

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 1999.03.24	(73) Titular(es): OTSUKA PHARMACEUTICAL CO., LTD. 9, KANDATSUKASA-CHO 2-CHOME CHIYODA- KU, TOKYO 101-0048 YOJI IKAWA	JP JP
(30) Prioridade(s): 1998.03.27 JP 10046798		
(43) Data de publicação do pedido: 2007.10.24		
(45) Data e BPI da concessão: 2011.06.08 162/2011	(72) Inventor(es): YOJI IKAWA SHUNTARO IKAWA MASUO OBINATA	JP JP JP
	(74) Mandatário: JOSÉ EDUARDO LOPES VIEIRA DE SAMPAIO R DO SALITRE 195 RC DTO 1250-199 LISBOA	PT

(54) Epígrafe: **ANTICORPOS CONTRA A PROTEÍNA P51**

(57) Resumo:

UMA PROTEÍNA RECOMBINANTE P51 COMPREENDENDO UMA SEQUÊNCIA DE AMINOÁCIDOS DA SEQ ID Nº:1, EM QUE A DITA PROTEÍNA P51 RECOMBINANTE É EXPRESSA A PARTIR DE UM TRANSFORMANTE COM O VECTOR DE EXPRESSÃO DE P51, UM ANTICORPO ISOLADO OU UM SEU FRAGMENTO QUE SE LIGA ESPECIFICAMENTE À DITA PROTEÍNA P51 E UM KIT COMPREENDENDO O DITO ANTICORPO OU O SEU FRAGMENTO.

Descrição

"Anticorpos contra a proteína p51"

Campo da Invenção

A presente invenção diz respeito a anticorpos contra a proteína p51.

Antecedentes da Invenção

A proteína p53 foi descoberta como uma proteína nuclear que se liga ao grande antigénio T do ADN do vírus do tumor SV40 e o seu gene (gene p53) foi clonado. No início, o gene p53 foi considerado ser um oncogénio porque a transferência deste gene e dos genes ras conjuntamente para as células resultou na transformação de células embrionárias. Contudo, estudos posteriores revelaram que o gene p53 inicialmente clonado era do tipo mutante e que o tipo selvagem suprimia bastante a actividade transformadora do tipo mutante. Até agora, supressões e anomalias no gene p53 foram detectadas em muitos cancros humanos e foi também descoberta e uma mutação de gâmeta do gene p53 no síndrome de Li-Fraumeni que é conhecido por ser uma doença hereditária com um elevado risco de conversão maligna. Devido a estes e outros factos, o gene p53 foi até agora considerado como sendo um supressor oncogénio importante (Baker, S. J., et al., Science, 244, 217-221(1989); Nigro, J. M., Nature, 342, 705-708(1989)).

A proteína humana p53 consiste em 393 resíduos de aminoácidos e pode ser grosseiramente dividida no domínio N-terminal (a região dos primeiros 1-101 aminoácidos), o domínio nuclear (a região dos 102-292 aminoácidos), e o domínio C-terminal (a região de aminoácidos 293-393). O domínio N-terminal contém as sequências necessárias para a

regulação da transcrição, tais como aminoácidos acídicos e uma região de prolina elevada, e é considerada ser um domínio activador de transcrição. O domínio nuclear central contém 3 locais hidrofóbicos e é um domínio associado com a sequência nucleotídica de ligação ao ADN específica. O domínio C terminal contém muitos aminoácidos básicos e uma sequência necessária para a tetramerização e é considerada ser responsável pelo reconhecimento da ligação de ADN não específica e danificação do ADN e inibição da transformação.

Muitas das anormalidades do gene p53 detectadas nas células do cancro humano são mutações de sentido errado e a maior parte delas encontram-se concentradas no domínio nuclear correspondente ao 100°-300° aminoácido da sequência N terminal particularmente na região designada "local quente" ("hot spot") que foi conservada entre as espécies. A região do local quente do domínio nuclear é a sequência associada com a ligação entre a proteína p53 e o ADN e, actualmente, a mutação desta região resulta na inibição da ligação específica a ADN.

Torna-se claro a partir do que foi referido acima que a proteína p53 desempenha o papel de um factor de controlo transcricional que se liga especificamente a outros genes para modular a expressão dos genes.

O gene cuja transcrição é induzida pela proteína p53 inclui, entre outras, o gene p21(conhecido como WAF1, CIP1, ou SDI1 (El-Dairy, W. S., et al., Cell, 75, 817(1993)); MDM2(Wu. X., et al., Genes Dev., 7, 1126 (1993)); MCK [Weintraub. H., et al., Proc. Natl. Acad. Sci. USA, 88, 4570 (1991); Zambetti. G. P., et al., Genes Dev., 6, 1143 (1992)], GAD95 [Kastan, M. B., et al., Cell, 71, 587 (1992)], Cyclin G [Cyclin G: Okamoto, K., EMBO J., 13, 4816 (1994)], BAX [Miyashita, T., et al., Cell, 80, 293(1995)], e factor de crescimento semelhante à insulina de ligação à

proteína 3 [IGF-BP3: Buckbinder, L., et al., *Nature*, 377, 646 (1995)].

A proteína codificada pelo gene p21 é uma proteína inibidora da cinase dependente da ciclina (CDK), e verificou-se que a proteína p53 do tipo selvagem regula o ciclo celular de uma forma inibidora através de p21 [Harper, J. W., et al., *Cell*, 75, 805(1993): Xiong Y., et al., *Nature*, 366, 707(1993):Gu, Y. et al., *Nature*, 366, 701(1993)]. Além disso, foi reportado que o gene p21 liga-se ao antigénio nuclear da célula em proliferação (PCNA) para inibir directamente a replicação do ADN [Waga, S., et al., *Nature*, 369, 574(1994)]. Além disso, verificou-se que o gene p21 era o mesmo gene que o gene SDI1 que induz a senescência das células para inibir a síntese de ADN [Noda., A., et al., "*Exp. Cell Res.*", 211, 90(1994)].

O MDM2 liga-se à proteína p53 para inactivar a actividade de regulação transcricional da proteína do gene, conduzindo à conclusão putativa de que MDM2 actua com um factor de regulação de resposta negativo.

IGF-BP3 é um factor de regulação negativo na sinalização de IGF. Por isso, o aumento do gene IGF-BP3 pela proteína p53 sugere o possível resultado de que a proteína p53 induz a supressão do crescimento de células dependentes de IGF.

Entretanto, a proteína p53 do tipo selvagem foi reportada induzir a apoptose de células de leucemia mielocíticas [Yonish-Rouach, E., et al., *Nature*, 352, 345(1991)]. A indução da apoptose timocítica por irradiação não tem lugar em ratinhos p53 defectivos [Lowe, S. W., *Nature*, 362, 847(1993): Clarke, A. R., et al., *Nature* 362, 849 (1993)] e, na lente cristalina, retina e no cérebro, a proteína p53 induz a morte por apoptose de células privadas da actividade do gene do blastoma retinal (gene RB) [Pan,

H., e Griep, A. E., *Genes Dev.*, 8, 1285 (1999): Morgenbesser, S. D., et al., *Nature* 371, 72(1999): Howes, K. A., *Genes Dev.*, 8, 1300 (1994): Symons, H., e tal., *Cell*, 78, 703(1994)]. E. White propõe que a proteína p53 é útil para uma vigilância da mutação do gene RB e que a proteína provavelmente induz apoptose das células nas quais uma mutação do gene RB se encontra envolvida [White, E., *Nature*, 371, 21 (1994)].

Além disso, na linha celular da leucemia eritroide do ratinho em que o gene p53 sensível à temperatura é apenas expresso, uma queda na temperatura resulta na reconversão do gene p53 mutante do tipo selvagem para induzir apoptose e o gene p53 isolado a partir dele confere a capacidade de crescer em meio de agár suave para dar uma linha de fibroblastos p53 defectiva (confere independência de ancoragem) [Xu et al. *Jpn. J. Cancer Res.* 86:284-291 (1995); Kato et al., *Int. J. Oncol.* 9:269-277].

BAX é capaz de se ligar a bcI-2, que é um inibidor da apoptose, e encoraja a morte por apoptose das células [Oltvai, Z. M., et al., *Cell*, 74, 609 (1993)]. O aumento no gene BAX e diminuição em bcI-2 pela proteína p53 estão envolvidos na apoptose da linha celular M1 da leucemia do ratinho [Miyashita, T., et al., *Oncogene*, 9, 1799(1994)] e Fas, que é um dos transdutores de sinais da apoptose é aumentado no cancro de não pequenas células do pulmão e eritroleucemia [Owe-Schaub, L. B., et al., *Mol. Cell Biol.*, 15, 3032 (1995)].

As muitas investigações acima referidas revelaram que a proteína p53 activa ou reprime a transcrição de vários genes não estando limitada ao gene p21. Além disso, mesmo a proteína mutante p53 danificada na função de regulação da transcrição é capaz de interagir com outras proteínas

intracelulares para transmitir sinais e descarregar uma função de reparação dos danos de ADN.

Entre as funções da proteína p53 que foram até agora identificadas encontra-se uma transcrição da função reguladora, uma função de transdução do sinal, através da ligação a outras proteínas intracelulares, um elemento constituinte de um complexo proteico relacionado com a replicação do ADN, uma função de ligação do ADN, e actividade de exonuclease, e conjectura-se que seja o resultado de uma interacção composta destas funções que provocam a paragem do ciclo celular nas células, indução de apoptose, reparação de ADN, regulação da replicação de ADN, e indução de diferenciação.

Além disso, não é verdade que as funções da proteína p53 sejam expressas apenas no caso da danificação de um gene, mas é divulgado que quando o tecido vivo é sujeito a várias tensões, tais como infecção viral, estimulação de citocinas, hipoxia, uma alteração no conjunto nucleotídico, anormalidade metabólica induzida pelos fármacos, etc., os estímulos desencadeiam alterações quantitativas ou qualitativas na proteína p53. A proteína p53 sujeita à regulação quantitativa ou qualitativa expressa as suas funções, tais como transdução de sinal, através de interacções com outras proteínas e controlo da transcrição de outros genes, para regular a replicação de ADN em células do tecido vivo sujeito a tensões biológicas, reparação da célula suspendendo o ciclo celular, eliminação de células por apoptose, ou promover a diferenciação de células contribuindo deste modo para a protecção do tecido vivo contra tensões [Ganman, C. E., et al., Genes Dev., 9, 600-611(1995); Graeber, T. G., et al., Nature, 379, 88-91(1996); Linke, S. P., et al., Genes Dev., 10, 934-947(1996); Xiang, H., et al., J. Neurosci., 16, 6753-6765(1996)].

Tendo em vista a existência de mutações do gene de p53 em metade dos tumores humanos, a aplicação clínica do gene p53 e da sua proteína produto para o diagnóstico e terapia dos tumores foi um assunto de estudo nos anos recentes. O processo de detectar as células tumorais que invadem o nódulo linfático ou fluido corporal efectuando uma PCR utilizando iniciadores que reconhecem especificamente o local de mutação do gene p53 pode ser uma técnica de diagnóstico eficaz para estimar o âmbito da invasão do tumor ou para prever a recorrência do tumor [Hayashi, H., et al., Lancet, 345, 1257-1259 (1995)].

Além disso, aproveitando a actividade indutora da apoptose da proteína p53, uma terapia génica compreendendo a introdução de um gene p53 do tipo selvagem na célula tumoral, através de um vector de vírus tem sido praticada nos Estados Unidos da América e a sua eficácia foi divulgada [Roth, J. A., et al., Nature Med., 2, 985-991 (1996)]. Recentemente, também no Japão, esta terapia génica foi iniciada em várias localizações.

Entretanto, mais do que a maioria dos tumores humanos não estão associados com a mutação do gene p53 e com base neste facto foi salientada a possibilidade da existência de outras proteínas inibidoras da tumorigénese análogas à proteína p53.

Os inventores da presente invenção verificaram previamente que uma mutação no gene da p53 não pode ser um indicador premonitório útil do linfoma não-Hodgkin (NHL).

Recentemente, foi identificado um novo gene, designado p73 que possui uma elevada homologia com o dito gene p53 [Kaghad, M., et al., Cell, 90, 809-819 (1997)]. De acordo com a informação disponível aos presentes inventores, a proteína p73 apresenta 29% de homologia com a proteína p53 humana no domínio de activação transcricional (a região do 1º-45º aminoácido). Além disso, esta proteína p73 possui

uma homologia de 63% no domínio de ligação do ADN (região do 113°-290° aminoácido) possuindo 6 sequências complementares conservadas designadas locais quentes de mutação; e uma homologia de 38% no domínio da oligomerização (região dos aminoácidos 319°-363°). No que diz respeito ao domínio C-terminal, contudo, não foi reconhecida nenhuma homologia significativa entre a proteína p73 e p53.

É reportado que a expressão excessiva da proteína inibe o crescimento de uma linha celular do neuroblastoma e das células SAOS2 (linha celular do osteosarcoma) e que a expressão transiente da proteína p73 promove a apoptose de células SAOS2 de células de rim de hamster bebé [Bruce Clurman e Mark Groudine, *Nature*, 389, 122-123(1997); Christine, A., et al., *Nature*, 389, 191-194(1987)].

Contudo, a proteína p73 é de algum modo diferente da proteína p53, em que a primeira é expressa apenas a níveis baixos em tecidos normais. Além disso, a proteína p 73 é diferente da proteína p53 em que a expressão da primeira proteína numa linha celular do neuroblastoma não é induzida pela irradiação UV ou uma dose baixa de actinomicina D.

Por isso, não é verdade que a proteína p73 possua exactamente as mesmas funções que as da proteína p53 e, actualmente muito depende de mais investigação. Existe um relatório que discute que com base nas observações até agora efectuadas, esta p73 pode ser categorizada como um factor supressor putativo do tumor no neuroblastoma. Ver também a divulgação do intermediário seguinte [Osada et al., *Nature Medecine*, 4, 839-843 (Julho 1992) e Yang et al., *Molecular Cell*, 2, 305-316 (Setembro 1998)].

A presente invenção possui como seu objectivo disponibilizar anticorpos contra a proteína p51.

Breve Descrição das Figuras

A Fig. 1 é um diagrama que ilustra as características dos domínios estruturais da proteína p51A, conjuntamente com aquelas da proteína p53 e proteína p73 β . No diagrama, "TA" representa um diagrama de activação transcrricional; "ligação de ADN" representa um domínio de ligação de ADN; e "oligo" representa um domínio de oligomerização.

A Fig. 2 é um diagrama que apresenta a relação de homologia com a sequência de aminoácidos codificada pelo gene da p51A humana, a sequência de aminoácidos da proteína p53, e a sequência de aminoácidos da proteína p73 β . Os aminoácidos que são comuns entre as três sequências estão indicados em blocos.

A Fig. 3 é um diagrama que apresenta a relação de homologia da sequência de aminoácidos codificada pelo gene da p51B humano e a sequência de aminoácidos da proteína p73 α . Os aminoácidos que são comuns a ambas as sequências estão indicados em blocos.

A Fig. 4 é um diagrama esquemático que compara a estrutura da variante de união alternativa (p51A, p51B) da proteína p51 com uma estrutura da variante de união alternativa (p73 α , p73 β) da proteína p73.

A Fig. 5 é uma fotografia em vez de um desenho que mostra o padrão de expressão de ARNm de p51 em vários tecidos humanos utilizando a transferência de Northern (utilizando um filtro de Clontech) electroforetograma. As bandas representam os resultados para 1: coração, 2: cérebro, 3: placenta, 4: pulmão, 5:fígado, 6:músculo esquelético, 7:baço, 8:pâncreas, respectivamente.

A Fig. 6 é uma fotografia em vez de um desenho que mostra o padrão de expressão de ARNm de p51 em vários tecidos humanos utilizando um electroforetograma de transferência de Northern (utilizando um filtro preparado utilizando o

ARN adquirido à Clontech). As bandas representam os resultados para 1: glândula mamária, 2: próstata, 3: glândula salivar, 4: estômago, 5: timus, 6: tiróide, 7: traqueia, e 8: útero, respectivamente.

A Fig. 7 é uma fotografia em vez de um desenho que mostra a actividade anti-formação de colónias do gene de p51A. Mais especificamente é uma fotografia em vez de um desenho, que mostra como comparação as actividades de formação de anti-colónias das células transformadas com o plasmídeo de expressão p51A (p51A), com o plasmídeo de expressão p53 (p53), plasmídeo de expressão p51A marcado com HA (HAp51A), e apenas o vector (RcCMV), respectivamente.

A Fig. 8 é uma representação esquemática que mostra as construções do gene repórter utilizadas no Exemplo Experimental 2. No diagrama, o promotor "WAF-1 luc" representa uma construção do promotor p21WAF1 do tipo selvagem com dois elementos de regulação p53 retidos; "del 1" representa uma construção semelhante em que um elemento a montante foi suprimido; e "del 2" representa uma construção em que ambos os elementos foram suprimidos.

A Fig. 9 é um diagrama que apresenta a actividade de transactivação encontrada na transferência do plasmídeo de expressão p51A (p51A), plasmídeo de expressão p53 (p53) ou vector de controlo (Rc/CMV) que alberga as várias construções repórter apresentadas na Fig. 8 em células SAOS2 [cf Exemplo Experimental 2].

A Fig. 10 é um diagrama que apresenta a actividade de transactivação encontrada na transferência do plasmídeo de expressão p51A (p51A), plasmídeo de expressão p51A marcado com HA (HAp51A), plasmídeo de expressão p53 (p53), ou vector de controlo (Rc/CMV) que alberga a construção PGC repórter, cuja resposta p53 foi demonstrada experimentalmente em células SAOS2 [ver Exemplo Experimental 2].

A Fig. 11 é uma fotografia (um electroforegrama de um gel de agarose corado com brometo de etídio), em vez de um desenho que mostra os resultados de fragmentação de ADN efectuados com células 1C1 e 4B1 contendo o gene p51A humano e células 1-2-3 que não contêm o gene p51A cultivado a temperaturas diferentes de 32°C e 37°C.

No diagrama "1-2-3 células" representa células de controlo em que foi introduzido apenas o vector e que não contém o gene p51A e as "células 1C1" ou "células 4B1" representa as células 1-2-3 contendo p51A transformadas com o vector de expressão que alberga o gene p51A (pRcCMV/p51A). O " χ /HindIII" representa produtos de digestão tal como digeridos com a enzima de restrição Hind III do ADN do fago χ e são marcadores de tamanho de ADN (produto da New England Biolabs. Ind.). A "escada de 100 pb" representa marcadores de tamanho compreendendo fragmentos de ADN que possuem tamanhos que correspondem a múltiplos de 100 pb (produto de GIBCO-BRL).

As Figs. 12-14 mostram um diagrama que compara a sequência nucleotídica (linha de baixo) da região de codificação do gene humano p51B com a sequência correspondente (linha de cima) do homólogo de ratinho (gene p51B de ratinho). Os nucleótidos comuns entre as duas sequências são indicados pelo asterisco no diagrama.

A Fig. 15 é um diagrama que compara as sequências de aminoácidos da proteína p51B e da proteína p51B do ratinho codificada pelo gene p51B humano e pelo gene p51B do ratinho, ambos apresentados nas Figs. 12-14, respectivamente. Os aminoácidos comuns a ambas sequências são indicados por um asterisco no diagrama.

Descrição Detalhada da Invenção

Uma vez que mais do que a maioria dos tecidos tumorais humanos possuem mutantes do gene p53 que é um gene supressor oncogénico como mencionado acima, foi sugerida a possibilidade da existência de outros produtos génicos (proteínas) diferentes da proteína p53 que descarregam a função que inibe a tumorigénese.

Por isso, os inventores da presente invenção efectuaram investigações intensas na procura de novos genes e produtos génicos que poderiam estar associados com a dita função inibidora da tumorigénese. Como resultado constatarem a existência de um novo gene de origem humana que codifica para uma proteína que apresenta uma actividade semelhante à da dita proteína p53 e confirmou que o gene em particular ou produto génico encontra-se particularmente associado com a apoptose. A presente invenção possui a sua base nesta constatação.

A presente invenção é por isso direccionada para o seguinte.

1. Um anticorpo isolado que se liga especificamente a uma proteína p51 compreendendo a sequencia de aminoácidos da SEQ ID N°: 1, em que a dita proteína p51 é produzida através de tecnologia recombinante ou síntese química.

2. O anticorpo isolado do item 1, em que a ligação é determinada utilizando imunoprecipitação, imunocoloração, transferência de Western, ELISA, ou ensaios rádio-imunológicos (RIA).

3. Um kit para detectar a proteína p51 compreendendo
(i) um anticorpo isolado do item 1, e
(ii) um meio para detectar o anticorpo.

4. O kit do item 3, em que os meios para detectar o anticorpo são um ou vários de enzimas, um co-enzima, um substrato, um polipéptido, um anticorpo, ou um marcador detectável.

5. O kit do item 4, em que os meios para detectar o anticorpo são um segundo anticorpo que se liga especificamente ao anticorpo.

6. Kit do item 3, em que o anticorpo é um anticorpo monoclonal.

7. Kit do item 3, em que o anticorpo encontra-se numa forma de uma preparação purificada do anticorpo policlonal.

Também descritos são (como referência): um vector que alberga o dito gene de p51, células hospedeiras transformadas com o dito vector, e um processo de produzir a proteína p51 caracterizado por fazer crescer as ditas células hospedeiras num meio e colhendo a proteína a partir da cultura resultante.

Deveria ser compreendido que a designação p51 é utilizada apenas por uma questão de conveniência nesta referência da memória descritiva.

Além disso, o termo "gene(ADN)" no contexto da presente memória descritiva significa não apenas um ADN de dupla cadeia, mas também um ADN de cadeia simples, inclusive do componente da cadeia de sentido e cadeia de anti-sentido, e não é de modo algum limitativo do seu comprimento. Por isso, a não ser que indicado de outro modo, o gene (ADN) inclui o ADN de cadeia dupla, inclusive do ADN genómico humano, um ADN de cadeia simples compreendendo o ADNc (cadeia de sentido), um ADN de cadeia simples que possui uma sequência complementar à dita cadeia

de sentido (cadeia anti-sentido), e todos os seus fragmentos.

A representação dos aminoácidos, péptidos, sequências nucleótídicas, ácidos nucleicos, etc. por abreviaturas nesta memória descritiva está de acordo com as recomendações da IUPAC-IUB, as "Orientações para a Preparação de Especificações etc. que contêm Sequências de Ácidos Nucleicos ou Sequências de Aminoácidos " (Instituto de Patentes, Japonês, do Estados Unidos da América e Europeu da Trindade), e as convenções na utilização dos símbolos no campo do estado da técnica relacionado.

(1) O gene p51 e seu equivalente (apenas para referência)

A presente memória descritiva diz respeito a novos genes humanos que codificam para proteínas que possuem acções ou funções idênticas ou equivalentes às acções ou funções da proteína p53.

O gene foi adquirido efectuando uma PCR utilizando iniciadores recentemente estabelecidos após explorações energéticas utilizando regiões específicas seleccionadas a partir das sequências do gene conhecido até agora p53 e do gene p73 com esforços judiciosos.

Mais particularmente, efectuando uma PCR utilizando os iniciadores novos descritos nos Exemplos que aparecem a seguir, foi obtido um fragmento génico que não é idêntico, mas semelhante ao gene p53 e ao gene 73. Utilizando este fragmento de ADN como sonda, um clone de ADNc que codifica para uma proteína nova que possui uma homologia elevada com a sequência de aminoácidos da proteína p53 foi isolado com sucesso a partir de clones de ADNc seleccionados aleatoriamente a partir de uma biblioteca de ADNc do musculo esquelético humano.

A massa molecular calculada a partir da sequência de aminoácidos deduzida a partir do ADNc obtido deste modo foi cerca de 50894 Da e, por isso, os presentes inventores designaram este ADNc (ADN) "gene p51A humano (ou abreviadamente, gene p51A)" e a proteína que possui a sequência de aminoácidos codificada por este gene "proteína p51A (ou abreviadamente, proteína p51A)" por questões de conveniência.

Investigação posterior revelou que o gene codificado pelo clone de p51 ADNc possui variantes de união alternativas. Além disso, uma investigação do padrão de expressão e produção dos transcritos do gene em vários tecidos humanos revelou que os produtos de expressão (proteínas) existem unidos principalmente numa forma curta e numa forma longa.

Com base na formação de aminoácidos deduzida a partir das variantes de união do p51 ADNc, a variante curta de união é o gene (p512A gene) que codifica para a proteína que possui a dita sequência de 448 aminoácidos (massa molecular de ca 50,9 kDa) e a variante de união de forma longa é um gene que codifica para uma proteína que possui uma sequência de 641 aminoácidos (massa molecular cerca de 71,9 kDa). Nesta memória descritiva, por razões de conveniência, o último gene é designado "gene p51B humano (ou brevemente, gene p51B)" e a proteína que possui a sequência de aminoácidos codificada pelo dito gene "proteína humana p51B (ou brevemente, proteína p51B)".

Além disso, nesta memória descritiva, o dito gene p51A e gene p51B são designados colectivamente como "gene p51" e as proteínas p51A e p51B são designadas colectivamente como "proteína p51".

No que diz respeito à dita variante de união do gene p51, foi confirmada a existência de uma pluralidade de

variantes inclusive do gene defectivo de uma parte do domínio TA.

Uma investigação dos produtos de expressão destes genes p51 revelou que o produto génico p51 (proteína p51) da presente invenção apresenta actividade de activação transcricional, actividade inibidora do crescimento celular e actividade indutora de apoptose que são semelhantes às actividades da proteína p53. Além disso, a expressão do gene p51 em tecidos humanos foi verificada ser mais específica dos tecidos do que a expressão do gene p53 e, que comparada com a expressão do gene p73 que é também específica do tecido era mais ampla na distribuição no tecido apesar de existir uma sobreposição do padrão de expressão entre eles. Além disso, a mutação do gene p51 foi encontrada nos tecidos tumorais humanos ou linhas de células tumorais.

Os factos acima sugeriram fortemente que o gene humano p51 é um novo membro da família do gene p53 supressor de tumores.

Como exemplo específico do gene p51 pode ser mencionado um que possui a sequência de ADN possuída pelos clones (p51A p51B) descrita no Exemplo 1 que aparece a seguir.

Como gene que possui o clone p51A, pode ser mencionado o gene (1344 nucleótidos) que codifica para a proteína com 448 resíduos da SEQ ID N°:1 da Listagem de Sequências que aparece a seguir. Especificamente, este é um gene que possui a sequência nucleotídica que corresponde aos nucleótidos 145°-1488° da SEQ ID N°:2 que corresponde a uma grelha de leitura aberta.

A sequência nucleotídica de comprimento completo do p51A ADNc consiste em 2816 nucleótidos como indicado na SEQ ID N°:2. O gene p51A inclui os genes que contêm esta sequência nucleotídica apresentada como SEQ ID N°:2. Na

sequência nucleotídica apresentada como SEQ ID N° 2, o codão de iniciação (ATG) encontra-se situado na posição nucleotídica 145-147 e o sinal de poliadenilação (AATAA) encontra-se situado na posição 2786-2791.

A sequência de aminoácidos do resíduo 448 da proteína p51A codificado pelo gene p51A é apresentada na SEQ ID N°:1. Como apresentado, esta proteína possui um domínio de activação transcricional correspondente aos aminoácidos número 1-59, um domínio de ligação do ADN correspondente aos aminoácidos número 142-321, e um domínio de oligomerização correspondente aos aminoácidos números 353-397.

A homologia de cada um dos ditos domínios da proteína p51A com o domínio correspondente das proteínas conhecidas p53 ou p73 foi investigado com o programa FASTA utilizando o software GCG (Wisconsin Sequencing Package, Genetics Computer Group) [Person, W. R. e Lipman, D. J., Proc. Natl. Acad. Sci. USA, 85, 1435-1441(1998)]. Os resultados são apresentados na Tabela 1 (ver Figs 1 e 2). Como referência as homologias entre a proteína p53 e a proteína p73 β determinadas pelo mesmo processo são indicadas na Tabela 1.

Tabela 1

	Sequência de comprimento completo	Domínio de activação da transcrição	Domínio de ligação do ADN	Domínio de oligomerização
p51A $\leftrightarrow$ p53	36%	22%	60%	37%
p51A $\leftrightarrow$ p73 β	42%	30%	87%	65%
p53 $\leftrightarrow$ p73	28%	27%	63%	83%

Por outro lado, como gene possuído pelo clone p51B, pode ser mencionado o gene (1923 nucleótidos) que codifica para a proteína de 641 resíduos apresentada na SEQ ID N°:4 na Listagem de Sequências apresentada a seguir.

Especificamente, isto é um gene que possui a sequência nucleotídica identificada pelos nucleótidos números 145-2067 da SEQ ID N°:5, que corresponde a uma grelha de leitura aberta.

A sequência nucleotídica de comprimento completo deste p51B ADNc consiste em 2270 nucleótidos como apresentado na SEQ ID N°:5. O gene p51B inclui genes contendo a sequência nucleotídica apresentada na SEQ ID N° 5.

A sequência de aminoácidos do resíduo 641 da proteína p51B codificada pelo gene p51B é apresentada na SEQ ID N°:4. Esta proteína possui um domínio de ativação transcricional correspondente aos aminoácidos número 1-59, um domínio de ligação do ADN correspondente aos aminoácidos números 142-321, e um domínio de oligomerização correspondente aos aminoácidos números 353-397. Além disso, existe uma sequência adicional (domínio SAM) na região C-terminal, apesar dos números de aminoácidos correspondentes não poderem ser identificados. Nesta memória descritiva, a região dos aminoácidos número 353-641 inclusive deste domínio SAM é considerada como um domínio de oligomerização em sentido lato.

Tal como no caso da proteína p51A, a homologia da sequência de aminoácidos de cada um dos ditos domínios da proteína p51B em relação à sequência dos aminoácidos correspondentes da proteína conhecida p73 α foi investigada com o programa FASTA utilizando o software GCG. Os resultados são indicados na Fig. 3., as partes encaixilhadas são sequências de aminoácidos comuns à proteína p51B e proteína p73 α . É por isso claro, que a sequência de aminoácidos da proteína p51B é homóloga à sequência da proteína p73 α numa gama larga.

Assim, o gene p51 inclui um gene p51A humano que possui uma sequência nucleotídica que codifica para uma proteína que possui a sequência de aminoácidos apresentada

na SEQ ID N°:1 e um gene humano p51B que possui uma sequência nucleotídica que codifica para uma proteína que possui a sequência de aminoácidos apresentada na SEQ ID N°:4. Contudo, o gene p51 não se encontra limitado aqueles genes, mas inclui todos os homólogos daqueles genes humanos p51.

O termo “homólogo do gene p51 humano” significa qualquer membro de um grupo de genes inter-relacionados que são análogos ao dito gene p51A ou gene p51B na sequência nucleotídica e/ou características estruturais e padrão de expressão gênica ou análogos um ao outro nas suas próprias funções biológicas e produtos gênicos (proteínas) e, como tal, podem ser considerados como constituindo uma família de genes. Neste sentido, variantes de união e alelos do gene p51 humano são evidentemente incluídos no conceito do dito “homólogo”.

Como um exemplo do dito homólogo, pode ser mencionado um gene que codifica para uma proteína que possui uma sequência de aminoácidos que resulta da mutação ou modificação de um ou de uma pluralidade de locais da sequência de aminoácidos apresentada na SEQ ID N°:1 e possuindo ações ou funções semelhantes àquelas da dita proteína p51A que possui a sequência de aminoácidos apresentada na SEQ ID N°:1. O preferido é um gene que codifica para uma sequência de aminoácidos que retém pelo menos um determinado grau de homologia com a sequência de aminoácidos apresentada na SEQ ID N°:1.

O grau de homologia na sequência de aminoácidos pode ser geralmente não menos do que cerca de 45%, preferencialmente não menos do que cerca de 50%, em termos da sequência de comprimento completo de acordo com o determinado pelo programa FASTA utilizando o dito software GCG. Preferencialmente, a homologia não deveria ser inferior a um determinado valor em pelo menos um domínio de

activação transcricional, domínio de ligação de ADN e domínio de oligomerização. Por exemplo, a homologia no domínio de activação transcricional pode ser cerca de 35% ou mais elevado, preferencialmente não inferior a 45%, a homologia no domínio de ligação de ADN pode ser 88% ou mais elevado, preferencialmente, não inferior a cerca de 90%, e a homologia no domínio de oligomerização pode ser cerca de 70% ou superior, preferencialmente não inferior a cerca de 80%.

Assim, este gene inclui qualquer gene que possui uma sequência nucleotídica que codifica para uma proteína que possui uma sequência de aminoácidos derivada da SEQ ID N°:1 por supressão, substituição ou adição de um ~ aproximadamente alguns ou uma pluralidade de aminoácidos com a condição de que as qualificações acima mencionadas são satisfeitas.

A extensão de “supressão, substituição ou adição de aminoácidos” e o local ou locais envolvidos não são particularmente tão restringidos desde que a proteína modificada deste modo seja funcionalmente equivalente à proteína (proteína p51A ou p51B) possuindo a sequência de aminoácidos da SEQ ID N°:1 ou 4. Assim, o termo “actividade p51”, quando utilizado nesta especificação significa as actividades e funções possuídas pela proteína p51, representada pela proteína p51A ou p51B, da presente invenção, incluindo deste modo a actividade inibidora do crescimento da célula tumoral, actividade indutora da apoptose e função de regulação da transcrição em células entre outras.

A proteína p51 compreende a sequência de aminoácidos da SEQ ID N°:1, e é considerada possuir acções semelhantes aquelas da proteína p53 que é conhecida por ser um factor de inibição da proliferação celular. Por isso, o termo “actividade p51”, quando utilizado nesta especificação

referindo-se a acções e funções da proteína p51 pode ser novamente expresso nos mesmos termos que os aplicados às actividades conhecidas e características funcionais da proteína p53.

Como acções e funções da proteína p53, podem ser mencionadas uma função de regulação da transcrição, uma função de transdução de sinal que é expressa através da sua ligação a outras proteínas intracelulares, a função como um componente do complexo de proteína relacionado com a replicação de ADN, uma função de ligação de ADN, actividade de exonuclease, etc., e como as funções expressas pela interacção composta das ditas várias funções, uma função que interrompe um ciclo celular, uma função indutora de apoptose, e uma função de reparação de ADN, uma função de controlo da replicação e/ou uma função indutora de diferenciação nas células. É considerado que a proteína p51 da presente invenção possui algumas ou todas estas acções e funções.

A modificação de uma sequência de aminoácidos pode ser espontânea, por ex. uma mutação espontânea ou modificação pós-tradução, mas pode ser induzida artificialmente com base num gene nativo.

A presente memória descritiva compreende todos os genes modificados que codificam para proteínas que possuem as características acima mencionadas da proteína p51 da invenção independentemente da causa ou meios de mutação ou de modificação.

Os meios de efectuar tais modificações artificiais incluem técnicas de engenharia genética tais como mutagénese específica do local (dirigida) [Methods in Enzymology, 154:350, 367-382(1987); ditto 100:468(1983); Nucleic Acid Res., 12:9441(1984); Zoku Seikagaku Jikken Koza 1 "Idenshi Kenkyuho II"[Experimental Biochemistry Series 1 "Methods for Gene Research II"(editado pela

Sociedade Bioquímica Japonesa), p. 105 (1986)], etc. e técnicas de química sintética, tal como o processo de fosfotriéster e o processos do fosfoamidato [J. Am. Chem. Soc., 89:4801(1967); ditto 91:3350(1968); Science, 150:178(1968); Tetrahedron Lett., 22:1859(1981); ditto 24:245(1983)] assim como uma combinação adequada de tais técnicas. Mais especificamente a síntese do ADN pode ser efectuada quimicamente pelo processo de fosforamídida ou pelo processo do triéster, ou num sintetizador de oligonucleótidos automático comercial. O fragmento de cadeia dupla pode ser obtido sintetizando cadeias complementares e hibridizando-as em condições adequadas, ou pode ser obtido através de uma cadeia simples sintetizada quimicamente adicionando uma cadeia de ADN utilizando uma ADN polimerase conjuntamente com sequências de iniciadores adequadas.

Como exemplos específicos do gene pode ser mencionado o gene que possui uma sequência nucleotídica correspondente aos nucleótidos número 145-1488 da sequência nucleotídica apresentada na SEQ ID N°:2 e o gene que possui uma sequência nucleotídica correspondente aos nucleótidos números 145-2067 da SEQ ID N°:5. Cada uma destas sequências nucleotídicas representa um exemplo de combinação dos codões que codificam para os resíduos de aminoácidos respectivos da sequência de aminoácido apresentada na SEQ ID N°:1 ou 4. Por isso, o gene não se encontra limitado a genes que possuem tais sequências nucleotídicas específicas, mas podem ter sequências nucleotídicas desenhadas utilizando uma combinação de codões opcionais para cada resíduo de aminoácidos. Os codões de selecção podem ser efectuados de uma forma rotineira, por exemplo com referência à frequência de utilização de cada codão pelo hospedeiro a ser utilizado [Nucleic Acids Res., 9, 43(1981)].

Além disso, como mencionado acima, o gene inclui uma sequência nucleotídica que possui um grau definido de homologia em relação à sequência nucleotídica correspondente aos nucleótidos números 145-1488 [aqui referido algumas vezes brevemente como sequência nucleotídica (145-1488)] da sequência nucleotídica apresentada na SEQ ID N°:2.

Como exemplo de um tal gene, pode ser mencionado um gene que possui uma sequência nucleotídica capaz de se hibridizar com um ADN que possui a dita sequência nucleotídica (145-1488) em condições severas, por exemplo em 0,1% SDS- contendo 0,2XSSC a 50°C ou em 0,1% SDS contendo 1xSSC a 60°C.

O gene pode ser facilmente produzido e adquirido através das técnicas padrão de engenharia genética [Molecular Cloning 2d Ed, Cold Spring Harbor Lab. Press (1989); Zoku Seikagaku Jikken Koza 1 "Idenshi Kenkyuho I, II, III" [Supplemental Biochemical Experimental Series 1 "Methods for Gene Research I, II, III" (editado pela Sociedade de Bioquímica Japonesa), (1986), etc.] baseada na informação da sequência nos exemplos específicos apresentados na SEQ ID N°:2.

Mais particularmente, o gene objecto pode ser adquirido construindo uma biblioteca de ADNc a partir de uma fonte adequada em que o gene da invenção pode ser expresso e seleccionando o clone desejado a partir desta biblioteca de ADNc utilizando uma sonda adequada ou anticorpo específico do gene da invenção da maneira conhecida *per se* [Proc. Natl. Acad. Sci., USA, 78:6613(1981); Science, 222:778(1983), etc.].

No procedimento acima, a fonte de ADNc inclui, mas não se encontra limitada a várias células ou tecidos em que o gene da invenção é expresso e células desta origem são

cultivadas. Isolamento do ARN completo de uma tal fonte, isolamento e purificação de ARNm, aquisição de ADNc, e sua clonagem podem ser efectuadas de forma rotineira. Bibliotecas de ADN encontram-se também comercialmente disponíveis. Na prática da presente invenção, tais bibliotecas de ADNc comerciais, por exemplo aqueles disponibilizados pela Clontech Lab Inc. podem também ser utilizadas.

O processo de rastreio do gene de uma biblioteca de ADNc não se encontra também particularmente restringido, mas os processos convencionais podem ser utilizados selectivamente.

Especificamente, a título de exemplo, pode-se mencionar a selecção de um clone de ADNc através de uma técnica de imunorastreio utilizando um anticorpo específico contra a proteína produzida pelo ADNc, a técnica de hibridização em placa, ou hibridização de colónia utilizando uma sonda possuindo uma afinidade de ligação selectiva para a sequência de ADN objectivo, ou uma sua combinação.

Quanto à sonda a ser utilizada no procedimento acima é geralmente vantajoso utilizar um ADN sintetizado quimicamente de acordo com a informação da sequência nucleotídica no gene, mas este mesmo gene que já foi adquirido, ou um seu fragmento pode evidentemente ser utilizado com vantagem como a dita sonda. Além disso, o iniciador de sentido e o iniciador anti-sentido estabelecido com base na informação da sequência nucleotídica no gene p51 da presente invenção pode ser utilizado como uma sonda de rastreio.

Para a aquisição da amplificação do ADN/ARN do gene através do processo de PCR [Science, 230, 1350(1985)], ou uma sua modificação pode também ser utilizada com vantagem. Particularmente em circunstâncias em que um ADNc de

comprimento completo pode dificilmente ser obtido a partir de uma biblioteca, o processo RACE [amplificação rápida de extremidades de ADNc] [Jikken Igaku (Experimental Medicine), 12(6):35(1994)], em particular o processo 5'-RACE [Frohman, M. A., et al., Proc. Natl. Acad. Sci., USA., 8:8998(1988)], pode ser utilizado com vantagem.

Os iniciadores para serem utilizados em tais processos PCR podem ser judiciosamente estabelecidos de acordo com a informação da sequência no gene p51 que foi aqui revelada e que pode ser sintetizado pelo processo convencional. O isolamento e a purificação do fragmento de ADN amplificado ou ARN podem ser efectuados através de técnicas convencionais como mencionado anteriormente por electroforese em gel ou hibridização.

A sequência nucleótídica do gene p51 ou qualquer dos vários fragmentos de ADN que podem ser obtidos como acima podem ser determinados de forma rotineira, por exemplo, através do processo de didesóxi [Proc. Natl. Acad. Sci.; USA., 74: 5463(1977)], o processo de Maxam-Gilbert [Methods in Enzymology, 65:499(1980)] ou, de forma mais expediente, através de um kit comercial de sequenciação.

Com o gene p51, por exemplo utilizando uma sequência nucleótídica parcial ou de comprimento completo deste gene a expressão ou a não expressão do gene p51 num corpo humano ou outro corpo individual ou em vários dos seus tecidos pode ser especificamente detectada.

Esta detecção pode ser efectuada de forma rotineira. Por exemplo, determinação a nível celular por amplificação de ARN por RT-PCR [reação em cadeia da polimerase transcrita reversa; E. S. Kawasaki et al. Amplification of ARN In PCR Protocol, A Guide to Methods and applications, Academic Press, Inc., San Diego, 21-27(1991)], análise de transferência de Northern [Molecular Cloning, Cold Spring

Harbor Lab.(1989)], in situ RT-PCR [Nucl. Acids Res., 21, 3159-3166(1993)] ou hibridização in situ, o processo NASBA [amplificação baseada na sequência de ácidos nucleicos, Nature, 350, 91-92(1991)] e outras técnicas podem ser mencionadas. O preferido é o processo RT-PCR-SSCP.

Os iniciadores utilizados no procedimento da PCR não se encontram particularmente restringidos desde que o gene p51 (inclusive de um ADN parcial) possam ser especificamente amplificados, e possam ser judiciosamente estabelecidos com base na informação da sequência no gene p51. Geralmente, por exemplo, podem ser utilizados iniciadores possuindo cada um uma sequência parcial do gene p51 e um comprimento que vai de cerca de 10 a 35 nucleótidos, preferencialmente 15-30 nucleótidos.

Assim, o gene inclui fragmentos de ADN que podem ser utilizados como iniciadores específicos e/ou sondas específicas para a detecção do gene humano p51.

Um tal fragmento de ADN pode ser definido como o ADN caracterizado pela sua capacidade de se hibridizar com um ADN que possui a dita sequência nucleotídica (145-1488) em condições severas. As condições severas mencionadas acima podem ser as condições convencionais utilizadas para iniciadores e sondas e, neste caso não se encontram particularmente restringidas, mas as condições acima mencionadas, nomeadamente 0,1% de SDS contendo 0,2XSSC a 50°C ou 0,1% SDS contendo 1XSSC a 60°C, podem por exemplo ser mencionadas.

Com o gene p51 humano pode ser produzida facilmente a proteína compreendendo o produto génico correspondente (proteína p51), numa grande escala, e com uma boa reprodutibilidade utilizando a tecnologia convencional de engenharia genética.

(2) A proteína p51

A presente invenção disponibiliza por isso a proteína p51 codificada pelo gene acima descrito. A proteína da presente invenção é a proteína p51A compreendendo a sequência de aminoácidos apresentada na SEQ ID N°:1. Deverá no entanto ser entendido que a proteína da presente invenção não se encontra limitada à dita proteína p51A e proteína p51B, mas inclui os seus homólogos.

Mais particularmente, podem ser mencionados os produtos génicos dos ditos homólogos do gene p51 (genes relacionados p51 incluindo variantes de união e alelos).

A proteína da presente invenção pode ser preparada, através da tecnologia de ADN recombinante convencional [por ex. Science, 224, 1431(1984); Biochem. Biophys. Res. Comm., 130, 692(1985); Proc. Natl. Acad. Sci., USA., 80, 5990(1983), etc.] com base na informação da sequência do gene p51 humano disponibilizada pela presente invenção.

(3) O polipéptido contendo um ou vários domínios funcionais da proteína p51

A presente especificação é ainda dirigida para um polipéptido contendo uma região parcial da dita proteína p51.

O polipéptido possui preferencialmente a sequência de aminoácidos correspondente a qualquer uma das ditas várias regiões funcionais da proteína p51 e, especificamente pode ser mencionado um polipéptido possuindo uma sequência de aminoácidos correspondente a pelo menos um domínio seleccionado a partir do grupo constituído pelo domínio de activação transcricional, domínio de ligação de ADN e domínio de oligomerização da proteína p51.

Como mencionado acima, as localizações do domínio de activação transcricional, o domínio de ligação do ADN e

domínio de oligomerização da proteína p51 podem ser identificados pelos aminoácidos número 1-59, aminoácidos número 142-321 e pelos aminoácidos número 359-397, respectivamente, da sequência de aminoácidos da proteína p51A que é apresentada na SEQ ID N°:1.

Por isso, o polipéptido inclui o seguinte

(i) Um polipéptido possuindo a sequência de aminoácidos correspondente aos aminoácidos número 1-59 da SEQ ID N°: 1(aqui designado brevemente como sequência de aminoácidos 1 (1-59) e seu equivalente.

O equivalente mencionado mesmo acima inclui qualquer polipéptido que possui uma sequência de aminoácidos derivada da dita sequência de aminoácidos 1 (1-59) por supressão, substituição ou adição de um ou de uma pluralidade de aminoácidos e possuindo uma função de activação transcricional. A extensão da modificação ou da mutação da sequência de aminoácidos não se encontra particularmente restringida desde que o polipéptido modificado retenha a dita função de activação transcricional. Preferencialmente, contudo, a homogeneidade da sequência modificada deste modo em relação à sequência de aminoácidos 1 (1-59) não é menos do que cerca de 35%, em particular não inferior a cerca de 45%.

(ii) Um polipéptido possuindo uma sequência de aminoácidos identificada pelos aminoácidos número 142-321 da SEQ ID N°: 1(aqui designado brevemente como sequência de aminoácidos 1 (142-321) e seus equivalentes.

O equivalente mencionado acima inclui qualquer polipéptido que possui uma sequência de aminoácidos derivada da dita sequência de aminoácidos 1 (142-321) por supressão, substituição ou adição de um ou de uma pluralidade de aminoácidos e possuindo a dita função de ligação do ADN. A extensão da modificação ou da mutação da sequência de aminoácidos não se encontra particularmente restringida

desde que a sequência assim modificada retenha a dita actividade de ligação do ADN, mas a sequência preferencialmente possui uma homologia não inferior a cerca de 88%, mais preferencialmente não menos do que 90% em relação à sequência de aminoácidos 1 (142-321).

(iii) Um polipéptido possuindo uma sequência de aminoácidos identificada pelos aminoácidos número 353-397 da SEQ ID N°: 1(aqui designado brevemente como sequência de aminoácidos 1 (353-397) e seus equivalentes.

Este equivalente inclui um polipéptido que possui uma sequência de aminoácidos derivada da dita sequência de aminoácidos 1 (353-397) por supressão, substituição ou adição de um ou de uma pluralidade de aminoácidos e possuindo uma função de oligomerização, por exemplo o domínio de oligomerização em sentido lato (i.e. os aminoácidos número 353-641 da SEQ ID N°4) da proteína p51. A extensão da modificação ou da mutação da sequência de aminoácidos não está particularmente restringida desde que a sequência assim modificada retenha a função de oligomerização, mas preferencialmente retenha uma homologia não inferior a cerca de 70%, mais preferencialmente não menos do que 80% em relação à sequência de aminoácidos 1 (353-397).

O polipéptido pode ser um polipéptido que contém qualquer das ditas sequências de aminoácido 1(1-59) ou um seu equivalente, a dita sequência de aminoácidos 1(142-321) ou um seu equivalente, e dita sequência de aminoácidos 1(353-397) ou um seu equivalente numa região ou um polipéptido que contém duas ou mais das ditas sequências de aminoácidos numa combinação opcional como uma região contínua, ou uma região descontínua.

A presente especificação inclui ainda genes(ADN) que possuem sequências oligonucleótídicas que codificam para tais polipéptidos. Mais particularmente, a sequência

nucleotídica que codifica para a dita sequência de aminoácidos 1(1-59) é a sequência nucleotídica correspondente aos nucleótidos números 145-321 da SEQ ID N°:2; a sequência nucleotídica que codifica para a dita sequência de aminoácidos 1(142-321) é a sequência nucleotídica correspondente aos nucleótidos números 568-1107 da SEQ ID N°:2; e a sequência nucleotídica que codifica para a dita sequência de aminoácidos 1(353-397) é a sequência nucleotídica correspondente aos nucleótidos números 1201-1335 da SEQ ID N°:2.

(4) Processo para produzir a proteína p51 e os materiais para serem utilizados na sua produção

A presente memória descritiva disponibiliza ainda um processo para produzir a dita proteína p51 e os materiais a serem utilizados para a sua produção, por exemplo, um vector que alberga o dito gene e células hospedeiras transformadas com o dito vector.

Mais particularmente, a produção da dita proteína é efectuada de acordo com o processo que compreende construir um vector recombinante (vector de expressão) em que o gene que codifica para a proteína desejada pode ser expresso, transformando células hospedeiras com a construção resultante, cultivando os transformantes obtidos deste modo, e colhendo a proteína desejada a partir da cultura obtida.

Como ditas células hospedeiras podem ser utilizadas quer células eucarióticas, ou células procarióticas.

As células eucarióticas incluem células hospedeiras de vertebrados e de leveduras entre outras. Entre as primeiras células a linha celular do macaco COS [Cell, 23:175(1981)], células do ovário do hamster chinês e a sua linha defectiva em dihidrofolase reductase [Proc. Natl. Acad. Sci., USA.,

77:4216(1980)] podem ser mencionadas como exemplos. Em relação às últimas células, células de leveduras que pertencem ao gene Saccharomyces podem ser mencionadas como exemplos destas e não são escolhas exclusivas.

Como hospedeiro procariótico, vários procariontes que são habitualmente utilizados, tais como Eschericia coli e Bacillus subtilis podem ser brevemente utilizados. As células hospedeiras preferidas são aquelas derivadas de Eschericia coli, em particular células de E. Coli K12.

O vector de expressão não se encontra particularmente restringido desde que albergue o gene da presente invenção e permita a expressão do dito gene, mas seja geralmente seleccionado em relação aos tipos de células hospedeiras.

Quando as células de um vertebrado são utilizadas como células hospedeiras, geralmente um vector de expressão que possui uma região promotora a montante do gene da invenção, podem ser utilizados quando necessário o local de união do ARN, local de poliadenilação e sequência de terminação da transcrição e quando necessário pode incluir ainda a origem de replicação. Como exemplo de um tal vector de expressão pode ser mencionado pSV2dhfr [Mol. Cell. Biol., 1,854(1981)] possuindo um promotor precoce SV40.

Quando as células eucarióticas de microrganismos tais como leveduras são utilizadas como células hospedeiras, o vector de expressão que pode ser utilizado inclui pAM82 [Proc. Natl. Acad. Sci., USA., 80, 1(1983)] que possui o promotor do gene da fosfatase ácida, e o vector para utilização na presente invenção pode ser preparado inserindo o gene da invenção a montante deste promotor. Preferencialmente, pode ser utilizado um vector de fusão obtido por hibridização com um gene procariótico como exemplos específicos de um tal vector podem ser mencionados, pGEX-2TK e pGEX-4T-2 possuindo cada um domínio

GST com um peso molecular de 26000 (derivado de *S. Japonicum*).

Quando são utilizadas células procarióticas como células hospedeiras, o vector de expressão pode ser por exemplo um vector equipado com uma região promotora e uma sequência SD(Shine-Dalgarno) a montante do gene de modo que o gene possa ser aí expresso e ainda um codão de iniciação (por ex. ATG) necessário para a iniciação da síntese de proteína. Particularmente, quando são utilizadas células de Eschericia coli (por ex. E. Coli K12) como células hospedeiras, geralmente pBR322 como tal ou modificada é frequentemente utilizada como vector. Contudo, estas não são escolhas exclusivas, mas outras estirpes bacterianas exclusivas e vectores conhecidos podem também ser utilizados. Como promotores, podem ser utilizados o promotor triptofano (trp), promotor Ipp, o promotor lac, PL/PR, etc.

O processo para introduzir o dito vector de expressão nas células hospedeiras (processo de transformação) não se encontra também particularmente restringido, mas podem ser utilizados quaisquer dos vários processos padronizados.

A cultura do transformante resultante pode também ser efectuada de uma forma rotineira. Por uma tal cultura a proteína objecto codificada pelo gene é expressa, produzida, e acumulada na célula transformada ou secretada de forma extracelular, ou sobre a membrana celular.

O meio para a dita cultura pode ser judiciosamente seleccionado entre os meios de cultura convencionais de acordo com o tipo das células hospedeiras adoptadas, e a cultura pode também ser efectuada em condições adequadas para o crescimento das células hospedeiras.

A proteína recombinante produzida deste modo pode ser opcionalmente isolada e purificada através de vários procedimentos de isolamento utilizando as suas propriedades

físicas, químicas ou outras [Seikagaku (Biochemical) Data Book II, pág. 1175-1259, 1ª Ed., 1ª Impressão, 23 de Junho de 1980, Tokyo Kagaku Dojin; Biochemistry, 25(25):8274(1986); Eur. J. Biochem., 163:313(1987); etc.].

Os procedimentos mencionados acima incluem especificamente o tratamento padrão de reconstituição, tratamento com um agente de precipitação de proteína (precipitação por sais), centrifugação, processo de choque osmótico, quebra sónica, ultrafiltração, vários tipos de cromatografia, por ex. cromatografia por peneiros moleculares (filtração por gel), cromatografia de adsorção, cromatografia de permuta iónica, cromatografia de afinidade, cromatografia líquida de elevada eficiência (HPLC), etc. diálise, e suas combinações. O processo particularmente preferido é a cromatografia de afinidade utilizando uma coluna conjugada com um anticorpo específico contra a proteína.

No desenho do gene objectivo que codifica para a proteína da sequência nucleotídica do gene humano p51A identificado pela sequência nucleotídica (145-1488) na SEQ ID N°:2 ou a sequência do gene p51B humano identificado pela sequência nucleotídica (145-2067) na SEQ ID N°:5 podem ser utilizados vantajosamente. Se desejado, este gene pode ser utilizado com os codões que designam os respectivos resíduos de aminoácidos alterados judiciosamente.

Além disso, a modificação parcial da sequência de aminoácidos codificada pelo gene 51A humano ou gene p51B por substituição, supressão, ou adição de algum dos resíduos de aminoácidos ou uma determinada sequência parcial pode ser conseguida através de várias técnicas mencionadas anteriormente, por exemplo através de mutagénese específica do local.

A proteína pode ser sintetizada através de tecnologia padrão para síntese química de acordo com a sequência de aminoácidos apresentada na SEQ ID N°:1.

Esta tecnologia inclui processos de fase líquida e de fase sólida para a síntese de péptidos.

Mais particularmente, a tecnologia sintética inclui o chamado processo de alongamento por passos em que um aminoácido após outro são sequencialmente acoplados de acordo com a informação da sequência de aminoácidos e o processo de condensação dos fragmentos em que cada fragmento, cada um constituído por vários aminoácidos são sintetizados previamente e depois acoplados. O polipéptido pode ser sintetizado, através de qualquer dos processos alternativos acima.

O processo de condensação para ser utilizado na síntese de péptidos acima pode também ser o convencional que inclui, mas não se encontra limitado ao processo da azida, processos do anidrido ácido misturado, processo DCC, processo do éster activado, processo redox, processo DPPA (difenílfosforilo azida), processo de DCC+aditivo (1-hidróxibenzotriazole, N-hidróxisuccinamida, N-hidróxi-5-norbornene-2,3-dicarboximida ou semelhantes) e processo de Woodward.

O solvente para ser utilizado nestes processos pode ser judiciosamente seleccionado a partir dos solventes habituais que são bem conhecidos dos peritos no estado da técnica da condensação de péptidos. A título de exemplo, N,N-dimetilformamida (DMF), dimetilsulfóxido (DMSO), hexafluoramida, dioxano, tetrahydrofurano(THF), acetato de etilo, etc. e misturas destes solventes. Os grupos carboxilo de aminoácidos ou péptidos que não são para serem envolvidos na reacção para a dita síntese peptídica podem ser geralmente protegidos por esterificação, por exemplo na forma de um éster alquílico inferior, por ex. éster

metílico, éster etílico, éster ter-butílico ou análogos, ou um éster aralquílico, por ex. éster benzílico, éster p-metóxibenzílico, éster p-nitrobenzílico ou análogos.

O aminoácido que possui um grupo funcional na sua cadeia lateral, por exemplo grupo hidróxilo do resíduo da tirosina, pode ser protegido com um acetilo, benzilo, benzilóxicarbonilo, ter-butil ou outro grupo, apesar desta protecção não ser indispensável. Além disso, o grupo guanidina de um resíduo de arginina, por exemplo, pode ser protegido com um grupo protector adequado, tal como nitro, tosil, p-metóxibenzilsulfonilo, metileno-2-sulfonil, benzilóxi-carbonilo, isobornilóxicarbonilo, adamantiloxicarbonilo, ou semelhantes.

As reacções de desprotecção de tais aminoácidos protegidos, péptidos e proteína produto final da presente invenção para a remoção dos grupos protectores podem também ser efectuadas através do processo convencional, por exemplo o processo de redução catalítica, ou o processo que utiliza amónia líquida/sódio, fluoreto de hidrogénio, brometo de hidrogénio, cloreto de hidrogénio, ácido trifluoroacético, ácido acético, ácido fórmico, ácido metanossulfónico ou semelhantes.

A proteína produzida deste modo pode ser purificada através do processo que é convencionalmente utilizado no campo da química dos péptidos, tais como os ditos vários processos, tal como cromatografia de permuta iónica, cromatografia de partição, cromatografia de permeação em gel, cromatografia de distribuição em contra-corrente, etc.

A proteína pode ser utilizada com vantagem como um imunogénio na preparação de um anticorpo específico em relação à proteína p51, e utilizando um tal imunogénio, o anti-soro desejado (anticorpo policlonal) e o anticorpo monoclonal podem ser adquiridos.

A tecnologia de produção de anticorpos como tal é bem compreendida pelos peritos no estado da técnica e, na prática da presente invenção, também os processos convencionais da presente invenção podem ser utilizados [por ex. Zoku Seikagaku Jikken Koza (Supplemental Biochemical Experimental Series), Methods for Immunobiochemical Research, ed. pela Sociedade de Bioquímica Japonesa (1986)]. O anticorpo obtido deste modo pode ser utilizado com vantagem, por exemplo na purificação da proteína p51 e em ensaios imunológicos ou na caracterização da proteína.

Além disso, a proteína encontra aplicação no campo farmacêutico, no fabrico de produtos farmacêuticos contendo-os como um componente activo.

(5) Composições farmacêuticas contendo a proteína p51

A presente especificação é por isso dirigida ainda a produtos farmacêuticos contendo a proteína.

A proteína acima mencionada inclui o seu sal farmaceuticamente aceitável. Um tal sal inclui metais alcalinos não tóxicos, metais alcalino-terrosos e sais de amónio, tal como sódio, potássio, lítio, cálcio, magnésio, bário e sais de amónio. Além disso, o dito sal inclui sais de adição não tóxicos possíveis de obter fazendo reagir o péptido da invenção com um ácido orgânico ou inorgânico adequado. Os sais de adição não tóxicos representativos são cloridrato, bromohidrato, bissulfato, acetato, oxalato, valerato, oleato, laurato, borato, benzoato, lactato, fosfato, p-toluenossulfonato (tosilato), citrato, maleato, fumarato, succinato, tartarato, sulfonato, glicolato, ascorbato, benzenossulfonato e naftalato, entre outros. A presente especificação compreende ainda uma composição farmacêutica ou forma farmacêutica que contém uma

quantidade farmacêuticamente eficaz da proteína como princípio activo conjuntamente com um veículo ou diluente farmacêutico não tóxico adequado.

O veículo farmacêutico que pode ser utilizado na dita composição farmacêutica (ou forma farmacêutica) inclui os diluentes e excipientes que são habitualmente utilizados de acordo com o modo de utilização da preparação farmacêutica, tal como enchimento, formador de volume, ligante, humidificante, desintegrante, lubrificante, etc., e estes são selectivamente utilizados de acordo com a forma unitária da preparação farmacêutica.

A preparação farmacêutica preferida particularmente é produzida utilizando várias substâncias de formulação que podem ser incorporadas na preparação convencional da proteína, tal como o estabilizante, o bactericida, o tampão, agente de isotonização, o agente quelante, o agente de controlo do pH, agente tensioactivo, etc., em proporções adequadas.

O estabilizador mencionado acima inclui, mas não se encontra limitado à albumina do soro humano, vulgares L-aminoácidos, sacarídeos e derivados de celulose e estes podem ser utilizados independentemente ou em combinação com um agente tensioactivo ou semelhantes. Particularmente na combinação utilizada, a estabilidade do princípio activo pode ser ainda melhorada em determinados casos.

Os L-aminoácidos mencionados acima não se encontram particularmente restringidos, mas podem ser glicina, cisteína, ácido glutâmico etc.

Os sacarídeos acima mencionados incluem monossacarídeos tais como glucose, manose, galactose, frutose, etc. álcoois de açúcares tais como manitol, inositol, xilitol, etc., dissacarídeos tais como sacarose, maltose, lactose etc., polissacarídeos tais como dextrano, hidróxipropilamido, condroitina, sulfato, ácido

hialurónico, etc., e seus derivados. O agente tensioactivo não se encontra também particularmente restringido. Assim, podem ser mencionados os agentes tensioactivos iónicos e os agentes tensioactivos não iónicos, tais como agentes tensioactivos do éster alquílico de sorbitano de polióxido de etileno glicol, éter alquílico de polióxido de etileno, éster de monoacilo de sorbitano e a série de glicerídeos de ácidos gordos.

Os derivados de celulose não se encontram também particularmente restringidos, mas também podem ser mencionados metilcelulose, etilcelulose, hidróxietylcelulose, hidróxiethylcelulose, hidróxiethylmetilcelulose, carbóximetilcelulose de sódio etc.

O nível adequado de adição do sacarídeo por 1 µg de princípio activo não é menos do que 0,0001 mg, preferencialmente cerca de 0,01 ~ 10 mg. O nível de adição de agente tensioactivo por 1 pg de princípio activo pode ser de forma adequada não inferior a cerca de 0,00001 mg, preferencialmente cerca de 0,0001 ~ 0,1 mg. O nível de adição de albumina do soro humano por 1 µg de princípio activo pode ser de forma adequada não inferior a cerca de 0,0001 mg, preferencialmente algures na gama de cerca de 0,001-0,1 mg. O nível de adição do dito aminoácido por pg do princípio activo é adequadamente de cerca de 0,001-10 mg. O nível de adição do derivado de celulose por pg de princípio activo é de forma adequada não menos do que cerca de 0,00001 mg, preferencialmente cerca de 0,001-0,1 mg.

A quantidade de princípio activo na forma farmacêutica pode ser liberalmente seleccionada a partir de uma gama larga, mas pode ser judiciosamente seleccionada a partir da gama de geralmente de cerca de 0,00001-70% em peso, preferencialmente de cerca de 0,0001-5 % em peso.

A forma farmacêutica pode ser suplementada com vários aditivos, tais como um tampão, um agente de isotonização, um agente quelante, etc. O tampão mencionado acima inclui ácido bórico, ácido fosfórico, ácido acético, ácido cítrico, ácido ϵ -aminocaproico, ácido glutâmico, etc. e/ou os sais correspondentes (sais de metais alcalinos ou sais de metais alcalino-terrosos, por ex., o sal de sódio, sal de potássio, sal de cálcio, sal de magnésio, etc.). O agente de isotonização inclui, mas não está limitado a cloreto de sódio, cloreto de potássio, açúcares e glicerina. O agente quelante inclui edetato de sódio e ácido cítrico, entre outros.

A composição farmacêutica pode ser utilizada não apenas na forma que está de uma solução, mas também disponibilizada na forma de um produto liofilizado que pode ser preservado e reconstituído extemporaneamente com água ou uma solução tampão inclusive de soro fisiológico numa concentração adequada.

A forma unitária da composição farmacêutica pode ser seleccionada a partir de várias alternativas de acordo com o propósito terapêutico e inclui formas farmacêuticas sólidas, tais como comprimidos, pílulas, pós, pós finos, grânulos, cápsulas, etc. e formas líquidas, tais como soluções, suspensões, emulsões, xaropes e elixires. Estas formas de dosagem podem ser ainda classificadas como oral, parenteral, transnasal, vaginal, rectal (supositório) e formas sublinguais, unguentos e outros produtos, e cada produto pode ser manufacturado de acordo com a formulação estabelecida e processo de moldagem/processamento.

Tomando o fabrico dos comprimidos como exemplo, o veículo farmacêutico que pode ser utilizado inclui vários excipientes tais como lactose, sacarose, cloreto de sódio, glucose, ureia, amido, carbonato de cálcio, caolino, celulose cristalina, ácido silícico, fosfato de potássio,

etc.; ligantes tais como água, etanol, propanol, xarope simples, solução de glucose, solução de amido, solução de gelatina, carboximetilcelulose, hidróxipropilcelulose, metilcelulose, polivinilpirrolidona, etc. desintegrantes tais como carboximetilcelulose de sódio, carboximetilcelulose de cálcio, hidróxipropilcelulose de baixo grau de substituição, amido seco, alginato de sódio, pó de agar, pó de laminaran, hidrogenocarbonato de sódio, carbonato de cálcio, etc.; agentes tensioactivos tais como óxido de polietileno sorbitano, ésteres de ácidos gordos, lauril sulfato de sódio, estearil monoglicerídeo, etc.; inibidores de desintegração, tais como sacarose, estearina, manteiga de cacau, óleo hidrogenado, etc.; promotores de absorção tais como bases de amónio quaternárias, laurilsulfato de sódio, etc., humidificantes, tais como glicerina, amido, etc.; adsorventes tais como amido, lactose, caolino, bentonite, sílica coloidal, etc.; e lubrificantes tais como talco purificado, sais de estearato, pó de ácido bórico, polietileno glicol etc.

Além disso, os comprimidos podem ser opcionalmente revestidos com um material de revestimento habitual para fornecer comprimidos revestidos com açúcar, comprimidos revestidos com gelatina, comprimidos entéricos revestidos, comprimidos revestidos com película, etc. ou mesmo processados em comprimidos de multicamadas, tais como comprimidos de dupla camada.

Pílulas podem ser fabricadas utilizando vários veículos farmacêuticos inclusivé de excipientes tais como glucose, lactose, amido, manteiga de cacau, óleo vegetal hidrogenado, caolino, talco, etc.; ligantes tais como goma arábica, pó de tragacanto, gelatina, etanol, etc.; e desintegrantes tais como laminaran, agár etc.

As cápsulas podem ser preparadas misturando o ingrediente activo da presente invenção com os ditos vários

veículos farmacêuticos e invólucros de cápsulas de enchimento, tais como invólucros de cápsulas de gelatina dura, ou invólucros de cápsulas de gelatina mole com a composição resultante.

Formas líquidas para administração oral incluem soluções farmacêuticamente aceitáveis, emulsões, xaropes, elixires, etc. preparadas utilizando o diluente inerte convencional, tal como água, e pode conter ainda agentes auxiliares, tais como agentes molhantes, emulsionantes e agentes de suspensão. Estas formas podem ser fabricadas, através do processo convencional.

Formas líquidas para administração parenteral, ou não parenteral tais como soluções estéreis aquosas, ou não aquosas, emulsões e suspensões podem ser fabricadas utilizando um diluente tal como água, álcool etílico, propileno glicol, polietileno glicol, álcool isoestearílico etoxilado, álcool isoestearílico polioxilado, éster de ácido gordo de polióxido de etileno sorbitano, um óleo vegetal, por ex. azeite, e pode ser formulado com um éster orgânico injectável, tal como oleato de etilo. Além disso, tais preparações podem ser suplementadas com um solubilizante convencional, tampão, agente molhante, emulsionante, agente de suspensão, conservante, dispersante e outros aditivos.

A esterilização pode ser efectuada por filtração através de um filtro bacteriano, formulação de um bactericida, irradiação, aquecimento, ou semelhante. Além disso, as ditas preparações podem ser processadas em formas sólidas estéreis que podem ser extemporaneamente dissolvidas em água estéril, ou num meio de esterilização adequado.

No fabrico de supositórios rectais ou de uma forma para administração vaginal pode ser utilizado uma base de unguento farmacêutico, tal como polietileno glicol,

manteiga de cacau, um álcool superior, um éster de álcool superior, gelatina, glicerídeo semi-sintético ou semelhante.

No fabrico de unguentos, inclusive de pastas, cremes e geles podem ser utilizados tais diluentes, tais como petrolato, parafina, glicerina, derivados de celulose, propileno glicol, polietileno glicol, óleo de silicone, bentonite e óleos vegetais, tais como azeite.

Podem ser preparadas composições para administração transnasal ou sublingual com excipientes padrão bem conhecidos de maneira convencional.

Quando necessário, agentes de coloração, conservantes, agentes aromatizantes, correctores, adoçantes, e outras substâncias medicinais podem ser incorporadas em produtos farmacêuticos da presente invenção.

O método de administração do dito produto farmacêutico não se encontra particularmente restringido, mas deverá ser judiciosamente seleccionado de acordo com a forma de dosagem, factores do doente, tal como idade e sexo, gravidade da doença e outros factores. Por exemplo, os ditos comprimidos, pílulas, soluções, suspensões, emulsões, grânulos e cápsulas são administrados oralmente. O produto parenteral é utilizado isoladamente ou misturado com uma convenção convencional, tal como glucose ou infusão de aminoácidos, e administrada intravenosamente ou, quando necessário, administrada isoladamente por via intramuscular, por via intradérmica, por via subcutânea, ou por via intraperitoneal. Os supositórios são administrados por via rectal; a preparação vaginal é administrada dentro da vagina. A preparação transnasal é administrada nas narinas; comprimidos sub-linguais são administrados na boca; e unguentos são administrados topicamente por via transdérmica.

A quantidade de proteína no produto farmacêutico e na sua forma farmacêutica não se encontra particularmente restringida, mas pode ser judiciosamente seleccionada a partir de uma larga gama de acordo com o efeito terapêutico esperado, método de administração, decurso do tratamento, ou duração, factores do doente, tais como idade e sexo, e outros factores. Geralmente, a dosagem é geralmente cerca de 0,01 µg -10 mg/kg de peso corporal/dia, preferencialmente cerca de 0,1 µg -1 mg/kg b. d./dia, e esta quantidade pode ser administrada uma vez ou nalgumas doses divididas diariamente.

(6) Terapia génica

A presente especificação disponibiliza ainda um processo de terapia génica utilizando o gene p51 humano. Este processo terapêutico pode ser considerado como um método para conferir as funções do gene p51 do tipo selvagem às células que albergam um gene mutante p51. Conferindo as funções normais possuídas inerentemente pelo gene p51 do tipo selvagem ou produto génico a células pode ser inibido o crescimento neoplásico das células receptoras/alvo. O gene p51 do tipo selvagem mencionado acima pode ser transferido para as células objectivo, através de um vector ou plasmídeo capaz de manter o gene de forma extra-cromossomal. Neste caso, o gene em particular é expresso fora do cromossoma.

A introdução do gene p51 do tipo selvagem para as células que albergam um tal gene p51 mutante para se ter a uma proteína p51 normal expressa nas células receptoras, o gene p51 não precisa de ter uma sequência de comprimento completo, mas pode ser o dito gene modificado desde que o último possua as funções substancialmente homólogas em relação ao gene não modificado. Como alternativa, pode ser

utilizado o gene que possui uma sequência parcial que retém algumas das tais funções. Como exemplo do gene mencionado mesmo acima, pode ser mencionado um gene que codifica para uma porção da proteína p51 que é necessária para um crescimento não tumoral de células (inibição do crescimento celular).

O gene p51 do tipo selvagem ou um seu fragmento é preferencialmente introduzido em células mutantes de tal maneira que uma recombinação vai ocorrer com o gene mutante p51 endógeno. Para uma tal recombinação, a ocorrência de uma recombinação dupla que corrige a dita mutação do gene p51 é dita ser necessária.

Os vectores que podem ser utilizados na transferência do gene objecto tanto para a recombinação como para a manutenção do gene fora do cromossoma já são conhecidas no estado da técnica e qualquer dos vectores conhecidos podem ser utilizados na sua prática. Por exemplo, pode ser mencionado um vector viral ou vector plasmídico que alberga uma cópia do gene p51 ligado a um elemento de controlo de expressão e é capaz de assegurar a expressão do produto génico nas células alvo. Como tal vector, os vectores de expressão mencionados acima podem ser geralmente utilizados, mas preferencialmente os vectores construídos utilizando tais vectores de origem como os vectores revelados na Patente de invenção US5252479 e PCTW093/07282 (pWP-7A, pwP-19, pWU-1, pWP-8A; pWP-21 e/ou pRSVL, entre outros) ou pRC/CMV (Invitrogen). Os mais preferidos são os vários vectores virais descritos a seguir.

Como promotor do vector a ser utilizado em terapia génica, os promotores intrínsecos aos tecidos alvo afectados em várias doenças podem ser utilizados com vantagem.

Podem ser citados os exemplos específicos abaixo. Podem ser mencionados a título de exemplo para o fígado,

por exemplo, albumina, α -fetoproteína, α 1-anti-tripsina, transferrina, transtiretin, etc. Podem ser mencionados para o cólon, carboxil anidrase I, antigénio carcinoembriogénico, etc. Para o útero e para a placenta, estrogénio, citocromo P450 aromatase, P450 que quebra a cadeia lateral do colesterol, P450 17 α -hidroxilase, etc.

Podem ser mencionados para a próstata, antigénio prostático, gene gp91-fox, kallikrein específico da próstata, etc. Podem ser mencionados para a mama erb-B2, erb-B3, β -caseína, β -lactoglobina, soro do leite, etc. Podem ser mencionados para os pulmões a proteína activadora C uroglobulina, entre outros. Podem ser mencionados para a pele K-14-queratina, queratina humana 1 ou 6, leucrina, etc.

Podem ser mencionados para o cérebro proteína ácida das fibras neurogliais, proteína específica dos astrocitos maduros, mielina, vilina pancreática da tirosina hidroxilase, glucagona, polipéptido amilóide do ilhéu de Langerhans, etc. Podem ser mencionados para a tiróide tiroglobina, calcitonina, etc. Pode ser mencionado para o osso, colagénio α 1, osteocalcina, sialoglicoproteína do osso, etc. Pode ser mencionado para o rim, renina, fosfatase alcalina fígado/osso/rim, eritropoietina, etc. Pode ser mencionado para o pâncreas a amilase, PAP1, etc.

O gene (completo ou um fragmento) a ser utilizado na construção de um vector de transferência de gene pode ser facilmente produzido e adquirido através de tecnologia de engenharia genética padrão baseada na informação da sequência nucleotídica sobre o gene p51 mencionado anteriormente.

A introdução de um tal vector de transferência de gene em células pode ser efectuada através de várias técnicas alternativas conhecidas pelos peritos no estado da técnica, tais como electroporação, coprecipitação de fosfato de

cálcio, transdução viral etc. As células transformadas com o gene p51 do tipo selvagem podem ser utilizadas como estão na forma isolada como agente para suprimir ou inibir a metástase do cancro, ou como um sistema modelo para investigação terapêutica.

Na terapia génica, o dito vector de transferência génica pode ser introduzido nas células tumorais de um doente por administração tópica no local do tumor ou por administração sistémica ao doente por injeção. Por administração sistémica, o gene pode ser feito chegar a quaisquer células tumorais metastáveis ou a outros locais. Se as células transformadas não puderem ser recebidas nos cromossomas das células tumorais alvo, a administração acima pode ser repetida periodicamente.

O processo da terapia génica inclui tanto o processo *in vivo* que compreende a administração de um material para a dita transferência génica (vector de transferência génica) directamente para o corpo e o processo *ex vivo* que compreende a separação das células tumorais do corpo do doente introduzir o gene extra-corporalmente, e devolvendo a células ao corpo.

Uma outra alternativa possível é uma terapia génica utilizando um ribozima que compreende a introdução do gene p51 humano directamente nas células e cindindo a cadeia de ARN com o ribozima que é uma molécula activa.

O vector de transferência génica que alberga o gene p51 humano ou um seu fragmento e um agente terapêutico génico compreendendo as células transformadas com o gene p51 humano, como princípio activo, através do dito vector são dirigidos especialmente para a terapia dos cancros, mas a terapia génica (tratamento) mencionada acima pode ser também aplicada à terapia de doenças hereditárias e doenças virais tais como SIDA, assim como para o objectivo de marcação de genes.

As células alvo para as quais o gene é transferido, podem ser judiciosamente seleccionadas de acordo com o objecto da terapia génica (tratamento). Por exemplo, podem ser mencionadas como células alvo, não apenas células de cancro e tecidos tumorais, mas também linfócitos, fibroblastos, hepatócitos, células embrionárias hematopoiéticas e outras células.

O processo de introduzir o gene nas células na terapia génica inclui um processo de transferência viral e um processo de não transferência viral.

Como processo de transferência viral, pode ser utilizado o processo que utiliza um retrovírus, por exemplo, tendo em vista o facto de que o gene p51 humano é um gene estranho que é expresso em células normais. Como outros vectores virais podem ser mencionados o vector de adenovírus HIV (vírus da imunodeficiência humana), vector do vírus adeno-associado (AAV), vector do vírus de herpes, vírus do herpes simplex (HSV) e vírus de Epstein-Barr (EBV), etc.

O processo não viral de transferência do gene inclui o processo de coprecipitação do fosfato de cálcio; o processo de membrana-lipossoma de fusão que compreende a fusão de lipossomas contendo ADN com um vírus Sendai inactivado, quando exposto a radiação UV para a destruição do gene para construir membranas-lipossomas de fusão e introduzir o ADN em células por fusão directa com a membrana celular [Kato. K., et al., J. Biol. Chem., 266, 22071-22074(1991)]; o processo que compreende o revestimento do plasmídeo de ADN com ouro e introduzindo o ADN fisicamente em células por descarga de alta voltagem [Yang, N. S. et al., Proc. Natl. Acad. Sci., 87, 9568-9572(1990)]; o processo do ADN nu em que o ADN do plasmídeo é directamente injectado num órgão ou tumor in vivo [Wolff, J. A., et al., Science, 247, 1465-

1467(1990)]; o processo catiónico do lipossoma em que o gene aprisionado em lipossomas positivamente carregados multilamelares é introduzido nas células [Yagi, K. *Advance in Medecine*, vol. 175, Nº 9, 635-637(1995)]; e o processo completo do ligando-ADN em que um ligando que se liga a um receptor expresso nas células alvo é acoplado ao ADN de modo que o gene possa ser introduzido exclusivamente nas células seleccionadas e não noutras células e o complexo resultante é administrado [Frindeis, et al., *Trends Biotechnology.*, 11, 201(1993); Miller et al., *FASEB J.*, 9, 190(1995)] entre outros.

O processo do complexo ligando-ADN mencionado acima inclui o processo que compreende a utilização do receptor asialoglicoproteína expresso nas células de fígado como alvo e uma asialoglicoproteína como ligando [Wu, et al., *J. Bio. Chem.*, 266, 14338 (1991); Ferkol, et al., *FASEB J.*, 7, 1081-1091(1993)] e o processo que compreende a utilização do receptor da transferrina expressa em níveis elevados em células tumorais como alvo e a transferrina como ligando [Wagner et al., *Proc. Natl. Acad. Sci., USA.*, 87, 3410(1990)], entre outros.

Além disso o processo de transferência génica que pode ser utilizado pode ser uma combinação útil dos ditos processos de transferência de genes biológicos e físicos. Como um destes processos de combinação pode ser mencionado o processo em que um plasmídeo de ADN de um determinado tamanho é utilizado em combinação com um anti-corpo de polilisina conjugado específico da proteína do adenovírus hexon. De acordo com este processo, o complexo formado encontra-se ligado ao vector de adenovírus e o complexo trimolecular resultante pode ser utilizado para infectar as células e transferindo deste modo o gene. De acordo com este processo, ligação eficaz, endogeneização e degradação

do endossoma pode ocorrer antes do ADN acoplado ao vector de adenovírus ser danificado. Além disso, o dito complexo de lipossoma/ADN pode mediar a transferência génica *in vivo*.

O processo para construir um vector viral para transferência do gene e o processo de introduzir o gene nas células alvo ou tecido alvo é agora descrito.

O sistema de vectores de retrovírus compreende o vector viral e células auxiliares (células de empacotamento). As células auxiliares mencionadas acima são células em que os genes tais como a proteína estrutural do retrovírus gag (a proteína estrutural na partícula viral), pol (transcriptase reversa) e env (proteína de envólucro) foram expressas, mas não formaram partículas virais. Por outro lado, o vector viral possui um sinal de empacotamento e LTR (terminais de repetição longos), mas estão isentos de genes estruturais necessários para a replicação viral, tal como gag, pol e env. O sinal de empacotamento é uma sequência que serve como marcador na montagem de uma partícula viral e os genes selectivos (neo, hyg) e o gene desejado (gene p51 ou um seu fragmento) incorporado no local de clonagem são inseridos em vez do gene viral. Aqui para se ter um elevado título de partículas virais é importante efectuar a inserção tão curta como possível, alargar o sinal de empacotamento incluindo uma parte do gene gag e ter cuidado para não deixar o ATG do gene gag.

Introduzindo o vector de ADN que alberga o gene objecto p51 nas células auxiliares, o ARN do genoma do vector é empacotado com a proteína estrutural do vírus das células auxiliares para formar e secretar partículas de vírus. Após a partícula do vírus como vírus recombinante ter infectado a célula alvo, o ADN transcrito de forma reversa a partir do ARN genómico do vírus é integrado no núcleo da células e o gene inserido no vector é expresso.

Como processo de melhorar a eficiência de transferência do gene objecto, pode ser utilizado o processo que utiliza um fragmento que contém o domínio de adesão da célula da fibronectina, o local de ligação da heparina e o segmento de conjugação [Hanenberg, H., et al., Exp. Hemat., 23, 747(1995)].

Como um exemplo de um vector para utilização no sistema do vector do retrovírus acima, pode ser mencionado o retrovírus derivado do vírus do ratinho da leucemia [McLachlin, J. R., et al., Proc. Natl. Acad. Res. Molec. Biol., 38, 91-135(1990)].

O processo compreendendo a utilização do vector de adenovírus é agora descrito em detalhe.

A construção do dito vector de adenovírus pode ser efectuada de acordo com os processos de Berkner [Berkner, K. L., Curr. Topics Microbiol. Immunol., 158, 39-66 (1992)], Setoguchi, Y. et al. [Setoguchi, Y., et al., Blood, 84, 2946-2953(1994)], Kanegae, H. et al. [Experimental Medicine, 12, 28-34(1994)], e Ketner et al. [Ketner, G., et al., Proc. Natl. Acad. Sci., USA, 91, 6186-6190(1994)].

Por exemplo, para a construção de um vector de adenovírus não proliferativo, o gene precoce E1 e/ou regiões do gene E3 do adenovírus são removidos primeiro. Depois, um vector plasmídico que alberga a unidade de expressão do gene objecto estranho (constituído pelo gene a ser transferido, que é o gene p51 um promotor para a transcrição do gene, e poli-A que confere estabilidade ao transcrito) e uma porção do ADN genómico do adenovírus e um plasmídeo que alberga o genoma do adenovírus são utilizados para transferir células concorrentemente, por ex. 293 células. Provocando a ocorrência de uma recombinação homóloga entre os dois e substituindo deste modo a unidade

de expressão do gene para E1, o vector de adenovírus não proliferativo que alberga o gene p51 pode ser construído. É também possível integrar o ADN genómico do adenovírus no vector cosmídeo para construir um vector de adenovírus do terminal 3' com a proteína terminal adicionada. Além disso, o vector YAC podem também ser utilizado na construção de um vector de adenovírus recombinante.

A produção de um vector de vírus adenoassociado (AAV) é agora brevemente descrita. O AAV foi descoberto como um vírus de pequeno tamanho contaminante de um sistema de cultura de adenovírus. Deste vírus, o género parvovirus que não necessita de um vírus auxiliar para replicação, mas que prolifera autonomamente na célula hospedeira e o depenvirus que necessita de um vírus auxiliar foram confirmados. Este AAV é um dos vírus comuns que possui uma gama larga de hospedeiro e infecta uma variedade de células. O seu genoma é um ADN de cadeia simples linear constituído por 4680 nucleótidos e os 145 nucleótidos em cada terminal possui uma sequência característica designada ITR (repetição terminal invertida). Esta região ITR é um ponto de iniciação de replicação e desempenha o papel de um iniciador. Além disso, o seu ITR é indispensável ao empacotamento da partícula do vírus e à integração no ADN cromossomal da célula hospedeira. Além disso, no que diz respeito à proteína do vírus, a metade esquerda do genoma codifica para uma proteína não estrutural que é a proteína reguladora Rep que controla a replicação e a transcrição.

A construção de um AAV recombinante pode ser efectuada utilizando a propriedade do AAV para ser integrado no ADN cromossomal, em que um vector para a transferência do gene desejado pode ser construído. Mais particularmente, de acordo com este método, um plasmídeo (vector plasmídico AAV) que alberga o gene objecto a ser transferido (gene p51 humano) inserido entre os ITRs remanescentes em ambos os

terminais 5'- e 3' do tipo selvagem AAV é primeiro construído. Por outro lado, a proteína do vírus necessária para a replicação do vírus e para a construção da partícula do vírus é fornecida a partir de um plasmídeo auxiliar independente. É assegurado que não existe nenhuma sequência nucleotídica comum a ambos os plasmídeos de modo que um vírus recombinante do tipo selvagem não vai emergir. Então, ambos os plasmídeos são introduzidos por transfecção em por exemplo, 293-células, que são ainda infectadas com adenovírus (que podem ser do tipo não proliferativo, quando as células 293 são utilizadas) como vírus auxiliar, em que o AAV recombinante objectivo do tipo não proliferativo é produzido. Uma vez que este AAV recombinante existe no núcleo, é recuperado por congelamento-descongelamento e o adenovírus contaminante é inactivado por aquecimento a 56°C. Quando necessário, o AAV recombinante é isolado e concentrado por ultracentrifugação com cloreto de céσιο. Desta forma, o AAV recombinante objectivo para transferência do gene objecto pode ser adquirido.

A construção do vector HIV pode ser efectuada tipicamente de acordo com o processo de Shimada et al. [Shimada, T., et al., J. Clin. Invest. 88, 1043-1047 (1991)].

Uma vez que o adenovírus infecta especificamente células T auxiliares com CD4 como receptor, um vector HIV de transferência de um gene específico do tecido adaptado para uma introdução específica de um gene em células CD4 positivas humanas pode ser construído. Este vector HIV é óptimo para a terapia génica da SIDA.

A construção de um vector HIV recombinante pode ser efectuada tipicamente como se segue. Primeiro o plasmídeo de empacotamento CGPE é construído de tal maneira que os genes estruturais gag, pol e env e os genes de controlo (tat, ver, etc.) necessários para a sua expressão podem ser

expressos com o promotor do citomegalovírus (CMV) e o sinal poli A do gene da globina humana (poli A). Depois o vector plasmídico HXN pode ser construído de forma a permitir uma proliferação eficiente em células COS inserindo o gene bacteriano resistente à neomicina (neoR) que possui um promotor para a timidina quinase (TK) como um gene marcador entre os dois LTRs de HIV e inserindo ainda um mecanismo de replicação de SV40 no vector de plasmídeo basal. Como o plasmídeo de empacotamento acima CGPE e o vector plasmídico HXN são introduzidos concorrentemente por transfecção em células COS, o gene objectivo neoR - integrado no vírus recombinante é produzido e libertado no meio de cultura em grande quantidade.

A produção do dito vector EBV pode ser efectuada tipicamente de acordo com o processos de Shimidzu et al [Shimidzu, N., Saibou Kougaku (Cell Engineering), 14(3), 280-287(1995)].

A produção de um vector EBV para transferência do gene é agora brevemente descrita. O vírus EB (vírus de Epstein-Bar: EBV) é um vírus que pertence à família do herpes e foi isolado a partir de células cultivadas derivadas do linfoma de Burkitt por Epstein e colaboradores em 1964 [Kieff, E. e Liebowitz, D.: Virology, 2ª ed. Raven Press, Nova Iorque, 1990, p. 1989-1920]. O EBV possui uma actividade transformadora de células, por isso, de modo a que possa ser utilizada como um vector de transferência de gene, o vírus defectivo desta actividade transformadora tem que ser preparado. Isto pode ser feito como se segue.

Assim, em primeiro lugar, um genoma de EBV próximo do ADN alvo em que o gene estranho desejado é para ser integrado é clonado. Com este clone, um fragmento de ADN de um gene estranho e um gene resistente a um fármaco são integrados para preparar um vector para a produção de um vírus recombinante. Depois, o vector para a construção de

um vírus recombinante é excisado com enzimas de restrição adequadas e introduzido por transfecção em células Akata EBV positivas. O vírus recombinante produzido pela recombinação homóloga pode ser recuperado, conjuntamente com o EBV Akata do tipo selvagem, através da simulação da produção de um vírus, através de tratamento anti-superfície de imunoglobulina. Isto é utilizado para infectar células Akata EBV negativas e uma estirpe resistente é seleccionada na presença do fármaco para obter as células Akata desejadas infectadas exclusivamente pelo vírus recombinante e libertadas do EBV do tipo selvagem. Seguidamente, através da indução da actividade viral nas células Akata recombinantes infectadas pelo vírus pode ser produzido o vector viral recombinante objectivo em grande quantidade.

A produção de um vector não viral para a introdução do gene desejado nas células alvo sem a utilização de um vector de vírus recombinante pode ser efectuada, através da técnica de transferência génica utilizando uma preparação de membrana com lipossomas de fusão. Esta é uma técnica em que conferindo actividade de fusão a um lipossoma de membrana (uma vesícula possuindo uma estrutura de bicamada lipídica), o conteúdo do lipossoma é directamente introduzido na célula.

A introdução do gene, através de tais lipossomas de fusão da membrana pode ser efectuada tipicamente de acordo com o processo de Nakanishi et al. [Nakanishi, M., et al., *Exp. Cell Res.*, 159, 399-499(1995); Nakanishi, M., et al., *Gene introduction into animal tissues. In Trends and Future Perspectives in Peptide and Protein Drug Delivery* (ed. By Lee, V. H. et al.)] Harwood Academic Publishers GmbH. Amsterdam, 1995, p. 337-349].

Este processo de transferência de gene, através dos ditos lipossomas de fusão de membrana é brevemente descrito abaixo. Assim, lipossomas nos quais o vírus de Sendai com o

seu gene inactivado por irradiação UV, o gene objecto e uma substância de elevado peso molecular, tal como uma proteína foram aprisionados são fundidos a 37°C. Este lipossoma de fusão de membrana possui uma estrutura designada “pseudovírus” que consiste numa cavidade interna derivada de um lipossoma e uma estrutura de espigão exterior semelhante à do envelope do vírus. Após purificação por centrifugação por gradiente de densidade de sacarose, os lipossomas de membrana de fusão são provocados a serem absorvidos em células de cultura alvo a 4°C. Seguidamente, quando a temperatura é aumentada para 37°C, o conteúdo dos lipossomas é introduzido na célula em que o gene desejado pode ser transferido para as células alvo. Neste caso, o lípido para o dito lipossoma é um fosfolípido sintético composto por 50% (em moles) de colesterol e lecitina e possuindo uma carga negativa e é preferencialmente formado como um lipossoma unilamelar com um diâmetro de 300 nm.

Como um processo alternativo de introdução do gene na célula alvo, através de lipossomas, o processo de transferência de gene utilizando lipossomas catiónicos pode ser mencionado. Este processo pode ser praticado de acordo com o processo de Yagi et al. [Yagi, K., et al., B. B. R. C., 196, 1042-1048(1993)]. Assim, com atenção dada ao facto de que plasmídeos e células encontram-se ambos negativamente carregados, uma carga positiva é conferida a ambas as superfícies externa e internas da membrana do lipossoma de modo a que a absorção do plasmídeo seja aumentada por electricidade estática para aumentar a interacção com as células. O lipossoma utilizado neste caso é preferencialmente uma grande vesícula multilamelar (MLV) que possui uma carga positiva, apesar de ser possível utilizar uma grande vesícula unilamelar (LUV) ou uma vesícula unilamelar pequena (SUV) para construir um

complexo com o plasmídeo para a introdução do gene desejado.

O processo de preparar um plasmídeo contendo MLV catiónico é brevemente descrito abaixo.

Em primeiro lugar, é preparada uma solução de clorofórmio contendo o lípido TMAG cloreto de (N-(α -trimetil-amónioacetil)-didodecil-D-glutamato), DLPC (dilauroil fosfatidilcolina) e DOPE (dioleil fosfatidiletanolamina) numa proporção molar de 1:2:2 (concentração lipídica: 1 mM). Seguidamente um total de 1 μ mol é colocado num tubo de centrifuga e o clorofórmio é destilado sob pressão utilizando um rotovapor para preparar o filme fino de lípido. O clorofórmio residual é completamente removido sob pressão reduzida e o filme é seco. Seguidamente, adiciona-se 0,5 ml de Mg e Ca contendo solução de tampão fosfato salino de Dulbecco contendo 20 pg do plasmídeo de transferência do gene e após purga de azoto a mistura é agitada com um misturador de vórtice durante 2 minutos para dar uma suspensão do gene que alberga o plasmídeo contendo MLV catiónico.

O seguinte é um exemplo da utilização de plasmídeo contendo MLV catiónico como agente de terapia génica. Por exemplo, o plasmídeo de expressão integrado com o ADNc do gene objecto é aprisionado numa quantidade de 0,6 pg como ADN por 30 nmole de lipossoma lipídico no MLV catiónico referido acima e o lipossoma é suspenso em 2 μ L tampão fosfato salino. Esta suspensão é administrada às células alvo extraídas do doente ou do tecido alvo do doente em dias alternados.

Neste contexto, as orientações estabelecidas pelo Ministério da Saúde e Bem Estar do Japão, a terapia génica é definida como “administrar um gene ou uma célula integrada com um gene no corpo humano para a terapia de uma doença”.

Contudo, a terapia génica no presente contexto abrange não apenas a terapia que falha sob a definição acima, mas também a terapia de várias doenças inclusivé de cancro que compreende a introdução de um gene caracterizado como um gene supressor de tumor, tal como o gene p51 humano nas ditas células alvo e a prática que compreende a introdução de um gene marcador ou células integradas com um tal gene marcador no corpo humano.

Na terapia génica, o processo de introduzir o gene objecto nas células alvo ou tecidos inclui os dois processos representativos seguintes.

O primeiro processo compreende a utilização das células alvo do doente a ser tratado fazer crescer as células fora do corpo, por exemplo na presença de interleucina-2 (IL-2), ou semelhante, introduzindo o gene p51 ligado ao vector do retrovírus nas células, e tornando a transplantar as células resultantes (processo ex vivo). Este processo é adequado para a terapia do síndrome da deficiência ADA, doenças hereditárias e cancros associados com genes defectivos, SIDA, e outras doenças.

Um segundo processo é um processo de transferência directa do gene que compreende a injeção do gene objecto (gene p51 humano) directamente no corpo do doente ou local alvo, tal como num tecido tumoral (processo directo).

Mais particularmente, o primeiro processo acima da terapia génica pode ser efectuado tipicamente como se segue. Assim, as células mononucleares isoladas do doente são separadas dos monócitos com um separador de sangue, as células colhidas são cultivadas na presença de IL-2 num meio adequado tal como meio AIM-V durante cerca de 72 horas, e o vector que alberga o gene a ser introduzido (gene p51 humano) é adicionado. Para uma eficiência de transferência génica acrescida, o sistema pode ser centrifugado a 2500 rpm na presença de protamina a 32°C

durante 1 hora e incubado sob 10% de gás dióxido de carbono a 37°C durante 24 horas. Após o procedimento acima ser repetido algumas vezes, as células são ainda cultivadas na presença de IL-2 em AIM-V ou outro meio durante 48 horas. As células são lavadas com soro fisiológico, as células viáveis são contadas, e a eficiência de transferência do gene é avaliada efectuando a dita PCR in situ ou, quando a função objecto é actividade enzimática, testando a actividade para conferir o efeito de transferência do gene.

Além disso, uma verificação de segurança, compreendendo a cultura de bactérias e fungos que contaminam as células cultivadas e teste da infecção de micoplasma e para a endotoxina é efectuada para confirmar a segurança. Seguidamente, as células cultivadas integradas com uma dose do gene (p 51 humano) prevista como sendo eficaz são devolvidas ao doente por injeção intravenosa em gotas. Este processo é repetido num intervalo de algumas semanas ou de alguns meses para terapia génica.

A dosagem do vector viral pode ser judiciosamente seleccionada de acordo com o tipo da célula alvo. Geralmente, em termos do título viral é utilizada uma dose na gama de 1×10^3 cfu - 1×10^8 cfu por 1×10^8 células alvo.

Uma versão alternativa do primeiro processo acima que pode ser utilizada compreende a co-cultura de células que produzem vírus contendo um vector de retrovirus que alberga o gene desejado (gene p51 humano) com, por exemplo as células do doente para introduzir o gene (gene humano p51) nas células alvo.

Para efectuar o segundo processo de terapia génica (processo directo) uma experiência preliminar, particularmente uma experiência ex vivo é efectuada preferencialmente para confirmar se o gene objecto (gene p51 humano) pode ser ou não de facto transferido por um PCR do gene do vector do ADNc ou de um ensaio PCR in situ, ou

para confirmar o efeito desejado de terapia resultante da transferência do gene objecto (gene p51 humano), por exemplo uma elevação na actividade específica ou um acréscimo ou supressão do crescimento de células alvo. Além disso, quando é utilizado um vector viral é evidentemente importante para conduzir uma terapia génica, para confirmar a segurança de introdução do gene efectuando uma PCR para pesquisar o retrovírus proliferativo, determinando a actividade da transcriptase reversa, ou para monitorizar o gene da proteína de membrana (env) por uma técnica PCR.

Quando o processo de terapia génica é aplicado a cancros ou tumores malignos em particular, um protocolo típico para a terapia do cancro compreende o isolamento de células de cancro do doente, tratar as células com uma enzima ou semelhante para estabelecer uma linha de células cultivadas, introduzindo o gene desejado nas células de cancro alvo, através de um retrovírus ou semelhante, efectuando um rastreio com células G418, determinando a quantidade de expressão de IL-12 ou semelhante (*in vivo*), sujeitando as células a tratamento por radiação e inoculando as células no tumor do doente ou para-tumor.

O gene do vírus do herpes simplex-timidina cinase (HSV-TK) é reportado provocar morte celular devido à divisão por envelhecimento, em particular convertendo o nucleótido análogo ganciclovir (GCV) num intermediário tóxico, e é conhecida uma terapia génica utilizando este gene em tumores [Patente de invenção US 5631236; JP Kohyo H9-504784]. Este processo é um processo de terapia génica que utiliza o fenómeno que quando células capazes de produzir um vector de retrovírus que alberga o dito gene HSV-TK, conhecido como gene suicida são injectados e, uma semana mais tarde, o agente anti-viral GCV é administrado, o GCV nas células transformadas com gene é activado por fosforilação para induzir a morte destas células e morte

das células circundantes não transformadas com gene devido ao contacto celular, através da junção gap. O vector de transferência do gene ou células contendo este vector podem também ser utilizados na terapia génica acima.

Um processo alternativo de terapia génica compreende a preparação de imunolipossomas contendo o gene (gene p51 humano) acoplado ao anti-corpo capaz de se acoplar com a superfície da célula alvo para introduzir o ADNc aprisionado nas células alvo selectivamente e com boa eficiência. É também possível de realizar uma terapia génica que compreende a administração do dito gene da citocina que alberga o vector viral e o dito gene suicida que alberga o adenovírus de uma vez e ao mesmo tempo. Estes processos caem invariavelmente no âmbito dos peritos no estado da técnica.

(7) Composição farmacêutica para a terapia génica

A presente memória descritiva disponibiliza ainda uma composição farmacêutica ou agente (agente génico terapêutico) compreendendo uma quantidade farmacologicamente eficaz da quantidade de células para o qual o vector de transferência génica ou o gene objecto (por ex. o gene humano p51 ou semelhante) foi transferido como um princípio activo conjuntamente com um veículo ou diluente farmacêutico adequado não tóxico.

O veículo farmacêutico que pode ser formulado na composição farmacêutica (preparação) da presente invenção inclui os diluentes convencionais e excipientes, tais como enchimentos, formador de volume, ligante, humidificante, desintegrante, agente tensioactivo, lubrificante, etc., que são habitualmente utilizados de acordo com o processo de utilização da preparação e estes veículos podem ser

utilizados selectivamente em relação à forma farmacêutica unitária desejada.

A forma unitária da composição farmacêutica inclui as mesmas formas que aquelas mencionadas para a proteína p51 e podem ser judiciosamente seleccionadas de acordo com o objectivo terapêutico.

Por exemplo, uma preparação farmacêutica contendo o vector de transferência do gene pode ser disponibilizada na forma do dito vector aprisionado em lipossomas ou na forma de células cultivadas infectadas com um vírus contendo um vector de retrovírus que alberga o gene objectivo.

Estas podem ser preparadas como soluções na forma de tampão fosfato salino (pH 7,4), a injeção de Ringer ou uma composição para uma injeção intracelular de fluido, por exemplo, ou de tal forma que possa ser administrada conjuntamente com a substância capaz de aumentar a eficiência da transferência génica, tais como a protamina.

O processo de administração da preparação farmacêutica acima não se encontra particularmente restringido, mas é seleccionado de acordo com a forma de dosagem, da idade do doente, sexo e outros factores, a gravidade da doença, e outras condições.

A quantidade de ingrediente activo a ser incorporado na dita composição farmacêutica ou preparação não se encontra particularmente restringida e pode ser liberalmente seleccionada a partir de uma gama larga de acordo com o efeito terapêutico desejado, método de administração, duração do tratamento, antecedentes do doente inclusivé idade e sexo e outras condições.

Falando geralmente, a dose diária do gene albergado pelo vector de retrovírus na forma de preparação farmacêutica por kilograma de peso corporal pode por exemplo ser cerca de 1×10^3 pfu a 1×10^{15} pfu em termos do título de retrovírus.

No caso de células em que o gene objecto tinha sido introduzido a dosagem pode ser judiciosamente seleccionada a partir da gama de 1×10^4 células/corpo a cerca de 1×10^{15} células/corpo.

O produto farmacêutico pode ser administrado uma vez por dia ou nalgumas doses divididas por dia, e pode ser administrado intermitentemente, por exemplo com um espaço de uma a várias semanas. Além disso, pode ser vantajosamente administrado em combinação com uma substância capaz de aumentar a eficiência da transferência génica, tal como a protamina, ou uma preparação farmacêutica contendo a dita substância.

Quando a terapia génica é aplicada ao tratamento de um cancro, os vários processos de terapia génica mencionados acima podem ser utilizados numa combinação adequada (combinação de terapia génica) e/ou em combinação com a terapia do cancro convencional, terapia de radiação, imunoterapia e/ou outra terapia. Além disso, a terapia génica pode ser efectuada com base nas orientações NIH, inclusive do seu aspecto de segurança [ver Recombinant DNA Advisory Committee, Human Gene Therapy, 4, 365-389 (1993)].

(8) Aplicação ao diagnóstico do tumor

De acordo com a presente memória descritiva, a presença de um gene p51 mutante que promove a tumorigénese em células humanas pode ser detectada, através de um procedimento que compreende a preparação do sangue, soro, ou outra amostra biológica, extraíndo opcionalmente o ácido nucleico, e analisando-o, quanto à presença ou ausência de um gene mutante p51 sensível. Além disso, a existência de uma alteração neoplásica em células ou tecidos, um marcador da progressão até a um estado pródromo de malignidade, ou um marcador de prognóstico pode ser detectado, através do

procedimento que compreende a preparação de uma amostra biológica contendo desordem e analisando-o, quanto à presença ou ausência de um gene mutante p51 neoplásico. Através do procedimento acima, a presença de um neoplasma em células ou tecidos, um marcador da progressão até a um estado pródromo de malignidade, ou um marcador de prognóstico pode ser detectado permitindo estabelecer um diagnóstico, por exemplo o diagnóstico de um cancro, avaliar o efeito da terapia do cancro, ou prever o prognóstico dos casos.

De acordo com este processo de detecção, baseado na informação do gene mutante p51 e na informação da sequência mutante, o fragmento relevante de ADN mutante é preparado e desenhado de modo a que possa ser utilizado no rastreio do gene mutante e/ou da sua ampliação. Mais particularmente, a sonda a ser utilizada na hibridização em placa, hibridização em colónia, transferência de Southern. Transferência de Northern, etc. ou a sonda para ampliação por PCR do fragmento de ADN mutante pode ser construída. Para tais objectivos, um iniciador que possui a mesma sequência que a mutação é primeiro preparado e feito reagir, como uma sonda de rastreio com uma amostra biológica (amostra de ácido nucleico), em que a presença de um gene possuindo uma sequência do gene p51 mutada pode ser confirmada. Para facilitar a detecção da sequência alvo, a dita amostra de ácido nucleico pode ser preparada utilizando várias técnicas, tais como a lise, digestão por enzima de restrição, electroforese ou transferência por pontos.

Fazendo referência a um processo de rastreio mencionado acima, a utilização de um processo PCR é particularmente preferido do ponto de vista de sensibilidade, e este processo não se encontra particularmente restringido desde que os fragmentos de p51 mutantes sejam utilizados como iniciadores. Assim,

quaisquer técnicas até agora conhecidas [Science, 230, 1350-1354 (1995)] e versões de PCR que foram recentemente desenvolvidas ou que vão ser utilizadas no futuro [Sakaki, Y. et al. (ed): Experimental Medicine, Supplemental Issue, 8(9)(1990), Yōdo-sha; Protein, Nucleic Acid and Enzyme, Special Supplemental Issue, 35(17)(1990), Kyoritsu Shuppan] podem ser utilizadas.

Os fragmentos de ADN que podem ser utilizados como iniciadores são oligoADNs sintetizados quimicamente, e estes oligoADNs podem ser sintetizados utilizando um hardware de síntese automático, tal como o sintetizador de ADN da Pharmacia LKB Gene Assembler Plus (Pharmacia). O comprimento do iniciador sintetizado deste modo (iniciador de sentido ou iniciador anti-sentido) é preferencialmente equivalente a cerca de 10-30 nucleótidos. A sonda para ser utilizada no dito rastreio é usualmente marcada, mas pode ser uma não marcada, e a detecção pode ser efectuada directamente, ou indirectamente por ligação específica a um ligando marcado. A marcação adequada e o processo de marcação da sonda ou do ligando são conhecidos dos peritos no estado da técnica, e a marcação radioactiva, biotina, grupo fluorescente, grupo quimioluminescente, enzima, anti-corpo, etc. que pode ser incorporado, através das técnicas conhecidas tal como tradução nick, iniciação aleatória ou tratamento com cinase estão também incluídas na tecnologia relevante.

O processo PCR para ser utilizado na dita detecção inclui RT-PCR, por exemplo, e várias modificações de PCR que são utilizadas no estado da técnica e podem ser aplicadas igualmente.

É também possível detectar o gene p51 do tipo selvagem e/ou gene p51 mutante e quantidades dos ADNs destes genes. Esta tecnologia inclui, mas não se encontra limitada ao ensaio competitivo tal como MSSA [Kinoshita, M. et al.,

CCA, 228, 83-90(1994)] e PCR-SSCP que é conhecido por ser uma técnica de detecção de mutação utilizando a alteração na mobilidade associada a uma alteração na estrutura de ordem superior de um ADN de cadeia simples [Orita, M. et al., Genomics, 5, 874-879(1989)].

Os processos analíticos acima mencionados, através de um exemplo podem ser efectuados como se segue. Por exemplo, um ou uma pluralidade de iniciadores contendo a mutação p51 (por ex. a sequência mutada baseada na informação do local de mutação obtida a partir de um doente com cancro ou semelhante) são preparados primeiro e hibridizados com o ADN obtido a partir de uma amostra biológica. Depois a mobilidade e área do pico medida por análise SSCP do fragmento padrão de ADN de p51 do tipo selvagem são comparadas com a mobilidade e área do pico na amostra teste como produto de amplificação utilizando os ditos iniciadores para detectar deste modo a mutação numa região específica de p51 e quantificar simultaneamente o produto da mutação.

A amostra teste contendo o gene mutante p51 a ser medido não se encontra particularmente restringida desde que contenha o dito gene mutante, incluindo assim vários materiais biológicos, tais como sangue, soro, urina e tecidos excisados. O gene mutante p51 pode ser extraído de tais amostras teste, purificado e preparado de forma rotineira. Por isso, comparando a mobilidade do dito fragmento de ADN padrão da invenção, determinada previamente com a mobilidade do produto de amplificação na amostra teste de acordo com o obtido na amplificação PCR do ADN de p51 da amostra teste utilizando um par de iniciadores de p51 mutante, a mutação numa região específica do ADN de p51 pode ser detectada de forma expediente e precisa.

Além disso, quando são utilizados os padrões em quantidade estabelecidos em passos conhecidos a quantificação do mutante p51 na amostra teste pode ser efectuada ao mesmo tempo comparando as áreas do pico dos padrões com a área do pico do produto de amplificação do ADN p51 na amostra teste no passo de amplificação de PCR utilizando o conjunto de iniciadores mutantes mencionado acima. O conjunto de iniciadores, padrões, análise PCR-SSCP e meios de detecção podem ser liberalmente modificados pelos peritos no estado da técnica.

A tecnologia do ensaio acima é agora descrita mais especificamente. Para começar, o ADN é extraído a partir do soro de um doente com cancro, através do procedimento de rotina, tal como o tratamento alcalino ou ácido. Depois um conjunto de iniciadores compreendendo uma sequência parcial de cadeia menos de um comprimento definido consistindo numa parte de uma sequência nucleotídica (145-1488) apresentada na SEQ ID N°:1 e uma sequência parcial de cadeia mais de um comprimento definido constituída numa parte pela sequência nucleotídica marcada com fluorescência (145-1488), assim como uma ADN polimerase resistente ao calor são provocadas para actuar na solução de ADN obtida acima para amplificar o fragmento de ADN marcado.

Por outro lado, um ou uma pluralidade de fragmentos de ADN contendo uma sequência mutante quimicamente sintetizada de acordo com a informação do local de mutação de p51 obtida a partir de por exemplo, um doente com cancro são integrados respectivamente em vectores plasmídicos e a E. coli é transformada. Após cultura da massa e recombinação, os plasmídeos recombinantes purificados são utilizados para preparar por ex. 10^3 cópias, 10^4 cópias, 10^5 cópias, 10^6 , cópias, 10^7 cópias e 10^8 cópias de padrões. O dito conjunto de iniciadores compreendendo uma sequência parcial de cadeia menos constituída por uma sequência parcial definida

da dita sequência nucleotídica (145-1488) e uma sequência parcial da cadeia mais consistindo numa sequência parcial definida da sequência nucleotídica marcada com fluorescência (145-1488) assim como uma ADN polimerase resistente ao calor para amplificar o fragmento de ADN marcado. A solução de ADN amplificada acima é aquecida a cerca de 95°C durante cerca de 5 minutos depois arrefecida imediatamente em gelo, e uma análise SSCP é efectuada utilizando um sequenciador automático, tal como o Sequenciador Automático ALF (Pharmacia), em que o pico fluorescente pode ser detectado. A forese nesta análise SSCP é efectuada preferencialmente a cerca de 30°C±1°C.

O pico (mobilidade) do ADN obtido a partir do soro do doente é comparado com os picos (mobilidades) dos padrões e o pico de acordo com um padrão é determinado a partir do tempo de migração. Deste modo, o tipo (espécie) de mutação do p51 do doente pode ser determinado. Além disso, através do cálculo das áreas dos picos dos padrões e da construção de uma curva de calibração, o ADN de p51 pode ser quantificado a partir da área calculada do pico do ADN do doente.

(9) Processo para detectar a mutação do gene p51 e vários métodos de ensaios

A presente especificação, disponibiliza por isso um protocolo de teste expedito para a detecção concorrente e quantificação de mutação numa determinada região do ADN do p51 na amostra teste, através da medição acima.

O processo de ensaios pode ser efectuado de forma conveniente utilizando um kit de reagente para detectar o gene p51 do tipo selvagem e gene p51 mutante numa amostra.

Por isso, a presente memória descritiva fornece um kit de reagentes para a detecção de p51 do tipo selvagem e p51

mutante caracterizada por compreender o dito fragmento de ADN do tipo selvagem e fragmento de ADN mutante de p51.

Este kit de reagente, desde que contenha um fragmento de ADN capaz de se hibridizar com uma parte da sequência nucleotídica completa (145-1488) apresentada na SEQ ID N°:2 ou a sua sequência oligonucleotídica complementar, ou um fragmento de ADN capaz de se hibridizar com uma parte, ou com a sequência mutante completa da sequência nucleotídica (145-1488), ou uma sequência nucleotídica complementar à dita sequência pode conter outros componentes tais como um agente de marcação, reagentes essenciais para a PCR (por ex. TaqADN polimerase, desoxinucleótido trifosfato, iniciadores, etc.). Em vez da sequência nucleotídica (145-1488) apresentada na SEQ ID N°:2, a sequência nucleotídica (145-2067) apresentada na SEQ ID N°:5 pode ser utilizada.

Como agente de marcação pode ser utilizado um radioisótopo, ou um modificador químico tal como uma substância fluorescente, apesar do próprio fragmento de ADN poder ter sido previamente conjugado com o agente de marcação. Além disso, este kit de reagente pode ainda compreender um solvente de reacção adequado, anticorpo padrão, tampão, solução de lavagem, reagente de paragem da reacção, etc. por razões de conveniência de realização de um ensaio.

A presente memória descritiva compreende ainda um processo de diagnóstico utilizando a dita técnica de ensaio e um agente de diagnóstico (para referência) e um kit de diagnóstico para utilização no dito processo de diagnóstico.

Além disso, através de sequenciação directa ou indirecta da sequência do p51 mutante obtido a partir da amostra teste, através do procedimento acima é possível descobrir novos genes relacionados com p51 possuindo uma elevada homologia com o tipo selvagem p51.

A presente especificação, por isso, fornece ainda um processo de rastreio de genes humanos relacionados com p51 nas amostras teste, através do ensaio acima descrito e sequenciação do ADN mutante de p51 em tais amostras teste.

Além disso, o p51 do tipo selvagem e/ou mutante p51 pode ser identificado sintetizando a proteína codificada pelo gene p51A humano da SEQ ID N°:1 ou a proteína correspondente à sequência de aminoácidos derivada da sequência da SEQ ID N°:1 por supressão, substituição ou adição de uma ou de uma pluralidade de aminoácidos ou uma sua sequência parcial, ou sintetizando o anti-corpo contra qualquer uma de tais proteínas. Além disso, em vez da proteína codificada pelo dito gene p51 humano, pode ser utilizada a proteína codificada pelo gene humano p51B apresentado na SEQ ID N°:4.

Por isso é disponibilizado um processo de ensaios de anticorpos contra o tipo selvagem p51 e/ou mutante p51 e para o ensaio do antigénio. Através deste processo de ensaio, o grau de distúrbio neoplásico ou a malignidade de um tumor maligno pode ser estimada com base na alteração da proteína p51 do tipo selvagem. A alteração mencionada acima pode ser determinada pela análise da sequência de p51, através da dita tecnologia de rotina, mas mais preferencialmente a alteração na proteína p51 ou a presença, ou ausência da proteína p51 ser detectada utilizando um anticorpo (um anticorpo policlonal ou um anticorpo monoclonal). Um exemplo específico de um processo de ensaio é o seguinte. Com o anticorpo p51, a proteína p51 pode ser imunoprecipitada a partir de uma solução contendo um material biológico humano isolado a partir de um ser humano, tal como sangue ou soro, e o anticorpo pode ser feito reagir com a proteína p51 num gel de poliacrilamida de transferência de Western ou de imunotransferência. Além disso, com o anticorpo p51 a proteína p51 numa secção de

parafina ou numa secção de tecido congelado pode ser detectada, através de uma técnica de imunohistoquímica. A tecnologia para a produção de anticorpo e purificação é bem conhecida do estado da técnica e as técnicas conhecidas podem ser utilizadas selectivamente.

Os modos específicos preferidos de praticar o processo para detectar p51 do tipo selvagem ou um seu mutante inclui o ensaio de imunosorção da enzima ligada (ELISA) inclusivé da técnica de sandwich utilizando um anticorpo monoclonal e/ou um anticorpo monoclonal, ensaio rádio-imunológico (RIA), ensaio de imunoradiometrix (IRMA) e ensaio de imunoenzimometrix (IEMA).

Além disso, é possível fornecer uma fracção de célula de membrana que possui uma actividade de ligação a p51 para a proteína p51 ou para o receptor de p51 presente na superfície celular. Para adquirir o dito receptor p51, a proteína p51 marcada é conjugada numa amostra biológica contendo a fracção da membrana celular, o conjugado p51 resultante é isolado por extracção e purificado, e a sequência de aminoácidos do isolado é determinada. A aquisição e sequenciação desta proteína receptora p51 cai no âmbito dos conhecimentos de um perito no estado da técnica.

(10) Aplicação ao rastreio de fármacos

A presente publicação pode ser aplicada ao rastreio de compostos (produtos de reacção do receptor p51: o composto pode ser um composto de baixo peso molecular, um composto de elevado peso molecular, uma proteína, um fragmento de proteína, um antigénio, um anticorpo ou semelhante) utilizando um polipéptido receptor de p51 ou um seu fragmento de ligação para o rastreio de vários fármacos. Preferencialmente a proteína receptora p51 é utilizada. O

polipéptido receptor de p51 ou um seu fragmento para ser utilizado em tal teste de rastreio pode ser imobilizado num suporte sólido ou utilizado na forma de uma suspensão num fluido transportado para a superfície da célula. Para mencionar um exemplo de pesquisa de fármacos, podem ser utilizadas células hospedeiras eucarióticas ou procarióticas transformadas de forma estável com um polipéptido recombinante e que expressam o polipéptido ou o seu fragmento, preferencialmente num ensaio de ligação competitivo. Além disso, tais células no estado livre ou imobilizado podem ser utilizadas num ensaio de ligação padrão. Mais preferencialmente, a formação de um complexo entre o polipéptido receptor p51, ou o seu fragmento e uma substância teste é quantificada para detectar o grau de inibição da dita formação de um complexo entre o polipéptido receptor p51 ou fragmento e o polipéptido p51 ou fragmento da substância teste, através do qual a pesquisa de um composto pode ser conseguida.

Assim, a presente invenção fornece um processo de rastreio de fármacos caracterizado por se fazer contactar uma tal substância com o polipéptido receptor p51 ou um seu fragmento, através de uma técnica conhecida *per se* e, depois, detectar a presença de um complexo entre a dita substância e o dito polipéptido receptor p51, ou fragmento ou a presença de um complexo entre o dito polipéptido receptor p51 ou o fragmento e um ligando. Além disso, a actividade do receptor p51 é medida para ver se a dita substância pode antagonizar o receptor p51 para exhibir as actividades p51 definidas anteriormente, por exemplo a actividade para modificar o ciclo celular ou modular a indução da apoptose. Especificamente, para efectuar um tal ensaio de ligação competitivo, o polipéptido receptor p51 ou o seu fragmento é marcado. O polipéptido receptor p51 livre ou fragmento é separado do complexo proteína-

proteína. Seguidamente a quantidade de marcador livre (que não forma complexo) pode ser uma medida da ligação entre o factor a ser testado e o receptor p51 ou inibição da ligação do receptor p51 -polipéptido p51. O péptido pequeno (pseudopéptido) da polipéptido p51 é assim analisado, e um que possui actividade inibidora p51 é determinado.

Outro processo de rastreio de fármacos é um processo de rastreio de compostos que possuem uma afinidade de ligação adequada para o polipéptido receptor p51. Resumidamente, um grande número de diferentes compostos peptídicos de teste é sintetizado sobre suportes sólidos tais como alfinetes plásticos ou outras superfícies. Seguidamente, os compostos peptídicos teste são feitos reagir com o polipéptido receptor p51, seguindo-se uma lavagem. A seguir, o polipéptido receptor p51 que reagiu e que se encontra ligado é detectado, através de uma técnica conhecida *per se* [PCT:W084-03564]. O receptor p51 purificado pode ser revestido directamente numa placa para ser utilizado na dita pesquisa de fármacos. Contudo, o polipéptido receptor p51 pode ser imobilizado numa fase sólida por suplementação de anticorpo utilizando um anticorpo de não neutralização contra o polipéptido. É ainda revelado um ensaio de pesquisa de fármacos competitivo em que um anticorpo que se liga especificamente ao polipéptido receptor p51 e um composto de teste são feitos competir um com o outro para a ligação do polipéptido receptor p51 ou fragmento. Através deste ensaio competitivo utilizando o anticorpo, a presença dos péptidos que possuem um ou vários epitopos ou locais antigénicos determinantes do polipéptido receptor p51 podem também ser detectados.

Fazendo referência ainda à pesquisa de fármacos, ainda um outro processo compreende a utilização de uma linha celular eucariótica hospedeira ou células que contêm um

gene p51 não funcional. Assim, depois da linha celular hospedeira ou células serem deixadas crescer na presença de um fármaco durante um tempo pré-determinado, a velocidade de proliferação das células hospedeiras é medida para ver se o composto pode modular a apoptose, ou o ciclo celular. Como meio para medir a velocidade de proliferação é possível medir a actividade biológica do receptor p51.

Além disso, com o objectivo de desenvolver os derivados peptídicos p51 mais activos ou estáveis ou fármacos que vão potenciar ou interferir com a função do polipéptido p51 in vivo podem ser construídos vários polipéptidos interactivos biologicamente activos ou análogos estruturais, por ex. agonistas p51, antagonistas p51, inibidores p51, etc. Tais análogos estruturais podem ser caracterizados por exemplo analisando estruturas tridimensionais de complexos entre p51 e outras proteínas por cristalografia de raios-X, modulação computadorizada, ou uma combinação de tais técnicas. Além disso, a informação sobre análogos estruturais pode também ser gerada por modelação de proteínas baseada nas estruturas de proteínas homólogas.

O processo de obter derivados de polipéptidos p51 ainda mais activos ou estáveis pode envolver por exemplo uma análise de varrimento de alanina. Neste processo, Ala é substituída por cada resíduo de aminoácido e o efeito da substituição na actividade peptídica é determinada. Cada resíduo de aminoácido num péptido é assim analisado e a região ou regiões de importância para a actividade ou estabilidade do péptido são determinadas. Utilizando este processo podem ser desenhados os derivados de p51 mais activos ou estáveis.

É também possível isolar o anticorpo específico alvo seleccionado através de um ensaio funcional e analisar a sua estrutura cristalina. Como regra, um núcleo

farmacológico que disponibiliza a base para a continuação do desenho do fármaco pode ser obtido, através desta estratégia. É possível identificar ou isolar um péptido a partir de um banco de péptidos construído quimicamente ou biologicamente provocando a formação de um anticorpo anti-ideotipo contra um anticorpo funcional farmacologicamente activo. Por isso, espera-se que o péptido seleccionado sirva como núcleo farmacológico.

Desta maneira, fármacos que possuem actividade melhorada ou estabilidade de p51 ou fármacos que actuam como inibidores, agonistas ou antagonistas da actividade p51 podem ser desenhados e desenvolvidos.

De acordo com a sequência p51 clonada, uma quantidade suficiente de polipéptido p51 pode ser procurada e submetida a cristalografia de raios-X e outra pesquisa analítica. Além disso, a proteína p51 que possui a sequência de aminoácidos da SEQ ID N°:1 fornecida pela presente invenção possibilita o estabelecimento de um programa de modelação por computador que pode tomar o lugar da cristalografia de raios-X ou completar a última técnica.

Além disso, construindo um ratinho inactivado (ratinho mutante) que transporta um gene p51 humano é possível detectar quais são os locais da sequência do gene p51 humano que vão influenciar as várias actividades p51 in vivo isto é, quais são as funções que o produto génico p51 e produtos génicos mutantes p51 vão ter in vivo.

Esta é uma tecnologia para modificar intencionalmente a informação genética dos organismos utilizando recombinações homólogas dos genes, e o processo utilizando células estaminais embrionárias de ratinho (células ES) pode ser mencionado como um exemplo [Capecchi, M. R. Science, 244, 1288-1292(1989)].

O processo de construir tais ratinhos mutantes como referido acima é bem conhecido dos peritos no estado da

técnica, e aplicando o gene p51 humano do tipo selvagem ou gene p51 mutante à tecnologia acima modificada (Noda, T. (ed.), Experimental Medicine, Supplemental Issue, 14(20)(1996), Yōdo-sha), podem ser estabelecidos facilmente ratinhos mutantes. Por isso, utilizando a tecnologia acima, os fármacos que possuem actividade ou estabilidade de p51 acrescida ou fármacos que actuam como inibidores, agonistas, ou antagonistas da actividade p51 podem ser desenhados e desenvolvidos.

A presente invenção compreende o seguinte.

1. Um anticorpo isolado que se liga especificamente a uma proteína p51 compreendendo a sequência de aminoácidos da SEQ ID N°:1, em que a dita proteína p51 é produzida por tecnologia recombinante ou síntese química.
2. O anticorpo isolado do item 2, em que a ligação é determinada utilizando imunoprecipitação, imunocoloração, transferência de Western, ELISA, ou ensaio rádio-imunológico (RIA).
3. Um kit para detectar a proteína p51 compreendendo
(i) um anticorpo isolado de acordo com a reivindicação 2, e
(ii) um meio para detectar o anticorpo.
4. O kit do item 4, em que o meio para detectar o anticorpo é uma ou várias enzimas, uma co-enzima, um substrato, um polipéptido, um anticorpo ou um marcador detectável.
5. O kit do item 5, em que o meio para detectar o anticorpo é um segundo anticorpo que se liga especificamente ao anticorpo.

6. O kit do item 4, em que o anticorpo é um anticorpo monoclonal.

7. O kit do item 4, em que o anticorpo encontra-se na forma de uma preparação purificada de anticorpo policlonal.

Melhor Processo de Efectuar a Invenção

Os exemplos seguintes e exemplos experimentais são ainda ilustrativos da presente invenção, deverá contudo ser entendido que o âmbito da invenção não se encontra de modo algum restringido por estes exemplos e exemplos experimentais.

Exemplo 1 Isolamento do gene p51 humano (apenas como referência)

(1) Clonagem e sequenciação de ADN do gene humano p51

(a) Os presentes inventores efectuaram uma amplificação por PCR utilizando o iniciador de sentido seguinte p73-F1 e iniciador p73-R1 anti-sentido e depois uma segunda amplificação através de uma PCR em ninho utilizando o iniciador de sentido p73-F2 e o iniciador anti-sentido p73-R2.

p73-F1: 5'-TA(CGT)GCA(CGT)AAA(G)ACA(CGT)TGC(T)CC-3'
 p73-R1: 3'-TGC(T)GCA(CGT)TGC(T)CCA(CGT)GGA(CGT)A(C)G-5'
 p73-F2: 5'-TA(CGT)ATA(CT)A(C)GA(CGT)GTA(CGT)GAA(G)GG-3'
 p73-R2: 3'-ATGAAC(T)A(C)GA(CGT)A(C)GA(CGT)CCA(CGT)AT-5'

Mais especificamente, o ADNc foi sintetizado a partir do músculo esquelético humano poliA+ARN (Clontech) utilizando um iniciador aleatório e um iniciador oligo dT. Seguidamente uma biblioteca de ADNc constituída por cerca de 10⁷ placas construída utilizando ZipLox (Gibco BRL) como vector foi amplificada e o ADN foi extraído. Utilizando

0,21 µg de ADNc como molde e o dito p73-F1 e p73-R1 como iniciadores foi efectuada uma reacção de amplificação em 25 ciclos de 94°C, 30 s, 45°C, 30 s e 72°C de acordo com o manual da Tag polimerase (Gibco-BRL). Utilizando seguidamente 1/100 do produto de amplificação como molde e os ditos p73-F2 e p73-R2 como iniciadores foi efectuada mais uma amplificação sob as mesmas condições.

Uma vez que foi obtida uma banda de 172 pb deduzida a partir da estrutura do gene p53 foi preparado um mapa de cisão da enzima de restrição da banda. Como resultado, a presença de um gene diferente do gene p53 foi detectada. Esta banda foi subclonada em pGEM7(Promega) e utilizando um sequenciador automático AB1377 (ABI), a sequência nucleótídica foi determinada de maneira rotineira. Como resultado foi verificado ser um fragmento de ADN derivado de um novo gene que parece o gene p53 mas que tem uma sequência nucleótídica nova.

Separadamente, uma análise semelhante foi efectuada utilizando bibliotecas de ADNc derivadas de outros órgãos (por ex. do cérebro). Como resultado, um fragmento de ADN derivado de um outro gene novo diferente que se assemelha ao gene p53 foi detectado, mas verificou-se ser um fragmento derivado do gene p73.

Este fragmento de ADN subclonado foi excisado e utilizando um kit de marcação BcaBest (Takara); foi construída uma sonda marcada. O ensaio de hibridização em placa de uma biblioteca não amplificada de $2,4 \times 10^6$ placas construída utilizando o iniciador oligo dt isoladamente da mesma maneira que a construção da dita biblioteca de ADNc deu origem a 8 clones positivos. Uma vez que o χ ZipLox pode ser facilmente convertido num plasmídeo utilizando o sistema Cre-LoxP, sequenciação do plasmídeo obtido por conversão foi efectuada de maneira rotineira utilizando o

sequenciador automático LICOR e sequenciador automático ABI377 (ABI).

Seguidamente foi efectuada uma pesquisa de homologia entre a sequência nucleotídica do gene obtido e as sequências nucleotídicas dos genes p53 e p73 com um programa FASTA utilizando software GCG (Wisconsin Sequencing Package, Genetics Computer Group) [Person, W. R. e Lipman, D. J., Proc. Natl. Acad. Sci. USA, 85, 1435-1441(1988)].

Como resultado da dita pesquisa de homologia, dois dos clones seleccionados pelo processo acima e sequenciados verificou-se possuírem uma elevada homologia em relação ao gene p53 e gene p73. As massas moleculares calculadas a partir dos aminoácidos deduzidos codificados pela sequência génica destes 2 clones foram de 50,894 Da e cerca de 71900 Da, respectivamente.

Os presentes inventores designaram estes clones p51A e p51B, respectivamente.

A sequência nucleotídica de comprimento completo do gene (gene p51A) possuída pelo clone p51A obtido como acima referido é apresentada na SEQ ID N°:2 e a sequência nucleotídica de comprimento completo do gene (gene p51B) possuído pelo clone p51B é apresentada na SEQ ID N°: 5.

Como apresentado na SEQ ID N°:2, o clone p51A foi verificado possuir um gene que possui uma sequência nucleotídica (1344 nucleótidos) que codifica para a sequência de aminoácido (448 aminoácidos) da SEQ ID N°: 1, como uma grelha de leitura aberta na posição 145^a-1488^a. Além disso, a sequência de aminoácidos deduzida codificada pela sequência nucleotídica do gene possuído por este clone possuía um domínio de activação transcricional na posição 1^a -59^a, o domínio de ligação do ADN na posição 142^a-321^a e o domínio de oligomerização na posição 359^a-397^a.

Por outro lado, como apresentado na SEQ ID N°:5, o clone p51B foi verificado ter um gene que possui uma sequência nucleotídica (1923) que codifica para a sequência de aminoácidos (641 aminoácidos) da SEQ ID N°:4, como uma grelha de leitura aberta na posição 145^a-2067^a. Além disso, a sequência de aminoácidos deduzida codificada pela sequência nucleotídica do gene possuído por este clone possuía um domínio de activação transcricional na 1^a-59^a posição, um domínio de ligação de ADN na posição 142^a-321^a, e uma posição de oligomerização na posição 353^a-397^a. Além disso, verificou-se que esta sequência possuía uma sequência adicional, (domínio SAM) na região C-terminal e a região 353^a-641^a inclusive desta sequência adicional pode ser considerada como um domínio de oligomerização no sentido lato do termo.

A sequência de aminoácidos codificada pelo gene p51A foi comparada com as sequências de aminoácidos da proteína p53 e da proteína p73 β para comparação da homologia entre as três sequências (Fig. 2). No diagrama, os aminoácidos comuns às 3 sequências estão encaixilhados.

A Fig. 1 é um diagrama esquemático que apresenta características dos domínios estruturais da proteína p51A conjuntamente com aqueles da proteína p53 e proteína p73 β . No diagrama "TA", representa um domínio de activação transcricional, "ligação de ADN" representa um domínio de ligação do ADN, e "Oligo" representa um domínio de oligomerização. As características estruturais da proteína p51 e da proteína p73 β foram deduzidas a partir das características estruturais da proteína p53.

Como resultado, a homologia das sequências de aminoácidos deduzidas da proteína p51A, proteína p53 e p73 β em cada sequência de comprimento completo, domínio de activação transcricional, domínio de ligação do ADN, e domínio de oligomerização foi respectivamente como se

segue: entre a proteína p51A e a proteína p53, 36%, 22%, 60% e 37%, respectivamente; entre a proteína p51A e p73, 42%, 30%, 87% e 65%, respectivamente; e entre a proteína p53 e a proteína p73, 28%, 27%, 63% e 83%, respectivamente (Tabela 1).

Além disso, apesar da estrutura de 448 resíduos da proteína p51A ser mais curta do que a estrutura de 636 resíduos da proteína p73 α , a estrutura total da proteína p51A assemelha-se à proteína p73 com a região do terminal C cortada.

Estes resultados indicam que apesar da sequência de aminoácidos putativa da proteína p51A se assemelhar às sequências deduzidas de ambas as proteínas p53 e p73 β , a sua homologia com a sequência de aminoácidos da proteína p73 β é superior à homologia entre a proteína p53 e a proteína p73 β nas regiões diferentes do domínio de oligomerização. Além disso, entre a proteína p51A e proteína p73 β , a homologia foi encontrada na região em que nenhuma homologia foi encontrada entre a proteína p63 e proteína p73 β , ou entre a proteína p53 e a proteína p51A. Estes resultados sugerem que ao nível da sequência de aminoácidos, a proteína p51A pode ser dita estar mais perto da proteína p73 β do que a proteína p53.

De forma semelhante, a sequência de aminoácidos codificada pelo gene p51B foi comparada com a sequência de aminoácidos da proteína p73 α para comparação da homologia (Fig. 3). No diagrama, os aminoácidos comuns às duas sequências estão encaixilhados.

A Fig. 4 é um diagrama esquemático que apresenta características dos domínios estruturais das variantes de união codificadas pelos genes p51 (A e B) conjuntamente com os das proteínas p73 (α e β). Enquanto que o ponto de divergência entre a proteína p51A e a proteína p51B inicia-

se no intrão 10, e o ponto de divergência entre a proteína p73 α e p73 β inicia-se no intrão 13.

Exemplo 2 Confirmação da expressão de p51ARNm em tecido humano normal (para referência)

(1) Análise de transferência de Northern

A expressão de p51mARN no tecido normal foi avaliada por transferência de Northern utilizando, como sonda um clone humano de ADNc marcado pelo processo de iniciação oligonucleotídica aleatória.

A análise de transferência de Northern foi efectuada utilizando a transferência de Northern de tecido múltiplo humano (Clontech, Palo Alto, Calif., EUA) de acordo com o manual do produto.

Assim, o fragmento EcoRI(600 pb: correspondente ao terminal 5' do ADNc) de um produto de amplificação PCR do clone de ADN obtido no Exemplo 1 foi marcado com [³²p]-dCTP (kit de marcação de ADN de iniciação aleatória, Boehringer Mannheim) para ser utilizado como uma sonda.

A transferência foi efectuada utilizando Express-Hyb Hybridization Solution (Clontech) sob as condições prescritas pelo manual do utilizador, e a detecção foi efectuada utilizando BAS2000 (FUJI).

Os resultados são apresentados na Fig. 5 e na Fig. 6.

A Fig. 5 apresenta o resultado da hibridização de Northern efectuada com o filtro adquirido à Clontech, e a Fig. 6 apresenta o resultado da hibridização de Northern efectuada com um filtro construído pelos presentes inventores utilizando o ARN adquirido à Clontech. A Fig. 5 apresenta o electroforegrama com 2 μ g de poli A+ARN adicionado para cada banda e a Fig. 6 representa o

electroforegrama com 0,5 µg de poli A+ARN adicionado para cada banda.

As bandas na Fig. 5 representam o resultado para 1: coração, 2: cérebro, 3: placenta, 4: pulmão, 5: fígado, 6: músculo esquelético, 7:baço, e 8:pâncreas. As bandas na Fig. 6 representam o resultado para 1:glândula mamária, 2:próstata, 3:glândula salivar, 4:estômago, 5:timus, 6:tiróide, 7: traqueia, e 8: útero.

Verificou-se que a distribuição da expressão do ARNm (4.4 kb) do gene designado "gene p51 humano" estava bastante confinado em contraste com a expressão ubíqua de p53 ARNm com o nível de expressão mais elevado no músculo esquelético, e diminuía progressivamente na placenta, traqueia, glândula mamária, próstata, glândula salivar, timus, útero, estômago, pulmão, cérebro, e coração na ordem mencionada. Noutros tecidos (por ex. glândula adrenal, intestino delgado, medula espinal, baço), não pôde ser detectada a expressão de p51ARNm.

O padrão de expressão do gene p73 encontra-se também restringido pelo tecido. Contudo verificou-se que enquanto a expressão do gene p51 sobrepunha-se à expressão do gene p73 (expressão no mesmo tecido), a distribuição da expressão foi mais larga do que a distribuição da expressão do gene p73.

A diferença acima na distribuição do tecido de expressão entre o gene p51 humano, gene p53 e gene p73 sugeria que apesar da semelhança na actividade biológica entre estes genes eles são diferentes em função dependendo dos tecidos in vivo.

A continuação da investigação também revelou que em vários tecidos humanos, o ARNm de p51, tal como no caso da proteína p73, existe em formas submetidas a união selectiva (variantes de união alternativas), nomeadamente uma forma

curta que codifica para a proteína p51A e uma forma longa que codifica para a proteína p51B.

Verificou-se que a última forma longa que codifica para p51B possuía homologia com o factor designado ket que tinha sido acidentalmente descoberto numa pesquisa do receptor glutamato na língua. O ARNm de 3 kb que é um transcrito principal no músculo esquelético foi o ARNm mais abundante observado em todos os tecidos investigados. A forma curta do clone de ADNc suspeitou-se derivar do seu transcrito. É interessante que em contraste com o ARN observado nos tecidos normais esta forma curta (p51A) de ARNmp51 foi verificada ser expressa em muitas linhas celulares tumorais.

A Fig. 4 é um diagrama esquemático que compara a estrutura das variantes de união alternativas da proteína p51 e da proteína p73. Este ARNmp51B codificou para uma proteína que possui um massa molecular (calculada) semelhante à da p73 α .

Diferenças funcionais entre p51A e p51B permanecem desconhecidas.

Exemplo 3 Mapeamento do cromossoma do gene p51 (para referência)

Utilizando um painel de radiação híbrido (GeneBridge 4 Radiation Hybrid Panel; Research Genetics), o gene p51 foi mapeado no cromossoma humano. Como resultado o gene p51 foi localizado na região 3q28-ter entre os marcadores AFBM327YD9 e WI-1189 (5.66cR do anterior marcador).

Exemplo 4 Mutação de p51 em várias linhas celulares do cancro humano e tumores humanos (para referência)

A questão mais intrigante sobre o gene p51 é a questão se as características do gene supressor do tumor p53 são partilhadas pelo gene p51, assim como a relação de uma mutação do presente gene com a morfogénese de um tumor humano.

Por isso, utilizando as várias linhas celulares tumorais foi feita uma pesquisa, quanto à presença ou ausência da mutação do gene p51. Para esta pesquisa, foi utilizado o processo para análise funcional dos alelos separados em leveduras (FASAY), que foi previamente utilizado pelos presentes inventores na identificação e caracterização da mutação p53 [Ishioka et al., Nat. Genet. 5, 124-129(1993)].

Um fragmento de comprimento completo de ADN complementar que codifica para o gene p51A humano foi amplificado pelo mesmo processo PCR que o utilizado na determinação até agora descrita para adquirir a sequência nucleotídica do produto de amplificação que cobre a sequência de comprimento completo que codifica para o gene p51A e esta sequência nucleotídica foi determinada, através de sequenciação directa para detectar a presença ou ausência de uma mutação.

As células tumorais foram respectivamente cultivadas em Meio Essencial de Dulbecco Modificado suplementado com 5→10% de soro de vitela num meio de CO₂ a 5%. Uma vez que todos os clones p51AADNc podiam amplificar o ADNcp53 na análise prévia, a qualidade dos ADNcs das linhas celulares foi garantida.

Das 102 linhas celulares, 67 linhas analisadas foram capazes de amplificar o fragmento de ADN p51A. A sequência nucleotídica foi determinada por sequenciação directa de 35 linhas celulares referidas acima.

Foram encontradas mutações em duas linhas celulares, nomeadamente Ho-1-u-1 (JCRB0828) que é uma linha celular do

cancro da cabeça e pescoço, e SKG-IIIa (JCRB0611) que é uma linha celular de cancro cervical.

A mutação foi Ser¹⁴⁵→Leu na primeira e Gln¹⁶⁵→Leu na última. No que diz respeito à proteína p53 era provável que a função normal da proteína p53 tivesse sido danificada por mutação na primeira e pela infecção pelo vírus do papilloma humano na última. Além disso, foram notadas nos ARNms derivados de células tumorais, várias variantes de união.

No que diz respeito aos cancros primários humanos, as sequências nucleótídicas dos produtos de amplificação de ADN obtidos por SSCP e pelas técnicas de RT-PCR foram determinadas por sequenciação directa na pesquisa da mutação do gene p51A. Numa população total de 66 casos de tumores humanos, nomeadamente 8 casos de neuroblastoma, 8 casos de cancro do cólon, 8 casos de cancro do peito, 8 casos de cancro do pulmão, 8 casos de tumor do cérebro, 8 casos de cancro esofágico, 8 casos de cancro hepatocelular, 6 casos de cancro pancreático, e 4 casos de cancro renal, foi detectada uma mutação Ala¹⁴⁸→Pro num caso de cancro do pulmão.

A análise dos 3 casos acima foi invariavelmente uma análise do ADNc e foi claro que a expressão tinha origem num único local do cromossoma.

Exemplo Experimental 1 Supressão da formação de colónias por transformação de p51

A proteína p53 possui uma capacidade de bloquear células na fase G1 ou de induzir apoptose.

Para investigar a actividade inibidora da formação de colónias da proteína p51 da invenção, a linha celular do osteosarcoma SAOS2 (número de acesso: ATCC HTB85) foi cotransfectada com um plasmídeo de expressão resistente à puomicina (pBABEpuro: Morgenstern J. Nuc. Acids Ru, 18,

3587, 1990) assim como a construção de expressão p51A, uma construção de expressão marcada com HA (ATGTATCCATATGATGTTCCAGATTATGCT que codifica para a sequência de aminoácidos MYPYDVPDYA), uma construção de expressão p53, e um vector e a capacidade de formar colónias foi avaliada.

Foram construídos os vectores de expressão referidos acima clonando o fragmento da região de codificação do ADN de p51A (nucleótidos 2816; na SEQ ID N°:2, oligonucleótidos número 1-2816), o fragmento preparado adicionando a tag HA ao ADNc de p51A, e o fragmento da região de codificação de p53ADNc (nucleótidos 1698; oligonucleótidos número 62-1760), respectivamente. Seguidamente a linha celular do osteossarcoma SAOS2 foi cultivada em Meio Essencial de Dulbecco Modificado suplementado com 10% de soro fetal de vitela num ambiente de CO₂ a 5%. Uma placa de 6 cm foi inoculado com as células SAOS2 referidas acima (1x10⁶ células/disco), e, após 24 horas, as células foram transfectadas com um vector de expressão p51 do tipo selvagem contendo a cadeia de ADNc de p51A (pRcCMV/p51A). Foram efectuadas transformações semelhantes utilizando ADNc marcado com HA, o gene p53 do tipo selvagem e, como controlo, apenas o vector de expressão pRcCMV.

Utilizando o Kit de Transfecção de Mamíferos (Stratagene), 1 µg de pBABE puro foi introduzido nas células. As células resultantes foram fixas e coradas com violeta de cristal. As colónias coradas foram fotografadas. Cada transfecção foi efectuada duas vezes para analisar a formação de colónias.

Como resultado, diminuições significativas no número de colónias foram observadas nas placas de cultura de células transfectadas com genes p53 e células transfectadas com o gene p51. Em contraste, na placa de cultura de células tratadas apenas com o vector foi observado um

grande número de colónias. Verificou-se assim, que o gene p51 tinha uma capacidade de suprimir a formação de colónias, mas que esta capacidade era ligeiramente pobre, quando comparada com a capacidade do gene p53. Por outro lado, o gene p51 marcado com HA mostrou uma capacidade supressora da formação de colónias comparável com a do gene p53 (Fig. 7).

Exemplo Experimental 2 Teste da activação transcricional da proteína p51

Uma vez que a actividade da proteína p53 para parar o crescimento celular na fase G1 ou para induzir apoptose encontrava-se dependente da função de activação transcricional da proteína p53, foi efectuado um teste para ver se a proteína p51 iria exhibir tal actividade.

A jusante do promotor Waf1, que é conhecido por ser controlado pela função de activação transcricional de p53, e pela sequência RGC (cluster do gene ribossomal), um plasmídeo repórter de luciferase, assim como uma construção de expressão do gene p51A foram introduzidos através do processo descrito no Exemplo 5. Especificamente, as células SAOS2 foram co-transfectadas com o dito plasmídeo repórter da luciferase e o vector de expressão de p51A, vector de expressão p53 ou vector de controlo e a actividade da luciferase do lisado obtido a partir do transformante resultante foi ensaiada. A actividade da luciferase foi calculada com um sistema de ensaio da luciferase repórter dual (ProMega) considerando a eficiência da transfecção.

A Fig. 8 é um diagrama esquemático que mostra a construção repórter utilizada na experiência. São mostradas no diagrama 3 construções de luciferase cada uma ligada a jusante com os vários promotores p21WAF1. No diagrama, “o promotor WAF-1 luc” representa uma construção do promotor

p21WAF1 do tipo selvagem que retém os dois elementos de controlo p53; “del 1” representa a construção privada a montante dos ditos elementos; e “del 2” representa a construção privada de ambos os ditos elementos.

Os resultados são indicados na Fig. 9 e na Fig. 10. A actividade relativa das ordenadas representa a actividade da luciferase calculada utilizando o sistema de ensaio repórter da luciferase dual tomando em consideração a eficiência de transfecção.

A Fig. 9 apresenta a actividade de transactivação encontrada quando o plasmídeo de expressão p51 (p51A), o plasmídeo de expressão p53 (p53) ou apenas o vector (Rc/CMV) que tinha sido ligado a cada uma das várias construções repórter apresentadas na Fig. 8 foi introduzido em células SAOS2. Os resultados mostraram que tal como a proteína p53, a proteína p51 possui actividade para induzir um expressão dependente do número da sequência p53 reactiva.

A Fig. 10 mostra os resultados de uma experiência semelhante utilizando o plasmídeo de expressão de p51A (p51A), o plasmídeo de expressão p51A marcado com HA (Hap51A), o plasmídeo de expressão p53 (p53) ou o vector (RcCMV), que foi ligado à construção PGC repórter que foi demonstrada experimentalmente possuir reactividade p53. Como na experiência indicada na Fig. 9, os resultados acima indicaram que tanto p51A e HAp51A semelhantes a p53 possuíam actividade para induzir a expressão dependente do número da sequência p53 reactiva. A actividade fraca encontrada, quando o plasmídeo de expressão p51A foi utilizado pode ser atribuída ao facto de uma vez que o plasmídeo foi construído no vector de expressão com a sequência leader retida, a quantidade de expressão era pequena.

Quando a sequência leader foi removida numa experiência posterior, a proteína p51A mostrou uma expressão de acção indutora mais forte do que a proteína p53 e, também no dito ensaio de inibição de formação de colónias verificou-se que esta proteína possuía uma forte actividade.

É evidente a partir dos resultados acima que a proteína p51 possuía a capacidade de induzir a transcrição através do seu domínio de regulação da transcrição. A constatação de que a actividade de transcrição foi perdida na mutação induzida deste elemento sugere que a proteína p51 também utiliza a mesma sequência de reconhecimento que a proteína p53.

Seguidamente foi investigado se a mesma relação de transcrição também permanece verdadeira in vivo. Uma expressão da construção do gene p51A que possui um epitopo marcado com HA foi introduzida nas células SAOS2 durante um curto espaço de tempo. A verificação da tomada do gene p51A pelas células indicava que o p51A se encontra localizado no núcleo e verificou-se que todas as células elevaram o nível de p21Waf1. Isto indicou que a proteína p51 é também capaz de induzir p21Waf1 que é conhecido por ser controlado pela proteína p53.

Exemplo Experimental 3 Mutação do gene p51 nos tumores in situ (para referência)

A mutação do gene p51 foi investigada nas células de cancro in situ de 66 doentes (neuroblastoma, 8 casos, cancro do cólon, 8 casos; cancro da mama, 8 casos; cancro do pulmão, 8 casos; tumor do cérebro, 8 casos; cancro do esófago, 8 casos, hepatocitoma, 8 casos; cancro pancreático, 6 casos, e cancro renal, 4 casos), através do processo de PCR de transcrição reversa - polimorfismo de

estrutura de cadeia simples (RT-PCR-SSCP) e processo de sequenciação de ADN.

(1) Preparação de ARN

As amostras de tumores frescas foram isoladas cirurgicamente, imediatamente congeladas e armazenadas a -80°C até serem utilizadas. O ARN foi extraído através do processo descrito no relatório de Nakagawa et al. [Nakagawa, A. et al., N. Engl. J. Med., 328, 847-854(1993)].

(2) RT-PCR-SSCP e Sequenciação de ADN

O ADN total, 5 g foi transcrito em ADNc utilizando Superscript II transcriptase reversa-(Gibco-BRL) e iniciadores aleatórios. O 20° ADNc deste produto reaccional foi utilizado para amplificação por PCR. Foi efectuada PCR-SSCP de acordo com o processo de Mashiyama et al. [Mashiyama S. et al., Oncogene, 6, 1313-1318(1991)]. Especificamente, o produto PCR foi amplificado utilizando 3 iniciadores para o p51A ADNc.

As sequência nucleótídicas dos iniciadores utilizados para PCR são as seguintes.

```
p51-F1: 5'-AAAGAAAGTTATTACCGATG-3'
p51-R1: 5'-CGCGTGCTCTGTGTTATAGG-3'
p51-F2: 5'-CATGGACCAGCAGATTCAGA-3'
p51-R2: 5'-CATCACCTTGATCTGGATG-3'
p51-F3: 5'-CCACCTGGACGTATTCCACT-3'
p51-R3: 5'-TGGCTCATAAGGTACCAG-3'
p51-F4: 5'-CATGAGCTGAGCCGTGAAT-3'
p51-R4: 5'-TATCTTCATCCGCCTTCCTG-3'
p51-F5: 5'-ATGAACCGCCGTCCAATT-3'
p51-R5: 5'-GTGCTGAGGAAGGTACTGCA-3'
p51-F6: 5'-TGAAGATCAAAGAGTCCCTG-3'
p51-R6: 5'-CTAGTGGCTTTGTGCCTTTG-3'
```

Seguidamente, o 32PdCTP foi diluído com tampão de carga 1:10. Após 5 minutos de continuação da desnaturação a

98°C, a separação foi efectuada em 5% de glicerol/5% de gel de poliacrilamida a 200 volts à temperatura ambiente durante 12-14 horas. Após a electroforese, o gel foi seco e exposto contra um filme de raios-X de um dia para o outro de modo que as bandas de migração deveriam ficar definitivamente visíveis. Para confirmar a presença, ou a ausência da mutação, o produto PCR foi subclonado em pGEH-T Easy Vector (Promega), seguido por sequenciação com o sequenciador de ABI377 ADN.

Como resultado, no tecido de cancro do pulmão que pertence ao tipo de cancro de células escamosas altamente diferenciadas, uma substituição de aminoácidos pontual (Ala¹⁴⁸→Pro) foi encontrada na região de ligação do ADN deduzida da proteína p51A. Este tumor mostrou metástase para traqueal do nódulo linfático e invasão pleural. Uma vez que todos os 5 clones seleccionados aleatoriamente tinham a mesma mutação, o gene p51 possuído por esta célula tumoral foi sugerido ser um alelo único ou ter sido expresso mono-alelicamente.

Exemplo Experimental 4 Indução de apoptose pela introdução do p51ADNc (para referência)

Foi explorado se como a proteína p53, a proteína p51 iria induzir a apoptose celular.

O teste de indução de apoptose com a proteína p51 foi efectuado pelo processo dos presentes inventores, nomeadamente o processo referido acima, compreendendo a utilização de uma linha celular de euritroleucemia de um ratinho transgénico (linha celular 1-2-3) que apresenta as características típicas da apoptose, quando cultivadas a 32°C [Kato, M. V., et al., Int. J. Oncol., 9,269 (1996)].

Esta linha celular de euritroleucemia de um ratinho (linha celular 1-2-3) foi estabelecida a partir de

eritroleucemia derivada ratinhos transgénicos - de genes gp55 do vírus que se forma no foco do baço de Friend formador de vírus [Xu et al., Jpn J. Cancer Res. 68, 284-291(1995); Kato et al., Int. JP Oncol. 9, 269-277] e é uma estirpe de células que expressa apenas uma proteína p53 mutante sensível à temperatura (st) (mutação pontual: Ala1353Val). Esta é uma proteína p53 mutante st localizada no citoplasma a 37°C, que é uma temperatura de cultura habitual, e, por isso, não exhibe a função de controlo da molécula p53 que é intrinsecamente descarregada no núcleo, mas a 32°C migra para o núcleo de modo a que a actividade p53 é induzida [Levine, A. J. et al., Nature 351, 453-456(1991)]. Já foi reportado que nesta linha celular é induzida uma apoptose lenta a 32°C.

As células 1-2-3 foram cultivadas em meio 1640 RPMI suplementado com 10% de soro fetal de vitela num meio de CO₂ a 5%. Depois utilizando pRc/CMV como o vector de expressão, o gene p51A foi introduzido nas células acima. As células foram então cultivadas num meio selectivo e utilizando a resistência à neomicina como teste (Neo^t), foram seleccionadas células resistente G418. Foi então efectuado um estudo de indução em células que expressam p51A.

Assim, duas estirpes de células transfectadas com p51A (a seguir referidas como células 1-2-3 "1C1" e células "4B1") transfectadas com o gene p51A que albergada pelo vector de expressão (pRcCMV/p51A) e, como controlo, células 1-2-3 transformadas apenas com o vector e que não contêm o gene p51A foram inoculadas respectivamente em placas de 10 cm (dia) a uma concentração de 1×10^5 /ml e cultivadas a duas temperaturas alternativas de 37°C e 32°C durante 24 horas. As células foram então recolhidas e tratadas com proteinase K e Rnase A para preparar amostras de ADN. As amostras de

ADN foram sujeitas a electroforese em agarose. As imagens coradas com brometo de etídio são apresentadas na Fig. 11.

Como pode ser observado a partir da Fig. 11 na cultura a 37°C, em que não foi detectado nenhum fragmento de ADN nas células 1-2-3 (banda 1), pôde ser detectada fragmentação de ADN em oligómeros de 180pb nas células 1C1 e 4B1 transfectadas com o gene p51A (bandas 2 e 3).

Na cultura a 32°C foi detectada fragmentação de ADN em células 1-2-3 (banda 4) e a fragmentação de ADN foi promovida em células 1C1 e células 4B1 (bandas 5 e 6). Estes resultados estiveram em consonância com os resultados da observação morfológica da apoptose e os resultados do teste de inibição do crescimento de células p-51 introduzidas (32°C, 37°C).

A presença, ou ausência de alterações apoptóticas nas células foi estudada fixando as células respectivas em lamelas de vidro e, após coloração com Gimsa observando a morfologia celular e o grau de coloração microscopicamente. A contagem de células viáveis foi efectuada através de coloração com azul de tripano e contagem.

Como resultado, as células feitas crescer a 32°C possuíam projecções de superfície e apresentavam uma forma encolhida em tensão ou constringida. Além disso, nos espécimens celulares corados com Gimsa foi observada uma condensação de cromatina à volta do envelope nuclear ou das massas intracelulares. Em contraste, nas células cultivadas a 37°C não foi observada uma tal alteração na morfologia.

Em 24 horas de cultura a 32°C, as células que sofrem morte por apoptose e células que continuam o ciclo celular e crescimento estavam presentes em mistura, e após 24 horas a contagem de células viáveis em células que expressam p51 foi de $10^5/\text{ml}$ e a contagem de células 1-2-3 foi de $1,7 \times 10^5/\text{ml}$.

O referido anteriormente indicou que as células contendo os genes p51 tratados à temperatura de 32°C experimentaram uma apoptose repentina na presença de p53. Foi então confirmado que a proteína p51, tal como a proteína p53 induz a apoptose de forma significativa.

Exemplo Experimental 5

Foi preparado um anticorpo específico contra o gene C-terminal (570^a-641^a posições da sequência de aminoácidos) da proteína humana p51B as células humanas foram imunocoradas.

Assim, a região de codificação (as posições 185^a -206^a da sequência nucleotídica) do ADN humano p51B foi ligada ao vector de expressão da proteína de fusão GST pGEX-1 χ T (Pharmacia) e uma proteína de fusão foi sintetizada em Escherichia coli. Utilizando esta proteína de fusão e ratinhos BALB/C, foi preparado um anti-soro (anticorpo policlonal) de forma rotineira e absorvido com a proteína GST para disponibilizar um anticorpo específico contra a região C-terminal da proteína p51B.

O anticorpo acima foi sujeito a uma reacção primária com um espécimen de tecido de pele humana congelado e depois a uma reacção secundária com um anticorpo marcado com IgG de cabra anti-ratinho.

Como resultado, este anticorpo corou especificamente o núcleo da célula da camada da célula espinal até à camada base da pele humana. A especificidade foi encontrada porque o dito tratamento com a proteína de fusão anulou esta reacção, enquanto que o pré-tratamento com a proteína GST falhou em anular esta reacção.

Aplicabilidade Industrial

Os genes podem ser caracterizados como genes relacionados do gene p53 que é conhecido como um gene supressor do tumor. Com estes genes, os níveis de expressão e funções nas células podem ser analisados e espera-se que analisando os seus produtos de expressão, morbidez em várias doenças associadas com estes genes (por exemplo tumores malignos) possam ser elucidadas e diagnosticadas e modalidades terapêuticas possam ser estabelecidas.

Além disso, os genes são expressos em tecidos glandulares (próstata, glândula mamária), tecidos dos músculos, e timus e outros sistemas imunitários em contraste com p73 que é expresso no sistema nervoso e estão provavelmente envolvidos nas anormalidades destes tecidos. Por isso, espera-se que estes genes vão contribuir para o desenvolvimento de inibidores ou supressores de tais anormalidades. É fornecido um novo gene humano p51 que é de valor como gene supressor da proliferação celular. O novo gene assemelha-se ao gene que codifica para a proteína p53 ou proteína p73. Por isso o gene pode ser utilizado em estudos sobre as relações das funções analisadas destes genes relacionados com as várias doenças e utilizados em estudos para a aplicação ao diagnóstico de genes e à utilização medicinal destes genes em várias doenças. Além disso, utilizando o gene, o padrão da expressão do gene em vários tecidos humanos pode ser explorado e as suas funções no corpo humano podem ser analisadas.

Além disso, com este gene, a proteína p51 humana codificada pelo gene pode ser produzida em grandes quantidades pela tecnologia de engenharia humana. Assim, o gene e fragmentos de gene disponibilizados podem ser integrados com vectores de expressão para construir proteínas recombinantes humanas p51 e estudar a actividade da proteína p51 e as funções por ex. a actividade de ligação da proteína p51.

Além disso, a proteína p51 é útil para a elucidação patológica, diagnóstico e terapia de doenças associadas com o gene p51 ou o seu produto (por ex. doenças relacionadas com a actividade de transcrição de células e de várias doenças relacionadas com a apoptose, em particular de cancro).

A proteína p51 possui acções fisiológicas ou funções semelhantes às da proteína p53. Por isso, quando vários stresses biológicos tais como infecção viral, estimulação de citocinas, hipoxia, uma alteração no conjunto de nucleótidos, desarranjos metabólicos induzidos por fármacos, etc. actuam sobre tecidos vivos, a proteína da invenção exibe tais funções como transduções de sinais, através de interacções com outras proteínas e controlo transcrricional sobre os outros genes para modular deste modo a replicação do ADN da célula nos tecidos sujeitos aos ditos stresses biológicos, interromper o seu ciclo celular para reparar as células, eliminar as células através da apoptose, ou promover a diferenciação das células para controlar a defesa dos tecidos vivos contra os ditos stresses.

É fornecido um vector de transferência de genes contendo o gene p51 humano ou um seu alelo que é útil para a terapia génica, células em que o dito gene p51 ou alelo foi introduzido, um agente génico terapêutico contendo o dito vector ou células como princípios activos, e um processo de terapia génica para os explorar.

Além disso, é fornecido um produto farmacêutico contendo a proteína p51 como princípio activo que possui actividade para suprimir o crescimento de várias células cancerosas e encontra aplicação no tratamento de várias doenças neoplásicas e sintomas associados, através da dita actividade.

As regiões funcionais do gene humano p51 e o gene do ratinho correspondente são completamente idênticos com exceção dos 3 aminoácidos no domínio TA reflectindo assim um elevado grau de conservação e sugerindo a importância do presente gene [as sequências nucleótidas dos dois genes são mostradas de forma correlacionada nas Fig. 12-14 e na Fig. 15].

Listagem das Sequências

<110> Ikawa, Yoji
Otsuka Pharmaceutical Co. Ltd.

<120> Human p51 gene and its product

<130> P99-16

<140>
<141>

<150> JP P1998-10046T
<151> 1998-03-27

<160> 23

<170> PatentIn Ver. 2.0

<210> 1
<211> 448
<212> PRT
<213> Human

<220>
<221> DOMAIN
<222> (1).. (59)
<223> transactivation domain

<220>
<221> DNA-BIND
<222> (142).. (321)
<223> DNA binding domain

<220>
<221> DOMAIN
<222> (353).. (397)
<223> oligomerization domain

<400> 1

180	185	190
Lys Lys Ala Glu His Val Thr	Glu Val Val Lys Arg Cys Pro Asn His	
195	200	205
Glu Leu Ser Arg Glu Phe Asn Glu Gly Gln Ile Ala Pro Pro Ser His		
210	215	220
Leu Ile Arg Val Glu Gly Asn Ser His Ala Gln Tyr Val Glu Asp Pro		
225	230	235
Ile Thr Gly Arg Gln Ser Val Leu Val Pro Tyr Glu Pro Pro Gln Val		
245	250	255
Gly Thr Glu Phe Thr Thr Val Leu Tyr Asn Phe Met Cys Asn Ser Ser		
260	265	270
Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Val Thr Leu		
275	280	285
Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Cys Phe Glu Ala Arg		
290	295	300
Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp Ser Ile		
305	310	315
Arg Lys Gln Gln Val Ser Asp Ser Thr Lys Asn Gly Asp Gly Thr Lys		
325	330	335
Arg Pro Phe Arg Gln Asn Thr His Gly Ile Gln Met Thr Ser Ile Lys		
340	345	350
Lys Arg Arg Ser Pro Asp Asp Glu Leu Leu Tyr Leu Pro Val Arg Gly		
355	360	365
Arg Glu Thr Tyr Glu Met Leu Leu Lys Ile Lys Glu Ser Leu Glu Leu		
370	375	380
Met Gln Tyr Leu Pro Gln His Thr Ile Glu Thr Tyr Arg Gln Gln Gln		
385	390	395
Gln Gln Gln His Gln His Leu Leu Gln Lys His Leu Leu Ser Ala Cys		
405	410	415
Phe Arg Asn Glu Leu Val Glu Pro Arg Arg Glu Thr Pro Lys Gln Ser		
420	425	430
Asp Val Phe Phe Arg His Ser Lys Pro Pro Asn Arg Ser Val Tyr Pro		
435	440	445

```

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

```

<210> 2

<211> 2816

<212> DNA

<213> Human

<220>

<221> CDS

<222> (145)..(1488)

<220>

<221> polyA signal

<222> (2786)..(2791)

<400> 2

```

tcgttgatct caaagacagt tgaaggaat gaattttgaa acttcacggt gtgccacct 60
acagtactgc cctgacctt acaaccagcg ttctgtgaa accagctca ttctcttgg 120
aaagaaggtt attaccgata cacc atg tcc cag agc acc cag acc aat gaa 171
Met Ser Gln Ser Thr Gln Thr Asn Glu

```

1										5										
ttc	cic	agt	cca	gag	gtt	ttc	cag	cat	atc	tcg	gat	ttt	cig	gaa	cag					219
Phe	Leu	Ser	Pro	Glu	Val	Phe	Gln	His	Ile	Trp	Asp	Phe	Leu	Glu	Gln					
10					15					20					25					
ccg	ata	tgt	tca	gtt	cag	ccc	att	gac	tig	acc	ttt	gtg	gat	gaa	cca					267
Pro	Ile	Cys	Ser	Val	Gln	Pro	Ile	Asp	Leu	Asn	Phe	Val	Asp	Glu	Pro					
				30				35						40						
tca	gaa	gat	ggt	gcc	aca	aac	aag	att	gag	att	agc	atg	gac	tgt	atc					315
Ser	Glu	Asp	Gly	Ala	Thr	Asn	Lys	Ile	Glu	Ile	Ser	Met	Asp	Cys	Ile					
			45				50						55							
ccg	atg	cag	gac	tcg	gac	cig	agt	gac	ccc	atg	tcg	cca	cag	tac	acc					363
Arg	Met	Gln	Asp	Ser	Asp	Leu	Ser	Asp	Pro	Met	Trp	Pro	Gln	Tyr	Thr					
		60				65					70									
aac	cig	ggg	cic	cig	aac	agg	atg	gac	cag	cag	att	cag	aac	ggc	tcc					411
Asn	Leu	Gly	Leu	Leu	Asn	Ser	Met	Asp	Gln	Gln	Ile	Gln	Asn	Gly	Ser					
	75				80						85									
tcg	tcc	acc	agt	ccc	tat	aac	aca	gac	cac	ggg	cag	aac	ggc	gtc	acc					459
Ser	Ser	Thr	Ser	Pro	Tyr	Asn	Thr	Asp	His	Ala	Gln	Asn	Ser	Val	Thr					
	90				95				100					105						
ggg	ccc	tcg	ccc	tac	gca	cag	ccc	agg	tcc	acc	ttc	gat	gct	cic	tct					507
Ala	Pro	Ser	Pro	Tyr	Ala	Gln	Pro	Ser	Ser	Thr	Phe	Asp	Ala	Leu	Ser					
				110				115						120						
cca	tca	ccc	ggc	atc	ccc	tcc	aac	acc	gac	tac	cca	ggc	ccg	cac	agt					555
Pro	Ser	Pro	Ala	Ile	Pro	Ser	Asn	Thr	Asp	Tyr	Pro	Gly	Pro	His	Ser					
			125				130					135								
ttc	gac	gtg	tcc	ttc	cag	cag	tcg	agg	acc	ggc	aag	tcg	ggc	acc	tcg					603
Phe	Asp	Val	Ser	Phe	Gln	Gln	Ser	Ser	Thr	Ala	Lys	Ser	Ala	Thr	Trp					
	140					145					150									
acc	tat	tcc	act	gaa	cig	aag	aaa	cic	tac	tgc	caa	att	gca	aag	aca					651
Thr	Tyr	Ser	Thr	Glu	Leu	Lys	Lys	Leu	Tyr	Cys	Gln	Ile	Ala	Lys	Thr					
	155				160					165										
tgc	ccc	atc	cag	atc	aag	gtg	atg	acc	cca	ccg	ccg	cag	gga	gct	gtt					699
Cys	Pro	Ile	Gln	Ile	Lys	Val	Met	Thr	Pro	Pro	Pro	Gln	Gly	Ala	Val					
	170				175				180					185						
atc	cgc	ggc	atg	ccg	gtc	tac	aaa	aaa	gct	gag	cac	gtc	acc	gag	gtg					747
Ile	Arg	Ala	Met	Pro	Val	Tyr	Lys	Lys	Ala	Glu	His	Val	Thr	Glu	Val					
			190					195					200							
gtg	aag	ccg	tcg	ccc	aac	cat	gag	cig	agg	cgt	gaa	ttc	aac	gag	gga					795
Val	Lys	Arg	Cys	Pro	Asn	His	Glu	Leu	Ser	Arg	Glu	Phe	Asn	Glu	Gly					
			205				210					215								
cag	att	ggc	ccg	ccg	agt	cat	tig	att	cga	gta	gag	ggg	aac	agg	cat					843
Gln	Ile	Ala	Pro	Pro	Ser	His	Leu	Ile	Arg	Val	Glu	Gly	Asn	Ser	His					
		220				225					230									
ggc	cag	tat	gta	gaa	gat	ccc	atc	aca	gga	aga	cag	agt	gtg	cig	gta					891
Ala	Gln	Tyr	Val	Glu	Asp	Pro	Ile	Thr	Gly	Arg	Gln	Ser	Val	Leu	Val					
	235				240				245											
ccg	tat	gag	cca	ccc	cag	gtt	ggc	act	gaa	ttc	acc	aca	gtc	tig	tac					939
Pro	Tyr	Glu	Pro	Pro	Gln	Val	Gly	Thr	Glu	Phe	Thr	Thr	Val	Leu	Tyr					
	250				255				260				265							
aat	ttc	atg	tgt	aac	agg	agt	tgt	gtt	gga	ggg	atg	aac	cgc	cgt	cca					987
Asn	Phe	Met	Cys	Asn	Ser	Ser	Cys	Val	Gly	Gly	Met	Asn	Arg	Arg	Pro					
			270				275					280								
att	tta	atc	att	gtt	act	cig	gaa	acc	aga	gat	ggg	caa	gtc	cig	ggc					1035
Ile	Leu	Ile	Ile	Val	Thr	Leu	Glu	Thr	Arg	Asp	Gly	Gln	Val	Leu	Gly					
			285				290					295								

cgg cgc tgc tti gag gcc cgg atc tgt gct tgc cca gga aga gac agg 1083
 Arg Arg Cys Phe Glu Ala Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg
 300 305 310

aag gcg gat gaa gat agc atc aga aag cag caa gti tgc gac agt aca 1131
 Lys Ala Asp Glu Asp Ser Ile Arg Lys Gln Gln Val Ser Asp Ser Thr
 315 320 325

aag aac ggt gat ggt acc aag cgc cgc tti cgt cag aac aca cat ggt 1179
 Lys Asn Gly Asp Gly Thr Lys Arg Pro Phe Arg Gln Asn Thr His Gly
 330 335 340 345

atc cag atg aca tcc atc aag aaa cga aga tcc cca gat gat gaa ctg 1227
 Ile Gln Met Thr Ser Ile Lys Lys Arg Arg Ser Pro Asp Asp Glu Leu
 350 355 360

tta tac tta cca gtg agg ggc cgt gag act tat gaa atg ctg tta agg 1275
 Leu Tyr Leu Pro Val Arg Gly Arg Glu Thr Tyr Glu Met Leu Leu Lys
 365 370 375

atc aaa gag tcc ctg gaa ctc atg cag tac ctt cct cag cac aca att 1323
 Ile Lys Glu Ser Leu Glu Leu Met Gln Tyr Leu Pro Gln His Thr Ile
 380 385 390

gaa acc tac agg caa cag caa cag cag cag cac cag cac tta ctt cag 1371
 Glu Thr Tyr Arg Gln Gln Gln Gln Gln Gln His Gln His Leu Leu Gln
 395 400 405

aaa cat ctc ctt tca gcc tgc ttc agg aat gag ctt gtg gag ccc cgg 1419
 Lys His Leu Leu Ser Ala Cys Phe Arg Asn Glu Leu Val Glu Pro Arg
 410 415 420 425

aga gaa act cca aaa caa tct gac gtc ttc tti aga cat tcc aag ccc 1467
 Arg Glu Thr Pro Lys Gln Ser Asp Val Phe Phe Arg His Ser Lys Pro
 430 435 440

cca aac cga tca gtg tac cca tagagcccta tcctatatatt ttaagtgtgt 1518
 Pro Asn Arg Ser Val Tyr Pro
 445

gttgtgtatt tccatgtgta taltgtgtgt taltgtgtgt taltgtgtgt cgtgtgtatc 1578
 tagccctcat aaacaggact tgaagacact ttggctcaga gacccaactg ctcaaggcca 1638
 caaagccact agttagagaa tcttttgaag gactcaaac ctttacaaga aaggtgttll 1698
 tctgcagatt ttgtatcctt agaccgcca ttgttgggtg aggaaccact gtgtttctct 1758
 gtgagcttcc tttgttttcc tggaggggag aggtcaggta gggaaagggg cattagatg 1818
 ttatttgaa cctttttctg tcttctctg ttgttttctt aaatttcaca ggaagcttll 1878
 tggcaggctc tcaacctaa gttgtcttll taagaaaggg agaaasagtt tttttgtg 1938
 tttgtatag taagtgtag gttactgaga gactcagta gacctttta atgtctgtca 1998
 tttatataa ttgcaagtag taagaaagga aggtgtcag ttatctgtg gcaagcaggg 2058
 tttatctac caaagtaat caactttgt gttggagagt tctttgtag aactgtcatl 2118
 atttgttcc tccctcatg tttaggtaga acatttctta atgtgtgtta cctgtctgtg 2178
 ccactgtatg ttggatctg ttatctttaa gtttttctg tacatgaac cctggagagc 2238
 ctactacaas aaactgtgt ttggccccc atagcagggt aactatttll gtgtttttaa 2298
 tagaaagaca aatccacccc agtaatatgt cctttacgta gtgttttacc attattcaas 2358
 gttcaaaata gaatttgaag cctcttcaca aaattcttga ttattttgtt taattagagc 2418
 ttctatccct caagcctacc taccataaaa ccagccatct tacgtatctt gtctagtgca 2478
 tttagccagg agacttact tttaggttag ttgattccaa gcagacgtgt taattcagc 2538

```

atccaggac tggasttaa agatgaaag gtagactac tticliitt ilactaaa 2598
agittagaga atcicggtt cttccattt taanaacaa ittaagata atagataaa 2658
gacttanaa agttccicc cctccatct cccacacca gtcacagca cgtcttttc 2718
tgccaccaag acaagattt ctgtttatg aggtcttgc ttigiggat agtgattt 2778
attttccat aaacttgcg atctggtt aaagaaa 2816

```

```

<210> 3
<211> 448
<212> PRT
<213> Human

<400> 3

```

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

```

Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp Ser Ile
305              310              315              320
Arg Lys Gln Gln Val Ser Asp Ser Thr Lys Asn Gly Asp Gly Thr Lys
              325              330              335
Arg Pro Phe Arg Gln Asn Thr His Gly Ile Gln Met Thr Ser Ile Lys
              340              345              350
Lys Arg Arg Ser Pro Asp Asp Glu Leu Leu Tyr Leu Pro Val Arg Gly
              355              360              365
Arg Glu Thr Tyr Glu Met Leu Leu Lys Ile Lys Glu Ser Leu Glu Leu
              370              375              380
Met Gln Tyr Leu Pro Gln His Thr Ile Glu Thr Tyr Arg Gln Gln Gln
385              390              395              400
Gln Gln Gln His Gln His Leu Leu Gln Lys His Leu Leu Ser Ala Cys
              405              410              415
Phe Arg Asn Glu Leu Val Glu Pro Arg Arg Glu Thr Pro Lys Gln Ser
              420              425              430
Asp Val Phe Phe Arg His Ser Lys Pro Pro Asn Arg Ser Val Tyr Pro
              435              440              445

```

<210> 4
 <211> 641
 <212> PRT
 <213> Human

<220>
 <221> DOMAIN
 <222> (1), (59)
 <223> transactivation domain

<220>
 <221> DNA_BIND
 <222> (142), (321)
 <223> DNA binding domain

<220>
 <221> DOMAIN
 <222> (353), (397)
 <223> oligomerization domain

<400> 4

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

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

```

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

```

<210> 5

<211> 2270

<212> DNA

<213> Human

<220>

<221> CDS

<222> (145)..(2067)

<400> 5

```

tcgttgatct caagacagc tgaaggaat gaattttgaa acttcacggc gtcacacct 60
acagtaacgac cctgacctt acatccagcg tticgtagaa acccagctca tttctcttgg 120
aaagaaagtt attaccgac cacc atg tcc cag agc aca cag aca aat gaa 171
Met Ser Gln Ser Thr Gln Thr Asn Glu
1 5
ttc ctc agt cca gag gtt ttc cag cat atc tgg gat ttt ctg gaa cag 219
Phe Leu Ser Pro Glu Val Phe Gln His Ile Trp Asp Phe Leu Glu Gln
10 15 20 25
cct ata tgt tca gtt cag ccc att gac ttg aac ttt gtg gat gaa cca 267
Pro Ile Cys Ser Val Gln Pro Ile Asp Leu Asn Phe Val Asp Glu Pro
30 35 40
tca gaa gat ggt gcg aca aac aag att gag att agc atg gac tgt atc 315
Ser Glu Asp Gly Ala Thr Asn Lys Ile Glu Ile Ser Met Asp Cys Ile
45 50 55
cgc atg cag gac tgg gac ctg agt gac ccc atg tgg cca cag tac acg 363
Arg Met Gln Asp Ser Asp Leu Ser Asp Pro Met Trp Pro Gln Tyr Thr
60 65 70
aac ctg ggg ctc ctg aac agc atg gac cag cag att cag aac ggc tcc 411
Asn Leu Gly Leu Leu Asn Ser Met Asp Gln Gln Ile Gln Asn Gly Ser
75 80 85
tcg tcc acc agt ccc tat aac aca gac cac gcg cag aac agc gtc acg 459
Ser Ser Thr Ser Pro Tyr Asn Thr Asp His Ala Gln Asn Ser Val Thr
90 95 100 105
gcg ccc tgg ccc tac gca cag ccc agc tcc acc ttc gat gct ctc tct 507
Ala Pro Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Ala Leu Ser
110 115 120
cca tca ccc gcc atc ccc tcc aac acc gac tac cca ggc ccg cac agt 555
Pro Ser Pro Ala Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His Ser
125 130 135

```

ttc gac gtc tcc ttc cag cag tcc agc acc gcc aag tcc gcc acc tag	603
Phe Asp Val Ser Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp	
140 145 150	
acc tat tcc act gaa ctg aag aaa ctc tac tgc cca att gca aag aca	651
Thr Tyr Ser Thr Glu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr	
155 160 165	
tgc ccc atc cag atc aag gtc atc acc cca cct cct cag gga gct gtt	699
Cys Pro Ile Gln Ile Lys Val Met Thr Pro Pro Pro Gln Gly Ala Val	
170 175 180 185	
atc cgc gcc atg cct gtc tac aaa aaa gct gag cac gtc acg gag gtc	747
Ile Arg Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Glu Val	
190 195 200	
gtg aag cag tgc ccc aac cat gag ctg agc cgt gaa ttc aac gag gga	795
Val Lys Arg Cys Pro Asn His Glu Leu Ser Arg Glu Phe Asn Glu Gly	
205 210 215	
cag att gcc cct cct agt cat ttg att cga gtc gag ggg aac agc cat	843
Gln Ile Ala Pro Pro Ser His Leu Ile Arg Val Glu Gly Asn Ser His	
220 225 230	
gcc cag tat gtc gaa gat ccc atc aca gga aga cag agt gtc ctg gta	891
Ala Gln Tyr Val Glu Asp Pro Ile Thr Gly Arg Gln Ser Val Leu Val	
235 240 245	
cct tat gag cca ccc cag gtt gcc act gaa ttc acg aca gtc ttg tac	939
Pro Tyr Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Val Leu Tyr	
250 255 260 265	
aat ttc atg tgt aac agc agt tgt gtt gga ggg aig aac cgc cgt cca	987
Asn Phe Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro	
270 275 280	
att tta atc att gtt act ctg gaa acc aca gat ggg caa gtc ctg gcc	1035
Ile Leu Ile Ile Val Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly	
285 290 295	
caa cgc tgc ttt gag gcc cgc atc tct gct tgc cca gga aga gac agg	1083
Arg Arg Cys Phe Glu Ala Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg	
300 305 310	
aag gcc gat gaa gat agc atc aga aag cag caa gtt tcc gac agt aca	1131
Lys Ala Asp Glu Asp Ser Ile Arg Lys Gln Gln Val Ser Asp Ser Thr	
315 320 325	
aag aac ggt gat ggt acg aag cgt ccg ttt cgt cag aac aca cat ggt	1179
Lys Asn Gly Asp Gly Thr Lys Arg Pro Phe Arg Gln Asn Thr His Gly	
330 335 340 345	
atc cag atg aca tcc atc aag aaa cga aga tcc cca gat gat gaa ctg	1227
Ile Gln Met Thr Ser Ile Lys Lys Arg Arg Ser Pro Asp Asp Glu Leu	
350 355 360	
tta tac tta cca gtc agc gcc cgt gag act tat gaa atg ctg ttg aag	1275
Leu Tyr Leu Pro Val Arg Gly Arg Glu Thr Tyr Glu Met Leu Leu Lys	
365 370 375	
atc aaa gag tcc ctg gaa ctc atg cag tac ctt cct cag cac aca att	1323
Ile Lys Glu Ser Leu Glu Leu Met Gln Tyr Leu Pro Gln His Thr Ile	
380 385 390	
gaa acg tac agg caa cag caa cag cag cag cac cag cac tta ctt cag	1371
Glu Thr Tyr Arg Gln Gln Gln Gln Gln His Gln His Leu Leu Gln	
395 400 405	
aaa cag acc tca ata cag tct cca tct tca tat ggt aac agc tcc cca	1419
Lys Gln Thr Ser Ile Gln Ser Pro Ser Ser Tyr Gly Asn Ser Ser Pro	
410 415 420 425	
cct ctg aac aaa atg aac agc atg aac aag ctg cct tct gtc agc cag	1467

Pro Leu Asn Lys Met Asn Ser Met Asn Lys Leu Pro Ser Val Ser Gln
 430 435 440
 ctt att aac cct cag cag cgc aac gcc ctc act cct acc acc att cct 1515
 Leu Ile Asn Pro Gln Gln Arg Asn Ala Leu Thr Pro Thr Thr Ile Pro
 445 450 455
 gat gcc atg gga gcc aac att ccc atg atg gcc acc cac atg cca atg 1563
 Asp Gly Met Gly Ala Asn Ile Pro Met Met Gly Thr His Met Pro Met
 460 465 470
 gct gga gac atg aat gga ctc acc ccc acc cag gca ctc cct ccc cca 1611
 Ala Gly Asp Met Asn Gly Leu Ser Pro Thr Gln Ala Leu Pro Pro Pro
 475 480 485
 ctc tcc atg cca tcc acc tcc cac tgc aca ccc cca cct cag tat ccc 1659
 Leu Ser Met Pro Ser Thr Ser His Cys Thr Pro Pro Pro Pro Tyr Pro
 490 495 500 505
 aca gat tgc acc att gtc agt ttc tta gcc agc ttg gcc tct tca tca 1707
 Thr Asp Cys Ser Ile Val Ser Phe Leu Ala Arg Leu Gly Cys Ser Ser
 510 515 520
 tet ctc gac tat ttc acc acc cag ggg ctc acc acc atc tat cag att 1755
 Cys Leu Asp Tyr Phe Thr Thr Gln Gly Leu Thr Thr Ile Tyr Gln Ile
 525 530 535
 gag cat tac tcc atg gat gat ctc gca agt ctc aaa atc cct gag cca 1803
 Glu His Tyr Ser Met Asp Asp Leu Ala Ser Leu Lys Ile Pro Glu Gln
 540 545 550
 tit cga cat gcc atc tgg aag gcc atc ctc gac cac cga cag ctc cac 1851
 Phe Arg His Ala Ile Trp Lys Gly Ile Leu Asp His Arg Gln Leu His
 555 560 565
 gaa ttc tcc tcc cct tct cat ctc ctc cgg acc cca agc agt gcc tct 1899
 Glu Phe Ser Ser Pro Ser His Leu Leu Arg Thr Pro Ser Ser Ala Ser
 570 575 580 585
 aca gtc agt gtc gcc tcc aat gag acc cca ggt gag cgt gtt att gat 1947
 Thr Val Ser Val Gly Ser Ser Glu Thr Arg Gly Glu Arg Val Ile Asp
 590 595 600
 gct gtc cga ttc acc ctc cgc cag acc atc tct ttc cca ccc cga gat 1995
 Ala Val Arg Phe Thr Leu Arg Gln Thr Ile Ser Phe Pro Pro Arg Asp
 605 610 615
 gag tgg aat gac ttc aac ttt gat atg gat gct cgc cgc aat aag cca 2043
 Glu Trp Asn Asp Phe Asn Phe Asp Met Asp Ala Arg Arg Asn Lys Gln
 620 625 630
 cag cgc atc aaa gag gag ggg gag tgaaccac catgtgagct ctctctatcc 2097
 Gln Arg Ile Lys Glu Glu Gly Glu
 635 640
 ctctctaac tgcagcccc ctasagcac tctgcctaa tctcaagc ctctctccca 2157
 gctctctccc tctctctgt ctagttctt agggaaagga gaagtaagag gctactcttt 2217
 acctaacac tgaactggca tctaatctg attctgxtt taagcttca aaa 2270

<210> 6
 <211> 641
 <212> PRT
 <213> Human
 <400> 6

Met Ser Gln Ser Thr Gln Thr Asn Glu Phe Leu Ser Pro Glu Val Phe
 1 5 10 15
 Gln His Ile Trp Asp Phe Leu Glu Gln Pro Ile Cys Ser Val Gln Pro
 20 25 30

Ile Asp Leu Asn Phe Val Asp Glu Pro Ser Glu Asp Gly Ala Thr Asn
 35 40 45
 Lys Ile Glu Ile Ser Met Asp Cys Ile Arg Met Gln Asp Ser Asp Leu
 50 55 60
 Ser Asp Pro Met Trp Pro Gln Tyr Thr Asn Leu Gly Leu Leu Asn Ser
 65 70 75 80
 Met Asp Gln Gln Ile Gln Asn Gly Ser Ser Ser Thr Ser Pro Tyr Asn
 85 90 95
 Thr Asp His Ala Gln Asn Ser Val Thr Ala Pro Ser Pro Tyr Ala Gln
 100 105 110
 Pro Ser Ser Thr Phe Asp Ala Leu Ser Pro Ser Pro Ala Ile Pro Ser
 115 120 125
 Asn Thr Asp Tyr Pro Gly Pro His Ser Phe Asp Val Ser Phe Gln Gln
 130 135 140
 Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Thr Glu Leu Lys
 145 150 155 160
 Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln Ile Lys Val
 165 170 175
 Met Thr Pro Pro Pro Gln Gly Ala Val Ile Arg Ala Met Pro Val Tyr
 180 185 190
 Lys Lys Ala Glu His Val Thr Glu Val Val Lys Arg Cys Pro Asn His
 195 200 205
 Glu Leu Ser Arg Glu Phe Asn Glu Gly Gln Ile Ala Pro Pro Ser His
 210 215 220
 Leu Ile Arg Val Glu Gly Asn Ser His Ala Gln Tyr Val Glu Asp Pro
 225 230 235 240
 Ile Thr Gly Arg Gln Ser Val Leu Val Pro Tyr Glu Pro Pro Gln Val
 245 250 255
 Gly Thr Glu Phe Thr Thr Val Leu Tyr Asn Phe Met Cys Asn Ser Ser
 260 265 270
 Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Val Thr Leu
 275 280 285
 Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Cys Phe Glu Ala Arg
 290 295 300
 Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp Ser Ile
 305 310 315 320
 Arg Lys Gln Gln Val Ser Asp Ser Thr Lys Asn Gly Asp Gly Thr Lys
 325 330 335
 Arg Pro Phe Arg Gln Asn Thr His Gly Ile Gln Met Thr Ser Ile Lys
 340 345 350
 Lys Arg Arg Ser Pro Asp Asp Glu Leu Leu Tyr Leu Pro Val Arg Gly
 355 360 365
 Arg Glu Thr Tyr Glu Met Leu Leu Lys Ile Lys Glu Ser Leu Glu Leu
 370 375 380
 Met Gln Tyr Leu Pro Gln His Thr Ile Glu Thr Tyr Arg Gln Gln Gln
 385 390 395 400
 Gln Gln Gln His Gln His Leu Leu Gln Lys Gln Thr Ser Ile Gln Ser
 405 410 415
 Pro Ser Ser Tyr Gly Asn Ser Ser Pro Pro Leu Asn Lys Met Asn Ser
 420 425 430

```

Met Asn Lys Leu Pro Ser Val Ser Gln Leu Ile Asn Pro Gln Gln Arg
435                      440                      445

Asn Ala Leu Thr Pro Thr Thr Ile Pro Asp Gly Met Gly Ala Asn Ile
450                      455                      460

Pro Met Met Gly Thr His Met Pro Met Ala Gly Asp Met Asn Gly Leu
465                      470                      475                      480

Ser Pro Thr Gln Ala Leu Pro Pro Pro Leu Ser Met Pro Ser Thr Ser
485                      490                      495

His Cys Thr Pro Pro Pro Pro Tyr Pro Thr Asp Cys Ser Ile Val Ser
500                      505                      510

Phe Leu Ala Arg Leu Gly Cys Ser Ser Cys Leu Asp Tyr Phe Thr Thr
515                      520                      525

Gln Gly Leu Thr Thr Ile Tyr Gln Ile Glu His Tyr Ser Met Asp Asp
530                      535                      540

Leu Ala Ser Leu Lys Ile Pro Glu Gln Phe Arg His Ala Ile Trp Lys
545                      550                      555                      560

Gly Ile Leu Asp His Arg Gln Leu His Glu Phe Ser Ser Pro Ser His
565                      570                      575

Leu Leu Arg Thr Pro Ser Ser Ala Ser Thr Val Ser Val Gly Ser Ser
580                      585                      590

Glu Thr Arg Gly Glu Arg Val Ile Asp Ala Val Arg Phe Thr Leu Arg
595                      600                      605

Gln Thr Ile Ser Phe Pro Pro Arg Asp Glu Trp Asn Asp Phe Asn Phe
610                      615                      620

Asp Met Asp Ala Arg Arg Asn Lys Gln Gln Arg Ile Lys Glu Glu Gly
625                      630                      635                      640

Glu

```

<210> 7

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p73-F1 sense primer

<400> 7

tacgtgcaag taagacaag ttgtcc 27

<210> 8

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p73-R1 antisense primer

<400> 8

tgtgcaogl tgtccaogl ggaogtaog 29

<210> 9

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p73-F2 sense primer

<400> 9

tacgtatact aacacgtgta cgtgaaggg 29

<210> 10

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p73-R2 antisense primer

<400> 10

atgaactaag aagtaagaag tcaacgtat 29

<210> 11

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:HA-labeled expression construct

<400> 11

atgtatccat atgatgttcc agattaigct 30

<210> 12

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-F1 sense primer

<400> 12

aaagaaggtt attacgatg 20

<210> 13

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-R1 antisense primer

<400> 13

cggttggtct gtgttatagg 20

<210> 14

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-F2 sense primer

<400> 14

catggaccag cagattcaga 20

<210> 15

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-R2 antisense primer

<400> 15

catcaccitg atciggatg 19

<210> 16

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-F3 sense primer

<400> 16

ccaactggac gtattccact 20

<210> 17

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-R3 antisense primer

<400> 17

tggtccataa ggtaccag 18

<210> 18

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-F4 sense primer

<400> 18

catgagctga gccgtgaat 19

<210> 19

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-R4 antisense primer

<400> 19
tatcttcac cgccttctg 20

<210> 20
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:p51-F5 sense primer

<400> 20
atgaaccgcc gtaacaatt 18

<210> 21
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:p51-R5 antisense primer

<400> 21
gagctgagga aggtactgca 20

<210> 22
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:p51-F6 sense primer

<400> 22
tgaagatcaa agagtcctg 20

<210> 23
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:p51-R6 antisense primer

<400> 23
ctagtggctt tgtgctttg 20

SEQUENCE LISTING

[0339]

<110> Ikawa, Yoji
Otsuka Pharmaceutical Co. Ltd.

<120> Human p51 gene and its product

<130> P99-16

<140>

<141>

<150> JP P1998-100467

<151> 1998-03-27

<160> 23

<170> PatentIn Ver. 2.0

<210> 1

<211> 448

<212> PRT

<213> Human

<220>

<221> DOMAIN

<222> (1).. (59)

<223> transactivation domain

<220>

<221> DNA BIND

<222> (142).. (321)

<223> DNA binding domain

<220>

<221> DOMAIN

<222> (353).. (397)

<223> oligomerization domain

<400> 1

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

180										185										190									
Lys	Lys	Ala	Glu	His	Val	Thr	Glu	Val	Val	Lys	Arg	Cys	Pro	Asn	His														
		195						200				205																	
Glu	Leu	Ser	Arg	Glu	Phe	Asn	Glu	Gly	Gln	Ile	Ala	Pro	Pro	Ser	His														
		210				215					220																		
Leu	Ile	Arg	Val	Glu	Gly	Asn	Ser	His	Ala	Gln	Tyr	Val	Glu	Asp	Pro														
		225			230					235					240														
Ile	Thr	Gly	Arg	Gln	Ser	Val	Leu	Val	Pro	Tyr	Glu	Pro	Pro	Gln	Val														
			245						250					255															
Gly	Thr	Glu	Phe	Thr	Thr	Val	Leu	Tyr	Asn	Phe	Met	Cys	Asn	Ser	Ser														
			260					265					270																
Cys	Val	Gly	Gly	Met	Asn	Arg	Arg	Pro	Ile	Leu	Ile	Ile	Val	Thr	Leu														
		275				280						285																	
Glu	Thr	Arg	Asp	Gly	Gln	Val	Leu	Gly	Arg	Arg	Cys	Phe	Glu	Ala	Arg														
		290				295					300																		
Ile	Cys	Ala	Cys	Pro	Gly	Arg	Asp	Arg	Lys	Ala	Asp	Glu	Asp	Ser	Ile														
		305			310					315					320														
Arg	Lys	Gln	Gln	Val	Ser	Asp	Ser	Thr	Lys	Asn	Gly	Asp	Gly	Thr	Lys														
			325					330					335																
Arg	Pro	Phe	Arg	Gln	Asn	Thr	His	Gly	Ile	Gln	Met	Thr	Ser	Ile	Lys														
			340					345					350																
Lys	Arg	Arg	Ser	Pro	Asp	Asp	Glu	Leu	Leu	Tyr	Leu	Pro	Val	Arg	Gly														
		355				360						365																	
Arg	Glu	Thr	Tyr	Glu	Met	Leu	Leu	Lys	Ile	Lys	Glu	Ser	Leu	Glu	Leu														
		370				375					380																		
Met	Gln	Tyr	Leu	Pro	Gln	His	Thr	Ile	Glu	Thr	Tyr	Arg	Gln	Gln	Gln														
		385			390					395					400														
Gln	Gln	Gln	His	Gln	His	Leu	Leu	Gln	Lys	His	Leu	Leu	Ser	Ala	Cys														
			405					410						415															
Phe	Arg	Asn	Glu	Leu	Val	Glu	Pro	Arg	Arg	Glu	Thr	Pro	Lys	Gln	Ser														
		420						425					430																
Asp	Val	Phe	Phe	Arg	His	Ser	Lys	Pro	Pro	Asn	Arg	Ser	Val	Tyr	Pro														
		435				440						445																	

<210> 2

<211> 2816

<212> DNA

<213> Human

<220>

<221> CDS

<222> (145).. (1488)

<220>

<221> polyA signal

<222> (2786).. (2791)

<400> 2

```

tctgtgtat caagacagt tgaggaaat gaattttaa acttcacggt gggccaccct 60
acagtactgc ccgaccctt acaaccagtg ttctgtagaa accagctca ttctcttgg 120
aaagasagti attaccgatc cacc atg tcc cag agc aca cag aca aat gaa 171
Met Ser Gln Ser Thr Gln Thr Asn Glu

```

	1	5	
ttc ctc agt cca gag gtt ttc cag cal atc tgg gat ttt ctg gaa cag	219		
Phe Leu Ser Pro Glu Val Phe Gln His Ile Trp Asp Phe Leu Glu Gln			
10 15 20 25			
cct ata tgt tca gtt cag ccc att gac ttg aac ttt gtc gat gaa cca	267		
Pro Ile Cys Ser Val Gln Pro Ile Asp Leu Asn Phe Val Asp Glu Pro			
30 35 40			
tca gaa gat ggt gcg aca aac aag att gag att agc alg gac tgt atc	315		
Ser Glu Asp Gly Ala Thr Asn Lys Ile Glu Ile Ser Met Asp Cys Ile			
45 50 55			
ccc atg cag gac tgc gac ctg agt gac ccc atg tgg cca cag tac acg	363		
Arg Met Gln Asp Ser Asp Leu Ser Asp Pro Met Trp Pro Gln Tyr Thr			
60 65 70			
aac ctg ggg ctg ctg aac agc alg gac cag cag att cag aac ggc tcc	411		
Asn Leu Gly Leu Leu Asn Ser Met Asp Gln Gln Ile Gln Asn Gly Ser			
75 80 85			
tgc tcc acc agt ccc tat aac aca gac cac gcg cag aac agc gtc acc	459		
Ser Ser Thr Ser Pro Tyr Asn Thr Asp His Ala Gln Asn Ser Val Thr			
90 95 100 105			
gtg ccc tgc ccc tac gca cag ccc agc tcc acc ttc gat gct ctg tct	507		
Ala Pro Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Ala Leu Ser			
110 115 120			
cca tca ccc gcc atc ccc tcc aac acc gac tac cca gcc ccg cac agt	555		
Pro Ser Pro Ala Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His Ser			
125 130 135			
ttc gac gtg ttc ttc cag cag tgc agc acc gtc aag tgc gcc acc tgc	603		
Phe Asp Val Ser Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp			
140 145 150			
acc tat tcc act gaa ctg aag aaa ctg tac tgc caa att gca aag aca	651		
Thr Tyr Ser Thr Glu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr			
155 160 165			
tgc ccc atc cag atc aag gtc alg acc cca cct cct cag gga gct gtt	699		
Cys Pro Ile Gln Ile Lys Val Met Thr Pro Pro Pro Gln Gly Ala Val			
170 175 180 185			
atc cgc gcc atg cct gtc tac aaa aaa gct gag cac gtc acc gag gtg	747		
Ile Arg Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Glu Val			
190 195 200			
gtg aag cgg tgc ccc aac cat gag ctg agc cgt gaa ttc aac gag gga	795		
Val Lys Arg Cys Pro Asn His Glu Leu Ser Arg Glu Phe Asn Glu Gly			
205 210 215			
cag att gcc cct cct agt cat ttg att cga gta gag ggg aac agc cat	843		
Gln Ile Ala Pro Pro Ser His Leu Ile Arg Val Glu Gly Asn Ser His			
220 225 230			
gcc cag tat gta gaa gat ccc atc aca gga aga cag agt gtg ctg gta	891		
Ala Gln Tyr Val Glu Asp Pro Ile Thr Gly Arg Gln Ser Val Leu Val			
235 240 245			
cct tat gag cca ccc cag gtt gcc act gaa ttc acg aca gtc ttg tac	939		
Pro Tyr Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Val Leu Tyr			
250 255 260 265			
aat ttc atg tgt aac agc agt tgt gtt gga ggg alg aac cgc cgt cca	987		
Asn Phe Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro			
270 275 280			
att tta atc att gtt act ctg gaa acc aga gat ggg caa gtc ctg gcc	1035		
Ile Leu Ile Ile Val Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly			
285 290 295			

cca cgc tgc ttt gag gcc cgg atc tgt gct tgc cca gga aga gac agc 1083
 Arg Arg Cys Phe Glu Ala Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg
 300 305 310

agc ggc gat gaa gat agc atc aga aag cag caa gtt tgc gac agt aca 1131
 Lys Ala Asp Glu Asp Ser Ile Arg Lys Gln Gln Val Ser Asp Ser Thr
 315 320 325

agc aac ggt gat ggt agc aag cgc cgc ttt cgt cag aac aca cat ggt 1179
 Lys Asn Gly Asp Gly Thr Lys Arg Pro Phe Arg Gln Asn Thr His Gly
 330 335 340 345

atc cag atg aca tcc atc aag aaa cga aga tcc cca gat gat gaa ctg 1227
 Ile Gln Met Thr Ser Ile Lys Lys Arg Arg Ser Pro Asp Asp Glu Leu
 350 355 360

tta tsc ita cca gtc agc ggc cgt gag aci tat gaa atg ctg ttg aag 1275
 Leu Tyr Leu Pro Val Arg Gly Arg Glu Thr Tyr Glu Met Leu Leu Lys
 365 370 375

atc aaa gag tcc ctg gaa ctg atg cag tac ctt cct cag cac aca att 1323
 Ile Lys Glu Ser Leu Glu Leu Met Gln Tyr Leu Pro Gln His Thr Ile
 380 385 390

gaa acc tac agc caa cag caa cag cag cag cac cag cac tta ctt cag 1371
 Glu Thr Tyr Arg Gln Gln Gln Gln Gln Gln His Gln His Leu Leu Gln
 395 400 405

aag cat ctg ctt tca gcc tgc ttc aag aat gag ctt gtc gag ccc cgg 1419
 Lys His Leu Leu Ser Ala Cys Phe Arg Asn Glu Leu Val Glu Pro Arg
 410 415 420 425

aga gaa act cca aaa caa tci gac gtc ttc ttt aga cat tcc aag ccc 1467
 Arg Glu Thr Pro Lys Gln Ser Asp Val Phe Phe Arg His Ser Lys Pro
 430 435 440

cca aac cga tca gtc tac cca tagagccca tcctatatt ttaagtgtgt 1518
 Pro Asn Arg Ser Val Tyr Pro
 445

gtgttgtatt tcaatgtgta taigtgagtg tctgtgtgtg tatgtgtgtg cgtgtgtatc 1578

tagccctcat aacagagact tgaagacact ttggtcaga gacccactg ctcaagcca 1636

caagccact agtgaagaa tctttgaag ggaactcaac ctctacaaga aagatgttt 1698

tctgcagatt ttgtatcctt agaccggcca ttgttggtg aggaaccact gtgtttgtct 1758

gtgagcttc ttgttttcc tggaggggg agtcaggig gggagaggg caataagtg 1818

ttatttgaa ccttttctg tcttctctg ttgttttct aaattacaa gggagcctt 1878

tggcaggtc tcaacttaa ggtgtcttt taagaaagg agaaaaagi ttatttgtc 1938

ttgcatang taagtgtag gtactgaga gactcagica gaccttita atctggica 1998

tgaatata ttgcaagtag taaganacga aggtgtcaag tgaactgtg ggcagcagc 2058

tgaatata caaaglaat caacttttg gttgagaggt tctttgtag aactgtcat 2118

atttgtgtc tccctcatg tttaggtaga acatttcta atctgtgta cctgctctg 2178

ccactgtat ttgcactcg ttatgclaa gtitttctg taatgaaac cctggaagc 2238

ctactacaa aaactgttg ttggccccc atagcaggig aactatttt gtcttttaa 2298

tagaagaca aatccacccc agtaatatg cctttacga gtgtttacc attaticaa 2358

gctcaaaata gaattgaag cctctcaca aaatctgga ttatttgtt taattagac 2418

ttctatcct caagcctacc taactaaaa ccagccatct tactgatact gttaagtga 2478

tttagccagg agacttact tttagttag tgagatcaa gcagacgtgt taatatcagc 2538

```

atcctcggac tggaaatcaa agatigaag gtagactac ttctcttctt ttactcaca 2598
agtttagaga atcctcgttt ctctccattt taaaacata tttaagata aagcataaa 2658
gactttaaaa aigtccccc cctccatctt ccacaccca gtcaccagca ctgtatttc 2718
tgacaccaag acaatgattt ctgttatig aggcctgtgc ttigttagat gigtgatit' 2778
aatittcaat aaactttgc atcttggttt aaagaaa 2816

```

```

<210> 3
<211> 448
<212> PRT
<213> Human

<400> 3

```

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

```

Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp Ser Ile
305                      310                      315                      320
Arg Lys Gln Gln Val Ser Asp Ser Thr Lys Asn Gly Asp Gly Thr Lys
                      325                      330                      335
Arg Pro Phe Arg Gln Asn Thr His Gly Ile Gln Met Thr Ser Ile Lys
                      340                      345                      350
Lys Arg Arg Ser Pro Asp Asp Glu Leu Leu Tyr Leu Pro Val Arg Gly
                      355                      360                      365
Arg Glu Thr Tyr Glu Met Leu Leu Lys Ile Lys Glu Ser Leu Glu Leu
                      370                      375                      380
Met Gln Tyr Leu Pro Gln His Thr Ile Glu Thr Tyr Arg Gln Gln Gln
                      385                      390                      395                      400
Gln Gln Gln His Gln His Leu Leu Gln Lys His Leu Leu Ser Ala Cys
                      405                      410                      415
Phe Arg Asn Glu Leu Val Glu Pro Arg Arg Glu Thr Pro Lys Gln Ser
                      420                      425                      430
Asp Val Phe Phe Arg His Ser Lys Pro Pro Asn Arg Ser Val Tyr Pro
                      435                      440                      445

```

<210> 4
 <211> 641
 <212> PRT
 <213> Human

<220>
 <221> DOMAIN
 <222> (1).. (59)
 <223> transactivation domain

<220>
 <221> DNA BIND
 <222> (142).. (321)
 <223> DNA binding domain

<220>
 <221> DOMAIN
 <222> (353).. (397)
 <223> oligomerization domain

<400> 4

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

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

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

<210> 5

<211> 2270

<212> DNA

<213> Human

<220>

<221> CDS

<222> (145).. (2067)

<400> 5

ttc gac gtg tcc ttc cag cag tcc agc acc gcc aag tcc gcc acc tgg Phe Asp Val Ser Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp 140 145 150	603
acg tat tcc act gaa ctg aag aaa ctc tac tgc caa att gca aag aca Thr Tyr Ser Thr Glu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr 155 160 165	651
tgc ccc atc cag atc aag gtg atg acc cca cct cct cag gga gct gtt Cys Pro Ile Gln Ile Lys Val Met Thr Pro Pro Gln Gly Ala Val 170 175 180 185	699
atc cgc gcc atg cct gtc tac aaa aaa gct gag ccc gtc acg gag gtc Ile Arg Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Glu Val 190 195 200	747
gig aag cgg tgc ccc aac cat gag ctg agc cgt gaa ttc aac gag gga Val Lys Arg Cys Pro Asn His Glu Leu Ser Arg Glu Phe Asn Glu Gly 205 210 215	795
cag att gcc cct cct agt cat ttg att cga gta gag gga aac agc cat Gln Ile Ala Pro Pro Ser His Leu Ile Arg Val Glu Gly Asn Ser His 220 225 230	843
gcc cag tat gta gaa gat ccc atc aca gga aga cag agt gtg ctg gta Ala Gln Tyr Val Glu Asp Pro Ile Thr Gly Arg Gln Ser Val Leu Val 235 240 245	881
cct tat gag cca ccc cag gtt ggc act gaa ttc acg aca gtc ttg tac Pro Tyr Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Val Leu Tyr 250 255 260 265	939
aat ttc atg tgt aac agc agt tgt gtt gga gga atg aac cgc cgt cca Asn Phe Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro 270 275 280	987
att tta atc att gtt act ctg gaa acc aga gat gga caa gtc ctg ggc Ile Leu Ile Ile Val Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly 285 290 295	1035
cga cgc tgc ttt gag gcc cgg atc tgt gct tgc cca gga aga gac agg Arg Arg Cys Phe Glu Ala Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg 300 305 310	1083
aag ggc gat gaa gat agc atc aga aag cag caa gtt tcc gac agt aca Lys Ala Asp Glu Asp Ser Ile Arg Lys Gln Gln Val Ser Asp Ser Thr 315 320 325	1131
aag aac ggt gat ggt acg aag cgc cgg ttt cgt cag aac aca cat gct Lys Asn Gly Asp Gly Thr Lys Arg Pro Phe Arg Gln Asn Thr His Gly 330 335 340 345	1179
atc cag atg aca tcc atc aag aaa cga aga tcc cca gat gat gaa ctg Ile Gln Met Thr Ser Ile Lys Lys Arg Arg Ser Pro Asp Asp Glu Leu 350 355 360	1227
tta tac tta cca gtc aag ggc cgt gag act tat gaa atg ctg ttg aag Leu Tyr Leu Pro Val Arg Gly Arg Thr Tyr Glu Met Leu Leu Lys 365 370 375	1275
atc aaa gag tcc ctg gaa ctg atg cag tac ctt cct cag cac aca att Ile Lys Glu Ser Leu Glu Leu Met Gln Tyr Leu Pro Gln His Thr Ile 380 385 390	1323
gaa acg tac agg caa cag caa cag cag cag cac cag cac tta ctt cag Glu Thr Tyr Arg Gln Gln Gln Gln Gln His Gln His Leu Leu Gln 395 400 405	1371
aaa cag acc tca ata cag tct cca tct tca tat ggt aac agc tcc cca Lys Gln Thr Ser Ile Gln Ser Pro Ser Ser Tyr Gly Asn Ser Ser Pro 410 415 420 425	1419
cct ctg aac aaa atg aac agc atg aac aag ctg cct tct gtg agc cag	1467

```

Pro Leu Asn Lys Met Asn Ser Met Asn Lys Leu Pro Ser Val Ser Gln
430                               435                               440

cct atc aac cct cag cag cgc aac ggc ctr cct cct aca acc att cct 1515
Leu Ile Asn Pro Gln Gln Arg Asn Ala Leu Thr Pro Thr Thr Ile Pro
445                               450                               455

gat ggc atg gga gcc aac att ccc atg atg ggc acc cxc atg cxc atg 1563
Asp Gly Met Gly Ala Asn Ile Pro Met Met Gly Thr His Met Pro Met
460                               465                               470

gct ggc gac atg aat gga ctc agc xxx acc cag gca ctc cct ccc cxa 1611
Ala Gly Asp Met Asn Gly Leu Ser Pro Thr Gln Ala Leu Pro Pro Pro
475                               480                               485

cic tcc atg cca tcc acc tcc cax tgc cxa ccc cca cct ccg tat ccc 1659
Leu Ser Met Pro Ser Thr Ser His Cys Thr Pro Pro Pro Pro Tyr Pro
490                               495                               500

aca gat tgc agc att gtc agt ttc tta ggc agg ttc ggc tgt tca tca 1707
Thr Asp Cys Ser Ile Val Ser Phe Leu Ala Arg Leu Gly Cys Ser Ser
510                               515                               520

tgt ctc gac tat ttc acc acc cag ggc ctc acc acc atc tat cag att 1755
Cys Leu Asp Tyr Phe Thr Thr Gln Gly Leu Thr Thr Ile Tyr Gln Ile
525                               530                               535

gag cat tcc tcc atg gat gat ctc ggc xgt ctc aca atc cct gag caa 1803
Glu His Tyr Ser Met Asp Asp Leu Ala Ser Leu Lys Ile Pro Glu Gln
540                               545                               550

ttt cca cat ggc atc tgg aag ggc atc ctc gar cax cgg cag ctc cxc 1851
Phe Arg His Ala Ile Trp Lys Gly Ile Leu Asp His Arg Gln Leu His
555                               560                               565

gaa ttc tcc tcc cct tct cat ctc ctc ggc acc cca agc agt ggc tct 1899
Glu Phe Ser Ser Pro Ser His Leu Leu Arg Thr Pro Ser Ser Ala Ser
570                               575                               580

aca gtc agt gtc ggc tcc agt gag acc cgg ggt gag cgt gtt att gat 1947
Thr Val Ser Val Gly Ser Ser Glu Thr Arg Gly Glu Arg Val Ile Asp
590                               595                               600

gct gtc cga ttc acc ctc cgc cag acc atc tct ttc cca ccc cga gat 1995
Ala Val Arg Phe Thr Leu Arg Gln Thr Ile Ser Phe Pro Pro Arg Asp
605                               610                               615

gag tgg aat gac ttc aac ttt gac atg gat gct cgc cgc aat aag caa 2043
Glu Trp Asn Asp Phe Asn Phe Asp Met Asp Ala Arg Arg Asn Lys Gln
620                               625                               630

cag cgc atc aaa gag gag ggc gag tggatcacc catgagact cttctatc 2097
Gln Arg Ile Lys Glu Glu Gly Glu
635                               640

ctctccaaac tggcagcccc ctaaaagcac tctctcttaa tctccaaagc ctctcccta 2157
gtctctcccc tctctcttgt ctgattcttt agggcagga gaagaaagag gctaccctt 2217
acctaccatc tggcctggca tctctcttg attcggctt tggccttca aac 2270

```

<210> 6

<211> 641

<212> PRT

<213> Human

<400> 6

```

Met Ser Gln Ser Thr Gln Thr Asn Glu Phe Leu Ser Pro Glu Val Phe
1           5           10           15

```

```

Gln His Ile Trp Asp Phe Leu Glu Gln Pro Ile Cys Ser Val Gln Pro
20           25           30

```

Ile Asp Leu Asn Phe Val Asp Glu Pro Ser Glu Asp Gly Ala Thr Asn
 35 40 45
 Lys Ile Glu Ile Ser Met Asp Cys Ile Arg Met Gln Asp Ser Asp Leu
 50 55 60
 Ser Asp Pro Met Trp Pro Gln Tyr Thr Asn Leu Gly Leu Leu Asn Ser
 65 70 75 80
 Met Asp Gln Gln Ile Gln Asn Gly Ser Ser Ser Thr Ser Pro Tyr Asn
 85 90 95
 Thr Asp His Ala Gln Asn Ser Val Thr Ala Pro Ser Pro Tyr Ala Gln
 100 105 110
 Pro Ser Ser Thr Phe Asp Ala Leu Ser Pro Ser Pro Ala Ile Pro Ser
 115 120 125
 Asn Thr Asp Tyr Pro Gly Pro His Ser Phe Asp Val Ser Phe Gln Gln
 130 135 140
 Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Thr Glu Leu Lys
 145 150 155 160
 Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln Ile Lys Val
 165 170 175
 Met Thr Pro Pro Pro Gln Gly Ala Val Ile Arg Ala Met Pro Val Tyr
 180 185 190
 Lys Lys Ala Glu His Val Thr Glu Val Val Lys Arg Cys Pro Asn His
 195 200 205
 Glu Leu Ser Arg Glu Phe Asn Glu Gly Gln Ile Ala Pro Pro Ser His
 210 215 220
 Leu Ile Arg Val Glu Gly Asn Ser His Ala Gln Tyr Val Glu Asp Pro
 225 230 235 240
 Ile Thr Gly Arg Gln Ser Val Leu Val Pro Tyr Glu Pro Pro Gln Val
 245 250 255
 Gly Thr Glu Phe Thr Thr Val Leu Tyr Asn Phe Met Cys Asn Ser Ser
 260 265 270
 Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Val Thr Leu
 275 280 285
 Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Cys Phe Glu Ala Arg
 290 295 300
 Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp Ser Ile
 305 310 315 320
 Arg Lys Gln Gln Val Ser Asp Ser Thr Lys Asn Gly Asp Gly Thr Lys
 325 330 335
 Arg Pro Phe Arg Gln Asn Thr His Gly Ile Gln Met Thr Ser Ile Lys
 340 345 350
 Lys Arg Arg Ser Pro Asp Asp Glu Leu Leu Tyr Leu Pro Val Arg Gly
 355 360 365
 Arg Glu Thr Tyr Glu Met Leu Leu Lys Ile Lys Glu Ser Leu Glu Leu
 370 375 380
 Met Gln Tyr Leu Pro Gln His Thr Ile Glu Thr Tyr Arg Gln Gln Gln
 385 390 395 400
 Gln Gln Gln His Gln His Leu Leu Gln Lys Gln Thr Ser Ile Gln Ser
 405 410 415
 Pro Ser Ser Tyr Gly Asn Ser Ser Pro Pro Leu Asn Lys Met Asn Ser
 420 425 430

Met Asn Lys Leu Pro Ser Val Ser Gln Leu Ile Asn Pro Gln Gln Arg
435 440 445

Asn Ala Leu Thr Pro Thr Thr Ile Pro Asn Gly Met Gly Ala Asn Ile
450 455 460

Pro Met Met Gly Thr His Met Pro Met Ala Gly Asp Met Asn Gly Leu
465 470 475 480

Ser Pro Thr Gln Ala Leu Pro Pro Pro Leu Ser Met Pro Ser Thr Ser
485 490 495

His Cys Thr Pro Pro Pro Tyr Pro Thr Asn Cys Ser Ile Val Ser
500 505 510

Phe Leu Ala Arg Leu Gly Cys Ser Ser Cys Leu Asn Tyr Phe Thr Thr
515 520 525

Gln Gly Leu Thr Thr Ile Tyr Gln Ile Glu His Tyr Ser Met Asp Asp
530 535 540

Leu Ala Ser Leu Lys Ile Pro Glu Gln Phe Arg His Ala Ile Trp Lys
545 550 555 560

Gly Ile Leu Asp His Arg Gln Leu His Glu Phe Ser Ser Pro Ser His
565 570 575

Leu Leu Arg Thr Pro Ser Ser Ala Ser Thr Val Ser Val Gly Ser Ser
580 585 590

Glu Thr Arg Gly Glu Arg Val Ile Asp Ala Val Arg Phe Thr Leu Arg
595 600 605

Gln Thr Ile Ser Phe Pro Pro Arg Asp Glu Trp Asn Asp Phe Asn Phe
610 615 620

Asp Met Asp Ala Arg Arg Asn Lys Gln Gln Arg Ile Lys Glu Glu Gly
625 630 635 640

Glu

<210> 7

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p73-F1 sense primer

<400> 7

taagtgacag taaagacag tgcctc 27

<210> 8

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p73-R1 antisense primer

<400> 8

tgtgtacagt tgcctcaagt ggaagtaag 29

<210> 9

<211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:p73-F2 sense primer

<400> 9
 taagtatact acgaagtga cgtgaaggg 29

<210> 10
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:p73-R2 antisense primer

<400> 10
 atgaactaag acgtacgaag tcaacgtat 29

<210> 11
 <211> 30
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:HA-labeled expression construct

<400> 11
 atgtatccat atgatgttcc agattatgct 30

<210> 12
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:p51-F1 sense primer

<400> 12
 aaagaaggtt attacgaag 20

<210> 13
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:p51-R1 antisense primer

<400> 13
 cgogtgggtat gtgtatagg 20

<210> 14
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-F2 sense primer

<400> 14

catggaccag cagattcaga 20

<210> 15

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-R2 antisense primer

<400> 15

catcaccttg atctggatg 19

<210> 16

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-F3 sense primer

<400> 16

ccacctggac gttaccact 20

<210> 17

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-R3, antisense primer

<400> 17

tggctcataa ggtaccag 18

<210> 18

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-F4 sense primer

<400> 18

catgagctga gcogtgaat 19

<210> 19

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:p51-R4 antisense primer

<400> 19
tatcttcac ccgcttcctg 20

<210> 20
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:p51-F5 sense primer

<400> 20
atgaacccgc gtccaatt 18

<210> 21
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:p51-R5 antisense primer

<400> 21
gtgctgagga aggtactgca 20

<210> 22
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:p51-F6 sense primer

<400> 22
tgaagatcaa agagtcctg 20

<210> 23
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:p51-R6 antisense primer

<400> 23
ctagtggcct tgtgcccttg 20

SEQUENCE LISTING

[0340]

<110> Ikawa, Yoji
Otsuka Pharmaceutical Co. Ltd.

<120> Human p51 gene and its product

<130> P99-16

<140>

<141>

<150> JP P1998-100467

<151> 1998-03-27

<160> 23

<170> PatentIn ver.2.0

<210> 1

<211> 448

<212> PRT

<213> Human

<220>

<221> DOMAIN

<222> (1)..(59)

<223> transactivation domain

<220>

<221> DNA BIND

<222> (142)..(321)

<223> DNA binding domain

<220>

<221> DOMAIN

<222> (353)..(397)

<223> oligomerization domain

<400> 1

```

Met Ser Gln Ser Thr Gln Thr Asn Glu Phe Leu Ser Pro Glu Val Phe
 1      5      10      15
Gln His Ile Trp Asp Phe Leu Glu Gln Pro Ile Cys Ser Val Gln Pro
 20      25      30
Ile Asp Leu Asn Phe Val Asp Glu Pro Ser Glu Asp Gly Ala Thr Asn
 35      40      45
Lys Ile Glu Ile Ser Met Asp Cys Ile Arg Met Gln Asp Ser Asp Leu
 50      55      60
Ser Asp Pro Met Trp Pro Gln Tyr Thr Asn Leu Gly Leu Leu Asn Ser
 65      70      75      80
Met Asp Gln Gln Ile Gln Asn Gly Ser Ser Ser Thr Ser Pro Tyr Asn
 85      90      95
Thr Asp His Ala Gln Asn Ser Val Thr Ala Pro Ser Pro Tyr Ala Gln
100      105      110
Pro Ser Ser Thr Phe Asp Ala Leu Ser Pro Ser Pro Ala Ile Pro Ser
115      120      125
Asn Thr Asp Tyr Pro Gly Pro His Ser Phe Asp Val Ser Phe Gln Gln
130      135      140
Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Thr Glu Leu Lys
145      150      155      160
Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln Ile Lys Val
165      170      175
Met Thr Pro Pro Pro Gln Gly Ala Val Ile Arg Ala Met Pro Val Tyr
180      185      190
Lys Lys Ala Glu His Val Thr Glu Val Val Lys Arg Cys Pro Asn His
195      200      205
Glu Leu Ser Arg Glu Phe Asn Glu Gly Gln Ile Ala Pro Pro Ser His
210      215      220
Leu Ile Arg Val Glu Gly Asn Ser His Ala Gln Tyr Val Glu Asp Pro

```

```

225      230      235      240
Ile Thr Gly Arg Gln Ser Val Leu Val Pro Tyr Glu Pro Pro Gln Val
Gly Thr Glu Phe 245 Thr Val Leu Tyr 250 Phe Met Cys Asn 255 Ser
Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Val Thr Leu
Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Cys Phe Glu Ala Arg
Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp Ser Ile
Arg Lys Gln Gln Val Ser Asp Ser Thr Lys Asn Gly Asp Gly Thr Lys
Arg Pro Phe Arg Gln Asn Thr His Gly Ile Gln Met Thr Ser Ile Lys
Lys Arg Arg Ser Pro Asp Asp Glu Leu Tyr Leu Pro Val Arg Gly
Arg Glu Thr Tyr Glu Met Leu Leu Lys Ile Lys Glu Ser Leu Glu Leu
Met Gln Tyr Leu Pro Gln His Thr Ile Glu Thr Tyr Arg Gln Gln Gln
Gln Gln Gln His Gln His Leu Leu Gln Lys His Leu Leu Ser Ala Cys
Phe Arg Asn Glu Leu Val Glu Pro Arg Arg Glu Thr Pro Lys Gln Ser
Asp Val Phe Arg His Ser Lys Pro Pro Asn Arg Ser Val Tyr Pro

```

<210> 2

<211> 2818

<212> DNA

<213> Human

<220>

<221> CDS

<222> (145)..(1488)

<220>

<221> polyA_signal

<222> (2786)..(2791)

<400> 2

```

tcgttgatat caaagacagt tgaaggaaat gaattttgaa acttcacggt gtgccaccct 60
acagtactgc cctgaccctt acatccagcg tttcgtagaa acccagctca ttctcttgg 120
aaagaaagt attaccgatc cacc atg tcc cag agc aca cag aca aat gaa 171
Met Ser Gln Ser Thr Gln Thr Asn Glu
1 5
ttc ctc agt cca gag gtt ttc cag cat atc tgg gat ttt ctg gaa cag 219
Phe Leu Ser Pro Glu Val Phe Gln His Ile Trp Asp Phe Leu Glu Gln
10 15 20 25
cct ata tgt tca gtt cag ccc att gac ttg aac ttt gtg gat gaa cca 267
Pro Ile Cys Ser Val Gln Pro Ile Asp Leu Asn Phe Val Asp Glu Pro
30 35 40
tca gaa gat ggt gcg aca aac aag att gag att agc atg gac tgt atc 315
Ser Glu Asp Gly Ala Thr Asn Lys Ile Glu Ile Ser Met Asp Cys Ile
45 50 55
cgc atg cag gac tcg gac ctg agt gac ccc atg tgg cca cag tac acg 363
Arg Met Gln Asp Ser Asp Leu Ser Asp Pro Met Trp Pro Gln Tyr Thr
60 65 70
aac ctg ggg ctc ctg aac agc atg gac cag cag att cag aac ggc tcc 411
Asn Leu Gly Leu Leu Asn Ser Met Asp Gln Gln Ile Gln Asn Gly Ser
75 80 85
tcg tcc acc agt ccc tat aac aca gac cac gcg cag aac agc gtc acg 459
Ser Ser Thr Ser Pro Tyr Asn Thr Asp His Ala Gln Asn Ser Val Thr
90 95 100 105
gcg ecc tcg ccc tac gca cag ccc agc tcc acc ttc gat gct ctc tct 507
Ala Pro Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Ala Leu Ser
110 115 120

```

cca tca ccc gcc atc ccc tcc aac acc gac tac cca ggc ccg cac agt	555
Pro Ser Pro Ala Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His Ser	
125 130 135	
ttc gac gtg tcc ttc cag cag tcc acc gcc aag tcc gcc acc tgg	603
Phe Asp Val Ser Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp	
140 145 150	
acg tat tcc act gaa ctg aag aaa ctc tac tgc caa att gca aag aca	651
Thr Tyr Ser Thr Glu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr	
155 160 165	
tgc ccc atc cag atc aag gtg atg acc cca cct cct cag gga gct gtt	699
Cys Pro Ile Gln Ile Lys Val Met Thr Pro Pro Gln Gly Ala Val	
170 175 180 185	
atc cgc gcc atg cct gtc tac aaa aaa gct gag cac gtc acg gag gtg	747
Ile Arg Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Glu Val	
190 195 200	
gtg aag cgg tgc ccc aac cat gag ctg agc cgt gaa ttc aac gag gga	795
Val Lys Arg Cys Pro Asn His Glu Leu Ser Arg Glu Phe Asn Glu Gly	
205 210 215	
cag att gcc cct cct agt cat ttg att cga gta gag ggg aac agc cat	843
Gln Ile Ala Pro Pro Ser His Leu Ile Arg Val Glu Gly Asn Ser His	
220 225 230	
gcc cag tat gta gaa gat ccc atc aca gga aga cag agt gtg ctg gta	891
Ala Gln Tyr Val Glu Asp Pro Ile Thr Gly Arg Gln Ser Val Leu Val	
235 240 245	
cct tat gag cca ccc cag gtt ggc act gaa ttc acg aca gtc ttg tac	939
Pro Tyr Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Val Leu Tyr	
250 255 260 265	
aat ttc atg tgt aac agc agt tgt gtt gga ggg atg aac cgc cgt cca	987
Asn Phe Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro	
270 275 280	
att tta atc att gtt act ctg gaa acc aga gat ggg caa gtc ctg ggc	1035
Ile Leu Ile Ile Val Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly	
285 290 295	
cga cgc tgc ttt gag gcc cgg atc tgt gct tgc cca gga aga gac agg	1083
Arg Arg Cys Phe Glu Ala Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg	
300 305 310	
aag gcg gat gaa gat agc atc aga aag cag caa gtt tcc gac agt aca	1131
Lys Ala Asp Glu Asp Ser Ile Arg Lys Gln Gln Val Ser Asp Ser Thr	
315 320 325	
aag aac ggt gat ggt acg aag cgc ccg ttt cgt cag aac aca cat ggt	1179
Lys Asn Gly Asp Gly Thr Lys Arg Pro Phe Arg Gln Asn Thr His Gly	
330 335 340 345	
atc cag atg aca tcc atc aag aaa cga aga tcc cca gat gat gaa ctg	1227
Ile Gln Met Thr Ser Ile Lys Lys Arg Arg Ser Pro Asp Asp Glu Leu	
350 355 360	
tta tac tta cca gtg agg ggc cgt gag act tat gaa atg ctg ttg aag	1275
Leu Tyr Leu Pro Val Arg Gly Arg Glu Thr Tyr Glu Met Leu Leu Lys	
365 370 375	
atc aaa gag tcc ctg gaa ctc atg cag tac ctt cct cag cac aca att	1323
Ile Lys Glu Ser Leu Glu Leu Met Gln Tyr Leu Pro Gln His Thr Ile	
380 385 390	
gaa acg tac agg caa cag caa cag cag cac cag cac tta ctt cag	1371
Glu Thr Tyr Arg Gln Gln Gln Gln Gln Gln His Gln His Leu Leu Gln	
395 400 405	
aaa cat ctc ctt tca gcc tgc ttc agg aat gag ctt gtg gag ccc cgg	1419
Lys His Leu Leu Ser Ala Cys Phe Arg Asn Glu Leu Val Glu Pro Arg	
410 415 420 425	
aga gaa act cca aaa caa tct gac gtc ttc ttt aga cat tcc aag ccc	1467
Arg Glu Thr Pro Lys Gln Ser Asp Val Phe Arg His Ser Lys Pro	
430 435 440	
cca aac cga tca gtg tac cca tagagcccta tctctatatatt ttaagtgtgt	1518
Pro Asn Arg Ser Val Tyr Pro	
445	
gtgtgtgtatt tccatgtgta tatgtgagtg tgtgtgtgtg tatgtgtgtg cgtgtgtatc	1578
tagccctcat aaacaggact tgaagacact ttggctcaga gacccaactg ctcaaaggca	1638
caaaagccact agtgagagaa tcttttgaag ggactcaaac ctttacaaga aaggatgttt	1698
tctgcagatt ttgtatcctt agaccggcca ttgggtgggtg aggaaccact gtgtttgtct	1758
gtgagccttc tgttgtttcc tgggagggag gggtcaggtg gggaaagggg carttaagatg	1818

tttattggaa	cccttttctg	tcttcttctg	ttgtttttct	aaaattcaca	gggaagcttt	1878
tgagcaggtc	tcaaacttaa	gatgtctttt	taagaaaagg	agaaaaaagt	tgttattgtc	1938
tgtagcataag	taagttgtag	gtgactgaga	gactcagtc	gacctttta	atgctgggtc	1998
tgtaataata	ttgcaagtag	taagaaacga	aggtgtcaag	tgtactgctg	ggcagcgagg	2058
tgatcattac	caaaagtaat	caactttgtg	gggtggagagt	tcittgtag	aacttgcatt	2118
atttgtgtcc	tcccttcacg	tgtaggtaga	acattttcta	atgctgtgta	cctgcctctg	2178
ccactgfatg	ttggcatctg	ttatgctaaa	gtttttcttg	tacatgaaac	cctggaagac	2238
ctactacaaa	aaaactgttg	tttggcccc	atagcaggtg	aactcatttt	gtgcttttaa	2298
tagaaaagaca	aatccacccc	agtaatatgg	cccttacgta	gttgtttacc	attattcaaa	2358
gcicaaaata	gaatttgaag	ccctctcaca	aaatctgtga	ttaatttgct	taattagagc	2418
ttctatcctt	caagcctacc	taccataaaa	ccagccatat	tactgatact	gttcagtgca	2478
tttagccagg	agacttacgt	tttgagtaag	tgagatccaa	gcagacgtgt	taaaatcagc	2538
actcctggac	tggaaattaa	agattgaaaag	ggtagactac	ttttcttttt	tttactcaaa	2598
agtttagaga	atctctgttt	ctttccattt	taaaaacata	ttttaagata	atagcataaa	2658
gacttttaaa	atgttctctc	cctccatctt	cccacacca	gtcaccagca	ctgtattttc	2718
tgtcaccaag	acaatgattt	cttgtttattg	aggctgtttg	ttttgtggat	gtgtgatttt	2778
aattttcaat	aaacttttgc	atcttgggtt	aaaagaaa			2816

<210> 3

<211> 448

<212> PRT

<213> Human

<400> 3

Met Ser Gln Ser Thr Gln Thr Asn Glu Phe Leu Ser Pro Glu Val Phe
 1 5 10 15
 Gln His Ile Trp Asp Phe Leu Glu Gln Pro Ile Cys Ser Val Gln Pro
 20 25 30
 Ile Asp Leu Asn Phe Val Asp Glu Pro Ser Glu Asp Gly Ala Thr Asn
 35 40 45
 Lys Ile Glu Ile Ser Met Asp Cys Ile Arg Met Gln Asp Ser Asp Leu
 50 55 60
 Ser Asp Pro Met Trp Pro Gln Tyr Thr Asn Leu Gly Leu Leu Asn Ser
 65 70 75 80
 Met Asp Gln Gln Ile Gln Asn Gly Ser Ser Ser Thr Ser Pro Tyr Asn
 85 90 95
 Thr Asp His Ala Gln Asn Ser Val Thr Ala Pro Ser Pro Tyr Ala Gln
 100 105 110
 Pro Ser Ser Thr Phe Asp Ala Leu Ser Pro Ser Pro Ala Ile Pro Ser
 115 120 125
 Asn Thr Asp Tyr Pro Gly Pro His Ser Phe Asp Val Ser Phe Gln Gln
 130 135 140
 Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Thr Glu Leu Lys
 145 150 155 160
 Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln Ile Lys Val
 165 170 175
 Met Thr Pro Pro Pro Gln Gly Ala Val Ile Arg Ala Met Pro Val Tyr
 180 185 190
 Lys Lys Ala Glu His Val Thr Glu Val Val Lys Arg Cys Pro Asn His
 195 200 205
 Glu Leu Ser Arg Glu Phe Asn Glu Gly Gln Ile Ala Pro Pro Ser His
 210 215 220
 Leu Ile Arg Val Glu Gly Asn Ser His Ala Gln Tyr Val Glu Asp Pro
 225 230 235 240
 Ile Thr Gly Arg Gln Ser Val Leu Val Pro Tyr Glu Pro Pro Gln Val
 245 250 255
 Gly Thr Glu Phe Thr Thr Val Leu Tyr Asn Phe Met Cys Asn Ser Ser
 260 265 270
 Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Val Thr Leu
 275 280 285
 Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Cys Phe Glu Ala Arg
 290 295 300
 Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp Ser Ile
 305 310 315 320
 Arg Lys Gln Gln Val Ser Asp Ser Thr Lys Asn Gly Asp Gly Thr Lys
 325 330 335
 Arg Pro Phe Arg Gln Asn Thr His Gly Ile Gln Met Thr Ser Ile Lys
 340 345 350

Lys Arg Arg Ser Pro Asp Asp Glu Leu Leu Tyr Leu Pro Val Arg Gly
 355 360 365
 Arg Glu Thr Tyr Glu Met Leu Leu Lys Ile Lys Glu Ser Leu Glu Leu
 370 375 380
 Met Gln Tyr Leu Pro Gln His Thr Ile Glu Thr Tyr Arg Gln Gln Gln
 385 390 395 400
 Gln Gln Gln His Gln His Leu Leu Gln Lys His Leu Leu Ser Ala Cys
 405 410 415
 Phe Arg Asn Glu Leu Val Glu Pro Arg Arg Glu Thr Pro Lys Gln Ser
 420 425 430
 Asp Val Phe Phe Arg His Ser Lys Pro Pro Asn Arg Ser Val Tyr Pro
 435 440 445

<210> 4

<211> 641

<212> PRT

<213> Human

<220>

<221> DOMAIN

<222> (1)..(59)

<223> transactivation domain

<220>

<221> DNA_BIND

<222> (142)..(321)

<223> DNA binding domain

<220>

<221> DOMAIN

<222> (353)..(397)

<223> oligomerization domain

<400> 4

```

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

```

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

<210> 5

<211> 2270

<212> DNA

<213> Human

<220>

<221> CDS

<222> (145)..(2067)

<400> 5

tcgttgatat	caaagacagt	tgaaggaaat	gaattttgaa	acttcacggt	gtgccaccct	60
acagtactgc	cttgaccctt	acatccagcg	tttcgtagaa	acccagctca	ttctcttgg	120
aaagaaagtt	attaccgatc	cacc atg tcc	cag agc aca	cag aca aat	gaa	171
		Met Ser Gln Ser Thr	Gln Thr Asn Glu			
		1	5			
ttc ctc agt	cca gag gtt	ttc cag cat	atc tgg gat	ttt ctg gaa	cag	219
Phe Leu Ser	Pro Glu Val	Phe Gln His	Ile Trp Asp	Phe Leu Glu	Gln	
10	15	20	25			
cct ata tgt	tca gtt cag	ccc att gac	ttg aac ttt	gtg gat gaa	cca	267
Pro Ile Cys	Ser Val Gln	Pro Ile Asp	Leu Asn Phe	Val Asp Glu	Pro	


```

395          400          405
aaa cag acc tca ata cag tct cca tct tca tat ggt aac agc tcc cca 1419
Lys Gln Thr Ser Ile Gln Ser Pro Ser Ser Tyr Gly Asn Ser Ser Pro
410
cct ctg aac aaa atg aac agc atg aac aag ctg cct tct gtg agc cag 1467
Pro Leu Asn Lys Met Asn Ser Met Asn Lys Leu Pro Ser Val Ser Gln
430
cct atc aac cct cag cag cgc aac gcc ctc act cct aca acc att cct 1515
Leu Ile Asn Pro Gln Gln Arg Asn Ala Leu Thr Pro Thr Thr Ile Pro
445
gat ggc atg gga gcc aac att ccc atg atg ggc acc cac atg cca atg 1563
Asp Gly Met Gly Ala Asn Ile Pro Met Met Gly Thr His Met Pro Met
460
gct gga gac atg aat gga ctc agc ccc acc cag gca ctc cct ccc cca 1611
Ala Gly Asp Met Asn Gly Leu Ser Pro Thr Gln Ala Leu Pro Pro Pro
475
ctc tcc atg cca tcc acc tcc cac tgc aca ccc cca cct ccg tat ccc 1659
Leu Ser Met Pro Ser Thr Ser His Cys Thr Pro Pro Pro Tyr Pro
490
aca gat tgc agc att gtc agt ttc tta gcc agg ttg ggc tgt tca tca 1707
Thr Asp Cys Ser Ile Val Ser Phe Leu Ala Arg Leu Gly Cys Ser Ser
510
tgt ctg gac tat ttc acg acc cag ggg ctg acc acc atc tat cag att 1755
Cys Leu Asp Tyr Phe Thr Thr Gln Gly Leu Thr Thr Ile Tyr Gln Ile
525
gag cat tac tcc atg gat gat ctg gca agt ctg aaa atc cct gag caa 1803
Glu His Tyr Ser Met Asp Asp Leu Ala Ser Leu Lys Ile Pro Glu Gln
540
ttt cga cat gcc atc tgg aag gcc atc ctg gac cac cgg cag ctc cac 1851
Phe Arg His Ala Ile Trp Lys Gly Ile Leu Asp His Arg Gln Leu His
555
gaa ttc tcc tcc cct tct cat ctc ctg cgg acc cca agc agt gcc tct 1899
Glu Phe Ser Ser Pro Ser His Leu Leu Arg Thr Pro Ser Ser Ala Ser
570
aca gtc agt gtg gcc tcc agt gag acc cgg ggt gag cgt gtt att gat 1947
Thr Val Ser Val Gly Ser Ser Glu Thr Arg Gly Glu Arg Val Ile Asp
590
gct gtg cga ttc acc ctc cgc cag acc atc tct ttc cca ccc cga gat 1995
Ala Val Arg Phe Thr Leu Arg Gln Thr Ile Ser Phe Pro Pro Arg Asp
605
gag tgg aat gac ttc aac ttt gac atg gat gct cgc cgc aat aag caa 2043
Glu Trp Asn Asp Phe Asn Phe Asp Met Asp Ala Arg Arg Asn Lys Gln
620
cag cgc atc aaa gag gag ggg gag tgagcctcac catgtgagct cttcctatcc 2097
Gln Arg Ile Lys Glu Glu Gly Glu
635
ctctcctaac tgccagcccc ctaaaagcac tcttgcttaa tcttcaaagc cttctccta 2157
gctctcccc ttcctcttgt ctgatttctt aggggaagga gaagtaagag gctacctctt 2217
acctaacatc tgacctggca tctaattctg attctggcct taagccttca aaa 2270

```

<210> 6
 <211> 641
 <212> PRT
 <213> Human

<400> 6

```

Met Ser Gln Ser Thr Gln Thr Asn Glu Phe Leu Ser Pro Glu Val Phe
1          5          10          15
Gln His Ile Trp Asp Phe Leu Glu Gln Pro Ile Cys Ser Val Gln Pro
20          25          30
Ile Asp Leu Asn Phe Val Asp Glu Pro Ser Glu Asp Gly Ala Thr Asn
35          40          45
Lys Ile Glu Ile Ser Met Asp Cys Ile Arg Met Gln Asp Ser Asp Leu
50          55          60
Ser Asp Pro Met Trp Pro Gln Tyr Thr Asn Leu Gly Leu Leu Asn Ser
65          70          75          80
Met Asp Gln Gln Ile Gln Asn Gly Ser Ser Ser Thr Ser Pro Tyr Asn
85          90          95

```

Thr	Asp	His	Ala	Gln	Asn	Ser	Val	Thr	Ala	Pro	Ser	Pro	Tyr	Ala	Gln
			100					105					110		
Pro	Ser	Ser	Thr	Phe	Asp	Ala	Leu	Ser	Pro	Ser	Pro	Ala	Ile	Pro	Ser
		115					120					125			
Asn	Thr	Asp	Tyr	Pro	Gly	Pro	His	Ser	Phe	Asp	Val	Ser	Phe	Gln	Gln
	130				135						140				
Ser	Ser	Thr	Ala	Lys	Ser	Ala	Thr	Trp	Thr	Tyr	Ser	Thr	Glu	Leu	Lys
145					150					155					160
Lys	Leu	Tyr	Cys	Gln	Ile	Ala	Lys	Thr	Cys	Pro	Ile	Gln	Ile	Lys	Val
			165						170					175	
Met	Thr	Pro	Pro	Pro	Gln	Gly	Ala	Val	Ile	Arg	Ala	Met	Pro	Val	Tyr
			180					185					190		
Lys	Lys	Ala	Glu	His	Val	Thr	Glu	Val	Val	Lys	Arg	Cys	Pro	Asn	His
		195					200					205			
Glu	Leu	Ser	Arg	Glu	Phe	Asn	Glu	Gly	Gln	Ile	Ala	Pro	Pro	Ser	His
	210					215					220				
Leu	Ile	Arg	Val	Glu	Gly	Asn	Ser	His	Ala	Gln	Tyr	Val	Glu	Asp	Pro
225					230					235					240
Ile	Thr	Gly	Arg	Gln	Ser	Val	Leu	Val	Pro	Tyr	Glu	Pro	Pro	Gln	Val
			245						250					255	
Gly	Thr	Glu	Phe	Thr	Thr	Val	Leu	Tyr	Asn	Phe	Met	Cys	Asn	Ser	Ser
			260					265					270		
Cys	Val	Gly	Gly	Met	Asn	Arg	Arg	Pro	Ile	Leu	Ile	Ile	Val	Thr	Leu
	275						280					285			
Glu	Thr	Arg	Asp	Gly	Gln	Val	Leu	Gly	Arg	Arg	Cys	Phe	Glu	Ala	Arg
	290				295						300				
Ile	Cys	Ala	Cys	Pro	Gly	Arg	Asp	Arg	Lys	Ala	Asp	Glu	Asp	Ser	Ile
305					310					315					320
Arg	Lys	Gln	Gln	Val	Ser	Asp	Ser	Thr	Lys	Asn	Gly	Asp	Gly	Thr	Lys
			325						330					335	
Arg	Pro	Phe	Arg	Gln	Asn	Thr	His	Gly	Ile	Gln	Met	Thr	Ser	Ile	Lys
			340					345					350		
Lys	Arg	Arg	Ser	Pro	Asp	Asp	Glu	Leu	Leu	Tyr	Leu	Pro	Val	Arg	Gly
		355					360					365			
Arg	Glu	Thr	Tyr	Glu	Met	Leu	Leu	Lys	Ile	Lys	Glu	Ser	Leu	Glu	Leu
	370					375					380				
Met	Gln	Tyr	Leu	Pro	Gln	His	Thr	Ile	Glu	Thr	Tyr	Arg	Gln	Gln	Gln
385					390					395					400
Gln	Gln	Gln	His	Gln	His	Leu	Leu	Gln	Lys	Gln	Thr	Ser	Ile	Gln	Ser
			405						410					415	
Pro	Ser	Ser	Tyr	Gly	Asn	Ser	Ser	Pro	Pro	Leu	Asn	Lys	Met	Asn	Ser
			420					425					430		
Met	Asn	Lys	Leu	Pro	Ser	Val	Ser	Gln	Leu	Ile	Asn	Pro	Gln	Gln	Arg
		435					440					445			
Asn	Ala	Leu	Thr	Pro	Thr	Thr	Thr	Ile	Pro	Asp	Gly	Met	Gly	Ala	Asn
	450					455						460			
Pro	Met	Met	Gly	Thr	His	Met	Pro	Met	Ala	Gly	Asp	Met	Asn	Gly	Leu
465					470					475					480
Ser	Pro	Thr	Gln	Ala	Leu	Pro	Pro	Pro	Leu	Ser	Met	Pro	Ser	Thr	Ser
			485						490					495	
His	Cys	Thr	Pro	Pro	Pro	Pro	Tyr	Pro	Thr	Asp	Cys	Ser	Ile	Val	Ser
			500					505					510		
Phe	Leu	Ala	Arg	Leu	Gly	Cys	Ser	Ser	Cys	Leu	Asp	Tyr	Phe	Thr	Thr
		515					520					525			
Gln	Gly	Leu	Thr	Thr	Ile	Tyr	Gln	Ile	Glu	His	Tyr	Ser	Met	Asp	Asp
	530					535					540				
Leu	Ala	Ser	Leu	Lys	Ile	Pro	Glu	Gln	Phe	Arg	His	Ala	Ile	Trp	Lys
545					550					555					560
Gly	Ile	Leu	Asp	His	Arg	Gln	Leu	His	Glu	Phe	Ser	Ser	Pro	Ser	His
			565						570					575	
Leu	Leu	Arg	Thr	Pro	Ser	Ser	Ala	Ser	Thr	Val	Ser	Val	Gly	Ser	Ser
			580					585					590		
Glu	Thr	Arg	Gly	Glu	Arg	Val	Ile	Asp	Ala	Val	Arg	Phe	Thr	Leu	Arg
		595					600					605			
Gln	Thr	Ile	Ser	Phe	Pro	Pro	Arg	Asp	Glu	Trp	Asn	Asp	Phe	Asn	Phe
	610					615					620				
Asp	Met	Asp	Ala	Arg	Arg	Asn	Lys	Gln	Gln	Arg	Ile	Lys	Glu	Glu	Gly
625					630					635					640

glu

<210> 7

<211> 27

<212> DNA

<213> Artificial sequence

<220>

<223> Description of Artificial Sequence : p73-F1 sense primer

<400> 7

taoatgcaag taaagacag ttgtcc 27

<210> 8

<211> 29

<212> DNA

<213> Artificial sequence

<220>

<223> Description of Artificial Sequence : p73-R1 antisense primer

<400> 8

tgatgcaagt tgatcaagt ggacgtacg 29

<210> 9

<211> 29

<212> DNA

<213> Artificial sequence

<220>

<223> Description of Artificial Sequence : p73-F2 sense primer

<400> 9

taagtatact acgaagtgtg cgtgaaggg 29

<210> 10

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence : p73-R2 antisense primer

<400> 10

atgaactacg acgtacgaag tccacgtat 29

<210> 11

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence : HA-labeled expression construct

<400> 11

atgtatccat atgatgttc agataatgct 30

<210> 12
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence : p51-F1 sense primer

<400> 12
 aaagaaagtt attacgaig 20

<210> 13
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence : p51-R1 antisense primer

<400> 13
 ogcgtgtct ggttatagg 20

<210> 14
 <211> 20
 <212> DNA

<213> Artificial sequence

<220>
 <223> Description of Artificial Sequence : p51-F2 sense primer

<400> 14
 catggaccag cagattcaga 20

<210> 15
 <211> 19
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence : p51-R2 antisense primer

<400> 15
 catcaccctg atciggatg 19

<210> 16
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence : p51-F3 sense primer

<400> 16
 ccacccggac gttaccact 20

<210> 17
 <211> 18
 <212> DNA

<213> Artificial sequence

<220>

<223> Description of Artificial Sequence : p51-R3 antisense primer

<400> 17

gggtcataa ggtaccag 18

<210> 18

<211> 19

<212> DNA

<213> Artificial sequence

<220>

<223> Description of Artificial Sequence : p51-F4 sense primer

<400> 18

catgagctga gcggtgaat 19

<210> 19

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence : p51-R4 antisense primer

<400> 19

tatcttcacg cgccttcctg 20

<210> 20

<211> 18

<212> DNA

<213> Artificial sequence

<220>

<223> Description of Artificial Sequence : p51-F5 sense primer

<400> 20

atgaaccgcc gccaatt 18

<210> 21

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> Description of Artificial Sequence : p51-R5 antisense primer

<400> 21

gtgtgagga aggtactgca 20

<210> 22

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence : p51-F6 sense primer

<400> 22

tgaagatcaa agagtccttg 20

<210> 23

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> Description of Artificial Sequence : p51-R6 antisense primer

<400> 23

ctagtggtt tggccttg 20

Lisboa, 11 de Agosto de 2011.

Reivindicações

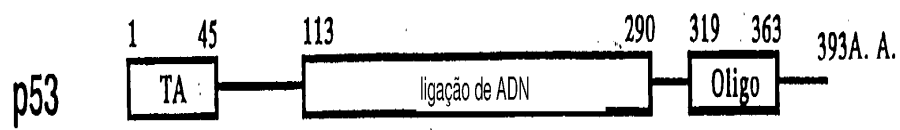
1. Anticorpo isolado que se liga especificamente a uma proteína p51 compreendendo a sequência de aminoácidos da SEQ ID N°:1, em que a dita proteína p51 é produzida por tecnologia recombinante ou síntese química.
2. Anticorpo isolado de acordo com a reivindicação 1, em que a ligação é determinada utilizando imunoprecipitação, imunocoloração, transferência de Western, ELISA ou ensaio rádio-imunológico (RIA).
3. Kit para detectar a proteína p51 compreendendo
(i) um anticorpo isolado de acordo com a reivindicação 1, e
(ii) meio de detectar o anticorpo.
4. Kit de acordo com a reivindicação 3, em que o meios para detectar o anticorpo é uma ou várias enzimas, uma co-enzima, um substrato, um polipéptido, um anticorpo ou um marcador detectável.
5. Kit de acordo com a reivindicação 4, em que o meio para detectar o anticorpo é um segundo anticorpo que se liga especificamente ao anticorpo.
6. Kit de acordo com a reivindicação 3, em que o anticorpo é um anticorpo monoclonal.
7. Kit de acordo com reivindicação 3, em que o anticorpo se encontra na forma de uma preparação purificada ou anticorpo policlonal.

Lisboa, 11 de Agosto de 2011.

Resumo

"Anticorpos contra a proteína p51"

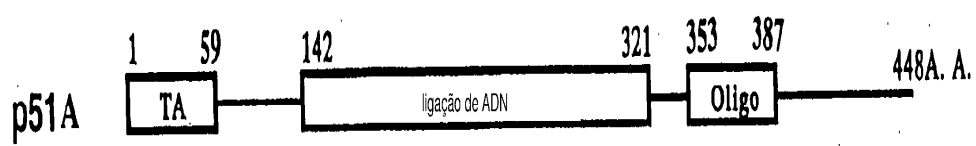
Uma proteína recombinante p51 compreendendo uma sequência de aminoácidos da SEQ ID N°:1, em que a dita proteína p51 recombinante é expressa a partir de um transformante com o vector de expressão de p51, um anticorpo isolado ou um seu fragmento que se liga especificamente à dita proteína p51 e um kit compreendendo o dito anticorpo ou o seu fragmento.



22%

60%

37%



30%

87%

65%

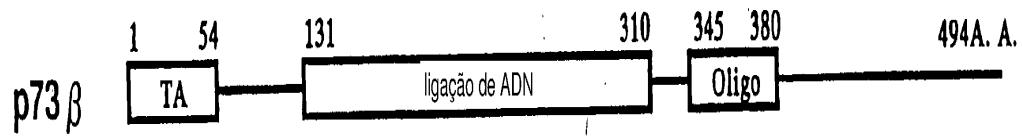


FIG. 1

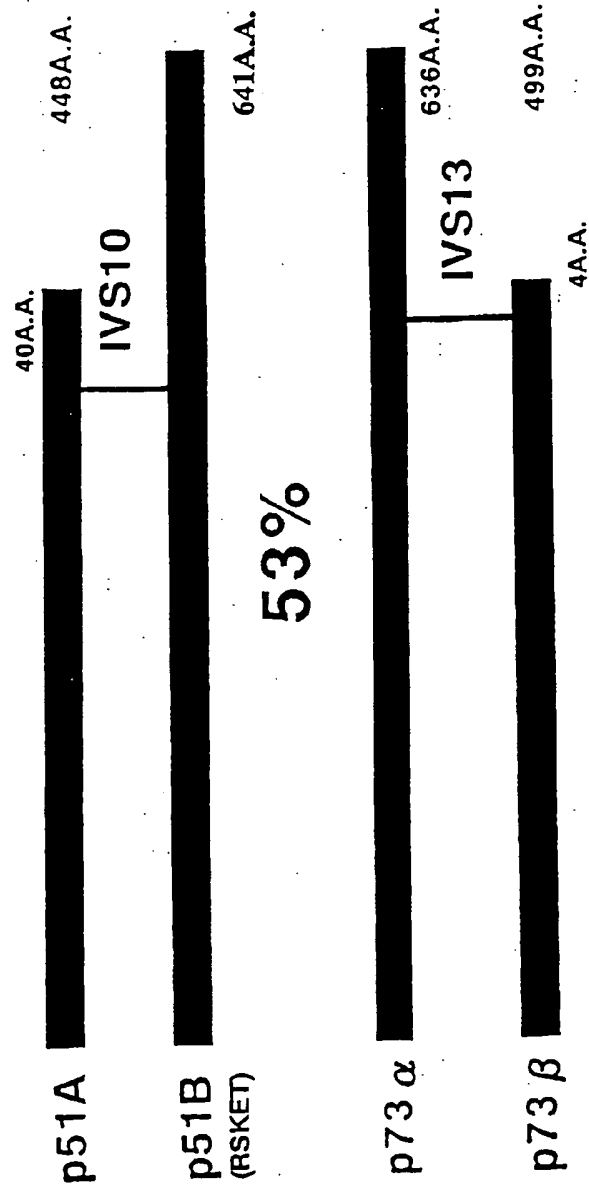
FIG. 2

p51A	MSOSTQTNEF	LSPEVFQHIW	DFLEQPICSV	QPIQUNFVDE	PSQDGATNKI	50
p73B	MAQSTATSP-	DGGTTFEHLW	SSLE-POSTY	--FQUP-QSS	RGHNEVVGGT	45
p53	KEEQSDPS-	VEPPLSQETF	SQU-----	--NKL-----L	PENNVLSPLP	36
Consensus	W.QST.T...	..P..FQH.W	S.LE-P....	--.LL.....	P.NN.....	50
p51A	EISMCIRMQ	DSGLSDPMMP	QYTMGLLNS	MDQOIQNGSS	STSPRYMTHA	100
p73B	DSIMCVFHL	--GHTTSVMA	CF---NLLSS	THQDMSSRAA	SASFYTPENA	90
p53	SQAHQDLMLS	----PQDI-E	QM-----FTED	PGPDEAPRMP	EAAAF-RVAHA	76
Consensus	..SMD...L	D....	Q....LL.S	...Q...R..	SASEY...NA	100
p51A	QNSVTAPSPY	ACPSSTFDAL	SPSPAIPSN	DTFQPHSDV	SFQOSSTAKS	150
p73B	-ASVPTHSPY	ACPSSTFDTH	SPAPVPSNT	DTFQPHREV	TFQOSSTAKS	139
p53	-PAAPTAA-	PAPAPSN-PL	SSS--VPSQK	TYQCSYQRL	QFLHSOTAKS	121
Consensus	-SVPTSPSY	ACPSSTFD.L	SPSP..PSNT	DTFQPH..V	TFQOSSTAKS	150
p51A	ATMTYSTELX	KLYCQIAKTC	PQIINVTTP	PQCAVIRAMP	VYKKAELITE	200
p73B	ATMTYSPUUX	KLYCQIAKTC	PQIINVSTTP	PQCAVIRAMP	VYKKAELITE	189
p53	VYMTYSPAUN	KMFQIAKTC	PQCLVNSTP	PQCTRVIRAMP	VYKQSCNITE	171
Consensus	ATMTYSP ILX	KLYCQIAKTC	PQIIN..TTP	PQCT..IRAMP	VYKKAELITE	200
p51A	VNRCPNHEL	SREFNEQQA	PSSHILIRVEG	NSHAQVVEEP	ITIRQSVAVF	250
p73B	VNRCPNHEL	GRDFNEQQA	PASHILIRVEG	NHLSQVVEEP	ITIRQSVAVF	239
p53	VNRCPNHER	CSD-SQC-LA	PPOHLIRVEG	NLRVHVLQOR	MTIRQSVAVF	219
Consensus	VNRCPNHEL	.RDFNEQQA	PSSHILIRVEG	N...QAVVEEP	ITIRQSVAVE	250
p51A	YEPRVQTEF	TTILYNNMCH	SSQGQWRR	PILITITLET	RQGNLGRRC	300
p73B	YEPRVQTEF	TTILYNNMCH	SSQGQWRR	PILITITLEN	RQGNLGRRS	289
p53	YEPRVQSDC	TTILYNNMCH	SSQGQWRR	PILITITLED	SSQNLGRNS	269
Consensus	YEPRVQTEF	TTILYNNMCH	SSQGQWRR	PILITITLET	RQGNLGRRS	300
p51A	FEARTCACPG	RDRKAEQDSI	RKQVSD--S	TKNGDGTKRP	FRQNTHGI-Q	347
p73B	FEARTCACPG	RDRKAEQDHY	REQQALNESS	AKNGAASKRA	EKQSPPAVPA	339
p53	FEVPCACPG	RDRRTBEENL	RKKGEPIH-H	ELPPGSTKRA	LPNNTSS---	314
Consensus	FEARTCACPG	RDRKAEQD..	REQQ....-S	.KNG..TKRA	F.QNT....	350
p51A	M-TSINKRRS	PDELLYLVP	RGRBTYEMLL	KINBSLELMQ	YLPQHTIETY	396
p73B	LGAGVKKRRH	QEDTYYLQV	RGRNFEILM	KLBSLELME	LVPOPLVDSY	389
p53	--SPQPKKKP	LDGEYFYLIQI	RGRBRFEMFR	ELHBALELKD	AQAGKEPGGS	362
Consensus	KRRR.	LD.E..YLQV	RGRB..EMML	KLBSLELM.	..PQ.....Y	400
p51A	RQQQQQQHQ-	HL-----	---LQK-H-	LLSACFRNE	LV-EPRRE--	427
p73B	RQQQQLLQRP	SHLQPPSYGP	VLSPMNVHVG	GANKLPSVNO	LVGQPPPHSS	439
p53	RAHSS-----	HL-----	---K-----	SKKGQ	---STSRH--	380
Consensus	RQQQQ.....	HL-----	...K-H-	NQ	LV-.P.RH--	450
p51A	--TPKQSDV-	--FFR---HS	KPPN-----	-----RSVYP-	448	
p73B	AATPNLGPVG	PGMLNHGHA	VPANGEMSSS	MSAQSMVSGS	HCTPPPPYHA	489
p53	---KKL----	--MFK-----T	EGPD-----	-----SD--	393	
Consensus	--TPKL..V-	--MF..--H.	.PPN-----	-----S.Y.-	500	
p51A	-----	-----	-----	-----	448	
p73B	DPSLVRTWGP	-----	-----	-----	499	
p53	-----	-----	-----	-----	393	
Consensus	-----	-----	-----	-----	510	

FIG. 3

p51B	MAQSTQTNF LSPENFCH R DKLQFICSV QP DUNFVDE PSEDGATNKI	50
p73d	MAQSTATSP- DGGTTFCHW SSLE-POSTY --RULP-QSS RGNHEVVGCT	45
Consensus	MAQST... LSPEN... R DKLQFICSV QP DUNFVDE PSEDGATNKI	50
p51B	EISMCIRMQ DSDLSDFMMP QYTNLGLLMS MOOQIONGSS STSPYNTDHA	100
p73d	DSMEVFHLE --GNTTSVMA QF---MLLSS TMOQSSRAA SASPYTPBHA	90
Consensus	EISMCIRMQ DSDLSDFMMP QYTNLGLLMS MOOQIONGSS STSPYNTDHA	100
p51B	QMSVTAQSPY AQPSSTFOAL SPSSAIPSNT DYPGPHFQV SFQOSSTAKS	150
p73d	-ASVPTHSPY AQPSSTFOTM SPFAVTPSNT DYPGPHFQV TFQOSSTAKS	139
Consensus	QMSVTAQSPY AQPSSTFOAL SPSSAIPSNT DYPGPHFQV SFQOSSTAKS	150
p51B	ATWTYSTLK KLYCQIAKTC PIQIKVTPP PQGAVIRAMP VYKKAEMVTE	200
p73d	ATWTYSTLK KLYCQIAKTC PIQIKVTPP PQGAVIRAMP VYKKAEMVTE	189
Consensus	ATWTYSTLK KLYCQIAKTC PIQIKVTPP PQGAVIRAMP VYKKAEMVTE	200
p51B	VVKRCPMHEL SREFNEGCA PSHLIRVEG KSHAQVMDP ITGROSVNVR	250
p73d	VVKRCPMHEL SREFNEGCA PSHLIRVEG KSHAQVMDP ITGROSVNVR	239
Consensus	VVKRCPMHEL SREFNEGCA PSHLIRVEG KSHAQVMDP ITGROSVNVR	250
p51B	YEPPQVGTEF TTYLYNFMCN SSCVGQNNRR PILIITLET RDGQVLGRRC	300
p73d	YEPPQVGTEF TTYLYNFMCN SSCVGQNNRR PILIITLET RDGQVLGRRC	289
Consensus	YEPPQVGTEF TTYLYNFMCN SSCVGQNNRR PILIITLET RDGQVLGRRC	300
p51B	FEARICACPG RDRKADEDSI RQCVSD-S TKNDGTRP FQNTNGI-Q	347
p73d	FEARICACPG RDRKADEDSI RQCVSD-S TKNDGTRP FQNTNGI-Q	339
Consensus	FEARICACPG RDRKADEDSI RQCVSD-S TKNDGTRP FQNTNGI-Q	350
p51B	M-TSIKKRRS PODELYLVV RGRBTBLL RIKESLELMQ YIPQHTIET	396
p73d	LGAGVKRRH QEDTYLVV RGRBTBLL RIKESLELMQ YIPQHTIET	389
Consensus	M-TSIKKRRS PODELYLVV RGRBTBLL RIKESLELMQ YIPQHTIET	400
p51B	RQQQQCHH LLOKQTSICS PPSYQSSHP LMMH-SHMK LPSVQLINP	445
p73d	RQQQQCHH LLOKQTSICS PPSYQSSHP LMMH-SHMK LPSVQLINP	433
Consensus	RQQQQCHH LLOKQTSICS PPSYQSSHP LMMH-SHMK LPSVQLINP	450
p51B	QQRNALTRIT IPDQGANIP MAGTHMPMAG DNGLSPTQA LPPPLSAPST	495
p73d	PPPHSSAATP NLGPNCPG-- MLNMHGHAVP ANGENSSSH-- --SACSMSG	478
Consensus	QQRNALTRIT IPDQGANIP MAGTHMPMAG DNGLSPTQA LPPPLSAPST	500
p51B	SHCTPPPPYP TQSTVQFLA ALGSSQLOY FTIQGUTIIY QIEHYSMDOL	545
p73d	SHCTPPPPYP TQSTVQFLA ALGSSQLOY FTIQGUTIIY QIEHYSMDOL	528
Consensus	SHCTPPPPYP TQSTVQFLA ALGSSQLOY FTIQGUTIIY QIEHYSMDOL	550
p51B	ASLKIPEQHR MARMQIUCH RQHERSPS HLLATPSSAS TVSGSSETR	595
p73d	GALKIPEQHR MARMQIUCH RQHERSPS HLLATPSSAS TVSGSSETR	578
Consensus	ASLKIPEQHR MARMQIUCH RQHERSPS HLLATPSSAS TVSGSSETR	600
p51B	GRVMDAVRF TUPOTISHP R-----DE RNDHFMQDA RNNKOFIKE	638
p73d	RQVMEAVRF RVPHTITIFN RGGPGGGDE MADRFQLPD CKARNOFIKE	628
Consensus	GRVMDAVRF TUPOTISHP R-----DE RNDHFMQDA RNNKOFIKE	650
p51B	--EGR--	641
p73d	EFTBATH	636
Consensus	--EGR--	658

F I G. 4



F I G. 5

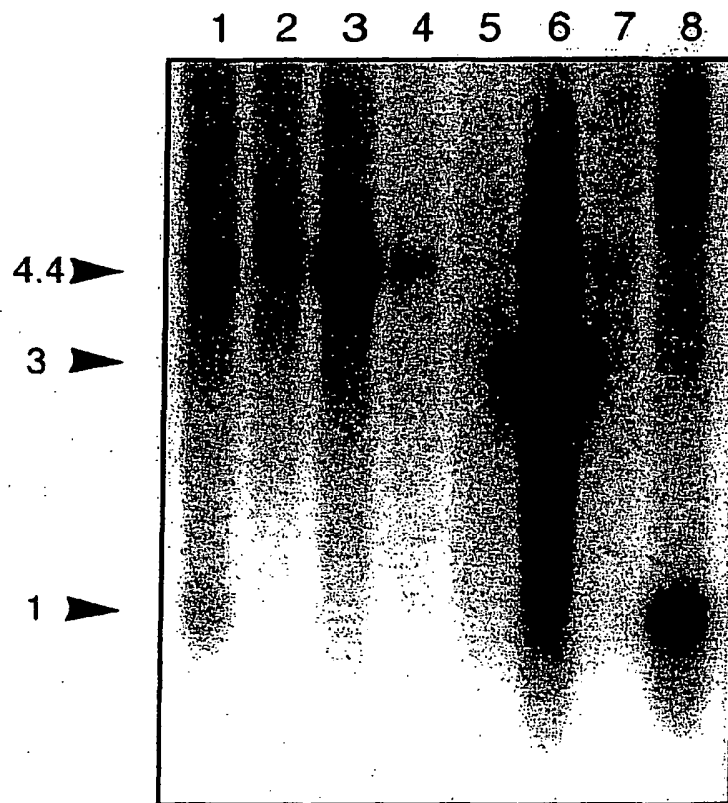


FIG. 6

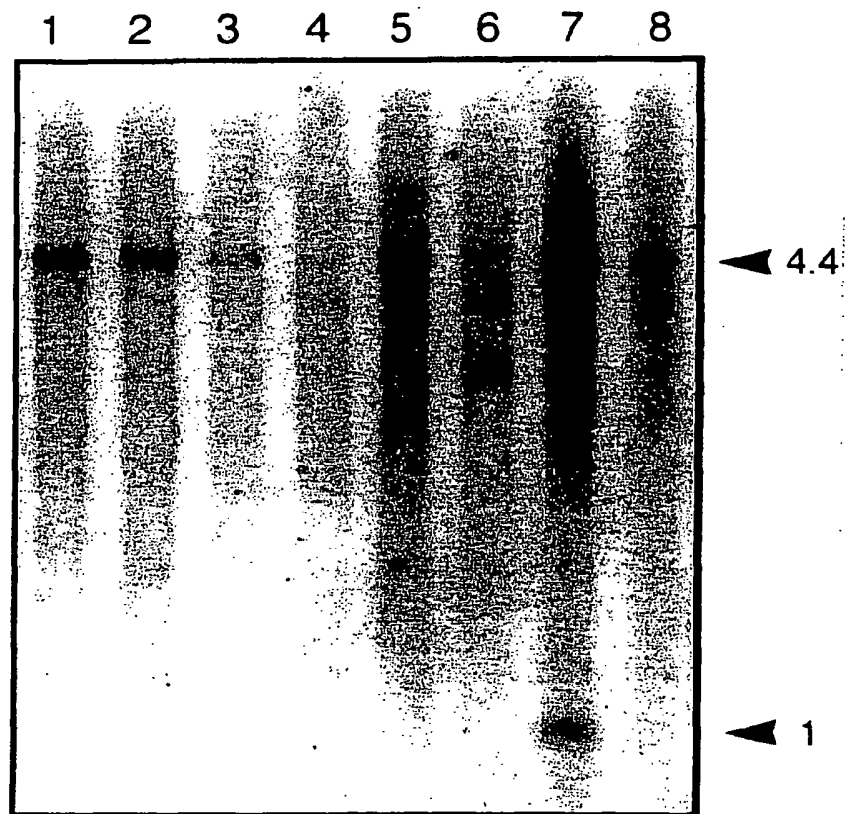
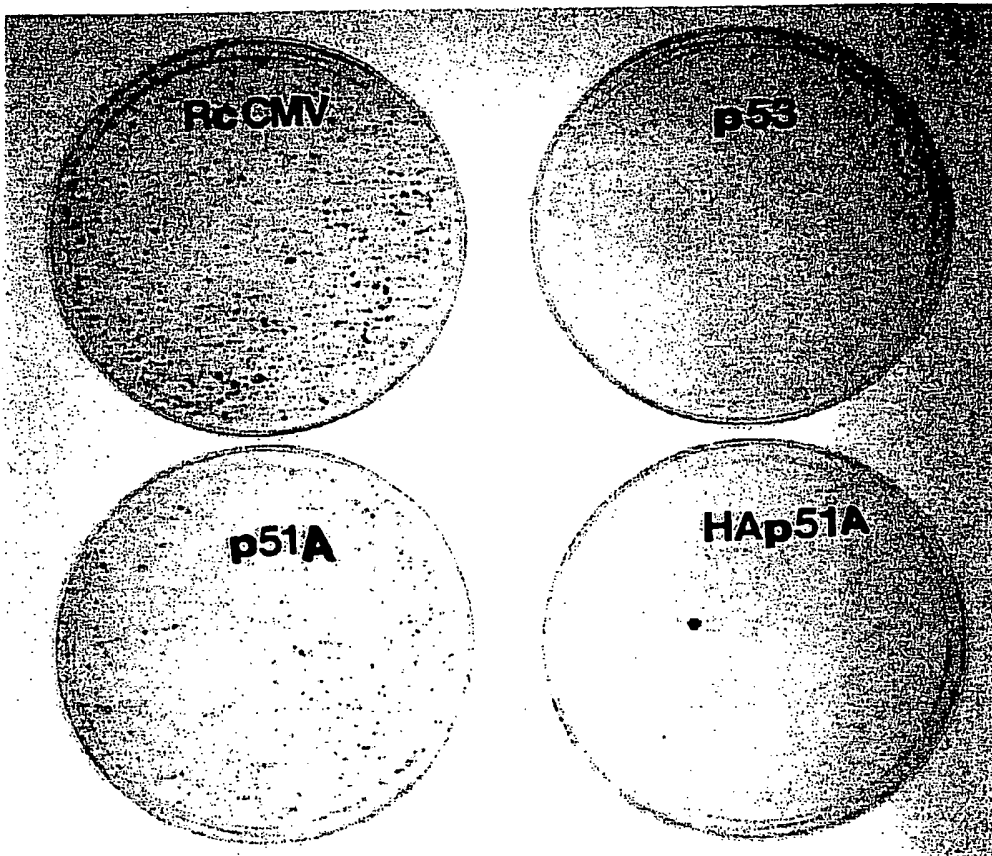


FIG. 7



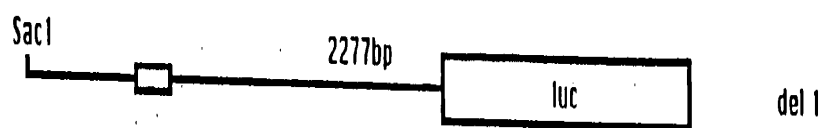
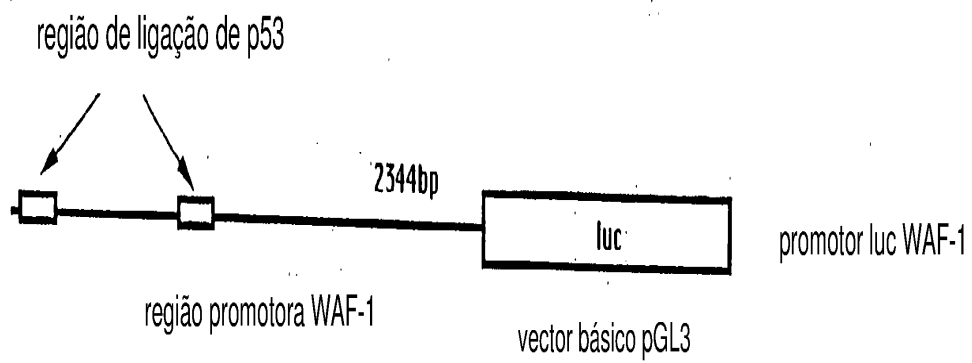


FIG. 8

FIG. 9

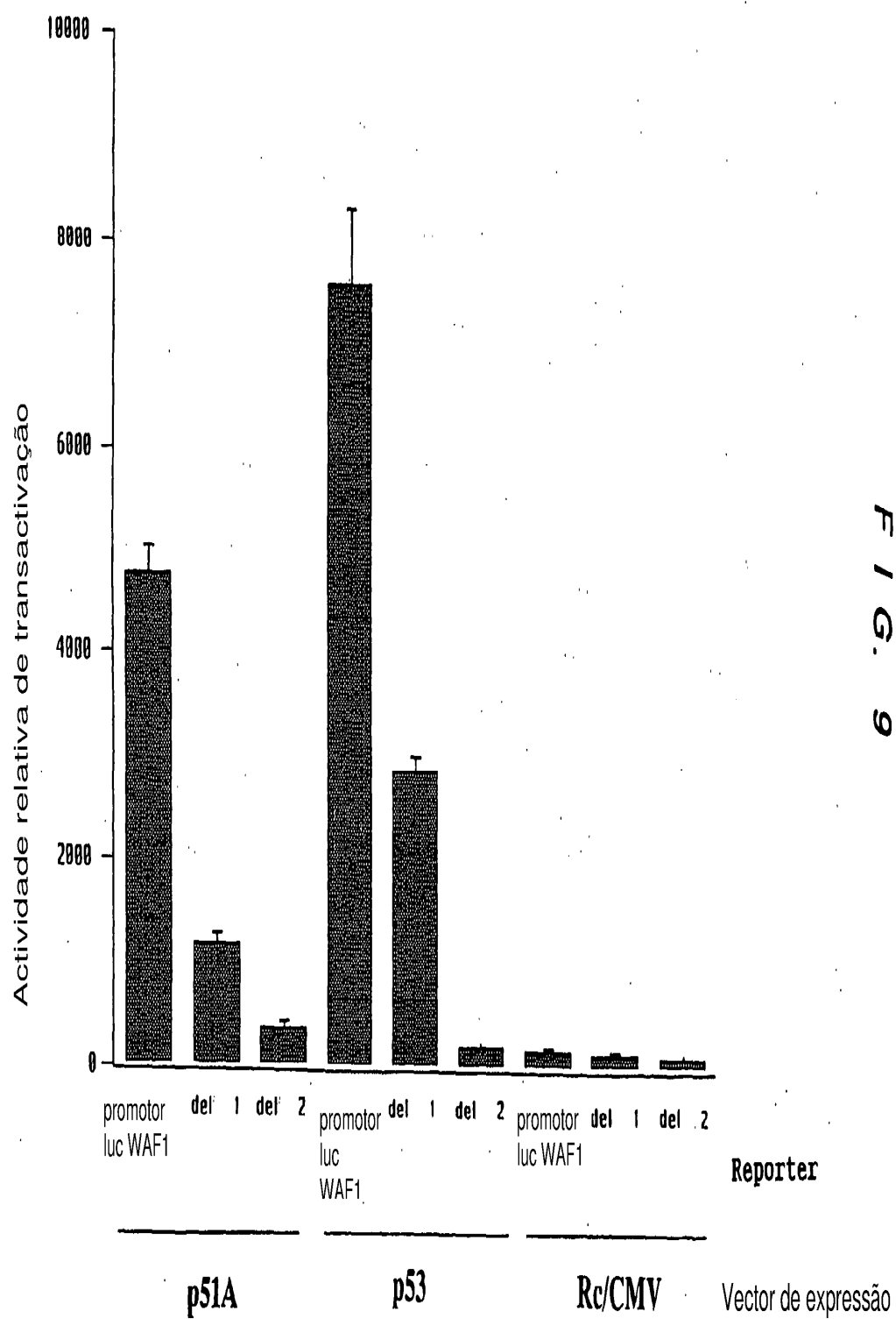


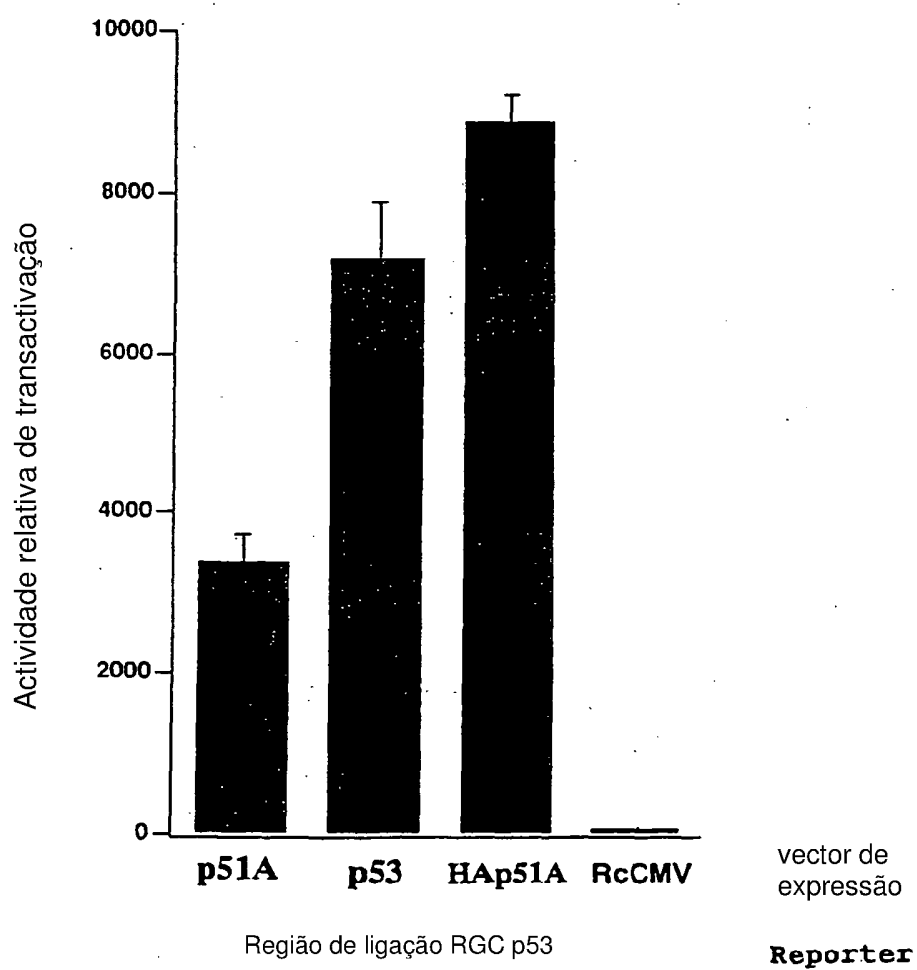
FIG. 10

FIG. 11

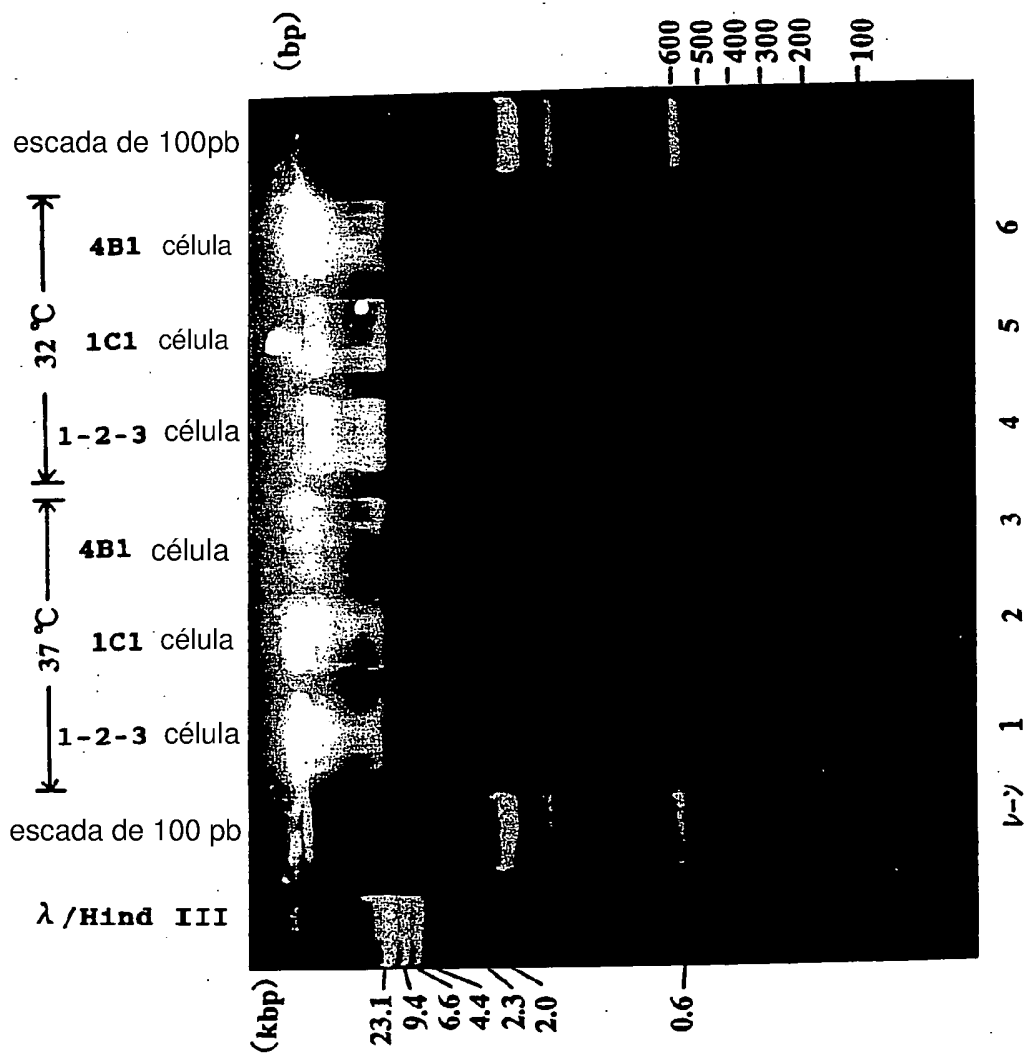


FIG. 12

```

mp51 1 ATGTCGCAGAGCAGCCAGACAAAGCGAGTTCTCAGCCAGAGGTCTTCCAGCATATCTGG 60
p51B 1 ATGTCGCAGAGCAGCAGACAAATGAATTCTCAGTCCAGAGGTCTTCCAGCATATCTGG 60
*****

mp51 61 GATTTTCTGGAACAGCCTATATGCTCAGTACAGCCCATCGAGTTGAACCTTTGTGGATGAA 120
p51B 61 GATTTTCTGGAACAGCCTATATGTTCAATTACAGCCCATGACTTGAACCTTTGTGGATGAA 120
*****

mp51 121 CCTTCCGAAATGGTGCAACAAACAGATTGAGATTAGCATGGATTGTATCCGCATGCAA 180
p51B 121 CCATCAGAAGATGGTGCGACAAACAGATTGAGATTAGCATGGACTGTATCCGCATGCA 180
***

mp51 181 GACTCAGACCTCAGTGACCCCATGTGGCCACAGTACACGAACCTGGGGCTCCTGAACAGC 240
p51B 181 GACTCGGACCTCAGTGACCCCATGTGGCCACAGTACACGAACCTGGGGCTCCTGAACAGC 240
*****

mp51 241 ATGGACCAGCAGATTTCAGAACGGCTCCTCGTCCACCAGCCCTACAACACAGACCAGCA 300
p51B 241 ATGGACCAGCAGATTTCAGAACGGCTCCTCGTCCACCAGTCCCTATAACACAGACCAGCG 300
*****

mp51 301 CAGAATAGCGTGACGGCGCCCTCGCCCTATGCACAGCCAGCTCCACCTTTGATGCCCTC 360
p51B 301 CAGAACAGCGTCACGGCGCCCTCGCCCTACGCACAGCCAGCTCCACCTTCGATGCTCTC 360
*****

mp51 361 TCTCCATCCCTGCCATTCCCTCCAACACAGATTACCGGGCCCAACAGCTTCGATGTG 420
p51B 361 TCTCCATACCCGCCATCCCTCCAACACAGACTACCCAGGGCCCAACAGTTTCGACGTG 420
*****

mp51 421 TCCTTCCAGCAGTCAAGCACTGCCAAGTCAGCCACCTGGACGTATTCCACCGAAGTGAAG 480
p51B 421 TCCTTCCAGCAGTCGAGCAACGCCAAGTCGGCCACCTGGACGTATTCCACTGAAGTGAAG 480
*****

mp51 481 AAGCTGTACTGCCAGATTGCGAAGACATGCCCCATCCAGATCAAGGTGATGACCCACCC 540
p51B 481 AAATCTACTGCCAATTGCAAAGACATGCCCCATCCAGATCAAGGTGATGACCCACCT 540
**

mp51 541 CCACAGGGCGCTGTTATCCGTGCCATGCCTGTCTACAAGAAAGCTGAGCATGTCACCGAG 600
p51B 541 CCTCAGGGAGCTGTTATCCGCGCCATGCCTGTCTACAAAAAGCTGAGCACGTCACCGAG 600
**

mp51 601 GTTGTGAAACGATGCCCTAACCATGAGCTGAGCCGTGAGTTCAATGAGGGACAGATTGCC 660
p51B 601 GTGGTGAAGCGGTGCCCAACCATGAGCTGAGCCGTGAATTCAACGAGGGACAGATTGCC 660
**

mp51 661 CCTCCAGTCATCTGATTGAGTAGAAGGGAACAGCCATGCCAGTATGTAGAAGATCCT 720
p51B 661 CCTCCTAGTCATTGATTGAGTAGAGGGGAACAGCCATGCCAGTATGTAGAAGATCCC 720
*****

mp51 721 ATCACGGGAAGGCAGAGCGTGCTGGTCCCTTATGAGCCACCACAGGTTGGCACTGAATTC 780
p51B 721 ATCACAGGAAGACAGAGTGTGCTGGTACCTTATGAGCCACCCAGGTTGGCACTGAATTC 780
*****

```

F I G. 1 3

```

mp51 781 ACAACAGTCCTGTACAATTTCTATGTGTAAACAGCAGCTGCGTCCGAGGAATGAACAGACGT 840
p51B 781 ACCACAGTCCTGTACAATTTCTATGTGTAAACAGCAGTGTGTGAGGGATGAACCGCGGT 840
      ** ***** **

mp51 841 CCAATTTTAATCATCTGTACTCTGGAAACCAAGAGATGGGCAAGTCTGGGCCGACGGTGC 900
p51B 841 CCAATTTTAATCATCTGTACTCTGGAAACCAAGAGATGGGCAAGTCTGGGCCGACGGTGC 900
      ***** **

mp51 901 TTTGAGGCCCGGATCTGTGCTTGGCCAGGAAGAGACCGGAAGGCAGATGAAGACAGCATC 960
p51B 901 TTTGAGGCCCGGATCTGTGCTTGGCCAGGAAGAGACCGGAAGGCAGATGAAGACAGCATC 960
      ***** **

mp51 961 AGAAAGCAGCAAGTATCGGACAGCGCAAGAAGCGGCGATGGTACGAAGCGCCCTTTCCGT 1020
p51B 961 AGAAAGCAGCAAGTTTCGGACAGTCAAGAAGCGGTGATGGTACGAAGCGCCCTTTCCGT 1020
      ***** **

mp51 1021 CAGAATACACACGGGAATCCAGATGACTTCCATCAAGAAACGGAGATCCCCAGATGATGAG 1080
p51B 1021 CAGAACACACATGGTATCCAGATGACATCCATCAAGAAACGAGATCCCCAGATGATGAA 1080
      ***** **

mp51 1081 CTGCTGTACCTACCAGTGAGAGGTCTGTGAGACGTACGAGATGTTGCTGAAGATCAAAGAG 1140
p51B 1081 CTGTTATACTTACCAGTGAGGGGCCGTGAGACTTATGAAATGCTGTTGARGATCAAAGAG 1140
      *** * ** ***** **

mp51 1141 TCACTGGAGCTCATGCAGTACCTCCCTCAGCACACGATCGAAACGTACAGGCAGCAGCAG 1200
p51B 1141 TCCCTGGAACTCATGCAGTACCTTCCTCAGCACACAATTGAACGTACAGGCAACAGCAA 1200
      ** ***** **

mp51 1201 CAGCAGCAGCACCAGCACCTACTTCAGAAACAGACCTCGATGCAGTCTCAGTCTTCATAT 1260
p51B 1201 CAGCAGCAGCACCAGCACTTACTTCAGAAACAGACCTCAATACAGTCTCCATCTTCATAT 1260
      ***** **

mp51 1261 GGCAACAGTTCCTCCCTCTGAACAAAATGAACAGCATGAACAAGCTGCCTTCCGTGAGC 1320
p51B 1261 GGTAAACAGTTCCTCCCTCTGAACAAAATGAACAGCATGAACAAGCTGCCTTCTGTGAGC 1320
      ** ***** **

mp51 1321 CAGCTTATCAACCCACAGCAGCGCAANTGCCCTCACTCCACCCACCATGCCGTGAGGGCATG 1380
p51B 1321 CAGCTTATCAACCCCTCAGCAGCGCAACGCCCTCACTCCTACACCATTCTCTGATGGCATG 1380
      ***** **

mp51 1381 GGAGCCAAACATTCCCTATGATGGGCACTCAATGCCCAATGGCTGGAGACATGAATGGACTC 1440
p51B 1381 GGAGCCAAACATTCCCATGATGGGCACCCACATGCCCAATGGCTGGAGACATGAATGGACTC 1440
      ***** **

mp51 1441 AGCCCTACCCAAAGCTCTCCCTCCTCCACTCTCCATGCCCTCCACCTCCCACTGCACCCCA 1500
p51B 1441 AGCCCCACCCAGGCACTCCCTCCCCCACTCTCCATGCCATCCACCTCCCACTGCACCCCA 1500
      ***** **

mp51 1501 CCACCGCCCTACCCACAGACTGCAGCATTTGTCAATTCTTAGCAAGGTTGGGCTGCTCA 1560
p51B 1501 CCACCTCCGTATCCACAGATTGCAGCATTTGTCAATTCTTAGCGAGGTTGGGCTGTTCA 1560
      ***** **

```

F I G. 1 4

```

mp51 1561 TCATGCCTGGACTATTTACGACCCAGGGGCTGACCACCATCTATCAGATTGAGCATTAC 1620
p51B 1561 TCATGCTCTGGACTATTTACGACCCAGGGGCTGACCACCATCTATCAGATTGAGCATTAC 1620
*****

mp51 1621 TCCATGGATGATTTGGCAAGTCTGAAGATCCCTGAACAGTTCCGACATGCCATCTGGAAG 1680
p51B 1621 TCCATGGATGATCTGGCAAGTCTGAAATCCCTGAGCAATTTCGACATGCGATCTGGAAG 1680
*****

mp51 1681 GGCATCCTGGACCACAGGCAGCTGCACGACTTCTCCTCACCTCCTCATCTCCTGAGGACC 1740
p51B 1681 GGCATCCTGGACCACCGGCAGCTCCACGAATTCTCCTCCCTTCTCATCTCCTGCGGACC 1740
*****

mp51 1741 CCAAGTGGTGCCTCTACCGTCAGTGTGGGCTCCAGTGAGACCCGGTGAACGTGTGATC 1800
p51B 1741 CCAAGCAGTGCCTCTACAGTCAGTGTGGGCTCCAGTGAGACCCGGGCTGAGCGTGTATT 1800
*****

mp51 1801 GATGCCGTGCGCTTTACCCCTCCGCCAGACCATCTCTTTCCACCCCGTGACGAGTGAAT 1860
p51B 1801 GATGCTGTGCGATTACCCCTCCGCCAGACCATCTCTTTCCACCCCGAGATGAGTGAAT 1860
*****

mp51 1861 GATTTCAACTTTGACATGGATTCTCGTCGCAACAAGCAGCAGCGTATCAAAGAGGAAGGA 1920
p51B 1861 GACTTCAACTTTGACATGGATGCTCGCCGCAATAAGCAACAGCGCATCAAAGAGGAGGGG 1920
**

mp51 1921 GAA 1923
p51B 1921 GAG 1923
**

```

FIG. 15

```

mp51BnAA  MSQSTQTSEFLSPEVFOHIWDFLEQPICSVQPIELNFVDEPSENGATNKIEISMDCIRMQ  60
p51B aa    MSQSTQTNEFLSPEVFOHIWDFLEQPICSVQPIDLNFVDEPSEDGATNKIEISMDCIRMQ  60
*****

mp51BnAA  DSDLSDPMWPOYTNLGLLNSMDQIQNGSSSTSPYNTDHAQNSVTAPSPYAQPSSTFDAL  120
p51B aa    DSDLSDPMWPOYTNLGLLNSMDQIQNGSSSTSPYNTDHAQNSVTAPSPYAQPSSTFDAL  120
*****

mp51BnAA  SPSPALPSNTDYPGPHSFDVSFQSSSTAKSATWTYSTELRKLYCQIAKTCPIQIKVMTFP  180
p51B aa    SPSPALPSNTDYPGPHSFDVSFQSSSTAKSATWTYSTELRKLYCQIAKTCPIQIKVMTFP  180
*****

mp51BnAA  PQGAVIRAMPVYKKAESHVTEVVKRCPNHELSTREFNEGQIAPPSHLIRVEGNSHAQYVEDP  240
p51B aa    PQGAVIRAMPVYKKAESHVTEVVKRCPNHELSTREFNEGQIAPPSHLIRVEGNSHAQYVEDP  240
*****

mp51BnAA  ITGRQSVLVPIEPPQVGTEFTTVLYNFMCNSSCVGGMNRRPILIIIVTLETRDGQVLGRRC  300
p51B aa    ITERQSVLVPIEPPQVGTEFTTVLYNFMCNSSCVGGMNRRPILIIIVTLETRDGQVLGRRC  300
*****

mp51BnAA  FEARICACPGDRKAEDEDSIRKQVSDSAKNGDGTKRPFRQNTBGIQMTSIXKRRSPDDE  360
p51B aa    FEARICACPGDRKAEDEDSIRKQVSDSTKNGDGTKRPFRQNTBGIQMTSIXKRRSPDDE  360
*****

mp51BnAA  LLYLPVRGRET YEMLLKIKESLELMQYLPQHTIETIRQQQQQQHQLLQKQTSIQSSSY  420
p51B aa    LLYLPVRGRET YEMLLKIKESLELMQYLPQHTIETIRQQQQQQHQLLQKQTSIQSSSY  420
*****

mp51BnAA  GNSSPPLNKMNSMKNLPSVSQLINPQORNALTPTTMPEGMGANIPMMGTHMPMAGDMNGL  480
p51B aa    GNSSPPLNKMNSMKNLPSVSQLINPQORNALTPTTIPDGMGANIPMMGTHMPMAGDMNGL  480
*****

mp51BnAA  SPTQALPPPLSMPSTSHCTPPPPYPTDCSIVSFLARLGCSSCLDYFTTQGLTTIYQIEHY  540
p51B aa    SPTQALPPPLSMPSTSECTPPPPYPTDCSIVSFLARLGCSSCLDYFTTQGLTTIYQIEHY  540
*****

mp51BnAA  SMDDLASLKIFEQFRHAIWKGILDHROLEDFFSPPHLLRTPSGASTVSVGSSETRGERVI  600
p51B aa    SMDDLASLKIFEQFRHAIWKGILDHROLEDFFSPPHLLRTPSSASTVSVGSSETRGERVI  600
*****

mp51BnAA  DAVRFTLRQTISFFPRDEWNDNFNDMDSRNKQQRKEEGE  641
p51B aa    DAVRFTLRQTISFFPRDEWNDNFNDMDARNKQQRKEEGE  641
*****

```