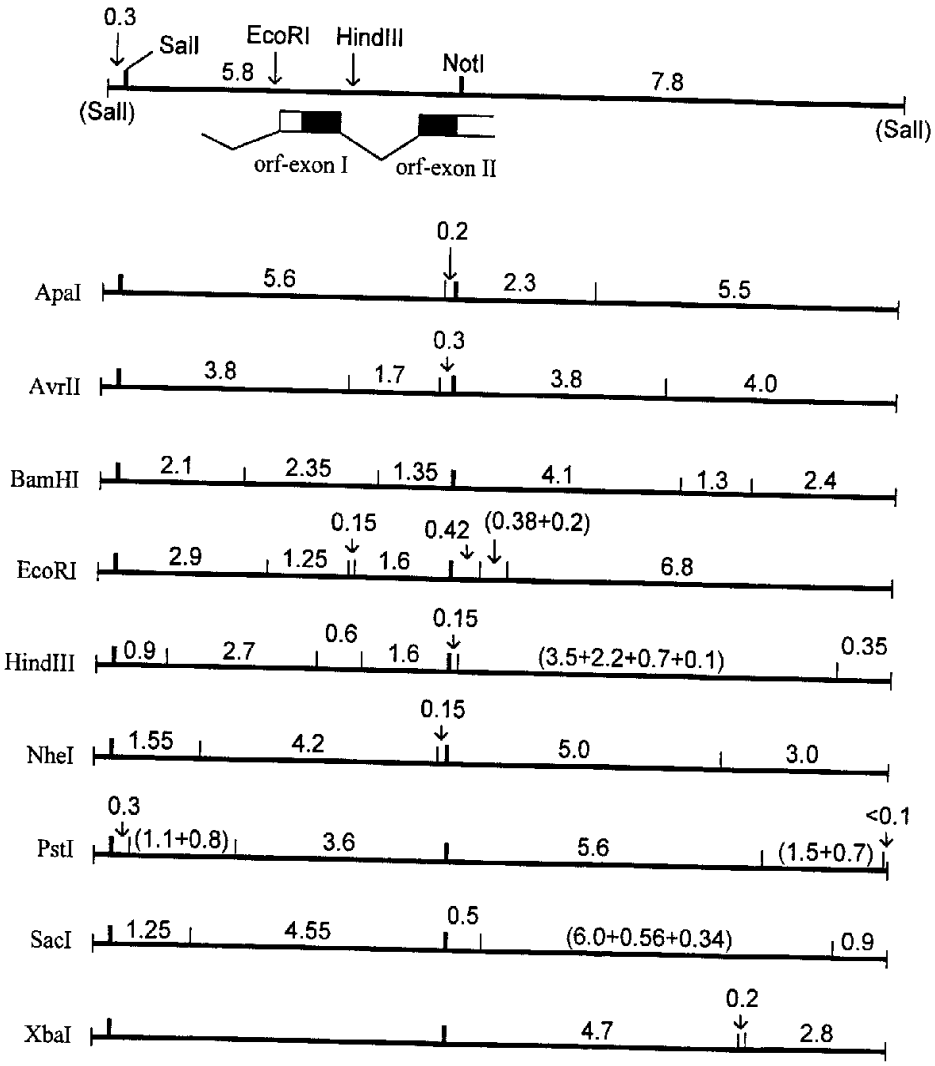
0.3
SalI
EcoRI
HindIII
5.8
NotI
7.8
(SalI)
(SalI)
orf-exon I
orf-exon II
ApaI
5.6
0.2
2.3
5.5
AvrII
3.8
1.7
0.3
3.8
4.0
BamHI
2.1
2.35
1.35
4.1
1.3
2.4
EcoRI
2.9
1.25
0.15
1.6
0.42
(0.38+0.2)
6.8
HindIII
0.9
2.7
0.6
1.6
0.15
(3.5+2.2+0.7+0.1)
0.35
NheI
1.55
4.2
0.15
5.0
3.0
PstI
0.3
(1.1+0.8)
3.6
5.6
(1.5+0.7)
<0.1
SacI
1.25
4.55
0.5
(6.0+0.56+0.34)
0.9
XbaI
4.7
0.2
2.8

Figure 6

```
p16: Ile Gln Val Met Met Met Gly Ser Ala Arg Val Ala Glu Leu Leu Leu Leu His Gly Ala Glu Pro Asn
p15: Ile Gln Val Met Met Met Gly Ser Ala Arg Val Ala Glu Leu Leu Leu Leu His Gly Ala Glu Pro Asn
p13:             Met Met Met Gly Asn Val His Val Ala Ala Leu Leu Leu Asn Tyr Gly Ala Asp Ser Asn

     Cys Ala Asp Pro Ala Thr Leu  *  Thr Arg Pro Val His Asp Ala Ala Arg Glu Gly Phe Leu Asp Thr
     Cys Ala Asp Pro Ala Thr Leu  *  Thr Arg Pro Val His Asp Ala Ala Arg Glu Gly Phe Leu Asp Thr
     Cys Glu Asp Pro  *  Thr Thr Phe Ser Arg Pro Val His Asp Ala Ala Arg Glu Gly Phe Leu Asp Thr

     Leu Val Val Leu His Arg Ala Gly Ala Arg Leu Asp Val Arg Asp Ala Trp Gly Arg Leu Pro Val Asp
     Leu Val Val Leu His Arg Ala Gly Ala Arg Leu Asp Val Arg Asp Ala Trp Gly Arg Leu Pro Val Asp
     Leu Val Val Leu His Gly Ser Gly Ala Arg Leu Asp Val Arg Asp Ala Trp Gly Arg Leu Pro Leu Asp

     Leu Ala Glu Glu Leu Gly His Arg Asp Val Ala Arg Tyr Leu Arg Arg Ala Ala Ala Gly Gly Thr ....
     Leu Ala Glu Glu Arg Gly His Arg Asp Val Ala Gly Tyr Leu Arg Thr Ala Thr Gly Asp
     Leu Ala Gln Glu Arg Gly His Gln Asp Ile Val Arg Tyr Leu Arg Ser Ala  *  Gly Cys Ser ....
```