(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 December 2005 (08.12.2005)

PCT

(10) International Publication Number
WO 2005/115458 A2

(51) International Patent Classification[7]: A61K 39/42

(21) International Application Number:
PCT/US2005/011740

(22) International Filing Date: 7 April 2005 (07.04.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/560,471 7 April 2004 (07.04.2004) US

(71) Applicant *(for all designated States except US)*: **UAB RESEARCH FOUNDATION** [US/US]; 701 South 20th Street, Suite 1120G, Birmingham, AL 35294 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHOW, Louise, T.** [US/US]; 3012 Westmoreland Drive, Mountain Brook, AL 3522-2728 (US). **BROKER, Thomas, R.** [US/US]; 3012 Westmoreland Drive, Mountain Brook, AL 3522-2728 (US). **DENG, Wentao** [US/US]; 417-A Honey Locust Lane, Birmingham, AL 35209 (US). **LIN, Biing, Yuan** [US/US]; 5577 Surrey Lane, Birmingham, AL 35242 (US). **YU, Jei-Hwa** [US/US]; 433 Cambrian Ridge Trail, Pelham, AL 35124 (US).

(74) Agents: **CLEVELAND, Janell, T.** et al.; Needle & Rosenberg, P.C., Suite 1000, 999 Peachtree Street, Atlanta, GA 30309-3915 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHODS AND COMPOSITIONS RELATED TO REGULATION OF NUCLEO-CYTOPLASMIC LOCALIZATION OF E1 DNA HELICASE IN PAPILLOMAVIRUSES

(57) **Abstract:** This invention relates generally to methods and compositions for regulating nucleocytoplasmic localization of E1 DNA helicase.

WO 2005/115458 A2

METHODS AND COMPOSITIONS RELATED TO REGULATION OF NUCLEO-CYTOPLASMIC LOCALIZATION OF E1 DNA HELICASE IN PAPILLOMAVIRUSES

BACKGROUND OF THE INVENTION

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of priority of U.S. Provisional Application 60/560,471, filed April 7, 2004, which application is incorporated herein by this reference in its entirety.

This invention was funded by the National Institutes of Health, grant numbers CA83679, CA107338-02, CA36200-17, and CA36200. Therefore, the United States Government has certain rights in this invention.

FIELD OF THE INVENTION

This invention relates generally to methods and compositions for regulating nucleocytoplasmic localization of papillomavirus E1 DNA helicase. The invention has broad applicability in protection against viral diseases or processes involving modulating the nuclear localization of E1 DNA helicase.

GENERAL BACKGROUND

Papillomaviruses are non-enveloped DNA viruses that induce hyperproliferative lesions of the epithelia. The papillomaviruses are widespread in nature and have been identified in higher vertebrates. Viruses have been characterized, amongst others, from humans, cattle, rabbits, sheep, horses, and dogs. Most animal papillomaviruses are associated with purely epithelial proliferative lesions, of both the cutaneous and mucosal epithelia and most lesions in animals are cutaneous. In the human, more than 100 types of papillomavirus have been identified, and they have been catalogued by site of infection: cutaneous epithelium and mucosal epithelium (oral-respiratory and anogenital mucosa) (de Villiers et al. (2004) Virology 324: 17-27.) The cutaneous-related diseases include common warts, flat warts, and plantar warts. The mucosal-related diseases include laryngeal papillomas and anogenital condylomas. Some infection may progress to cervical, penile and anal carcinomas (Fields, 1996, Virology, 3rd ed. Lippincott--Raven Pub., Philadelphia, N.Y.).

PCT/US2005/011740

Current methods of treatment for genital warts include physical removal such as cryotherapy, CO2 and other laser, electrosurgery, and surgical excision. Cytotoxic agents such as trichloroacetic acid (TCA), podophyllin or podofilox are also used. Immunotherapy is also available such as Interferon or Imiquimod. These treatments are not completely effective in eliminating all infected cells, and there is either a high cost incurred or uncomfortable side effects related thereto. In fact, there are currently no effective antiviral treatments for papillomavirus infection since recurrence of lesions is are common with all current therapies (Beutner & Ferenczy, 1997, Amer. J. Med., 102(5A), 28-37). What is needed in the art are compounds capable of interfering with viral DNA replication.

SUMMARY OF THE INVENTION

In accordance with the purposes of this invention, as embodied and broadly described herein, this invention, in one aspect, relates to methods of inhibiting papillomavirus replication in a subject.

In another aspect, the invention relates to methods of screening for an agent that inhibits nuclear localization of E1.

In yet another aspect, the invention relates to nucleic acids, polypeptides, vectors, cells, and cell lines relating to E1.

In yet another aspect, the invention relates to a composition used in the methods described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments of the invention and together with the description, serve to explain the principles of the invention.

Figure 1 shows *in vivo* phosphorylation and subcellular localization of GFP-E1 and mutations. Figure 1A shows functional domains and critical sequences regulating HPV-11 E1 nucleo-cytoplasmic localization. LRR, Localization Regulatory Region defined in this work. DBD, DNA binding domain and helicase domain were defined previously. Stars, candidate CDK/MAP kinase phosphorylation sites (serine or threonine followed by a proline residue). Cross hatch, a leucine-rich nuclear export sequence (NES). Underlined, the consensus cyclin binding motif (RRL). Filled boxes, bipartite nuclear localization sequence

(NLS). Figure 1B shows *in vivo* phosphorylation of EE-E1 (an E1 protein tagged with an Glu-rich epitope at the amino terminus). EE-E1 protein purified from different sources with (+) or without (-) prior treatment with protein phosphatase 1 (PP1) was blotted with MPM-2 antibody which recognizes phospho-S/T-P. The membrane was stripped and reprobed with an antibody which recognizes the EE epitope tag of the E1 protein. Figure 1C shows subcellular localization of GFP-E1 and phosphorylation site mutations in transfected COS7 cells (upper panels). Lower panels show the DAPI stained nuclei.

Figure 2 shows *in vivo* phosphorylation and subcellular localization of GFP-E1 and mutations in the presence of inhibitors of CDKs or CRM1. All images were taken from transfected COS7 cells except in Figure 2C. Figure 2C shows that GFP-E1 phosphorylation is blocked by the mutation in the cyclin binding motif (RRL to RRA, KRA or ARA, SEQ ID NOS: 1 and 11-13). GFP-E1 wild type or mutated proteins were immunoprecipitated using polyclonal GFP antibody blotted with MPM-2, and reprobed with monoclonal anti-GFP after stripping. Figure 2B shows GFP1-E1 was localized to the cytoplasm (left panel) after CDK activity was inhibited by roscovatine (right panels). Immunoprecipitation and Western blotting were performed as described above. Figure 2C shows GFP-E1 was localized to the cytoplasm after CDK2 activity was inhibited upon p21cip1 induction by IPTG in the cell line p21-9 (which was stably transduced with an inducible p21cip1 gene). Figure 2D shows that inhibition of CRM1 successfully restores nuclear localization of GFP-E1 phosphorylation mutations. In this experiment, 20 ng/ml leptomycin B (LMB), a specific inhibitor for CRM1-mediated protein nuclear export, were added to culture medium for 6 hours.

Figure 3 shows identification of the leucine-rich NES necessary and sufficient for E1 export. All images shown were taken from transfected COS7 cells. Figure 3A shows that an E1 Localization Regulatory Region (LRR), with the key motifs conserved among human papillomaviruses, contains the CDK/MAP kinase phosphorylation sites, the cyclin binding motif (RRL), a putative bipartite NLS (dashed underlines), and a leucine-rich NES. Underlined, the 10-aa (residue 106-115, SEQ ID NO: 14) core of the leucine-rich NES. The minimal functional nuclear export signal sequences were determined in this study. The consensus NES sequence is shown at the top. NEm indicates the mutations in NES. Mutations on the cyclin binding motif are illustrated on the right. Alignment with other HPVs is listed at the bottom (and is expanded in the attached sequence alignment tables).

Figure 3B shows subcellular localization of GFP-E1 NES mutations (NEm). Figure 3C shows subcellular localization of EE-E1 and mutations as revealed by indirect immunofluorescence (FITC) with anti-EE antibody. Figure 3 D shows delineation of the functional leucine-rich NES. For purposes of testing subcellular localization, E1 peptides were fused downstream of GFP. The positions of the appended E1 peptides are indicated in parentheses.

Figure 4 shows activities of GFP-E1 mutations in transient replication assays. Transient replication assay was performed as described in Example 1. 293 cells were co-transfected with 0.5 μg of E1 plasmid, 5 μg of pMT2-E2 plasmid and 0.5 μg of a plasmid containing the HPV origin of replication. Low molecular weight plasmids were isolated 48 hours post-transfection, digested with restriction enzyme(s) and detected by Southern blotting using HPV origin DNA plasmid as probe. Linearized ori (arrow) in restriction enzyme DpnI + lanes indicates the newly replicated origin DNA. Figure 4A shows ori replication by the wild type GFP-E1 and by phosphorylation site mutations, and by the corresponding NES mutations (NEm). Figure 4B shows signals of replicated origin DNA quantified and normalized to the wild type E1 to obtain the relative replication activities.

Figure 5 shows subcellular localization of GFP-E1 (Fig. 5A) and GFP-E1 S89,93,107A (Fig. 5B) in the absence and presence of the origin recognition protein E2. COS7 cells were cotransfected with 5 μg each of expression plasmids of GFP-E1 and native E2. Cells were fixed 24 hrs post-transfection. Immunostaining of E2 was performed using anti-E2 polyclonal antibody and anti-rabbit IgG-Texas Red conjugated secondary antibody. Top and bottom panels are typical patterns exhibited by the majority and minority of cotransfected cells, respectively.

Figure 6 shows that phosphorylation determines E1 localization by regulating its nuclear export signal.

Figure 7 shows that the E1 protein forms first a hexamer, then a dihexamer, on the origin.

Figure 8 shows immunogold-labeling of heat shock/chaperone proteins in E1: DNA complexes. On the left is Hsp70, and on the right is Hsp40.

Figure 9 shows helicase unwinding the origin DNA. Unwinding intermediates with double loops coated with single-stranded DNA binding protein are shown. DNA is pumped inward while being unwound to enable bi-directional replication.

PCT/US05/11740

Figure 10 diagrammatically shows assembly and activation of papillomavirus pre-replication complexes.

Figure 11 diagrammatically shows HPV-11 E1 and E2 genes, proteins, and regulatory sequences.

Figure 12 shows a eukaryotic cell cycle, and the role of cyclin expression and replication.

Figure 13 shows E1 localization regulatory sequences.

Figure 14 shows a putative MAPK docking domain alignment of papillomavirus E1 helicases.

Figure 15 shows the sequence of HPV-18 and E1/E2 interactions.

Figure 16 shows the relationship between the MAP kinases.

Figure 17 shows that MAP kinases phosphorylate HPV-11 E1 *in vitro*.

Figure 18 shows the mutations of the putative E1 MAPK docking domains.

Figure 19 shows that multiple mutations in S89, S93, or S107 abolish E1 nuclear localization.

Figure 20 shows that phosphorylation determines E1 localization by regulating its NES.

Figure 21 shows phosphorylation by MAPK and cyclin/CDK regulates HPV E1 protein nucleo-cytoplasmic localization and retention and therefore controls viral DNA replication.

Figure 22 shows MAPK phosphorylation on E1 can also affect E1-mediated viral DNA replication.

Figure 23 shows tetracycline dosage curves in GFP-11E1dm cells compared to the levels of the house-keeping protein actin as a sample loading standard. The E1 helicase is maximally induced by adding as little as 62.5 nanograms/ml of tetracycline to the cell culture media. Lower concentrations may also be used.

Figure 24 shows tetracycline dosage curves in GFP-11E1dm cells compared to actin levels as a function of time. The induction can be sustained for at least 22 hours of continual tet-treatment.

Figure 25 shows GFP-11E1dm wild type Tet-on 293 cells, where the gene has an inactivating mutation of the RNA splice donor site, thereby increasing the amount of E1 messenger RNA.

Figure 26 shows that HPV-11 E1 MAPK docking domain mutations cannot support HPV DNA replication.

Figure 27 shows that E1 is phosphorylated at one or more candidate cyclin/CDK sites *in vivo*.

Figure 28 shows that single as well as combined mutations in S89, S93, and S107 significantly reduce E1 nuclear localization.

Figure 29 shows that the inability to bind cyclins and be phosphorylated by CDKs leads to cytoplasmic E1.

Figure 30 shows that inhibition of cyclin/CDK1 or 2 with roscovitine leads to cytoplasmic E1.

Figure 31 shows that inhibition of cyclin E or A/CDK2 with p21cip1 leads to cytoplasmic E1.

Figure 32 shows that LMB, an inhibitor of CRM1-mediated nuclear export, largely restores nuclear localization of E1 phosphorylation mutations.

Figure 33 shows that a putative nuclear export signal is conserved among HPV E1 genotypes.

Figure 34 shows that leucine-rich nuclear export signal is necessary for efficient E1 export.

Figure 35 shows the regulation of nucleo-cytoplasmic of GFP-E1 is verified by testing key mutant forms of the E1 protein, each tagged with the EE epitope. Detection of EE-E1 is conducted by using antibody to the EE epitope.

Figure 36 shows HPV-11 E1 has a bipartite nuclear localization sequence.

Figure 37 shows Arg124, 125 in the cyclin binding motif (RxL) are part of the nuclear localization sequence.

Figure 38 shows single mutations in 3 CDK/MAP kinase phosphorylation sites significantly reduce E1 nuclear localization.

DETAILED DESCRIPTION OF THE INVENTION

The present invention may be understood more readily by reference to the following detailed description of preferred embodiments of the invention and the Examples included therein and to the Figures and their previous and following description.

Definitions

As used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a small molecule" includes mixtures of one or more small molecules, and the like.

Ranges may be expressed herein as from "about" one particular value, and/or to "about" another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint.

The terms "higher," "increases," "elevates," or "elevation" refer to increases above control levels. The terms "low," "lower," "inhibits," "inhibition," "reduces," or "reduction" refer to decreases below control levels.

The term "mediate" or "mediation" and "modulate" or "modulation" means to regulate, or control, in particular to increase, enhance, elevate, or alternatively to lower, inhibit, or reduce. The terms "mediate" and "modulate" are used interchangeably throughout.

As used throughout, by a "subject" is meant an individual. Thus, the "subject" can include domesticated animals, such as cats, dogs, etc., livestock (e.g., cattle, horses, pigs, sheep, goats, etc.), laboratory animals, and birds. Preferably, the subject is a mammal such as a primate, and, more preferably, a human.

The terms "control levels" or "control cells" are defined as the standard by which a change is measured, for example, the controls are not subjected to the experiment, but are instead subjected to a defined set of parameters, or the controls are based on pre- or post-treatment levels.

General Description

Human papillomavirus (HPV) is a family of small DNA virus specific for mucosal or cutaneous epithelia. It causes hyperproliferation ranging from benign warts to malignant cervical cancers (zur Hausen, H., and E.-M. de Villiers *Annu. Rev. Microbiol.* 48:427-447 (1994)). Infections can be latent but the viral DNA persists in the long-living basal cells and

can be activated during episodes of temporary or long-term immunosuppression. In the basal and parabasal cells that maintain the ability to divide, viral DNA exists as extrachromosmal plasmids in low copy numbers. Viral DNA amplification takes place only in a subset of differentiated cells in the upper strata. For initiation of viral DNA replication, HPV encodes its own replicative helicase E1 (Hughes, F. J., and M. A. Romanos *Nucleic Acids Res.* 21:5817-5823 (1993), Lin et al. *Mol. Cell. Biol.* 22:6592-6604 (2002), Seo et al. *Proc. Natl. Acad. Sci. USA* 90:702-706 (1993b), Yang et al. *Proc. Natl. Acad. Sci. USA* 90:5086-5090 (1993)) and origin-recognition protein E2 (Chiang et al. *J. Virol.* 66:5224-5231 (1992a), Mohr et al. *Science* 250:1694-1699 (1990), Seo et al. *Proc. Natl. Acad. Sci. USA* 90:2865-2869 (1993)). All other proteins necessary for replication come from host cells, including DNA polymerase α / primase, DNA polymerase δ, RPA, PCNA, and topoisomerases (Kuo et al. *J. Biol. Chem.* 269:24058-24065 (1994), Yang et al. *Nature* 353:628-632 (1991)). The virus expresses two oncoproteins, E6 and E7, the expression of which is elevated upon squamous differentiation. Their function is to reestablish a S phase environment capable of supporting viral DNA amplification (Chow, L. T., and T. R. Broker, p. 267-301. *In* N. Nathanson (ed.), Viral Pathogenesis. Lippincott-Raven Publishers, Philadelphia (1997)).

For replication, the viral E2 protein of 40 kDa recruits the E1 protein of 70 kDa to the origin of replication (ori), on which the E1 protein assembles into a dihexameric, bidirectional DNA helicase (Fouts et al. *J. Biol. Chem.* 274:4447-4458 (1999), Lin et al. *Mol. Cell. Biol.* 22:6592-6604 (2002), Liu et al. *J. Biol. Chem.* (1998), Sedman, J., and A. Stenlund. *J. Virol.* 72:6893-6897 (1998)). Assembly is facilitated by heat shock chaperone proteins. In the presence of RPA and topoisomerase I, ATP and an ATP regenerating system, the E1 protein then unwinds the double-stranded ori DNA (Lin et al. *Mol. Cell. Bio.* 22:6592-6604 (2002)) and recruits the host DNA polymerase α / primase to the replication fork to initiate replication. Thus, E2 is only required for the assembly of the pre-RC, whereas E1 is essential throughout initiation and elongation (Liu et al. 1995. J. Biol. Chem. 270:27283-27291). The HPV E1 protein also interacts with multiple cyclins and is phosphorylated *in vitro* by cyclin/CDK complexes.

Eukaryotic DNA replication is strictly controlled by mechanisms that regulate cell cycle to ensure both genetic inheritance and stability. DNA replication is initiated at a precise time when the cellular replication machinery is ready, and then the genome is

replicated for a single round in each cell cycle. Cell cycle entry and progression is regulated by cyclin/ CDKs that phosphorylate key regulatory proteins (Bell, S. P. et al. *Annu. Rev. Biochem.* 71:333-374 (2002); Pines, J. *Nat. Cell Biol.* 1:E73-79 (1999); Sherr, C. J. et al. *Genes Dev.* 13:1501-1512 (1999); Weis, K. *Cell* 112:441-451 (2003)). At G1/S phase transition, cyclin/CDK complexes phosphorylate the some of the components of the DNA pre-replication complex (Pre-RC), while others including ORC, Cdc6, Cdt1, and MCM2-7 could be phosphorylated by another kinase, leading to the activation of replication initiation and at the same time preventing re-initiation in the same cell cycle (Bell et al. (2002), Blow et al. *Trends Cell Biol.* 12:72-78 (2002), Nishitani et al. *Cells* 7:523-534 (2002)). E1 phosphorylation by cyclin/CDK is necessary for efficient viral DNA replication (Ma et al., 1999). E1 is phosphorylated by CDKs *in vivo,* and phosphorylation regulates its nucleo-cytoplasmic localization (Deng et al., 2004).

Precise subcellular localization of proteins is essential for their biological functions, and reciprocally, the control of protein localization provides the cells with a convenient way to regulate their functions. Post-translational modification of proteins plays critical roles in these aspects. For instance, three of the most important outcomes after protein phosphorylation are the changes in protein localization, stability, or activity (Laney et al. *Cell* 97:427-430 (1999)). Proteins that shuttle between the cytoplasm and nucleus often possess short peptide sequences that are recognized by transporting factors. Proteins with nuclear localization sequence (NLS) can be transported into the nucleus by a family of importin α/β heterodimers, while those with nuclear export sequence (NES) can form a complex with export receptors (exportins) and RanGTP to be transport out of the nucleus (Jans et al.. *Bioessays* 22:532-544 (2000), Komeili et al. *Annu. Rev. Genet.* 35:341-364 (2001), Nakielny et al. *Cell Biol.* 9:420-429 (1997), Ullman et al. *Cell* 90:967-970 (1997)). The activities of NLS and NES are often regulated by post-translational modifications, including phosphorylation, that can control the affinity for and accessibility by the protein transporters (Jans et al. (2000), Jans, D. A., and S. Hubner *Physiol. Rev.* 76:651-685 (1996)).

A conserved localization regulatory region (Example 7) was identified which contains a dominant leucine-rich nuclear export sequence (NES), the cyclin binding motif, three CDK substrate serine residues, and a putative bipartite nuclear localization sequence (NLS). E1 is exported from the nucleus by a CRM1-dependent mechanism unless the NES

is inactivated by CDK phosphorylation. It has been demonstrated that CDK phosphorylation controls E1 nuclear localization to support viral DNA amplification. Thus, papillomviruses adopt and adapt the cellular regulatory mechanism to regulate the E1 protein and to complete their reproductive cycle.

The HPV DNA helicase E1 contains a potent leucine-rich NES, which promotes E1 nuclear export in a CRM1-dependent manner. The E1 protein is shuttled out of the nucleus unless the NES is inactivated by CDK phosphorylation on three (3) substrate sites. Thus, one mechanism by which CDK controls the timing and extent of viral DNA replication is through the regulation on E1 subcellular localization. These results reveal that HPV has adapted the cellular regulatory mechanism to support its own DNA amplification.

Compositions

Disclosed herein are E1 polypeptides, fragments, mutants, and nucleic acids that can be used with the methods disclosed throughout the application. These and other materials are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these materials are disclosed that, while specific reference of each various individual and collective combinations and permutation of these compounds may not be explicitly disclosed, each is specifically contemplated and described herein. For example, if a particular compound is disclosed and discussed and a number of modifications that can be made to a number of molecules including the amino acids are discussed, specifically contemplated is each and every combination and permutation of the polypeptide and the modifications that are possible unless specifically indicated to the contrary. Thus, if a class of molecules A, B, and C are disclosed as well as a class of molecules D, E, and F and an example of a combination molecule, A-D is disclosed, then even if each is not individually recited each is individually and collectively contemplated meaning combinations, A-E, A-F, B-D, B-E, B-F, C-D, C-E, and C-F are considered disclosed. Likewise, any subset or combination of these is also disclosed. Thus, for example, the sub-group of A-E, B-F, and C-E would be considered disclosed. This concept applies to all aspects of this application including, but not limited to, steps in methods of making and using the disclosed compositions. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific embodiment or combination of embodiments of the disclosed methods.

The E1 protein is 70 kDa (SEQ ID NO: 1). It assembles at the origin of replication (ori) into a dihexameric, bidirectional DNA helicase. A conserved region of 44 amino acids (residues 83-126, SEQ ID NO: 2) at the HPV E1 amino-terminal portion, a localization regulatory region (LRR) (Fig. 3A), is described herein. LRR contains not only the NES (residues 96-116, SEQ ID NO: 5), it also spans the previously identified consensus cyclin binding motif RxL (residues 124-126, SEQ ID NO: 6) (Ma et al. (1999)), the three CDK/MAP kinase phosphorylation sites (S89, S93, S107), and a bipartite NLS (residues 83-85, KRK (SEQ ID NO: 7) and residues 120-125, KKVKRR, (SEQ ID NO: 8)). These sequence elements are highly conserved among E1 proteins of many and very likely almost all HPV genotypes (Fig. 3A), showing that a ubiquitous regulatory mechanism among E1 proteins of papillomaviruses can exist. The cell localization regulatory functions of this region are consistent with the report that the N-terminal 166 residues of HPV-11 E1 is dispensable for cell-free replication (Amin et al. 2000).

Also disclosed are mutants of E1 (SEQ ID NO: 1), such as SEQ ID NO: 9 (wherein leucine at position 110 is replaced with alanine); and SEQ ID NO: 10 (same as SEQ ID NO: 1, wherein the isoleucine at 113 is replaced with alanine). The serine at amino acid 89 can be mutated, the serine at amino acid 93 can be mutated, or the serine at amino acid 107 can be mutated. In one example, alanine is substituted for serine in the mutations discussed above. Also disclosed is a non-naturally occurring polypeptide comprising SEQ ID NO: 1, wherein the consensus cyclin binding motif RxL has been mutated to RRA (SEQ ID NO: 11), KRA (SEQ ID NO: 12), or ARA (SEQ ID NO: 13).

Homology/identity

It is understood that one way to define any known variants and derivatives or those that might arise from the disclosed genes and proteins herein is through defining the variants and derivatives in terms of homology to specific known sequences. For example SEQ ID NO: 1 sets forth a particular sequence of full-length E1, and SEQ ID NO: 2 sets forth a particular sequence of a functional fragment of E1. Specifically disclosed are variants of these and other genes and proteins herein disclosed which have at least, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 percent homology to the stated sequence. Those of skill in the art readily understand how to determine the homology of two proteins or nucleic acids, such as genes. For example, the

homology can be calculated after aligning the two sequences so that the homology is at its highest level.

Another way of calculating homology can be performed by published algorithms. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman Adv. Appl. Math. 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch, J. MoL Biol. 48: 443 (1970), by the search for similarity method of Pearson and Lipman, Proc. Natl. Acad. Sci. U.S.A. 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection.

The same types of homology can be obtained for nucleic acids by for example the algorithms disclosed in Zuker, M. *Science* 244:48-52, 1989, Jaeger et al. *Proc. Natl. Acad. Sci. USA* 86:7706-7710, 1989, Jaeger et al. *Methods Enzymol.* 183:281-306, 1989 which are herein incorporated by reference for at least material related to nucleic acid alignment.

Peptides

Protein variants

As discussed herein there are numerous variants of the E1 protein that are known and herein contemplated. In addition to the known functional E1 strain variants, there are derivatives of the E1 proteins which also function in the disclosed methods and compositions. Protein variants and derivatives are well understood to those of skill in the art and in can involve amino acid sequence modifications. For example, amino acid sequence modifications typically fall into one or more of three classes: substitutional, insertional or deletional variants. Insertions include amino and/or carboxyl terminal fusions as well as intrasequence insertions of single or multiple amino acid residues. Insertions ordinarily will be smaller insertions than those of amino or carboxyl terminal fusions, for example, on the order of one to four residues. Immunogenic fusion protein derivatives, such as those described in the examples, are made by fusing a polypeptide sufficiently large to confer immunogenicity to the target sequence by cross-linking *in vitro* or by recombinant DNA technology transformed with DNA encoding the fusion. Deletions are characterized by the removal of one or more amino acid residues from the protein sequence. These variants ordinarily are prepared by site specific mutagenesis of nucleotides in the DNA encoding the protein, thereby producing DNA encoding the variant, and thereafter

expressing the DNA in recombinant cell culture. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, for example M13 primer mutagenesis and PCR mutagenesis. Amino acid substitutions are typically of single residues, but can occur at a number of different locations at once. Deletions or insertions preferably are made in adjacent pairs, i.e. a deletion of 2 residues or insertion of 2 residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a final construct. The mutations must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure. Substitutional variants are those in which at least one residue has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Table 1 and are referred to as conservative substitutions.

TABLE 1:Amino Acid Substitutions
Original Residue Exemplary Conservative Substitutions, others are known in the art.
Ala/Ser
Arg/Lys/Gln
Asn/Gln/His
Asp/Glu
Cys/Ser
Gln/Asn/Lys
Glu/Asp
Gly/Pro
His/Asn/Gln
Ile/Leu/Val
Leu/Ile/Val
Lys/Arg/Gln
Met/Leu/Ile
Phe/Met/Leu/Tyr
Ser/Thr
Thr/Ser
Trp/Tyr
Tyr/Trp/Phe
Val/Ile/Leu

Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those in Table 1, i.e., selecting residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide

backbone in the area of the substitution, for example as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in the protein properties will be those in which (a) a hydrophilic residue, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; c) a residue having an electropositive side chain, e.g., lysine, arginine, or histidine, is substituted for (or by) an electronegative residue, e.g., glutamic acid or aspartic acid; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine, in this case, (e) by increasing the number of sites for sulfation and/or glycosylation.

For example, the replacement of one amino acid residue with another that is biologically and/or chemically similar is known to those skilled in the art as a conservative substitution. For example, a conservative substitution would be replacing one hydrophobic residue for another, or one polar residue for another. The substitutions include combinations such as, for example, Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. Such conservatively substituted variations of each explicitly disclosed sequence are included within the mosaic polypeptides provided herein.

It is understood that the description of conservative mutations and homology can be combined together in any combination, such as embodiments that have at least 70% homology to a particular sequence wherein the variants are conservative mutations.

As this specification discusses various proteins and protein sequences it is understood that the nucleic acids that can encode those protein sequences are also disclosed. This would include all degenerate sequences related to a specific protein sequence, i.e. all nucleic acids having a sequence that encodes one particular protein sequence as well as all nucleic acids, including degenerate nucleic acids, encoding the disclosed variants and derivatives of the protein sequences. Thus, while each particular nucleic acid sequence may not be written out herein, it is understood that each and every sequence is in fact disclosed and described herein through the disclosed protein sequence. For example, one of the many nucleic acid sequences that can encode the protein sequence set forth in SEQ ID NO: 1 is set forth in SEQ ID NO: 3. Another nucleic acid sequence that encodes the same protein sequence set forth in SEQ ID NO: 2 is set forth in SEQ ID NO: 4. It is also understood that

while no amino acid sequence indicates what particular DNA sequence encodes that protein within an organism, where particular variants of a disclosed protein are disclosed herein, the known nucleic acid sequence that encodes that protein in the particular sequence from which that protein arises is also known and herein disclosed and described.

It is understood that there are numerous amino acid and peptide analogs which can be incorporated into the disclosed compositions. For example, there are numerous D amino acids or amino acids which have a different functional substituent then the amino acids shown in Table 1. The opposite stereo isomers of naturally occurring peptides are disclosed, as well as the stereo isomers of peptide analogs. These amino acids can readily be incorporated into polypeptide chains by charging tRNA molecules with the amino acid of choice and engineering genetic constructs that utilize, for example, amber codons, to insert the analog amino acid into a peptide chain in a site specific way (Thorson et al., Methods in Molec. Biol. 77:43-73 (1991), Zoller, Current Opinion in Biotechnology, 3:348-354 (1992); Ibba, Biotechnology & Genetic Enginerring Reviews 13:197-216 (1995), Cahill et al., TIBS, 14(10):400-403 (1989); Benner, TIB Tech, 12:158-163 (1994); Ibba and Hennecke, Bio/technology, 12:678-682 (1994) all of which are herein incorporated by reference at least for material related to amino acid analogs).

Molecules can be produced that resemble peptides, but which are not connected via a natural peptide linkage. For example, linkages for amino acids or amino acid analogs can include CH_2NH--, --CH_2S--, --CH_2--CH_2 --, --CH=CH-- (cis and trans), --$COCH_2$ --, --$CH(OH)CH_2$--, and --CHH_2SO—(These and others can be found in Spatola, A. F. in Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983); Spatola, A. F., Vega Data (March 1983), Vol. 1, Issue 3, Peptide Backbone Modifications (general review); Morley, Trends Pharm Sci (1980) pp. 463-468; Hudson, D. et al., Int J Pept Prot Res 14:177-185 (1979) (--CH_2NH--, CH_2CH_2--); Spatola et al. Life Sci 38:1243-1249 (1986) (--CH H_2--S); Hann J. Chem. Soc Perkin Trans. I 307-314 (1982) (--CH--CH--, cis and trans); Almquist et al. J. Med. Chem. 23:1392-1398 (1980) (--$COCH_2$--); Jennings-White et al. Tetrahedron Lett 23:2533 (1982) (--$COCH_2$--); Szelke et al. European Appln, EP 45665 CA (1982): 97:39405 (1982) (--$CH(OH)CH_2$--); Holladay et al. Tetrahedron. Lett 24:4401-4404 (1983) (--$C(OH)CH_2$--); and Hruby Life Sci 31:189-199 (1982) (--CH_2--S--); each of which is incorporated herein by reference. A particularly preferred non-peptide linkage is --CH_2NH--. It is understood that

peptide analogs can have more than one atom between the bond atoms, such as b-alanine, g-aminobutyric acid, and the like.

Amino acid analogs and analogs and peptide analogs often have enhanced or desirable properties, such as, more economical production, greater chemical stability, enhanced pharmacological properties (half-life, absorption, potency, efficacy, etc.), altered specificity (e.g., a broad-spectrum of biological activities), reduced antigenicity, and others.

D-amino acids can be used to generate more stable peptides, because D amino acids are not recognized by peptidases and such. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (e.g., D-lysine in place of L-lysine) can be used to generate more stable peptides. Cysteine residues can be used to cyclize or attach two or more peptides together. This can be beneficial to constrain peptides into particular conformations. (Rizo and Gierasch Ann. Rev. Biochem. 61:387 (1992), incorporated herein by reference).

Nucleic acids

There are a variety of molecules disclosed herein that are nucleic acid based, including for example the nucleic acids that encode, for example, E1, as well as any other proteins disclosed herein, as well as various functional nucleic acids. The disclosed nucleic acids are made up of for example, nucleotides, nucleotide analogs, or nucleotide substitutes. Non-limiting examples of these and other molecules are discussed herein. It is understood that, for example, when a vector is expressed in a cell, the expressed mRNA will typically be made up of A, C, G, and U. Likewise, it is understood that if, for example, an antisense molecule is introduced into a cell or cell environment through for example exogenous delivery, it is advantageous that the antisense molecule be made up of nucleotide analogs that reduce the degradation of the antisense molecule in the cellular environment.

Nucleotides and related molecules

A nucleotide is a molecule that contains a base moiety, a sugar moiety and a phosphate moiety. Nucleotides can be linked together through their phosphate moieties and sugar moieties creating an internucleoside linkage. The base moiety of a nucleotide can be adenin-9-yl (A), cytosin-1-yl (C), guanin-9-yl (G), uracil-1-yl (U), and thymin-1-yl (T). The sugar moiety of a nucleotide is a ribose or a deoxyribose. The phosphate moiety of a nucleotide is pentavalent phosphate. A non-limiting example of a nucleotide would be 3'-AMP (3'-adenosine monophosphate) or 5'-GMP (5'-guanosine monophosphate).

A nucleotide analog is a nucleotide which contains some type of modification to either the base, sugar, or phosphate moieties. Modifications to nucleotides are well known in the art and would include for example, 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, and 2-aminoadenine as well as modifications at the sugar or phosphate moieties.

Nucleotide substitutes are molecules having similar functional properties to nucleotides, but which do not contain a phosphate moiety, such as peptide nucleic acid (PNA). Nucleotide substitutes are molecules that will recognize nucleic acids in a Watson-Crick or Hoogsteen base-pairing manner, but which are linked together through a moiety other than a phosphate moiety. Nucleotide substitutes are able to conform to a double helix type structure when interacting with the appropriate target nucleic acid.

It is also possible to link other types of molecules (conjugates) to nucleotides or nucleotide analogs to enhance for example, cellular uptake. Conjugates can be chemically linked to the nucleotide or nucleotide analogs. Such conjugates include but are not limited to lipid moieties such as a cholesterol moiety. (Letsinger et al., Proc. Natl. Acad. Sci. USA, 1989, 86: 6553-6556).

A Watson-Crick interaction is at least one hydrogen-bonding interaction with the Watson-Crick face of a nucleotide, nucleotide analog, or nucleotide substitute. The Watson-Crick face of a nucleotide, nucleotide analog, or nucleotide substitute includes the C2, N1, and C6 positions of a purine based nucleotide, nucleotide analog, or nucleotide substitute and the C2, N3, C4 positions of a pyrimidine based nucleotide, nucleotide analog, or nucleotide substitute.

A Hoogsteen interaction is the interaction that takes place on the Hoogsteen face of a nucleotide or nucleotide analog, which is exposed in the major groove of duplex DNA. The Hoogsteen face includes the N7 position and reactive groups (NH2 or O) at the C6 position of purine nucleotides.

Sequences

There are a variety of amino acid and nucleic acid sequences related to, for example, E1, as well as any other protein or nucleic acid disclosed herein that are disclosed on Genbank, and these sequences and others are herein incorporated by reference in their entireties as well as for individual subsequences contained therein.

A variety of sequences are provided herein and these and others can be found in Genbank, at www.pubmed.gov. Those of skill in the art understand how to resolve sequence discrepancies and differences and to adjust the compositions and methods relating to a particular sequence to other related sequences. Primers and/or probes can be designed for any sequence given the information disclosed herein and known in the art.

Primers and probes

Disclosed are compositions including primers and probes, which are capable of interacting with the genes disclosed herein. In certain embodiments the primers are used to support DNA amplification reactions. Typically the primers will be capable of being extended in a sequence specific manner. Extension of a primer in a sequence specific manner includes any methods wherein the sequence and/or composition of the nucleic acid molecule to which the primer is hybridized or otherwise associated directs or influences the composition or sequence of the product produced by the extension of the primer. Extension of the primer in a sequence specific manner therefore includes, but is not limited to, PCR, DNA sequencing, DNA extension, DNA polymerization, RNA transcription, or reverse transcription. Techniques and conditions that amplify the primer in a sequence specific manner are preferred. In certain embodiments the primers are used for the DNA amplification reactions, such as PCR or direct sequencing. It is understood that in certain embodiments the primers can also be extended using non-enzymatic techniques, where for example, the nucleotides or oligonucleotides used to extend the primer are modified such that they will chemically react to extend the primer in a sequence specific manner. Typically the disclosed primers hybridize with the nucleic acid or region of the nucleic acid or they hybridize with the complement of the nucleic acid or complement of a region of the nucleic acid.

Functional Nucleic Acids

Disclosed herein are functional nucleic acids that can be used with the methods disclosed herein to regulate E1 nuclear localization. Functional nucleic acids are nucleic acid molecules that have a specific function, such as binding a target molecule or catalyzing a specific reaction. Functional nucleic acid molecules can be divided into the following categories, which are not meant to be limiting. For example, functional nucleic acids include antisense molecules, aptamers, ribozymes, triplex forming molecules, and external guide sequences. The functional nucleic acid molecules can act as effectors, inhibitors,

modulators, and stimulators of a specific activity possessed by a target molecule, or the functional nucleic acid molecules can possess a de novo activity independent of any other molecules.

Functional nucleic acid molecules can interact with any macromolecule, such as DNA, RNA, polypeptides, or carbohydrate chains. Thus, functional nucleic acids can interact with the mRNA of E1 or the genomic DNA of E1, or they can interact with the polypeptide E1. Often functional nucleic acids are designed to interact with other nucleic acids based on sequence homology between the target molecule and the functional nucleic acid molecule. In other situations, the specific recognition between the functional nucleic acid molecule and the target molecule is not based on sequence homology between the functional nucleic acid molecule and the target molecule, but rather is based on the formation of tertiary structure that allows specific recognition to take place.

Antisense molecules are designed to interact with a target nucleic acid molecule through either canonical or non-canonical base pairing. The interaction of the antisense molecule and the target molecule is designed to promote the destruction of the target molecule through, for example, RNAseH mediated RNA-DNA hybrid degradation. Alternatively the antisense molecule is designed to interrupt a processing function that normally would take place on the target molecule, such as transcription or replication. Antisense molecules can be designed based on the sequence of the target molecule. Numerous methods for optimization of antisense efficiency by finding the most accessible regions of the target molecule exist. Exemplary methods would be in vitro selection experiments and DNA modification studies using DMS and DEPC. It is preferred that antisense molecules bind the target molecule with a dissociation constant (k_d) less than or equal to 10^{-6}, 10^{-8}, 10^{-10}, or 10^{-12}. A representative sample of methods and techniques which aid in the design and use of antisense molecules can be found in the following non-limiting list of United States patents: 5,135,917, 5,294,533, 5,627,158, 5,641,754, 5,691,317, 5,780,607, 5,786,138, 5,849,903, 5,856,103, 5,919,772, 5,955,590, 5,990,088, 5,994,320, 5,998,602, 6,005,095, 6,007,995, 6,013,522, 6,017,898, 6,018,042, 6,025,198, 6,033,910, 6,040,296, 6,046,004, 6,046,319, and 6,057,437.

Aptamers are molecules that interact with a target molecule, preferably in a specific way. Typically aptamers are small nucleic acids ranging from 15-50 bases in length that fold into defined secondary and tertiary structures, such as stem-loops or G-quartets.

Aptamers can bind small molecules, such as ATP (United States patent 5,631,146) and theophiline (United States patent 5,580,737), as well as large molecules, such as reverse transcriptase (United States patent 5,786,462) and thrombin (United States patent 5,543,293). Aptamers can bind very tightly with k_ds from the target molecule of less than 10^{-12} M. It is preferred that the aptamers bind the target molecule with a k_d less than 10^{-6}, 10^{-8}, 10^{-10}, or 10^{-12}. Aptamers can bind the target molecule with a very high degree of specificity. For example, aptamers have been isolated that have greater than a 10000 fold difference in binding affinities between the target molecule and another molecule that differ at only a single position on the molecule (United States patent 5,543,293). It is preferred that the aptamer have a k_d with the target molecule at least 10, 100, 1000, 10,000, or 100,000 fold lower than the k_d with a background binding molecule. It is preferred when doing the comparison for a polypeptide for example, that the background molecule be a different polypeptide. For example, when determining the specificity of aptamers, the background protein could be E1. Representative examples of how to make and use aptamers to bind a variety of different target molecules can be found in the following non-limiting list of United States patents: 5,476,766, 5,503,978, 5,631,146, 5,731,424 , 5,780,228, 5,792,613, 5,795,721, 5,846,713, 5,858,660 , 5,861,254, 5,864,026, 5,869,641, 5,958,691, 6,001,988, 6,011,020, 6,013,443, 6,020,130, 6,028,186, 6,030,776, and 6,051,698.

Ribozymes are nucleic acid molecules that are capable of catalyzing a chemical reaction, either intramolecularly or intermolecularly. Ribozymes are thus catalytic nucleic acid. It is preferred that the ribozymes catalyze intermolecular reactions. There are a number of different types of ribozymes that catalyze nuclease or nucleic acid polymerase type reactions which are based on ribozymes found in natural systems, such as hammerhead ribozymes, (for example, but not limited to the following United States patents: 5,334,711, 5,436,330, 5,616,466, 5,633,133, 5,646,020, 5,652,094, 5,712,384, 5,770,715, 5,856,463, 5,861,288, 5,891,683, 5,891,684, 5,985,621, 5,989,908, 5,998,193, 5,998,203, WO 9858058 by Ludwig and Sproat, WO 9858057 by Ludwig and Sproat, and WO 9718312 by Ludwig and Sproat) hairpin ribozymes (for example, but not limited to the following United States patents: 5,631,115, 5,646,031, 5,683,902, 5,712,384, 5,856,188, 5,866,701, 5,869,339, and 6,022,962), and tetrahymena ribozymes (for example, but not limited to the following United States patents: 5,595,873 and 5,652,107). There are also a number of ribozymes that

are not found in natural systems, but which have been engineered to catalyze specific reactions de novo (for example, but not limited to the following United States patents: 5,580,967, 5,688,670, 5,807,718, and 5,910,408). Preferred ribozymes cleave RNA or DNA substrates, and more preferably cleave RNA substrates. Ribozymes typically cleave nucleic acid substrates through recognition and binding of the target substrate with subsequent cleavage. This recognition is often based mostly on canonical or non-canonical base pair interactions. This property makes ribozymes particularly good candidates for target specific cleavage of nucleic acids because recognition of the target substrate is based on the target substrates sequence. Representative examples of how to make and use ribozymes to catalyze a variety of different reactions can be found in the following non-limiting list of United States patents: 5,646,042, 5,693,535, 5,731,295, 5,811,300, 5,837,855, 5,869,253, 5,877,021, 5,877,022, 5,972,699, 5,972,704, 5,989,906, and 6,017,756.

Triplex forming functional nucleic acid molecules are molecules that can interact with either double-stranded or single-stranded nucleic acid. When triplex molecules interact with a target region, a structure called a triplex is formed, in which there are three strands of DNA forming a complex dependant on both Watson-Crick and Hoogsteen base-pairing. Triplex molecules are preferred because they can bind target regions with high affinity and specificity. It is preferred that the triplex forming molecules bind the target molecule with a k_d less than 10^{-6}, 10^{-8}, 10^{-10}, or 10^{-12}. Representative examples of how to make and use triplex forming molecules to bind a variety of different target molecules can be found in the following non-limiting list of United States patents: 5,176,996, 5,645,985, 5,650,316, 5,683,874, 5,693,773, 5,834,185, 5,869,246, 5,874,566, and 5,962,426.

External guide sequences (EGSs) are molecules that bind a target nucleic acid molecule forming a complex, and this complex is recognized by RNase P, which cleaves the target molecule. EGSs can be designed to specifically target a RNA molecule of choice. RNAse P aids in processing transfer RNA (tRNA) within a cell. Bacterial RNAse P can be recruited to cleave virtually any RNA sequence by using an EGS that causes the target RNA:EGS complex to mimic the natural tRNA substrate. (WO 92/03566 by Yale, and Forster and Altman, Science 238:407-409 (1990)).

Similarly, eukaryotic EGS/RNAse P-directed cleavage of RNA can be utilized to cleave desired targets within eukarotic cells. (Yuan et al., Proc. Natl. Acad. Sci. USA 89:8006-8010 (1992); WO 93/22434 by Yale; WO 95/24489 by Yale; Yuan and Altman,

EMBO J 14:1590-168 (1995), and Carrara et al., Proc. Natl. Acad. Sci. (USA) 92:2627-2631 (1995)). Representative examples of how to make and use EGS molecules to facilitate cleavage of a variety of different target molecules be found in the following non-limiting list of United States patents: 5,168,053, 5,624,824, 5,683,873, 5,728,521, 5,869,248, and 5,877,162.

Hybridization/selective hybridization

Also provided herein are sequences that selectively hydridize with nucleic acids that encode SEQ ID NO: 1, 2, or 5-8 under stringent conditions. The term hybridization typically means a sequence driven interaction between at least two nucleic acid molecules, such as a primer or a probe and a gene. Sequence driven interaction means an interaction that occurs between two nucleotides or derivatives or two polynucleotides or derivatives in a sequence specific manner. For example, G interacting with C or A interacting with T are sequence driven interactions. Typically sequence driven interactions occur on the Watson-Crick face or Hoogsteen face of the nucleotide. The hybridization of two nucleic acids is affected by a number of conditions and parameters known to those of skill in the art. For example, the salt concentrations, pH, and temperature of the reaction all affect whether two nucleic acid molecules will hybridize.

Parameters for selective hybridization between two nucleic acid molecules are well known to those of skill in the art. For example, in some embodiments selective hybridization conditions can be defined as stringent hybridization conditions. For example, stringency of hybridization is controlled by both temperature and salt concentration of either or both of the hybridization and washing steps. For example, the conditions of hybridization to achieve selective hybridization may involve hybridization in high ionic strength solution (6X SSC or 6X SSPE) at a temperature that is about 12-25°C below the Tm (the melting temperature at which half of the sequences in a nucleic acid polymer dissociate from their hybridization partners) followed by washing at a combination of temperature and salt concentration chosen so that the washing temperature is about 5°C to 20°C below the Tm. The effects of temperature and the concentrations of salts and of chaotropic solutes are readily determined empirically in preliminary experiments in which samples of reference DNA immobilized on filters are hybridized to a labeled nucleic acid of interest and then washed under conditions of different stringencies. Hybridization temperatures are typically higher for DNA-RNA and RNA-RNA hybridizations. The

conditions can be used as described above to achieve stringency, or as is known in the art. (Chow, L.T., and T.R. Broker. 1981. Electron Microscopy in Biology, Vol. 1, (Ed. J.D. Griffith). John Wiley and Sons, N.Y. pp. 139-188; Chow, L.T. and T.R. Broker. 1989. Meth. Enzymol. 180: 239-261. Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989; Kunkel et al. Methods Enzymol. 1987:154:367, 1987 which is herein incorporated by reference for material at least related to hybridization of nucleic acids). A preferable stringent hybridization condition for a DNA:DNA hybridization can be at about 68°C (in aqueous solution) in 6X SSC or 6X SSPE followed by washing at 68°C. Stringency of hybridization and washing, if desired, can be reduced accordingly as the degree of complementarity desired is decreased, and further, depending upon the G-C or A-T richness of any area wherein variability is searched for. Likewise, stringency of hybridization and washing, if desired, can be increased accordingly as homology desired is increased, and further, depending upon the G-C or A-T richness of any area wherein high homology is desired, all as known in the art.

Another way to define selective hybridization is by looking at the amount (percentage) of one of the nucleic acids bound to the other nucleic acid. For example, in some embodiments selective hybridization conditions would be when at least about, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the limiting nucleic acid is bound to the non-limiting nucleic acid. Typically, the non-limiting primer is in for example, 10 or 100 or 1000 fold excess. This type of assay can be performed under conditions where both the limiting and non-limiting primer are for example, 10 fold or 100 fold or 1000 fold below their k_d, or where only one of the nucleic acid molecules is 10 fold or 100 fold or 1000 fold or where one or both nucleic acid molecules are above their k_d.

Another way to define selective hybridization is by looking at the percentage of primer that is enzymatically manipulated under conditions where hybridization is required to promote the desired enzymatic manipulation. In some embodiments selective hybridization conditions would be when at least about, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the primer is enzymatically manipulated under conditions which promote the enzymatic manipulation. For example, if the enzymatic manipulation is DNA extension, then selective

hybridization conditions would be when at least about 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the primer molecules are extended. Preferred conditions also include those suggested by the manufacturer or indicated in the art as being appropriate for the enzyme performing the manipulation.

Just as with homology, it is understood that there are a variety of methods herein disclosed for determining the level of hybridization between two nucleic acid molecules. It is understood that these methods and conditions may provide different percentages of hybridization between two nucleic acid molecules, but unless otherwise indicated meeting the parameters of any of the methods would be sufficient. For example if 80% hybridization was required and as long as hybridization occurs within the required parameters in any one of these methods it is considered disclosed herein.

It is understood that those of skill in the art understand that if a composition or method meets any one of these criteria for determining hybridization either collectively or singly, it is a composition or method that is disclosed herein.

Methods of Treatment

Disclosed herein are methods of inhibiting papillomavirus replication in a subject comprising the steps of identifying a subject with, or at risk of contracting, papillomavirus; and contacting the subject with a substance that reduces nuclear localization of E1, thereby inhibiting papillomavirus replication. The point of contact with the subject can be at or near an infected area or an area susceptible to infection. Alternatively, the point of contact can be systematic.

Nuclear localization of E1 can be inhibited by inhibiting cyclin/cdk2 maintenance of phosphorylation. Cyclin-dependent kinase inhibitors include p21cip1 and p27kip1. When either of these become elevated, the cyclin/cdk2 complexes are inhibited. For example, the key tumor suppressor protein p53 normally coordinates cellular defense responses against "unscheduled DNA synthesis" or aberrant DNA structures such as stalled replication forks, unresolved recombination structures, or broken chromosomes. p53, among its many properties, is an inducer of p21cip1 transcription in cycling cells, thus inhibiting cyclin E/cdk2 and cyclin A/cdk2 activites. In non-cycling, differentiated cells, p21cip1 mRNA transcription is p53-independent and the protein is steadily translated. If there is no perceived evidence of

DNA damage or unscheduled replication, p21cip1 is ubiquitinated and transported to the cytoplasm to be degraded by the 26S proteosomes. It has been shown that treatment of differentiated keratinocytes in squamous epithelia with the drug lactacystin, a potent inhibitor of the 26S proteasome, results in the rapid accumulation of p21cip1 and the inhibition of viral and cellular replication. An example of another CDK inhibitor is roscovitine.

The cellular replicative helicase is composed of six subunits (mcm proteins 2,3,4,5,6,7 termed the mcm2-7 complex). The cellular helicase is allowed only one round of activity and then the phosphorylated subunits are inactivated or exported from the nucleus to the cytoplasm where they cannot function and may be degraded by the proteosomes. Thus papillomaviruses have adopted and adapted the cyclin E/cdk2 post-translational modification of helicases, with the unique feature that the phosphorylation of the E1 protein enables it to remain nuclear and able to reassemble on new templates to carry out multiple rounds of viral DNA replication and amplification, leading to production of large numbers of viruses from each infected and productive cell.

There exists a stochastic distribution between infected cells that begin and permit the breakthrough of viral replication and another subset of cells in which p27kip1 and p21cip1 gain the upper hand and successfully inhibit viral replication. Thus, natural lesions (benign warts, papillomas and condylomas) are a mosaic of neighboring producer and non-producer cells.

Nuclear importation of E1 can also be inhibited by reducing the interaction of E1 with importin-α or importin-β. Studies in vertebrates indicate that the importin proteins are the nuclear localization signal (NLS) receptors that bind to NLS-containing proteins in the cytoplasm and transport them into the nucleus. Once inside the nucleus, the importins are released from the NLS-containing proteins and exported to the cytoplasm for another cycle of import. Therefore, by disrupting the interaction of importin-α or importin-β with the E1 protein, nuclear importation can be reduced or prevented, and viral replication can be reduced.

Nuclear localization of E1 can also be reduced by stimulating CRM1. CRM1 is a member of the importin-β family of transport receptors. CRM1 interacts with several effector proteins modulating its affinity to different cargo molecules, as well as the master regulator protein of nuclear transport, the small GTPase Ran. E1 contains a nuclear export seqeuence which is recognized by the nuclear exportin CRM1. Serine 107 of E1 is in the middle of the CRM1 binding site, and ser 89 and ser 93 are just upstream in the primary sequence and very

close in the folded E1 structure. Phosphorylation on any of these sites and especially on ser 107 blocks the binding of CRM1 (Example 6). Accordingly, the E1 helicase remains nuclear when phosphorylated. To increase the exportation of E1 from the nucleus, the amount of CRM1 present in the cell can be increased, thereby increasing the amount of E1 exported from the cell.

By "papillomavirus" is meant any of the papillomavirus types. Sequence alignment data shows comparable sequences of many genotypes of papillomavirus (Table 2). The mechanisms of E1 regulation is universal amongst papillomaviruses, so one of skill in the art can readily identify and apply the methods and compositions disclosed herein to any of papillomavirus.

TABLE 2

PV Sequence Alignments

HPV54
MADNQGTEE.....EGTGCNGWFFVEAIVERKTGD.IISDDEPEDVE.DS.GLDMVDFID
NSVSQVEGQE.N.PQAL 67
HPV32
MADDTGTE.....EGLGCSGWFSVEAIVERTTEN.TISDDEDENVE.DS.GLDLVDFVD
DSRIIP.TNQLK.AQAL 66
HPV42
MADDTGTE.....EGLGCSGWFCVEAIVDKTTEN.AISDDEDENVD.DS.GLDLVDFVD
NSTV..IHTKQVHAQAL 66
CPV1
MADDTGTDN.....EGTGCSGWFLVEAIVDKTTGE.QVSDDEDETVE.DS.GLDMVDFI
DDRPI..THNSLE.AQAL 66
HPV11R
MADDSGTEN.....EGSGCTGWFMVEAIVEHTTGT.QISEDEEEEVE.DS.GYDMVDFI
DDRHI..TQNSVE.AQAL 66
HPV13
MAEDTGTNN.....EGTGCSGWFLVEAVVERTTGQ.QISDDEDETVE.DS.GLDMVDFI
DDRPI..THNSVE.AQAL 66
HPV44
MADNTGTE.....GTGCSGWFLVEAIVENTTGQ.QISEDEDEAVE.DS.GLDMVDFIDD
RPI..THNSME.AQAL 64
HPV55
MADNTGTE.....GTGCSGWFLVEAIVEKTTGQ.QISEDEDEAVD.DS.GLDMVDFID
DRPI..THNSME.AQAL 64
HPV6bR
MADDSGTEN.....EGSGCTGWFMVEAIVQHPTGT.QISDDEDEEVE.DS.GYDMVDFI
DDSNI..THNSLE.AQAL 66
HPV34
MADS.G.NW.....EG.RCSGWFNVEAIVERKTGD.AIPADENYDGD.DTEDSEMGDFID
NAHISNIYSQQEIAQAL 67
HPV73
MADS.G.NW.....EG.RCTGWFNVEAIVERKTGD.PIPEDENYDGG.DTDESEMGDFID
NAHIPNIYAQQEIAQAL 67
RHPV1R
M.DPEGTPG.....EGVGCTGWFNVEAIVERKTGD.VVSEDEDDT.E.DT.GIDLVDFIDD
TCGSV.QTGDEAPGAL 66
HPV10
MDDNTGTEGGA..CSESERAGGWFIVEAIVDRRTGD.PISSDDDEEED.EA.GEDFVD
FIDDTRSLGDGQE.V.AQEL 71
HPV28
MDDTSGTEGDE..CSELERAGGWFMVEAIVDRRTGD.KPSSDEDEDEDADE.GEDF
VDFIDDRPVGD.GQE.V.AQEL 71
HPV29
MADNSGTEGEEEDCSEAERAGGWFMVEAIVDRRTGD.TISSDEDEE...DE.GEDMV
DFIDDRPIGD.GQE.V.AQEL 70
HPV3
MDDTSGTEGE...CSELERAGGWFMVEAIVDRRTGD.TVSSDEDEE.E.DG.GEDLVDF

IDDRPVGD.GQE.V.AQEL 68

HPV77

MADNSGTEGEEEDCSEAERAGGWFIVEAIVDKRTGD.TISSDEDEEGE....GEDMVD FIDDRPIGD.GQE.V.AQEL 70

HPV61

MADSEGTESGD.GTEAAERAGGWFLVEAVVDRTTGY.QVSSDEEDNSI.DT.GEDLV DFIDTRRPGD.GQE.V.PLAL 71

HPV2a

MEDSEGTDGT...EEDGCRAGGWFHVEAII..THGQRQVSSDEDEDET.ET.GEDL.DFI DNRVPGD.GQE.I.PLQL 67

HPV27

MEDSEGTDGT...EEDGCRAGGWFHVEAII..THGQRQVSSDEDEDCT.ET.GEDV.DFI DNRVPGD.GQE.I.PLQL 67

HPV57

MEDSEGTDGT...DEDGCRAGGWFHVEAII..THGQSQVSSDEDEDET.ET.REDL.DFI DNRVPGD.GQE.V.PLQL 67

HPV26

M.DCEGTNE.....EGRGCTGWFSVEAIVEKHTGD.TISDDETDNSS.DT.GSDLIGFIDD SSISD.YAEQEVTQAL 67

HPV51

M.DCEGTED.....EGAGCNGWFFVEAIVEKKTGD.NVSDDEDENAD.DT.GSDLINFID SETSICSQAEQETARAL 68

HPV30

MASPEGTDD.....EGGGCTGWFHVEAVVKKRTGD.IISEDETEEDE.GT.ASDLDGFL DNSNVITTQADRETAQQL 69

HPV53

MASPEGTDD.....EG.GCRGWFHVEAIVKKRTGD.VISEDETDE.E.ST.ESDLDGFIDN SNIISTQAERETAQQL 67

HPV56

MASPEGTDG.....EGKGCCGWFEVEAIVEKKTGD.KISDDESDEED.EI.DTDLDGFID DSYIQNIQADAETGQQL 69

HPV66

MASPEGTDG.....EGMGCCGWFQVEAIVERKTGD.TISDDESEEEN.ET.DTDVDGFID NTLINNTQEDRETAQQL 69

HPV18R

MADPEGTDG.....EGTGCNGWFYVQAIVDKKTGD.VISDDEDENAT.DT.GSDMVDF IDTQGTFCEQAELETAQAL 69

HPV39

MANREGTDG.....DGSGCNGWFLVQAIVDKQTGD.TVSEDEDENAT.DT.GSDLADFI DDSTDICVQAERETAQVL 69

HPV45

MADPEGTDG.....EGTGCNGWFFVETIVEKKTGD.VISDDEDETAT.DT.GSDMVDFI DTQLSICEQAEQETAQAL 69

HPV59

MADSEGTDG.....EGTGCNGWFFVQAIVDKKTGD.KISDDEDENAT.DT.GSDLVDFI DDTTTICVQAERETAQAL 69

HPV70

MANCEGTDG.....DGSGCNGWFLVQAIVDKQTGD.TVSEDEDENAT.DT.GSDLADFI DDTTDICVQAERETAQVL 69

PCT/US05/11740

HPV40
MADSPGTED.....GGAGCSGWFVVEAVVDKQTGD.AVSEDEDEEDIEDS.GFDMIDF
IDNSVVAEEHVELSNAQAL 70
HPV7
MADDSGTED.....VGSGCSGWFLVEAVVDKQTGD.VVSEDEDEDAIEDS.GYDMVD
FINDTVVSE.HEELSNAQAL 69
HPV16R
MADPAGTNG.....EEGTGCNGWFYVEAVVEKKTGD.AISDDENENDS.DT.GEDLVD
FIVNDNDYLTQAETETAHAL 70
HPV31
MADPAGTDG.....EGTGCNGWFYVEAVIDRQTGD.NISEDENEDSS.DT.GEDMVDFI
DNCNVYNNQAEAETAQAL 69
HPV33
MADPEGTNG.....AGMGCTGWFEVEAVIERRTGD.NISEDEDETAD.DS.GTDLLEFID
DSMENSIQADTEAARAL 69
HPV35h
MADPAGTDE.....GEGTGCNGWFFVEAVVSRRTGD.PVSEDENEDDC.DR.GEDMVD
FINDTDILNIQAETETAQAL 70
HPV52
MEDPEGTEG.....EREGCTGWFEVEAIIEKQTGD.NISEDEDENAY.DS.GTDLIDFIDD
SNINNEQAEHEAARAL 69
HPV58
MDDPEGTNG.....VGAGCTGWFEVEAVIERRTGD.NISDDEDETAD.DS.GTDLIEFID
DSVQSTTQAEAEAARAL 69
HPV67
MEDPEGTDG.....EQGGCTGWFSVEAIIEKRTGD.KVSEDEDEDAS.DT.GSDLIGFID
DTHIENIQADTQAARAL 69

HPV54
LHAQQLQADVEAVQQLKRKYIG.SPY.VSPVANSEPCVEKDLSPRLGAISLGRRSAK
AKRRLFD..KAQPPPNGHTD. 140
HPV32
LNRQQAHADKEAVQALKRKLLG.SPY.ESPASDLQESINKELSPRLGGLQLCRGSQG
AKRRLFQSLENRDSGYGYSE. 141
HPV42
LNKQQAHADQEAVQALKRKLLG.SPY.ESPVSDSQHSIDNELSPRLGGLTLCRGSQG
AKRRLFQSLENRDSGYGYSE. 141
CPV1
LNEQEADAHYAAVQDLKRKYLG.SPY.VSPLGHIEQSVDCDISPRLDAIQLSRKPKK
VKRRLFQSREITDSGYGYSE. 141
HPV11R
FNRQEADAHYATVQDLKRKYLG.SPY.VSPISNVANAVESEISPRLDAIKLTTQPKK
VKRRLFETRELTDSGYGYSE. 141
HPV13
LNEQEADAHYAAVQDLKRKYLG.SPY.VSPLGHVEQSVDCDISPRLDAIKLSRNSK
KVKRRLFQSREITDSGYGYSE. 141
HPV44
LNEQEADAHYAAVQDLKRKYLG.SPY.VSPLSNIEQAVECDISPRLDAITLSRQPKK
VKRRLFDRPELTDSGYGNTE. 139

PCT/US05/11740

HPV55
LNEQEADAHYAAVQDLKRKYLG.SPY.VSPLSNIKEAVECDISPRLDAIKLSRQPKK VKRRLFERPELTDSGYGNID. 139
HPV6bR
FNRQEADTHYATVQDLKRKYLG.SPY.VSPINTIAEAVESEISPRLDAIKLTRQPKKV KRRLFQTRELTDSGYGYSE. 141
HPV34
YHSQQVNADNEAIRVLKRKFAG.SAG.SSPDSKRHELKHKQRSPHILTIRDTNTT..ST HLLC...EEQDSGYGNTE. 137
HPV73
YQSQQANADNEAIRVLKRKFTG.SPG.GSPDMKRDEFIDKQLSPQINVLSISSGRSTS KRRLF...EEQDSGYGNTE. 139
RHPV1R
LHAQETQAHAEAVQVLKRKFVG.SPA.VSPLGNYNPCVDRDLSPRLNEISLNQGSG QAKRRLF....LPDSGYGNTE. 137
HPV10
FQQQTAADDDVAVQTVKRKFAP.SPY.FSPVCE.QASIEHELSPRLDAIKLGRQSAK AKRRLF...ELPDSGYGQTQ. 142
HPV28
LLQQAAADDDVAVQAVKRKFAP.SPY.FSPVCM.QPSIENELSPRLDAIKLGRQSGT AKRRLF...QLPDSGYGQTQ. 142
HPV29
LLQQAAADDDEAVHTVKRKFAP.SPY.FSPVCV..PSIEHELSPRLDAIKLGRQSSKAK RRLF...QLPDSGYGQTQ. 140
HPV3
LLQQAAADDDVEVQTVKRKFAP.SPY.FSPVCV.HPSIENELSPRLDAIKLGRQTSKA KRRLF...ELPDSGYGQTQ. 139
HPV77
LLQQAAADDDVAVQAVKRKFTH.SPY.FSPLCV.DPSIEHELSPRLDAIKLGRESAKA KRRLF...QLPDSGYGQTQ. 141
HPV61
FVQQNAQDDAATVQALKRKYTC.SPA.SSTCV...SLVDSELSPRLDAIRIHRGQDRA RRRLF...EQ.DSGYGHTQ. 139
HPV2a
YAQQTAQDDEATVQALKRKFVA.SPL.SACSC.....IENDLSPRLDAISLNRKSEKAKR RLFE.TEPPDSGYGNTQM 137
HPV27
YTQQIAQDDEATVQALKRKFVA.SPL.SACSC.....IENDLSPRLDAISLNRKSEKAKRR LFE.TEPPDSGYGNTQM 137
HPV57
YAQQIAQDDEATVQALKRKFVA.SPL.SACSC.....IENDLSPRLDAISLNRKSEKAKR RLFE.TEPPDSGYGNTQM 137
HPV26
FQAQQKQANTKAVRNLKRKLLG.SQN..SPLQD.ITNQHRQQSDSQQNTHQVNNS.Q AKRRAVD..SVPDSGYGYTE. 137
HPV51
FQAQELQANKEAVHQLKRKFLV.SPR.SSPLGD..ITNQNN....THSHSQANES.QVKRR LLD..SYPDSGYGNTQ. 134
HPV30

LHAQNTYADTQTLHNLKRKYLG.SPL.GDISNQ..QFVCRE.....GVKRRIID.TDV
ADSGYGNTLE 127
HPV53
LHAQNTHADTQTLQKLKRKYLG.SPL.GDISNQ..QTVCRE.....AVKRRLID.TE
VPDSGYGNTL. 124
HPV56
LQVQTAHADKQTLQKLKRKYIA.SPL.RDISNQ..QTVCRE.....GVKRRLIL.SDL
QDSGYGNTL. 126
HPV66
LQVQTAHADAQTLQKLKRKYIG.SPL.SDISNQ..QTVYRE.....EVKRRLI...LSED
SGYGNTL. 124
HPV18R
FHAQEVHNDAQVLHVLKRKFAGGSTE.NSPLGE.RLEVDTELSPRLQEISLNSGQKK
AKRRLF...TISDSGYGCSE. 141
HPV39
LHMQEAQRDAQAVRALKRKYTD.SSGDTRPYG...KKVGRNTRGTLQEISLNVSSTQ
ATQTVY...SVPDSGYGNME. 139
HPV45
FHAQEVQNDAQVLHLLKRKFAGGSKE.NSPLGE.QLSVDTDLSPRLQEISLNSGHKK
AKRRLF...TISDSGYGCSE. 141
HPV59
FNVQEAQRDAREMHVLKRKFGC.SIE.NSSEKA..AAGKKAKSP.LQEISVNVNHPKV
KRRLI...TVPDSGYGYSE. 138
HPV70
YNMQEAQRDAQSVRALKRKYGG.SNLNKSPCAK.PPGVHREQRVTLQELPVNICN
KQARTNVY...SVPDSGYGNME. 141
HPV40
LHVQQTCADAADLCELKRKYI..SPY.VSPIQYSEPSIDGDLSPRLHAIRLGGGQ.KAK
RRLFQRVEQRDSGYGYSE. 143
HPV7
LHAQQTCADAVELCELKRKYI..SPY.VSPIQCSEPSVDGDLSPRLHAIKLGGGK.KA
KRRLFERLEQRDSGYGYSQ. 142
HPV16R
FTAQEAKQHRDAVQVLKRKYLG.SPL.SDISG....CVDNNISPRLKAICIEKQSRAAKR
RLF...ESEDSGYGNTE. 138
HPV31
FHAQEAEEHAEAVQVLKRKYVG.SPL.SDISS....CVDYNISPRLKAICIENNSKTAKR
RLF...ELPDSGYGNTE. 137
HPV33
FNIQEGEDDLNAVCALKRKFAA.CSQ.SAAED....VVDRAANPCRTSINKNKECTYR
KRKID...ELEDSGYGNTE. 137
HPV35h
FHAQEEQTHKEAVQVLKRKYAS.SPL.SSVSL....CVNNNISPRLKAICIENKNTAAKR
RLF...ELPDSGYGNSE. 138
HPV52
FNAQEGEDDLHAVSAVKRKFTS.SPE.SAGQD....GVEKHGSPRAKHICVNTECVLPK
RKPC...HVEDSGYGNSE. 137
HPV58
FNVQEGVDDINAVCALKRKFAA.CSE.SAVED....CVDRAANVCVSWKYKNKECTH

RKRKII...ELEDSGYGNTE. 137
HPV67
FNLQEEEDDLNAVSALKRKFTG.IQA...CGGN.SNGIENQ.....ACTAAKRRAY...DI
EDSGYGNTE. 126

HPV54 VEA...A..VEVNT.....EGTD.....ETETDQ..... 159
HPV32 VEI..RQ..EQVEN.....GHG.....APDGSM..G.....NGGG.....M
166
HPV42
VEV..QQ..TQVEH.....GHG.....AVHGTM.GN.....GGAV.....G 167
CPV1 VET...A..TQVER.....YGEP.....ENGCGG.GG.....DGRE....KE
168
HPV11R
VEA...A..TQVEK.....HGDP.....ENGGDG.QE.....RDTGR..DIE 170
HPV13
VEA...E..TQVER.....NGEP.....ENDCGG.GG.....HGRD....KE 168
HPV44 VEA...E..TQVER.....NGEP.....EDCGGG..G.....QGRD....TE
165
HPV55 VEA...E..TQVER.....NGEP.....EDCGGG..G.....QGRD....TE
165
HPV6bR
VEA..GTG.TQVEK.....HGVP.....ENGGDG.QE.....KDTG....RD 170
HPV34
VET..YE..RQVPG.....PGGC.....LQSTSS.SN.....NGSQMA.SPG 168
HPV73
VET..YE..TEVPG.....LGAGVG.CLQNVNEE.GN.....QIVS....PR 171
RHPV1R VET..SL..LQVAG.....GGG.....QDVQAG..... 156
HPV10
VDT..ESGPKQVQGSSETQDGRQDDDEGSVVQSTLDTGNQNGRQNND.....EGSGRN
VGE.....HGSQ....EE 202
HPV28
VDT..ESGPLQVQDICETG.....TQDGRQDAD.....EGSGRNVGG.....NGGQ....E
E 183
HPV29
VDT..DTGPSQVQDGCETG.....DQNGRQQYK.....EGSGTKDGE.....NGSQ....
EE 181
HPV3
VDT..ESGPKQVQDICKTS.....QQDGCQGAD.....EGRGRNVGG.....NGSQ....E
E 180
HPV77
VDTDSGP..SQVQDICEAG.....GQDGRQTNT.....EGSGTEQGE.....NVSR....EE
182
HPV61
VEI..GASESQVPG.....DAQH.....EGGGES..V.....QEAE.....E 167
HPV2a
VVG..TP..EEVTGD.....EESQG.....GRPVED.QE.....EERQ.....G 166
HPV27
VVG..TP..EEVTGD.....ENSEG.....GRPVED.KE.....EERQ.....G 166
HPV57

VVG..TP..EEVTGE.....DNSQG.....GRPVDV.RE.....EERQ.....G 166
HPV26 VET.LTP..VQVDK.....QYE.....ENGGLP..S.....VCSQ.....G 163
HPV51 VET..VEATLQVDG.....QHG.....GSQNS.....VCSS.....G 159
HPV30
VEA..TQ...QVQD.....NTYG....SGKQQDG.GS.....QTSVC...SR 156
HPV53
.ET..VEATQQVQE.....QYVR.....EASGGKEQN.....GGSQHSVCSR 158
HPV56
.ET..LETPEQVDE.....EVQG.....RGCGNT.QN.....GGSQN..STY 157
HPV66
.ET..LETSQQVEY.....EKG.....NGCGSS.QN.....GGSQ....NS 152
HPV18R
VEA..TQ..IQVTT.....NGEH.....GGNVCS.GGSTEAIDNGGTE.....G 175
HPV39
VET..AEV.EEVTV.....ATNT.....NGDAEG.EH.....GGSV....RE 168
HPV45 VEA..AE..TQVTV.....NTNA.....ENGGSV.....HSTQ.....S 166
HPV59
VEM..LE..TQVTV.....ENTG.....NGDSNG..S.....VCSD....SQ 165
HPV70
VET..AE..VEVTVVN.....NTNGEE.....EGENGG.EN.....GGSI....RE 173
HPV40
VET..TE..RQVET.....EHGGP.....EDTVGGSGR.....VTTD....EA 173
HPV7
VET..TE..TQVEE.....EHGEP.....EGIEGGSGR.....AATV....ET 172
HPV16R
VET..QQM.LQVEG.....RHETETPCSQYSGGS.GG.....GCSQY..SSG 174
HPV31 VET..QQM.VQVEE.....QQTT.....LSCNGS..... 158
HPV33
VET..QQMVQQVES.....QNGD.....TNLNDL..E.....SSGV.....G 165
HPV35h VEI..QQI.QQVEG.....HDTV.....EQCSMG.....S 160
HPV52
VEA..QQMADQVDG.....QNGD.....WQSNSS.QS.....SGVG.....A 166
HPV58
VET..EQMAHQVES.....QNGD.....ADLNDS..E.....SSGV.....G 165
HPV67
VET..QETMVQVEG.....QNGD.....MQCSSQ.CS.....TGASS...TG 157

HPV54
..VQTV.SGET...TTD....SL..GRQQITELIHNTN.IRVALFGMFKDLYGLSFMDLARPFK SDKTVCTDWVIAAF 224
HPV32
GSVHGV.QENQ...EIGT...NT..PTTRVVELLKCKN.LQATLLGKFKELFGLSFGDLVRQ FKSDKSSCTDWVVAAF 234
HPV42

SELGVQ.ENEE...GSTT...ST..PTTRVVELLKCKN.LHATLLGKFKELFGVSFGDLVRQ
FKSDKSSCTDWVIAAF 235
CPV1
GEGQVH.TEVHTESEIEQ...HT..GTTRVLELLKCKD.VRATLHGKFKECYGLSFKDLT
REFKSDKTTCGDWVVAGF 239
HPV11R
GEGVEH.REAEA.VDDSTR.EHA..DTSGILELLKCKD.IRSTLHGKFKDCFGLSFVDLI
RPFKSDRTTCADWVVAGF 242
HPV13
GEGQVH.TEVHTGSQIEE...HT..GTTRVLELLKCKD.VRATLYGKFKDCYGLSFTDLI
RPFKSDKTTCGDWVVAAF 239
HPV44
GVEQVE.TEVQTHSNTQQ...HT..GTTRVLELLKCKN.IRATLLGKFKDCYGLSYTDLI
RQFKSDKTTCGDWVIAAF 236
HPV55
EGVEQV.ETEVQ..THSDTQLHT..ETTRVVELLKCKN.IRATLLGKFKDCYGLSYTDL
IRQFKSDKTTCGDWVIAAF 237
HPV6bR
IEGEEH.TEAEA.PTNSVR.EHA..GTAGILELLKCKD.LRAALLGKFKECFGLSFIDLIR
PFKSDKTTCLDWVVAGF 242
HPV34
ETNSGS.SSISN.MDIDM..EST..PITDITNILKSSN.VKATLLAKFKEVYGLSYMELVRP
YKSDKTQCQDWVCAVF 239
HPV73
ESSSGS.SSISN.MDIET..EST..PITDITNLLQRNN.AKAALLAKFKEVYGLSYMELVRP
YKSDKTHCQDWVCAVF 242
RHPV1R
....GK.ENTR...PDD.....GGGDATQLLRCSN.LKATLLSKFKSVYGVSFSELVRSFKSD
RTTCADWVVGAA 217
HPV10
ERAGGD.GEESD..LQST...STGKGAGGVVEILRASN.KKATLLGKFKEQFGLGYNELI
RHFKSDRTSCADWVVCVF 273
HPV28
ERAGGD.GEESQ..TQGV...QTDKAACGVLAILRASN.QKATLLGKFKEQFGLGFNEL
IRHFKSSKTVCLDWVVCVF 254
HPV29
ERAGGD.GEESQ..PLST...ETEKGACGVLSILKASN.QKATLLGKFKEQFGLGYNELV
RHFKSSRTACVDWVVCVF 252
HPV3
ERAGGD.GEESQ..TESV...QTDTTACGVLAILKASN.HKATLLGKFKEQFGLGFNELI
RHFKSNKTVCSDWVVCVF 251
HPV77
ERAGGD.GEESQ..PLGA...ETERGACGVLSILKASN.QRATLLGKFKEQFGLGYNEL
VRHFKSNRTACADWVVCVF 253
HPV61
ERGGGD.GEAEATGNQET...QAQEQAADILEVFKVSN.LKAKLLYKFKDLFGLAFGE
LVRNFKSDKSICGDWVICAF 240
HPV2a
GDGEAD.LTVHT..PQS....GT.DAAGSVLTLLRSSN.LKATLLSKFKDLFGVGFYELVR

QFKSSKTACADWVVCAY 235

HPV27

GDGEAD.LTVQT..PQS....GT.DAAGSVLTLLRSSN.LKATLLSKFKELFGVGYYELVR QFKSSKTACADWVVCAF 235

HPV57

GDGEAD.LTVHT..PQS....GT.DAAGSVLTLLKSSN.LKATLLSKFKELYGVGYYELV RQFKSSRTACADWVVCAF 235

HPV26

GSNASV.EDID...VDT....HV.NSVTQICELLKCSN.VKAALLSKFKTVYGVSFAELVRV FKSDKTCCSDWVCAAF 231

HPV51

GGSVMD.VET....TESC...AN.VELNSICEVLKSSN.AKATLMAKFKELYGISYNELVR VFKSDKTCCIDWVCALF 227

HPV30

ENSIEA.DSDM...DIG....AT..PPQQIQELLKSSN.VQAKLCYKFKELFGIPFSELVRTFK SDSTCCHDWICAMF 223

HPV53

DGSIGS.GSDMD..VDRQ...DI.MPLQQIQDILKCSN.VQAKLYCKFKDIFGIPFSELVRT FKSDSTCCHDWICAIF 228

HPV56

SNNSED.SVIHM..DIDRN.NET..PTQQLQDLFKSSN.LQGKLYYKFKEVYGIPFSELVR TFKSDSTCCNDWICAIF 228

HPV66

NCSEHS.VSNM...DIDTN.MET..PTHQLQELFKSSN.VQGRLHFKFKEVYGVPYTELV RTFKSDSTCCNDWICAIF 222

HPV18R

NNSSVD.GTSDN.SNIENVNPQC..TIAQLKDLLKVNN.KQGAMLAVFKDTYGLSFTD LVRNFKSDKTTCTDWVTAIF 248

HPV39

ECSSVD.SAIDS.ENQDP...KS..PTAQIKLLLQSNN.KKAAMLTQFKETYGLSFTDLVR TFKSDKTTCTDWVAAIF 238

HPV45

SGGDSS.DNAE...NVDP...HC..SITELKELLQASN.KKAAMLAVFKDIYGLSFTDLVRN FKSDKTTCTDWVMAIF 234

HPV59

IDCSDS.SNMDV.ENIV....PT.SPTNQLLQLLHSKN.KKAAMYAKFKELYGLSFQDLV RTFKSDRTTCSDWVTAIF 235

HPV70

ECSSVD.SAIDS.ENQDP...QS..PTAQLKTVLQANN.QKAILLSQFKHTYGLAFNDLVR TFKSDKTICTDWVAAIC 243

HPV40

EAVEVV.EDGSH..VID....HC.SPRTQLIELFKCKD.LNAKLYGKFKELYGVGFGDLVR QFKSDKSTCTDWVYAVF 242

HPV7

EAVEVL.EESSD..VIQ....QL.SPRTQVVELFKCKD.LNAKLCGKFKELFGVGFHDLVR QFKSDKSTCTDWVYAVF 241

HPV16R

SGGEGV.SERH...TIC....QT..PLTNILNVLKTSN.AKAAMLAKFKELYGVSFSELVRPF KSNKSTCCDWCIAAF 241

PCT/US05/11740
HPV31
..DGTH.SERE...NET.....PTRNILQVLKTSN.GKAAMLGKFKELYGVSFMELIRPFQS
NKSTCTDWCVAAF 221
HPV33
DDSEVS.CETN...VDSC...EN.VTLQEISNVLHSSN.TKANILYKFKEAYGISFMELVRP
FKSDKTSCTDWCITGY 234
HPV35h
GDSITS.SSDE...RHD....ET..PTRDIIQILKCSN.ANAAMLAKFKELFGISFTELIRPFKSD
KSTCTDWCVAAF 227
HPV52
SNSDVSCTSIED..NEEN...SN.RTLKSIQNI.MCENSIKTTVLFKFKETYGVSFMELVRP
FKSNRSSCTDWCIIGM 237
HPV58
ASSDVS.SETD...VDSC...NT.VPLQNISNILHNSN.TKATLLYKFKEAYGVSFMELVRP
FKSDKTSCTDWCITGY 234
HPV67
NSVDMQ.SESNS..SQEQ...SM..PLQTVENIMHVNN.IKATLMHKFKEAYGVTFTQLIR
PFKSDRTSCTDWCVTAF 226

HPV54
GIYHGITDGFKTLLEPHCLYGHIQWLTCRW..GMVLLLLTRFKCGKNRLTVSKCLG
MLLNIPETQMLIDPPKLRTPAA 300
HPV32
GVHHSIAEGFNTLIKAEALYTHIQWLTCTW..GMVLLMLIRFKCGKNRTTVSKGMC
KLLNIPANQLLIEPPRLQSVAA 310
HPV42
GVNHSIAEGFNTLIKADSLYTHIQWLTCTW..GMVLLMLIRFKCGKNRTTVSKGLSK
LLNIPTNQLLIEPPRLQSVAA 311
CPV1
GVHHSVSEAFQKLIQPLSTYSHIQWLTNYKCMGMVLLVLLRFKVNKNRCTVARTL
ATLLNIPEDHMLIEPPKIQSSVA 317
HPV11R
GIHHSIADAFQKLIEPLSLYAHIQWLTNAW..GMVLLVLIRFKVNKSRCTVARTLGT
LLNIPENHMLIEPPKIQSGVA 318
HPV13
GIHHSVSEAFEKLMQPLTTYMHIQWLTNAW..GMVLLVLIRFKVNKSRCTVARTLA
TFLNIPEDHMLIEPPKIQSSVA 315
HPV44
GVHHSVSEAFQNLIQPVTTYSHIQWLTNAW..GMVLLALVRFKVNKNRCTVARM
MATRLNIPEDHMLIEPPKIQSGVA 312
HPV55
GVHHSVSEAFQNLIQPVTTYSHIQWLTNAW..GMVLLALLRFKVNKNRCTVARMM
ATRLNIPEDHMLIEPPKIQSGVA 313
HPV6bR
GIHHSISEAFQKLIEPLSLYAHIQWLTNAW..GMVLLVLLRFKVNKSRSTVARTLATL
LNIPENQMLIEPPKIQSGVA 318
HPV34
GVAPSLAESLKSLLTQYCLYIHLQCLTCSW..GIIVLLLARFKCNKNRLTVQKLLHGL
LNVTQEYMLIEPPRLRSTPC 315

PCT/US05/11740

HPV73
GVIPSLAESLKSLLTQYCMYIHLQCLTCTW..GIIVLVLVRFKCNKNRLTVQKLLSSL
LNVTQERMLIEPPRLRSTPC 318
RHPV1R
GVHHSVAEGLKQLIQPFCSYAHIQCLTCDW..GVYLLLLARFKCGKNRLTVSKCMS
TLLNVQETHMLIEPPKLRSAAA 293
HPV10
GVFCTVAEGIKTLIQPLCDYAHIQVLPCQW..GMTVLMLVRYKRAKNRETVAKGLS
TLLNVPESQMLIEPPKLRSGPA 349
HPV28
GVYCTLAEGIKTLIQPQCDYAHIQVLSCQW..GMTVLMLVRYKRAKNRETVAKGL
STLLNVPESHMLIEPPKLRSGPA 330
HPV29
GVYCTVAEGIKQLIQPLCEYAHIQVLPCQW..GMTVLMLVRYKRAKNRETVAKGL
STLLNVPESHMLIEPPKLRSSPA 328
HPV3
GVYCTLAESFKTLIQPQCEYAHIQVLSCQW..GMTVLTLVRFKRAKNRETVAKGFS
TLLNVPENHMLIEPPKLRSAPA 327
HPV77
GVYCTVAEGIKQLIQPLCDYAHIQVLPCQW..GMTVLMLLRYKRAKNRETVAKGL
STLLNVPESHMLIEPPKLRSGPA 329
HPV61
GVYHAVAEAVKTLIQPICVYAHIQIQTCQW..GMVILMLVRYKCGKSRETVAHSMG
KLLNIPERQMLIEPPKIRSAPC 316
HPV2a
GVYYAVAEGLKKLIQPHTQYAHIQVQTSSW..GMVVFMLLRYNCAKNRDSVSKN
MSMLLNIPEKHMLIEPPKLRSTPA 311
HPV27
GVYYAVAEGIKQLIQPHTQYAHIQVLTCSW..GMVVFMLLRYNCAKNRDTVSKNM
SMLLNIPEKHMLIEPPKLRSTPA 311
HPV57
RVYYAVAEGIKQLIQPHTQYAHIQIQTSSW..GMVVFMLLRYNCAKNRDTVSKNM
SMLLNIPEKHMLIEPPKLRSTPA 311
HPV26
GVAGSVAESIKSLIQQYCLYYHIQCLTCNW..GVIVLMLVRFTCAKNRTTIKNCLCM
LLNVPETQLLIEPPKLRSTAV 307
HPV51
GVSPMVAENLKTLIKPFCMYYHIQCLSCDW..GTIVLMLIRFSCAKNRTTIAKCLSTL
VNIPQSQMFIEPPKLRSTPV 303
HPV30
GVNETLAEALKTIIKSQCMYYHIQCLTCTW..GVVILMLIRYTCGKNRKTIIKSLSSIV
NVPSEQMLVQPPKIRSPAV 299
HPV53
GVNETLAEALKTIIKTQCIYYHMQCLTCTW..GVVILLLIRYTCGKNRKTIVKSLASI
LNVPTEQMLVQPPKIRSPAV 304
HPV56
GVNETLAEALKTIIKPHCMYYHMQCLTCTW..GVIVMMLIRYTCGKNRKTIAKALS
SILNVPQEQMLIQPPKIRSPAV 304
HPV66

GVNETLAEALKTILKPQCVYYHMQCLTCSW..GVIVMMLIRYICGKNRKTITKSLSS
ILNVPQEQMLIQPPKLRSPAV 298
HPV18R
GVNPTIAEGFKTLIQPFILYAHIQCLDCKW..GVLILALLRYKCGKSRLTVAKGLSTL
LHVPETCMLIQPPKLRSSVA 324
HPV39
GVHPTIAEGFKTLINKYALYTHIQSLDTKQ..GVLILMLIRYTCGKNRVTVGKGLSTL
LHVPESCMLLEPPKLRSPVA 314
HPV45
GVNPTVAEGFKTLIKPATLYAHIQCLDCKW..GVLILALLRYKCGKNRLTVAKGLST
LLHVPETCMLIEPPKLRSSVA 310
HPV59
GVNPTVAEGFKTLIQPYVLYAHIQCLDCAW..GVVILALLRYKCGKNRITVAKGLST
LLHVPDTCMLIEPPKLRSGVA 311
HPV70
GVNPTIAEGFKTLIQPYALYTHIQCLDTKY..GVYILLLIRYKCGKNRITVGKGLSKLL
HVPESCMLIEPPKLRSPVA 319
HPV40
GVNPTIAEGFHTLLKRQALYLHTQWTSCKW..GMVLLALCRYKVGKNRETVVRQL
SKMLNVPDNQILVQPPKLQSPPA 318
HPV7
GVNPTIAEGFHTLLKGQALYLHTQWTTCRW..GMVLLALCRYKVAKNRETVVRQL
AKMLNVPDNQLMVQPPKLQSSAA 317
HPV16R
GLTPSIADSIKTLLQQYCLYLHIQSLACSW..GMVVLLLVRYKCGKNRETIEKLLSKL
LCVSPMCMMIEPPKLRSTAA 317
HPV31
GVTGTVAEGFKTLLQPYCLYCHLQSLACSW..GMVMLMLVRFKCAKNRITIEKLLE
KLLCISTNCMLIQPPKLRSTAA 297
HPV33
GISPSVAESLKVLIKQHSLYTHLQCLTCDR..GIIILLLIRFRCSKNRLTVAKLMSNLLS
IPETCMVIEPPKLRSQTC 310
HPV35h
GIAPSVAESLKTLIKPYCLYIHIQCLSCSW..GMVILALLRFKCAKNRTTIEKLLSKLL
CISAASMLIQPPKLRSTPA 303
HPV52
GVTPSVAEGLKVLIQPYSIYAHLQCLTCDR..GVLILLLIRFKCGKNRLTVSKLMSQL
LNIPETHMVIEPPKLRSATC 313
HPV58
GISPSVAESLKVLIKQHSIYTHLQCLTCDR..GIILLLLIRFKCSKNRLTVAKLMSNLLS
IPETCMIIEPPKLRSQAC 310
HPV67
GITPSVAESLKVLIKPQTLYTHLQCLTCDR..GIIILLLVRFKCAKNRLTVSKLMSNLL
SIPETHMIIEPPKIRSTTC 302

HPV54
ALYWYRQGLSNASEIFGTPPEWLARQTVIEYSLADSQFDLSKMVQWAYDHNYID
DSIIALEYAKLADIDENAAAFLGS 378
HPV32

AIYWFRAGISNASVVTGETPEWIQRQTIVEHCFADTQFNLTEMVQWAYDNDLTED
SDIAYEYAQRADTDSNAAAFLKS 388
HPV42
AIYWFRSGISNASIVTGDTPEWIQRQTILEHCFADAQFNLTEMVQWAYDNDITEDS
DIAYEYAQRADRDSNAAAFLKS 389
CPV1
ALYWFRTSISNASIVTGDTPEWIARQTIVEHGLADNQFKLTEMVQWAYDNDYCDE
SDIAFEYAQRADFDSNAKAFLNS 395
HPV11R
ALYWFRTGISNASTVIGEAPEWITRQTVIEHSLADSQFKLTEMVQWAYDNDICEES
EIAFEYAQRGDFDSNARAFLNS 396
HPV13
ALYWFRTGISNASIVTGETPEWIKRQTIVEHGLADNQFKLTEMVQWAYDNDFCDE
SEIAFEYAQRGDFDSNARAFLNS 393
HPV44
ALYWFRSGISNASIVTGETPEWITRQTIVEHGLADNQFKLADMVQWAYDNDFCEE
SEIAFEYAQRADIDANARAFLNS 390
HPV55
ALYWFRSSISNASIVTGETPEWITRQTIVEHGLGDNQFKLTEMVQWAYDNDFCEES
EIAFEYAQRADIDANARAFLNS 391
HPV6bR
ALYWFRTGISNASTVIGEAPEWITRQTVIEHGLADSQFKLTEMVQWAYDNDICEES
EIAFEYAQRGDFDSNARAFLNS 396
HPV34
ALYWYRTSLSNISETVGEVPEWIKRQTVVQHSLEDCQFDLSQMVQWAFDNDITND
CEIAYKYALLASEDSNAAAFLKS 393
HPV73
ALYWYRTSLSNISEIVGDTPEWIKRQTLVQHSLDDSQFDLSQMIQWAFDNDITDDC
EIAYKYALLGNVDSNAAAFLKS 396
RHPV1R
ALYWYRTGISNVSEVIGETPEWITRQTMFQHGLEDSIFDLSEMVQWAYDHDFTDD
SVIAYEYAQLAGIDSNAAAFLKS 371
HPV10
ALYWYKTSMSSCSDVYGETPEWIVRQTMVGHAMEDAQFSLSEMVQWAYDHDIT
DESTLAYEYALIADTDSNAAAFLSS 427
HPV28
ALYWYKTAMSNCSDVYGETPEWIVRQTMVGHALEEAQFSLSEMVQYAYDHDIT
DESMLAFEYALLADTDANAAAFLSS 408
HPV29
ALYWYKTSMSNISDVYGETPEWIVRQTMVGHALQEVQFSLSEMVQWAYDHDITD
EGTLAYEYALIADVDSNAAAFLAS 406
HPV3
ALYWFKTSLSNCSEVFGETPEWIVRQTVVGHALEEAQFSLSEMVQYAYDHDITDE
STLAYEYALQADTDANAAAFLAS 405
HPV77
ALYWYKTGMSNISEVYGDTPDWIVRQTIVGHALEETQFRLSDMVQWAYDHDITD
EGTLAYEYALIAEFDANAAAFLAS 407
HPV61
ALYWYRTAMGNASEVYGETPEWIVRQTVVGHAMQEAQFSLSMLVQWAYDNDIT

DESVLAYEYALLGNEDPNAAAFLAS 394

HPV2a

ALYWYKTAMGNGSEVYGETPEWIVRQTLVGHSMEDEQFRLSVMVQYAYDHDIV
EESVLAFEYAQLADVDANAAAFLNS 389

HPV27

ALYWYKTAMGNGSEVYGETPEWIVRQTLVGHSMEDEQFRLSVMVQFAYDHDIV
EESVLAFEYAQLADVDANAAAFLNS 389

HPV57

ALYWYKTSMGNGSEVYGETPEWIVRQTLIGHSMEDEQFKLSVMVQYAYDHDITD
ESALAFEYAQLADVDANAAAFLNS 389

HPV26

ALYFYKTGLSNISETYGDTPEWIVRQTQLEHSFDDATFDLSKMVQWAFDHDITDD
SEIAFKYAQLADIDSNAAAFLKS 385

HPV51

ALYFYRTGISNISNTYGETPEWITRQTQLQHSFEDSTFELSQMVQWAFDHEVLDDS
EIAFHYAQLADIDSNAAAFLKS 381

HPV30

ALYFYKTAMSNISDIYGETPEWIQRQTQIQHSFQDCQFELSKMVQWAFDNDVTDD
SDIAFYYAQLADVDSNAQAFLKS 377

HPV53

ALYFYKTSISNISDVYGSTPEWIERQTQLQHSFEDCQFELSKMVQWAFDNEVTDDS
QIAFHYAQLADVDSNAQAFLKS 382

HPV56

ALYFYKTAMSNISDVYGDTPEWIQRQTQLQHSLQDSQFELSKMVQWAFDNEVTD
DSQIAFQYAQLADVDSNAQAFLKS 382

HPV66

ALYFYKTAMSNISEVYGETPEWIQRQTQLQHSLQDNQFELSKMVQWAFDNEVTD
DSQIAFLYAQLADIDSNAQAFLKS 376

HPV18R

ALYWYRTGISNISEVMGDTPEWIQRLTIIQHGIDDSNFDLSEMVQWAFDNELTDES
DMAFEYALLADSNSNAAAFLKS 402

HPV39

ALYWYRTGISNISVVTGDTPEWIQRLTVIQHGIDDSVFDLSDMVQWAFDNEYTDE
SDIAFNYAMLADCNSNAAAFLKS 392

HPV45

ALYWYRTGISNISEVSGDTPEWIQRLTIIQHGIDDSNFDLSDMVQWAFDNDLTDES
DMAFQYAQLADCNSNAAAFLKS 388

HPV59

ALYWYRTGMSNISEVIGETPEWIQRLTIIQHGVDDSVFDLSEMIQWAFDNDLTDES
DIAYEYALIADSNSNAAAFLKS 389

HPV70

ALYWYRTGMSNISEVSGTTPEWIQRLTVIQHGIDDSVFDLSDMVQWAFDNDVTED
SDIAYGYALLADSNSNAAAFLKS 397

HPV40

ALFWFRAGMGNGSEVSGTTPEWIAKQTMLEHSFADTQFSLTDMVQWAYDNGHT
DECEIAYYYAQRADVDANAAAFLKS 396

HPV7

ALFWFRSGMGNGSEVSGTTPEWIAKQTMLEHSFAEAQFSLTQMVQWAYDNGHT
DECEIAYYYAQIADIDANAAAFLKS 395

PCT/US05/11740

HPV16R
ALYWYKTGISNISEVYGDTPEWIQRQTVLQHSFNDCTFELSQMVQWAYDNDIVD
DSEIAYKYAQLADTNSNASAFLKS 395
HPV31
ALYWYRTGMSNISDVYGETPEWIERQTVLQHSFNDTTFDLSQMVQWAYDNDVM
DDSEIAYKYAQLADSDSNACAFLKS 375
HPV33
ALYWFRTAMSNISDVQGTTPEWIDRLTVLQHSFNDNIFDLSEMVQWAYDNELTD
DSDIAYYYAQLADSNSNAAAFLKS 388
HPV35h
ALYWFKTAMSNISEVDGETPEWIQRQTVLQHSFNDAIFDLSEMVQWAYDNDFIDD
SDIAYKYAQLAETNSNACAFLKS 381
HPV52
ALYWYRTGLSNISEVYGTTPEWIEQQTVLQHSFDNSIFDFGEMVQWAYDHDITDD
SDIAYKYAQLADVNSNAAAFLKS 391
HPV58
ALYWFRTAMSNISDVQGTTPEWIDRLTVLQHSFNDDIFDLSEMIQWAYDNDITDD
SDIAYKYAQLADVNSNAAAFLRS 388
HPV67
ALYWFRTGMSNISEVSGQTPEWIERLTVLQHSFDDTIFDLGEMVQWAYDNDITDD
SEIAYQYAMLADVNSNAAAFLKS 380

HPV54
NCQAKYVKDCGTMCRHYIRAQKMQMTMSQWIKHRCDLVEEEGEWKEIVRFLRY
QHVDFISFMIALKQFLQGIPKHNCI 456
HPV32
NCQAKYVKDCGIMCRHYKKAQMKRMSMPQWIKHRSERTGDNGDWRPIVKFIRY
QGIDFLTFMSAFKKFLHNIPKKSCL 466
HPV42
NCQAKYVKDCGVMCRHYKKAQMRRMSMGAWIKHRSAKIGDSGDWKPIVKFIR
YQQIDFLAFMSAFKKFLHNIPKKSCL 467
CPV1
NCQAKYVKDCATMCKHYKNAEMKKMSIKQWIKYRSNKIDETGNWKPIVQFLRH
QGIEFISFLSKLKLWLHGTPKKNCI 473
HPV11R
NMQAKYVKDCAIMCRHYKHAEMKKMSIKQWIKYRGTKVDSVGNWKPIVQFLRH
QNIEFIPFLSKLKLWLHGTPKKNCI 474
HPV13
NCQAKYVKDCATMCKHYKNAEMKKMSMKQWITYRSKKIEEAGNWKPIVQFLR
HQNIEFIPFLSKLKLWLHGTPKKNCI 471
HPV44
NCQAKYVKDCATMCKHYKTAEMKKMNMKQWIKFRSSKFEDTGNWKPIVQFLR
HQNIEFIPFLTKLKMWLHGTPKKNCI 468
HPV55
NCQAKYVKDCATMCKHYKTAEMKKMSMKQWIKFRSSKYEETGNWKPIVQFLR
YQNIEFIPFLTKLKMWLHGTPKKNCI 469
HPV6bR
NMQAKYVKDCATMCRHYKHAEMRKMSIKQWIKHRGSKIEGTGNWKPIVQFLRH
QNIEFIPFLTKFKLWLHGTPKKNCI 474

PCT/US05/11740

HPV34
NAQAKYVKDCGTMCRHYKAAERKQMTMSQWITHRCDLIDDGGNWKHIVQFLR
YQQVEFVPFLIALKQFLKGIPKQNCI 471
HPV73
NAQAKYVKDCGTMCRHYKAAERKQMSMAQWIQHRCDLTNDGGNWKDIVLFLR
YQNVEFMPFLITLKQFLKGIPKQNCI 474
RHPV1R
NAQAKYVKDCATMCRHYKRAERQQMTMSQWIKQRCEKTDDGGDWRPIVQFLR
YQGVEFIAFLAALKLFLKGIPKKNCI 449
HPV10
NCQAKYLKDACTMCRHYKRGEQARMSMSEWIWFRGDKVQGDGDWKPIVQFLR
YQDVEFIPFLCAFKTFLQGVPKKSCL 505
HPV28
NCQAKYVKDACTMCRHYKRGEQARMNMSEWIWFRGDKVQGDGDWKPIVQFLR
YHDVEFIPFLCAFKTFLQGIPKKSCL 486
HPV29
NCQAKYVKDACTMCRHYKRGEQARMSMSEWIRFRSNKVQGEGDWKPIVHFLRY
QNVEFIPFLCAFKLFLQGIPKKSCL 484
HPV3
NCQAKYVKDACTMCRHYKRGEQARMNMSEWIKFRGDKIQGDGDWKPIVQYLR
YQDVEFIPFLCALKSFLQGIPKKSCI 483
HPV77
NCQAKYVRDACTMCRHYKRGEQARMTMSEWIKFRSDKIQGDGNWKPIVQYLRY
QDVEFVPFLCALKSFLQGIPKKSCL 485
HPV61
NCQAKYIKDAITMCKHYRRAEQAKMTMAQWITHRGRKVADTGDWKAIVKYLR
YQQVEFVPFISALKLFLKGVPKKSCM 472
HPV2a
NCQAKYVKDAVTMCRHYKRAEREQMSMSQWITFRGNKVSEEGDWKPIVRFLRH
QGVEFVSFLAAFKLFLKGVPKKNCI 467
HPV27
NCQAKYVKDAVTMCRHYKRAERAQMSMSQWITFRGNKVLEEGDWKPIVKFLRH
QGVEFVSFLAAFKLFLKGVPKKNCI 467
HPV57
NCQAKYLKDAVTMCRHYKRAEREQMSMSQWITFRGSKISEEGDWKPIVKFLRHQ
GVEFVSFLAAFKSFLKGVPKKNCI 467
HPV26
NCQAKYVKDCATMTRHYKRAQKRSMCMSQWLQYRCSKIEEGGSWKEIAKFLRF
QHVNFIYFLQVLKQFLKGTPKHNCI 463
HPV51
NCQAKYVKDCGTMARHYKRAQRKSLSMSAWIRYRCDRAKDGGNWREIAKFLR
YQGVNFMSFIQMFKQFLKGTPKHNCI 459
HPV30
NMQAKYVKDCGIMCRHYKRAQQQQMNMKQWITHICSKVDEGGDWRPIVQFLR
YQGVDFISFLSYFKLFLRGTPKHNCL 455
HPV53
NMQAKYVKDCGIMCRHYKRAQQQQMNMKQWIKHVCSKVDDGGDWKPIVQFL
RYQGVEFISFLSYFKLFLQGTPKHNCL 460
HPV56

NMQAKYVKDCGIMCRHYKRAQQQQMNMCQWIKHICSKTDEGGDWKPIVQFLR
YQGVDFISFLSYFKLFLQGTPKHNCL 460
HPV66
NMQAKYVKDCGIMCRHYKRAQQQQMNMCQWIKHICSKVDEGGDWKPIVQFLR
YQGVDFISFLSYFKLFLQGTPKHNCL 454
HPV18R
NCQAKYLKDCATMCKHYRRAQKRQMNMSQWIRFRCSKIDEGGDWRPIVQFLRY
QQIEFITFLGALKSFLKGTPKKNCL 480
HPV39
NCQAKYVKDCATMCKHYKRAQKRQMSMSQWIKFRCSKCDEGGDWRPIVQFLR
YQGIEFISFLCALKEFLKGTPKKNCI 470
HPV45
NCQAKYLKDCAVMCRHYKRAQKRQMNMSQWIKYRCSKIDEGGDWRPIVQFLRY
QGVEFISFLRALKEFLKGTPKKNCI 466
HPV59
NCQAKYLKDCAVMCRHYKRAQKRQMSMSQWIKWRCDKIEEGGDWKPIVQFLR
YQGVEFITFLCALKDFLKGTPKRNCI 467
HPV70
NCQAKYVRDCATMCRHYKRAQKKQMTMAQWIRFRCDKCDDGGDWRPIVQFLR
YQGVEFITFLCAFKEFLKGTPKKNCI 475
HPV40
NNQAKYVRDCASMCKHYRLAEMRRMSMAEWIKHRGEKC.DEGDWKPIVKLLRY
QHIDIIVFLAALKKWLQGIPKKNCI 473
HPV7
NNQAKYVRDCAAMCKHYRLAEMRRMSMADWIKHRGEKC.DEGDWKPIVKLLR
YQHIDIIVFLAALKKWLHGIPKKNCI 472
HPV16R
NSQAKIVKDCATMCRHYKRAEKKQMSMSQWIKYRCDRVDDGGDWKQIVMFLR
YQGVEFMSFLTALKRFLQGIPKKNCI 473
HPV31
NSQAKIVKDCGTMCRHYKRAEKRQMSMGQWIKSRCDKVSDEGDWRDIVKFLRY
QQIEFVSFLSALKLFLKGVPKKNCI 453
HPV33
NSQAKIVKDCGIMCRHYKKAEKRKMSIGQWIQSRCEKTNDGGNWRPIVQLLRYQ
NIEFTAFLGAFKKFLKGIPKKSCM 466
HPV35h
NSQAKIVKDCATMCRHYKRAEKREMTMSQWIKRRCEKVDDDGDWRDIVRFLRY
QQVDFVAFLSALKNFLHGVPKKNCI 459
HPV52
NSQAKIVKDCATMCRHYKRAERKHMNIGQWIQYRCDRIDDGGDWRPIVRFLRYQ
DIEFTAFLDAFKKFLKGIPKKNCL 469
HPV58
NAQAKIVKDCGVMCRHYKRAEKRGMTMGQWIQSRCEKTNDGGNWRPIVQFLRY
QNIEFTAFLVAFKQFLQGVPKKSCM 466
HPV67
NSQAKIVKDCGTMCRHYKRAEKRKMTIGQWIQARCEKVNDGGDWRTIVKLLRY
QNVEFTQFLATFKKFLKGIPKKSCM 458

HPV54

LLYGPPDTGKSNFAMSLISFLGGVVLSYVNSSSHFWLEPLADAKIAMLDDATTQC
WNYMDIYMRNALDGNPMCFDRKH 534

HPV32
VLIGPPNTGKSQFGMSLVKFLAGTVISFVNSHSHFWLQPLDSAKIAMLDDATPPCW
TYLDTYLRNLLDGNPCSIDRKH 544

HPV42
VLIGPPNTGKSQFGMSLINFLAGTVISFVNSHSHFWLQPLDSAKIAMLDDATPPCW
TYLDIYLRNLLDGNPCSIDRKH 545

CPV1
AIVGPPDTGKSAFCMSLIKFLGGTVISYVNSSSHFWLQPLCNAKVALLDDATQSCW
GYMDTYMRNLLDGNPMSIDRKH 551

HPV11R
AIVGPPDTGKSCFCMSLIKFLGGTVISYVNSCSHFWLQPLTDAKVALLDDATQPCW
TYMDTYMRNLLDGNPMSIDRKH 552

HPV13
AIVGPPDTGKSCFCMSLIKFLGGTVISYVNSSSHFWLQPLCNAKVALLDDATQSCW
VYMDTYMRNLLDGNPMSIDRKH 549

HPV44
AIVGPPDTGKSCFCMSLIKFLGGTVISYVNSSSHFWLQPLCNAKVALLDDVTQSCW
VYMDTYMRNLLDGNPMTIDRKH 546

HPV55
AIVGPPDTGKSCFCMSLIKFLGGTVISYVNSSSHFWLQPLCNAKVALLDDVTQSCW
VYMDTYMRNLLDGNPMTIDRKH 547

HPV6bR
AIVGPPDTGKSYFCMSLISFLGGTVISHVNSSSHFWLQPLVDAKVALLDDATQPCW
IYMDTYMRNLLDGNPMSIDRKH 552

HPV34
VIYGPPDTGKSHFGMSLMQFMQGVVISYVNSNSHFWLSPLADAKMALLDDATPA
CWTYIDRYLRNALDGNPMCLDRKH 549

HPV73
VLYGPPDTGKSHFGMSLIKFIQGVVISYVNSTSHFWLSPLADAKMALLDDATPGC
WTYIDKYLRNALDGNPICLDRKH 552

RHPV1R
VLFGPPNTGKSYFGMSLIHFLQGSIISYVNSNSHFWLQPLADAKVAMLDDATPQC
WSYIDNYLRNALDGNPISVDRKH 527

HPV10
VFYGPADTGKSYFCMSLLRFLGGAVISYANSSSHFWLQPLSEAKIGLLDDATSQC
WNYIDTYLRNALDGNQICVDRKH 583

HPV28
VFYGPADTGKSYFCMSLLRFLGGVVISYANSNSHFWLQPLADAKIGLLDDATSQC
WCYIDTYLRNALDGNQVCIDRKH 564

HPV29
VFYGPADTGKSYFCMSLLKFMGGVVISYANSHSHFWLQPLSEAKMGLLDDATSQ
CWSYVDTYLRNALDGNVMCIDRKH 562

HPV3
VFYGPADTGKSYFCMSLLKFLGGVVISYANSSSHFWLQPLAEAKIGLLDDATSQC
WCYIDTYLRNALDGNQVCIDRKH 561

HPV77
VFYGPADTGKSYFCMSLLRFMGGAVISYANSTSHFWLQPLSEAKMGLLDDATSQ

CWNYIDTYLRNALDGNVICIDRKH 563

HPV61
VFYGPSDTGKSLFCMSLLNFLGGAVISYVNSSSHFWLSPLADTKVGLLDDATYQC
WQYIDTYLRTVLDGNAISIDRKH 550

HPV2a
VFYGPADTGKSYFCMSLLQFLGGAVISYANSSSHFWLQPLSDSKIGLLDDATPQC
WSYIDIYLRNLLDGHPVSIDRKH 545

HPV27
VFYGPADTGKSYFCMSLLQFLGGAVISYANSSSHFWLQPLSDSKIGLLDDATPQC
WSYIDTYLRNLLDGNPVSIDRKH 545

HPV57
VFYGPADTGKSYFCMSLLQFLGGAVISYANSSSHFWLQPLADSKIGLLDDATAQC
WTYIDTYLRNLLDGNPFSIDRKH 545

HPV26
VIYGPPNTGKSQFAMSFIKFMQGSVISYVNSNSHFWLQPLEDAKVAVLDDATYSC
WLYIDKYLRNFLDGNPCCIDRKH 541

HPV51
VIYGPPNTGKSLFAMSLMKFMQGSIISYVNSGSHFWLQPLEDAKIALLDDATYGC
WTYIDQYLRNFLDGNPCSIDRKH 537

HPV30
VLYGPPNTGKSCFAMSLIQFFQGSVISYVNSHSHFWLQPLDNAKLGMLDDATDAC
WRYIDEYMRNLLDGNPVSLDRKH 533

HPV53
VIYGPPNTGKSCFAMSLINFFHGSVISYVNSHSHFWLQPLDNTKLGMLDDATEAC
WKYIDEYLRNLLDGNPVSLDRKH 538

HPV56
VLCGPPNTGKSCFAMSLIKFFQGSVISFVNSQSHFWLQPLDNAKLGLLDDATEICW
KYIDDYLRNLVDGNPISLDRKH 538

HPV66
VLCGPPNTGKSCFAMSLINFFQGSVISFVNSQSHFWLQPLDNAKLGLLDDATDTC
WRYIDDYLRNLLDGNPISLDRKH 532

HPV18R
VFCGPANTGKSYFGMSFIHFIQGAVISFVNSTSHFWLEPLTDTKVAMLDDATTTCW
TYFDTYMRNALDGNPISIDRKH 558

HPV39
VIYGPANTGKSHFCMSLMHFLQGTVISYVNSTSHFWLEPLADAKLAMLDDATGTC
WSYFDNYMRNALDGYAISLDRKY 548

HPV45
LLYGPANTGKSYFGMSFIHFLQGAIISFVNSNSHFWLEPLADTKVAMLDDATHTC
WTYFDNYMRNALDGNPISIDRKH 544

HPV59
VLCGPANTGKSYFGMSLLHFLQGTVISHVNSNSHFWLEPLTDRKLAMLDDATDSC
WTYFDTYMRNALDGNPISVDRKH 545

HPV70
VIQGPPNTGKSYFCMSLMHFLQGTVISYVNSTSHFWLEPLADAKVAMLDDATGTC
WSYFDTYMRNALDGNPISLDRKH 553

HPV40
CIVGPPDTGKSCFGMSLMHFMQGTIISYVNSCSHFWLQSLADAKVAMLDDVTAA
CWGYMDTHMRNLLDGNPTSIDRKH 551

PCT/US05/11740

HPV7
CIVGPPDTGKSCFGMSLMHFLQGTIISFVNSCSHFWLQSLVDAKVAMLDDVTSAC
WAYMDTHMRNLLDGNPTSIDRKH 550
HPV16R
LLYGAANTGKSLFGMSLMKFLQGSVICFVNSKSHFWLQPLADAKIGMLDDATVPC
WNYIDDNLRNALDGNLVSMDVKH 551
HPV31
LIHGAPNTGKSYFGMSLISFLQGCIISYANSKSHFWLQPLADAKIGMLDDATTPCW
HYIDNYLRNALDGNPVSIDVKH 531
HPV33
LICGPANTGKSYFGMSLIQFLKGCVISCVNSKSHFWLQPLSDAKIGMIDDVTPISWT
YIDDYMRNALDGNEISIDVKH 544
HPV35h
LIYGAPNTGKSLFGMSLMHFLQGAIISYVNSKSHFWLQPLYDAKIAMLDDATSPC
WAYIDQYLRNALDGNPISLDVKH 537
HPV52
VLYGPANTGKSYFGMSLIRFLSGCVISYVNSKSHFWLQPLTDAKVGMIDDVTPICW
TYIDDYMRNALDGNDISVDVKH 547
HPV58
LLCGPANTGKSYFGMSLIHFLKGCIISYVNSKSHFWLQPLSDAKLGMIDDVTAISW
TYIDDYMRNALDGNDISIDVKH 544
HPV67
VICGPPNTGKTYFAMSLIHFLQGCVISYVNAKSHFWLQPLSDAKIGMIDDVTAICW
TYIDDYLRNALDGNDISIDVKH 536

HPV54
RAMVQTKCPPLIVTSNINASTDDRWRYLHSRVKCFCFPNRFPFDSNGNPVYDLSN
KNWKSFFKRSWSRLALN..DNDN 610
HPV32
KALTVVKCPPLIITSNTDIRTEDRWKYLYSRISLFEFPNPFPLDKNGNPVYYVLNDEN
WKSFFQRLWSSLEFQ...ESE 619
HPV42
KALTVVKCPPLLITSNTDIRTNDKWKYLYSRVSLFEFPNPFPLDTNGNPVYELNDK
NWKSFFQRLWSSLEFQ...ESE 620
CPV1
KSLALIKCPPLLVTSNIDITTEERYKYLYSRVTLFKFPNPFPFDSNGNAVYELCDAN
WKCFFARLSASLDI....QDS 625
HPV11R
RALTLIKCPPLLVTSNIDISKEEKYKYLHSRVTTFTFPNPFPFDRNGNAVYELSDAN
WKCFFERLSSSLDI....EDS 626
HPV13
KSLALIKCPPLLVTSNVDITKDDKYKYLYSRVTTLTFPNPFPFDRNGNAVYELSDA
NWKCFFTRLSASLDI....QDS 623
HPV44
KSLALIKCPPLIVTSNIDITKEEKYKYLCSRVTLFTFPNPFPFDRNGNALYDLCETNW
KCFFARLSSSLDI....QTS 620
HPV55
KSLALIKCPPLIVTSNIDITKEDKYKYLCSRVTLFTFPNPFPFDRNGNALYDLCESNW
KCFFARLSTSLDI....QTS 621

PCT/US05/11740

HPV6bR
KALTLIKCPPLLVTSNIDITKEDKYKYLHTRVTTFTFPNPFPFDRNGNAVYELSNTN
WKCFFERLSSSLDI....QDS 626
HPV34
KHLLQIKCPPLLITSNTNPKADDTWKYLHSRMKVFTFSNPFPFDSNGNPLYQLTNE
NWKAFFTKTWSKLDLT...EDD 624
HPV73
KNLLQVKCPPLLITSNTNPKADDTWKYLHSRIKVFTFLNPFPFDSNGNPLYQLTNE
NWKAFFTKTWSKLDLT...EDD 627
RHPV1R
KNLVQMKCPPLLITSNTNAGQDDRWMYLHSRMVVFTFEQPFPFDQNGNPVYELN
DKNWKSFFSRTWSRLDLQ...EEE 602
HPV10
RALLQLKCPPLLITTNINPLTDERWKFLRSRLQLFTFKNPFPVTTQGEPMYTLNDQN
WKCFFRRLWARLSLT...DPE 658
HPV28
RALLQLKCPPLLITTNINPLEDDRWKYLRSRVQLFTFKNKFPLTTQGEPLYTLNDQ
NWKCFFRRLWARLSLT...DPD 639
HPV29
RSLLQLKCPPLLITTNVNPLEDDRWKYLRSRLQVFTFSNPCPLTSKGEPVYTLNDQ
NWKSFFQRLWARLSLT...DPD 637
HPV3
RALLQLKCPPLLITTNINPLGDERWKYLRSRLQVFTFNNKFPLTTQGEPLYTLNDQ
NWKSFFQRLWARLNLT...DPE 636
HPV77
RSLLQLKCPPLLITTNVNPLEDERWKYLRSRLQVFTFKNKFPVTSSGDPLYTLNDQ
NWKSFFQRLWARLRLT...DPD 638
HPV61
RNLTQLKCPPLMITTNINPLEDPTFKYLHSRIVVFQFLHKCPLNSNGDPVYTLNNEN
WKSFFRRSWARIEGSDQQEEE 628
HPV2a
KTLLQLKCPPLMITTNTNPLEEDRWKYLRSRLTVFTFKNPFPFASPGEPLYPINNAN
WKCFFQRSWSRLDLN...SPE 620
HPV27
KTLLQLKCPPLMITTNINPLEEDRWKYLRSRLTLFTFNNPFPFASPGEPLYPINNAN
WKCFFQRSWSRLDLN...SPE 620
HPV57
KTLLQIKCPPLMITTNINPLEEDRWKYLRSRVTLFKFTNPFPFASPGEPLYPINNAN
WKCFFQRSWSRLDLN...SPE 620
HPV26
RSLLQVTCPPLIITSNINPQEDNSLLYLHSRVTVIPFPNTFPFDSNGNPVYALTDVNW
KSFFSTTWSRLDL....EED 615
HPV51
RSLIQLVCPPLLITSNINPQEDANLMYLHTRVTVLKFLNTFPFDNNGNAVYTLNDE
NWKNFFSTTWSRLDL....EEE 611
HPV30
KQLVQIKCPPVIITTNINPLHDAKLQYLHSRIHVVPFLNPFPIDTNGNPVYQLNNVN
WKCFFERTWSRLDLN...NDE 608
HPV53

KQLVQIKCPPVEITTNINPMQDAKLRYLHSRIHVLQFLNPFPIDVNGNPVYQLNNA
NWKCFFERTWSRLDLD...NDE 613
HPV56
KQLVQIKCPPLLITTNINPMLDAKLRYLHSRMLVFQFQNPFPLDNNGNPVYELSNV
NWKCFFTRTWSRLNLD...NDE 613
HPV66
KQLVQIKCPPVIITTNVNPMQDAKLRYLHSRISVFKFENPFPLDNNGNPVYELSNVN
WKCFFERTWSRLNLD...NDE 607
HPV18R
KPLIQLKCPPILLTTNIHPAKDNRWPYLESRITVFEFPNAFPFDKNGNPVYEINDKN
WKCFFERTWSRLDLH...EEE 633
HPV39
KSLLQMKCPPLLITSNTNPVEDDRWPYLRSRLTVFKFPNAFPFDQNRNPVYTINDK
NWKCFFEKTWCRLDLQ...QDE 623
HPV45
KPLLQLKCPPILLTSNIDPAKDNKWPYLESRVTVFTFPHAFPFDKNGNPVYEINDKN
WKCFFERTWSRLDLH...EDD 619
HPV59
RHLVQIKCPPMLITSNTNPVTDNRWPYLNSRLMVFKFPNKLPFDKNRNPVYTINDR
NWKCFFERTWCRLDLN...EEE 620
HPV70
RHLIQIKCPPILITSNTNPVEENRWPYLTSRLTVFTFPNAFPFDQNRNPVYTINNKNW
KSFFQKTWCKLDLQ...QDE 628
HPV40
KPLAVIKCPPLLLTSNINITQDSKYQYLQSRVQVFEFPNPFPFDSNGNAVYELNDAN
WNSFFKRLASSLEL....QTP 625
HPV7
KSLAVIKCPPLLLTSNINIKHDCKYQYLQSRVTVFEFPNPFPFDSNGNAVYELSDAN
WNSFFKRLASSLEL....QTT 624
HPV16R
RPLVQLKCPPLLITSNINAGTDSRWPYLHNRLVVFTFPNEFPFDENGNPVYELNDK
NWKSFFSRTWSRLSLH...EDE 626
HPV31
KALMQLKCPPLLITSNINAGKDDRWPYLHSRLVVFTFPNPFPFDKNGNPVYELSDK
NWKSFFSRTWCRLNLH...EEE 606
HPV33
RALVQLKCPPLLLTSNTNAGTDSRWPYLHSRLTVFEFKNPFPFDENGNPVYAINDE
NWKSFFSRTWCKLDLI...EEE 619
HPV35h
KALVQLKCPPLLITSNINAGKDDRWPYLHSRVVVFTFHNEFPFDKNGNPVYGLND
KNWKSFFSRTWCRLNLH...EEE 612
HPV52
RALVQIKCPPLILTTNTNAGTDPRWPYLHSRLVVFHFKNPFPFDENGNPIYEINNEN
WKSFFSRTWCKLDLI...QEE 622
HPV58
RALVQLKCPPLIITSNTNAGKDSRWPYLHSRLTVFEFNNPFPFDANGNPVYKINDE
NWKSFFSRTWCKLGLI...EEE 619
HPV67
KALVQLKCPPLLLTSNIDVATDSRWPFLHSRVVVFRFNNPFPFDENGNPVYNLNDE

NWKSFTSRTWCQLDLL..EEE 611

HPV54 EEEE.NGDPSNTFRCVPGKASRPI.. 633
HPV32 DEEE.NGDTGQTFRCVPGTVVRTV.. 642
HPV42 DEED.YGETGQTFRCVPGTVVRTV.. 643
CPV1 EDED.DGDTSQAFRCVPGTVVRTV.. 648
HPV11R EDEE.DGSNSQAFRCVPGSVVRTL.. 649
HPV13 EDED.DGDNSQAFRCVPGTVVRTV.. 646
HPV44 EDED.DGDNSQAFRCVPGTVVRTV.. 643
HPV55 EDED.DGDNSQAFRCVPGTVVRTV.. 644
HPV6bR EDEE.DGSNSQAFRCVPGTVVRTL.. 649
HPV34 DKEN.DGDTVQTFKCVSGRNPRTV.. 647
HPV73 DKEN.DGDTVQTFKCVSGRNPRTV.. 650
RHPV1R ETEN.DGSTCRAFKCVAGQNLRTV.. 625
HPV10 DEEE.HGNPSEPFRCVPGQNARTI.. 681
HPV28 DEEE.NGNPSEPFRCVPGQNARTL.. 662
HPV29 DEEE.NGEPSEPFRCVPGQNTRTV.. 660
HPV3 DEED.NGNTSEPFRCVPGQNTRTV.. 659
HPV77 DEEE.NGEPSEPFRCVPGQNARTL.. 661
HPV61 EEEDEDGVTSRPFRCVPGEITRPL.. 652
HPV2a EQDD.NGNTGEPFRCVPGDVARTV.. 643
HPV27 EQDD.NGNTSEPFRCVPGDVARTL.. 643
HPV57 DQED.NGNTGEPFRCVPGDVARTV.. 643
HPV26 ADKE.NGEPLPAFKCVPGENTRLL.. 638
HPV51 EDKE.NGDPMPPFKCVPGENTRLL.. 634
HPV30 DKEN.HGDSMPTFRCVPGENSRLF.. 631
HPV53 DKEN.DGDAMPTFRCVPGENSRLF.. 636
HPV56 DKEN.NGDAFPTFKCVPEQNTRLF.. 636
HPV66 DKEN.NGDSIPTFRCVPEQNTRLL.. 630
HPV18R EDADTEGNPFGTFKCVAGQNHRPL.. 657
HPV39 DEGDNDENTFTTFKCVTGQNTRIL.. 647
HPV45 EDADTEGIPFGTFKCVTGQNTRPL.. 643
HPV59 EDADSDGHPFAAFKCVTGSNIRTL.. 644
HPV70 DEGDNDGNTIPTFKCVTGENTRTL.. 652
HPV40 GDE..DGESSQAPRFVPGTVVRTV.. 647
HPV7 EDE..DGETSQAPRFVPGTVVRTL.. 646
HPV16R DKEN.DGDSLPTFKCVSGQNTNTL.. 649
HPV31 DKEN.DGDSFSTFKCVSGQNIRTL.. 629
HPV33 DKEN.HGGNISTFKCSAGENTRSLRS 644
HPV35h DKEN.DGDAFPAFKCVSGQNTRTLRD 637
HPV52 DKEN.DGVDTGTFKCSAGKNTRSIRS 647
HPV58 DKEN.DGGNISTFKCSAGQNPRHIRS 644
HPV67 DKEN.HGGNCNTFKCSAGENSRCIRS 636

Human papillomaviruses (HPV) have been well documented. Mucosotropic HPV types are divided into "high risk" and "low risk" categories based on their potentials to cause cancers

of the anogenital tract of both males and females. HPV-6 and HPV-11, for instance, cause most cases of genital warts; they are considered "low risk" because they rarely lead to cancer. Other HPV strains, such as HPV 16, 31, 33, 35, 52 and 58 as well as HPV-18, 39, 45, and 70 are considered "high risk" because they have been linked with an increased risk for cervical, vaginal and anal cancers. Other examples of HPV strains include, but are not limited to HPV 2a, 3, 5, 6b, 7, 10, 11, 13, 16, 18, 26, 27, 28, 29, 30, 32, 34, 40, 44, 45, 51, 53, 55, 56, 57, 59, 61, 66, 68, 70, 72, 73, 74 and 83. Also contemplated herein are animal papillomaviruses.

Administration

The agents, substances, compounds, and compositions disclosed herein can be administered *in vivo* in a pharmaceutically acceptable carrier. By "pharmaceutically acceptable" is meant a material that is not biologically or otherwise undesirable, i.e., the material may be administered to a subject without causing undue undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition in which it is contained. The carrier would naturally be selected to minimize any degradation of the active ingredient and to minimize any adverse side effects in the subject, as would be well known to one of skill in the art.

The agents disclosed herein are of benefit to subjects who are experiencing viral infection or are at risk for infection Because the agents disclosed herein reduce the activity of E1, thereby reducing the severity or duration of the infection, any subject that can benefit from a reduction in the activity of E1 can be administered the agents disclosed herein.

The compositions comprising an agent disclosed herein in a pharmaceutically acceptable carrier may be administered orally, parenterally (e.g., intravenously), by intramuscular injection, by intraperitoneal injection, transdermally, extracorporeally, topically or the like, although topical intranasal administration or administration by inhalant is typically preferred. The exact amount of the compositions required will vary from subject to subject, depending on the species, age, weight and general condition of the subject, the severity of the disorder being treated, the particular nucleic acid or vector used, its mode of administration and the like. Thus, it is not possible to specify an exact amount for every composition. However, an appropriate amount can be determined by one of ordinary skill in the art using only routine experimentation given the teachings herein.

Parenteral administration of the composition, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution of suspension in liquid prior to injection, or as emulsions. A more recently revised approach for parenteral administration involves use of a slow release or sustained release system such that a constant dosage is maintained. See, e.g., U.S. Patent No. 3,610,795, which is incorporated by reference herein in its entirety for the methods taught.

The compositions may be in solution or in suspension (for example, incorporated into microparticles, liposomes, or cells). These compositions may be targeted to a particular cell type via antibodies, receptors, or receptor ligands. The following references are examples of the use of this technology to target specific proteins to given tissue (Senter, et al., Bioconjugate Chem., 2:447-451, (1991); Bagshawe, K.D., Br. J. Cancer, 60:275-281, (1989); Bagshawe, et al., Br. J. Cancer, 58:700-703, (1988); Senter, et al., Bioconjugate Chem., 4:3-9, (1993); Battelli, et al., Cancer Immunol. Immunother., 35:421-425, (1992); Pietersz and McKenzie, Immunolog. Reviews, 129:57-80, (1992); and Roffler, et al., Biochem. Pharmacol, 42:2062-2065, (1991)). Vehicles such as "stealth" and other antibody conjugated liposomes (including lipid mediated drug targeting to colonic carcinoma), receptor mediated targeting of DNA through cell specific ligands, lymphocyte directed tumor targeting, and highly specific therapeutic retroviral targeting of murine glioma cells *in vivo*. In general, receptors are involved in pathways of endocytosis, either constitutive or ligand induced. These receptors cluster in clathrin-coated pits, enter the cell via clathrin-coated vesicles, pass through an acidified endosome in which the receptors are sorted, and then either recycle to the cell surface, become stored intracellularly, or are degraded in lysosomes. The internalization pathways serve a variety of functions, such as nutrient uptake, removal of activated proteins, clearance of macromolecules, opportunistic entry of viruses and toxins, dissociation and degradation of ligand, and receptor-level regulation. Many receptors follow more than one intracellular pathway, depending on the cell type, receptor concentration, type of ligand, ligand valency, and ligand concentration. Molecular and cellular mechanisms of receptor-mediated endocytosis has been reviewed (Brown and Greene, DNA and Cell Biology 10:6, 399-409 (1991)).

Delivery

The disclosed agents can be delivered to the target cells in a variety of ways. For example, the agents can be delivered through electroporation, or through lipofection, or through calcium phosphate precipitation. The delivery mechanism chosen will depend in part on the type of cell targeted and whether the delivery is occurring for example *in vivo* or *in vitro*.

Thus, the agents can comprise, for example, lipids such as liposomes, such as cationic liposomes (e.g., DOTMA, DOPE, DC-cholesterol) or anionic liposomes. Liposomes can further comprise proteins to facilitate targeting a particular cell, if desired. Administration of a composition comprising a compound and a cationic liposome can be administered to the blood afferent to a target organ or inhaled into the respiratory tract to target cells of the respiratory tract. Regarding liposomes, see, e.g., Brigham et al. Am. J. Resp. Cell. Mol. Biol. 1:95-100 (1989); Felgner et al. Proc. Natl. Acad. Sci *USA* 84:7413-7417 (1987); U.S. Pat. No. 4,897,355. Furthermore, the compound can be administered as a component of a microcapsule that can be targeted to specific cell types, such as macrophages, or where the diffusion of the compound or delivery of the compound from the microcapsule is designed for a specific rate or dosage.

The agents of the present invention can also be administered using methods of delivering exogenous nucleic acids, such as in gene therapy. See, e.g., U.S. Patent No. 5,399,346, which is incorporated by reference herein in its entirely for the methods of delivery. Primary cells transfected with the gene for the agent of the present invention can additionally be transfected with tissue specific promoters to target specific organs, tissue, grafts, or cells.

Administration of the agents disclosed herein can occur in conjunction with other therapeutic agents. Thus, the agents of the present invention can be administered alone or in combination with one or more therapeutic agents. For example, a subject can be treated with the disclosed agent alone, or in combination with chemotherapeutic agents, antibodies, antivirals, steroidal and non-steroidal anti-inflammatories, conventional immunotherapeutic agents, cytokines, chemokines, and/or growth factors. Combinations may be administered either concomitantly (e.g., as an admixture), separately but simultaneously (e.g., via separate intravenous lines into the same subject), or sequentially (e.g., one of the compounds or agents is given first followed by the second). Thus, the term "combination" or "combined" is used

to refer to either concomitant, simultaneous, or sequential administration of two or more agents.

Pharmaceutically Acceptable Carriers

Delivery of the agents disclosed herein can be used therapeutically in combination with a pharmaceutically acceptable carrier. Pharmaceutical carriers are known to those skilled in the art. These most typically would be standard carriers for administration of drugs to humans, including solutions such as sterile water, saline, and buffered solutions at physiological pH. The compositions can be administered intramuscularly or subcutaneously. Other compounds will be administered according to standard procedures used by those skilled in the art.

Pharmaceutical compositions may include carriers, thickeners, diluents, buffers, preservatives, surface active agents and the like in addition to the molecule of choice. Pharmaceutical compositions may also include one or more active ingredients such as antimicrobial agents, anti-inflammatory agents, anesthetics, and the like.

The pharmaceutical composition may be administered in a number of ways depending on whether local or systemic treatment is desired, and on the area to be treated. Administration may be topically (including opthamalically, vaginally, rectally, intranasally), orally, by inhalation, or parenterally, for example by intravenous drip, subcutaneous, intraperitoneal or intramuscular injection. The disclosed compounds can be administered intravenously, intraperitoneally, intramuscularly, subcutaneously, intracavity, or transdermally.

Preparations for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like.

Formulations for topical administration may include ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

Compositions for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets, or tablets. Thickeners, flavorings, diluents, emulsifiers, dispersing aids or binders may be desirable.

Some of the compositions may potentially be administered as a pharmaceutically acceptable acid- or base- addition salt, formed by reaction with inorganic acids such as hydrochloric acid, hydrobromic acid, perchloric acid, nitric acid, thiocyanic acid, sulfuric acid, and phosphoric acid, and organic acids such as formic acid, acetic acid, propionic acid, glycolic acid, lactic acid, pyruvic acid, oxalic acid, malonic acid, succinic acid, maleic acid, and fumaric acid, or by reaction with an inorganic base such as sodium hydroxide, ammonium hydroxide, potassium hydroxide, and organic bases such as mono-, di-, trialkyl and aryl amines and substituted ethanolamines.

Therapeutic Uses

The dosage ranges for the administration of the agents disclosed herein are those large enough to produce the desired effect in which the symptoms of the disorder are affected. The dosage should not be so large as to cause adverse side effects, such as unwanted cross-reactions, anaphylactic reactions, and the like. Generally, the dosage will vary with the age, condition, sex and extent of the disease in the patient and can be determined by one of skill in the art. The dosage can be adjusted by the individual physician in the event of any contraindications. Dosage can vary, and can be administered in one or more dose administrations daily, for one or several days.

As described above, the agents disclosed herein can be administered together with other forms of therapy. For example, the molecules can be administered with antibodies, antibiotics, or other cancer treatment protocols as described above, or viral vectors. When the agent is in a vector, as described above, the vector containing the nucleic acid for therapeutic purposes can also contain the agent that modulates E1 activity.

Screening Methods

Disclosed herein are methods of screening for an agent that inhibits nuclear localization of E1, comprising the steps of transfecting a cell with a tagged E1 protein, fragment, or mutant thereof; contacting the cell with the agent to be screened; inducing

expression of the tagged E1 protein, fragment, or mutant thereof; and detecting a cellular location of the tagged E1 protein, fragment, or mutant thereof wherein the absence of E1 protein in the nucleus or its presence in the cytoplasm indicates an agent that inhibits nuclear localization of E1. The E1 protein can be SEQ ID NO: 1, SEQ ID NO: 2, or any fragment or variant thereof.

The method of identifying an E1 inhibitor can be performed *in vitro* in a cell assay, for example. Many reporter proteins are known to one of skill in the art. Fluorescent reporter proteins can also be used, such as green fluorescent protein (GFP), cyan fluorescent protein (CFP), red fluorescent protein (RFP), yellow fluorescent protein (YFP). Other examples include the green fluorescent protein from Aequorea coerelescens (AcGFP), DsRedExpress, and red coral fluorescent proteins (for example, AmCyan, ZsGreen, ZsYellow, AsRed2, DsRed2, and HcRed1). For example, by utilizing GFP, fluorescence is observed upon exposure to light at 489 nm without the addition of a substrate. The use of a reporter protein that, like GFP, are directly detectable without requiring the addition of exogenous factors are preferred for detecting or assessing gene expression or location. Fluorescent proteins can be isolated from many different species, including but not limited to, Aequorea victoria (Chalfie, et al., 1994), Zoanthus species (Matz, et al., 1999), and Renilla reniformis (Ward and Cormier, 1979).

The concept of the assay is to monitor the subcellular location of the E1 helicase in a high-throughput assay. In one example, a cell line containing a tetracycline-inducible GFP-E1 protein can be induced in all cells growing on the (bottom of the well) window of optical microtiter plates (eg. Corning Inc, Costar #3904 96 well flat bottom tissue culture-treated black plates with optically clear bottoms). The use of a fluorescent fusion protein which is directly detectable without requiring the addition of exogenous factors is preferred for detecting or assessing gene expression or location. For example, by utilizing GFP-tagged E1, fluorescence is observed upon exposure to light at 489 nm without the addition of a substrate.

In the absence of inhibitors, essentially all E1 protein will be nuclear. The nuclei can be marked with various DNA intercalating stains such as DAPI and various Hoechst dyes. If the nuclear dye (blue, in the case of DAPI and red in the case of Hoechst) is overlaid with green (from GFP-E1), then there are comparable signals in both color channels when viewed by charge-cooled device (CCD) digital cameras or comparable recording instrumentation. Conversely, if the presence of the test compound results in the largely or completely

cytoplasmic localization of the GFP-E1 protein, then the nuclear dye (eg. blue or red) gives a relatively pure to pure signal, without GFP overlay. Initial lead compounds may generate a mixed pattern with E1 in both compartments. Further modifications of the lead compounds can result in an largely or exclusive cytoplasmic GFP-E1. One skilled in the art will appreciate that it is desirable to screen test compounds using several different concentrations of the test compound. Computer-based image analysis software is known to those of skill in the art.

E1 can also be fused to other fluorescence proteins, such as cyan fluorescent protein (CFP), red fluorescent protein (RFP), yellow fluorescent protein (YFP) and blue fluorescent protein (BFP). Other examples include the green fluorescent protein from Aequorea coerelescens (AcGFP), DsRedExpress, and red coral fluorescent proteins (for example, AmCyan, ZsGreen, ZsYellow, AsRed2, DsRed2, and HcRed1). For example, fluorescent proteins can be isolated from many different species, including but not limited to, Aequorea victoria (Chalfie, et al., 1994), Zoanthus species (Matz, et al., 1999), and Renilla reniformis (Ward and Cormier, 1979).

High throughput screening methods can be used with the screening methods disclosed herein. Several commercial luminescence and fluorescence detectors are available that can simultaneously inject liquid into single or multiple wells such as the WALLAC VICTOR2 (single well), MICROBETA RTM JET (six wells), or AURORA VIPR (eight wells). Typically, these instruments require 12 to 96 minutes to read a 96-well plate in flash luminescence or fluorescence mode (1 min/well). An alternative method is to inject the test substance into all sample wells at the same time and measure the luminescence in the whole plate by imaging with a CCD camera, similar to the way that calcium responses are read by calcium-sensitive fluorescent dyes in the FLIPR or FLIPR-384 instruments. Other luminescence or fluorescence imaging systems include LEADSEEKER from AMERSHAM, the WALLAC VIEWLUX TM ultraHTS microplate imager, and the MOLECULAR DEVICES CLIPR imager. PE BIOSYSTEMS TROPIX produces a CCD-based luminometer, the NORTHSTAR TM HTS Workstation. This instrument is able to rapidly dispense liquid into 96-well or 384-well microtiter plates by an external 8 or 16-head dispenser and then can quickly transfer the plate to a CCD camera that images the whole plate. The total time for dispensing liquid into a plate and transferring it into the reader is about 10 seconds. Other

systems and methods for high throughput screening are known in the art and contemplated herein.

Also disclosed are agents identified by the screening methods disclosed herein. The agent can follow the guidelines of "Lipinski's Rule of Five." (Lipinski, 1997). Lipinski's Rule of Five is particularly useful when the goals of compound design are (i) to have less than 5 hydrogen donors, (ii) less than 10 hydrogen bond acceptors, (iii) molecular weight of less than 500 Daltons and (iv) the log of the partition coefficient, P (where P = the concentration of the compound in water divided by the concentration of the compound in 1 octanol) is less than 5. The Lipinski Rule of Five is a useful guideline, however, the composition is not limited to these parameters.

A wide variety of small molecular weight compounds can be used in the screening methods disclosed herein. Such compounds include, but are not limited to, any compositions which are being tested for drug discovery or development. Such compounds include, but are not limited to, nucleic acids including functional nucleic acids, amino acids including peptides and proteins and fragments thereof, and various other chemical compounds. Compounds can be aqueous- or lipid-soluble. Compounds can be dissolved or suspended within solution, or affixed to a solid-support. Solid supports may include, but are not limited to, insoluble polymer beads or a polymeric matrix coated with one or a plurality of individual compounds, or with combinatorial chemistries. Dosages and volumes which are administered in the screening methods can be varied so as to optimize dosages for further studies or to rank compounds as to their toxicity and/or potency. Information resulting from variations in conditions can be used to prioritize chemicals for further study, to delineate the relative toxicities of structurally related chemicals, and/or to identify the proper dose range for subsequent toxicity studies (see e.g., Harris, et al., Fundam. Appl. Toxicol. 19:186-196). Most importantly, the known structures of lead compounds can then be varied through (prior or subsequent) combinatorial chemistry to work toward more favorable properties (eg. higher efficacy and specificity, lower toxicity, desired solubilities, optimized bioavailabilities).

With respect to the generation of small molecular weight compound libraries, the combination of biochemical diversity is often synergistic with the metabolic diversity obtained from the *in vivo* production of "natural products". Collections of starting compounds, for example small molecule chemicals, can be administered to cultures of microorganisms. In accord, each microbial strain may potentially create numerous modified

"small molecule chemicals", thus generating a "metabolite library". Because each of these aforementioned cultures can contain a very complex mixture of metabolites, a highly efficacious method of screening is required (i.e., high throughput screening). An aliquot of the library is incubated with the cell lines induced for GFP-E1 protein expression. and the nucleo-cytopolasmic localization of the GFP-E1 protein will be recorded utilizing an HTS-based assay. Furthermore, natural product diversity can be screened by creating a mixture of combinatorially-tagged liposomes; wherein each liposome preferably encapsulates only one member or a simple mixture of a natural product compound library.

Vectors and Cells

Disclosed is an expression vector comprising SEQ ID NO: 1, or a fragment or variant thereof, operatively linked to a tag. Also disclosed is an expression vector comprising SEQ ID NO: 2, or a fragment or variant thereof, operatively linked to a tag. Also disclosed are cells and cell lines comprising the vectors described herein, which can be used with the screening methods disclosed herein. The tag can be a fluorescent tag, such as GFP or RFP, for example, as described above. Also disclosed is a plurality of cells stably transfected with the vectors described herein, wherein the expression of a polypeptide encoded by the vector is inducible. At least about 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the cells can be induced to express the polypeptide.

For example, contemplated is a green fluorescent protein chimeric tag fused to the amino-terminus of an E1 protein. The chimeric protein is active in transient replication assays, and the GFP moiety does not materially affect helicase location or function. Expression of the GFP-tagged helicase can be induced in all cells in a population. This was achieved in two stages: (1) creating a mutant version of E1 in which the dominant splice donor site (surrounding nucleotide 847) was mutated to curtail splicing of the primary transcript, and thereby to dramatically increase the translation of the full-length E1 helicase protein; and (2) driving the expression of GFP-tagged E1 helicase from a tetracyclin-inducible promoter. Cell lines were created to contain an integrated cassette consisting of a regulatable Tet-on promoter driving expression of the GFP-E1.

The GFP-11E1dm open reading frame (ORF) was inserted into the pcDNA4/TO vector and linearized by ScaI restriction endonuclease digestion. The linearized dsDNA was transfected by lipofectamine 2000 into the T-REx-293 or T-REx-COS 7 cell line which

contains the Tet repressor from pcDNA6/TR. Cells were selected by the antibiotic Zeocin, and single colonies were picked up and tested for their characteristics.

E1 helicase is maximally induced by adding as little as 62.5 nanograms/ml of tetracycline to the cell culture media (Figure 23). Lower concentrations may also be used. The induction can be sustained for at least 22 hours of continual tet-treatment (Figure 24). Pulsed, cyclical treatments can also be used. GFP-E1 helicase was induced in all cells, and it was nuclear, providing a highly controlled, robust means for expression of fluorescence-tagged E1 helicase. These tagged E1 proteins can also be made into vectors containing mutations, including but not limited to those in the various regulatory motifs described herein. These serve as both controls for cytoplasmic E1 localization in response to potential inhibitory drugs.

Organotypic epithelial keratinocyte raft cultures are also contemplated (Wilson et al. Cell Growth and Differentiation 3: 471-483; Dollard et al. Genes Development 6: 1131-1142; Chow et al. Clinics Dermatol. 15: 217-227). Complete stratification and differentiation of primary human keratinocytes has been achieved using these rafts. Preparation of organotypic raft cultures comprised of cells with induced (or also transient) expression of tagged E1 protein enables the development of stratified epithelial (eg. mucosal or cutaneous) skin equivalents with non-cycling cells in which the papillomaviral E1 helicase can be induced in the organotypic context, and that anti-E1 drug validation (specificity, selectivity, bioavailability, persistence, and (lack of) toxicity) can include such an organotypic testing environment. The raft cultures enables *ex vivo* validation of efficacy.

Kits

Disclosed herein are kits that are drawn to reagents that can be used in practicing the methods disclosed herein. The kits can include any reagent or combination of reagent discussed herein or that would be understood to be required or beneficial in the practice of the disclosed methods. For example, the kits could include E1, an attenuated virus, and a suitable container. The kit can also include combinatorial libraries of small molecules. The kit can also include a screening platform, such as well plates for screening molecules.

The present invention is more particularly described in the following examples, which are intended as illustrative only since numerous modifications and variations therein will be apparent to those skilled in the art.

Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

Although the present process has been described with reference to specific details of certain embodiments thereof, it is not intended that such details should be regarded as limitations upon the scope of the invention except as and to the extent that they are included in the accompanying claims.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how the compounds, compositions, articles, devices and/or methods claimed herein are made and evaluated, and are intended to be purely exemplary of the invention and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers (e.g., amounts, temperature, etc.), but some errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C or is at ambient temperature, and pressure is at or near atmospheric.

EXAMPLES

Example 1: Methods

For the examples, specific experiments conducted with HPV-11 and its E1 DNA helicase are described. The very high conservation of sequence motifs among HPV E1 proteins shows that all types have similar control over cytoplasmic-nuclear translocation and retention. Assays described and drugs eventually identified based on HPV-11 are fully applicable to the other HPV genotypes.

Plasmids and antibodies. The HPV-11 EE-E1 and E2 expression plasmids pMTX-EE-E1, pMT2-E2, and origin-containing plasmid p7730-99 have been described (Chiang et al. *Proc. Natl. Acad. Sci. USA* 89:5799-5803 (1992), Kuo et al. (1994)). The NES mutation, the phosphorylation site mutations, and the cyclin binding motif mutations were each constructed in the context of pGFP-11E1dm, which encodes a GFP-E1 fusion protein and is mutated at the dominate splice donor site at nt 847 without affecting the encoded amino acids (18). pMTX-EE-E1dm was derived from pMTX-EE-E1 by incorporating the E1dm mutation just described. pMTX EE-E1 S107A and pMTX EE-E1 S107A-NEm were

then prepared in pMTX-EE-E1dm. Oligonucleotide-directed mutagenesis was performed using standard PCR method. All PCR-amplified fragments were confirmed by DNA sequencing.

The anti-E2 polyclonal antibody and the anti-EE monoclonal antibody has been described previously (Chiang et al. (1992), Grussenmeyer et al. Proc. Natl. Acad. Sci. U S A. 82:7952-7954 (1985), Kuo et al. (1994)). For immunoprecipitation, EE was crosslinked to CNBr-activated Sepharose 4 FastFlow beads (Amersham Biosciences Corp., Piscataway, NJ) according to manufacturer's instruction. Phospho-Serine/Threonine-Proline specific monoclonal antibody MPM-2 was purchased from Upstate (Lake Placid, NY), the anti-GFP monoclonal antibody JL-8 and polyclonal antibody from Clontech Laboratories (Palo Alto, Ca), anti-mouse IgG-FITC conjugate from Molecular Probes (Eugene, OR) and anti-rabbit IgG-Texas Red conjugate from Vector Laboratories (Burlingame, CA).

Cell culture, transfection, and drug treatment. COS7, 293 and p21-9 (Chang et al. *Proc. Natl. Acad. Sci.* USA 97:4291-4296 (2000) were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum at 37°C and 5% CO_2. Cells were transfected by electroporation as described(Chiang et al. (1992), Zou et al. *J. Virol.* 74:3761-3770 (2000)). Cells were treated with roscovitine (Sigma Co., St. Louis, MO) at 20 μM 4 hrs post-transfection and cultured for another 20 hrs. For p21cip induction, isopropyl-β-thio-galactosidase (IPTG) (Invitrogen Life Technologies, Carlsbad, CA) was added to the culture medium to a final concentration of 50 μM 4 hrs post-transfection and maintained for 20 hrs. Leptomycin B (LMB) (Sigma) at 5 μg/ml was added to the culture medium to a final concentration of 20 ng/ml 24 hrs post-transfection. Cells were then cultured for 6 more hrs for fluorescence microscopy.

Fluorescence microscopy. After electroporation, $1x10^5$ cells were grown on 2-well chamber slides (Nalge Nunc International, Naperville, IL), treated with the chemicals as indicated above. Cells were then fixed at the indicated time point with 4% paraformaldehyde. For GFP-E1 observation, slides were directly overlaid with coverslips using mounting media with DAPI (4',6-diamidino-2-phenylindole) (Vector Laboratories, Inc., Burlingame, CA). Immunostaining of EE-E1 was performed using monoclonal anti-EE antibody followed with anti-mouse IgG-FITC conjugate. Immunostaining of E2 was performed using polyclonal anti-E2 antibody followed with anti-rabbit IgG-Texas Red conjugate. These slides were mounted with DAPI-containing media. Images were

captured and processed with an Olympus AX70 fluorescence microscope with FITC/DAPI filters using an AxioCam digital camera (Carl Zeiss, Inc., Germany).

Protein purification, immunoprecipitation, and protein phosphatase treatment. Purification of *E. coli*-expressed EE-E1 and Sf9-expressed EE-E1 protein were described (Kuo et al. (1994), Ma et al. (1999)). For EE-E1 immunoprecipitation, $5x10^7$ COS7 cells or $1.5x10^8$ 293 cells transfected with pMTX-EE-E1dm plasmid were lysed in 5 ml NETN buffer (20mM Tris, PH 8.0, 100mM NaCl, 1mM EDTA, 5mM NaF, 0.5% NP-40) containing 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1% of Protease Inhibitor Cocktail (Sigma). Lysates were mixed with 200 µl Sepharose 4 beads to which anti-EE antibody was coupled at 4°C overnight, then extensively washed five times with NETN buffer. Proteins were eluted with 400 µl of 0.1 M glycine buffer (PH 3.0) and neutralized with 4 µl of 1.5 M Tris buffer (PH 9.0). The final solution was concentrated using Nanosep Centrifugal columns (Pall Co., Ann Arbor, MI). The concentrations of EE-E1 proteins recovered were determined by western blotting with anti-EE antibody using purified EE-E1 as standards. 50 ng each were used for subsequent analysis. For phosphatase 1 treatment, 2 µl of purified or immunoprecipitated EE-E1 (50 ng) was mixed with 1 unit of protein phosphatase 1 catalytic subunit (Sigma) in reaction buffer (50 mM Tris, pH 7.5, 0.1mM EDTA, 1mM $MnCl_2$) to a final volume of 10 µl. The mixture was incubated at 30°C for an hour. All samples were resolved by electrophoresis through SDS-8% polyacrylamide gels, followed by immunoblotting with individual antibodies and detected by enhanced chemiluminescence (ECL) reagents (Amersham). The MPM-2 antibody was used to detect phosphorylated protein. The membranes were stripped and re-probed with monoclonal anti-EE antibody. GFP-E1 immunoprecipitation from transfected COS7 cells was performed similarly except a polyclonal anti-GFP antibody was used to recover the fusion protein and bound protein was eluted from protein A-Sepharose beads (Amersham) with 2 × SDS-PAGE sample buffer and subject to immunoblotting. MPM-2 was used to detect phosphorylated protein. The membranes were stripped and re-probed with monoclonal anti-GFP antibody (JL-8).

Transient replication assay. Transient replication assays were performed using 293 cells as described (10, 18, 84). In each assay, 0.5 µg of GFP-E1 expression plasmid, 5 µg of pMT2-E2 plasmid, 0.5 µg of HPV-11 origin-containing plasmid p7730-99 were cotransfected into $5x10^6$ cells by electroporation. Cells were harvested 48 hrs post-

transfection and low-molecular-weight DNA was isolated. One half of the recovered DNA was digested with Hind III to linearize the ori plasmid. The other half was digested together with Dpn I, which cut the unreplicated input plasmids into small fragments. Standard Southern blotting was then performed using ^{32}P-α-dCTP labeled probes from the ori plasmid p7730-99. The results were analyzed by PhosphorImager (Molecular Dynamics Inc., Sunnyvale, CA).

Example 2: E1 is phosphorylated by CDKs *in vivo*.

A predominant intragenic splice occurs within the HPV-11 E1 transcript, abolishing E1 expression. Consequently, pGFP-E1, which encodes the green fluorescence protein (GFP) fused to the N-terminus of the E1 sequence, expresses primarily a pancellular GFP variant of slightly higher molecular mass (Deng et al. *J. Virol.* 77:10213-10226 (2003)). A mutation at the donor splice site at nt 847 (E1dm), which does not alter the encoded amino acids, abrogates this splice and pGFP-E1dm encodes abundant, replication-competent GFP-E1 protein (Deng et al. (2003)). Thus, this mutation not only increases the quantity of the E1 protein expressed, it also allows the subcellular localization of the GFP-E1 protein to be followed. All E1 expression vectors described herein contain this donor site mutation.

HPV-11 E1 is phosphorylated by multiple cyclin/CDK complexes *in vitro*, except cyclin D/CDK4 (43). To assess whether the E1 protein is also phosphorylated *in vivo*, HPV-11 E1 tagged at its amino terminus with glu-rich (EE) epitope was expressed in insect Sf9 cells, monkey COS7 cells and human 293 cells. The EE-E1 functions as a potent DNA helicase in vitro and is replication competent in transient replication and cell-free replication assays (Kuo et al. (1994), Lin et al. (2002), Ma et al. (1999)). The protein was purified from Sf9 cells as described (Kuo et al. (1994), Lin et al. (2002), Liu et al. (1995), Liu et al. (1998)), or immunoprecipitated from transfected mammalian cells by using Sepharose beads conjugated to monoclonal antibody to the EE epitope. The recovered protein was then analyzed by SDS-PAGE and western blotted with an antibody specific for phospho-Ser/Thr-Pro resulting from phosphorylation by CDK kinases. Of several such antibodies tested, only MPM-2 specifically recognizes EE-E1. Fig. 1B shows that this antibody detected EE-E1 protein expressed in insect and mammalian cells (lanes 3, 5, 8). However, if the EE-E1 was first treated with protein phosphatase 1 (PP1), a serine/threonine phosphatase, reactivity with the MPM-2 antibody was abolished (Fig. 1B, lane 4, 6, 9) while reactivity with the anti-EE antibody was unchanged (Fig. 1B, lanes 3-6,

8, 9). EE-E1 purified from *E. coli* was only detected by the anti-EE antibody, not by MPM-2 antibody with or without PP1 treatment (Fig. 1B, lane 1 and 2). These results demonstrate that the HPV-11 E1 is indeed phosphorylated in eukaryotic cells.

To substantiate this conclusion, COS7 cells expressing EE-E1 were treated for 20 hours with 20 μM roscovitine. Roscovitine is a purine derivative that reversibly competes for ATP binding and specifically inhibits the activity of CDK1 and CDK2, but has no effect on CDK4 and CDK6 (De Azevedo et al. *Eur. J. Biochem.* 243:518-526 (1997), Meijer et al. *Eur. J. Biochem.* 243:527-536 (1997)), nor on the MAP kinases (Bain et al. *Biochem. J.* 371:199-204 (2003)), which have the same consensus substrate sites as CDKs. Upon treatment with roscovitine, the ability to detect EE-E1 with the MPM-2 antibody was abolished (Fig. 1B, lane 7). Thus, the HPV E1 protein is phosphorylated *in vivo* by one or more CDK complexes on the consensus CDK substrate site(s). Mutational analyses described below showed that CDK phosphorylation substrates in HPV-11 E1 helicase included S89, S93, and S107.

Example 3: Single mutations in S89, S93 or S107 significantly reduce nuclear localization.

Mutations at four CDK phosphorylation sites abrogated E1 function to support viral origin DNA replication *in vivo* (Ma et al. (1999)). In the majority of the transfected cells, GFP-E1 was restricted to the nucleus, whereas a small percentage of the cells exhibit cytoplasmic or pancellular signals (Deng et al. (2003)). To understand the mechanism by which CDK phosphorylation regulates the E1 replication activity, it was determined whether phosphorylation could affect its subcellular localization. The subcellular localization of GFP-E1 fusion protein was tracked after the incorporation of single or multiple Ser to Ala mutations at the candidate CDK phosphorylation sites. All direct fluorescence microscopy of GFP-E1 or mutations was conducted in monkey kidney epithelial cell line COS7 cells and in human kidney epithelial cell line 293 cells. The results were identical.

By fluorescence microscopy, the majority of cells expressing the S89A, S93A, or S107A single mutations exhibited cyctoplasmic GFP-E1 (Fig. 1C). S107A exhibited the most severe defect in E1 nuclear localization. More than 80% of the cells with GFP-E1 S107A contained only cytoplasmic signals, whereas in S89A and S93A expressing cells, about 60-70% had exclusivly cytoplasmic signals. In the remaining cells for all three

mutations, there was some nuclear signal in addition to a dominant cytoplasmic signal.

With the three double mutations S89,93A, S89,107A, or S93,107A, more than 90% of the transfected cells had exclusively cytoplasmic GFP-E1 protein (Fig. 1C). When all three serine residues or all four CDK substrates were mutated, as in S89,93,107A and S89,93,107,T468A, all the GFP-E1 protein was restricted to the cytoplasm (Fig. 1C). These observations show that CDK phosphorylation on S89, S93 and S107 are all required for effective retention of E1 in the nucleus.

Example 4: Lack of phosphorylation by cyclin/CDK leads to cytoplasmic E1.

E1 mutations unable to bind cyclins and are no longer able to be phosphorylated by CDK complexes should exhibit a phenotype similar to the phosphorylation site mutations. To demonstrate this, GFP-E1 mutated in the consensus cyclin binding motif RxL (Ma et al. (1999)) was constructed and tested. The RRL sequence was mutated to RRA, KRA or ARA (Fig. 2A). The ARA mutation has previously been shown to disrupt the binding to cyclins (43). HPV-11 E1 protein has a bipartite nuclear localization sequence (NLS), which partially overlap the cyclin binding motif, and mutation of the RR residues decreased but did not abolish E1 nuclear localization (see Fig. 2A). All three mutations were localized to the cytoplasm (Fig. 2A, left panel). Western blots with the phospho-S/TP-specific antibody MPM-2 failed to detect the GFP-E1 RRA, although the protein was detected at a similar level as the wild type protein by using an anti-GFP antibody (Fig. 2A, right panel, compare lanes 2 and 3). In contrast, the wild type GFP-E1, which is primarily nuclear (Fig. 1C), was detected by the MPM-2 and GFP antibodies, consistent with the observation with the EE-E1 protein (Fig. 1B). In addition, the GFP-E1 S89,93,107A,T468A protein, which is exclusively cytoplasmic (Fig. 1C), did not produce a detectable signal when probed with the MPM-2 antibody (Fig. 2A, right panel, lane 1). Collectively, these results show that E1 phosphorylation at S89, S93 and S107 is required for efficient nuclear localization.

Example 5: Inhibition of CDK activities results in cytoplasmic E1 protein.

To verify the above interpretation further, COS7 cells transfected with GFP-E1 were treated with 20 μM roscovitine for 20 hours. The results showed that specific inhibition of CDK activities resulted in a GFP-E1 which was no longer detectable with the MPM-2 antibody (Fig. 2B, right panel). On the cellular level, the GFP-E1 protein was observed exclusively in the cytoplasm in treated cultures (Fig. 2B, left panel, compared

with Fig. 1C). Similarly, roscovitine treatment of transfected cells also resulted in the cytoplasmic relocalization of the GFP-E1 T468A.

In another experiment, elevated p21cip1 expression was used to inhibit CDK activity. p21cip1 is a potent inhibitor of cyclin E/CDK2 and cyclin A/CDK2 complexes that regulate G1/S phase transition and also plays a key role in G1/S or G2 checkpoint after DNA damage (Abraham, R. T. *Genes Dev.* 15:2177-2196 (2001), Sherr et al. (1999), Sun et al. *J Gen Virol* 79:1651-1658 (1998)). GFP-E1 expression vectors were transfected into the p21-9 cell line. This cell line is derived from HT1080 human fibrosarcoma cells with an inducible p21cip1 gene under the control of the bacteria *lac* I repressor (Chang et al. (2000)). Maximal induction of the p21cip1 protein was achieved by adding isopropyl-β-thio-galactosidase (IPTG) to the medium at 50 μM for 14 to 24 hrs (Chang et al. (2000)). In the absence of IPTG, the GFP-E1 protein was primarily nuclear, as in COS7 or 293 cells. Upon induction with IPTG, GFP-E1 was observed in the cytoplasm in the majority of the cells (Fig. 2C). Collectively, these results show that phosphorylation of E1 by cyclin/CDK complexes, including CDK2, is required for E1 nuclear localization.

Example 6: Mutations in E1 phosphorylation sites promote CRM1 mediated nuclear export. To determine whether phosphorylation affected E1 nuclear import or nuclear export, cells were treated with leptomycin B (LMB). LMB is a potent antifungal antibiotic which binds to the nuclear export receptor CRM1 in a highly specific manner and inhibits nuclear export (Kudo et al. *Proc. Natl. Acad. Sci. USA* 96:9112-9117 (1999), Nishi et al. *J. Biol. Chem.* 269:6320-6324 (1994)). COS7 cells expressing wild type GFP-E1 proteins were treated with 20 ng/ml LMB for 6 hours (Fig. 2D). Virtually all cells were observed to have the GFP-E1 protein in the nucleus. In untreated cells, about 20-30% contained some cytoplasmic signal (Deng et al., 2003). In the presence of LMB, GFP-E1 S107A, GFP-E1 S89,93,107A, and GFP-E1 L126A (in the RxL motif) were also observed in the nucleus, in contrast to their original, primarily cytoplasmic distribution in the absence of LMB (compare Fig. 2D with Fig. 1C and 2A). Prolonged treatment with LMB increased nuclear localization and all the protein was localized to the nucleus at 11 hrs.. Similarly, all other E1 phosphorylation site mutations were also found in the nucleus upon LMB treatment. Collectively, these results demonstrate that cytoplasmic distribution for each of these E1 mutations was caused by the rapid, continuous export from nucleus mediated by CRM1 and that the dominant, CRM1-dependent nuclear export is inactivated

by phosphoryation by cyclin/CDK complexes. These observations also indicate that neither the loss of phosphorylation by CDK nor the mutation in the overlapping bipartite NLS and consensus cyclin binding motif (RRL) is able to abolish its nuclear entry completely.

Example 7: A conserved, putative leucine-rich NES controls E1 export.

CRM1 binds specifically to proteins with a leucine-rich NES and, in a complex with RanGTP, and transports the cargo out of the nucleus (Stade et al. (1997)). Leucine-rich NES is a short peptide sufficient to promote rapid protein export from the nucleus (Fischer et al. (1995), Wen et al. (1995)). It has a core sequence of about 10 residues that are highly enriched with large hydrophobic residues such as leucine, isoleucine and valine. A consensus sequence for the core peptide has been proposed to be L-X(2-3)-L/I/V/F/M-X(2-3)-L-X-L/I (SEQ ID NO: 16). Upon inspecting the whole E1 sequence with special attention to the region close to the three CDK substrates, one region (residues 83-126, SEQ ID NO: 2) was identified which is rich in leucine, isoleucine or valine residues (Fig. 3A, green box), including a putative NES core consensus sequence (green underline), ISPRLDAIKL (residues 106 to 115, SEQ ID NO: 14).

To test whether this region might be an NES, two hydrophobic residues (L110 and I113, SEQ ID NO: 9 and 10) were mutated in this apparent core sequence to alanine substitutions (designated NEm) in the wild type GFP-E1 and a number of CDK phosphorylation site mutations. These GFP-E1 mutations were transfected into COS7 cells. The nuclear localization of S89A-NEm, S93A-NEm, and S89,93,107A-NEm were restored to different extent. In contrast, the wild type E1-NEm and SA107A-NEm were both exclusively nuclear (Fig. 3B, compared with Fig. 1B). The dramatic difference between S107A and S107A-NEm, compared with the other phosphorylation sites mutations and the corresponding NEm, demonstrates that the primary function of phosphorylation of S107 is to inactivate the NES.

Example 8: CDK phosphorylation inactivates the NES of the E1 protein.

To substantiate the hypothesis that one of the critical roles of CDK phosphorylation is to inactivate the NES, the NEm was also introduced into the GFP-E1 L126A mutated in the cyclin binding motif. GFP-E1 L126A was cytoplasmic (Fig. 2A) and was not detected by the MPM-2 antibody in western blots (Fig. 2A). The introduction of the NEm did not restore CDK phosphorylation, as the protein did not react with the MPM-2 antibody (Fig. 2A, right panel, lanes 3, 4). Nevertheless, the GFP-E1 RRA-NEm protein was primarily

located in the nucleus (Fig. 3B). Thus, this observation shows that the L110A,I113A mutation inactivates the NES and that the NES mutation can at least partially substitute for the requirement for CDK phosphorylation for E1 nuclear retention. Cells transfected with the NEm mutations were also treated with roscovitine. This CDK inhibitor caused the wild type E1 protein to redistribute to the cytoplasm (Fig. 2B), but it had no effect on the subcellular localization of any of the NES mutations. These results show that the NES is effectively inactivated by the L110A,I113A mutation and hence CDK phosphorylation is no longer necessary for its inactivation. Collectively, these results show that the leucine-rich peptide spanning residues 106-115 (SEQ ID NO: 14) functions as an NES and one of the functions of CDK complexes is to inactivate it by phosphorylation.

The above localization experiments were conducted by direct observation of GFP-tagged E1 protein. This protein is fully functional in supporting transient replication (Deng et al. (2003)). To verify that these observations are not confounded by the GFP moiety, three EE-tagged E1 or mutations were expressed in COS7 cells. Indirect immunofluoescent staining of EE-E1, EE-E1 S107A, EE-E1 S107A-NEm using antibody against the EE-epitope showed that each of the protein exhibited exactly the same subcellular localization patterns as the corresponding GFP-E1 counterparts (Fig. 3C, compare with Figs. 1C and 3B). EE-E1 was observed in the nucleus in most of the cells. The phosphorylation site mutation S107A was primarily in the cytoplasm. The double mutation S107A-NEm was localized entirely in the nucleus. Thus, through observations made on both GFP-tagged and EE-tagged E1, it was demonstrated that E1 phosphorylation by CDKs determines its nuclear localization through the disruption of a leucine-rich NES.

Example 9: Delineation of a *bona fide* leucine-rich NES sufficient for protein export.

To delineate the NES of the E1 protein, a series of short peptides spanning the 10 amino acid NES core (106-115, SEQ ID NO: 14) were fused to the C-terminus of the GFP protein. The region between residues 86 and 119 was purposefully limited to circumvent any complications which might result from the NLS or from phosphorylation by cyclin/CDK, the association of which depends on the consensus cyclin binding motif (residues 124-126, SEQ ID NO: 6). As shown in Fig. 3D, because of it small size, GFP alone was observed both in the nucleus and in the cytoplasm. The localization of the fusion proteins showed that each of the E1 peptides spanning residues 96-115 (SEQ ID NO: 15)

conferred an exclusive cytoplasmic localization to GFP. Each of these peptides contains the 10-amino acid long core sequence (residues 106-115, SEQ ID NO: 14) and also includes two additional valine residues (Fig. 3A). In contrast, the 10 amino-acid core sequence was not sufficient (Fig. 3D). Thus a *bona fide* leucine-rich NES in HPV-11 E1 protein between residues 96 and 115 was defined and determined to be necessary and sufficient for export from the cell nucleus.

Example 10: E1 replication activity correlates with its subcellular localization.

If one of the functions of E1 phosphorylation is to determine its nuclear localization, this would explain why the E1 S89,93,107,T468A mutation supported virus origin DNA replication *in vivo* extremely poorly, if at all (Ma et al. (1999)). A prediction of this conclusion is that, if nuclear export could be abolished, a recovery of the replication activities of the phosphorylation site mutations should be observed. Therefore, transient replication assays were performed (Fig. 4). 293 cells were co-transfected with 0.5 μg of an E1 expression plasmid, 5 μg of pMT2-E2 expression plasmid and 0.5 μg of a HPV origin-containing plasmid. The extent of the viral origin DNA replication was evaluated by southern blotting of the newly synthesized origin plasmid (Dpn I + lanes) 48 hr post transfection. A subset of the GFP-E1 phosphorylation site mutations and their corresponding NES mutations (NEm) were tested. Fig. 4A is an autoradiogram of one of several experiments. The signals were quantified in Fig. 4B to obtain the relative replication activities. The data show that E1 proteins mutated at phosphorylation site(s) or cyclin binding motif had much reduced activities to support virus DNA replication (Fig. 4A, upper panel). The extent of decrease correlated largely with the percentages of cells containing primarily cytoplasmic E1 protein (see Fig. 1C). When nuclear localization of these mutations was restored through mutations in the NES (Fig. 3B), their abilities to support replication were increased, although to different extent (Fig. 4A, lower panel). These experiments were conducted using different amounts of GFP-E1 plasmids and a similar tendency was observed. These results clearly demonstrate that cyclin/CDK phosphorylation inactivates E1 NES, promotes its nuclear retention, and consequently controls virus DNA replication. The results also show that the phosphorylation site mutations per se did not completely abrogate the ability of the E1 protein to function as a replicative helicase or to interact with the host proteins to initiate replication. However, mutations involving S107 appeared to have additional functional impairment, as its

replication activity was the lowest among all the mutations even when NEm was introduced (Fig. 4).

Example 11: Altered subcellular localization of E1 S89,93,107A protein in the presence or absence of the E2 protein.

To investigate the reason for the poor replication activity of E1S89,93,107, subcellular localization of the GFP-E1 wild type protein and GFP-E1 S89,93,107A was examined in the absence and in the presence of the native E2 in transfected COS cells. E2 was revealed by indirect immunofluorescence detection with E2 antibody. In a great majority of the cells, the wild type GFP-E1 colocalized with E2 and formed small nuclear foci above diffuse E1 and E2 nuclear signal (Fig. 5A, top panels). Only in less than 5% of the cells positive for nuclear E2, cytoplasmic GFP-E1 was observed (Fig. 5A, lower cell in bottom panels). In the absence of E2, up to 20-30% of the cells exhibited some cytoplasmic signals. Thus, under the condition of high levels of protein expression, E2 was able to increase the percentage of cells with nuclear wild type GFP-E1 protein.

In contrast, in the majority of cells expressing GFP-E1 S89,93,107A and the native E2 protein, the GFP-E1 protein remained cytoplasmic (Fig. 5B, top panels). These results show that the E2 fail to overcome effectively the powerful NES in E1 to prevent the nuclear export of this E1 phosphorylation mutation. In addition, this mutation might have other defects. This is inferred from the pattern of E1 distribution in the remaining cells in which GFP-E1 was observed in the nucleus or in both compartments. When GFP-E1 S89,93,107A was observed in the nucleus, the two proteins colocalized in dots much brighter and larger than the foci formed by wild type E1 and E2 proteins. In addition, there was very little diffuse GFP-E1 remaining in the nucleoplasm (compare Fig. 5A top panels with Fig. 5B bottom panels). These observations indicate that in a low percentage of cells, this mutated E1 was able to interact with E2 and, in doing so, almost all of the GFP-E1 and most of the E2 protein were sequestered into these dots (Fig .5).

Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

REFERENCES

1. Abraham, R. T. 2001. Cell cycle checkpoint signaling through the ATM and ATR kinases. Genes Dev. 15:2177-2196.

2. Amin, A. A., S. Titolo, A. Pelletier, D. Fink, M. G. Cordingley, and J. Archambault. 2000. Identification of domains of the HPV11 E1 protein required for DNA replication in vitro. Virology 272:137-150.

3. Bain, J., H. McLauchlan, M. Elliott, and P. Cohen. 2003. The specificities of protein kinase inhibitors: an update. Biochem. J. 371:199-204.

4. Bell, S. P., and A. Dutta. 2002. DNA replication in eukayotic cells. Annu. Rev. Biochem. 71:333-374.

5. Blow, J. J., and B. Hodgson. 2002. Replication licensing -- Origin licensing: defining the proliferative state? Trends Cell Biol. 12:72-78.

6. Bogerd, H., R. Fridell, R. Benson, J. Hua, and B. Cullen. 1996. Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel in vivo randomization-selection assay. Mol. Cell. Biol. 16:4207-4214.

7. Chang, B.-D., K. Watanabe, E. V. Broude, J. Fang, J. C. Poole, T. V. Kalinichenko, and I. B. Roninson. 2000. Effects of p21Waf1/Cip1/Sdi1 on cellular gene expression: Implications for carcinogenesis, senescence, and age-related diseases. Proc. Natl. Acad. Sci. USA 97:4291-4296.

8. Chen, G., and A. Stenlund. 1998. Characterization of the DNA-Binding Domain of the Bovine Papillomavirus Replication Initiator E1. J. Virol. 72:2567-2576.

9. Chiang, C.-M., G. Dong, T. R. Broker, and L. T. Chow. 1992a. Control of human papillomavirus type 11 origin of replication by the E2 family of transcription regulatory proteins. J. Virol. 66:5224-5231.

10. Chiang, C.-M., M. Ustav, A. Stenlund, T. F. Ho, T. R. Broker, and L. T. Chow. 1992. Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins. Proc. Natl. Acad. Sci. USA 89:5799-5803.

11. Chow, L. T., and T. R. Broker. 1997. Small DNA tumor viruses, p. 267-301. In N. Nathanson (ed.), Viral Pathogenesis. Lippincott-Raven Publishers, Philadelphia.

12. Conger, K. L., J.-S. Liu, S.-R. Kuo, L. T. Chow, and T. S.-F. Wang. 1999. Human papillomavirus DNA replication. Interactions between the viral E1 protein and two subunits of human DNA polymerase a /primase. J. Biol. Chem. 274:2696-2705.

13. Coue, M., S. E. Kearsey, and M. Mechali. 1996. Chromotin binding, nuclear localization and phosphorylation of Xenopus cdc21 are cell-cycle dependent and associated with the control of initiation of DNA replication. EMBO J. 15:1085-1097.

14. Cueille, N., R. Nougarede, F. Mechali, M. Philippe, and C. Bonne-Andrea. 1998. Functional interaction between the bovine papillomavirus virus type 1 replicative helicase E1 and cyclin E-cdk2. J. Virol. 72:7255-7262.

15. Dalton, S., and L. Whitbread. 1995. Cell cycle-regulated nuclear import and export of Cdc47, a protein essential for initiation of DNA replication in budding yeast. Proc. Natl. Acad. Sci. U S A 92:2514-2518.

16. De Azevedo, W. F., S. Leclerc, L. Meijer, L. Havlicek, M. Strnad, and S. H. Kim. 1997. Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human cdk2 complexed with roscovitine. Eur. J. Biochem. 243:518-526.

17. Delmolino, L. M., P. Saha, and A. Dutta. 2001. Multiple mechanisms regulate subcellular localization of human CDC6. J. Biol. Chem. 276:26947-26954.

18. Deng, W., G. Jin, B.-Y. Lin, B. A. Van Tine, T. R. Broker, and L. T. Chow. 2003. mRNA splicingr regulates human papillomavirus type 11 E1 protein production and DNA replication. J. Virol. 77:10213-10226.

19. Fischer, U., J. Huber, W. C. Boelens, I. W. Mattaj, and R. Luhrmann. 1995. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell 82:475-483.

20. Fornerod, M., M. Ohno, M. Yoshida, and I. Mattaj. 1997. CRM1 is an export receptor for leucine-rich nuclear export signals. Cell 90:1051-1060.

21. Fouts, E. T., X. Yu, E. H. Egelman, and M. R. Botchan. 1999. Biochemical and electron microscopic image analysis of the hexameric E1 helicase. J. Biol. Chem. 274:4447-4458.

22. Gillitzer, E., G. Chen, and A. Stenlund. 2000. Separate domains in E1 and E2 proteins serve architectural and productive roles for cooperative DNA binding. EMBO J. 19:3069-3079.

23. Grussenmeyer, T., K. H. Scheidtmann, M. A. Hutchinson, W. Eckhart, and G. Walter. 1985. Complexes of polyoma virus medium T antigen and cellular proteins. Proc. Natl. Acad. Sci. U S A. 82:7952-7954.

24. Hendrickson, M., M. Madine, S. Dalton, and J. Gautier. 1996. Phosphorylation of MCM4 by cdc2 protein kinase inhibits the activity of the minichromosome maintenance complex. Proc. Natl. Acad. Sci. USA 93:12223-12228.

25. Hubner, S., C.-Y. Xiao, and D. A. Jans. 1997. The protein kinase CK2 site (Ser111/112) enhances recognition of the simian virus 40 large T-antigen nuclear localization sequence by importin. J. Biol. Chem. 272:17191-17195.

26. Hughes, F. J., and M. A. Romanos. 1993. E1 protein of human papillomavirus is a DNA helicase/ATPase. Nucleic Acids Res. 21:5817-5823.

27. Ishimi, Y., and Y. Komamura-Kohno. 2001. Phosphorylation of Mcm4 at specific sites by cyclin-dependent kinase leads to loss of Mcm4,6,7 helicase activity. J. Biol. Chem. 276:34428-34433.

28. Ishimi, Y., Y. Komamura-Kohno, Z. You, A. Omori, and M. Kitagawa. 2000. Inhibition of Mcm4,6,7 helicase activity by phosphorylation with Cyclin A/Cdk2. J. Biol. Chem. 275:16235-16241.

29. Jans, D., C. Xiao, and L. MH. 2000. Nuclear targeting signal recognition: a key control point in nuclear transport? Bioessays 22:532-544.

30. Jans, D. A., and S. Hubner. 1996. Regulation of protein transport to the nucleus: central role of phosphorylation. Physiol. Rev. 76:651-685.

31. Jiang, W., N. J. Wells, and T. Hunter. 1999. Multistep regulation of DNA replication by Cdk phosphorylation of HsCdc6. Proc. Natl. Acad. Sci. USA 96:6193-6198.

32. Komeili, A., and E. K. O'Shea. 2001. New perspectives on nuclear transport. Annu. Rev. Genet. 35:341-364.

33. Kudo, N., N. Matsumori, H. Taoka, D. Fujiwara, E. P. Schreiner, B. Wolff, M. Yoshida, and S. Horinouchi. 1999. Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. Proc. Natl. Acad. Sci. USA 96:9112-9117.

34. Kuo, S.-R., J.-S. Liu, T. R. Broker, and L. T. Chow. 1994. Cell-free replication of the human papillomavirus DNA with homologous viral E1 and E2 proteins and human cell extracts. J. Biol. Chem. 269:24058-24065.

35. Labib, K., J. F. Diffley, and S. E. Kearsey. 1999. G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. Nat. Cell Biol. 1:415-422.

36. Laney, J. D., and M. Hochstrasser. 1999. Substrate targeting in the ubiquitin system. Cell 97:427-430.

37. Leng, X., V. G. Wilson, and X. L. Xiao. 1994. Genetically defined nuclear localization signal sequence of bovine papillomavirus E1 protein is necessary and sufficient for the nuclear localization of E1-beta-galactosidase fusion proteins. J Gen Virol. 75:2463-2467.

38. Lentz, M. R., D. Pak, I. Mohr, and M. R. Botchan. 1993. The E1 replication protein of bovine papillomavirus type 1 contains an extended nuclear localization signal that includes a p34cdc2 phosphorylation site. J. Virol. 67:1414-1423.

39. Lin, B. Y., T. Ma, J.-S. Liu, S.-R. Kuo, G. Jin, T. R. Broker, J. W. Harper, and L. T. Chow. 2000. HeLa cells are phenotypically limiting in cyclin E/CDK2 for efficient human papillomavirus DNA replication. J. Biol. Chem. 275:6167-6174.

40. Lin, B. Y., A. M. Makhov, J. D. Griffith, T. R. Broker, and L. T. Chow. 2002. Chaperone proteins abrogate inhibition of the human papillomavirus (HPV) E1 replicative helicase by the HPV E2 protein. Mol. Cell. Biol. 22:6592-6604.

41. Liu, J.-S., S.-R. Kuo, T. R. Broker, and L. T. Chow. 1995. The functions of human papillomavirus type 11 E1, E2, and E2C proteins in cell-free DNA replication. J. Biol. Chem. 270:27283-27291.

42. Liu, J.-S., S.-R. Kuo, A. M. Makhov, D. M. Cyr, J. D. Griffith, T. R. Broker, and L. T. Chow. 1998. Human Hsp70 and Hsp40 chaperone proteins facilitate human papillomavirus-11 E1 protein binding to the origin and stimulate cell-free DNA replication. J. Biol. Chem. 273:30704-30712.

43. Ma, T., N. Zou, B. Y. Lin, L. T. Chow, and J. W. Harper. 1999. Interaction between cyclin-dependent kinases and human papillomavirus replication-initiation protein E1 is required for efficient viral replication. Proc. Natl. Acad. Sci. USA 96:382-387.

44. Masterson, P. J., M. A. Stanley, A. P. Lewis, and M. A. Romanos. 1998. A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase alpha -primase p68 subunit. J. Virol. 72:7407-7419.

45. McShan, G. D., and V. G. Wilson. 1997. Casein kinase II phosphorylates bovine papillomavirus type 1 E1 in vitro at a conserved motif. J. Gen. Virol. 78:171-177.

46. McShan, G. D., and V. G. Wilson. 2000. Contribution of bovine papillomavirus type 1 E1 protein residue 48 to replication function. J. Gen. Virol. 81:1995-2004.

47. Meijer, L., A. Borgne, O. Mulner, J. Chong, J. Blow, N. Inagaki, M. Inagaki, J. Delcros, and J. Moulinoux. 1997. Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5. Eur. J. Biochem. 243:527-536.

48. Mohr, I. J., R. Clark, S. Sun, E. J. Androphy, P. MacPherson, and M. R. Botchan. 1990. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. Science 250:1694-1699.

49. Nakielny, S., and G. Dreyfuss. 1997. Nuclear export of proteins and RNAs. Curr. Opin. Cell Biol. 9:420-429.

50. Nguyen, V. Q., C. Co, K. Irie, and J. J. Li. 2000. Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. Curr. Biol. 10:195-205.

51. Nguyen, V. Q., C. Co, and J. J. Li. 2001. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature 411:1068-1073.

52. Nishi, K., M. Yoshida, D. Fujiwara, M. Nishikawa, S. Horinouchi, and T. Beppu. 1994. Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression. J. Biol. Chem. 269:6320-6324.

53. Nishitani, H., and Z. Lygerou. 2002. Control of DNA replication licensing in a cell cycle. Genes to Cells 7:523-534.

54. Pelizon, C., M. A. Madine, P. Romanowski, and R. A. Laskey. 2000. Unphosphorylatable mutants of Cdc6 disrupt its nuclear export but still support DNA replication once per cell cycle. Genes Dev. 14:2526-2533.

55. Pereverzeva, I., E. Whitmire, B. Khan, and M. Coue. 2000. Distinct phosphoisoforms of the Xenopus Mcm4 protein regulate the function of the Mcm complex. Mol. Cell. Biol. 20:3667-3676.

56. Petersen, B. O., J. Lukas, C. S. Sorensen, J. Bartek, and K. Helin. 1999. Phosphorylation of mammalian CDC6 by Cyclin A/CDK2 regulates its subcellular localization. EMBO J. 18:396-410.

57. Pines, J. 1999. Four-dimensional control of the cell cycle. Nat. Cell Biol. 1:E73-79.

58. Rangasamy, D., and V. G. Wilson. 2000a. Bovine papillomavirus E1 protein is sumoylated by the host cell Ubc9 protein. J. Biol. Chem. 275:30487-30495.

59. Remm, M., R. Brain, and J. R. Jenkins. 1992. The E2 binding sites determine the efficiency of replication for the origin of human papillomavirus type 18. Nucleic Acids Res. 20:6015-6021.

60. Rihs, H. P., D. A. Jans, H. Fan, and R. Peters. 1991. The rate of nuclear cytoplasmic protein transport is determined by the casein kinase II site flanking the nuclear localization sequence of the SV40 T-antigen. EMBO J. 10:633-639.

61. Sedman, J., and A. Stenlund. 1998. The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities. J. Virol. 72:6893-6897.

62. Seo, Y.-S., F. Müller, M. Lusky, E. Gibbs, H.-Y. Kim, B. Phillips, and J. Hurwitz. 1993. Bovine papilloma virus (BPV)-encoded E2 protein enhances binding of E1 protein to the BPV replication origin. Proc. Natl. Acad. Sci. USA 90:2865-2869.

63. Seo, Y.-S., F. Müller, M. Lusky, and J. Hurwitz. 1993b. Bovine papilloma virus (BPV)-encoded E1 protein contains multiple activities required for BPV DNA replication. Proc. Natl. Acad. Sci. USA 90:702-706.

64. Sherr, C. J., and J. M. Roberts. 1999. CDK inhibitors: positive and negative regulators of G1-phase progression. Genes Dev. 13:1501-1512.

65. Stade, K., C. S. Ford, C. Guthrie, and K. Weis. 1997. Exportin 1 (Crm1p) is an essential nuclear export factor. Cell 90:1041-1050.

66. Sun, Y., H. Han, and D. McCance. 1998. Active domains of human papillomavirus type 11 E1 protein for origin replication. J Gen Virol 79:1651-1658.

67. Swindle, C. S., N. Zou, B. A. Van Tine, G. M. Shaw, J. A. Engler, and L. T. Chow. 1999. Human papillomavirus DNA replication compartments in a transient DNA replication system. J. Virol. 73:1001-1009.

68. Tanaka, S., and J. F. Diffley. 2002. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. Nat Cell Biol. 4:198-207.

69. Taylor, W. R., and G. R. Stark. 2001. Regulation of the G2/M transition by p53. Oncogene 20:1803-1815.

70. Titolo, S., K. Brault, J. Majewski, P. W. White, and J. Archambault. 2003. Characterization of the Minimal DNA Binding Domain of the Human Papillomavirus E1

Helicase: Fluorescence Anisotropy Studies and Characterization of a Dimerization-Defective Mutant Protein. J. Virol. 77:5178-5191.

71. Tye, B. K. 1999. MCM proteins in DNA replication. Annu. Rev. Biochem. 68:649-686.

72. Ullman, K. S., M. A. Powers, and D. J. Forbes. 1997. Nuclear export receptors: from importin to exportin. Cell 90:967-970.

73. Ustav, M., and A. Stenlund. 1991. Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. EMBO J. 10:449-457.

74. Weis, K. 2003. Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. Cell 112:441-451.

75. Wen, W., J. L. Meinkoth, R. Y. Tsien, and S. S. Taylor. 1995. Identification of a signal for rapid export of proteins from the nucleus. Cell 82:463-473.

76. White, P. W., A. Pelletier, K. Brault, S. Titolo, E. Welchner, L. Thauvette, M. Fazekas, M. G. Cordingley, and J. Archambault. 2001. Characterization of recombinant HPV6 and 11 E1 helicases. Effect of ATP on the interaction of E1 with E2 and mapping of a minimal helicase domain. J. Biol. Chem. 276:22426-22438.

77. Xiao, C.-Y., P. Jans, and D. A. Jans. 1998. Negative charge at the protein kinase CK2 site enhances recognition of the SV40 large T-antigen NLS by importin: effect of conformation. FEBS Letters 440:297-301.

78. Yan, H., A. M. Merchant, and B. K. Tye. 1993. Cell cycle-regulated nuclear localization of MCM2 and MCM3, which are required for the initiation of DNA synthesis at chromosomal replication origins in yeast. Genes Dev. 7:2149-2160.

79. Yang, J., E. S. G. Bardes, J. D. Moore, J. Brennan, M. A. Powers, and S. Kornbluth. 1998. Control of Cyclin B1 localization through regulated binding of the nuclear export factor CRM1. Genes Dev. 12:2131-2143.

80. Yang, J., H. Song, S. Walsh, E. S. G. Bardes, and S. Kornbluth. 2001. Combinatorial control of Cyclin B1 nuclear trafficking through phosphorylation at multiple sites. J. Biol. Chem. 276:3604-3609.

81. Yang, L., R. Li, I. J. Mohr, R. Clark, and M. R. Botchan. 1991. Activation of BPV-1 replication in vitro by the transcription factor E2. Nature 353:628-632.

82. Yang, L., I. Mohr, E. Fouts, D. A. Lim, M. Nohaile, and M. R. Botchan. 1993. The E1 protein of bovine papilloma virus 1 is an ATP-dependent DNA helicase. Proc. Natl. Acad. Sci. USA 90:5086-5090.

83. Zanardi, T. A., C. M. Stanley, B. M. Saville, S. M. Spacek, and M. R. Lentz. 1997. Modulation of bovine papillomavirus DNA replication by phosphorylation of the viral E1 protein. Virology 228:1-10.

84. Zou, N., B. Y. Lin, F. Duan, K.-Y. Lee, G. Jin, R. Guan, G. Yao, E. J. Lefkowitz, T. R. Broker, and L. T. Chow. 2000. The hinge of the human papillomavirus type 11 E2 protein contains major determinants for nuclear localization and nuclear matrix association. J. Virol. 74:3761-3770.

85. zur Hausen, H., and E.-M. de Villiers. 1994. Human papillomaviruses. Annu. Rev. Microbiol. 48:427-447.

What is claimed is:

1. A method of inhibiting papillomavirus replication in a subject comprising the steps of:
   a. identifying a subject with, or at risk of contracting, papillomavirus; and
   b. contacting the subject with a substance that reduces nuclear localization of E1, thereby inhibiting papillomavirus replication.
2. The method of claim 1, wherein the substance reduces nuclear localization by reducing the interaction of E1 with importin-α.
3. The method of claim 1, wherein the substance reduces nuclear localization by inhibiting cyclin/cdk2 maintenance of phosphorylation.
4. The method of claim 3, wherein the substance is roscovitine.
5. The method of claim 1, wherein the substance reduces nuclear localization by stimulating CRM1.
6. The method of claim 1, wherein the substance reduces nuclear localization by increasing p21cip1 or p27kip1.
7. The method of claim 8, wherein the substance increases p21cip1 by inhibiting 26S proteasome.
8. The method of claim 7, wherein the substance is lactacystin.
9. A method of screening for an agent that inhibits nuclear localization of E1, comprising the steps of:
   a. transfecting a cell with a tagged E1 protein, fragment, or mutant thereof;
   b. contacting the cell with the agent to be screened;
   c. inducing expression of the tagged E1 protein, fragment, or mutant thereof; and
   d. detecting a cellular location of the tagged E1 protein, fragment, or mutant thereof wherein E1's absence in the nucleus or presence in the cytoplasm indicating an agent that inhibits nuclear localization of E1.
10. The method of claim 9, wherein tagged E1 protein is SEQ ID NO: 1.
11. The method of claim 9, wherein the tagged E1 protein is SEQ ID NO: 2.
12. The method of claim 9, wherein the tag is a fluorescent tag.
13. The method of claim 12, wherein the tag is GFP.
14. The method of claim 12, wherein the tag is RFP.

15. The method of claim 9, wherein the method of screening comprises using a high throughput screen.
16. An agent identified by the method of claim 9.
17. An expression vector comprising SEQ ID NO: 3, or a fragment thereof, operatively linked to a tag.
18. An isolated cell comprising the expression vector of claim 17.
19. A cell line comprising the vector of claim 17.
20. The vector of claim 17, wherein the tag is a fluorescent tag.
21. The vector of claim 20, wherein the fluorescent tag is GFP.
22. The vector of claim 20, wherein the fluorescent tag is RFP.
23. An expression vector comprising SEQ ID NO: 4, or a fragment thereof, operatively linked to a tag.
24. An isolated cell comprising the expression vector of claim 23.
25. A cell line comprising the vector of claim 23.
26. The vector of claim 23, wherein the tag is a fluorescent tag.
27. The vector of claim 26, wherein the fluorescent tag is GFP.
28. The vector of claim 26, wherein the fluorescent tag is RFP.
29. A plurality of cells stably transfected with the vector of claim 25, wherein the expression of a polypeptide encoded by the vector is inducible.
30. The cells of claim 25, wherein at least about 80% of the cells are induced to express the polypeptide.
31. The cells of claims 25, wherein at least about 90% of the cells are induced to express the polypeptide.
32. The cells of claims 27, wherein at least about 95% of the cells are induced to express the polypeptide.
33. The cells of claims 25, wherein at least about 98% of the cells are induced to express the polypeptide.
34. A non-naturally occurring polypeptide comprising SEQ ID NO: 2.
35. A non-naturally occurring polypeptide comprising SEQ ID NO: 5.
36. A non-naturally occurring polypeptide comprising SEQ ID NO: 9.
37. A non-naturally occurring polypeptide comprising SEQ ID NO: 10.
38. A non-naturally occurring polypeptide comprising SEQ ID NO: 2.

39. The polypeptide of claim 38, wherein serine at amino acid 89 is mutated.

40. The polypeptide of claim 38, wherein serine at amino acid 93 is mutated.

41. The polypeptide of claim 38, wherein serine at amino acid 107 is mutated.

42. The polypeptide of any one of claims 39-41, wherein alanine is substituted for serine.

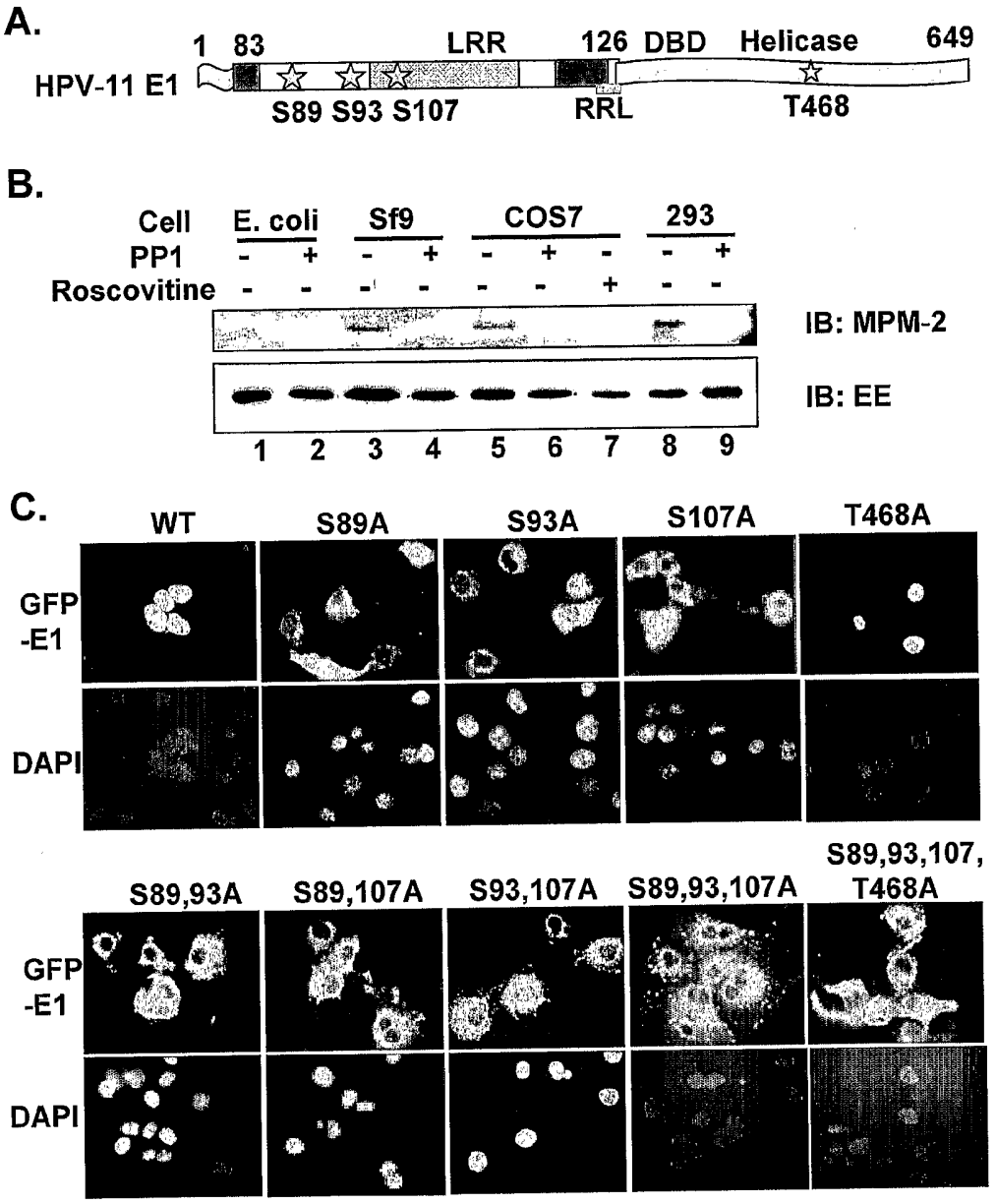

A.
1
83
LRR
126
DBD
Helicase
649
HPV-11 E1
S89 S93 S107
RRL
T468
B.
Cell
E. coli
Sf9
COS7
293
PP1
Roscovitine
IB: MPM-2
IB: EE
1 2 3 4 5 6 7 8 9
C.
WT
S89A
S93A
S107A
T468A
GFP -E1
DAPI
S89,93A
S89,107A
S93,107A
S89,93,107A
S89,93,107, T468A
GFP -E1
DAPI

FIG. 1

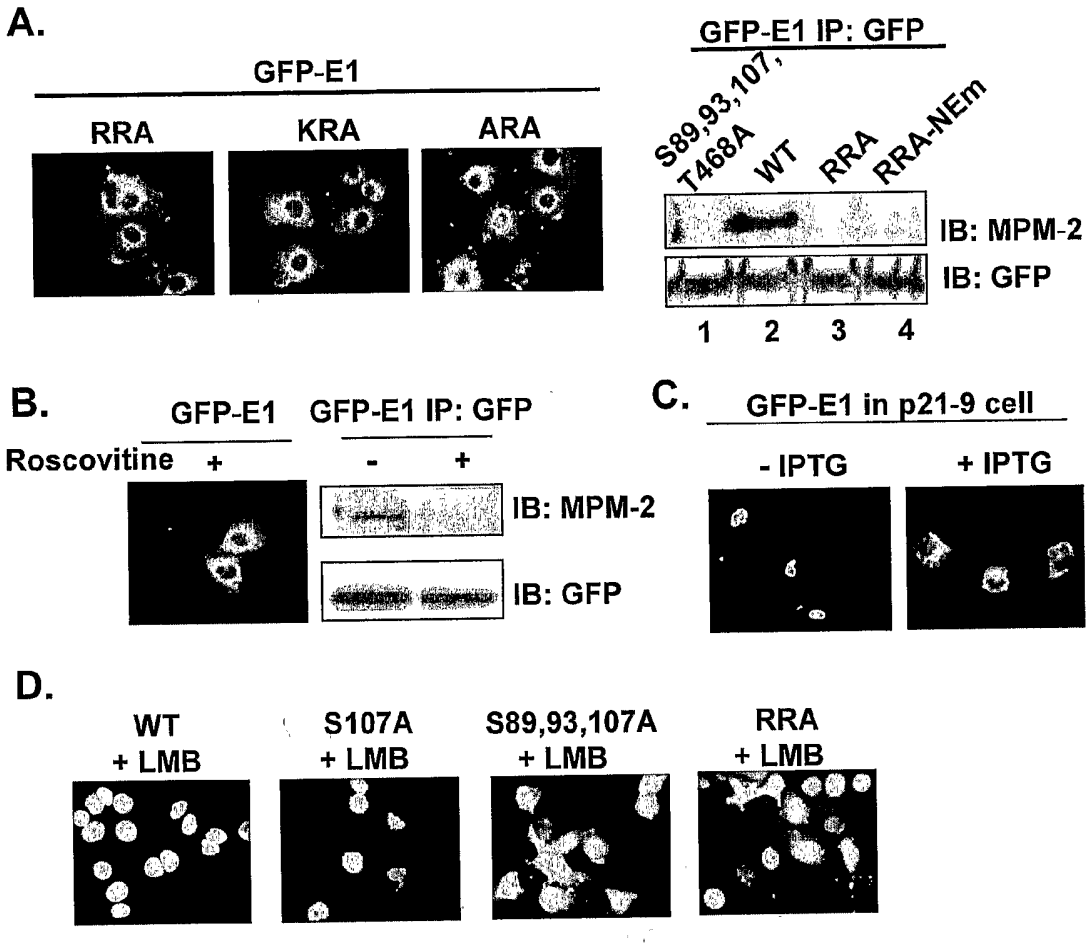

A.
GFP-E1
RRA
KRA
ARA
GFP-E1 IP: GFP
S89,93,107, T468A
WT
RRA
RRA-NEm
IB: MPM-2
IB: GFP
1 2 3 4
B.
GFP-E1
GFP-E1 IP: GFP
Roscovitine + - +
IB: MPM-2
IB: GFP
C.
GFP-E1 in p21-9 cell
- IPTG
+ IPTG
D.
WT + LMB
S107A + LMB
S89,93,107A + LMB
RRA + LMB

FIG. 2

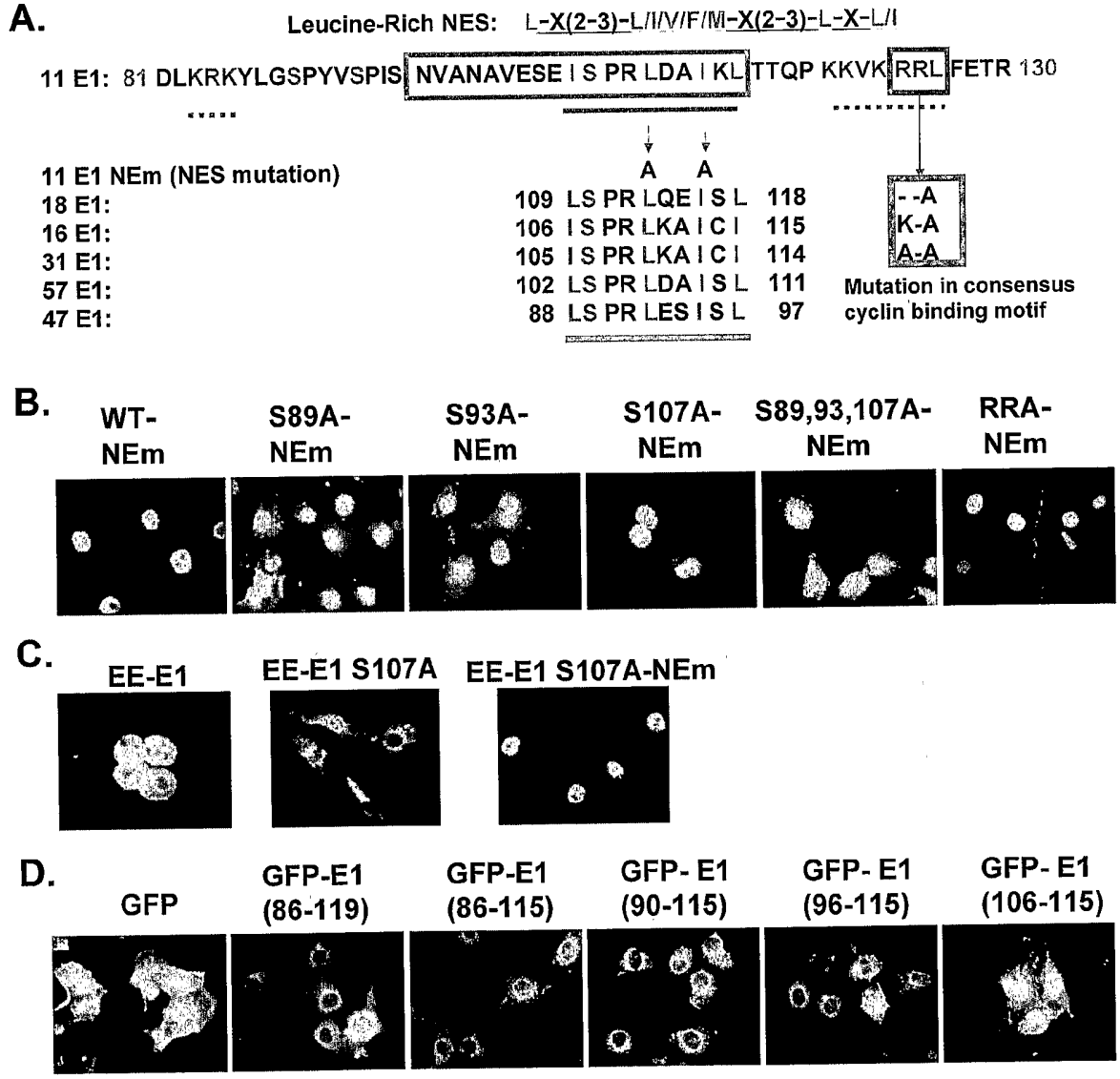

A.
Leucine-Rich NES: L-X(2-3)-L/I/V/F/M-X(2-3)-L-X-L/I
11 E1: 81 DLKRKYLGSPYVSPIS NVANAVESE I S PR LDA I KL TTQP KKVK RRL FETR 130
11 E1 NEm (NES mutation)
18 E1:
16 E1:
31 E1:
57 E1:
47 E1:
A A
109 LS PR LQE I S L 118
106 I S PR LKA I C I 115
105 I S PR LKA I C I 114
102 LS PR LDA I S L 111
88 LS PR LES I S L 97
- -A
K-A
A-A
Mutation in consensus cyclin binding motif
B.
WT-NEm
S89A-NEm
S93A-NEm
S107A-NEm
S89,93,107A-NEm
RRA-NEm
C.
EE-E1
EE-E1 S107A
EE-E1 S107A-NEm
D.
GFP
GFP-E1 (86-119)
GFP-E1 (86-115)
GFP- E1 (90-115)
GFP- E1 (96-115)
GFP- E1 (106-115)

FIG. 3

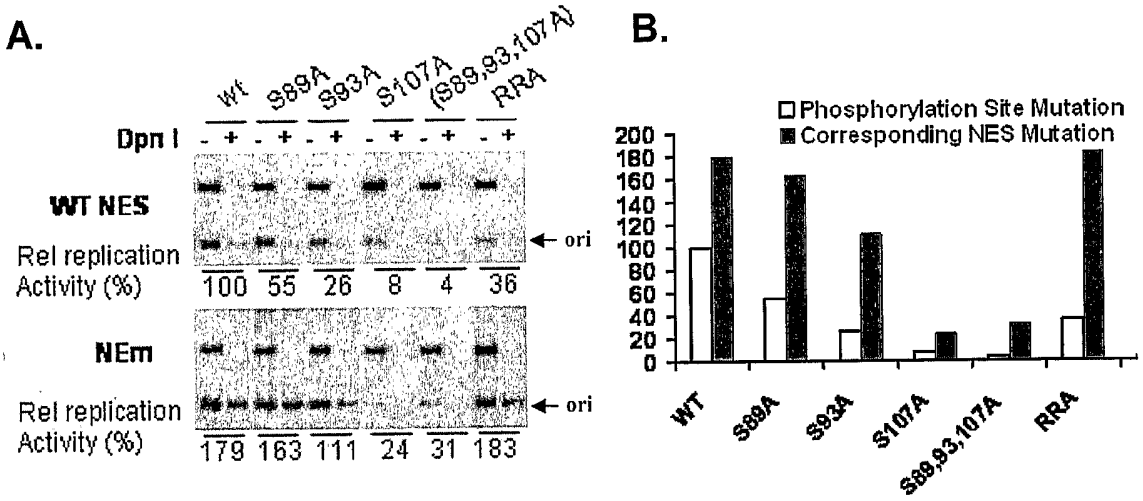

A.
Wt
S89A
S93A
S107A
(S89,93,107A)
RRA
Dpn I
- + - + - + - + - + - +
WT NES
Rel replication Activity (%)
100 55 26 8 4 36
ori
NEm
Rel replication Activity (%)
179 163 111 24 31 183
ori
B.
Phosphorylation Site Mutation
Corresponding NES Mutation
200
180
160
140
120
100
80
60
40
20
0
WT
S89A
S93A
S107A
S89,93,107A
RRA

FIG. 4

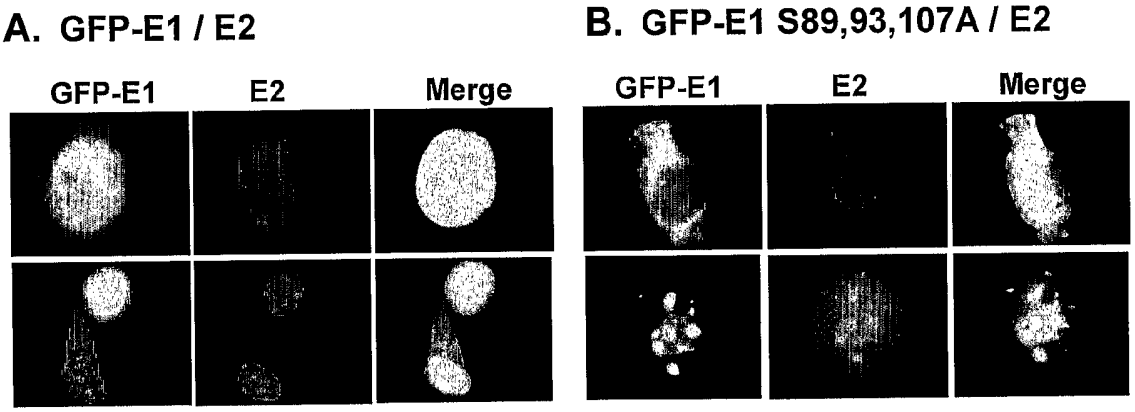


FIG. 5

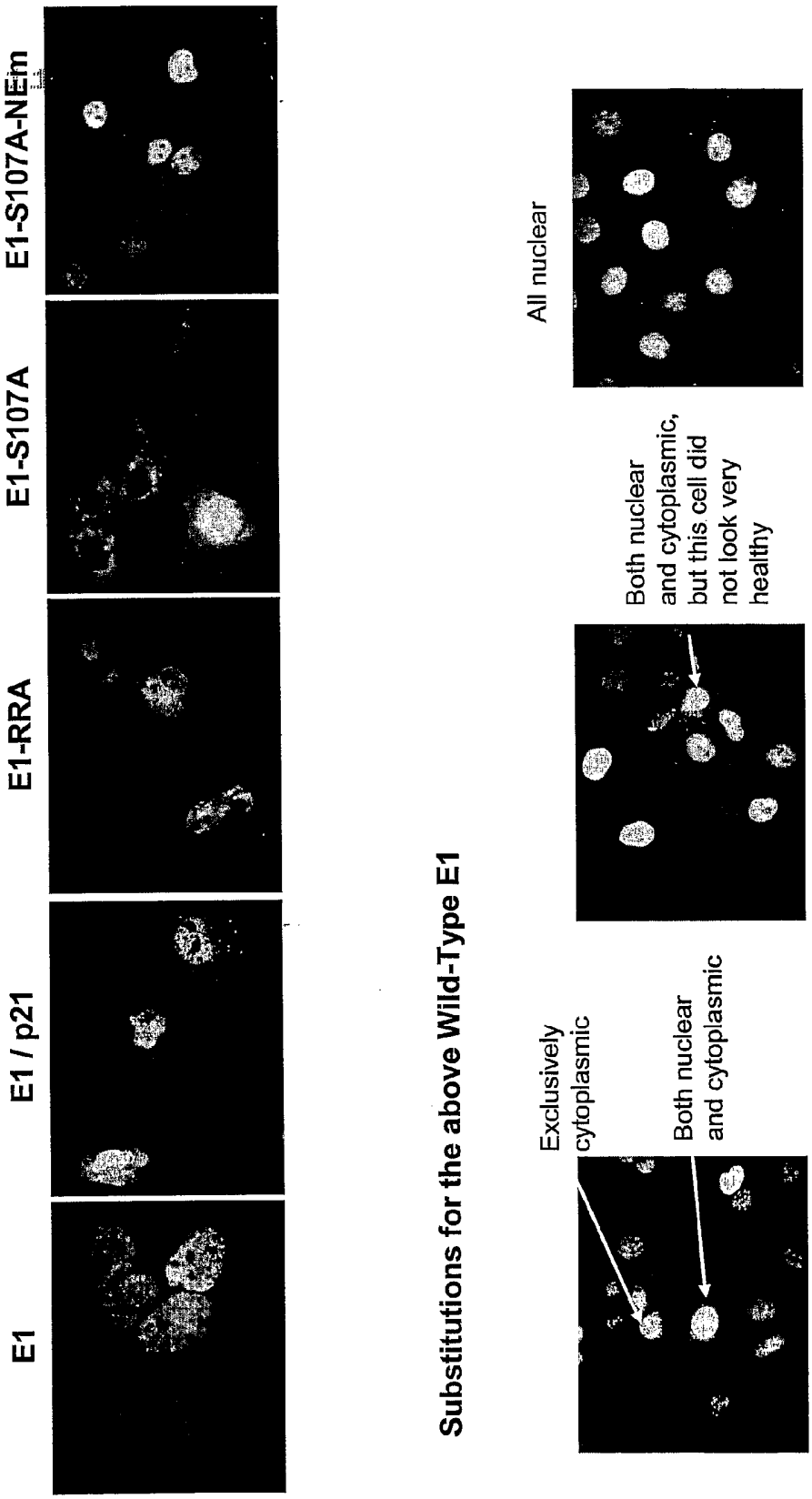

E1
E1 / p21
E1-RRA
E1-S107A
E1-S107A-NEm
Substitutions for the above Wild-Type E1
Exclusively cytoplasmic
Both nuclear and cytoplasmic
Both nuclear and cytoplasmic, but this cell did not look very healthy
All nuclear

FIG. 6

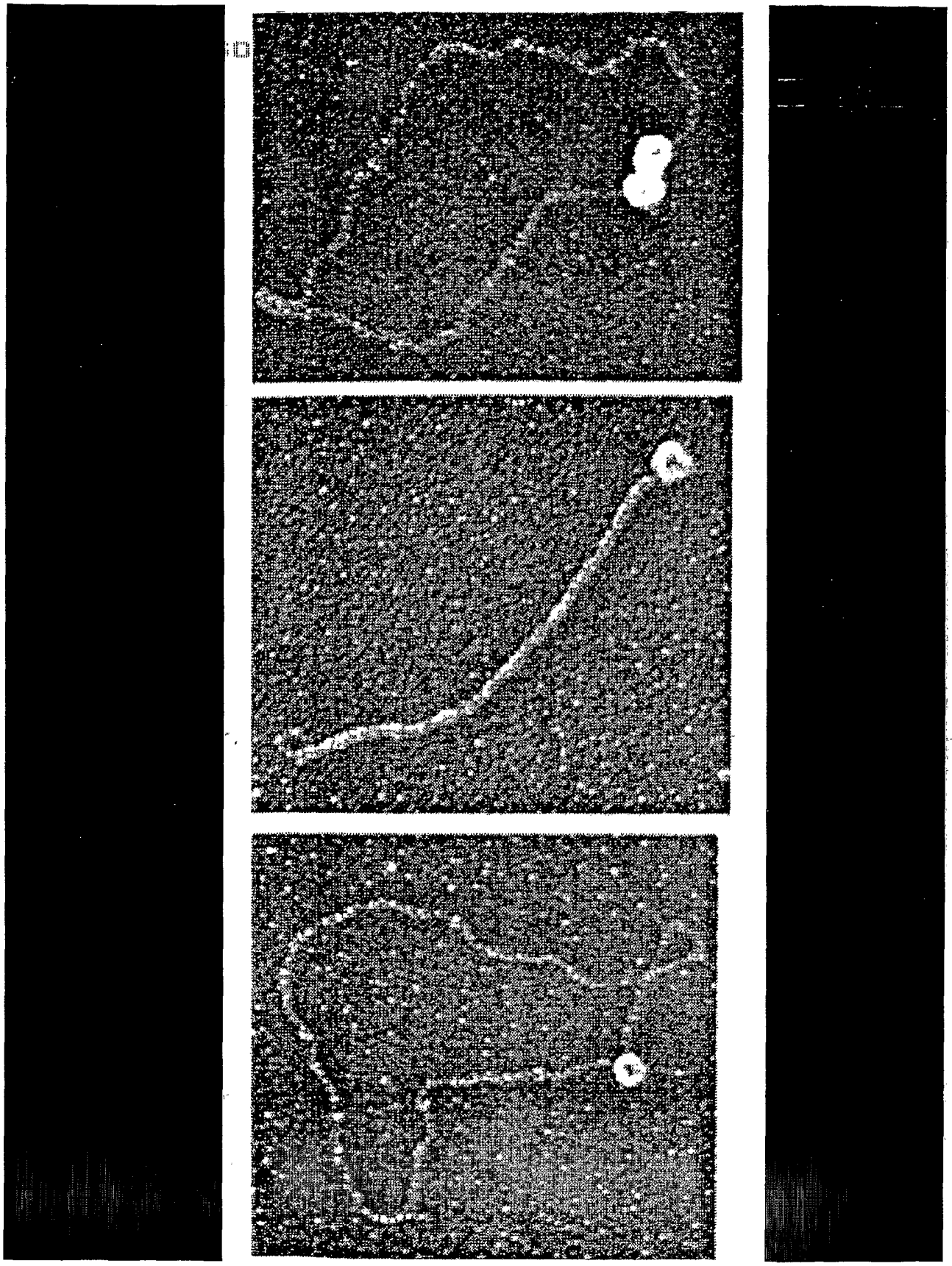


FIG. 7

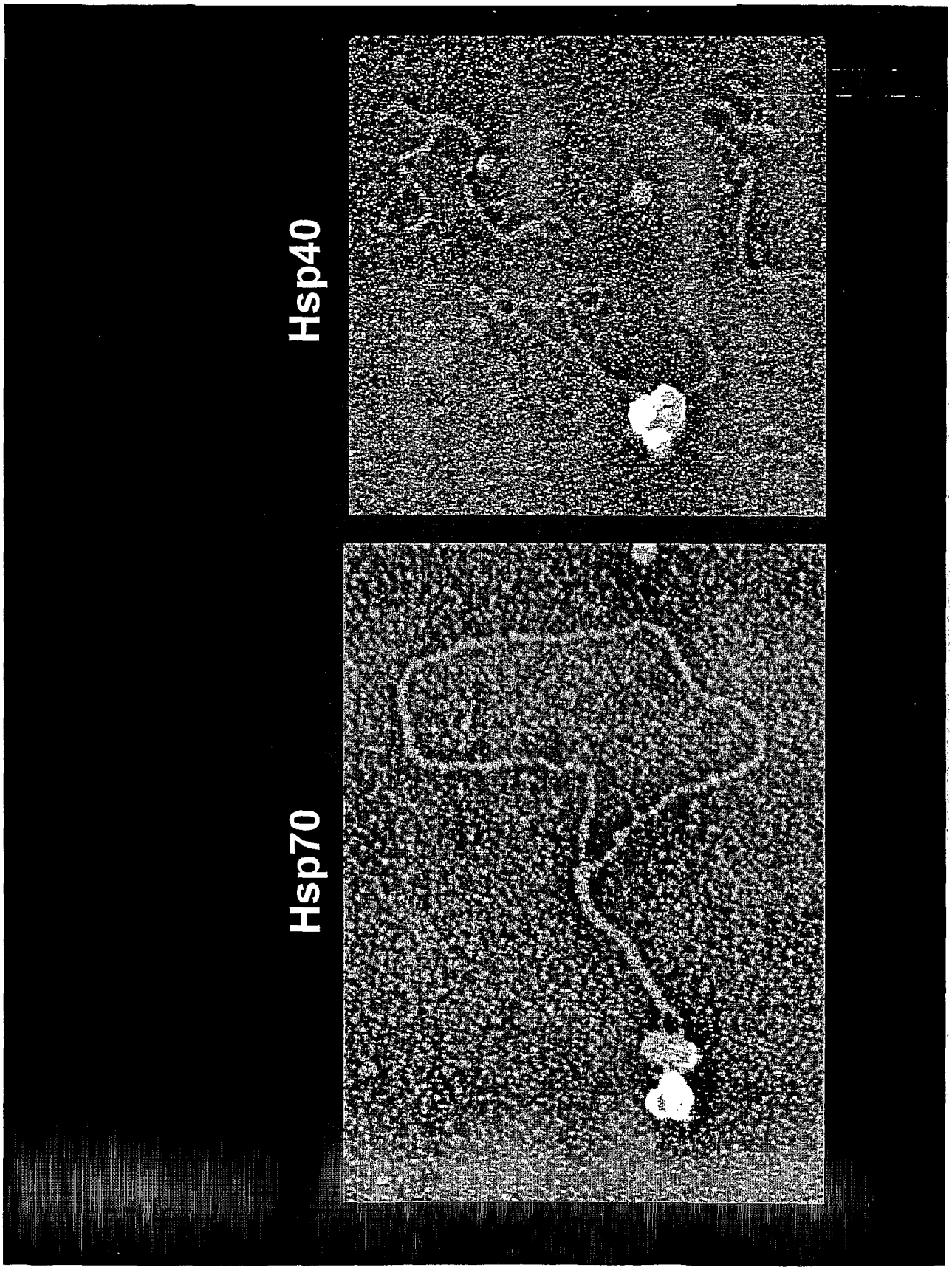


FIG. 8

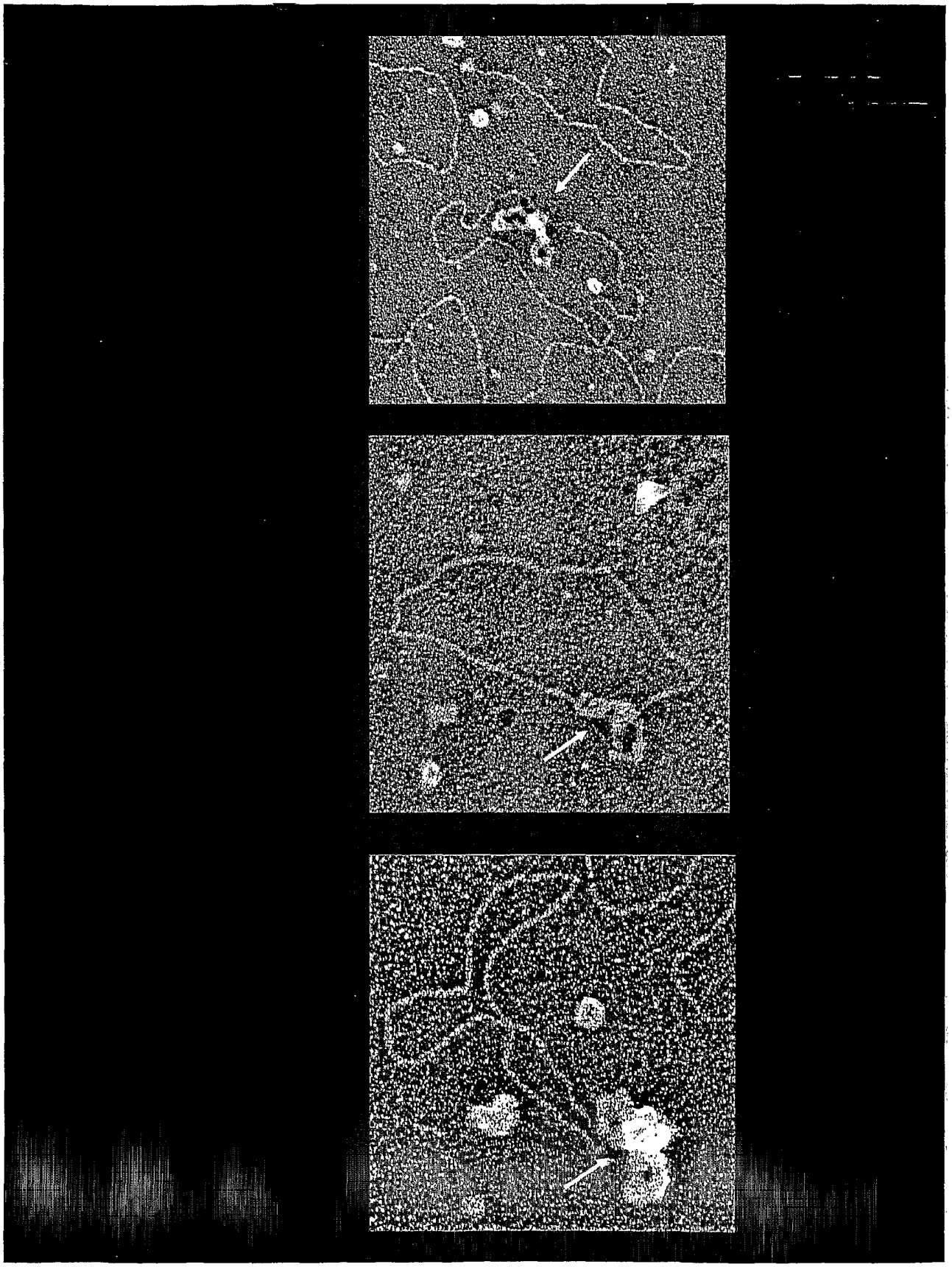


FIG. 9

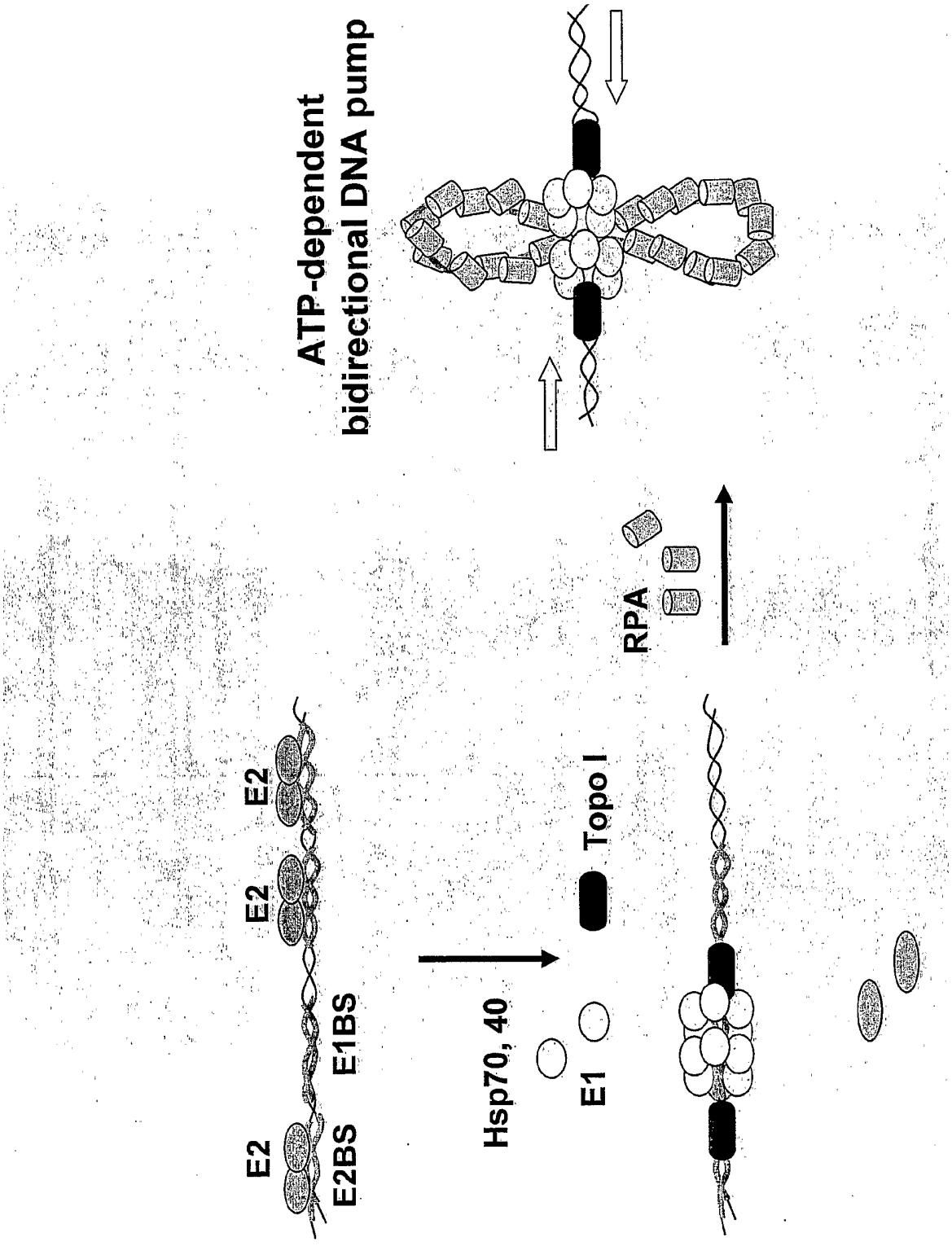

ATP-dependent
bidirectional DNA pump
RPA
Topo I
E1
Hsp70, 40
E2
E2
E2
E1BS
E2BS

FIG. 10

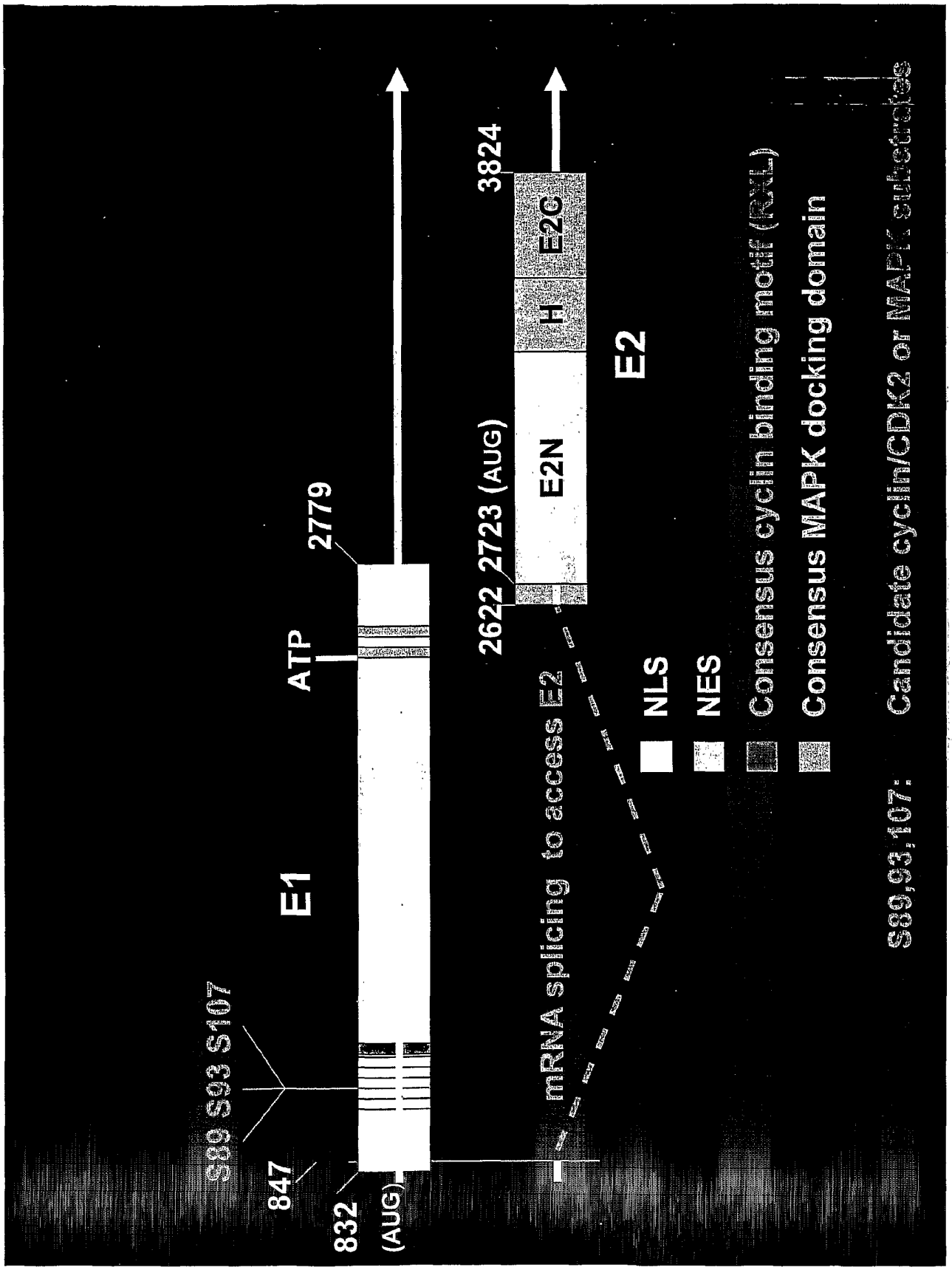

E1
ATP
2779
832
(AUG)
847
S89 S93 S107
tRNA splicing to access E2
2622 2723 (AUG)
E2N
H
E2C
E2
3824
NLS
NES
Consensus cyclin binding motif (RXL)
Consensus MAPK docking domain
S89,93,107: Candidate cyclin/CDK2 or MAPK substrates

FIG. 11

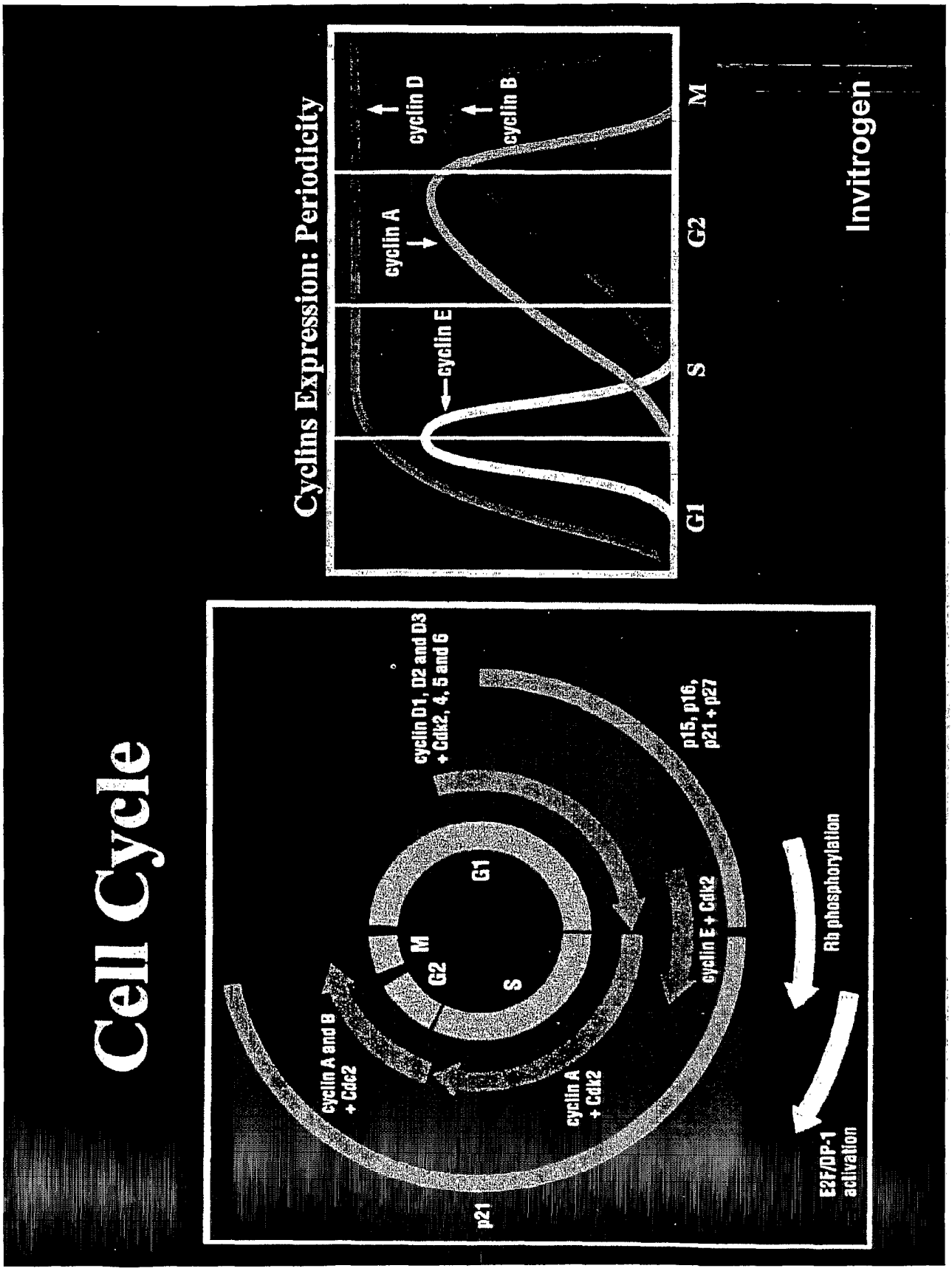

Cell Cycle
Cyclins Expression: Periodicity
cyclin A
cyclin D
cyclin B
cyclin E
G1
S
G2
M
Invitrogen
cyclin D1, D2 and D3
+ Cdk2, 4, 5 and 6
p15, p16,
p21 + p27
cyclin A and B
+ Cdc2
cyclin A
+ Cdk2
cyclin E + Cdk2
p21
Rb phosphorylation
E2F/DP-1
activation

FIG. 12

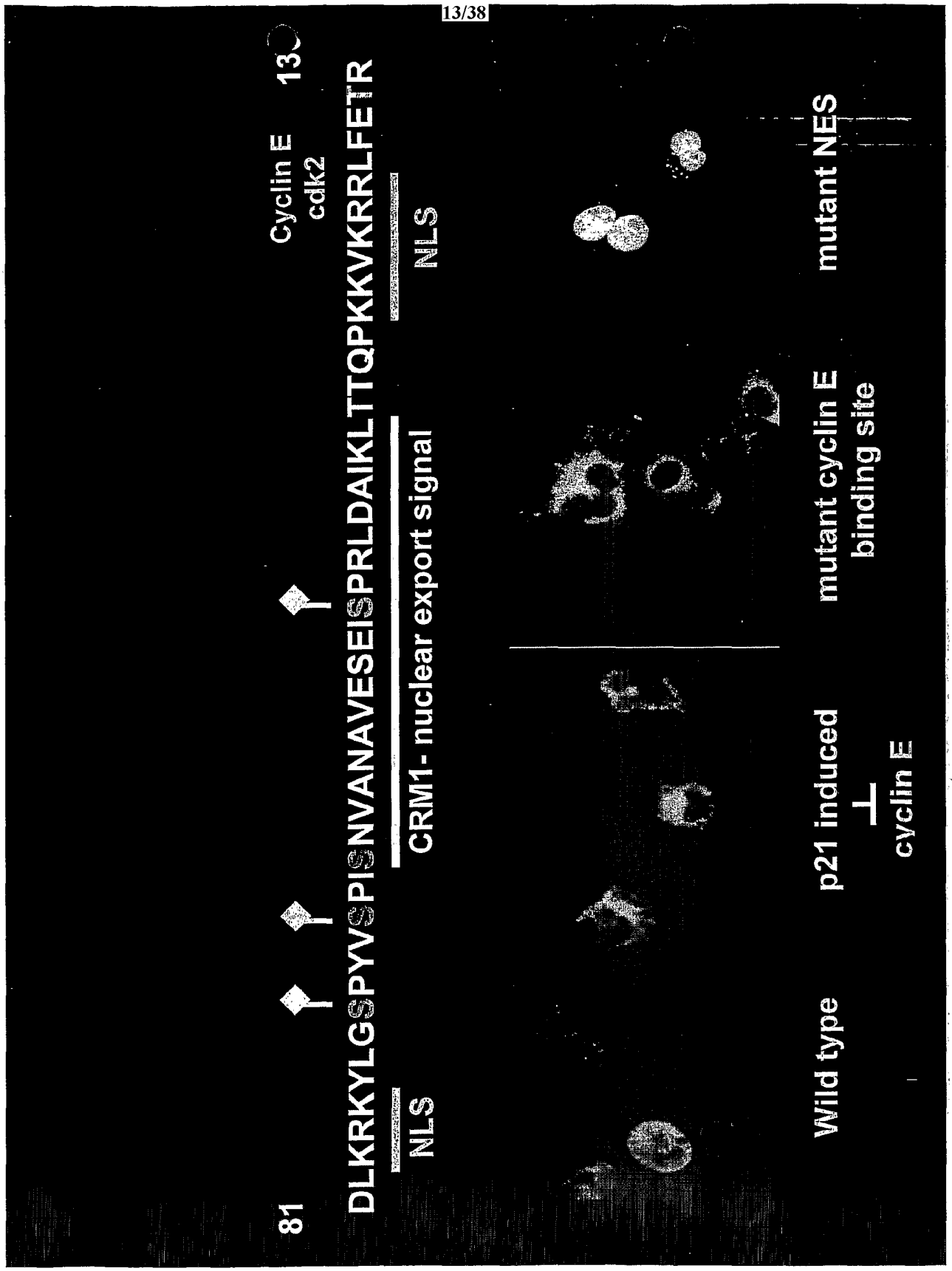

81
DLKRKYLGSPYVSPISNVANAVESEISPRLDAIKLTTQPKKVKRRLFETR
NLS
CRM1- nuclear export signal
NLS
Cyclin E
cdk2
13
Wild type
p21 induced
⊥
cyclin E
mutant cyclin E
binding site
mutant NES

FIG. 13

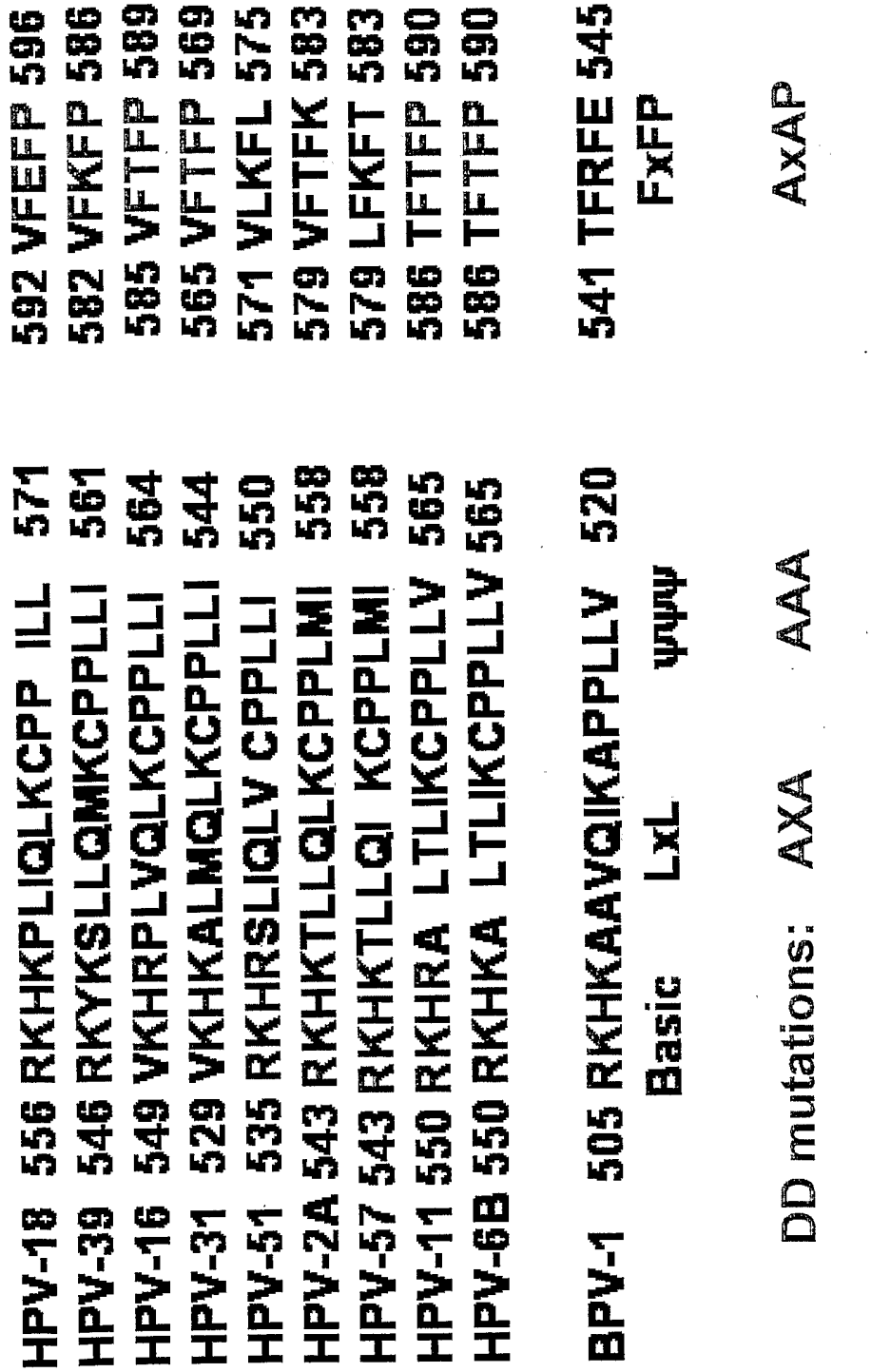

HPV-18 556 RKHKPLIQLKCPP ILL 571
HPV-39 546 RKYKSLLQMKCPPLLI 561
HPV-16 549 VKHRPLVQLKCPPLLI 564
HPV-31 529 VKHKALMQLKCPPLLI 544
HPV-51 535 RKHRSLIQLV CPPLLI 550
HPV-2A 543 RKHKTLLQLKCPPLMI 558
HPV-57 543 RKHKTLLQI KCPPLMI 558
HPV-11 550 RKHRA LTLIKCPPLLV 565
HPV-6B 550 RKHKA LTLIKCPPLLV 565
BPV-1 505 RKHKAAVQIKAPPLLV 520
Basic LxL ψψψ
DD mutations: AXA AAA
592 VFEFP 596
582 VFKFP 586
585 VFTFP 589
565 VFTFP 569
571 VLKFL 575
579 VFTFK 583
579 LFKFT 583
586 TFTFP 590
586 TFTFP 590
541 TFRFE 545
FxFP
AxAP

FIG. 14

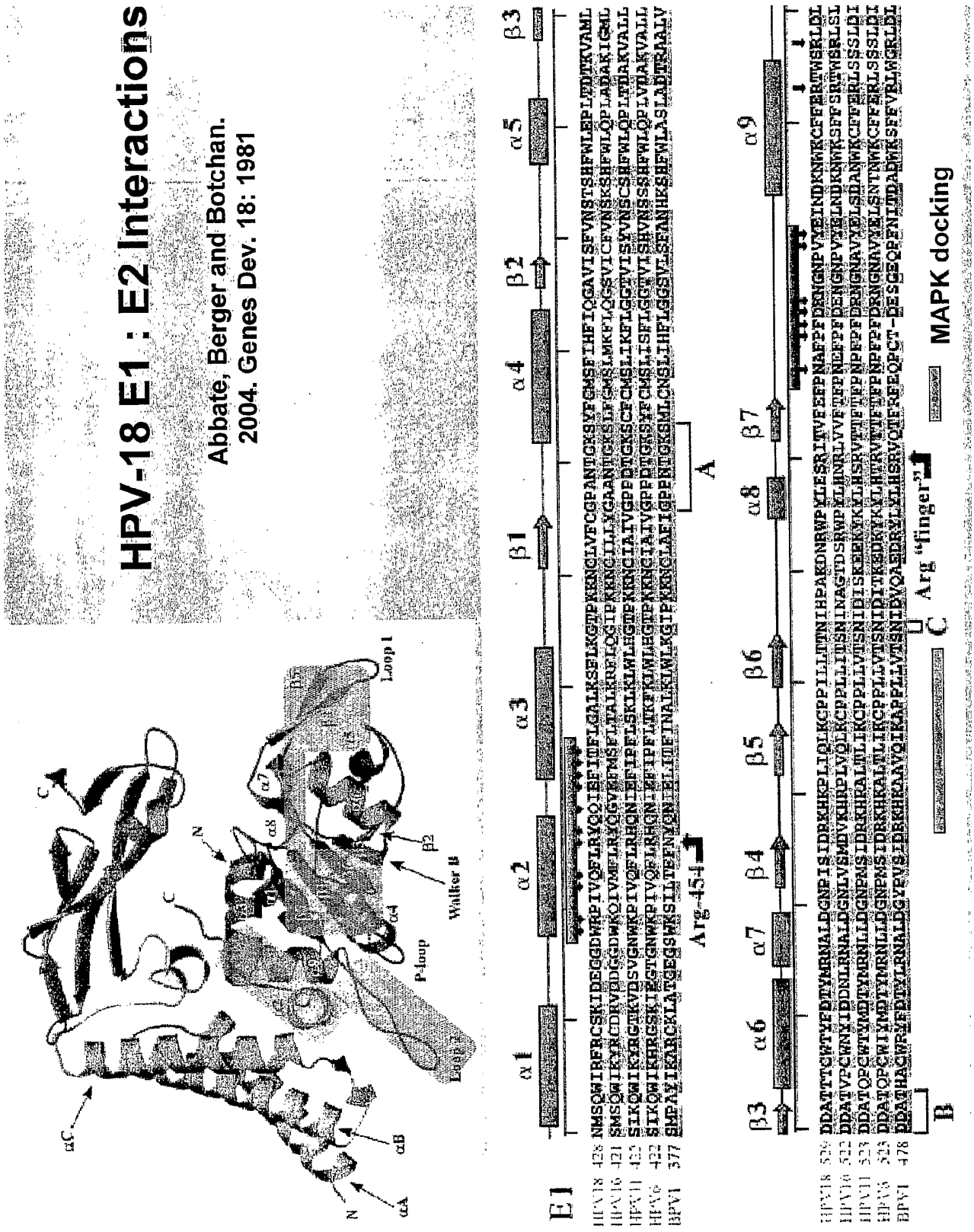

HPV-18 E1 : E2 Interactions
Abbate, Berger and Botchan.
2004. Genes Dev. 18: 1981
Loop 1
P-loop
Walker B
Arg-454
Arg "finger"
MAPK docking
E1
A
B
C

FIG. 15

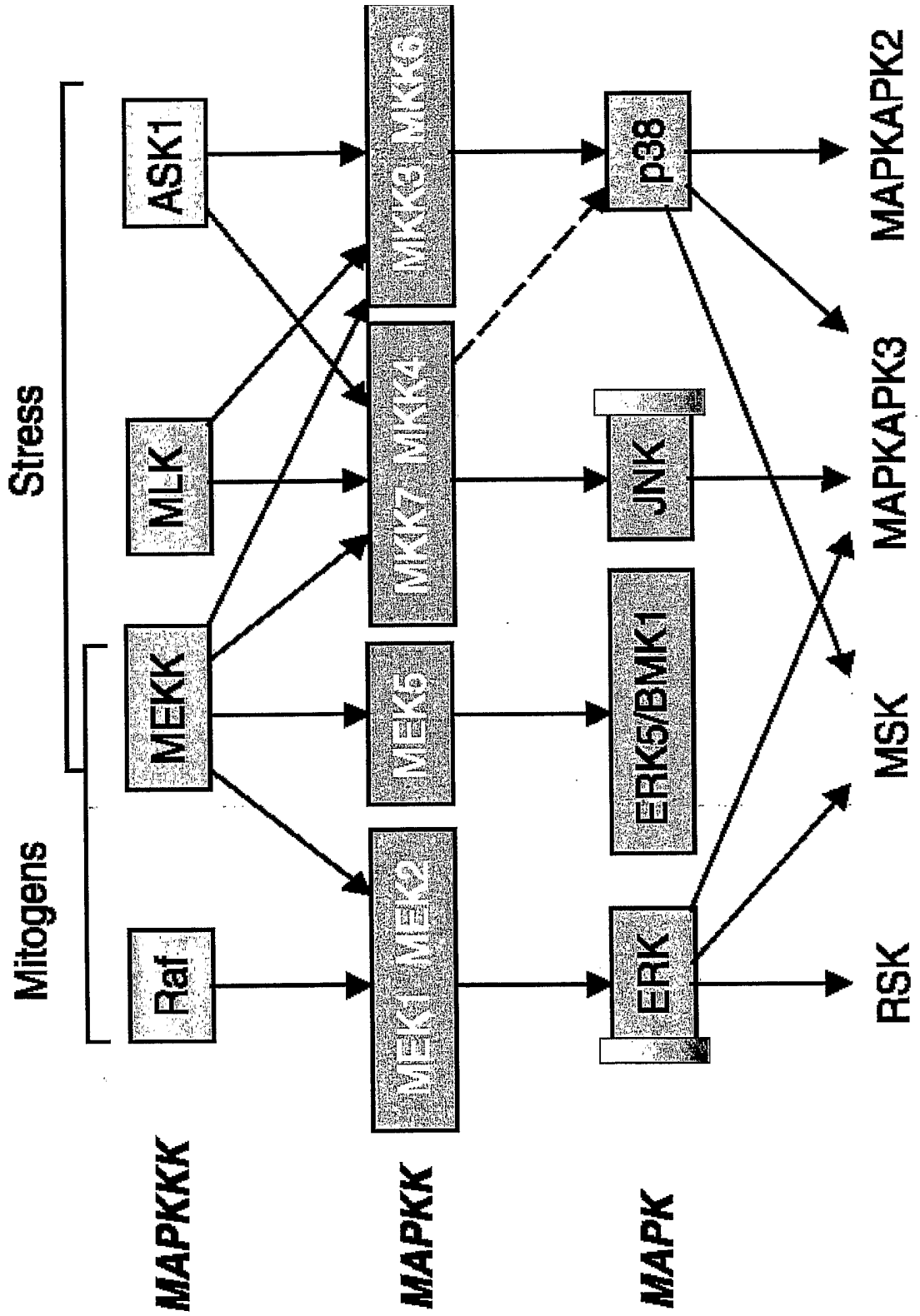

Stress
Mitogens
ASK1
MLK
MEKK
Raf
MKK3/MKK6
MKK4/MKK7
MEK5
MEK1/MEK2
p38
JNK
ERK5/BMK1
ERK
MAPKAPK2
MAPKAPK3
MSK
RSK
MAPKKK
MAPKK
MAPK

FIG. 16

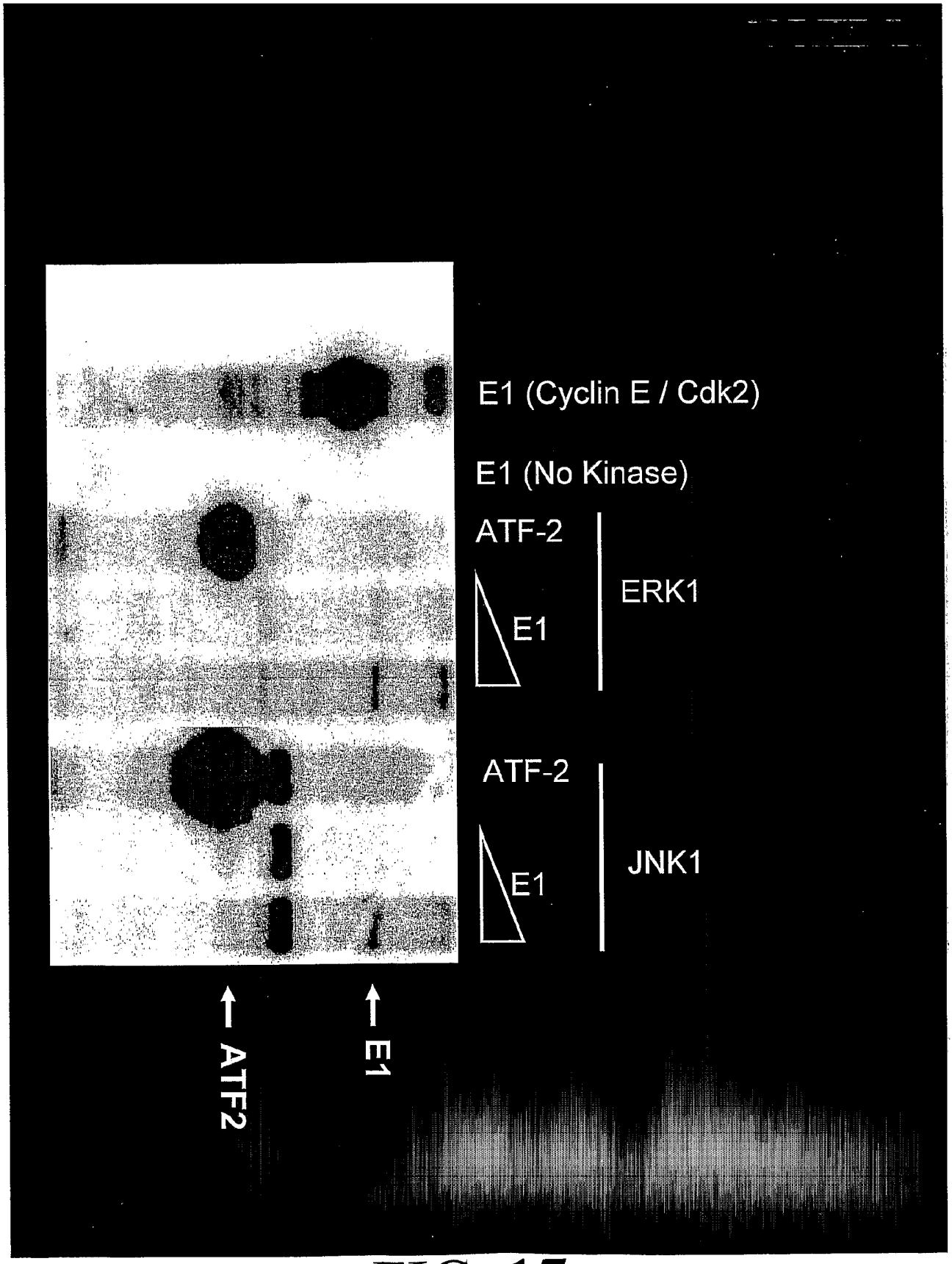
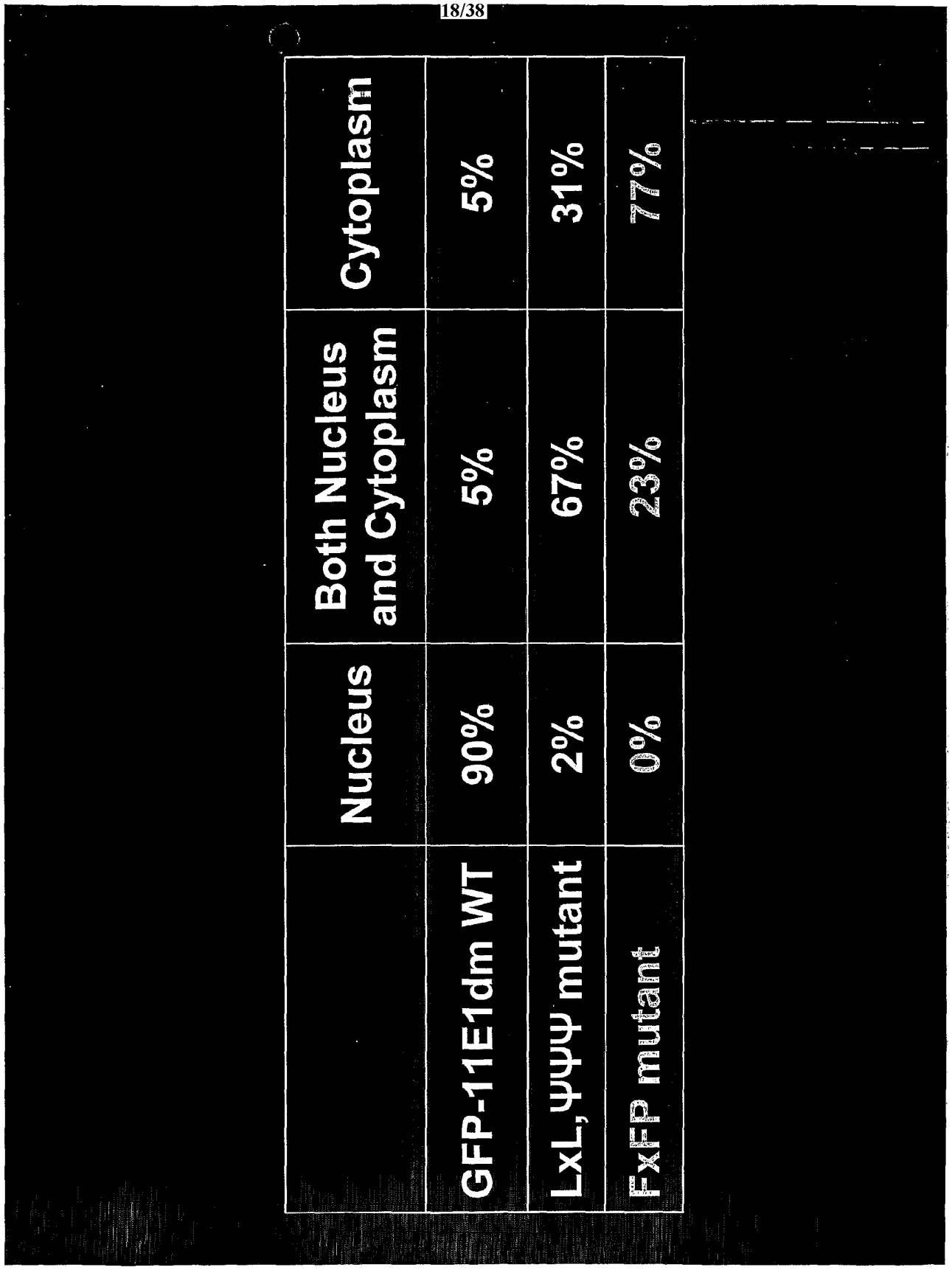


FIG. 17



	Nucleus	Both Nucleus and Cytoplasm	Cytoplasm
GFP-11E1dm WT	90%	5%	5%
LxL, ΨΨΨ mutant	2%	67%	31%
FxFP mutant	0%	23%	77%

FIG. 18

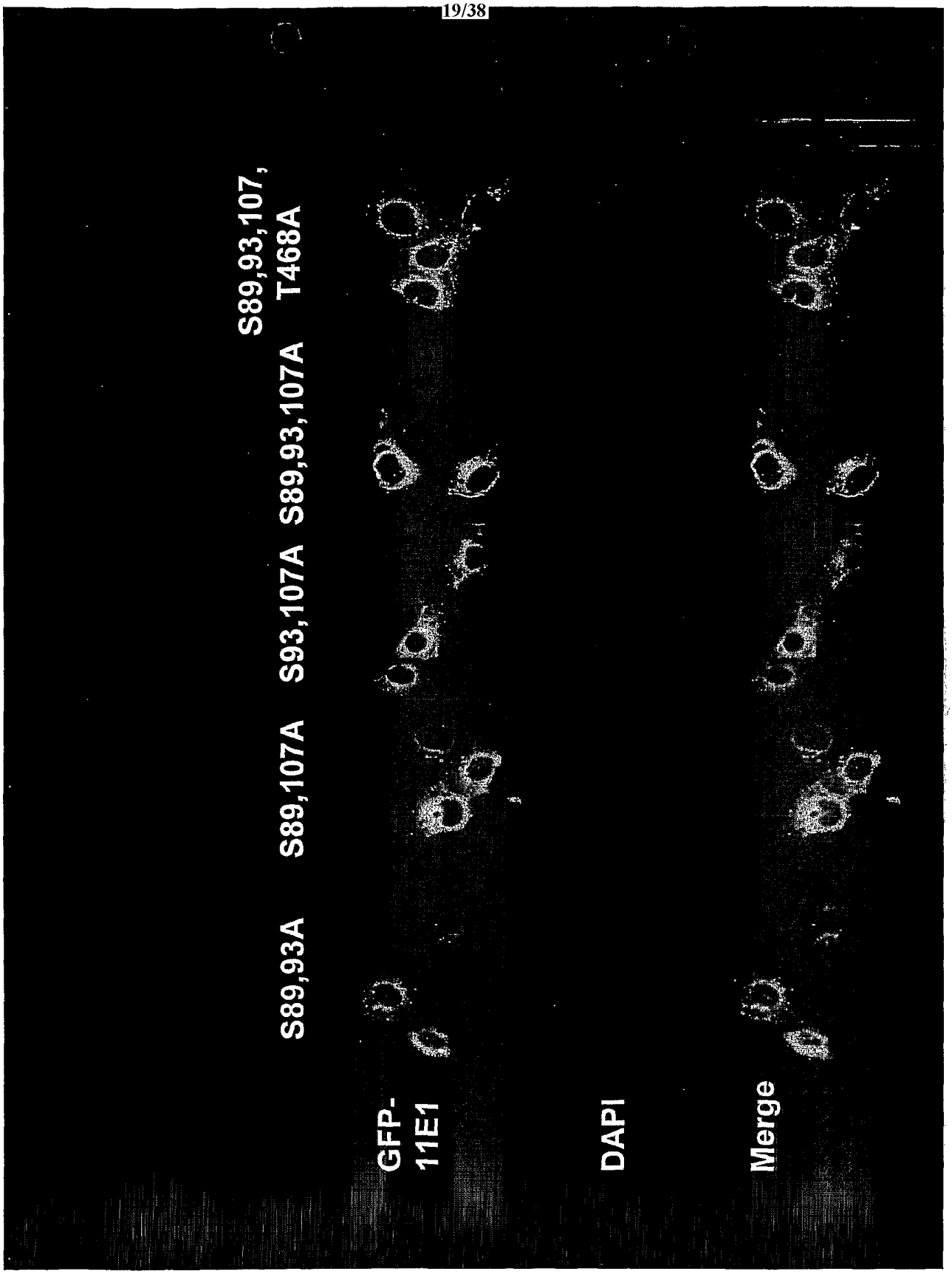

S89,93A
S89,107A
S93,107A
S89,93,107A
S89,93,107, T468A
GFP-11E1
DAPI
Merge

FIG. 19

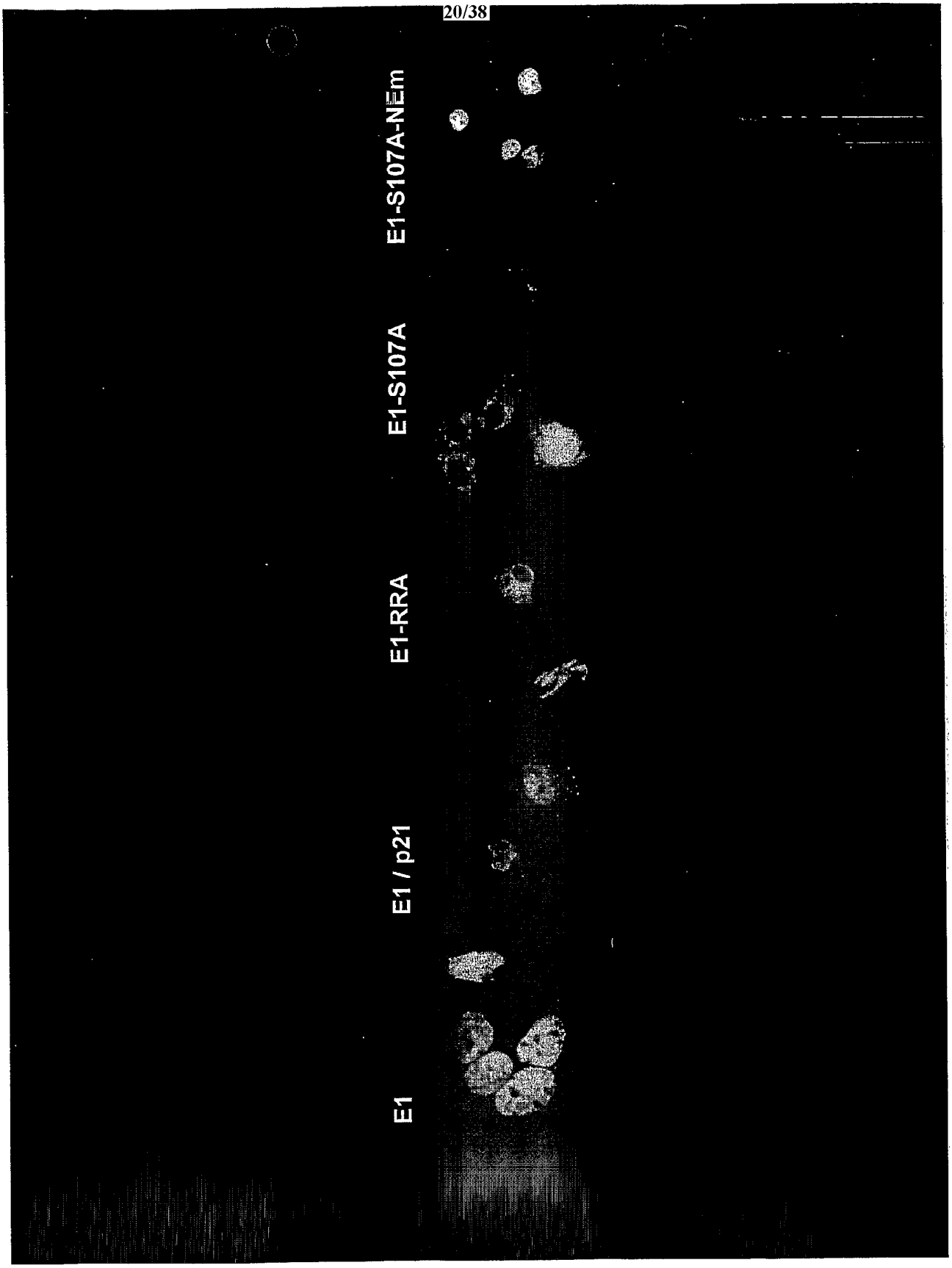


FIG. 20

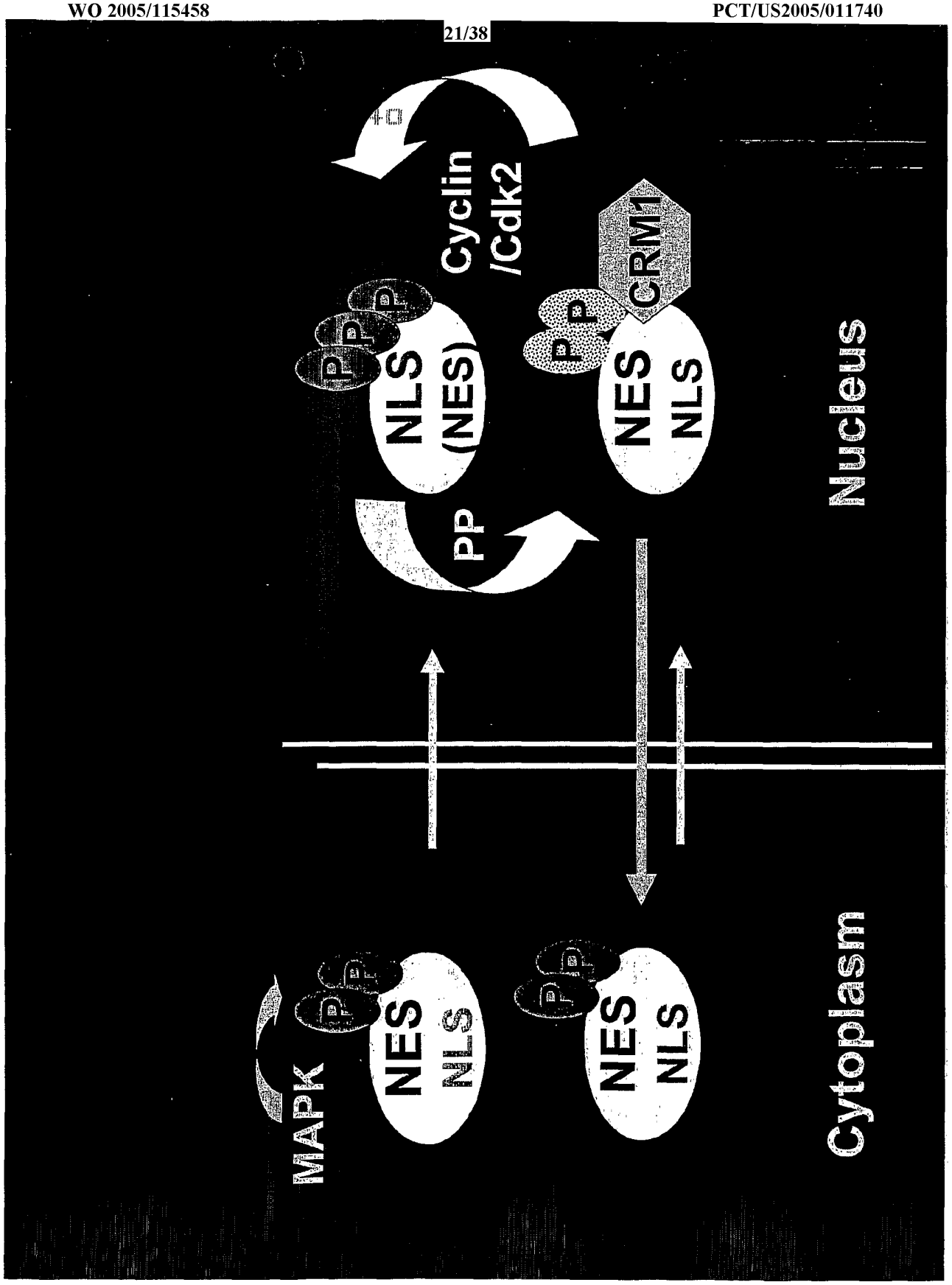

Cyclin
/Cdk2
CRM1
NLS
(NES)
NES
NLS
PP
Nucleus
MAPK
NES
NLS
NES
NLS
Cytoplasm

FIG. 21

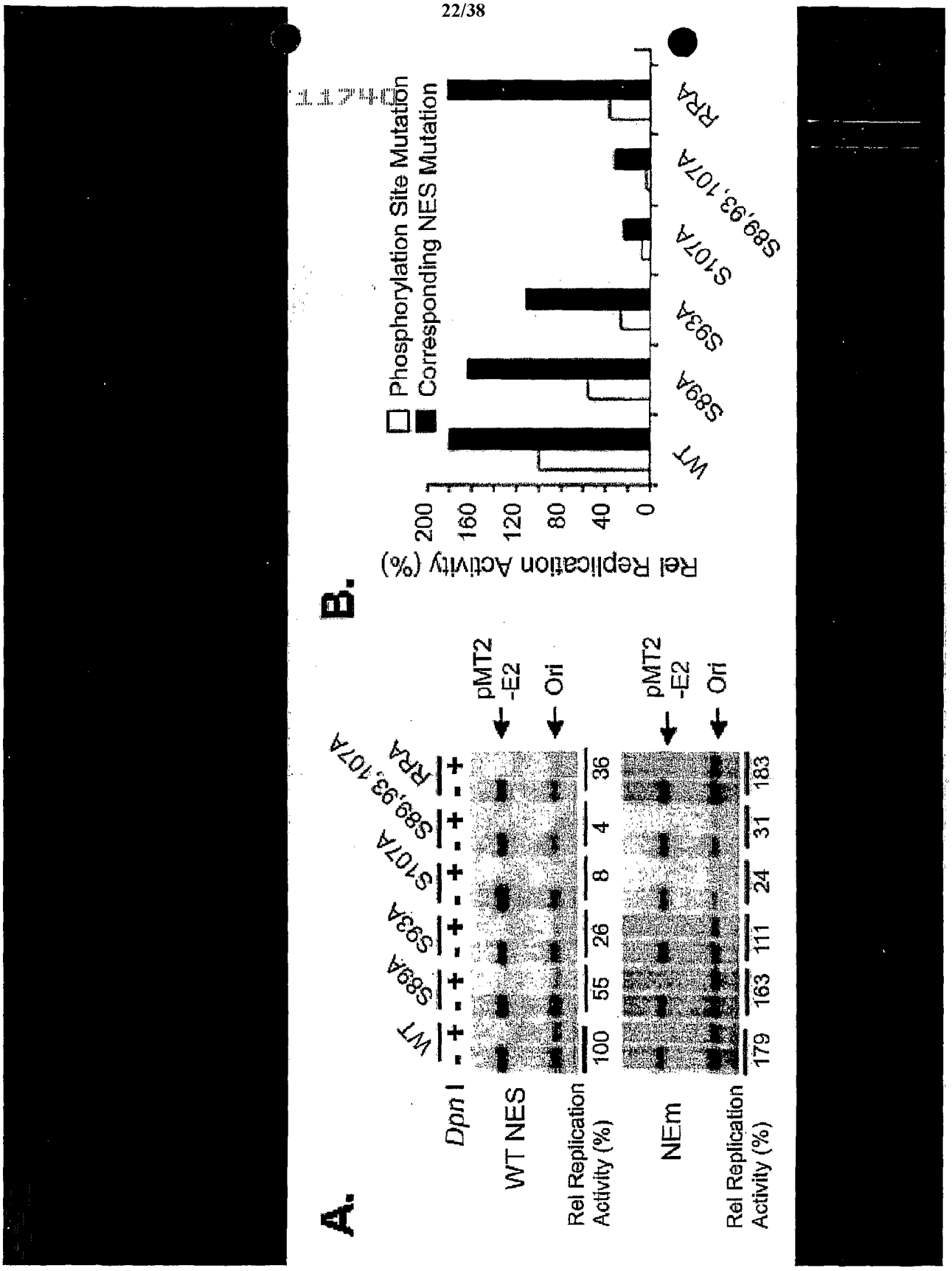

A.
Dpn I
WT
S89A
S93A
S107A
S89,93,107A
RRA
pMT2
-E2
Ori
WT NES
Rel Replication Activity (%)
100
55
26
8
4
36
NEm
179
163
111
24
31
183
B.
Phosphorylation Site Mutation
Corresponding NES Mutation
Rel Replication Activity (%)
200
160
120
80
40
0
WT
S89A
S93A
S107A
S89,93,107A
RRA

FIG. 22

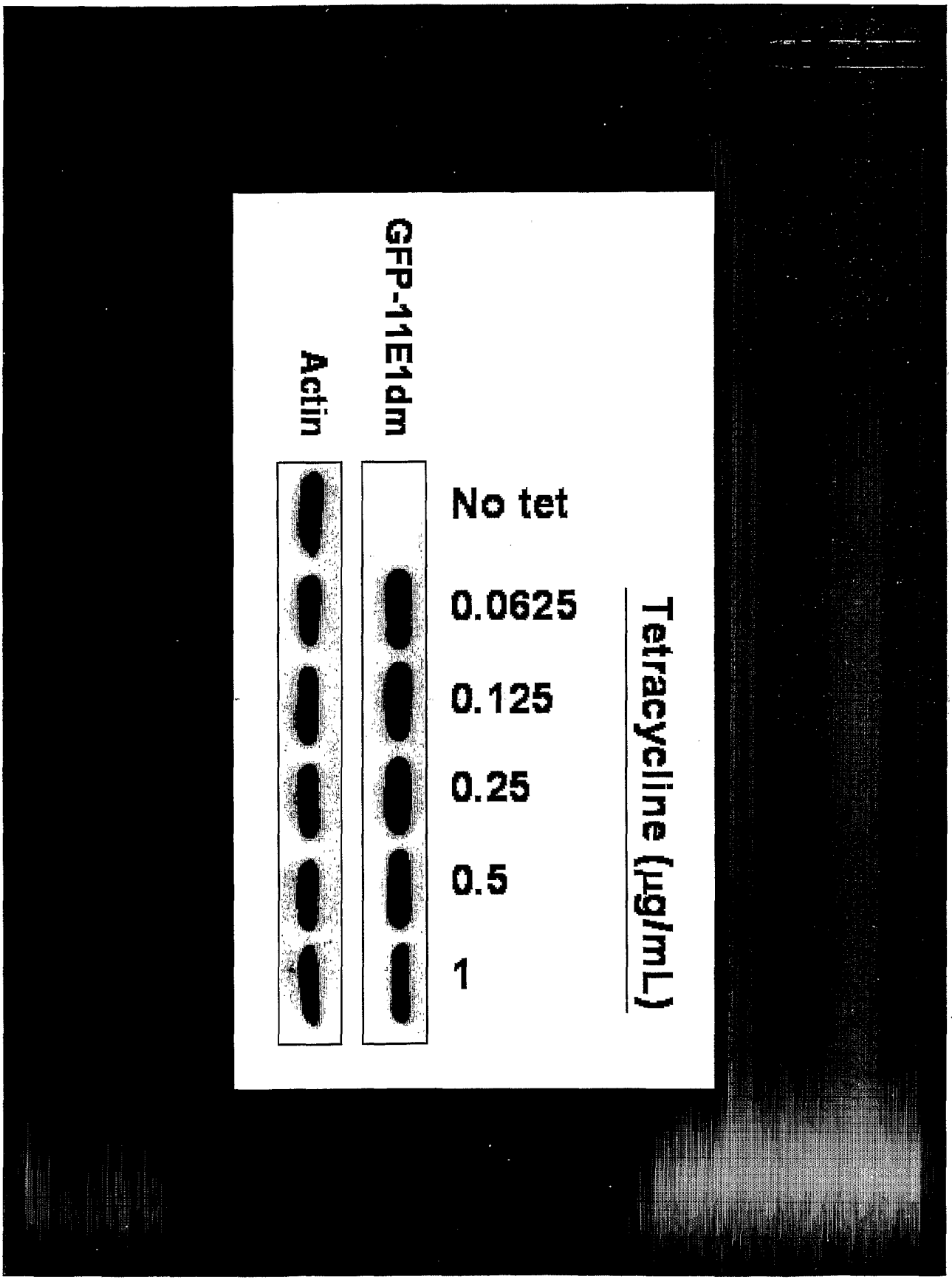

GFP-11E1dm
Actin
No tet
0.0625
0.125
0.25
0.5
1
Tetracycline (µg/mL)

FIG. 23

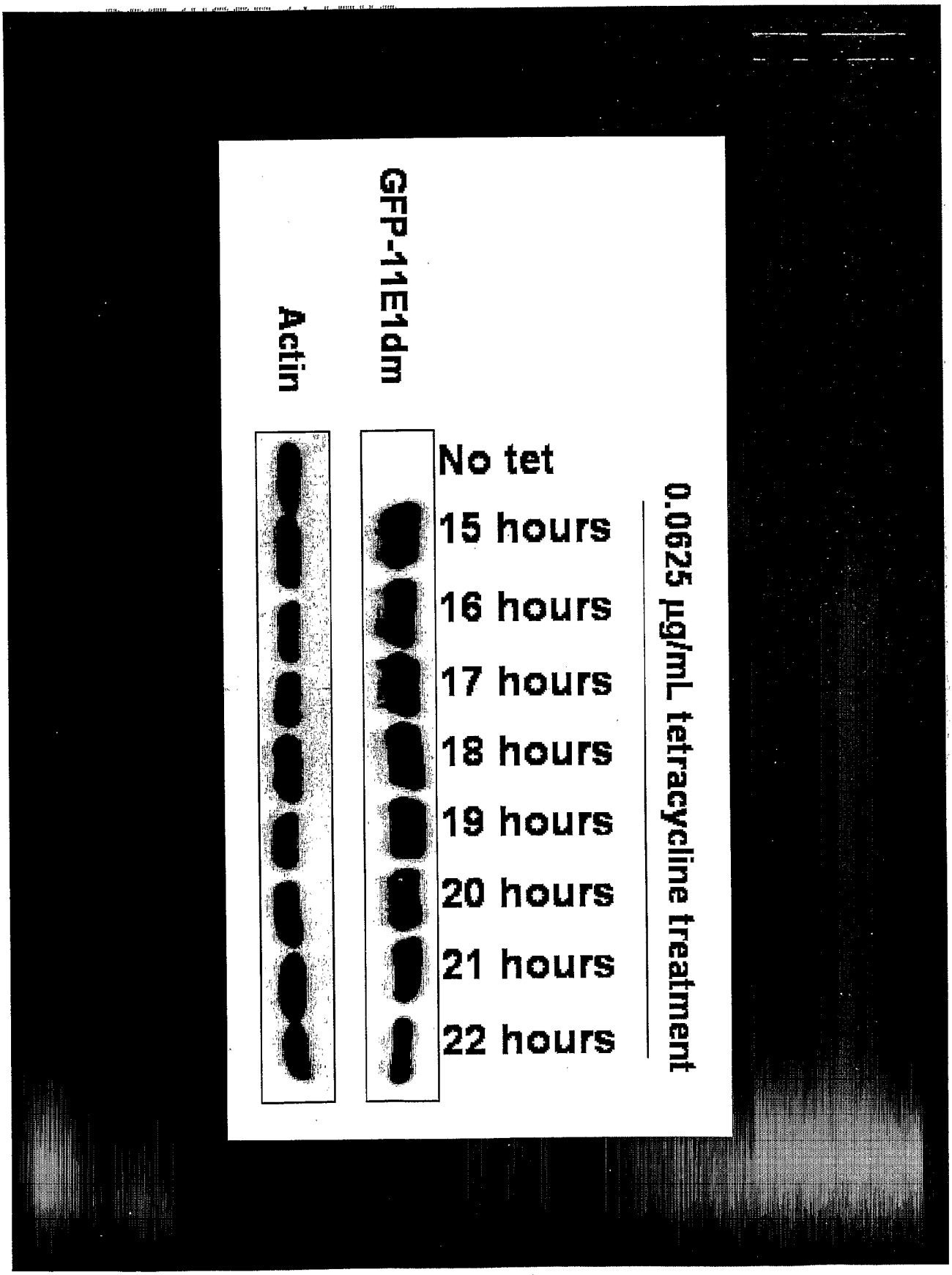

GFP-11E1dm
Actin
No tet
15 hours
16 hours
17 hours
18 hours
19 hours
20 hours
21 hours
22 hours
0.0625 µg/mL tetracycline treatment

FIG. 24

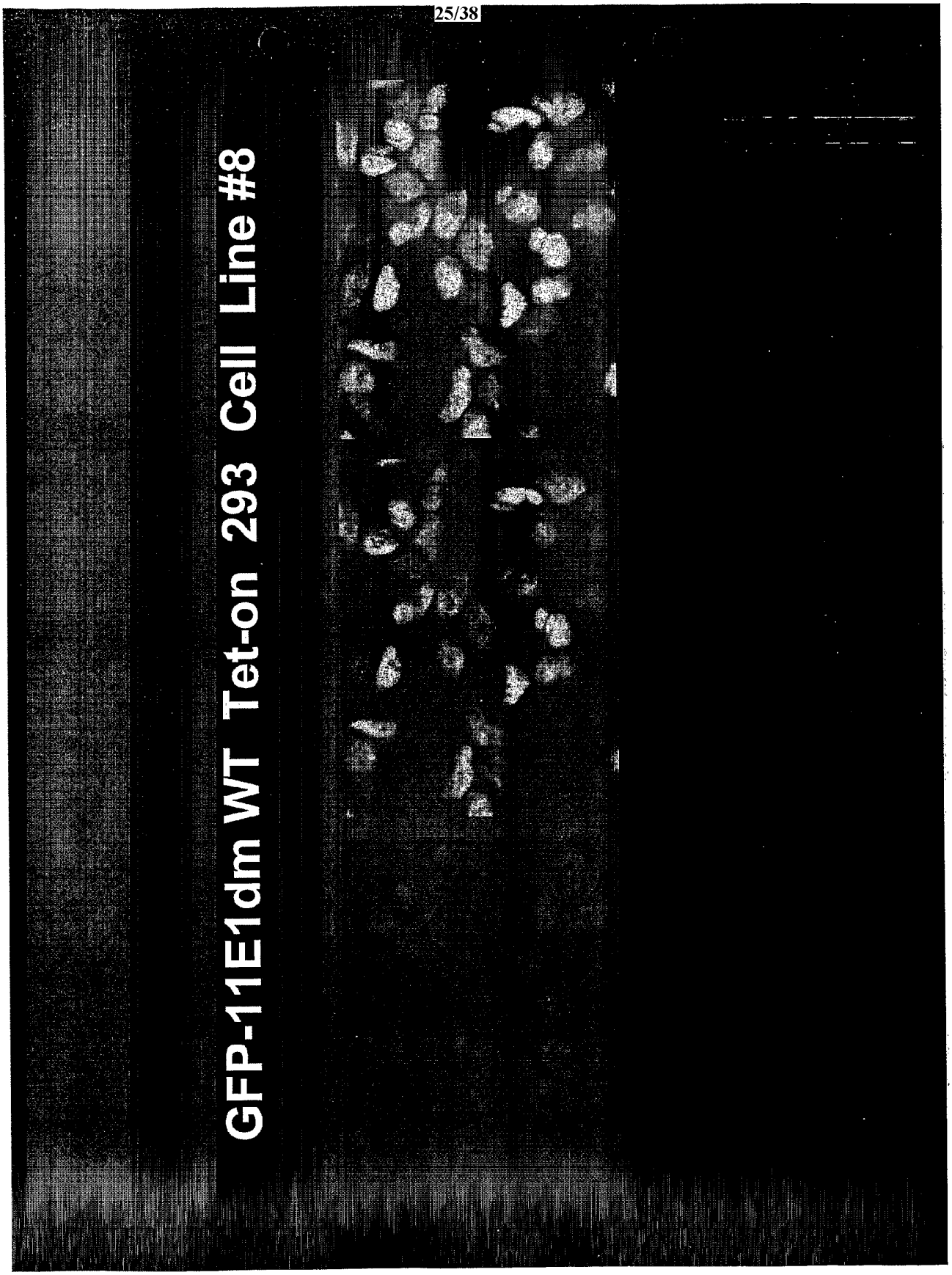


FIG. 25

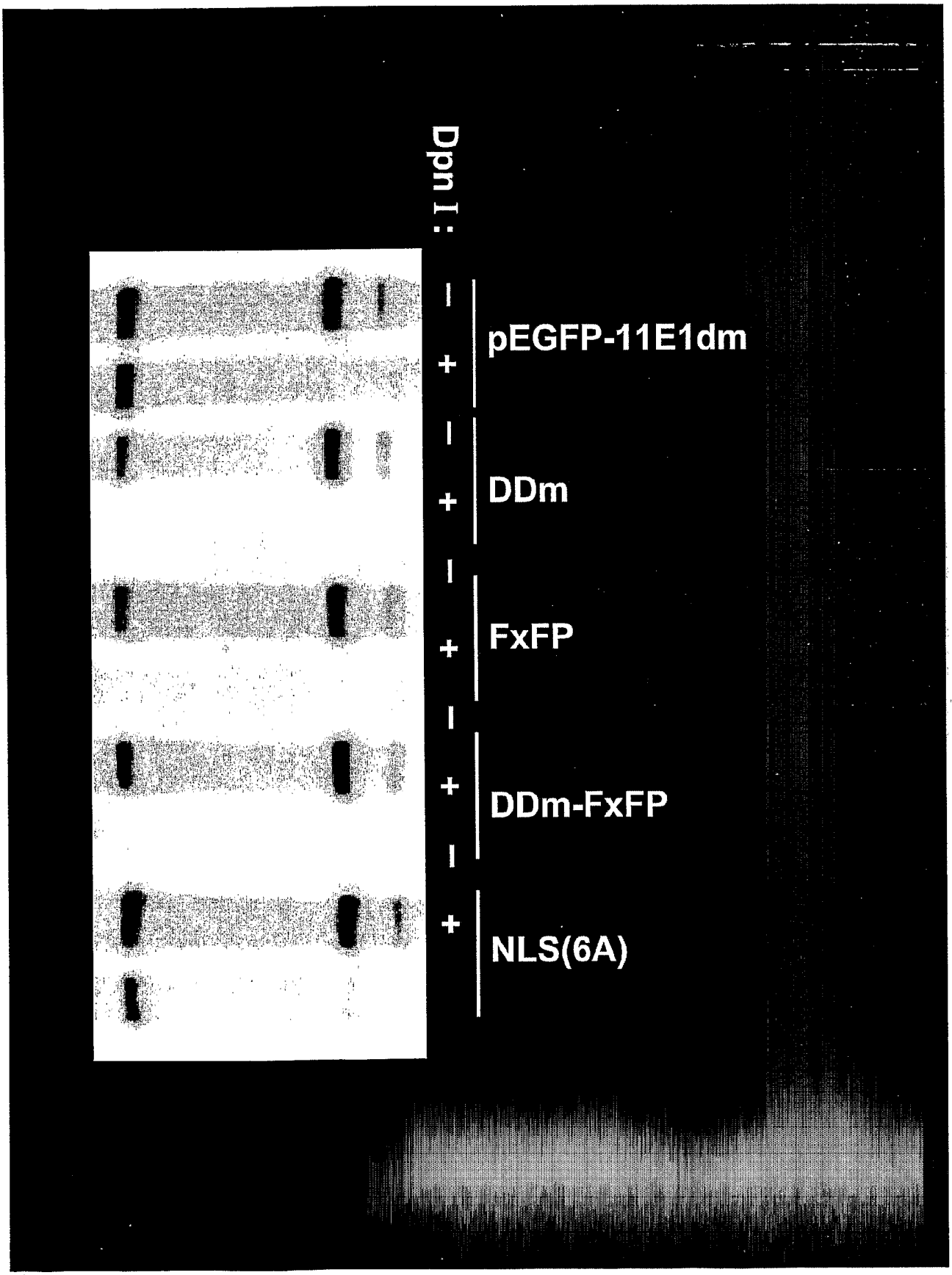


FIG. 26

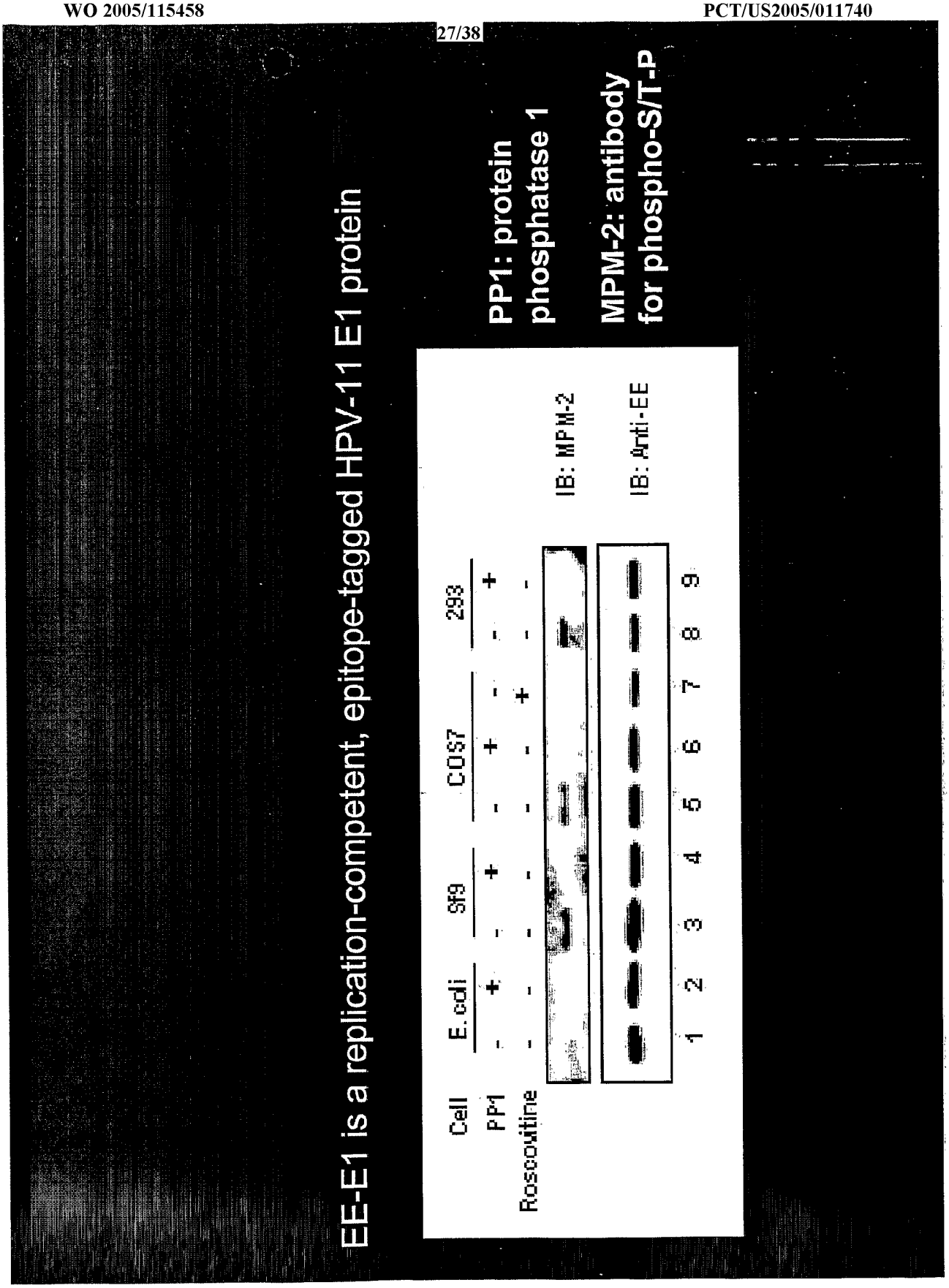

EE-E1 is a replication-competent, epitope-tagged HPV-11 E1 protein
PP1: protein phosphatase 1
MPM-2: antibody for phospho-S/T-P
Cell
PP1
Roscovitine
E. coli
Sf9
COS7
293
IB: MPM-2
IB: Anti-EE
1 2 3 4 5 6 7 8 9

FIG. 27

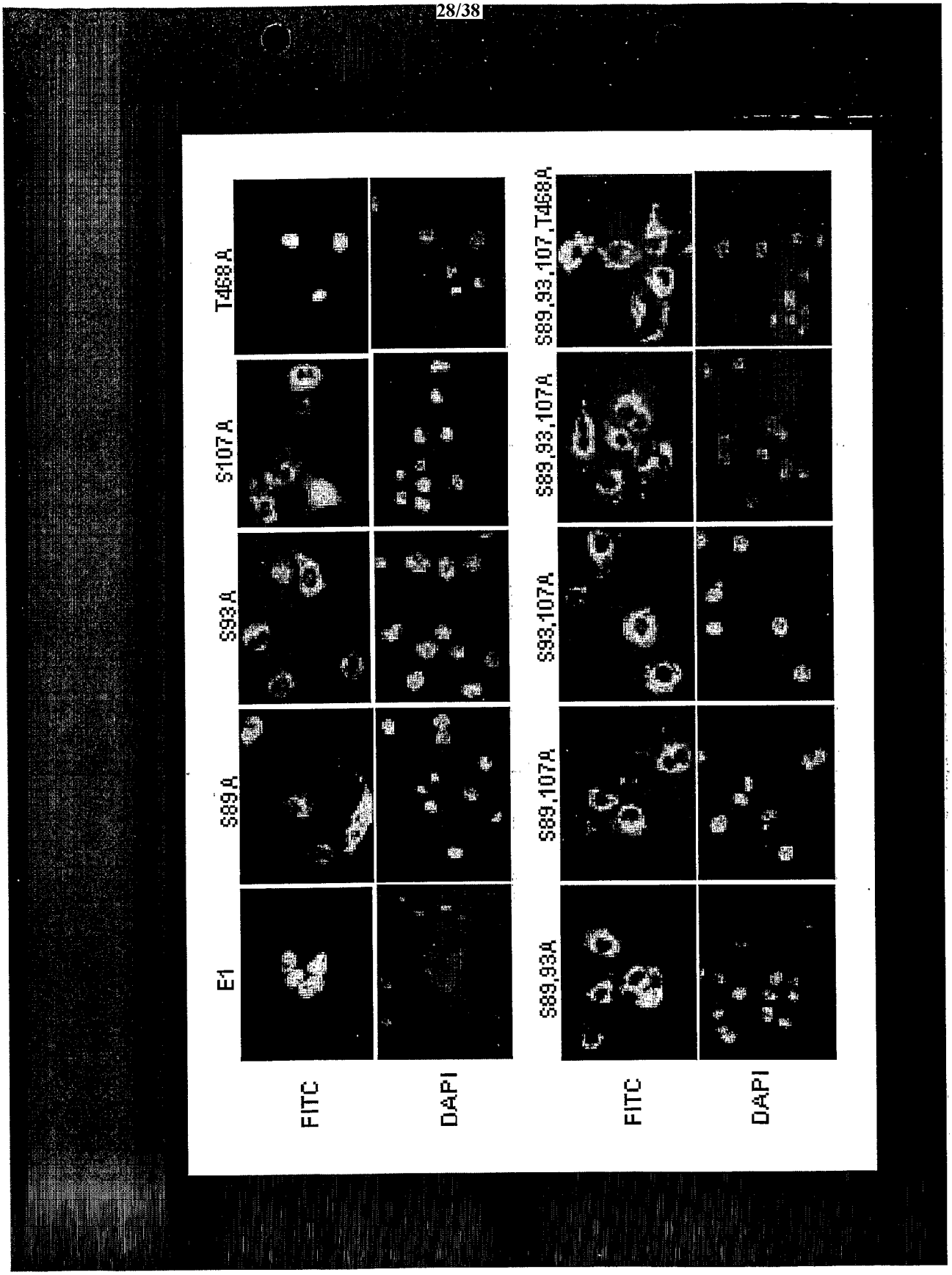

E1
S89A
S93A
S107A
T468A
S89,93A
S89,107A
S93,107A
S89,93,107A
S89,93,107,T468A
FITC
DAPI
FITC
DAPI

FIG. 28

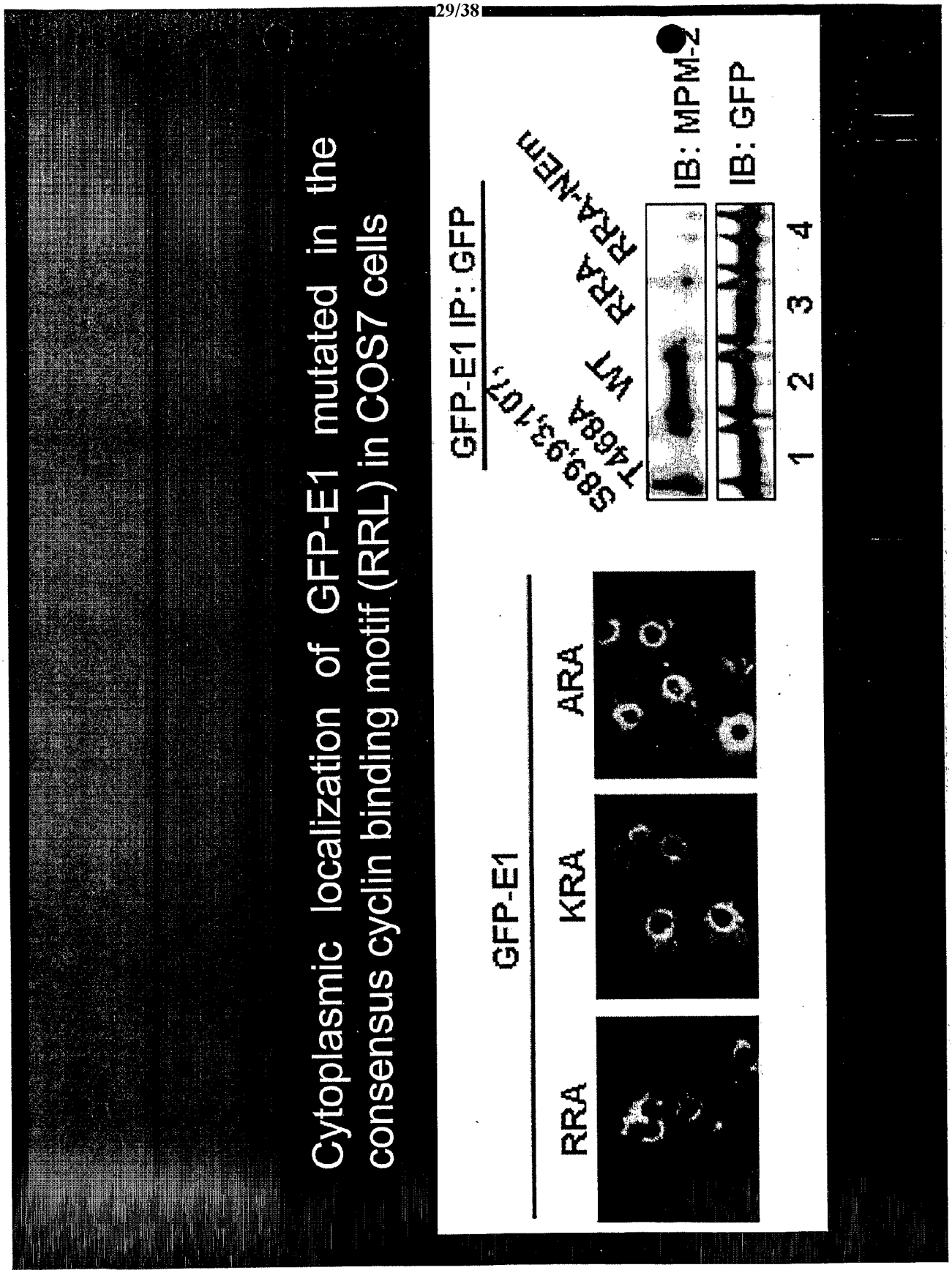

Cytoplasmic localization of GFP-E1 mutated in the consensus cyclin binding motif (RRL) in COS7 cells
GFP-E1
RRA
KRA
ARA
GFP-E1 IP: GFP
S89,93,107A
T468A
WT
RRA
RRA.NEM
IB: MPM-2
IB: GFP
1 2 3 4

FIG. 29

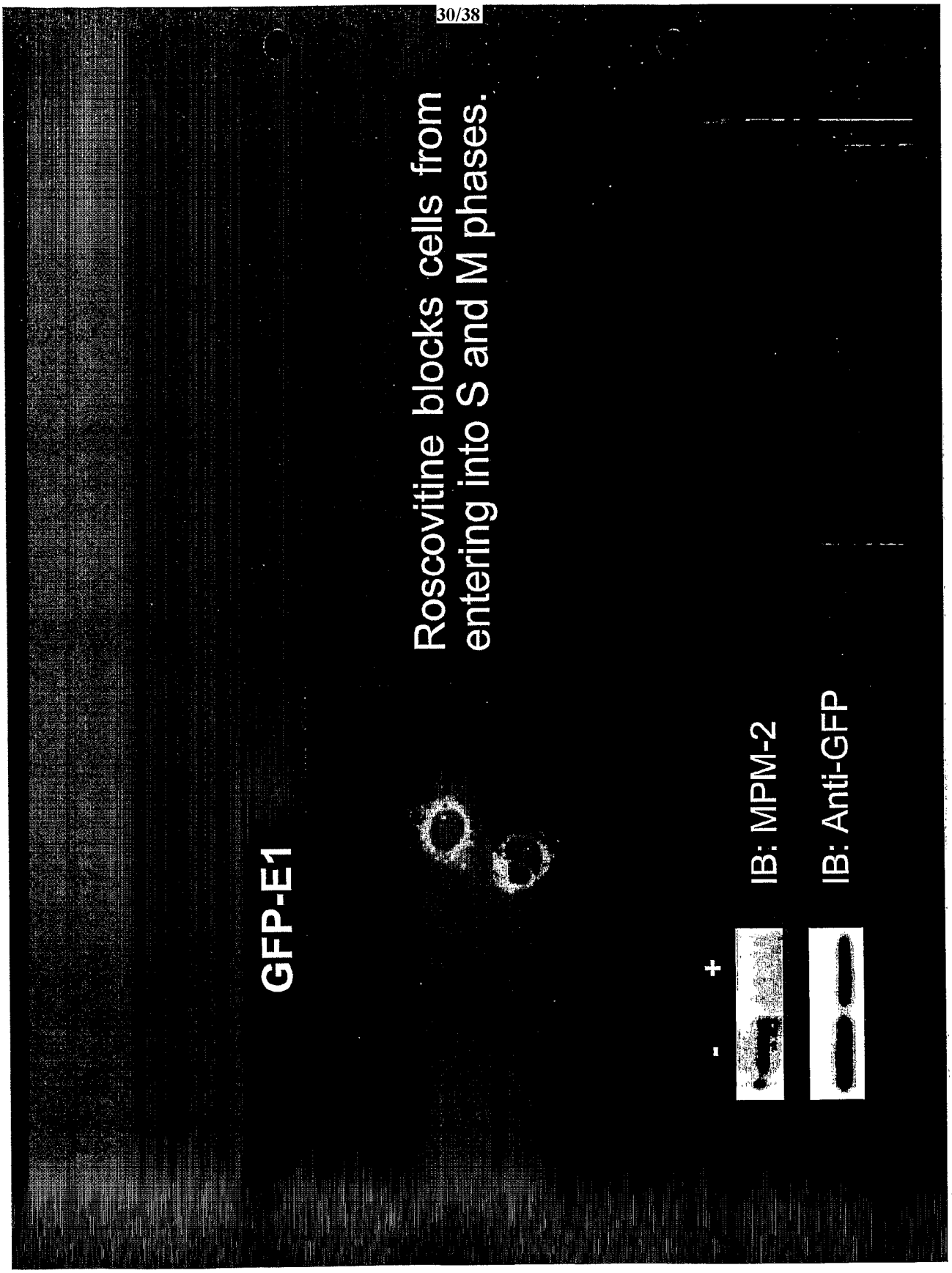

GFP-E1
Roscovitine blocks cells from entering into S and M phases.
- +
IB: MPM-2
IB: Anti-GFP

FIG. 30

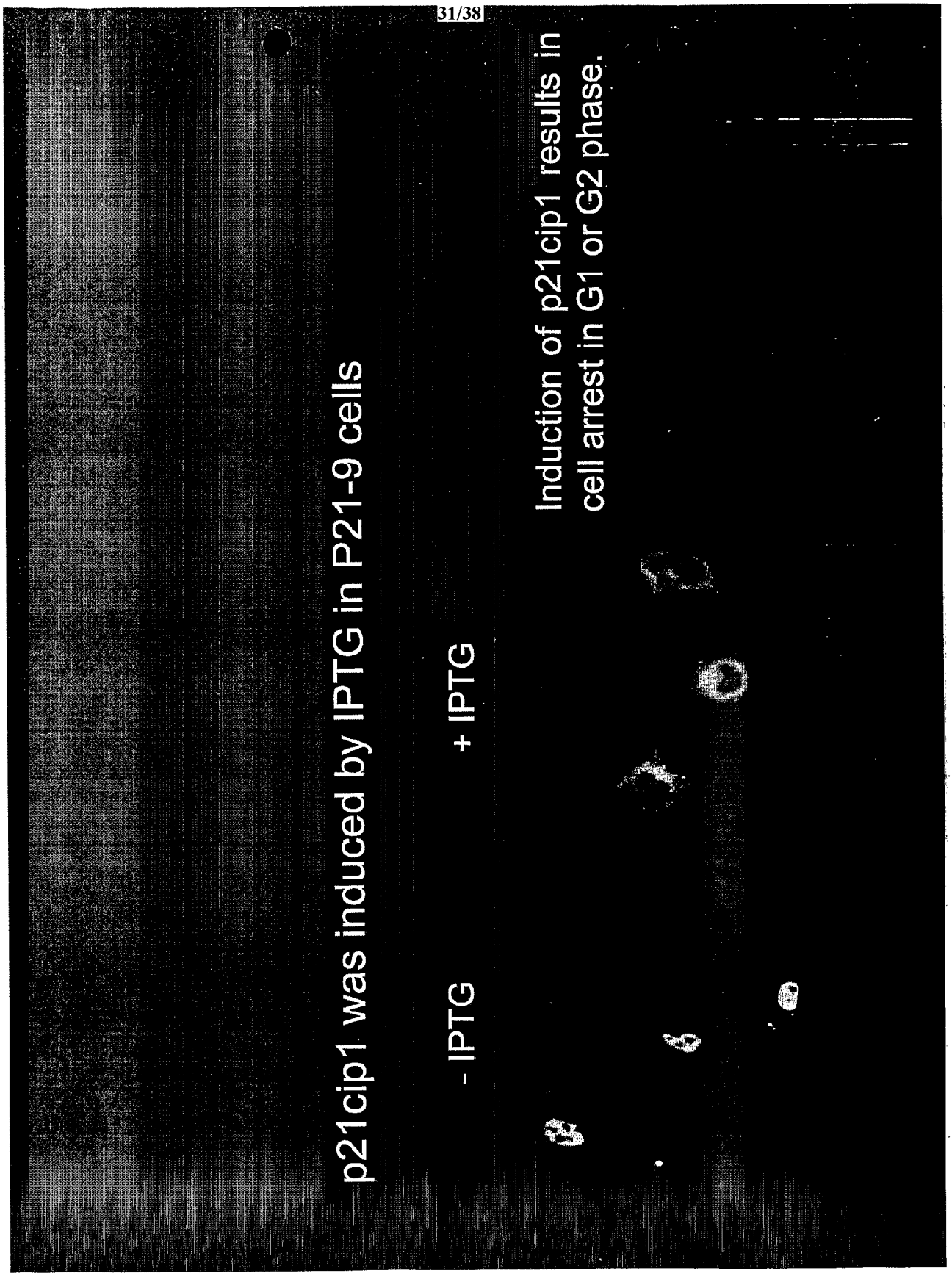

p21cip1 was induced by IPTG in P21-9 cells
- IPTG
+ IPTG
Induction of p21cip1 results in cell arrest in G1 or G2 phase.

FIG. 31

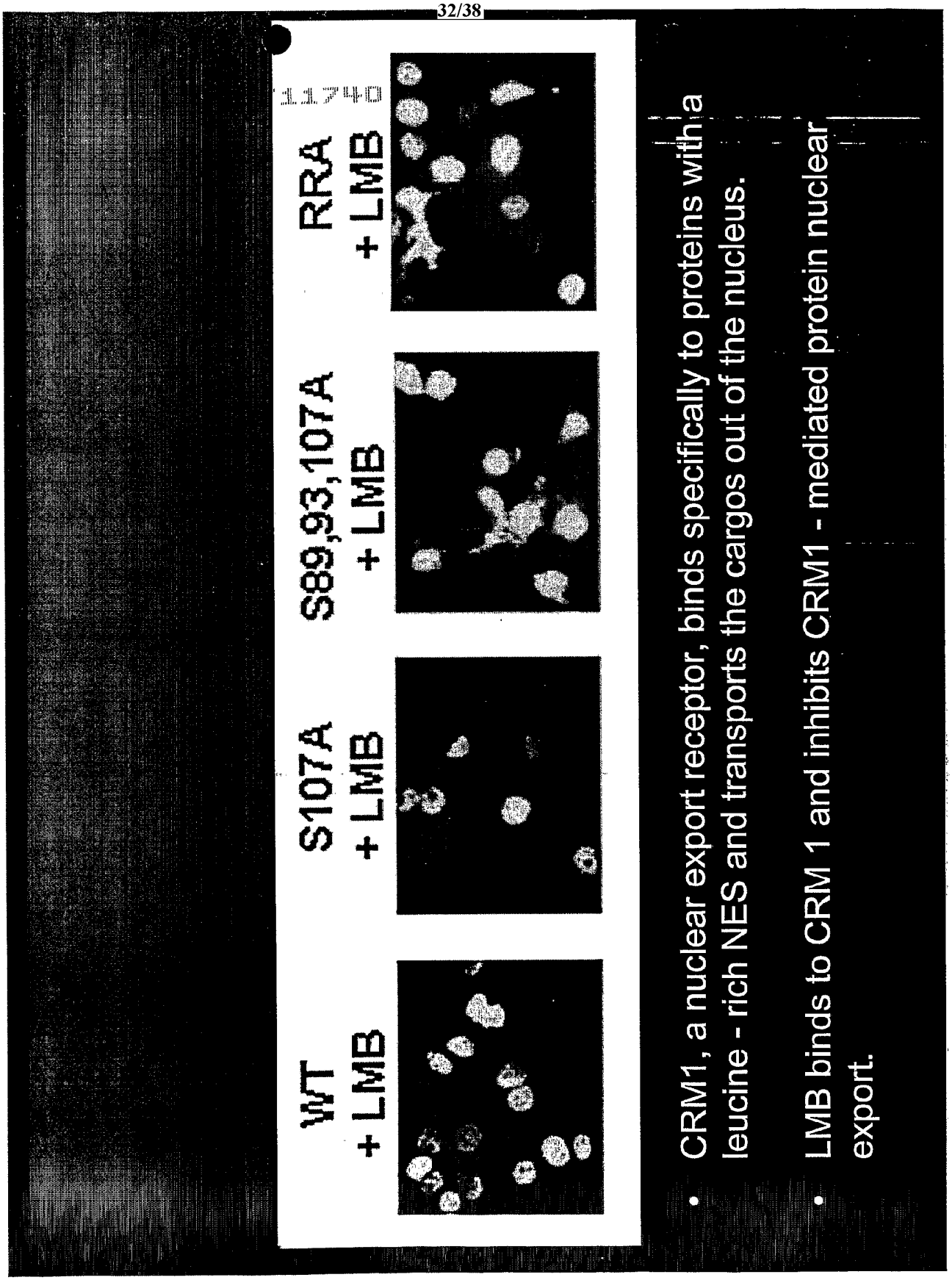

11740
WT
+ LMB
S107A
+ LMB
S89,93,107A
+ LMB
RRA
+ LMB
• CRM1, a nuclear export receptor, binds specifically to proteins with a leucine - rich NES and transports the cargos out of the nucleus.
• LMB binds to CRM 1 and inhibits CRM1 - mediated protein nuclear export.

FIG. 32

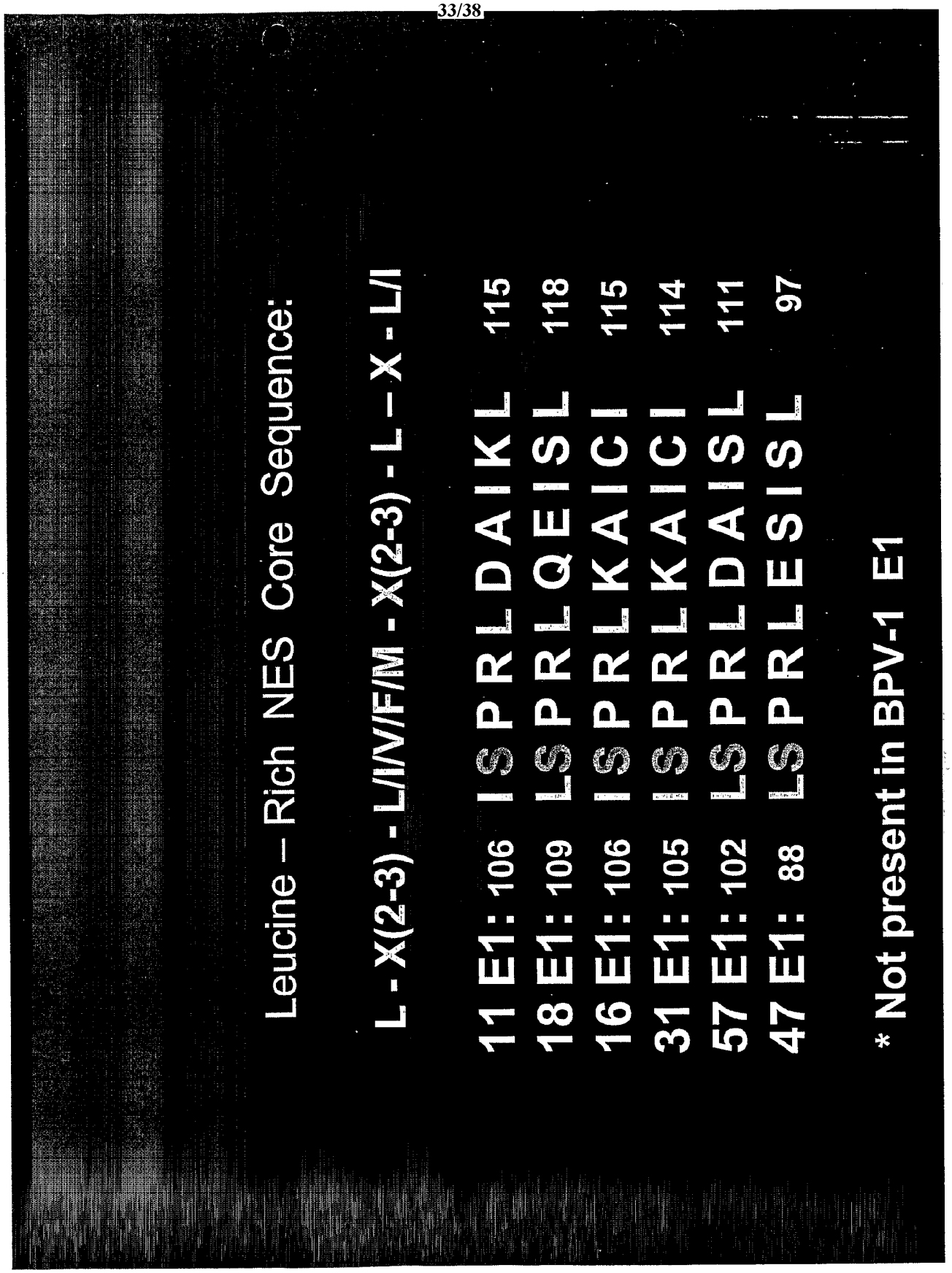

Leucine – Rich NES Core Sequence:
L - X(2-3) - L/I/V/F/M - X(2-3) - L – X - L/I
11 E1: 106 ISPRLDAIKL 115
18 E1: 109 LSPRLQEISL 118
16 E1: 106 ISPRLKAICI 115
31 E1: 105 ISPRLKAICI 114
57 E1: 102 LSPRLDAISL 111
47 E1: 88 LSPRLESISL 97
* Not present in BPV-1 E1

FIG. 33

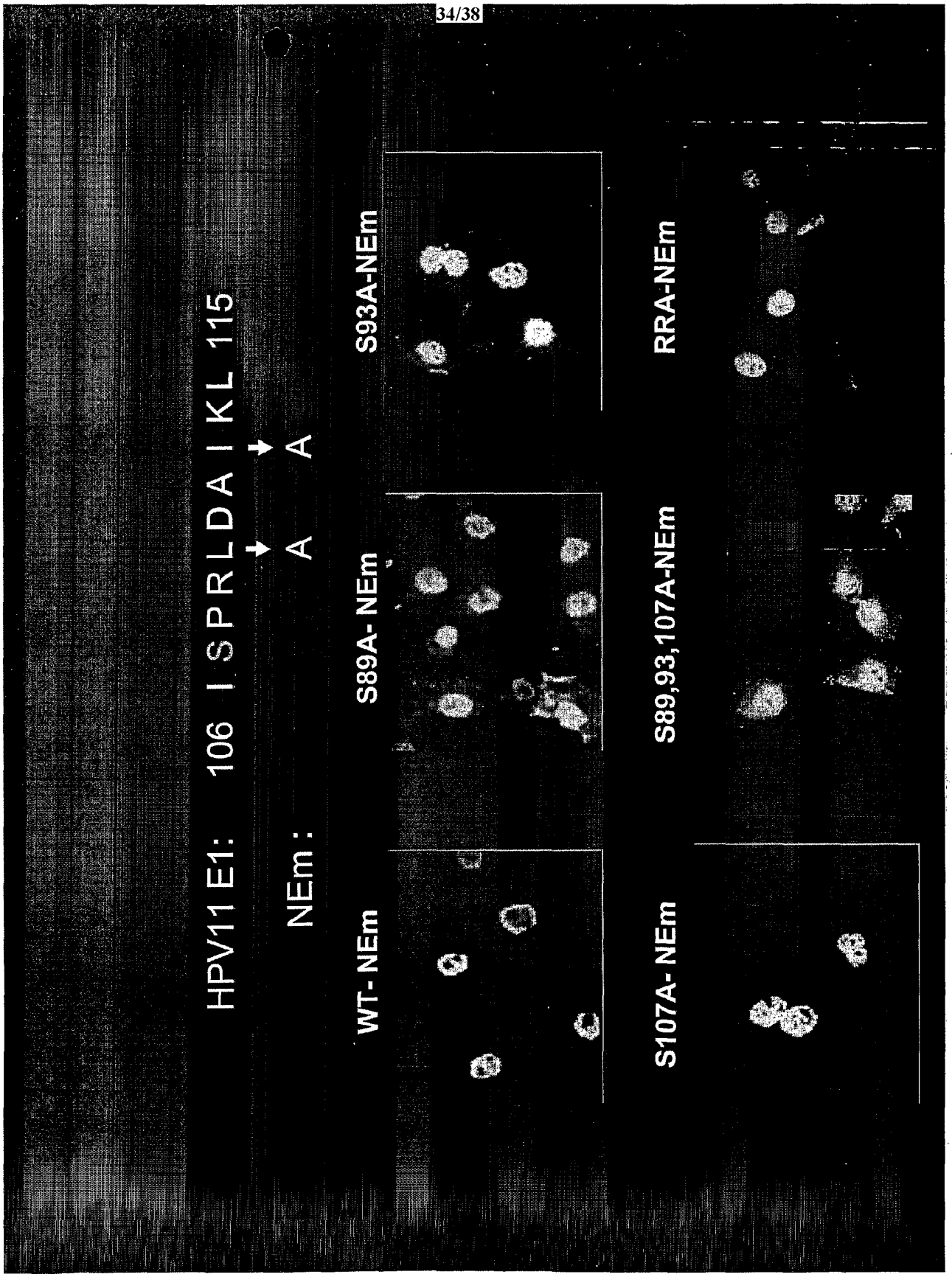

HPV11 E1: 106 I S P R L D A I K L 115
NEm : A A
WT- NEm
S89A- NEm
S93A-NEm
S107A- NEm
S89,93,107A-NEm
RRA-NEm

FIG. 34

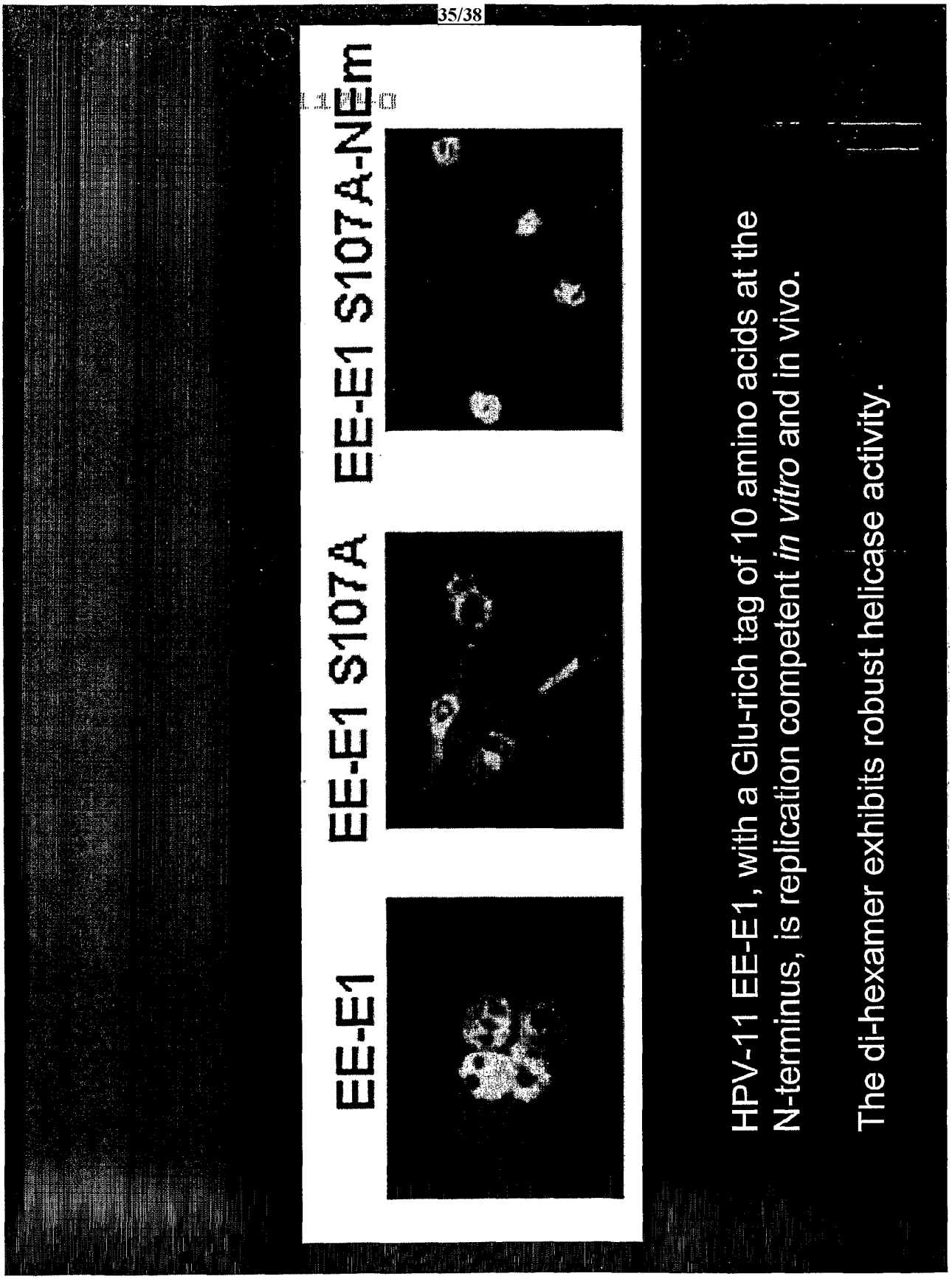

EE-E1
EE-E1 S107A
EE-E1 S107A-NEm
HPV-11 EE-E1, with a Glu-rich tag of 10 amino acids at the N-terminus, is replication competent *in vitro* and in vivo.
The di-hexamer exhibits robust helicase activity.

FIG. 35

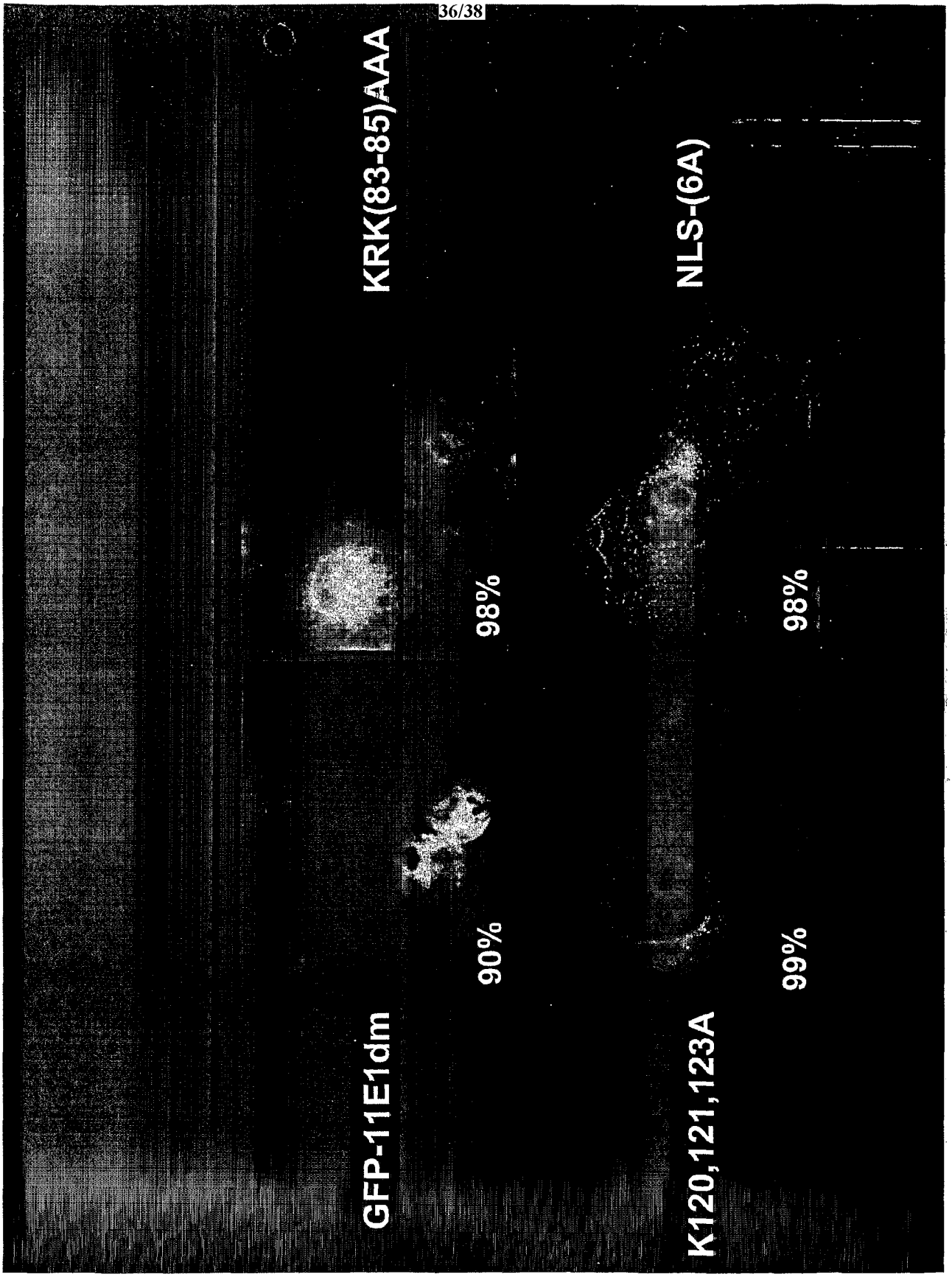


FIG. 36

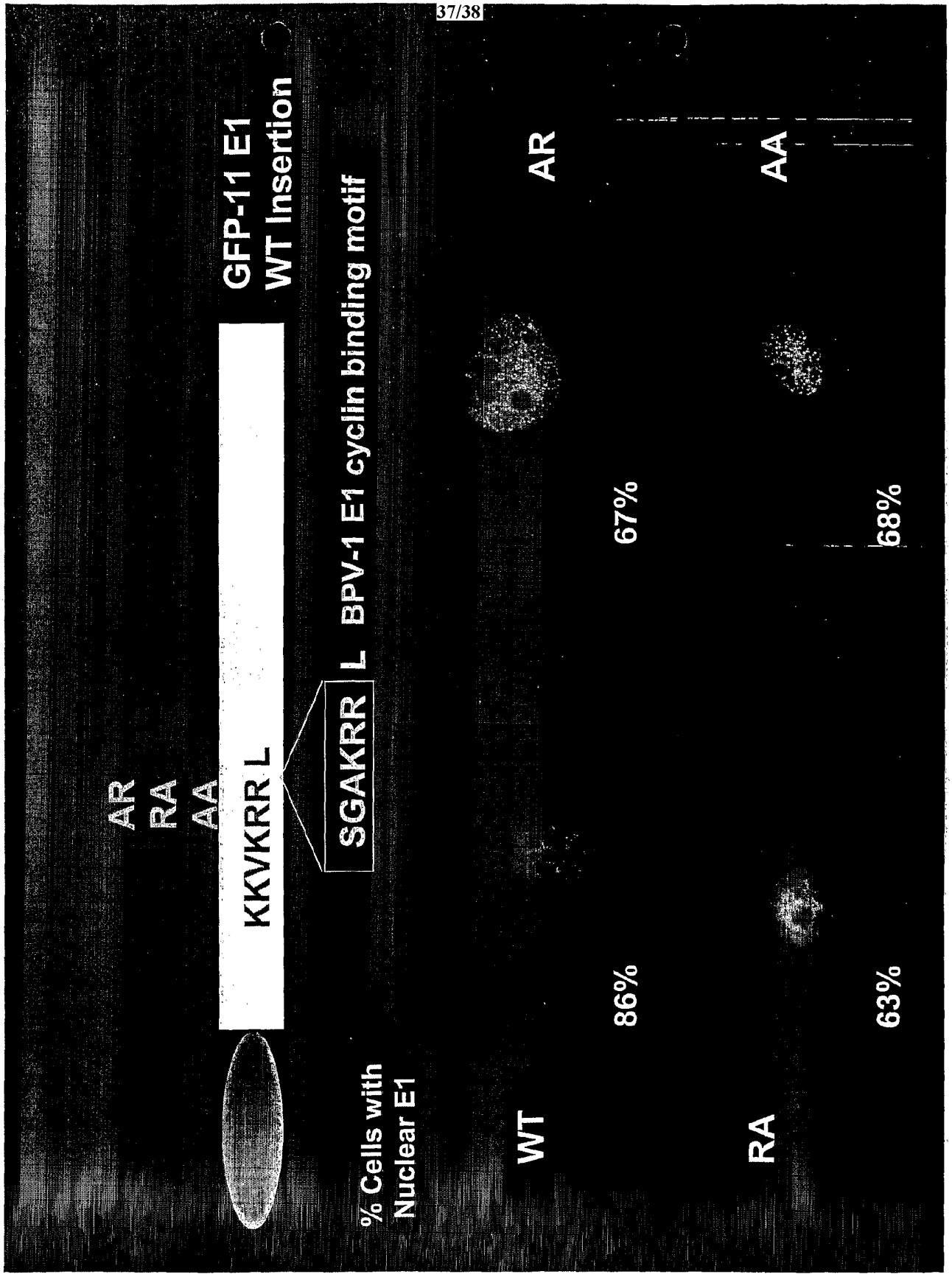

AR
RA
AA
KKVKRR L
SGAKRR L
BPV-1 E1 cyclin binding motif
GFP-11 E1
WT Insertion
% Cells with
Nuclear E1
WT
RA
86%
63%
67%
68%
AR
AA

FIG. 37

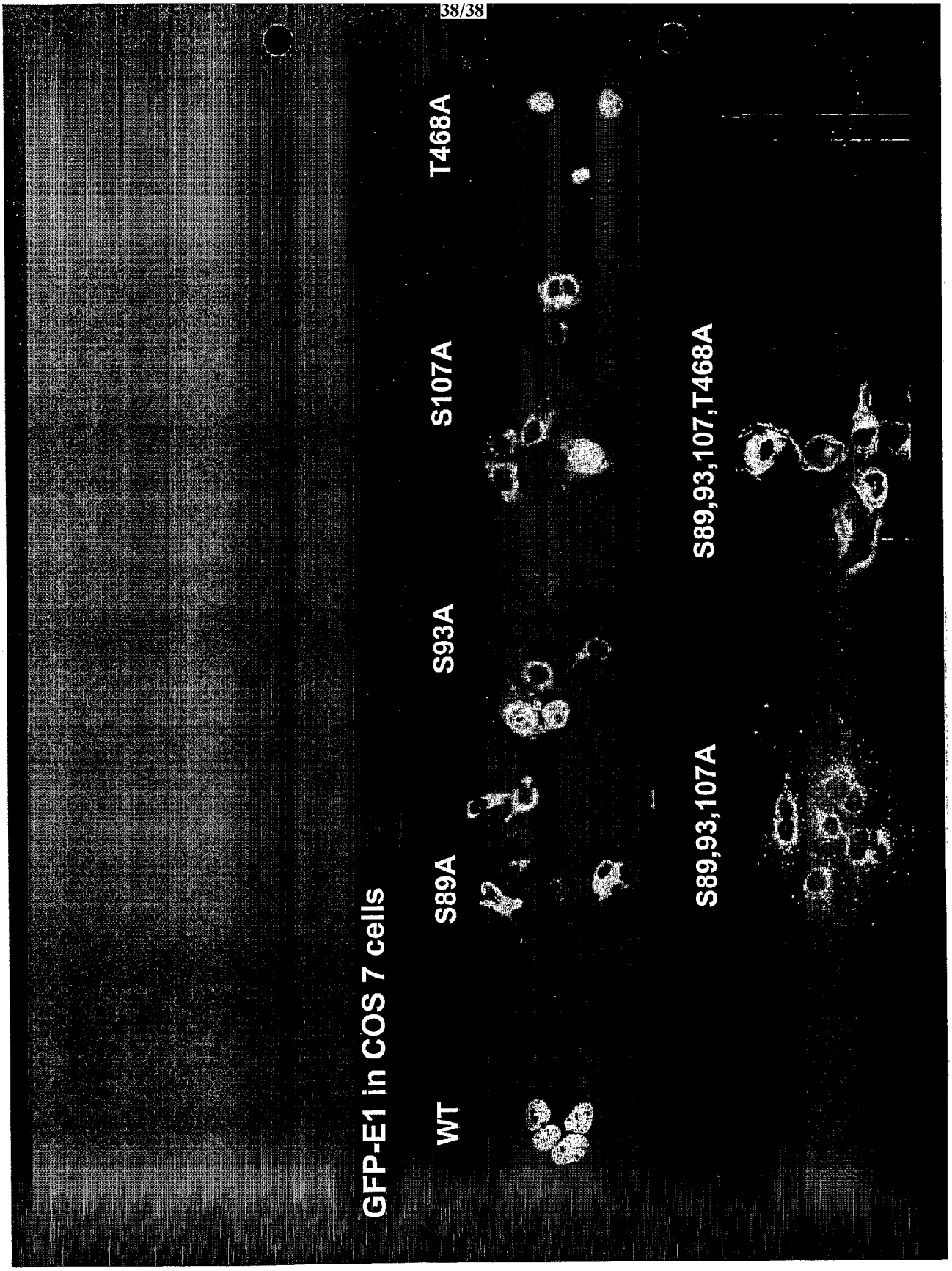

GFP-E1 in COS 7 cells
WT
S89A
S93A
S107A
T468A
S89,93,107A
S89,93,107,T468A

FIG. 38

SEQUENCE LISTING

<110> UAB RESEARCH FOUNDATION

Chow, Louise T.
Broker, Thomas R.
Deng, Wentao
Lin, Biing Yuan
Yu, Jei-Hwa

<120> METHODS AND COMPOSITIONS RELATED TO REGULATION OF NUCLEO-CYTOPLASMIC LOCALIZATION OF E1 DNA HELICASE IN PAPILLOMAVIRUSES

<130> 21085.0118P1

<140> Unassigned
<141> 2005-04-07

<150> 60/560,471
<151> 2004-04-07

<160> 16

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 649
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note = Synthetic Construct

<400> 1

```
Met Ala Asp Asp Ser Gly Thr Glu Asn Glu Gly Ser Gly Cys Thr Gly
 1               5                  10                  15
Trp Phe Met Val Glu Ala Ile Val Glu His Thr Thr Gly Thr Gln Ile
            20                  25                  30
Ser Glu Asp Glu Glu Glu Glu Val Glu Asp Ser Gly Tyr Asp Met Val
        35                  40                  45
Asp Phe Ile Asp Asp Arg His Ile Thr Gln Asn Ser Val Glu Ala Gln
    50                  55                  60
Ala Leu Phe Asn Arg Gln Glu Ala Asp Ala His Tyr Ala Thr Val Gln
65                  70                  75                  80
Asp Leu Lys Arg Lys Tyr Leu Gly Ser Pro Tyr Val Ser Pro Ile Ser
                85                  90                  95
Asn Val Ala Asn Ala Val Glu Ser Glu Ile Ser Pro Arg Leu Asp Ala
            100                 105                 110
Ile Lys Leu Thr Thr Gln Pro Lys Lys Val Lys Arg Arg Leu Phe Glu
        115                 120                 125
Thr Arg Glu Leu Thr Asp Ser Gly Tyr Gly Tyr Ser Glu Val Glu Ala
    130                 135                 140
Ala Thr Gln Val Glu Lys His Gly Asp Pro Glu Asn Gly Gly Asp Gly
145                 150                 155                 160
Gln Glu Arg Asp Thr Gly Arg Asp Ile Glu Gly Glu Gly Val Glu His
                165                 170                 175
```

```
Arg Glu Ala Glu Ala Val Asp Asp Ser Thr Arg Glu His Ala Asp Thr
            180                 185                 190
Ser Gly Ile Leu Glu Leu Leu Lys Cys Lys Asp Ile Arg Ser Thr Leu
        195                 200                 205
His Gly Lys Phe Lys Asp Cys Phe Gly Leu Ser Phe Val Asp Leu Ile
    210                 215                 220
Arg Pro Phe Lys Ser Asp Arg Thr Thr Cys Ala Asp Trp Val Val Ala
225                 230                 235                 240
Gly Phe Gly Ile His His Ser Ile Ala Asp Ala Phe Gln Lys Leu Ile
                245                 250                 255
Glu Pro Leu Ser Leu Tyr Ala His Ile Gln Trp Leu Thr Asn Ala Trp
            260                 265                 270
Gly Met Val Leu Leu Val Leu Ile Arg Phe Lys Val Asn Lys Ser Arg
        275                 280                 285
Cys Thr Val Ala Arg Thr Leu Gly Thr Leu Leu Asn Ile Pro Glu Asn
    290                 295                 300
His Met Leu Ile Glu Pro Pro Lys Ile Gln Ser Gly Val Ala Ala Leu
305                 310                 315                 320
Tyr Trp Phe Arg Thr Gly Ile Ser Asn Ala Ser Thr Val Ile Gly Glu
                325                 330                 335
Ala Pro Glu Trp Ile Thr Arg Gln Thr Val Ile Glu His Ser Leu Ala
            340                 345                 350
Asp Ser Gln Phe Lys Leu Thr Glu Met Val Gln Trp Ala Tyr Asp Asn
        355                 360                 365
Asp Ile Cys Glu Glu Ser Glu Ile Ala Phe Glu Tyr Ala Gln Arg Gly
    370                 375                 380
Asp Phe Asp Ser Asn Ala Arg Ala Phe Leu Asn Ser Asn Met Gln Ala
385                 390                 395                 400
Lys Tyr Val Lys Asp Cys Ala Ile Met Cys Arg His Tyr Lys His Ala
                405                 410                 415
Glu Met Lys Lys Met Ser Ile Lys Gln Trp Ile Lys Tyr Arg Gly Thr
            420                 425                 430
Lys Val Asp Ser Val Gly Asn Trp Lys Pro Ile Val Gln Phe Leu Arg
        435                 440                 445
His Gln Asn Ile Glu Phe Ile Pro Phe Leu Ser Lys Leu Lys Leu Trp
    450                 455                 460
Leu His Gly Thr Pro Lys Lys Asn Cys Ile Ala Ile Val Gly Pro Pro
465                 470                 475                 480
Asp Thr Gly Lys Ser Cys Phe Cys Met Ser Leu Ile Lys Phe Leu Gly
                485                 490                 495
Gly Thr Val Ile Ser Tyr Val Asn Ser Cys Ser His Phe Trp Leu Gln
            500                 505                 510
Pro Leu Thr Asp Ala Lys Val Ala Leu Leu Asp Asp Ala Thr Gln Pro
        515                 520                 525
Cys Trp Thr Tyr Met Asp Thr Tyr Met Arg Asn Leu Leu Asp Gly Asn
    530                 535                 540
Pro Met Ser Ile Asp Arg Lys His Arg Ala Leu Thr Leu Ile Lys Cys
545                 550                 555                 560
Pro Pro Leu Leu Val Thr Ser Asn Ile Asp Ile Ser Lys Glu Glu Lys
                565                 570                 575
Tyr Lys Tyr Leu His Ser Arg Val Thr Thr Phe Thr Phe Pro Asn Pro
            580                 585                 590
Phe Pro Phe Asp Arg Asn Gly Asn Ala Val Tyr Glu Leu Ser Asp Ala
        595                 600                 605
Asn Trp Lys Cys Phe Phe Glu Arg Leu Ser Ser Ser Leu Asp Ile Glu
    610                 615                 620
Asp Ser Glu Asp Glu Glu Asp Gly Ser Asn Ser Gln Ala Phe Arg Cys
625                 630                 635                 640
Val Pro Gly Ser Val Val Arg Thr Leu
                645
```

<210> 2
<211> 44
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 2
Lys Arg Lys Tyr Leu Gly Ser Pro Tyr Val Ser Pro Ile Ser Asn Val
1 5 10 15
Ala Asn Ala Val Glu Ser Glu Ile Ser Pro Arg Leu Asp Ala Ile Lys
20 25 30
Leu Thr Thr Gln Pro Lys Lys Val Lys Arg Arg Leu
35 40

<210> 3
<211> 1960
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 3
gatggcggac gattcaggta cagaaaatga ggggtcgggg tgtacaggat ggtttatggt 60
agaagccata gtagagcaca ctacaggtac acaaatatca gaagatgagg aagaggaggt 120
ggaggacagt gggtatgaca tggtggactt tattgatgac aggcatatta cacaaaattc 180
tgtggaagca caggcattgt ttaataggca ggaggcggat gctcattatg cgactgtgca 240
ggacctaaaa cgaaagtatt taggcagtcc atatgtaagt cctataagca atgtagctaa 300
tgcagtagaa agtgagataa gtccacggtt agacgccatt aaacttacaa cacagccaaa 360
aaaggtaaag cgacggctgt ttgaaacacg ggaattaacg gacagtggat atggctattc 420
tgaagtggaa gctgcaacgc aggtagagaa acatggcgac ccggaaaatg ggggagatgg 480
tcaggaaagg gacacaggga gggacataga gggtgagggg gtggaacata gagaggcgga 540
agcagtagac gacagcaccc gagagcatgc agacacatca ggaatattag aattactaaa 600
atgtaaggat atacgatcta cattacatgg taagtttaaa gactgctttg ggctgtcatt 660
tgttgattta attaggccat ttaaaagtga tagaaccaca tgtgccgatt gggtggttgc 720
aggatttggt atacatcata gcatagcaga tgcatttcaa aagttaattg agccattaag 780
tttatatgca catatacaat ggcttacaaa tgcatgggga atggtactat tagtattaat 840
aaggtttaaa gtaaataaga gcagatgtac cgtggcacgt acattaggta cgttattaaa 900
tatacctgaa aatcacatgt taattgagcc tcctaaaata caaagtggcg tacgagccct 960
gtattggttt aggacaggca tttcaaatgc aagtacagtt ataggggagg cgccggaatg 1020
gataacgcgc cagaccgtta ttgaacatag tttggctgac agtcaattta aattaactga 1080
aatggtgcag tgggcatatg ataatgatat ttgtgaagaa agtgagatag catttgaata 1140
tgcacagcgt ggagactttg actccaatgc aagggccttt ttaaatagta atatgcaggc 1200
taaatatgta aaagattgtg caattatgtg cagacattat aaacatgcag aaatgaaaaa 1260
gatgtctatt aaacaatgga ttaagtatag gggtactaaa gttgacagtg taggtaactg 1320
gaagccaatt gtgcagtttc taagacatca aaacatagaa tttattccat ttttaagcaa 1380
actaaaatta tggctgcacg gaacgcccaa aaaaaattgt atagccattg tagggccacc 1440
tgacactggg aagtcgtgct tttgcatgag tttaattaag tttttggggg gaacagttat 1500
tagttatgtt aattcctgca gccatttctg gctacagcca ctaacggatg caaaagtggc 1560
attattggat gatgccacac aaccatgttg gacatatatg gatacatata tgagaaacct 1620
attagatggt aatcctatga gcatagatag aaaacataga gcattaacat taattaagtg 1680
tccaccgcta ctggttacat caaatataga cattagcaaa gaggagaaat acaaatattt 1740
acatagtaga gttaccacat ttacatttcc aaatccattc ccctttgaca gaaatgggaa 1800
tgcagtatat gaactatcag atgcaaactg gaaatgtttc tttgaaagac tgtcgtccag 1860
cctagacatt gaggattcag aggacgagga agatggaagc aatagccaag cgtttagatg 1920
cgtgccagga tcagttgtta gaactttatg aagaaaacag 1960

<210> 4
<211> 138
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 4
ggacctaaaa cgaaagtatt taggcagtcc atatgtaagt cctataagca atgtagctaa 60
tgcagtagaa agtgagataa gtccacggtt agacgccatt aaacttacaa cacagccaaa 120
aaaggtaaag cgacggct 138

<210> 5
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 5
Ser Pro Ile Ser Asn Val Ala Asn Ala Val Glu Ser Glu Ile Ser Pro
1 5 10 15
Arg Leu Asp Ala Ile Lys Leu Thr
20

<210> 6
<211> 3
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 6
Arg Leu Leu
1

<210> 7
<211> 3
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 7
Lys Arg Lys
1

<210> 8
<211> 6
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 8
Lys Lys Val Lys Arg Arg
1 5

<210> 9
<211> 44
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 9
Lys Arg Lys Tyr Leu Gly Ser Pro Tyr Val Ser Pro Ile Ser Asn Val
1 5 10 15
Ala Asn Ala Val Glu Ser Glu Ile Ser Pro Arg Ala Asp Ala Ile Lys
20 25 30
Leu Thr Thr Gln Pro Lys Lys Val Lys Arg Arg Leu
35 40

<210> 10
<211> 44
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 10
Lys Arg Lys Tyr Leu Gly Ser Pro Tyr Val Ser Pro Ile Ser Asn Val
1 5 10 15
Ala Asn Ala Val Glu Ser Glu Ile Ser Pro Arg Leu Asp Ala Ala Lys
20 25 30
Leu Thr Thr Gln Pro Lys Lys Val Lys Arg Arg Leu
35 40

<210> 11
<211> 649
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 11
Met Ala Asp Asp Ser Gly Thr Glu Asn Glu Gly Ser Gly Cys Thr Gly
1 5 10 15
Trp Phe Met Val Glu Ala Ile Val Glu His Thr Thr Gly Thr Gln Ile
20 25 30
Ser Glu Asp Glu Glu Glu Glu Val Glu Asp Ser Gly Tyr Asp Met Val
35 40 45
Asp Phe Ile Asp Asp Arg His Ile Thr Gln Asn Ser Val Glu Ala Gln
50 55 60
Ala Leu Phe Asn Arg Gln Glu Ala Asp Ala His Tyr Ala Thr Val Gln
65 70 75 80

```
Asp Leu Lys Arg Lys Tyr Leu Gly Ser Pro Tyr Val Ser Pro Ile Ser
                85                  90                  95
Asn Val Ala Asn Ala Val Glu Ser Glu Ile Ser Pro Arg Leu Asp Ala
            100                 105                 110
Ile Lys Leu Thr Thr Gln Pro Lys Lys Val Lys Arg Arg Ala Phe Glu
        115                 120                 125
Thr Arg Glu Leu Thr Asp Ser Gly Tyr Gly Tyr Ser Glu Val Glu Ala
    130                 135                 140
Ala Thr Gln Val Glu Lys His Gly Asp Pro Glu Asn Gly Gly Asp Gly
145                 150                 155                 160
Gln Glu Arg Asp Thr Gly Arg Asp Ile Glu Gly Glu Gly Val Glu His
                165                 170                 175
Arg Glu Ala Glu Ala Val Asp Asp Ser Thr Arg Glu His Ala Asp Thr
            180                 185                 190
Ser Gly Ile Leu Glu Leu Leu Lys Cys Lys Asp Ile Arg Ser Thr Leu
        195                 200                 205
His Gly Lys Phe Lys Asp Cys Phe Gly Leu Ser Phe Val Asp Leu Ile
    210                 215                 220
Arg Pro Phe Lys Ser Asp Arg Thr Thr Cys Ala Asp Trp Val Val Ala
225                 230                 235                 240
Gly Phe Gly Ile His His Ser Ile Ala Asp Ala Phe Gln Lys Leu Ile
                245                 250                 255
Glu Pro Leu Ser Leu Tyr Ala His Ile Gln Trp Leu Thr Asn Ala Trp
            260                 265                 270
Gly Met Val Leu Leu Val Leu Ile Arg Phe Lys Val Asn Lys Ser Arg
        275                 280                 285
Cys Thr Val Ala Arg Thr Leu Gly Thr Leu Leu Asn Ile Pro Glu Asn
    290                 295                 300
His Met Leu Ile Glu Pro Pro Lys Ile Gln Ser Gly Val Ala Ala Leu
305                 310                 315                 320
Tyr Trp Phe Arg Thr Gly Ile Ser Asn Ala Ser Thr Val Ile Gly Glu
                325                 330                 335
Ala Pro Glu Trp Ile Thr Arg Gln Thr Val Ile Glu His Ser Leu Ala
            340                 345                 350
Asp Ser Gln Phe Lys Leu Thr Glu Met Val Gln Trp Ala Tyr Asp Asn
        355                 360                 365
Asp Ile Cys Glu Glu Ser Glu Ile Ala Phe Glu Tyr Ala Gln Arg Gly
    370                 375                 380
Asp Phe Asp Ser Asn Ala Arg Ala Phe Leu Asn Ser Asn Met Gln Ala
385                 390                 395                 400
Lys Tyr Val Lys Asp Cys Ala Ile Met Cys Arg His Tyr Lys His Ala
                405                 410                 415
Glu Met Lys Lys Met Ser Ile Lys Gln Trp Ile Lys Tyr Arg Gly Thr
            420                 425                 430
Lys Val Asp Ser Val Gly Asn Trp Lys Pro Ile Val Gln Phe Leu Arg
        435                 440                 445
His Gln Asn Ile Glu Phe Ile Pro Phe Leu Ser Lys Leu Lys Leu Trp
    450                 455                 460
Leu His Gly Thr Pro Lys Lys Asn Cys Ile Ala Ile Val Gly Pro Pro
465                 470                 475                 480
Asp Thr Gly Lys Ser Cys Phe Cys Met Ser Leu Ile Lys Phe Leu Gly
                485                 490                 495
Gly Thr Val Ile Ser Tyr Val Asn Ser Cys Ser His Phe Trp Leu Gln
            500                 505                 510
Pro Leu Thr Asp Ala Lys Val Ala Leu Leu Asp Asp Ala Thr Gln Pro
        515                 520                 525
Cys Trp Thr Tyr Met Asp Thr Tyr Met Arg Asn Leu Leu Asp Gly Asn
    530                 535                 540
Pro Met Ser Ile Asp Arg Lys His Arg Ala Leu Thr Leu Ile Lys Cys
545                 550                 555                 560
```

```
Pro Pro Leu Leu Val Thr Ser Asn Ile Asp Ile Ser Lys Glu Glu Lys
                565                 570                 575
Tyr Lys Tyr Leu His Ser Arg Val Thr Thr Phe Thr Phe Pro Asn Pro
            580                 585                 590
Phe Pro Phe Asp Arg Asn Gly Asn Ala Val Tyr Glu Leu Ser Asp Ala
        595                 600                 605
Asn Trp Lys Cys Phe Phe Glu Arg Leu Ser Ser Ser Leu Asp Ile Glu
    610                 615                 620
Asp Ser Glu Asp Glu Glu Asp Gly Ser Asn Ser Gln Ala Phe Arg Cys
625                 630                 635                 640
Val Pro Gly Ser Val Val Arg Thr Leu
                645
```

<210> 12
<211> 649
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

```
<400> 12
Met Ala Asp Asp Ser Gly Thr Glu Asn Glu Gly Ser Gly Cys Thr Gly
 1               5                  10                  15
Trp Phe Met Val Glu Ala Ile Val Glu His Thr Thr Gly Thr Gln Ile
            20                  25                  30
Ser Glu Asp Glu Glu Glu Glu Val Glu Asp Ser Gly Tyr Asp Met Val
        35                  40                  45
Asp Phe Ile Asp Asp Arg His Ile Thr Gln Asn Ser Val Glu Ala Gln
    50                  55                  60
Ala Leu Phe Asn Arg Gln Glu Ala Asp Ala His Tyr Ala Thr Val Gln
65                  70                  75                  80
Asp Leu Lys Arg Lys Tyr Leu Gly Ser Pro Tyr Val Ser Pro Ile Ser
                85                  90                  95
Asn Val Ala Asn Ala Val Glu Ser Glu Ile Ser Pro Arg Leu Asp Ala
            100                 105                 110
Ile Lys Leu Thr Thr Gln Pro Lys Lys Val Lys Lys Arg Ala Phe Glu
        115                 120                 125
Thr Arg Glu Leu Thr Asp Ser Gly Tyr Gly Tyr Ser Glu Val Glu Ala
    130                 135                 140
Ala Thr Gln Val Glu Lys His Gly Asp Pro Glu Asn Gly Gly Asp Gly
145                 150                 155                 160
Gln Glu Arg Asp Thr Gly Arg Asp Ile Glu Gly Glu Gly Val Glu His
                165                 170                 175
Arg Glu Ala Glu Ala Val Asp Asp Ser Thr Arg Glu His Ala Asp Thr
            180                 185                 190
Ser Gly Ile Leu Glu Leu Leu Lys Cys Lys Asp Ile Arg Ser Thr Leu
        195                 200                 205
His Gly Lys Phe Lys Asp Cys Phe Gly Leu Ser Phe Val Asp Leu Ile
    210                 215                 220
Arg Pro Phe Lys Ser Asp Arg Thr Thr Cys Ala Asp Trp Val Val Ala
225                 230                 235                 240
Gly Phe Gly Ile His His Ser Ile Ala Asp Ala Phe Gln Lys Leu Ile
                245                 250                 255
Glu Pro Leu Ser Leu Tyr Ala His Ile Gln Trp Leu Thr Asn Ala Trp
            260                 265                 270
Gly Met Val Leu Leu Val Leu Ile Arg Phe Lys Val Asn Lys Ser Arg
        275                 280                 285
Cys Thr Val Ala Arg Thr Leu Gly Thr Leu Leu Asn Ile Pro Glu Asn
    290                 295                 300
```

```
His Met Leu Ile Glu Pro Pro Lys Ile Gln Ser Gly Val Ala Ala Leu
305                 310                 315                 320
Tyr Trp Phe Arg Thr Gly Ile Ser Asn Ala Ser Thr Val Ile Gly Glu
                325                 330                 335
Ala Pro Glu Trp Ile Thr Arg Gln Thr Val Ile Glu His Ser Leu Ala
            340                 345                 350
Asp Ser Gln Phe Lys Leu Thr Glu Met Val Gln Trp Ala Tyr Asp Asn
        355                 360                 365
Asp Ile Cys Glu Glu Ser Glu Ile Ala Phe Glu Tyr Ala Gln Arg Gly
    370                 375                 380
Asp Phe Asp Ser Asn Ala Arg Ala Phe Leu Asn Ser Asn Met Gln Ala
385                 390                 395                 400
Lys Tyr Val Lys Asp Cys Ala Ile Met Cys Arg His Tyr Lys His Ala
                405                 410                 415
Glu Met Lys Lys Met Ser Ile Lys Gln Trp Ile Lys Tyr Arg Gly Thr
            420                 425                 430
Lys Val Asp Ser Val Gly Asn Trp Lys Pro Ile Val Gln Phe Leu Arg
        435                 440                 445
His Gln Asn Ile Glu Phe Ile Pro Phe Leu Ser Lys Leu Lys Leu Trp
    450                 455                 460
Leu His Gly Thr Pro Lys Lys Asn Cys Ile Ala Ile Val Gly Pro Pro
465                 470                 475                 480
Asp Thr Gly Lys Ser Cys Phe Cys Met Ser Leu Ile Lys Phe Leu Gly
                485                 490                 495
Gly Thr Val Ile Ser Tyr Val Asn Ser Cys Ser His Phe Trp Leu Gln
            500                 505                 510
Pro Leu Thr Asp Ala Lys Val Ala Leu Leu Asp Asp Ala Thr Gln Pro
        515                 520                 525
Cys Trp Thr Tyr Met Asp Thr Tyr Met Arg Asn Leu Leu Asp Gly Asn
    530                 535                 540
Pro Met Ser Ile Asp Arg Lys His Arg Ala Leu Thr Leu Ile Lys Cys
545                 550                 555                 560
Pro Pro Leu Leu Val Thr Ser Asn Ile Asp Ile Ser Lys Glu Glu Lys
                565                 570                 575
Tyr Lys Tyr Leu His Ser Arg Val Thr Thr Phe Thr Phe Pro Asn Pro
            580                 585                 590
Phe Pro Phe Asp Arg Asn Gly Asn Ala Val Tyr Glu Leu Ser Asp Ala
        595                 600                 605
Asn Trp Lys Cys Phe Phe Glu Arg Leu Ser Ser Ser Leu Asp Ile Glu
    610                 615                 620
Asp Ser Glu Asp Glu Glu Asp Gly Ser Asn Ser Gln Ala Phe Arg Cys
625                 630                 635                 640
Val Pro Gly Ser Val Val Arg Thr Leu
                645
```

<210> 13
<211> 649
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 13

```
Met Ala Asp Asp Ser Gly Thr Glu Asn Glu Gly Ser Gly Cys Thr Gly
 1               5                  10                  15
Trp Phe Met Val Glu Ala Ile Val Glu His Thr Thr Gly Thr Gln Ile
            20                  25                  30
Ser Glu Asp Glu Glu Glu Glu Val Glu Asp Ser Gly Tyr Asp Met Val
        35                  40                  45
```

```
Asp Phe Ile Asp Asp Arg His Ile Thr Gln Asn Ser Val Glu Ala Gln
    50                  55                  60
Ala Leu Phe Asn Arg Gln Glu Ala Asp Ala His Tyr Ala Thr Val Gln
65                  70                  75                  80
Asp Leu Lys Arg Lys Tyr Leu Gly Ser Pro Tyr Val Ser Pro Ile Ser
                85                  90                  95
Asn Val Ala Asn Ala Val Glu Ser Glu Ile Ser Pro Arg Leu Asp Ala
            100                 105                 110
Ile Lys Leu Thr Thr Gln Pro Lys Lys Val Lys Ala Arg Ala Phe Glu
        115                 120                 125
Thr Arg Glu Leu Thr Asp Ser Gly Tyr Gly Tyr Ser Glu Val Glu Ala
    130                 135                 140
Ala Thr Gln Val Glu Lys His Gly Asp Pro Glu Asn Gly Gly Asp Gly
145                 150                 155                 160
Gln Glu Arg Asp Thr Gly Arg Asp Ile Glu Gly Glu Gly Val Glu His
                165                 170                 175
Arg Glu Ala Glu Ala Val Asp Asp Ser Thr Arg Glu His Ala Asp Thr
            180                 185                 190
Ser Gly Ile Leu Glu Leu Leu Lys Cys Lys Asp Ile Arg Ser Thr Leu
        195                 200                 205
His Gly Lys Phe Lys Asp Cys Phe Gly Leu Ser Phe Val Asp Leu Ile
    210                 215                 220
Arg Pro Phe Lys Ser Asp Arg Thr Thr Cys Ala Asp Trp Val Val Ala
225                 230                 235                 240
Gly Phe Gly Ile His His Ser Ile Ala Asp Ala Phe Gln Lys Leu Ile
                245                 250                 255
Glu Pro Leu Ser Leu Tyr Ala His Ile Gln Trp Leu Thr Asn Ala Trp
            260                 265                 270
Gly Met Val Leu Leu Val Leu Ile Arg Phe Lys Val Asn Lys Ser Arg
        275                 280                 285
Cys Thr Val Ala Arg Thr Leu Gly Thr Leu Leu Asn Ile Pro Glu Asn
    290                 295                 300
His Met Leu Ile Glu Pro Pro Lys Ile Gln Ser Gly Val Ala Ala Leu
305                 310                 315                 320
Tyr Trp Phe Arg Thr Gly Ile Ser Asn Ala Ser Thr Val Ile Gly Glu
                325                 330                 335
Ala Pro Glu Trp Ile Thr Arg Gln Thr Val Ile Glu His Ser Leu Ala
            340                 345                 350
Asp Ser Gln Phe Lys Leu Thr Glu Met Val Gln Trp Ala Tyr Asp Asn
        355                 360                 365
Asp Ile Cys Glu Glu Ser Glu Ile Ala Phe Glu Tyr Ala Gln Arg Gly
    370                 375                 380
Asp Phe Asp Ser Asn Ala Arg Ala Phe Leu Asn Ser Asn Met Gln Ala
385                 390                 395                 400
Lys Tyr Val Lys Asp Cys Ala Ile Met Cys Arg His Tyr Lys His Ala
                405                 410                 415
Glu Met Lys Lys Met Ser Ile Lys Gln Trp Ile Lys Tyr Arg Gly Thr
            420                 425                 430
Lys Val Asp Ser Val Gly Asn Trp Lys Pro Ile Val Gln Phe Leu Arg
        435                 440                 445
His Gln Asn Ile Glu Phe Ile Pro Phe Leu Ser Lys Leu Lys Leu Trp
    450                 455                 460
Leu His Gly Thr Pro Lys Lys Asn Cys Ile Ala Ile Val Gly Pro Pro
465                 470                 475                 480
Asp Thr Gly Lys Ser Cys Phe Cys Met Ser Leu Ile Lys Phe Leu Gly
                485                 490                 495
Gly Thr Val Ile Ser Tyr Val Asn Ser Cys Ser His Phe Trp Leu Gln
            500                 505                 510
Pro Leu Thr Asp Ala Lys Val Ala Leu Leu Asp Asp Ala Thr Gln Pro
        515                 520                 525
```

```
Cys Trp Thr Tyr Met Asp Thr Tyr Met Arg Asn Leu Leu Asp Gly Asn
    530                 535                 540
Pro Met Ser Ile Asp Arg Lys His Arg Ala Leu Thr Leu Ile Lys Cys
545                 550                 555                 560
Pro Pro Leu Leu Val Thr Ser Asn Ile Asp Ile Ser Lys Glu Glu Lys
                565                 570                 575
Tyr Lys Tyr Leu His Ser Arg Val Thr Thr Phe Thr Phe Pro Asn Pro
            580                 585                 590
Phe Pro Phe Asp Arg Asn Gly Asn Ala Val Tyr Glu Leu Ser Asp Ala
        595                 600                 605
Asn Trp Lys Cys Phe Phe Glu Arg Leu Ser Ser Ser Leu Asp Ile Glu
    610                 615                 620
Asp Ser Glu Asp Glu Glu Asp Gly Ser Asn Ser Gln Ala Phe Arg Cys
625                 630                 635                 640
Val Pro Gly Ser Val Val Arg Thr Leu
                645
```

<210> 14
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 14
Ile Ser Pro Arg Leu Asp Ala Ile Lys Leu
1 5 10

<210> 15
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<400> 15
Ser Asn Val Ala Asn Ala Val Glu Ser Glu Ile Ser Pro Arg Leu Asp
1 5 10 15
Ala Ile Lys Leu
20

<210> 16
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
Synthetic Construct

<221> VARIANT
<222> 2, 8, 10
<223> Xaa = Any amino acid

<400> 16
Leu Xaa Leu Ile Val Phe Met Xaa Leu Xaa Leu Ile
1 5 10