



(11) Número de Publicação: **PT 760867 E**

(51) Classificação Internacional:
C07K 14/435 (2006.01) **C07K 14/47** (2006.01)
C12Q 1/68 (2006.01)

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

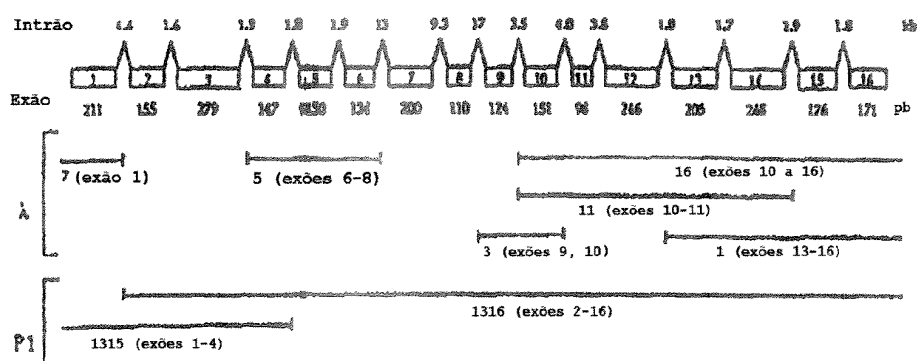
(22) Data de pedido: 1994.12.16	(73) Titular(es): OREGON HEALTH SCIENCES UNIVERSITY 3181 S.W. SAM JACKSON PARK ROAD PORTLAND, OR 97201 US DANA-FARBER CANCER INSTITUTE US
(30) Prioridade(s): 1993.12.17 US 0168877 1994.03.08 US 0209521 1994.12.09 US 0352902	
(43) Data de publicação do pedido: 1997.03.12	(72) Inventor(es): SEAN M. BAKER US RONI J. BOLLAG US RICHARD D. KOLODNER US C. ERIC BRONNER US ROBERT M. LISKAY US
(45) Data e BPI da concessão: 2006.06.14 011/2006	(74) Mandatário: JOSÉ EDUARDO LOPES VIEIRA DE SAMPAIO R DO SALITRE 195 RC DTO 1250-199 LISBOA PT

(54) Epígrafe: **PROCESSO PARA A DETECÇÃO DE ALTERAÇÕES NO MODO DE REPARAÇÃO DE ERROS NO EMPARELHAMENTO DO ADN**

(57) Resumo:

RESUMO

PROCESSO PARA A DETECÇÃO DE ALTERAÇÕES NO MODO DE REPARAÇÃO DE ERROS NO EMPARELHAMENTO DO ADN



Os requerentes verificaram agora que os eucariotos, incluindo os de mamíferos, têm um processo de reparação dos erros de emparelhamento do ADN análogo ao processo que existe nas bactérias. Os defeitos ou alterações neste processo de reparação dos erros de emparelhamento num mamífero, tal como nos seres humanos, resultarão na acumulação de sequências de ADN repetidas, instáveis. Esse fenótipo tem uma elevada correlação com o estado de doença num certo número de cancros, tal como os cancros hereditários do cólon. De acordo com isto, a descoberta de um defeito ou de uma alteração no processo pode ser o diagnóstico de uma pré-disposição para o cancro e o prognóstico para um cancro em particular. Os requerentes também descobriram e sequenciaram um dos genes neste processo num certo número de mamíferos, incluindo seres humanos. Este gene, referido aqui como MSH2, tem muitas aplicações. Pode ser utilizado em ensaios, para expressar o produto do gene, para avaliações de fármacos e para fins terapêuticos. Os requerentes também descrevem aqui um

DESCRIÇÃO

PROCESSO PARA A DETECÇÃO DE ALTERAÇÕES NO MODO DE REPARAÇÃO DE ERROS NO EMPARELHAMENTO DO ADN

A presente invenção foi feita com apoio do governo, suportada pelo acordo nº GM 32741 e o acordo HG00395/GM50006 concedido pela General Sciences Division do National Institute of Health. O governo tem alguns direitos sobre a presente invenção.

Âmbito da invenção

A presente invenção envolve os genes para a reparação de erros de emparelhamento do ADN. Em particular, a presente invenção tem por objecto a identificação de mutações e polimorfismos em genes para a reparação de erros de emparelhamento do ADN, a identificação e caracterização de tumores com defeitos que envolvem a reparação de erros de emparelhamento do ADN e a detecção de susceptibilidade genética para o cancro.

Antecedentes

Nos últimos anos, com o desenvolvimento das poderosas técnicas de clonagem e amplificação, tal como a reacção em cadeia de polimerase (RCP), em combinação com um corpo de informação rapidamente acumulado respeitante à estrutura e localização de numerosos genes e marcadores humanos, tornou-se prático e aconselhável recolher e analisar amostras de ADN ou de ARN de indivíduos que são membros de famílias que estão identificadas como portadoras de uma alta frequência de certos distúrbios transmissíveis geneticamente. Por exemplo, os processos de rastreio são

utilizados por rotina para rastrear os genes envolvidos na anemia dos drepanócitos, fibrose quística, síndrome frágil do cromossoma X e esclerose múltipla. Para alguns tipos de distúrbios, o diagnóstico precoce pode melhorar muito o prognóstico de longo prazo da pessoa adoptando, por exemplo, uma rotina de diagnóstico agressiva e/ou fazendo alterações no estilo de vida se for apropriado quer para prevenir ou preparar para um problema antecipado.

Uma vez identificada uma mutação particular de um gene humano e ligada a uma doença, o desenvolvimento de processos de rastreio para identificar indivíduos de alto risco pode ser relativamente directo. Por exemplo, depois de se compreender a estrutura e o papel fenotípico anormal do gene mutante, é possível preparar iniciadores para serem utilizados na RCP para se obterem quantidades amplificadas de genes amplificados de indivíduos para serem analisados. Contudo, a descoberta inicial de um gene de mutante, isto é, a sua estrutura, a sua localização e a ligação com um problema de saúde hereditário, conhecido, requer um esforço experimental substancial e estratégias de investigação criativas.

Uma abordagem para descobrir o papel de um gene mutante que causa uma doença, começa com estudos clínicos nos indivíduos que pertencem a famílias que exibem uma elevada frequência da doença. Nestes estudos, a localização aproximada do locus que causa a doença é determinada indirectamente por meio da pesquisa de um marcador do cromossoma que tende a segregar com o locus. Uma limitação principal desta abordagem é que, embora a localização genómica aproximada do gene possa ser determinada, geralmente não permite o isolamento actual nem a sequenciação do gene. Por exemplo, Lindblom et al.³ relatou

resultados de estudos de análise de ligação realizados com marcadores de PCSS (polimorfismo do comprimento de sequências simples, SSLP na terminologia inglesa) em indivíduos de uma família conhecida por exibir uma elevada incidência de cancro do cólon não poliposo hereditário (CCNPH). Lindblom et al. verificaram uma "ligação estreita" entre um marcador polimórfico no ramo curto do cromossoma 3 humano (3p21-23) e um locus de doença aparentemente responsável pelo aumento do risco de um indivíduo desenvolver cancro do cólon. Embora a 3p21-23 seja uma localização relativamente pouco específica em relação a todo o genoma, representa uma enorme região do ADN relativamente à pequena dimensão do gene mutante. O gene mutante podia ser separado dos marcadores identificando o locus de milhões de bases. No melhor, esses estudos de ligação têm apenas uma utilidade limitada para fins de rastreio porque, para prever o risco de uma pessoa, deve-se realizar a análise genética com marcadores genéticos fortemente ligados num certo número de indivíduos relacionados na família. É muita vez impossível obter essa informação, particularmente se os indivíduos da família afectada morreram. Também podem não existir marcadores informativos na família em análise. Sem conhecer a estrutura dos genes, não é possível fazer amostras, amplificá-las, sequenciá-las e determinar directamente se o indivíduo comporta o gene mutante.

Uma outra abordagem para descobrir o gene mutante que causa a doença, começa com a produção e a triagem dos iniciadores de RCP, com base na informação conhecida acerca da doença, por exemplo, teorias para estes mecanismos do estado da doença, estruturas e função das proteínas relacionadas, possíveis genes análogos em seres humanos ou noutras espécies, etc. O objectivo é isolar e sequenciar os

genes candidatos normais que se acredita que ocorrem algumas vezes nas formas mutantes tornando o indivíduo predisposto para a doença. Esta abordagem é altamente dependente do quanto se sabe acerca da doença ao nível molecular e da capacidade do investigador para construir estratégias e processos para encontrar genes candidatos. A associação de uma mutação num gene candidato com uma doença, deve ser demonstrada, em última análise, pela realização de ensaios nos membros de uma família que tem uma alta incidência da doença. A forma mais directa e definitiva para confirmar essa ligação em estudos de uma família é utilizar iniciadores de RCP que se destinam a amplificar porções de um gene candidato em amostras recolhidas dos membros dessa família. Os produtos dos genes amplificados são então sequenciados e comparados com a estrutura do gene normal com o fim de encontrar e caracterizar as mutações. Uma dada mutação está, em última análise, implicada quando mostra que os indivíduos afectados a têm, enquanto os indivíduos não afectados não a têm e que a mutação causa uma alteração na função da proteína que não é simplesmente um polimorfismo.

Outra maneira de mostrar uma elevada probabilidade de ligação entre uma mutação de um gene candidato e a doença é pela determinação da localização do cromossoma do gene, comparando assim a localização do mapa dos genes com regiões conhecidas de locus ligados à doença tal como o identificado por Lindblom et al. A localização coincidente no mapa de um gene candidato na região de um locus ligado à doença, identificado previamente, pode implicar fortemente uma associação entre uma mutação no gene candidato e a doença.

Há outras maneiras de mostrar que as mutações num gene candidato podem estar ligadas à doença. Por exemplo, podem introduzir-se formas mutantes, produzidas artificialmente, em animais. A incidência da doença em animais que comportam o gene mutante pode então ser comparada com animais com o genotipo normal. A incidência significativamente elevada da doença em animais com o genotipo mutante, relativamente a animais com o gene de tipo selvagem, podem suportar a teoria que as mutações no gene candidato são algumas vezes responsáveis pela ocorrência da doença.

Um tipo de doença que recebeu recentemente muita atenção por causa da descoberta de mutações de genes ligadas à doença é o Cancro do Cólon Não Polipóide Hereditário (CCNPH)^{1,2}. Membros de famílias com CCNPH também exibem uma susceptibilidade acrescida para outros cancros incluindo os do endométrio, dos ovários, gástricos e da mama. Crê-se que aproximadamente 10 % dos cancros colorectais sejam CCNPH. Tumores de pacientes com CCNPH exibem um defeito genético raro em que sequências de ADN repetidas, curtas, tal como as sequências repetidas de dinucleótidos encontradas no ADN dos cromossomas humanos ("ADN de micro-satélite"), que parecem ser instáveis. Esta instabilidade genómica das sequências de ADN repetidas, curtas, algumas vezes designadas por fenotipo "RER+", é também observada, numa proporção significativa, numa grande variedade de tumores esporádicos, sugerindo que muitos tumores esporádicos podem ter adquirido mutações que são semelhantes (ou idênticas) às mutações que são herdadas nos CCNPH.

Os estudos da ligação genética identificaram locus de CCNPH que se pensa que sejam da ordem de 90 % dos CCNPH. Os locus constituem o mapa dos cromossomas 2p15-16 (2p21) e

3p21-23. Estudos subsequentes identificaram o gene de reparação de erros de emparelhamento do ADN humano *hMSH2* como sendo o gene no cromossoma 2p21 em que as mutações são responsáveis por uma fracção significativa dos cancros CCNPH^{1, 2 12}. O *hMSH2* é um dos vários genes cuja função normal é identificar e corrigir os erros de emparelhamento do ADN incluindo os que seguem cada ciclo de replicação do cromossoma.

O processo de reparação dos erros de emparelhamento melhor definido é o processo de MutHLS de *E.coli* que promove uma reacção de reparação por excisão de um fragmento longo (aproximadamente 3 Kb) que depende dos produtos dos genes *mutH*, *mutL*, *mutS* e *mutU* (*uvrD*). O processo de MutHLS parece ser a via mais activa para a reparação dos erros de emparelhamento em *E.coli* e é conhecido tanto por aumentar a fidelidade da replicação do ADN como por actuar na recombinação de produtos intermédios contendo as bases desemparelhadas. O sistema tem sido reconstituído *in vitro*, e requer as proteínas de *mutH*, *mutL*, *mutS* e *mutU* (*uvrD*) (helicase II) em conjunto com a holoenzima polimerase III do ADN, a ligase de ADN, as proteínas de ligação do ADN de estrutura helicoidal simples (SSB) e uma das exonucleases de ADN de estrutura helicoidal simples, Exo I, Exo VII ou RecJ. *hMSH2* é homóloga do gene bacteriano *mutS*. Uma via similar em fungos inclui o gene de levedura *MSH2* e dois genes semelhantes a *mutL*, referidos como *PMS1* e *MLH1*.

Com o conhecimento de que as mutações num gene do tipo *mutS* humano (*hMSH2*) algumas vezes causa cancro e a descoberta de que os tumores CCNPH exibem instabilidade do ADN de micro-satélite, intensificou-se o interesse em outros genes e produtos de genes de reparação dos erros de

emparelhamento do ADN e nos seus possíveis papéis no CCNPH e/ou noutros cancros. Estima-se que pelo menos 1 em cada 200 indivíduos comporta uma mutação que no gene *hMSH2* ou noutros genes relacionados que codificam outras proteínas no mesmo processo de reparação dos erros de emparelhamento no ADN.

Um importante objectivo do trabalho dos requerentes tem sido identificar genes humanos que são úteis para a avaliação e identificação de indivíduos que estão em risco elevado de desenvolver cancro. Outros objectos são: determinar as sequências dos exões e as estruturas de intrões de flanqueio nesses genes; utilizar a informação estrutural para desenhar processos de ensaio com o fim de identificar e caracterizar mutações que resultam na ausência ou num defeito num produto de gene que confere susceptibilidade ao cancro; e para distinguir essas mutações das variações polimórficas "não perigosas". Um poutro objecto é utilizar a informação estrutural relacionada com os exões e as sequências de intrões de flanqueio de um gene ligado ao cancro, para diagnosticar tipos de tumores e prescrever a terapia apropriada. Um outro objecto é utilizar para o estudo a informação estrutural relacionada com um gene ligado ao cancro para identificar outros genes humanos candidatos relacionados.

Sumário da invenção

Com base no conhecimento dos requerentes sobre os mecanismos de reparação dos erros de emparelhamento do ADN em bactérias e fungos incluindo a conservação de genes de reparação de desemparelhamentos, os requerentes pensaram que homólogos de genes de reparação dos erros de emparelhamento de ADN devem existir e que essas mutações

nesses homólogos que afectam a função da proteína, causariam, provavelmente a instabilidade genética; levando possivelmente a um risco acrescido do desenvolvimento de certas formas de cancro humano.

Os requerentes isolaram e sequenciaram dois genes humanos, *hPMS1* e *hMLH1*, cada um dos quais codifica para uma proteína envolvida na reparação do desemparelhamento do ADN. *hPMS1* e *hMLH1* são homólogos do gene *mutL* encontrado na *E.coli*. Os estudos dos requerentes suportam-se fortemente numa associação entre as mutações nos genes de reparação dos erros de emparelhamento do ADN e a susceptibilidade a CCNPH. Assim, a informação sobre as sequências dos genes de reparação do desemparelhamento do ADN, nomeadamente as estruturas de ADNc e ADN genómico relacionadas com *hPMS1* e *hMLH1*, tornam possível um certo número de processos úteis que relacionam respeitantes à determinação do risco de cancro e no respectivo diagnóstico. Um grande número de estruturas de nucleótidos e de proteínas são úteis nesses processos.

Os requerentes mapearam a localização de *hMLH1* no cromossoma humano 3p21-23. Trata-se de uma região desse genoma humano, com base nos estudos sobre a família, que exhibe um locus que predispõe os indivíduos para CCNPH. Adicionalmente, os requerentes verificaram uma mutação numa região conservada do ADNc de *hMLH1* em indivíduos afectados com CCNPH numa família sueca. Não se encontrou a mutação em indivíduos não afectados da mesma família, nem é um simples polimorfismo. Os requerentes também verificaram que uma mutação homóloga em fungos resulta numa proteína de reparação de erros de emparelhamento do ADN defeituosa. Os requerentes também encontraram uma mutação numa secção alterada em indivíduos afectados com *hMLH1* de uma família

inglesa. A descoberta dos requerentes de mutações ligadas ao cancro no *hMLH1*, combinada com a posição dos genes no mapa que é coincidente com um locus ligado ao CCNPH previamente identificado, mais o provável papel do gene *hMLH1* na prevenção da mutação, faz do *hMLH1* um candidato de excelência subjacente a uma forma de cancro humano comum hereditário e um candidato primordial para avaliar e identificar indivíduos que tenham um elevado risco de desenvolvimento de cancro.

O *hMLH1* tem 19 exões e 18 intrões. Os requerentes determinaram a localização de cada um dos 18 intrões relativamente ao ADN de *hMLH1*. Os requerentes também determinaram a estrutura de todas as regiões da fronteira intrão/exão de *hMLH1*. O conhecimento das estruturas de todas as regiões da fronteira intrão/exão de *hMLH1* torna possível estabelecer regimes de avaliação eficientes para localizar mutações que afectam negativamente a estrutura e a função dos produtos dos genes. Além disso, os requerentes desenharam conjuntos completos de pares de iniciadores de oligonucleótidos que podem ser utilizados na RCP para amplificar exões individuais completos em conjunto com estruturas de fronteira que rodeiam o intrão.

Os requerentes mapearam a localização do *hPMS1* no cromossoma humano 7. Os estudos subsequentes de outros³⁹, confirmaram as previsões dos requerentes de que a mutação neste gene estava ligada ao CCNPH.

A utilização mais imediata da presente invenção, será em testes de avaliação em seres humanos que são membros de famílias que exibem uma elevada frequência excepcional de cancro de início precoce, por exemplo, CCNPH. De acordo com isto, um dos aspectos da presente invenção compreende um

processo de diagnóstico de susceptibilidade para o cancro do cólon num indivíduo, por meio da detecção de uma mutação num ácido nucleico do *hMLH1*, tal como está estabelecido na reivindicação 1 ou um ácido nucleico em *hPMS1*, tal como está estabelecido na reivindicação 3, num tecido de um indivíduo, em que a mutação é indicativa de uma susceptibilidade do indivíduo para o cancro do cólon.

O processo de diagnóstico compreende, preferencialmente, as etapas de: 1) amplificação de um ácido nucleico de um segmento do gene de reparação do erro de emparelhamento ou do produto do gene; 2) comparação do segmento amplificado com um segmento análogo de um alelo de tipo selvagem do gene de reparação do erro de emparelhamento ou do produto do gene; e 3) detecção da diferença entre o segmento amplificado e o segmento análogo, sendo a diferença indicativa de uma mutação no gene de reparação do erro de emparelhamento ou do produto do gene o que confere susceptibilidade do indivíduo para o cancro do cólon.

Um processo pode compreender a determinação se a diferença entre o segmento amplificado e o segmento análogo causa um fenotipo afectado, isto é, se a alteração da sequência afecta a capacidade do indivíduo para reparar os erros de emparelhamento do ADN.

O processo de diagnóstico pode incluir as etapas de: 1) transcrição reversa de toda ou parte de uma cópia de ARN do gene de reparação do erro de emparelhamento do ADN; e 2) amplificação de um segmento do ADN produzido por transcrição reversa. Uma etapa de amplificação pode compreender: a selecção de um par de iniciadores de oligonucleótidos capaz de hibridar com as estruturas

helicoidais opostas do gene de reparação do erro de emparelhamento, numa orientação oposta; e realizar uma reacção em cadeia de polimerase utilizando iniciadores de oligonucleótidos tais que o ácido nucleico da cadeia de reparação do erro de emparelhamento que intervém entre os iniciadores seja amplificado para se tornar no segmento amplificado.

O gene de reparação do erro de emparelhamento do ADN é o *hMLH1* ou o *hPMS1*. O segmento de ADN corresponde a uma única porção de uma sequência de nucleótido seleccionada no grupo que consiste nas SEQ. ID NOS: 6-24. Utilizam-se iniciadores de oligonucleótidos do "primeiro estágio" seleccionados no grupo que consiste nas SEQ. ID NOS: 44-82, na RCP para amplificar o segmento de AND are. Podem-se utilizar os iniciadores de oligonucleótidos incluídos no "segundo estágio" (SEQ. ID NOS: 83-122), com os iniciadores do primeiro estágio para permitir uma amplificação mais específica e a conservação do ADN de matriz.

Um processo de identificação e de classificação de um tumor que resulta de um defeito do gene reparador do erro de emparelhamento do ADN, pode compreender a detecção, no tumor, de uma mutação do gene de reparação do erro de emparelhamento ou do produto do gene, preferencialmente um homólogo de *mutL* (*hMLH1* ou *hPMS1*), sendo a mutação indicativa de um defeito no sistema de reparação do erro de emparelhamento do tumor.

Uma estrutura de nucleótido ou de proteína isolada pode incluir um segmento que corresponde, sob o ponto de vista da sequência, a uma única porção de um gene humano homólogo de *mutL* ou um produto do gene, preferencialmente derivado quer de *hMLH1* ou *hPMS1*.

Podem utilizar-se os iniciadores dos oligonucleótidos em conjunto numa reacção em cadeia de polimerase para amplificar especificamente um único segmento de um gene humano homólogo de *mutL*, preferencialmente, *hMLH1* ou *hPMS1*.

Uma sonda pode incluir uma sequência de nucleótido capaz de se ligar especificamente por emparelhamento de Watson/Crick com as bases complementares numa porção do gene homólogo de *mutL* humano; e uma parte marcada ligada à sequência em que a parte marcada tem uma propriedade seleccionada no grupo que consiste em fluorescência, radioactividade e quimioluminescência.

Os requerentes também isolaram e sequenciaram genes *MLH1* (*mMLH1*) e *PMS1* (*rPMS1*) de rato. Os requerentes utilizaram os seus conhecimentos dos genes de reparação de semparelhamentos em ratos para construir modelos de animais para o estudo do cancro. Os modelos serão úteis para identificar agentes oncogénicos adicionais e para estudar efeitos ambientais na mutagénese.

Os requerentes produziram anticorpos policlonais dirigidos à porção da proteína codificada pelo ADNc de *rPMS1*. Os anticorpos também reagem com a proteína de *hPMS1* e são úteis e são úteis para a detecção da presença da proteína codificada por um gene normal de *hPMS1*. Os requerentes também estão a produzir anticorpos monoclonais dirigidos a *hMLH1* ou *hPMS1*.

Para além das utilizações terapêuticas e de diagnóstico, o conhecimento dos requerentes sobre *hMLH1* ou *hPMS1* pode ser utilizado para investigar outros genes com funções relacionadas que são candidatos a desempenharem um papel em certas formas de cancro humano.

Descrição das figuras

A figura 1 é um diagrama de fluxo que mostra uma visão global da sequência das etapas experimentais que os requerentes utilizaram para isolar, caracterizar e utilizar genes de *PMS1* e *MLH1* humanos e de rato.

A figura 2 é um alinhamento de sequências de proteínas para os homólogos de *mutL* (SEQ. ID NOS: 1-3) mostrando duas regiões altamente conservadas (sublinhadas) que são utilizadas para criar oligonucleótidos de RCP degenerados para o isolamento adicional de homólogos de *mutL*.

A figura 3 mostra toda a sequência de nucleótidos de ADNc (SEQ. ID NO: 4) para o gene humano *MLH1* e a previsível sequência de aminoácidos correspondente (SEQ. ID NO: 5) para a proteína humana *MLH1*. As sequências de AND sublinhadas e as regiões de ADNc que correspondem aos iniciadores de RCP degenerados que foram originalmente utilizados para amplificar uma porção do gene *MLH1* (nucleótidos 118-135 e 343-359).

A figura 4A mostra as sequências de nucleótidos de 19 exões que correspondem colectivamente a toda a estrutura do ADNc de *MLH1*. Os exões são flanqueados pelas estruturas que rodeiam o intrão. Os sítios dos iniciadores estão sublinhados. Os exões com as suas estruturas de intrão de flanqueio correspondem às SEQ. ID NOS: 25-43.

A figura 4B mostra toda sequências de nucleótidos de pares de iniciadores que foram utilizados na RCP para amplificar os exões individuais. Os iniciadores de

amplificação do "segundo estágio" (SEQ. ID NOS: 83-122) são iniciadores "aninhados" que são utilizados para amplificar os exões alvo do produto de amplificação obtido com iniciadores de amplificação do "primeiro estágio" correspondentes (SEQ. ID NOS: 44-82). As estruturas na figura 4B correspondem às estruturas nos quadros 2 e 3.

A figura 5 é um alinhamento das sequências de aminoácidos previstas para as proteínas de *MLH1* humanas e de fungos (SEQ. ID NOS: 5 e 123). As identidades dos aminoácidos estão indicadas por caixas e os intervalos estão indicados por tracejado.

A figura 6 é uma árvore filogenética das proteínas relacionadas com *mutL*.

A figura 7 é uma fotografia com duas colunas. A primeira coluna (A) é uma metáfase dispersa que mostra a hibridação do gene *MLH1* humano do cromossoma 3. A segunda coluna (B) representa uma composição do cromossoma 3 de metáfases múltiplas dispersas alinhadas com um ideograma de um cromossoma 3 humano. A região de hibridação está indicada no ideograma por uma barra vertical.

A figura 8 é uma comparação dos cromatogramas da sequência de indivíduos afectados e não afectados que mostra a identificação de uma mutação de transição de C a T que produz uma substituição de aminoácido não conservadora na posição 44 da proteína de *MLH1* humano.

A figura 9 é um alinhamento da sequência de aminoácidos (SEQ. ID NOS: 124-131) de uma região

altamente conservada da família das proteínas MLH que rodeiam o sítio da previsível substituição de aminoácidos. Os caracteres a negrito indicam a posição previsível da serina com a substituição do aminoácido de fenilalanina nos indivíduos afectados. Também em evidência estão os resíduos de serina ou de alanina conservados nesta posição nas proteínas semelhantes a *MutL*. As bolas indicam as posições da mais alta conservação dos aminoácidos. Para a proteína *MLH1*, os pontos indicam que não se obteve a sequência. As sequências foram alinhadas tal como se descreve a seguir com referência à árvore filogenética da figura.

A figura 10 mostra toda a sequência de nucleótidos para PMS1 humano (SEQ. ID NO: 132).

A figura 11 é um alinhamento das sequências de aminoácidos previsíveis para as proteínas PMS1 humanas e de fungos (SEQ. ID NO: 133 e 134. respectivamente). As identidades dos aminoácidos estão indicadas por caixas e os intervalos estão indicados por tracejados.

A figura 12 é uma sequência parcial de nucleótido do ADNc de *MLH1* de rato ou *MLH1* de murino (SEQ. ID NO: 135).

A figura 13 é uma comparação da sequência de aminoácidos previsível para as proteínas MLH1 de murino e humanas, respectivamente).

A figura 14 mostra a sequência de nucleótidos do ADN para *PMS1* de rato (*rPMS1*) (SEQ. ID NO: 137).

A figura 15 é uma comparação das previsíveis sequências de aminoácidos para as proteínas de rPMS1 e HPMS1 (SEQ. ID NOS: 138 e 133, respectivamente).

Definições

Gene - "Gene" significa uma sequência de nucleótidos que contém uma sequência de codificação completa. Geralmente, "genes" também inclui sequências de nucleótidos encontradas a montante (por exemplo, sequências de promotores, melhoradores, etc.) ou a jusante (por exemplo, os sinais de terminação da transcrição, sítios de poliadenilação, etc.) da sequência de codificação que afectam a expressão do polipéptido codificado.

Produto do gene - Um "produto do gene" é uma cópia quer de ADN ou de ARN (ARNm) de uma porção de um gene ou uma sequência de aminoácido correspondente traduzidas no ARNm.

De tipo selvagem - A expressão "de tipo selvagem", quando aplicada a ácidos nucleicos e proteínas, significa uma versão de um ácido nucleico ou de uma proteína que funciona de uma forma indistinta de uma versão normal, de ocorrência natural desse ácido nucleico ou dessa proteína (isto é, um ácido nucleico ou uma proteína com actividade de tipo selvagem). Por exemplo, um alelo de "tipo selvagem" de um gene de reparação de um erro de emparelhamento é capaz de substituir funcionalmente uma cópia endógena, normal do mesmo gene dentro de uma célula hospedeira sem alterar de forma detectável a reparação de um erro de emparelhamento nessa célula. Diferentes versões de tipo

selvagem do mesmo ácido nucleico ou proteína podem ou não diferir estruturalmente umas das outras.

De tipo não selvagem - A expressão "de tipo não selvagem", quando aplicada a ácidos nucleicos e proteínas, significa que funciona de uma maneira distinta de uma versão normal, de ocorrência natural desse ácido nucleico ou dessa proteína. Alelos de tipo não selvagem de um ácido nucleico podem diferir estruturalmente de alelos de tipo selvagem do mesmo ácido nucleico em qualquer uma das variedades de vias, incluindo, mas não se limitando a diferenças na sequência de aminoácidos de um polipéptido codificado e/ou diferenças nos níveis de expressão de um transcrito de nucleótido codificado do produto do polipéptido.

Por exemplo, a sequência de nucleótidos de um alelo de tipo não selvagem de um ácido nucleico podem diferir de um alelo de tipo selvagem, por exemplo, por adição, eliminação, substituição e/ou re-arranjo de nucleótidos. Do mesmo modo, a sequência de aminoácidos de uma proteína de reparação de um erro de emparelhamento, de tipo não selvagem, pode diferir da proteína de reparação de um erro de emparelhamento, de tipo selvagem por exemplo, por adição, eliminação, substituição e/ou re-arranjo dos aminoácidos.

Os ácidos nucleicos ou proteínas particulares de tipo não selvagem que, quando introduzidos numa célula hospedeira normal, interferem com o processo endógeno de reparação de um erro de emparelhamento, são designados por ácidos nucleicos ou proteínas "negativos dominantes".

Homólogo - O termo "homólogo" refere-se a ácidos nucleicos ou polipéptidos que estão altamente relacionados com o nível de sequências de nucleótidos ou de aminoácidos. Os ácidos nucleicos ou polipéptidos que são homólogos uns dos outros são designados por homólogos.

O termo homólogo refere-se necessariamente a uma comparação entre duas sequências. Duas sequências de nucleótidos consideram-se como homólogas se os polipéptidos que elas codificam são pelo menos cerca de 50-60 % idênticos, preferencialmente cerca de 70 % idênticos, pelo menos numa extensão de pelo menos 20 aminoácidos. Preferencialmente, as sequências de nucleótidos homólogas são caracterizadas pela capacidade de codificar uma extensão de pelo menos 4-5 aminoácidos especificados de uma forma única. Tanto a identidade como o espaçamento aproximado destes aminoácidos relativamente uns aos outros devem ser considerados para que as sequências de nucleótidos sejam consideradas homólogas. Para as sequências de nucleótidos com um comprimento inferior a 60 nucleótidos, a homologia é determinada pela capacidade de codificar uma extensão de pelo menos 4-5 aminoácidos especificados de uma forma única.

A montante/a jusante - A expressão "a montante" e "a jusante" são termos compreendidos na técnica como referindo-se à posição de um elemento de uma sequência de nucleótidos. "A montante" significa um elemento que é mais 5' do que o elemento de referência. "A jusante" refere-se a um elemento que é mais 3' do que o elemento de referência.

Intrão/exão - Os termos "intrão" e "exão" são termos compreendidos na técnica como referindo-se a várias porções das sequências genómicas do gene. "Exões" são as porções de uma sequência de gene genómico que codifica a proteína. "Intrões" são sequências de nucleótidos encontradas entre os exões nas sequências genómicas do gene.

Afectado - O termo "afectado", tal como se utiliza aqui, refere-se aos membros de uma família que ou desenvolveu um cancro característico (por exemplo, cancro do cólon numa linhagem de CCNPH) e/ou é previsível, com base, por exemplo, em estudos genéticos, que comporte uma mutação que confira susceptibilidade ao cancro.

Único - Um segmento, fragmento ou porção de um gene ou de uma proteína "único/a" significa uma porção de um gene ou de uma proteína que é diferente, sob o ponto de vista da sequência, de qualquer outro segmento, fragmento ou porção de um gene ou de uma proteína no genoma de um indivíduo. Como matéria prática, um fragmento ou segmento único de um gene será normalmente um nucleótido de pelo menos cerca de 13 bases de comprimento e será suficientemente diferente de outros segmentos do gene de tal modo que os iniciadores de oligonucleótidos possam ser preparados e utilizados para amplificar selectivamente e especificamente o segmento. Um segmento único de uma proteína é normalmente uma sequência de aminoácido que pode ser traduzida de um segmento único de um gene.

Referências

As publicações que se seguem são referidas por um número no texto da descrição. Cada uma das publicações é incorporada aqui como referência.

1. **Fishel, R., et al.** Cell 75, 1027-1038 (1993).
2. **Leach, F., et al.** Cell 75, 1215-1225 (1993).
3. **Lindblom, A., Tannergard, Pl, Werelius, B. & Nordenskjold, M.** Nature Genetics 5, 279-282 (1993).
4. **Prolla, T. A., Christie, D. M. & Liskay, R. M.** Molec. and Cell. Biol. 14, 407-415 (1994).
5. **Strand, M. Prolla, T. A., Liskay, R. M. & Petes, T. D.** Nature 365, 274-276 (1993).
6. **Aaltonen, L. A., et al.** Science 260, 812-816 (1993).
7. **Han, H. J., Yanagisawa, A., Kato, Y., Park, J. G. & Nakamura, Y.** Cancer 53, 5087-5089 (1993).
8. **Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M.** Nature 363, 558-561 (1993).
9. **Risinger, J. I. et al.** Cancer 53, 5100-5103 (1993).
10. **Thibodeau, S. N., Bren, G. & Shaid, D.** Science 260, 816-819 (1993).
11. **Levinson, G. & Gutman, G. A.** Nucleic Acids Res. 15, 5323-5338 (1987).
12. **Parsons, R., et al.** Cell 75, 1227-1236 (1993).
13. **Modrich, P.** Ann. Rev. of Genet. 25, 229-53 (1991).
14. **Reenan, R. A. & Kolodner, R. D.** Genetics 132, 963-73 (1992).
15. **Bishop, D. K., Anderson, J. & Kolodner, R. D.** PNAS 86, 3713-3717 (1989).
16. **Kramer, W., Kramer, B., Williamson, M. S. & Fogel, S. J.** Bacteriol. 171, 5339-5346 (1989).

17. **Williamson, M. S., Game, J. C. & Fogel, S.,** Genetics 110, 609-646 (1985).
18. **Prudhomme, M., Martin, B., Mejean, V. & Claverys, J. J.** Bacteriol. 171, 5332-5338 (1989).
19. **Mankovich, J. A., McIntyre, C. A. & Walker, G. C.** J. Bacteriol. 171, 5325-5331 (1989).
20. **Lichter, P., et al.** Science 247, 64-69 (1990).
21. **Boyle, A., Feltquite, D. M., Dracopoli, N., Housman, D. & Ward, D. C.** Genomics 12, 106-115 (1992).
22. **Lyon, M. F. & Kirby, M. C.,** Mouse Genome 91, 40-80 (1993).
23. **Reenan, R. A. & Kolodner, R. D.** Genetics 132, 975-85 (1992).
24. **Latif, F. et al.** Cancer Research 52, 1451-1456 (1992).
25. **Naylor, S. L., Johnson, B. E., Minna, J. D. & Sakaguchi, A. Y.** Nature 329, 451-454 (1987).
26. **Ali, I. U., Lidereau, R. & Callahan, R.** Journal of the National Cancer Institute 81, 1815-1820 (1989).
27. **Higgins, D., Bleasby, A. & Fuchs, R.** Comput. Apple Biosci. 8, 189-191 (1992).
28. **Fields, S. & Song, O. K.** Nature 340, 245-246 (1989).
29. **Lynch, H. T., et al.** Gastroenterology 104, 1535-1549 (1993).
30. **Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M. & Davis, R. W.** Proc. Natl. Acad. Sci. U.S.A. 88, 1731-1735 (1991).
31. **Frohman, M.** Amplifications, a forum for PCR users 1, 11-15 (1990).
32. **Powell, S. M., et al.** New England Journal of Medicine 329, 1982-1987 (1993).
33. **Wu, D. Y., Nozari, G. Schold, M., Conner, B. J. & Wallace, R. B.** DNA 8, 135-142 (1989).

34. **Mullis, K. E. B. & Faloona, F. A.** Methods in Enzymology 155, 335-350 (1987).
35. **Bishop, T. D., Thomas, H.** Cancer Sur. 9, 585-604 (1990).
36. **Capecchi, M. R.** Scientific American 52-59 (March 1994).
37. **Erlich, H. A.** PCR Technology, Principles and Applications for DNA Amplification (1989).
38. **Papadopoulos et al.** Science 263, 1625-1629 (March 1994).
39. **Nicolaides et al.** Nature 371, 75-80 (September 1994).
40. **Tong et al.** Anal. Chem. 64, 2672-2677 (1992).
41. **Debuire et al.** Clin. Chem. 39, 1682-5 (1993).
42. **Wahlberg et al.** Electrophoresis 13, 547-551 (1992).
43. **Kaneoka et al.** Biotechniques 10, 30, 32, 34 (1991).
44. **Huhman et al.** Biotechniques 10, 84-93 (1991).
45. **Hultman et al.** Nuc. Acid. Res. 17, 4937-46 (1989).
46. **Zu et al.** Mutn. Res. 288, 232-248 (1993).
47. **Espelund et al.** Biotechniques 13, 74-81 (1992).
48. **Prolla et al.** Science 265, 1091-1093 (1994).
49. **Bishop et al.** Mol. Cell. Biol. 6, 3401-3409 (1986).
50. **Folger et al.** Mol. Cell. Biol. 5, 70-74 (1985).
51. **T. C. Brown et al.** Cell 54, 705-711 (1988).
52. **T. C. Brown et al.** Genome 31, 578-583 (1989).
53. **C. Muster-Nassal et al.** Proc. Natl. Acad. Sci. U.S.A. 83, 7618-7622 (1986).
54. **I. Varlet et al.** Proc. Natl. Acad. Sci. U.S.A. 87, 7883-7887 (1990).
55. **D. C. Thomas et al.** J. Biol. Chem. 266, 3744-3751 (1991).

56. **J. J. Holmes et al.** Proc. Natl. Acad. Sci. U.S.A. 87, 5837-5841 (1990).
57. **P. Branch et al.** Nature 362, 652-654 (1993).
58. **A. Kat et al.** Proc. Natl. Acad. Sci. U.S.A. 90, 6424-6428 (1993).
59. **K. Wiebauer et al.** Nature 339, 234-236 (1989).
60. **K. Wiebauer et al.** Proc. Natl. Acad. Sci. U.S.A. 87, 5842-5845 (1990).
61. **P. Neddermann et al.** J. Biol. Chem. 268, 21218-24 (1993).
62. **Kramer et al.** Mol. Cell Biol. 9:4432-40 (1989).
63. **Kramer et al.** J. Bacteriol. 171:5339-5346 (1989).

Descrição da invenção

Os requerentes encontraram genes de mamíferos que estão envolvidos na reparação de um erro de emparelhamento do ADN. Um dos genes, o *hPMS1*, codifica uma proteína que é homóloga de uma proteína de reparação de um erro de emparelhamento do ADN de uma levedura, *PMS1*. Os requerentes mapearam as localizações de *hPMS1*, no cromossoma 7 e humano e no gene *PMS1* de rato no cromossoma 5 de rato, na banda G. Outro gene, o *hMLH1* (homólogo de *MutL*) codifica uma proteína que é homóloga da proteína de *MLH1* (*MutL*reparação de um erro de emparelhamento do ADN de um fungo. Os requerentes mapearam as localizações de *hMLH1* no cromossoma humano 3p21.3-23 e no cromossoma 9 de rato, na banda E.

Estudos^{1, 2} demonstraram o envolvimento de um homólogo de um gene de reparação de um erro de emparelhamento do ADN humano, *hMSH2*, no cromossoma 2p em CCNPH. Com base nos dados de ligação foi assinalado um segundo locus de CCNPH no cromossoma 3p21.3-23³. O exame do ADN do tumor das famílias ligadas ao cromossoma 3 revelaram uma

instabilidade na repetição do nucleótido semelhante à observada noutras famílias⁶ com CCNPH e vários tipos de tumores esporádicos⁷⁻¹⁰. Como a instabilidade da repetição dos dinucleótidos é característica de um defeito na reparação^{5, 11, 12} de um erro de emparelhamento do ADN, os requerentes pensaram que CCNPH ligado ao cromossoma 3p21.3-23 podia resultar de uma mutação num segundo gene de reparação de um erro de emparelhamento do ADN.

A reparação de um erro de emparelhamento do ADN na *Escherichia coli* requer um certo número de genes incluindo defeitos de *mutS*, *mutL* e *mutH*, em qualquer um deles, o que resulta em taxas¹³ elevadas de mutação espontânea. A análise genética no fungo *Saccharomyces cerevisiae* identificou três genes de reparação de um erro de emparelhamento do ADN: um homólogo de *mutS*, *MSH2*¹⁴, e dois homólogos *mutL*, *PMS1*¹⁶ e *MLH1*⁴. Cada um destes três genes desempenha um papel indispensável na fidelidade da replicação do ADN, incluindo a estabilização das repetições de dinucleótidos⁵.

Os requerentes acreditam que *hMLH1* é o gene do CCNPH previamente ligado ao cromossoma 3p, com base na semelhança do produto do gene *hMLH1* com a proteína de reparação de um erro de emparelhamento do ADN, *MLH1*⁴, na localização coincidente do gene *hMLH1* e o locus de CCNPH no cromossoma 3 e as mutações de *hMLH1* sem sentido que os requerentes verificaram em indivíduos afectados de famílias de CCNPH ligadas ao cromossoma 3.

Os conhecimentos dos requerentes sobre a estrutura dos genes *MLH1* e *PMS1* humanos e de rato tem muitas utilizações importantes. A informação sobre a sequência dos genes pode ser utilizada para avaliar indivíduos quanto ao risco de

cancro. O conhecimento sobre a estrutura dos genes torna possível preparar facilmente iniciadores de RCP que podem ser utilizados para amplificar selectivamente porções dos genes *hMLH1* e *hPMS1* para comparações subsequentes com a sequência normal e análise do risco de cancro. Este tipo de ensaio também torna possível pesquisar e caracterizar mutações de *hMLH1* e *hPMS1* ligadas ao cancro com o fim de eventualmente focar o esforço de avaliação em locus específicos dos genes. A caracterização específica das mutações ligadas ao cancro em *hMLH1* e *hPMS1*, torna possível a produção de outras ferramentas de diagnóstico valiosas tais como sondas específicas de alelos que podem ser utilizadas em testes de avaliação para determinar a presença ou a ausência de mutações específicas dos genes.

Adicionalmente, a informação sobre a sequência do gene para *hMLH1* e/ou *hPMS1*, pode ser utilizada, por exemplo, num sistema de dois híbridos, para pesquisar outros genes com funções relacionadas que são candidatos para o envolvimento no cancro.

As estruturas dos genes *hMLH1* e *hPMS1* são úteis para a produção de proteínas que são utilizadas para desenvolver anticorpos dirigidos a porções específicas ou a toda a proteína *hMLH1* e *hPMS1*. Esses anticorpos podem então ser utilizados para isolar as proteínas correspondentes e possivelmente as proteínas relacionadas para fins de investigação e de diagnóstico.

As sequências dos genes *MLH1* e *PMS1* de rato são úteis para a produção de ratos com mutações nos respectivos genes. Os ratos mutantes são úteis para o estudo da função dos genes, particularmente na sua relação com o cancro.

Processos para o isolamento e caracterização de genes de *MLH1* e *PMS1* de mamíferos

Os requerentes isolaram e caracterizaram quatro genes de mamíferos, isto é, *MLH1* (*hMLH1*), *PMS1* humano (*hPMS1*), *MLH1* de rato (*rPMS1*) e *PMS1* de rato (*rPMS1*). Devido à semelhança estrutural entre estes genes, os processos utilizados pelos requerentes para isolá-los e caracterizá-los são geralmente os mesmos. A figura 1 mostra, em termos amplos, a abordagem experimental que os requerentes utilizaram para isolar e caracterizar os quatro genes. A discussão que se segue refere-se aos processos, nas suas várias etapas, mostrados na figura 1.

Etapa 1 Preparação de conjuntos de oligonucleótidos degenerados para RCP

Relatórios anteriores indicaram que porções de três proteínas semelhantes a MutL, duas de bactérias, MutL e HexB, e uma de fungos, PMS1, estão altamente conservadas^{16, 18, 19}. Depois da inspecção das sequências de aminoácidos das proteínas HexB, MutL e PMS1, como se mostra na figura 2, os requerentes prepararam conjuntos de pares de oligonucleótidos degenerados correspondendo a duas regiões altamente conservadas KELVEN e GFRGEA, das proteínas semelhantes a MutL. As sequências (SEQ. ID NOS: 135 e 136, respectivamente) dos oligonucleótidos degenerados que foram utilizados para isolar os quatro genes são:
5'-CTTGATTCTAGAGC(T/C)TCNCCNC(T/G)(A/G)AANCC-3' e
5'-AGGTCGGAGCTCAA(A/G)GA(A/G)(T/C)TNGTNGANAA-3'.

As sequências sublinhadas dentro dos iniciadores são os sítios de endonucleases de restrição *XbaI* e *SacI*, respectivamente. Foram introduzidos de modo a facilitar a

clonagem de fragmentos amplificados de RCP. Na preparação dos oligonucleótidos, os requerentes tiveram em conta o facto de um dado aminoácido poder ser codificado por mais do que um triplete de ADN (codão). A degenerescência dentro destas sequências está indicada por múltiplos nucleótidos dentro dos parêntesis ou N, para a presença de qualquer base nessa posição.

Etapa 2 Transcrição reversa e RCP em ARNm seleccionado de poli A+ isolado de células humanas

Os requerentes isolaram ARN mensageiro (enriquecido com poli A+) a partir de células humanas de cultura, ADNc de estrutura helicoidal dupla sintetizado a partir de ARNm e realizaram a RCP com os oligonucleótidos degenerados ⁴. Depois de tentarem um certo número de condições diferentes de RCP, por exemplo, ajustando a temperatura de têmpera, os requerentes amplificaram com sucesso um ADN da dimensão prevista (~210 pb) para uma proteína semelhante a MutL.

Etapa 3 Clonagem e sequenciação de fragmentos gerados por RCP; identificação de dois fragmentos de genes representando *MLH1* e *PMS1* humanos

Os requerentes isolaram material amplificado por RCP (~210 pb) de um gel de agarose e clonaram este material num plasmido (pUC19). Determinaram a sequência de ADN de vários clones diferentes. A sequência de aminoácidos inferida da sequência de ADN dos dois clones mostrou semelhanças com outras proteínas conhecidas semelhantes a MutL ^{4, 16, 18, 19}. A sequência de aminoácidos prevista para um dos clones foi a mais parecida com a proteína de fundo *PMS1*. Por isso, os requerentes designaram-na por *hPMS1*, para a *PMS1* humana. Verificou-se que o segundo clone codificava um polipéptido

que se parecia de mais perto com a proteína de fungo MLH1 e foi designada por *hMLH1*, para *MLH1* humana.

Etapa 4 Isolamento dos clones de ADNc de *PMS1* e *MLH1* humanos e de rato utilizando fragmentos da RCP como sondas

Os requerentes utilizaram fragmentos de 210 pb de RCP dos ADNc de *hPMS1* e *hMLH1*, como sondas para avaliar ambas as bibliotecas de ADNc humano e de rato (da Stratagene, ou tal como descrito na referência 30). Isolou-se um certo número de ADNcs que correspondiam a esses dois genes. Muitos dos ADNcs estavam truncados na extremidade 5'. Quando necessário, utilizaram-se técnicas de RCP³¹ para se obter a extremidade 5' do gene para além de outra avaliação das bibliotecas de ADNc. As sequências completas do ADNc composto foram utilizadas para prever as sequências de aminoácidos das proteínas *PMS1* e *MLH1* humanas e de ratos.

Etapa 5 Isolamento dos clones genómicos de *PMS1* e *MLH1* humanas e de ratos

As informações sobre a estrutura do ADN genómico e o ADNc dos genes de *PMS1* e *MLH1* humanas são necessárias de modo a avaliar cuidadosamente quanto a mutações em famílias predispostas para o cancro. Os requerentes utilizaram sequências de ADNc como sondas para isolar as sequências genómicas de *PMS1* e *MLH1* humanas. Isolaram quatro cosmidos e dois clones P1 para *hPMS1*, que, em conjunto, é provável que contenham a maior parte, se não toda a sequência do ADNc (exão). Para o *hMLH1* os requerentes isolaram quatro clones de fagos λ sobrepostos contendo as sequências genómicas 5'-*MLH1* e quatro clones P1 (dois clones de comprimento completo e dois que incluem a extremidade de codificação 5' mais porções da região do promotor). A

análise de Rcp utilizando pares de oligonucleótidos específicos para as extremidades 5' e 3' do ADNc de *hMLH1*, indica claramente que o clone P1 contém a informação completa do ADNc completo de *hMLH1*. Do mesmo modo, os clones genômicos para os genes de *PMS1* e *MLH1* de rato foram isolados e parcialmente caracterizados (descritos na etapa 8).

Etapa 6 Mapeamento posicional dos cromossomas dos genes *PMS1* e *MLH1* humanos e de ratos por meio de hibridação *in situ* de fluorescência

Os requerentes isolaram clones genômicos isolados a partir de genes *PMS1* e *MLH1* humanos e de ratos para a localização cromossômica por hibridação *in situ* de fluorescência (FISH)^{20, 21}. Os requerentes mapearam o gene *MLH1* humano com o cromossoma 3p21.3-23, mostrado na FIG. 7, tal como se discute com mais detalhe a seguir. Os requerentes mapearam o gene *MLH1* de rato para o cromossoma 9, banda E, uma região de sintonia entre o de rato e o humano²². Para além das técnicas de FISH, os requerentes utilizaram RCP com um par de oligonucleótidos específicos de *hMLH1* para analisar o ADN a partir de um painel de mapeamento de híbridos de células somáticas de roedores/seres humanos (Coriell Institute for Medical Research, Camden, N.J.). Os resultados da RCP obtidos pelos requerentes indicam que o *hMLH1* mapeia para o cromossoma 3. A posição de 3p21.3-23 de *hMLH1* é coincidente com uma região conhecida por arvorar um segundo locus para CCNPH com base nos dados de ligação.

Os requerentes mapearam o gene *hPMS1*, tal como se mostra na figura 12, no braço longo (q) do cromossoma 7 (quer 7q11 ou 7q22) e o *PMS1* de rato do cromossoma 5, banda

G, duas regiões de sintonia entre os de rato e os humanos²². Os requerentes realizaram a RCP utilizando oligonucleótidos específicos para *hPMS1* no ADN de um painel de células de roedor/ser humano. Em acordo com os dados de FISH, a localização de *hPMS1* foi confirmada como sendo no cromossoma 7. Estas observações asseguram aos requerentes que a sua posição no mapa humano para *hPMS1* no cromossoma 7 está correcta. A localização física de *hPMS1* é útil para o fim da identificação de famílias que podem ter potencialmente um cancro ligado a uma mutação em *hPMS1*.

Etapas 7 Utilização das sequências do ADN genómico e do ADNc para identificar mutações nos genes *hPMS1* e *hMLH1* para famílias com CCNPH.

Os requerentes analisaram amostras recolhidas de indivíduos em famílias CCNPH com o fim de para identificar mutações nos genes *hPMS1* e *hMLH1*. A abordagem dos requerentes consiste em preparar iniciadores de RCP com base no seu conhecimento das estruturas dos genes, para se obter segmentos de exão/intrão que os requerentes puderam comparar com as sequências normais conhecidas. Os requerentes referem esta abordagem como uma "avaliação do exão".

Utilizando a informação sobre as sequências de ADNc os requerentes prepararam e continuam a preparar oligonucleótidos específicos de *hPMS1* e *hMLH1* para delinear as fronteiras do exão/intrão dentro das sequências genómicas. Os oligonucleótidos específicos de *hPMS1* e *hMLH1* foram utilizados para sondar clones genómicos quanto à presença de exões contendo essas sequências. Os oligonucleótidos que hibridam foram utilizados como iniciadores para a sequenciação do ADN dos clones

genómicos. As junções exão/intrão foram identificadas comparando as sequências do ADN genómico com as do ADNc.

A amplificação de exões específicos do ADNc por meio de RCP e sequenciação dos produtos é um dos processos para avaliar as famílias CCNPH quanto às mutações ^{1, 2}. Os requerentes identificaram clones genómicos contendo informação sobre o ADNc de *hMLH1* e determinaram as estruturas de todas as regiões de fronteira do exão/intrão que flanqueiam os 19 exões de *hMCH1*.

Os requerentes utilizaram a abordagem da avaliação do exão para examinar o gene de *MLH1* de indivíduos de famílias com CCNPH mostrando uma ligação com o cromossoma 3 ³. Como será discutido com mais detalhe a seguir, os requerentes identificaram uma mutação no gene *MLH1* dessa família, consistindo numa substituição de C a T. Os requerentes previram que a mutação de C a T causa uma substituição de serina por fenilalanina numa região altamente conservada da proteína. Os requerentes continuam a identificação de famílias com CCNPH a partir das quais podem obter amostras de modo a encontrar mutações adicionais nos genes *hMLH1* e *hPMS1*.

Os requerentes estão também a utilizar uma segunda abordagem para identificar mutações em *hMLH1* e *hPMS1*. Esta abordagem consiste em preparar iniciadores de oligonucleótidos específicos de *hPMS1* ou *hMLH1* para produzir um ADNc com uma primeira estrutura helicoidal por transcrição reversa de ARN. A RCP, utilizando iniciadores específicos dos genes permitirá aos requerentes amplificar regiões específicas destes genes. A sequenciação do ADN dos fragmentos amplificados permitirá aos requerentes detectar mutações.

Etapa 8 Preparação de vectores para atingirem alvos para quebrar os genes de *PMS1* ou *MLH1* nas células ES; estudos em ratos deficientes na reparação de erros de emparelhamento.

Os requerentes construíram um vector para atingir um gene com base no seu conhecimento da estrutura do ASN genómico de *PMS1*. Os requerentes utilizaram o vector de *PMS1* para romper o gene em células embrionárias do estema de ratos ³⁶. As células foram injectadas em blastócitos de ratos que se desenvolveram em ratos que são quiméricos (misturas) para as células que comportam a mutação de *PMS1*. Os animais quiméricos serão utilizados para criar murganhos que são heterozigóticos e homozigóticos para a mutação de *PMS1*. Estes murganhos serão úteis para o estudo do papel do gene *PMS1* em todo o organismo.

***MLH1* humano**

A discussão que se segue é uma explicação mais detalhada de trabalho experimental dos requerentes com *MLH1*. Tal como foi mencionado ante, para clonar genes *MLH* de mamíferos, os requerentes utilizaram técnicas de RCP tal como as utilizadas para identificar os genes *MSH1*, *MSH2* e *MLH1* de fungos e o gene *MSH2* humano ^{1, 2, 4, 14}. Como matriz na RCP, os requerentes utilizaram ADNc de estrutura helicoidal dupla sintetizada a partir de ARN enriquecido com poli (A+) preparado a partir de fibroblastos humanos primários de cultura. Os oligonucleótidos degenerados foram atingidos nas sequências de aminoácidos de terminação N KELVEN e GFRGEA (ver FIG. 3), duas das regiões mais conservadas da família de MutL de proteínas previamente descrita para as bactérias e para os fungos ^{16, 18,19}. Identificaram-se dois produtos de RCP com a dimensão prevista, clonaram-se e eles mostraram codificar uma

sequência de aminoácidos prevista com homologia das proteínas semelhantes a MutL. Estes dois fragmentos gerados por RCP foram utilizados para isolar clones de ADNc humano e ADN genómico.

Os iniciadores de oligonucleótidos que os requerentes utilizaram para amplificar as sequências relacionadas com MutL humano foram 5'-CTTGATTCTAGAGC(T/C)TCNCCNC(T/G)(A/G)AANCC-3' (SEQ. ID NO: 135) e 5'-AGGTCGGAGCTCAA(A/G)GA(A/G)(T/C)TNGTNGANAA-3' (SEQ. ID NO: 136). A RCP foi realizada em misturas reacionais de 50 µL contendo uma matriz de ADNc, 1,0 µM de cada um dos iniciadores, 5 UI de Taq polimerase (C), KCl 50 mM, tampão Tris 10 mM a pH 7,5 e MgCl 1,5 mM. Realizou-se a RCP por 35 ciclos de 1 minuto a 94 °C, 1 minuto a 43 °C e 1,5 minutos a 62 °C. Os fragmentos da dimensão desejada, aproximadamente 212 pb, foram clonados em pUC19 e sequenciaram-se. Os produtos de RCP de *MLH1* clonados foram marcados com um kit de marcação aleatória de iniciadores (RadPrime, Gibco BRL) e utilizou-se para sondar o ADNc humano e bibliotecas de cosmido genómico por processos padrão. A sequenciação dos ADNs dos plasmidos de estrutura helicoidal dupla foi realizada como descrito previamente ¹.

A sequência de nucleótidos de ADNc de *hMLH1*, tal como se mostra na figura 3, codifica uma secção de leitura aberta de 2268 pb. Também se mostra na figura 3 a sequência de proteína prevista codificada pelo ADNc de *hMLH1*. As sequências de ADN sublinhadas são as regiões de ADNc que correspondem a iniciadores de RCP degenerados que foram originalmente utilizados para amplificar uma porção do gene de *MLH1* (nucleótidos 118-135 e 343-359).

A figura 4A mostra 19 sequências de nucleótidos que correspondem a porções de *hMLH1*. Cada sequência inclui um dos 19 exões, na sua totalidade, rodeados por sequências de flaqueio do intrão. Os sítios dos iniciadores alvo de RCP estão sublinhados. Mais detalhes relacionados com a derivação e utilizações das sequências mostradas na figura 4^a, tal como se estabelece a seguir.

Tal como se mostra na figura 5, a proteína *hMLH1* é constituída por 765 aminoácidos e partilha 41 % de identidade com o produto de proteína do gene de reparação do erro de emparelhamento do ADN de fungos *hMLH1* ⁴. As regiões da proteína *hMLH1* mais semelhantes a *MLH1* de fungos correspondem aos aminoácidos 11 a 317, mostrando 55 % de identidade e os últimos 13 aminoácidos que são idênticos entre as duas proteínas. A figura 5 mostra um alinhamento das sequências da prevista *hMLH1* e da proteína *MLH1* de *S. cerevisiae*. As identidades dos aminoácidos estão indicadas pelas caixas e os intervalos estão indicados pelos tracejados. O alinhamento da sequência de proteínas do elemento do par foi realizado com o DNASTar MegAlign utilizando o processo clustal ²⁷. Os parâmetros de alinhamento dos elementos do par foram um ktupre de 1, perda do intervalo de 3, janela de 5 e diagonais de 5. Além disso, tal como se mostra na figura 13, as sequências de aminoácidos previstas das proteínas *MLH1* humanas e de rato mostram pelo menos 74 % de identidade.

A figura 6 mostra uma árvore filogenética de proteínas relacionadas com MutL. A árvore filogenética foi construída utilizando as sequências de aminoácidos previstas de 7 proteínas relacionadas com MutL: *MLH1* humana; *MLH1* de rato; *MLH1* de *S. cerevisiae*; *PMS1* de *S. cerevisiae*; *E. coli*; *MutL*; *MutL* de *S. typhimurium* e *HexB* de *S. pneumoniae*. As

sequências necessárias foram obtidas do Genebank da versão 7.3. A árvore filogenética foi gerada com o programa PILEUP do software da Genetics Computer Group utilizando uma perda de intervalo de 3 a uma perda de comprimento de 0,1. As sequências de ADN registadas de *hMLH1* e *hPMS1* tinham sido submetidas ao GenBank.

Localização do intrão de *hMLH1* e estruturas das fronteiras do intrão/exão

No pedido de patente de invenção norte-americana anterior com o nº 08/209.521 dos requerentes, estes descreveram a sequência de nucleótidos de um clone de ADN complementar (ADNc) de um gene humano, *hMLH1*. A sequência de ADNc de *hMLH1* (SEQ. ID NO: 4) é apresentada no presente pedido de patente de invenção na figura 3. Os requerentes fazem notar que pode haver algumas variações entre as estruturas individuais do ADNc de *hMLH1*, resultantes dos polimorfismos dentro da população humana e a degenerescência do código genético.

No presente pedido de patente de invenção, os requerentes reportam os resultados dos seus estudos de sequenciação genómica. Especificamente, os requerentes clonaram a região genómica humana que inclui o gene *hMLH1*, com um foco específico nos exões individuais e nas estruturas das fronteiras do intrão/exão que os rodeiam. Tendo em vista este último objectivo de desenhar uma abordagem compreensiva e eficiente para identificar e caracterizar as mutações que conferem susceptibilidade ao cancro, os requerentes acreditam que é importante conhecer as sequências de tipo selvagem das estruturas do intrão que flanqueiam os exões no gene *hMLH1*. Uma vantagem em conhecer a sequência de intrões próxima das fronteiras dos exões, é

que isso torna possível preparar os pares de iniciadores para amplificar selectivamente os exões individuais de uma forma total. Mais importante ainda, é também possível que a mutação numa região do intrão, que, por exemplo, pode causar um erro no seccionamento do ARNm, podia resultar num produto de gene defeituoso, isto é, com susceptibilidade ao cancro, sem mostrar qualquer anomalia numa região do exão do gene. Os requerentes acreditam que uma abordagem de avaliação compreensiva requer a investigação de mutações, não apenas no exão ou no ADNc, mas também nas estruturas do intrão que flanqueiam as fronteiras dos exões.

Os requerentes clonaram a região genómica humana que inclui *hMLH1* utilizando abordagens que são conhecidas na técnica e outras abordagens conhecidas podiam ter sido utilizadas. Os requerentes utilizaram RCP para avaliar uma biblioteca genómica humana de P1 para o gene *hMLH1*. Os requerentes obtiveram quatro clones, dois que continham o gene completo e dois a que faltavam as terminações C. Os requerentes caracterizaram um dos clones de comprimento completo pelo ciclo de sequenciação, o que resultou na sua definição de todas as sequências de junção de intrões/exões para ambos os lados dos 19 exões de *hMLH1*. Prepararam assim múltiplos conjuntos de iniciadores de RCP para amplificar cada exão individual (iniciadores do primeiro estágio) e verificaram a sequência de cada exão e a sequência de flanqueio do intrão por meio da amplificação de várias amostras de ADN genómico diferentes e a sequenciação dos fragmentos resultantes utilizando um sequenciador ABI 373. Além disso, os requerentes determinaram as dimensões de cada exão de *hMLH1* utilizando processos de RCP. Finalmente, os requerentes projectaram um conjunto de iniciadores de RCP "aninhados" (iniciadores do segundo estágio) para uma nova amplificação dos exões individuais. Os requerentes

utilizaram iniciadores do segundo estágio num processo multiplex para a análise de famílias com CCNPH e tumores para as mutações de *hMLH1*. Geralmente, na abordagem do iniciador de RCP "aninhado", os requerentes realizaram uma primeira amplificação em multiplex com quatro a oito conjuntos dos iniciadores "do primeiro estágio", cada um deles dirigido a um exão diferente. Depois os requerentes voltaram a amplificar os exões individuais do produto ou a primeira etapa de amplificação, utilizando um único conjunto de iniciadores do segundo estágio. Exemplos e outros detalhes relacionados com a utilização dos requerentes dos iniciadores do primeiro e do segundo estádios estão indicados a seguir.

Através de estudos de sequenciação genómica dos requerentes, eles identificaram todos os dezanove exões dentro do gene *hMLH1* e mapearam as fronteiras do intrão/exão.

O quadro 1 apresenta as coordenadas do nucleótido (isto é, o ponto de inserção de cada intrão dentro da região de codificação do gene) dos exões de *hMLH1* (SEQ. ID NOS: 25-43). As coordenadas apresentadas baseiam-se na sequência do ADNc de *hMLH1*, a que se atribui a posição "1" com o "A" do início de "ATG" (em que A é o nucleótido 1 na SEQ. ID NO: 4).

Quadro 1

Número do intrão	Coordenadas da sequência de ADNc
intrão 1	116 & 117
intrão 2	207 & 208
intrão 3	306 & 307
intrão 4	380 & 381
intrão 5	453 & 454
intrão 6	545 & 546
intrão 7	592 & 593
intrão 8	677 & 678
intrão 9	790 & 791
intrão 10	884 & 885
intrão 11	1038 & 1039
intrão 12	1409 & 1410
intrão 13	1558 & 1559
intrão 14	1667 & 1668
intrão 15	1731 & 1732
intrão 16	1896 & 1897
intrão 17	1989 & 1990
intrão 18	2103 & 2104

Os requerentes também determinaram a sequência de nucleótidos das regiões do intrão que flanqueiam os exões do gene *hMLH1*. As SEQ. ID NOS: 6-24 são sequências individuais do intrão ligadas às suas respectivas sequências de exões a montante e a jusante. As mesmas estruturas de nucleótidos estão indicadas na fig. 4A, em que os exões estão numerados desde a terminação N até à terminação C no que respeita ao locus cromossômico. Os números de 5 dígitos indicam os iniciadores utilizados para amplificar o exão. Todas as sequências estão numeradas assumindo que o A do codão de ATG é o nucleótido 1. Os números entre () são as coordenadas dos nucleótidos da

sequência correspondente encontrada no exão indicado. O expoente é o intrão. O índice é o exão ou as sequências não traduzidas encontradas no clone de ARNm/ADNc. Os índices e as sequências sublinhadas correspondem aos iniciadores. O codão de paragem em 2269-2271 está em itálico e sublinhado.

O quadro 2 apresenta as sequências de pares de iniciadores (iniciadores do "primeiro estágio") que os requerentes utilizaram para amplificar exões individuais em conjunto com as estruturas de flaqueio do intrão:

Quadro 2

EXÃO NO.	LOCALIZAÇÃO DO INICIADOR	NO. DO INICIADOR	NO DA SEQ ID DO INICIADOR	SEQUÊNCIA DO NUCLEÓTIDO DO INICIADOR
1	a montante	18442	44	5'aggcactgaggtgattggc
1	a jusante	19109	45	5'tcgtagcccttaagtgagc
2	a montante	19689	46	5'aatatgtacattagagtagttg
2	a jusante	19688	47	5'cagagaaaggtcctgactc
3	a montante	19687	48	5'agagatttggaatgagtaac
3	a jusante	19786	49	5'acaatgtcatcacaggagg
4	a montante	18492	50	5'aacctttcccttggtagg
4	a jusante	18421	51	5'gattactctgagacctaggc
5	a montante	18313	52	5'gattttctctttcccttggg
5	a jusante	18179	53	5'caaacaaagcttcaacaattac
6	a montante	18318	54	5'gggttttattttcaagtacttctatg
6	a jusante	18317	55	5'gctcagcaactgttcaatgtatgagc
7	a montante	19009	56	5'ctagtgtgtgtttttggc
7	a jusante	19135	57	5'cataacctatctccacc
8	a montante	18197	58	5'ctcagccatgagacaataaatcc
8	a jusante	18924	59	5'ggttccaaataatgtgatgg
9	a montante	18765	60	5'caaaagcttcagaatctc
9	a jusante	18198	61	5'ctgtgg gtgttcctgtgagtgg
10	a montante	18305	62	5'catgactftgtgtgaatgtacacc
10	a jusante	18306	63	5'gaggagagcctgatagaacatctg
11	a montante	18182	64	5'ggg ctttttctccccctccc
11	a jusante	19041	65	5'aaaatctgggtctctcag
12	a montante	18579	66	5'aattatacctcatactagc
12	a jusante	18178	67	5'gttttattacagaataaaggagg
12	a jusante	19070	68	5'aagccaaagttagaaggca

EXÃO NO.	LOCALIZAÇÃO DO INICIADOR	NO. DO INICIADOR	NO DA SEQ ID DO INICIADOR	SEQUÊNCIA DO NUCLEÓTIDO DO INICIADOR
13	a montante	18420	69	5'tgcaaccacaaaatttggc
13	a jusante	18443	70	5' ctttctccatttccaaaacc
14	a montante	19028	71	5'tgggtgtctctagttctgg
14	a jusante	18897	72	5' cattgttgtagtag ctctgc
15	a montante	19025	73	5' cccatttgtcccaactgg
15	a jusante	18575	74	5'cggtcagttgaaatgtcag
16	a montante	18184	75	5' catttgatgctccgttaaagc
16	a jusante	18314	76	5' caccggctggaaattttatttg
17	a montante	18429	77	5' ggaaaggcactggagaaatggg
17	a jusante	18315	78	5'ccctccagcacacatgcatgtaccg
18	a montante	18444	79.	5'taagtagtctgtgatctccg
18	a jusante	18581	80	5'atgtatgaggtcctgtcc
19	a montante	18638	81	5'gacaccagtgtatgtttgg
19	a jusante	18637	82	5' gagaagaagaacacatccc

Adicionalmente, os requerentes prepararam um conjunto de iniciadores de amplificação do "segundo estágio", cujas estruturas se indicam a seguir no quadro 3. Os requerentes utilizam os iniciadores do segundo estágio em conjunto com os iniciadores do primeiro estágio num protocolo de amplificação, tal como se descreve a seguir.

Quadro 3

EXÃO NO.	LOCALIZAÇÃO DO INICIADOR	NO. DO INICIADOR	NO DA SEQ ID DO INICIADOR	SEQUÊNCIA DE NUCLEÓTIDOS DO INICIADOR
1	a montante	19295	83	5'tgtaaaacgacggccagtcact gaggtgattggctgaa
1	a jusante	19446	84	*5'tagcccttaagtgagcccg
2	a montante	18685	85	5'tgtaaaa cgacggccagttacat tagagtgttgacaga
2	a jusante	19067	86	* 5' aggtcctg actcttcca tg
3	a montante	18687	87	5'tgtaaaacgacggccagtttggg aatgagtaacatgatt
3	a jusante	19068	88	*5'tgtcatcacaggaggatat
4	a montante	19294	89	5'tgtaaaacgacggccagttcttc ccttgggtgagtgga
4	a jusante	19077	90	*5tactctgagacctaggccca
5	a montante	19301	91	5'tgtaaaacgacggccagttctct tttcccttgggattag
5	a jusante	19046	92	*5'acaaagcttcaacaatttactc t

6	a montante	19711	93	5 'tgtaaaa cgacggccagtggttt attttcaagtacttctatgaatt
6	a jusante	19079	94	*5'cagcaactgttcaatgtatgag cact
7	a montante	19293	95	5'tgtaaaacgacggccagtggtg tggttttggaac
7	a jusante	19435	96	*5'aaccttatctccaccagc
8	a montante	19329	97	5'tgtaaaacgacggccagtagcc atgagacaataaatccttg
8	a jusante	19450	98	*5'tcccaaatatgtgatggaatg
9	a montante	19608	99	5'tgtaaaacgacggccagtaagc ttcagaatctctttt
9	a jusante	19449	100	*5 'tggtgttttctgtgagtggtt
10	a montante	19297	101	5'tgtaaaacgacggccagtgctt gtgtgaatgtacacctgtg
10	a jusante	19081	102	*5'gagagcctgatagaacatctgt tg
11	a montante	19486	103	5'tgtaaaacgacggccagtgctttt ctcccectcccacta
11	a jusante	19455	104	* 5'tctgggctctcacgtct
12	a montante	20546	105	*5' cttattctgagtcctctcc
12	a jusante	20002	106	5'tgtaaaacgacggccagtggttg ctcagaggctgc
12	a montante	19829	107	*5'gatgggtctgtacagattcccg
12	a jusante	19385	108	5'tgtaaaacgacggccag ittatt acagaataaaggaggtag
13	a montante	19300	109	5'tgtaaaacgacggccagtaacc cacaaaatttggtctaa
13	a jusante	19078	110	* 5 'tctccatttccaaaaccttg
14	a montante	19456	111	*5'tgtctctagttctgggtgc
14	a jusante	19472	112	5'tgtaaaacgacggccagttggtg tagtagctctgcttg
15	a montante	19697	113	*5' atttgtcccaactgggtgta
15	a jusante	19466	114	5'tgtaaaacgacggccagttcagt tgaaatgtcagaaagtg
16	a montante	19269	115	5'tgtaaaacgacggccagtg
16	a jusante	19047	116	*5'ccggctggaattttatttgag
17	a montante	19298	117	5'tgtaaaacgacggccagtaggc actggagaaatgggatttg
17	a jusante	19080	118	*5'tccagcacacatgcatgtaccg aaat
18	a montante	19436	119	* 5' gtagtctgtg atctccgttt:
18	a jusante	19471	120	5'tgtaaaacgacggccagttatga ggctctgtcctag
19	a montante	19447	121	*5'accagtgatgttgggatg
19	a jusante	19330	122	5'tgtaaaacgacggccagtgaaa gaagaacacatcccaca

No quadro 3 um asterisco (*) indica que o nucleótido 5' é biotinilado. Os exões 1-7, 10, 13 e 16-19 podem ser especificamente amplificados nas reacções de RCP contendo quer MgCl_2 1,5 mM ou 3 mM. Os exões 11 e 14 podem ser especificamente amplificados em reacções de RCP contendo MgCl_2 3 mM. No que respeita ao exão 12, os iniciadores de amplificação do segundo estágio foram preparados de modo a que o exão 12 fosse novamente amplificado em duas metades. O conjunto de iniciadores 20546 e 20002 amplifica a metade do terminal N. O conjunto de iniciadores 19829 e 19835 amplifica a metade do terminal C. Um iniciador alternativo de 18178 é o 19070.

A informação sobre a sequência de *hMLH1* fornecida pelos estudos dos requerentes e descrita neste pedido de patente de invenção e os pedidos de patentes de invenção relacionados que a precederam, pode ser utilizada para preparar um grande número de iniciadores diferentes de oligonucleótidos para serem utilizados na identificação das mutações de *hMLH1* que estão correlacionadas com a susceptibilidade para o cancro e/ou com o desenvolvimento de tumores em indivíduos, incluindo os iniciadores que irão amplificar mais do que um exão (e/ou flanquear as sequências de intrões) numa única banda de produto.

Um especialista na matéria será familiar com considerações importantes para preparar os iniciadores de RCP para serem utilizados para amplificar os fragmentos ou genes desejados ³⁷. Estas considerações podem ser similares, embora não necessariamente idênticas às envolvidas na preparação de iniciadores de sequenciação, tal como se discutiu antes. Geralmente, é importante que os iniciadores hibridem de uma forma relativamente específica (isto é, com uma T_m superior a 55 graus C e

preferencialmente próxima dos 60 graus C). Na maior parte dos casos, os iniciadores com um comprimento entre cerca de 17 e 25 funcionam bem. Iniciadores mais compridos podem ser úteis para a amplificação de fragmentos mais compridos. Em todos os casos é desejável evitar utilizar iniciadores que sejam complementares a mais do que uma sequência no genoma humano, de modo a que cada par dos iniciadores de RCP amplifique apenas um único fragmento correcto. Apesar disso, é apenas absolutamente necessário que a banda correcta se possa distinguir de outras bandas de produto na reacção de RCP.

As condições exactas da RCP (por exemplo, a concentração do sal, o número de ciclos, o tipo de polimerase de ADN, etc.) podem variar como é sabido na técnica para melhorar, por exemplo, o rendimento ou a especificidade da reacção. Em particular, os requerentes verificaram que era valioso utilizar iniciadores "aninhados" nas reacções de RCP de modo a reduzir a quantidade de substrato de ADN necessária e a melhorar a especificidade da amplificação.

Seguem-se dois exemplos. O primeiro exemplo ilustra a utilização de um par de iniciadores de um primeiro estágio (SEQ. ID NOS: 69 e 70) para amplificar o segmento de intrão/exão (SEQ. ID NO: 18). O segundo exemplo ilustra a utilização de iniciadores do segundo estágio para amplificar um segmento alvo de intrão/exão a partir do produto de uma primeira etapa de amplificação de RCP utilizando iniciadores do primeiro estágio.

EXEMPLO 1: Amplificação dos clones genómicos de *hMLH1* a partir de uma biblioteca de fagos P1

Utilizou-se 25 ng de ADN genómico (ou pode-se utilizar 1 ng de fago P1) nas reacções de RCP incluindo:

dNTPs 0,05mM

KCl 50 mM

Mg 3 mM

Tris-HCl 10 mM a pH 8,5

gelatina a 0,01%

iniciadores 5 µM

As reacções foram realizadas num equipamento de realização de ciclos térmicos Cetus do modelo 9600 da Perkin-Elmer. Incubaram-se as misturas reaccionais a 95 graus C. durante 5 minutos, seguido de 35 ciclos (30 ciclos de um fago P1) de:

94 GRAUS C. durante 30 segundos

55 GRAUS C. durante 30 segundos

72 GRAUS C. durante 1 minuto.

Realizou-se então uma extensão final da reacção durante 7 minutos a 72 graus C. Os clones de P1 desejáveis foram os que produziram uma banda de produtos de aproximadamente um pb.

EXEMPLO 2: Amplificação das sequências de *hMLH1* a partir de ADN genómico utilizando iniciadores de RCP encastrados

Os requerentes realizaram uma amplificação por RCP em duas etapas das sequências de *hMLH1* a partir de ADN genómico, como se segue. Normalmente, a primeira amplificação foi realizada numa reacção de 25 microlitros incluindo:

25 ng de AND cromossómico

Tampão II de RCP da Perkin-Elmer (pode-se utilizar qualquer tampão apropriado) MgCl_2 3mM

50 μM de cada dNTP

Taq ADN polimerase

Iniciadores de 5 μM (SEQ 10 NOS: 69, 70)

e incubou-se a 95 graus C durante 5 minutos, seguido se 20 ciclos de

94 graus C durante 30 segundos

55 graus C durante 30 segundos.

A banda do produto foi normalmente suficientemente pequena (inferior a aproximadamente 500 pb) de tal modo que não se realizaram etapas de extensão como parte de cada ciclo. Em vez disso, realizou-se apenas uma etapa de extensão, a 72 graus C, durante 7 minutos, depois de os 20 ciclos estarem completos. Armazenaram-se os produtos da reacção a 4 graus C.

A segunda reacção de amplificação, normalmente 25 a 50 microlitros em volume, incluiu:

1 ou 2 microlitros (consoante o volume da reacção) do primeiro produto da reacção de amplificação

Tampão II de RCP da Perkin-Elmer (pode-se utilizar qualquer tampão apropriado)

MgCl_2 3mM

50 μM de cada dNTP

Taq ADN polimerase

Iniciadores de 5 μM encastrados (SEQ 10 NOS: 109, 110),

e incubou-se a 95 graus C durante 5 minutos, seguido se 20-25 ciclos de

94 graus C durante 30 segundos

55 graus C durante 30 segundos.

realizou-se uma única etapa de extensão, a 72 graus C, durante 7 minutos, depois de os ciclos estarem completos. Armazenaram-se os produtos da reacção a 4 graus C.

Pode-se utilizar qualquer conjunto de iniciadores capazes de amplificar uma sequência alvo de *hMLH1* na primeira reacção de amplificação. Os requerentes utilizaram cada um dos conjuntos de iniciadores apresentados no quadro 2 para amplificar um exão individual de *hMLH1* na primeira reacção de amplificação. Os requerentes utilizaram também combinações desses conjuntos de iniciadores, amplificando assim múltiplos exões individuais de *hMLH1* na primeira reacção de amplificação.

Os iniciadores encastrados utilizados na primeira etapa de amplificação foram produzidos relativamente aos iniciadores utilizados na reacção de amplificação. Isto é, quando se utilizou um único conjunto de iniciadores na primeira reacção de amplificação, os iniciadores utilizados na segunda reacção de amplificação deviam ser idênticos aos iniciadores utilizados na primeira reacção, excepto no facto de os iniciadores utilizados na segunda reacção não deverem incluir os principais nucleótidos de 5' dos iniciadores da primeira reacção de amplificação e deveriam amplificar-se suficientemente mais na extremidade 3' e pelo facto de a T_m dos iniciadores da segunda amplificação ser aproximadamente a mesma que a T_m dos iniciadores da primeira reacção de amplificação. Os iniciadores da segunda

reacção de amplificação perderam os 3 nucleótidos principais de 5' dos iniciadores da primeira reacção de amplificação e estenderam-se aproximadamente 3-6 nucleótidos mais na extremidade 3'. As SEQ ID NOS: 109, 110 são exemplos de pares de iniciadores aninhados que podem ser utilizados numa segunda reacção de amplificação quando se utilizaram as SEQ ID NOS: 69 e 70 na primeira reacção de amplificação.

Os requerentes também verificaram que pode ser valioso incluir uma sequência padrão na extremidade 5' de um dos iniciadores da segunda reacção de amplificação para iniciar as reacções de sequenciação. Adicionalmente, os requerentes verificaram que é útil biotinilar esse último nucleótido de um ou de ambos os iniciadores da segunda reacção de amplificação de modo a que a banda do produto possa ser facilmente purificada utilizando pérolas⁴⁰ magnéticas as reacções de sequenciação podem ser realizadas directamente nos produtos associados às pérolas ⁴¹⁻⁴⁵.

Para uma discussão adicional sobre a amplificação de multiplexes e os processos de sequenciação, ver referências de Zu et al. e Espelund et al.^{46,47}.

***hMLH1* liga-se ao cancro**

Como uma primeira etapa para determinar se foi um candidato para o locus CCNPH *hMLH1* no cromossoma humano 3p21-23 ³, os requerentes mapearam *hMLH1* por hibridação *in situ* por fluorescência (HISF, FISH na terminologia inglesa) ^{20,21}. Os requerentes utilizaram dois fragmentos genómicos separados (dados não mostrados) do gene *hMLH1* em análise por HISF. O exame de várias metáfases de cromossomas dispersos localizou *hMLH1* com o cromossoma 3p21.3.23.

A coluna A da figura 7 mostra a hibridação das sondas de *hMLH1* numa metáfase dispersa. Hibridaram-se sondas genómicas de *hMLH1* biotiniladas com cromossomas humanos em bandas de metáfase, tal como foi previamente descrito ^{20,21}. Realizou-se a detecção com avidina e conjugado de isotiocianato de fluoresceína (CITF, FITC na terminologia inglesa) (sinal verde); os cromossomas, que se mostraram a azul, foram contra-corados com 4'6-diamino-2-fenitindole (DAFI, DAPI na terminologia inglesa). Obtiveram-se imagens com uma câmara CCD arrefecida, melhorada, pseudocolorida e combinada com os seguintes programas: CCD Image Capture; NIH Image 1.4; Adobe Photoshop e Genejoin Maxpix respectivamente. A coluna B da figura 7 mostra uma composição do cromossoma 3 de múltiplas metafases dispersas alinhadas com o ideograma do cromossoma 3 humano. A região de hibridação (porção distal de 3p21.3-23) está indicada no ideograma por uma barra vertical.

Como confirmação independente da localização de *hMLH1* no cromossoma 3, os requerentes utilizaram tanto a RCP com um par de oligonucleótidos específicos de *hMLH1*, como a análise de mancha de Southern com uma sonda específica de *hMLH1* para analisar o ADN do painel de células de roedores/humanos NIGMS2 (Coriell Inst. Para Med. Res., Camden, NJ, EUA). Os resultados de ambas as técnicas indicaram a ligação do cromossoma 3. Os requerentes também mapearam o gene *hMLH1* por HSF com a banda E do cromossoma 9. Esta é uma posição de sintenia com o cromossoma humano 3p ²². Por isso, o gene *hMLH1* localiza-se no 3p21.3-23, dentro da região genómica implicada nas famílias de CCNPH, ligado ao cromossoma 3 ³.

Em seguida, os requerentes analisaram amostras de sangue de indivíduos afectados e não afectados de duas

famílias ³ candidatas a mutações do cromossoma 3. Uma das famílias, a família 1, mostrou uma forte ligação (pontuação lod = 3,01 na fracção de recombinação 0) entre CCNPH e um marcador em 3p. Para a segunda família, a família 2, a pontuação lod referida (1,02) estava abaixo do nível de significância geralmente aceite e assim só sugeria ligação ao mesmo marcador em 3p. A análise subsequente de ligação da família 2 com o marcador de micro-satélite D351298 em 3p21.3 deu uma pontuação lod mais significativa de 1,88 numa fracção de recombinação de 0. Inicialmente, os requerentes pesquisaram mutações nos dois exões amplificados por RCP do gene *hMLH1* por sequenciação directa do ADN (figura 4). Os requerentes examinaram estes dois exões de três indivíduos afectados da família 1 e não detectaram diferenças da sequência esperada. Na família 2, os requerentes observaram que quatro indivíduos afectados com cancro do cólon são heterozigóticos para uma substituição C para T num exão que codifica os aminoácidos 41-9, o que corresponde a uma região altamente conservada da proteína (figura 9). Para um indivíduo afectado, os requerentes pesquisaram ADNc amplificado por RCP no que respeita a mais diferenças da sequência. A informação combinada das sequências obtida a partir dos exões e do ADNc deste indivíduo afectado representa 95 % (isto é, apenas os primeiros 116 pb) da secção de leitura aberta. Os requerentes não observaram alterações dos nucleótidos para além da substituição de C por T. Além disso, verificou-se que quatro indivíduos da família 2, que se previa que fossem portadores com base nos dados da ligação e ainda não afectados com cancro do cólon, eram heterozigóticos para a mesma substituição de C por T. Dois dos previstos portadores estavam abaixo e dois estavam acima da idade média de aparecimento (50 anos) nesta família em particular. Dois indivíduos examinados, não afectados, da

mesma família, ambos previsivelmente não portadores, pelos dados da ligação, exibiram a sequência normal expectável nesta posição. As análises da ligação que incluem a substituição de C por T na família 2, dão uma pontuação lod de 2,23 numa fracção de recombinação 0. Utilizando critérios de diagnóstico de cancro pouco severos, os requerentes calcularam uma pontuação lod de 2.53. Estes dados indicam que a substituição de C por T mostra uma ligação significativa ao CCNPH na família 2.

A figura 8 mostra os cromatogramas da sequência indicando uma mutação da transição de C para T que produz uma substituição de aminoácidos não conservadora na posição 44 da proteína *hMLH1*. Apresenta-se a análise da sequência de um indivíduos não afectado (painéis do topo, estruturas helicoidais mais e menos) e um indivíduo afectado (painéis inferiores, estruturas helicoidais mais e menos). A posição do nucleótido heterozigótico está indicada por uma seta. A análise dos cromatógrafos da sequência indica que há suficiente sinal de T no pico C e suficiente sinal de A no pico G para que os indivíduos afectados sejam heterozigóticos neste sítio.

Para determinar se esta substituição de C por T foi um polimorfismo, os requerentes sequenciaram este mesmo exão amplificado a partir do ADN genómico de 48 indivíduos não aparentados e observou-se apenas a sequência normal. Os requerentes examinaram mais 26 indivíduos não aparentados utilizando a análise de hibridação dos oligonucleótidos específicos dos alelos (OEA) ³³. As sequências de OEA (SEQ ID NOS: 141 e 142, respectivamente) utilizadas pelos requerentes foram:

5' -ACTTGTGGA TTTTGC-3' e

5' -ACTTGTGAA TTITGC-3'.

Com base na sequenciação directa do ADN e na análise dos OEA, nenhum destes 74 indivíduos não aparentados comportava a substituição de C por T. Por isso, não é provável que a substituição de C por T observada nos indivíduos da família 2 seja um polimorfismo. Tal como se mencionou antes, os requerentes não detectaram esta mesma substituição de C por T em indivíduos afectados de uma segunda família ligada ao cromossoma 3, a família 1³. Os requerentes continuam a estudar indivíduos da família 1 para pesquisar mutações na *hMLH1*.

O quadro 4 a seguir resume as análises experimentais dos requerentes de amostras de sangue de indivíduos afectados e não afectados da família 2 e indivíduos não afectados.

Quadro 4

F A M Í L I A 2	Status	Número de indivíduos com a mutação de C para T / Número de indivíduos analisados
	Afectados	4/4
	Portadores previstos	4/4
	Não portadores previstos	0/2
	Indivíduos não aparentados	0/74

Com base em vários critérios, os requerentes sugerem que a substituição de C por T na região de codificação de *hMLH1* representa a mutação que está na base de CCNPH na família 2³. Em primeiro lugar, a sequência de ADN e a análise por OEA não detectaram a substituição de C por T em 74 indivíduos não aparentados. Assim, a substituição de C

por T não é simplesmente um polimorfismo. Em segundo lugar, é expectável que a substituição de C por T produza uma alteração de serina para fenilalanina na posição 44 (ver figura 9). Esta substituição do aminoácido é uma alteração não conservadora numa região conservada da proteína (figuras 3 e 9). As previsões da estrutura secundária, utilizando os parâmetros de Chou-Fasman sugerem uma estrutura em folha de hélice enrolada em beta com a posição 44 localizada na volta. A substituição observada de Ser por Fen, na posição 44, diminui consideravelmente esta previsão para este anel, sugerindo que a substituição prevista dos aminoácidos altera a conformação da proteína *hMLH1*. A sugestão de que a substituição de Ser por Fen é uma mutação que confere susceptibilidade ao cancro é ainda suportada pelas experiências dos requerentes que mostram uma substituição análoga (alanina por fenilalanina) num gene de *hMLH1* de fungos resulta numa proteína de reparação do desemparelhamento, não funcional. Nas bactérias e nos fungos, uma mutação que afecte a reparação do desemparelhamento do ADN causa aumentos comparáveis na taxa de mutação espontânea incluindo as adições e eliminações dentro das repetições dos dinucleótidos ^{4, 5,11,13,14,15,16}. Nos seres humanos, a mutação de *hMSH2* é a base de CCNPH do cromossoma 2 ^{1, 2}, tumores que mostram uma instabilidade do micro-satélite e um defeito aparente na reparação do desemparelhamento ¹². O CCNPH associado ao cromossoma 3 está também associado com a instabilidade das repetições dos dinucleótidos ³. Combinado com estas observações, o elevado grau de conservação entre a proteína humana *MLH1* e a proteína de reparação do desemparelhamento do ADN de fungos sugere a probabilidade da *hMLH1* funcionar na reparação do desemparelhamento do ADN. Durante o isolamento do gene de *hMLH1*, os requerentes identificaram o gene de *hPMS1*. Esta observação sugere que a reparação do

desemparelhamento do ADN em mamíferos, assim como em fungos⁴, pode requerer pelo menos duas proteínas semelhantes a MutL.

Deve notar-se que parece que famílias de CCNPH diferentes mostram diferentes mutações no gene de *MLH1*. Tal como se explicou antes, os indivíduos da família 1 mostraram uma "ligação estreita" entre CCNPH e um locus na região de 3p21-23. Contudo, os indivíduos afectados na família 1 não têm a mutação de C por T encontrada na família 2. Parece que os indivíduos afectados na família 1 têm uma mutação diferente no seu gene de *MLH1*. Além disso, os requerentes utilizaram a informação sobre a estrutura e os processos descritos neste pedido de patente de invenção para encontrar e caracterizar a mutação de *hMLH1* que, aparentemente, confere susceptibilidade ao cancro em portadores heterozigóticos do gene mutante numa grande família inglesa com CCNPH. A mutação de *hMLH1* na família inglesa é uma deslocação seccional de + 1 T que é previsível que leve à síntese de uma proteína *hMLH1* truncada. Ao contrário, por exemplo, a anemia drepanocitária, em que praticamente todos os indivíduos afectados conhecidos têm a mesma mutação, descobriram-se mutações múltiplas de *hMLH1* e ligadas ao cancro. Por isso, o conhecimento de toda a sequência de ADNc para *hMLH1* (e provavelmente para *hPMS1*), assim como as sequências genómicas particularmente as que rodeiam os exões, serão úteis e importantes para a caracterização das mutações em famílias identificadas como exibindo uma elevada frequência de cancro.

Na sequência da verificação dos requerentes de que um cancro confere mutação na *hMLH1*, estudos de outros resultaram na caracterização de pelo menos mais 5 mutações

em *hMLH1*, cada uma das quais parecendo ter conferido susceptibilidade ao cancro em indivíduos de pelo menos uma família CCNPH. Por exemplo, Papadopoulos et al. identificaram essa mutação, caracterizada pela eliminação de uma secção de 165 pares de bases entre os codões 578 a 632. Numa outra família, Papadopoulos et al. Observaram uma mutação de *hMLH1* caracterizada por uma deslocação de uma secção e a substituição de novos aminoácidos, nomeadamente uma eliminação de 4 pares de bases entre os codões 727 e 728. Papadopoulos et al. também reportaram uma mutação de ligada a cancro, caracterizada por uma extensão das terminações de COOH, nomeadamente uma inserção de 4 pares de bases entre os codões 755 e 756 ³⁸.

Em resumo, os requerentes mostraram que *hMLH1* do gene de reparação do desemparelhamento do ADN que é provavelmente o gene hereditário do cancro do cólon não poliposo previamente localizado por análise de ligação ao cromossoma 3p21-23 ³. A disponibilidade da sequência do gene de *hMLH1* facilitará a avaliação das famílias com CCNPH no que respeita às mutações ligadas ao cancro. Além disso, embora a perda da heterozigosidade (PEH) dos marcadores ligados não seja uma característica quer das formas de 2p ou de 3p de CCNPH ^{3,6}. A PEH que envolve a região de 3p21.3-23 tem sido observada em vários cancros humanos ²⁴⁻²⁶. Isto sugere a possibilidade de que a mutação de *hMLH1* pode desempenhar algum papel nestes tumores.

***PMS1* humana**

Isolou-se *PMS1* humana utilizando os processos discutidos com referência à figura 1. A figura 10 mostra toda a sequência de nucleótidos do ADNc de *hPMS1*. A figura 11 mostra um alinhamento das previsíveis sequências de

proteínas *PMS1* humanas e de fungos. Os requerentes determinaram por análise por HSF que a *PMS1* humana está localizada no cromossoma 7. Na sequência da descoberta de *hPMS1* pelos requerentes, outros identificaram mutações no gene que parece conferir susceptibilidade ao CCNPH ³⁹.

***MLH1* de rato**

Utilizando o processo sublinhado antes no que se refere à figura 1, os requerentes determinaram uma sequência parcial de nucleótidos do ADNc de *MLH1* de rato; como se mostra na figura 12 (SEQ 10 NO: 135). A figura 13 mostra a correspondente sequência de aminoácidos previsível para a proteína *MLH1* (SEQ 10 NO: 136) em comparação com a sequência previsível de *hMLH1* (SEQ 10 NO: 5). A comparação das proteínas *MLH1* humanas e de rato assim como a comparação de *hMLH1* com as proteínas *MLH1* de fungos, tal como se mostra na figura 9, indicam um elevado grau de conservação.

***PMS1* de rato**

Utilizando os processos discutidos antes no que se refere à figura 1, os requerentes isolaram e sequenciaram o gene de *PMS1* de rato, como se mostra na figura 14 (SEQ 10 NO: 137). Esta sequência de ADNc codifica uma proteína previsível de 864 aminoácidos (SEQ 10 NO: 138), tal como se mostra na figura 15, onde se compara a sequência de aminoácidos para *hPMS1* (SEQ 10 NO: 133). O grau de identidade entre as proteínas *PMS1* de rato e humanas é elevado, como seria expectável entre dois mamíferos. Do mesmo modo, tal como se fez notar antes, há uma forte semelhança entre a proteína *PMS1* humana e a proteína de reparação do desemparelhamento do ADN de fungos *PMS1*, tal

como se mostra na figura 11. O facto de a PMS1 e a MLH1 funcionarem nos fungos para reparar o desemparelhamento do ADN, sugere fortemente que a PMS1 e a MLH1 humanas e de ratos são também proteínas de reparação do desemparelhamento.

Utilizações das MLH1 e PMS1 de rato

Os requerentes crêem que o isolamento e a caracterização por eles feita dos genes de a *rPMS1* e *rMLH1* terão muitas aplicações na investigação. Por exemplo, tal como já se discutiu antes, os requerentes utilizaram os seus conhecimentos sobre o gene de *rPMS1* para produzir anticorpos que reagem especificamente com hPMS1. Os requerentes já explicaram que os anticorpos dirigidos às proteínas humanas, MLH1 ou PMS1 podem ser utilizados ambos para fins de investigação assim como para fins de diagnóstico.

Os requerentes também acreditam que o seu conhecimento de *rPMS1* e *rMLH1* será útil para a construção de modelos de ratos de modo a estudar as consequências dos defeitos de reparação do desemparelhamento do ADN. Os requerentes esperam que os ratos com defeitos de *rPMS1* e *rMLH1* terão uma elevada propensão para o cancro porque o CCNPH associado aos cromossomas 2p e 3p são ambos devidos a um defeito num gene de reparação do desemparelhamento ^{1,2}. Como se fez notar antes, os requerentes também produziram ratos quiméricos que comportam um gene defeituoso de *rPMS1*. Os requerentes estão neste momento a produzir ratos heterozigóticos quanto à mutação de *rPMS1* ou *rMLH1*. Estes ratos heterozigóticos devem providenciar modelos de animais úteis para estudar o cancro humano, em particular CCNPH. Os ratos serão úteis para a análise tanto dos

factores intrínsecos como dos extrínsecos que determina o risco e a progressão do cancro. Também os cancros associados com a deficiência da reparação do desemparelhamento podem responder de forma diferente à terapia convencional em comparação com outros cancros. Esses modelos de animais serão úteis para a determinação se existem diferenças e vão permitir o desenvolvimento de regimes para o tratamento efectivo deste tipo de tumores. Esses modelos de animais podem também ser utilizados para estudar a relação entre factores hereditários versus factores dietéticos na carcinogénese.

Distinção entre mutações e polimorfismos

Para os estudos sobre a susceptibilidade para o cancro e para a identificação e caracterização de tumores, é importante distinguir "mutações" de "polimorfismos". Uma "mutação" produz um "alelo de tipo não selvagem" de um gene. Um alelo de tipo não selvagem de um gene produz um transcrito e/ou um produto de proteína que não funciona normalmente dentro de uma célula. As "mutações" podem ser qualquer alteração na sequência de nucleótidos incluindo inserções, eliminações, substituições e re-arranjos.

"Polimorfismos", por outro lado, são diferenças da sequência que se encontram dentro da população de genes que funcionam normalmente (isto é, "de tipo selvagem"). Alguns polimorfismos resultam da degenerescência do código do ácido nucleico. Isto é, dado que a maior parte dos aminoácidos estão codificados por mais do que um codão de triplete, muitas sequências de nucleótidos diferentes podem codificar o mesmo polipéptido. Outros polimorfismos são simplesmente diferenças de sequências que não têm um efeito significativo na função do gene ou polipéptido codificado.

Por exemplo, os polipéptidos podem muitas vezes tolerar pequenas inserções ou eliminações ou substituições "conservadoras" na sua sequência de aminoácidos sem alterar significativamente a função do polipéptido.

As substituições "conservadoras" são aquelas em que um aminoácido particular é substituído por outro aminoácido de características químicas semelhantes. Por exemplo, os aminoácidos são muitas vezes caracterizados como "não polares (hidrofóbicos)" incluindo alanina, leucina, isoleucina, valina, prolina, fenilalanina, triptofano e metionina; "polar neutro", incluindo glicina, serina, treonina, cisteína, tirosina, asparagina e glutamina; "carregados positivamente (básicos)", incluindo arginina, lisina e histidina; e "carregados negativamente (ácidos)", incluindo ácido aspártico e ácido glutâmico. Uma substituição de um dos aminoácidos por outro no mesmo grupo é geralmente considerada como sendo "conservadora", particularmente se os grupos laterais dos dois aminoácidos relevantes forem de uma dimensão semelhante.

A primeira etapa na identificação de uma mutação ou polimorfismo numa sequência de genes de reparação de desemparelhamento envolve a identificação, utilizando técnicas disponíveis incluindo as aqui descritas, de um gene de reparação de desemparelhamento (ou fragmento de gene), sequência que difere de uma sequência normal (por exemplo de tipo selvagem) conhecida do mesmo gene de reparação de desemparelhamento (ou fragmento de gene). Por exemplo, podia-se identificar uma sequência de gene de *hMLH1* (ou fragmento de gene) que difere em pelo menos uma posição de nucleótido de uma normal sequência de *hMLH1*, conhecida (por exemplo de tipo selvagem), tal como uma qualquer das SEQ 10 NOS: 6-24.

As mutações podem distinguir-se dos polimorfismos utilizando qualquer um de uma variedade de processos, talvez o mais directo dos quais seja a recolha e correlação de dados com o desenvolvimento do tumor. Isto é, por exemplo, pode-se identificar um indivíduo cuja sequência de gene de *hMLH1* difere de uma sequência reportada nas SEQ 10 NOS: 6-24, mas que não tem cancro e não tem história de cancro na família. Particularmente se outros elementos da família dos indivíduos, em especial pessoas mais velhas, tiverem sequências de genes de *hMLH1* que diferem das SEQ 10 NOS: 6-24, da mesma forma, é provável que as sequências de genes de *hMLH1* desses indivíduos possam ser categorizadas como um "polimorfismo". Se outros indivíduos não aparentados forem identificados com a mesma sequência de gene de *hMLH1* e não tiverem uma história familiar de cancro, a categorização pode ser confirmada.

As mutações que são responsáveis por conferirem susceptibilidade genética ao cancro podem ser identificadas porque, entre outras coisas, essas mutações é provável que estejam presentes em todos os tecidos de um indivíduo não afectado e na linha de germes de pelo menos um desses parentes do indivíduo e não é provável que se encontre em famílias não aparentadas sem história de cancro.

Quando se quer distinguir as mutações dos polimorfismos, pode ser válido, algumas vezes, avaliar uma diferença particular da sequência na presença de pelo menos uma mutação conhecida do gene de reparação do desemparelhamento. Nalguns casos, uma alteração particular da sequência não terá efeito detectável (isto é, parecerá ser um polimorfismo) quando ensaiada isoladamente, mas fará, por exemplo, aumentar a pertinência de uma mutação conhecida, de tal modo que os indivíduos que são portadores

tanto da diferença do polimorfismo aparente e uma mutação conhecida têm uma probabilidade mais elevada de desenvolver cancro do que os indivíduos que são apenas portadores da mutação. As diferenças das sequências que têm esse efeito são consideradas apropriadamente como sendo mutações, embora sejam fracas.

Tal como se discutiu antes e previamente (pedidos de patentes de invenção norte-americanas U.S. n.ºs 08/168,877 e 08/209,521), as mutações nos genes de reparação de desemparelhamento ou produtos de genes produziram versões de tipo não selvagem desses genes ou produtos de genes. Algumas mutações podem por isso distinguir-se de polimorfismos pelas suas características funcionais em ensaio de reparação de desemparelhamentos *in vivo* ou *in vitro*. Qualquer ensaio de reparação de desemparelhamentos pode ser utilizado para analisar estas características ⁴⁹⁻
⁶³. É geralmente desejável utilizar mais do que um ensaio de reparação de desemparelhamentos antes de classificar uma alteração da sequência como um polimorfismo, dado que algumas mutações terão efeitos que não serão observados em todos os ensaios.

Por exemplo, não é expectável que um gene de reparação de desemparelhamento contendo uma mutação seja capaz de substituir uma cópia endógena do mesmo gene numa célula hospedeira sem afectar, de forma detectável, a reparação de desemparelhamento nessa célula; em que seria expectável que um gene de reparação de desemparelhamentos fosse capaz de substituir uma cópia endógena do mesmo gene numa célula hospedeira sem afectar, de forma detectável, a reparação de desemparelhamento nessa célula. O requerentes notaram que para esses estudos de "substituição", é geralmente desejável introduzir o gene a ser ensaiado numa célula

hospedeira da mesma espécie (ou pelo menos fortemente relacionada) que a célula de que derivou o gene de ensaio, para evitar complicações devidas, por exemplo, à inabilidade de um produto de gene de uma espécie para interagir com outros produtos de genes de reparação de desemparelhamentos de outras espécies. Do mesmo modo, não é expectável que uma proteína mutante de reparação de desemparelhamentos funcione normalmente num sistema de reparação de desemparelhamentos funcione normalmente num sistema de reparação de desemparelhamento *in vitro* (preferencialmente a partir de um organismo relacionado); em que seria expectável que uma proteína polimórfica de reparação de desemparelhamento funcionasse normalmente.

Os processos aqui descritos e os descritos previamente permitem a identificação de diferentes tipos de mutações de genes de reparação de desemparelhamentos. Os exemplos que se seguem ilustram protocolos para distinguir mutações de polimorfismos nos genes de reparação de desemparelhamentos do ADN.

EXEMPLO 3: Os requerentes desenvolveram um sistema para ensaiar no fungo *S. cerevisiae* o significado funcional de mutações encontradas quer nos genes de *hMLH1*, quer nos de *hPMS1*. O sistema está descrito no presente pedido de patente de invenção utilizando como exemplo a mutação causada pela substituição de serina (SER) por fenilalanina (FEN) em *hMLH1*, que se verificou numa família com CCNPH, tal como descrito antes. Os requerentes derivaram uma estirpe de fungo que foi praticamente eliminada do seu gene de *MLH1* e e por isso é um forte agente de mutação (isto é, 1000 vezes acima da taxa normal num ensaio simples de marcador genético que envolve a reversão da dependência do crescimento de um dado aminoácido para a independência

(reversão do alelo hom3-10, Prolla, Christie e Liskay, Mol Cell. Biol, 14: 407-415, 1994). Quando os requerentes colocaram o *MLH1* de um fungo normal, o gene (completo com todas as regiões de controlo conhecidas) num plasma de fungo que se manteve de forma estável como uma cópia única na estirpe eliminada de *MLH1*, o fenótipo do agente de mutação é completamente corrigido utilizando a reversão para o ensaio de independência do aminoácido. Contudo, se os requerentes introduziam uma cópia eliminada da *MLH1* de fungo, não há correcção. Os requerentes em seguida ensaiaram a mutação que, na família com CCNPH causou uma alteração de SER para FEN. Os requerentes verificaram que a proteína mutante de fungo resultante não pode corrigir o fenotipo do agente de mutação, sugerindo fortemente que a alteração da sequência do gene de tipo selvagem provavelmente confere susceptibilidade ao cancro e é por isso classificada como uma mutação, não um polimorfismo. Os requerentes em seguida ensaiaram as proteínas tratadas por engenharia genética para conter outros aminoácidos na posição "serena" e verificaram que a maior parte das alterações resultam num mutante completo ou pelo menos um fenotipo parcialmente mutante.

Como se verificam outras mutações "pontuais", nos genes de *MLH1* e de *PMS1*, em famílias de cancro, elas podem ser tratadas por engenharia genética no gene homólogo de fungo apropriado e estudou-se a sua consequência na função da proteína. Além disso, os requerentes identificaram um certo número de aminoácidos altamente conservados tanto nos genes de *MLH1* como *PMS1*. Os requerentes também puseram em evidência que *hMLH1* interage com *hPMS1* de fungos. Esta verificação levanta a possibilidade de que as mutações observadas no gene de *hMLH1* podem ser mais directamente ensaiadas no sistema de fungos. Os requerentes planeiam

fazer sistematicamente mutações que iriam alterar o aminoácido nestas posições conservadas e determinar quais as substituições de aminoácidos que são toleradas e quais as que não são. Recolhendo informação sobre as mutações relacionadas com *hMLH1* e *hPMS1*, tanto por determinação como por documentação actual, encontraram mutações em famílias com CCNPH e por meio de síntese artificial de mutantes para ensaiar em sistemas experimentais, pode eventualmente ser possível praticar um protocolo de ensaio da susceptibilidade ao cancro que, uma vez determinada a estrutura de *hMLH1* e *hPMS1* de indivíduos, requer apenas a comparação dessa estrutura com mutações conhecidas versus os dados do polimorfismo.

EXEMPLO 4: Um outro processo que os requerentes utilizaram para estudar as interacções físicas entre *hMLH1* e *hPMS1*, pode também ser utilizado para estudar se uma alteração particular num produto de gene resulta numa alteração do grau de interacção proteína-proteína. A informação respeitante às alterações na interacção de proteína-proteína pode demonstrar ou confirmar se uma variação genómica particular é uma mutação ou um polimorfismo. No seguimento das verificações laboratoriais dos requerentes sobre a interacção entre as proteínas *MLH1* e *PMS1* de fungos *in vitro* e *in vivo*, (pedido de patente de invenção norte-americana U.S. com o n° de série 08/168.877), ensaiou-se a interacção entre as contra-partes humanas destas duas proteínas de reparação do desemparelhamento do ADN. Ensaaiaram-se as proteínas humanas *MLH1* e *PMS1* quanto à interacção *in vitro* utilizando a cromatografia de afinidade da proteína de ligação da maltose (PLM). Preparou-se a proteína *hMLH1* como uma proteína de fusão de PLM, imobilizada ou numa coluna de resina de amilose por via de PLM ensaiou-se quanto à

ligação a *hPMS1*, sintetizada *in vitro*. A proteína *hPMS1* ligada à matriz de PLM - *hMLH1*, enquanto proteína de controlo não mostrou afinidade para a matriz. Quando se fez passar a proteína *hMLH1*, traduzida *in vitro*, sobre uma matriz de proteína de fusão de PLM - *hPMS1*, a proteína *hMLH1* ligou-se à matriz de PLM - *hPMS1*, enquanto as proteínas de controlo não se ligaram.

As interacções potenciais *in vivo* entre foram *MLH1* e *PMS1* ensaiadas utilizando o sistema de "dois híbridos" de fungo ²⁸. Os resultados iniciais dos requerentes indicam que *hMLH1* e *hPMS1* interagem *in vivo* em fungos. O mesmo sistema pode também ser utilizado para detectar alterações na interacção proteína-proteína que resulta das alterações na estrutura do gene ou do produto do gene e que ainda tem de ser classificada que como polimorfismo ou como mutação que confere susceptibilidade ao cancro.

Detecção de famílias com CCNPH e as suas mutações

Tem sido estimado que aproximadamente 1.000.000 de indivíduos nos Estados Unidos são portadores (SÃO heterozigóticos para) de um gene ²⁹ mutante de CCNPH. Além disso, as estimativas sugerem que 50-60 % das famílias segregam mutações no gene de *MSH2* que reside no cromossoma 2p.^{1, 2} Uma outra fracção significativa parece estar associada com o gene de CCNPH que mapeia o cromossoma 3p21-22, presumivelmente devido a mutações no gene de *hMLH1* tal como a transição de C para T discutida antes. A identificação de famílias que segregam alelos mutantes quer do gene de *hMSH2* ou *hMLH1* e a determinação dos indivíduos que nestas famílias actualmente têm a mutação serão de grande utilidade na intervenção precoce nestas doenças. Essa intervenção precoce irá provavelmente incluir a

detecção precoce através da avaliação e o tratamento de seguimento agressivo dos indivíduos afectados. Além disso, a determinação da base genética tanto para tumores familiares como esporádicos pode dirigir o processo de terapia no tumor primário ou nas recorrências.

Inicialmente, as famílias candidatas a CCNPH serão diagnosticadas parcialmente através do estudo das histórias da família, muito provavelmente ao nível local, por exemplo, por oncologistas hospitalares. Um critério para CCNPH é a observação de instabilidade de micro-satélites em tumores de indivíduos ^{3,6}. O paciente será avaliado quanto a mutações em *hMSH2*, *hMLH1*, *hPMS1* e outros genes envolvidos na reparação do desemparelhamento do ADN, tal como são identificados. Isto é feito mais facilmente por amostragem de sangue do indivíduo. Também altamente útil será o tecido de tumor congelado em fresco. É importante notar para o processo de avaliação, que os indivíduos afectados são heterozigóticos para a mutação ofensiva nos seus tecidos normais.

Os tecidos disponíveis, por exemplo, sangue e tumor, são trabalhados para a análise da mutação à base de RCP, utilizando um ou os dois processos que se seguem:

- 1) Análise de ligação com um marcador de micro-satélite fortemente ligado ao gene de *hMLH1*.

Uma abordagem para identificar famílias propensas ao cancro com uma mutação de *hMLH1* consiste em realizar análises de ligação com um marcador altamente polimórfico localizado dentro ou fortemente ligado ao *hMLH1*. Os micro-satélites são altamente polimórficos e por isso são muito úteis como marcadores na análise de ligação. Por os

requerentes disporem do gene de *hMLH1* num único fragmento genómico grande num clone de fago P1 (~100 kpb), é muito provável que um ou mais micro-sAtélites, por exemplo, traços de repetições de dinucleótidos existam dentro ou muito próximo do gene de *hMLH1*. Pelo menos um desses micro-satélites tem sido referenciado ³⁸. Uma vez identificados esses marcadores, preparam-se os amplificadores de RCP para amplificar as extensões de ADN contendo os micro-satélites. Vai-se avaliar o ADN de indivíduos afectados e não afectados de uma família com uma alta frequência de cancro para determinar a segregação dos marcadores de *MLH1* e a presença do cancro. Os dados resultantes podem ser utilizados para calcular uma pontuação lod e assim determinar a probabilidade de ligação entre *hMLH1* e a ocorrência de cancro. Uma vez estabelecida a ligação numa dada família, pode-se utilizar o mesmo marcador polimórfico para ensaiar outros membros aparentados quanto à probabilidade de serem portadores da mutação de *hMLH1*.

2) Sequenciação de ADNc de transcrição reversa

a) Faz-se a transcrição reversa (TR) do ARN de indivíduos afectados, indivíduos não afectados e não aparentados, seguida de RCP para amplificar o ADNc em 4-5 porções sobrepostas ^{34, 37}. Deve notar-se que para os fins de RCP, podem utilizar-se, potencialmente, muitas sequências de pares de iniciadores de oligonucleótidos diferentes, para amplificar porções relevantes de um gene de *hMLH1* ou *hPMS1* para fins de avaliação genética. Com o conhecimento das estruturas de ADNc para os genes, é um exercício linear construir o pares de iniciadores que provavelmente serão efectivos para amplificar especificamente porções seleccionadas do gene. Embora as sequências de iniciadores tenham normalmente entre 20 e 30 bases de comprimento pode

ser possível utilizar iniciadores mais curtos, potencialmente tão pequenos quanto aproximadamente 13 bases, para amplificar especificamente segmentos de genes seleccionados. A principal limitação quanto ao tamanho reduzido que uma sequência de iniciador pode ter é que deve ser suficientemente longa para hibridar especificamente com o segmento de gene atingido. A especificidade de RCP é normalmente melhorada aumentando o comprimento dos iniciadores e/ou utilizando pares de iniciadores incluídos.

Os produtos de RCP, que no total representam todo o ADNc são então sequenciados e comparados com sequências de tipo selvagem conhecidas. Na maior parte dos casos será observada uma mutação no indivíduo afectado. Idealmente, a natureza da mutação indicará que é provável que inactive o produto do gene. Se assim não for, deve-se determinar a possibilidade de que a alteração não seja simplesmente um polimorfismo.

b) Algumas mutações, por exemplo, as que afectam o seccionamento ou que resultam da translação de codões de paragem, podem desestabilizar o ARN mensageiro produzido a partir do gene mutante e por isso compreendem o processo de detecção da mutação com base na TR. Uma técnica recentemente referenciada pode circunscrever este problema, ensaiando se o ADNc mutante pode dirigir a síntese de proteínas de comprimento normal num sistema de transcrição/tradução acoplado *in vitro* ³².

3) Sequenciação directa do ADN genómico

Uma segunda via para detectar mutações baseia-se no exame de exões e das fronteiras de intrão/exão por meio de um ciclo de RCP para sequenciar directamente uma matriz de

ADN ^{1,2}. Este processo requer a utilização de pares de oligonucleótidos, tal como os descritos nos quadros 2 e 3 anteriores, o que amplifica os exões individuais para a sequenciação directa do ciclo de RCP. O processo depende da informação sobre a sequência do ADN genómico em cada fronteira de intrão/exão (50 pb ou superior para cada fronteira). A vantagem da técnica é de duas vezes. Primeiro, porque o ADN é mais estável do que o ARN, sendo que a condição do material utilizado para RCP não é tão importante como o é para os protocolos à base de ARN. Em segundo lugar, a maioria das mutações dentro da actual região do gene, incluindo as de um intrão que afecta o seccionamento, serão detectáveis.

Para cada gene candidato, a detecção da mutação pode exigir conhecimento tanto de toda a estrutura do ADNc como de todas as fronteiras de intrão/exão da estrutura genómica. Com essa informação, pode-se determinar o tipo de mutação causal numa família particular. Por sua vez, um esquema de detecção mutações mais específico e eficiente pode ser adaptado para a família em particular. A avaliação da doença (CCNPH) é complexa porque tem uma base geneticamente heterogénea no sentido de que mais do que um gene está envolvido e, para cada gene, estão envolvidos múltiplos tipos de mutações ². Qualquer família é muito provável que segregue uma mutação particular. Contudo, como se determina a natureza da mutação em múltiplas famílias, o espectro das mutações mais prevalentes na população será determinado. Em geral, a determinação das mutações mais frequentes será dirigida para a detecção da mutação.

Dado que o CCNPH é tão prevalente na população humana, a detecção de um portador à nascença poderia fazer parte dos ensaios padrão dos recém-nascidos. As famílias em risco

podem ser identificadas e todos os elementos não ensaiados previamente podem ser ensaiados. Eventualmente, pode determinar-se todos os indivíduos afectados.

Modo de rastreio e de ensaio da mutação

Ensaio à base de ADN

O ensaio inicial, incluindo a identificação de prováveis famílias de CCNPH por diagnóstico padrão e estudo da história da família, será provavelmente feito em laboratórios de diagnóstico de ADN mais pequenos e locais. Contudo, o ensaio em larga escala de múltiplos elementos da família e certamente um ensaio alargado da população, exigirão, em última análise, grandes faculdades comerciais centralizadas e eficientes.

Os ensaios serão desenvolvidos com base na determinação das mutações mais comuns para os genes principais subjacentes a CCNPH, incluindo pelo menos o gene de *hPMS2* no cromossoma 2p e o gene de *MLH1* no cromossoma 3p. Provavelmente vai desenvolver-se uma variedade de ensaios. Por exemplo, uma possibilidade é um conjunto de testes utilizando hibridações de oligonucleótidos que distinguem os alelos normais vs os mutantes ³³. Como já se fez notar, o conhecimento dos requerentes sobre as estruturas de nucleótidos para os genes de *hMLH1*, *hPMS1* and *hMSH2*, torna possível a preparação de numerosos pares de iniciadores de oligonucleótidos que podem ser utilizados para amplificar porções específicas de um gene de reparação de desemparelhamentos em indivíduos para a avaliação genética e análise do risco de cancro. O conhecimento dos requerentes sobre as estruturas dos genes também torna possível a preparação de sondas marcadas que podem ser

utilizadas rapidamente para determinar a presença ou a ausência de toda ou de uma porção dos genes de reparação do desemparelhamento do ADN. Por exemplo, as sondas de oligómeros específicas de alelos (OEA) podem ser preparadas para fazer a distinção entre alelos. Os OEAs são segmentos curtos de ADN que são idênticas na sequência excepto no facto de uma diferença numa única base que reflecte a diferença entre alelos normais e mutantes. Sob condições de hibridação apropriadas estas sondas podem reconhecer uma única diferença de bases entre duas sequências de alguma forma idênticas. As sondas podem ser marcadas radioactivamente ou com uma variedade de moléculas repórteres não radioactivas, por exemplo, partes fluorescente ou quimioluminescentes. As sondas marcadas são então utilizadas para analisar a amostra de RCP quanto à presença de alelo causador da doença. A presença ou a ausência de vários genes diferentes que causam a doença pode ser facilmente determinada numa única amostra. O comprimento da sonda deve ser suficientemente longo para evitar a ligação não específica às sequências de nucleótidos diferentes da sequência alvo. Todos os testes dependerão, em última análise, de informação precisa e completa sobre a estrutura relacionada com os genes de *hMLH1*, *hMSH2*, *hPMS1* e outros genes de reparação do desemparelhamento do ADN implicados no CCNPH.

Avaliação baseada na detecção de proteínas

Podem utilizar-se ensaios à base da funcionalidade do produto de proteína, *per se*. Os ensaios de exame da proteína utilizarão, mais provavelmente, reagentes de anticorpos específicos para cada uma das proteínas *hMLH1*, *hPMS1* e *hMSH2* ou outros produtos de gene relacionados com o "cancro", conforme são identificados.

Por exemplo, um espécimen de tumor congelado pode ser seccionado transversalmente e preparado para a coloração do anticorpo utilizando técnicas indirectas de fluorescência. É expectável que algumas mutações de genes alterem ou desestabilizem suficientemente a estrutura da proteína de tal modo que dêem um sinal alterado ou reduzido depois da coloração do anticorpo. É provável que esses ensaios sejam realizados em casos em que o envolvimento dos genes num cancro de uma família tenha ainda de ser estabelecida. Os requerentes estão em processo de desenvolvimento de anticorpos monoclonais de diagnóstico contra as proteínas MLH1 e PMS1 humanas. Os requerentes estão a sobre-expressar as proteínas humanas MLH1 e PMS1 em bactérias. Os requerentes purificaram as proteínas, injectaram-nas em ratos e derivaram os anticorpos monoclonais específicos da proteína que podem ser utilizados para fins de diagnóstico e de investigação.

Identificação e caracterização de tumores de reparação do desemparelhamento do ADN

Para além da sua utilidade no diagnóstico da susceptibilidade ao cancro num indivíduo, as sequências de nucleótidos que são homólogas de um gene de reparação de desemparelhamento podem ser valiosas para, entre outras coisas, serem utilizadas na identificação e na caracterização de tumores com defeitos da reparação de desemparelhamento. Essa identificação e caracterização é valiosa porque os tumores com defeitos da reparação de desemparelhamento podem responder melhor a regimes de terapia particulares. Por exemplo, os tumores com defeitos da reparação de desemparelhamento podem ser sensíveis aos agentes de danificação do ADN, especialmente quando administrados em combinação com outros agentes

terapêuticos.

Os defeitos nos genes de reparação de desemparelhamentos não precisam de estar presentes por todos os tecidos do indivíduo para contribuírem para a formação do tumor nesse indivíduo. A mutação espontânea de um gene de reparação de desemparelhamento numa célula ou num tecido particular pode contribuir para a formação do tumor nesse tecido. De facto, pelo menos nalguns casos, uma única mutação no gene de reparação de desemparelhamento não é suficiente para o desenvolvimento do tumor. Nesses casos, um indivíduo com uma única mutação num gene de reparação de desemparelhamento é susceptível ao cancro, mas não desenvolverá um tumor até ocorrer uma mutação secundária. Adicionalmente, nalguns casos, a mutação do gene de reparação de desemparelhamento que está estritamente associada a um tumor num indivíduo, será responsável por conferir susceptibilidade ao cancro numa família com uma predisposição hereditária para o desenvolvimento de cancro.

A informação sobre a sequência providenciada pelos requerentes pode ser utilizada com processos conhecidos na técnica para analisar tumores (ou linhas de células de tumores) e para identificar mutações associadas a tumores nos genes de reparação de desemparelhamentos. Preferencialmente, é possível demonstrar que estas mutações associadas a tumores não estão presentes nos tecidos que não são de tumor do mesmo indivíduo. A informação descrita neste pedido de patente de invenção é particularmente útil para a identificação de mutações do gene de reparação de desemparelhamento dentro de tumores (ou de linhas de células de tumor) que exibem instabilidade genómica dos elementos curtos e repetidos do ADN.

A informação sobre a sequência e os protocolos de ensaio podem também ser utilizados para determinar se dois tumores estão relacionados, isto é, se um segundo tumor é o resultado de metástases de um primeiro tumor encontrado mais cedo que exibe uma mutação particular do gene de reparação de desemparelhamento do ADN.

Isolamento de genes adicionais com uma função relacionada

As proteínas que interagem fisicamente quer com *hMLH1* e/ou *hPMS1*, estão provavelmente envolvidas na reparação do desemparelhamento do ADN. Por analogia com *hMLH1* e *hMSH2*, as mutações nos genes que codificam para essas proteínas serão fortes candidatos para uma ligação potencial ao cancro. Uma abordagem poderosa de genética molecular utilizando fungos, referida como "sistema de dois híbridos", permite uma detecção e um isolamento relativamente rápidos dos genes que codificam as proteínas que interagem com o produto do gene de interesse, por exemplo, *hMLH* ²⁸.

O sistema de dois híbridos envolve dois vectores de plasmido, cada um deles destinado a codificar uma proteína de fusão. Cada um dos dois vectores contém uma porção ou um domínio de um activador de transcrição. A célula de fungo utilizada no esquema de detecção contém um gene "repórter". O activador não pode activar sozinho a transcrição. Contudo, se os dois domínios forem colocados em grande proximidade, então a transcrição pode ocorrer. Insere-se o ADNc para a proteína de interesse, por exemplo *HMLH1*, dentro de uma secção de leitura em um dos vectores. Isto é designado por "isco" ("bait" na terminologia inglesa). Introduz-se uma biblioteca de ADNcs humanos, inserida num segundo vector de plasmido de modo a fazer fusões com o

outro domínio do activador de transcrição, nas células do fungo que exibem o vector "isco". Se uma célula particular de fungo recebe um elemento da biblioteca que contém um ADNc humano que codifica uma proteína que interage com a proteína *HMLH1*, esta interacção vai levar os dois domínios do activador de transcrição a uma grande proximidade, activa a transcrição do gene repórter e a célula do fungo vai ficar azul. Em seguida, a inserção é sequenciada para determinar se está relacionada com qualquer sequência na base de dados. Pode-se utilizar o mesmo processo para identificar proteínas de fungos em processos de reparação do desemparelhamento do ADN ou em processos relacionados. Realizar a "caça" em seres humanos e em fungos em paralelo tem certas vantagens. A função de novos homólogos de fungos pode ser rapidamente determinada em fungos por meio do rompimento do gene e o subsequente exame das consequências genéticas de ser defeituoso no novo gene encontrado. Estes estudos com fungos ajudarão a guiar a análise de novas proteínas humanas que "interagem com *hMLH1* ou *hPMS1* " mais ou menos da mesma maneira que os estudos sobre *PMS1* e *MLH1* influenciaram os estudos dos requerentes sobre os genes humanos de *PMS1* e *MLH1*.

Produção de anticorpos

Utilizando o conhecimento dos requerentes sobre as sequências de ADN para *hMLH1* e *hPMS1*, eles puderam sintetizar todas ou porções das estruturas de proteínas previstas para o fim da produção de anticorpos. Uma utilização importante para os anticorpos dirigida às proteínas *hMLH1* e *hPMS1* será a captura de outras proteínas que podem estar envolvidas na reparação do desemparelhamento do ADN. Por exemplo, utilizando técnicas de co-imuno-precipitação, podem precipitar-se os anticorpos

dirigidos quer a *hMLH1* ou *hPMS1*, em conjunto com outras proteínas associadas que estão funcionalmente e/ou fisicamente relacionadas. Outra utilização importante para os anticorpos terá por fim o isolamento das proteínas *hMLH1* e *hPMS1* de tecidos de tumores. As proteínas *hMLH1* e *hPMS1* de tecidos de tumores podem então ser caracterizadas com o fim de se determinar as estratégias de tratamento apropriadas.

Os requerentes estão em processo de desenvolvimento de anticorpos monoclonais dirigidos às proteínas *hMLH1* e *hPMS1*.

EXEMPLO 5: Os requerentes utilizaram também o processo que se segue para produzir anticorpos policlonais dirigidos para as formas humanas e de rato da proteína *PMS1*.

Os requerentes inseriram um fragmento de 3' do ADNc de *PMS1* de rato no vector do plasmido de expressão bacteriano, pET (Novagen, Madison, WI). A esperada porção expressa da proteína *PMS1* de rato corresponde a uma região de aproximadamente 200 aminoácidos no fim da proteína *PMS1*. Esta porção da *rPMS1* está conservada com *PMS1* de fungo mas não está conservada com nenhuma das proteínas *MLH1* humanas ou de rato. Uma das razões pela qual os requerentes seleccionaram esta porção de proteína *PMS1* para a produção de anticorpos é porque não queriam que os anticorpos reagissem de forma cruzada com *MLH1*. O fragmento de proteína *PMS1* de rato foi altamente expresso em *E. coli*., purificada a partir de um gel de poliacrilamida e preparou-se então a proteína eluída para injeções de animais. Enviaram-se aproximadamente 2 mg do fragmento de proteína *PMS1* para a Pocono Rabbit Farm (PA) para injeções em coelhos. Os soros dos coelhos foram titulados múltiplas

vezes em função do antigénio de PMS1 utilizando técnicas padronizadas de ELISA. Os anticorpos de coelhos específicos para a proteína PMS1 de rato foram purificados por afinidade utilizando colunas contendo a proteína PMS1 de rato imobilizada. A preparação do anticorpo policlonal purificado por afinidade foi ensaiada utilizando ainda a análise de mancha de Western e a mancha de pontos. Os requerentes verificaram que os anticorpos policlonais reconheceram não só a proteína PMS1 de rato, mas também a proteína PMS1 humana que é muito semelhante. Com base nas manchas de Western, não há indicação de que outras proteínas fossem fortemente reconhecidas pelos anticorpos dos requerentes, incluindo as proteínas *MLH1* quer de rato quer humanas.

Ratos deficientes na reparação do desemparelhamento do ADN.

EXEMPLO 6: Para criar um sistema de um modelo experimental para estudar os defeitos de reparação do desemparelhamento do ADN e o cancro resultante num modelo completo de animais, os requerentes derivaram ratos deficientes na reparação do desemparelhamento do ADN utilizando a tecnologia das células indiferenciadas embrionárias (IE). Utilizando o ADN genómico que contém uma porção do gene de *PMS1* de rato os requerentes construíram um vector que, após recombinação homóloga causa um rompimento do gene cromossómico de *PMS1* de rato. As células IE de rato de 129 estirpes de ratos foram confirmadas como contendo um alelo de *PMS1* de rato com rompimento. As células de IE foram injectadas em blastócitos hospedeiros de C57/BL6 para produzir animais que eram quiméricos ou uma mistura das células dos 129 e dos C57/BL6. A incorporação das células IE foi determinada pela presença de adesivos de coloração de revestimento de cutias (indicativo da

contribuição das células IE). Todos os machos quiméricos foram criados com ratos fêmea C57/BL6.

Em seguida, nasceram doze crias (F_2) em que se detectou a cor da pele de cutia indicando a transmissão da linha de células do material genético das células IE. As análises do ADN extraído das pontas das caudas das crias indicaram que seis dos animais eram heterozigóticos (continham um alelo de tipo selvagem e um alelo mutante) para a mutação de *PMS1* de rato. Dos seis animais heterozigóticos, três eram fêmeas (animais F_2 -8, F_2 -11 e F_2 -12) e três eram machos (F_2 , F_2 -10 e F_2 -13). Para a criação criaram-se cercados para se obter ratos que eram homozigóticos para a mutação de *rPMS1* e mais ratos heterozigóticos. O cercado de criação #1 que continha os animais F_2 -11 e F_2 -10, originaram um total de treze ratos em três ninhadas, quatro dos quais tinham sido genotipados. O cercado de criação #2 (animais F_2 -8 e F_2 -13) originaram vinte e dois animais e em três ninhadas, três dos quais tinham sido genotipados. Dos sete animais genotipados, identificaram-se três animais fêmeas homozigóticas. Um animal morreu às seis semanas de idade por causas desconhecidas. As restantes fêmeas homozigóticas estavam vivas e de saúde às doze semanas de idade. Os resultados indicam que os ratos homozigóticos com defeitos da *PMS1* de rato, são viáveis.

Utilizaram-se os cercados de criação #3 e #4 para cruzar novamente a mutação *PMS1* de rato nos ratos C57/BL6 anteriores. O cercado de criação #3 (animal F_2 -12 cruzado com um rato C57/BL6) produziu vinte e um animais em duas ninhadas, nove dos quais tinham sido genotipados. O cercado de criação #4 (animal F_2 -6 cruzado com um rato C57/BL6) deu

oito ratos. Além disso, o macho quimérico original (cercado de criação #5) tinha produzido mais trinta e uma crias.

Para genotipar os animais, desenvolveu-se uma série de iniciadores de RCP que foram utilizados para identificar genes de *PMS1* de rato mutantes e de tipo selvagem. Eles são: (SEQ ID NOS: 143-148, respectivamente)

Iniciador 1: 5'TTCGGTGACAGATTTGTAAATG-3'

Iniciador 2: 5'TTTACGGAGCCCTGGC-3'

Iniciador 3: 5'TCACCATAAAAATAGTTTCCCG-3'

Iniciador 4: 5'TCCTGGATCATATTTTCTGAGC-3'

Iniciador 5: 5'TTTCAGGTATGTCCTGTTACCC-3'

Iniciador 6: 5'TGAGGCAGCTTTTAAGAACTC-3'

Iniciadores 1 + 2 (5' marcado)

Iniciadores 1 + 3 (5' não marcado)

Iniciadores 4 + 5 (3' marcado)

Iniciadores 4 + 6 (3' não marcado)

Os ratos desenvolvidos pelos requerentes originam um sistema de modelo de animais para estudar as consequências dos defeitos na reparação do desemparelhamento do ADN e no CCNPH resultante. A sobrevivência de longo prazo dos ratos homozigóticos e heterozigóticos para a mutação de *PMS1* de rato e os tipos e tempos dos tumores nestes ratos serão determinados. Os ratos serão avaliados diariamente no que respeita a qualquer indicação de início de cancro tal como indicado por uma aparência de corcova em combinação com a deterioração das condições de revestimento. Estes ratos portadores da mutação de *PMS1* de rato serão utilizados para ensaiar os efeitos de outros factores, ambientais e genéticos, na formação do tumor. Por exemplo, o efeito da dieta no cólon e noutros tipos de tumores pode ser comparado para os animais normais versus os que são portadores da mutação de *PMS1* de rato quer no genotipo

heterozigótico quer homozigótico. Além disso, a mutação de *PMS1* de rato pode ser posta em diferentes contextos genéticos anteriores para se aprender acerca das interações entre os genes da reparação do desemparelhamento e outros genes envolvidos no cancro humano, por exemplo p53. Os ratos portadores de mutações de *PMS1* de rato serão úteis para ensaiar a eficácia da terapia somática de genes nos cancros que aparecem em ratos, por exemplo, os expectáveis cancros do cólon. Além disso, as linhas de células de fibroblastos isogénicos de *rPMS1* de ratos heterozigóticos e homozigóticos podem ser estabelecidas para serem utilizadas em vários estudos celulares, incluindo a determinação das taxas de mutação espontâneas.

Os requerentes estão actualmente a construir um vector para romper o gene *rMLH1* de rato para derivar ratos que comportam mutação em *rMLH1*. Os requerentes vão comparar ratos portadores de defeitos na *rPMS1* com ratos portadores de defeitos na *rMLH1*. Além disso, os requerentes vão produzir ratos que são portadores de mutações em ambos os genes de CCNPH. Outros estudos sobre os ratos mutantes em *rMLH1* serão tal como se descreveu antes para os ratos mutantes em *rPMS1*.

LISTAGEM DE SEQUÊNCIAS

(1) INFORMAÇÃO GERAL

i) REQUERENTE: Liskay, Robert M.
Bronner, C. Eric
Baker, Sean M.
Bollag, Roni J.
Kolodner, Richard D.

(ii) TÍTULO DA INVENÇÃO: COMPOSIÇÕES E PROCESSOS
RELACIONADOS COM OS GENES DE REPARAÇÃO DO DESEM-
PARELHAMENTO DO ADN

(iii) NUMERO DE SEQUÊNCIAS: 148

(iv) ENDEREÇO PARA CORRESPONDÊNCIA:

(A) ENDEREÇO: Kolisch, Hartwell, Oickinson,
McCormack & Heuser

(B) RUA: 520 S.W. Yamhill Street, Suite 200

(C) CIDADE: Portland

(D) ESTADO: Oregon

(E) PAÍS: E.U.A.

(F) CÓDIGO POSTAL: 97204

(v) FORMA LISÍVEL EM COMPUTADOR:

(A) TIPO DE MEIO: Floppy disk

(B) COMPUTADOR: compatível com IBM PC

(C) SISTEMA OPERATIVO: PC-DOS/MS-DOS

(D) SOFTWARE: Patentln Release #1.0, Versão #1.25

(vi) DADOS DO PRESENTE PEDIDO DE PATENTE DE INVENÇÃO:

(A) NÚMERO DO PEDIDO DE PATENTE DE INVENÇÃO:

(B) DATA DE REGISTO:

(C) CLASSIFICAÇÃO:

(viii) INFORMAÇÃO SOBRE O ADVOGADO /AGENTE:

(A) NOME: Van Ryseelberghe, Pierre C.

(B) NÚMERO DE REGISTO: 33.557

(C) REFERÊNCIA/NÚMERO DE ARMAZENAGEM: OHSU 306B

(ix) INFORMAÇÃO DE TELECOMUNICAÇÕES:

- (A) TELEFONE: (503) 224-6655
- (B) TELEFAX: (503) 295-6679
- (C) TELEX: 360619

(2) INFORMAÇÃO PARA A SEQ ID NO: 1:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 361 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 1:

Met	Pro	Ile	Gln	Val	Leu	Pro	Pro	Gln	Leu	Ala	Asn	Gln	Ile	Ala	Ala
1				5					10					15	
Gli	Glu	Val	Val	Glu	Arg	Pro	Ala	Ser	Val	Val	Lis	Glu	Leu	Val	Glu
			20					25					30		
Asn	Ser	Leu	Asp	Ala	Gli	Ala	Tre	Arg	Val	Asp	Ile	Asp	Ile	Glu	Arg
		35					40					45			
Gli	Gli	Ala	Lis	Leu	Ile	Arg	Ile	Arg	Asp	Asn	Gli	Cis	Gli	Ile	Lis
	50					55				60					
Lis	Glu	Glu	Leu	Ala	Leu	Ala	Leu	Ala	Arg	His	Ala	Tre	Ser	Lis	Ile
65				70					75					80	
Ala	Ser	Leu	Asp	Asp	Leu	Glu	Ala	Ile	Ile	Ser	Leu	Gli	Fen	Arg	Gli
			85					90					95		
Glu	Ala	Leu	Ala	Ser	Ile	Ser	Ser	Val	Ser	Arg	Leu	Tre	Leu	Tre	Ser
		100						105					110		
Arg	Tre	Ala	Glu	Gln	Ala	Glu	Ala	Trp	Gln	Ala	Tir	Ala	Glu	Gli	Arg
	115					120						125			
Asp	Met	Asp	Val	Tre	Val	Lis	Pro	Ala	Ala	His	Pro	Val	Gli	Tre	Tre
	130					135					140				
Leu	Glu	Val	Leu	Asp	Leu	Fen	Tir	Asn	Tre	Pro	Ala	Arg	Arg	Lis	Fen

145	150	155	160
Met Arg Tre Glu Lis Tre Glu Fen Asn His Ile Asp Glu Ile Ile Arg			
	165	170	175
Arg Ile Ala Leu Ala Arg Fen Asp Val Tre Leu Asn Leu Ser His Asn			
	180	185	190
Gli Lis Leu Val Arg Gln Tir Arg Ala Val Ala Lis Asp Gli Gln Lis			
	195	200	205
Glu Arg Arg Leu Gli Ala Ile Cis Gli Tre Pro Fen Leu Glu Gln Ala			
	210	215	220
Leu Ala Ile Glu Trp Gln His Gli Asp Lis Tre Lis Arg Gli Trp Val			
225	230	235	240
Ala Asp Pro Asn His Tre Tre Tre Ala Leu Tre Glu Ile Gln Tir Cis			
	245	250	255
Tir Val Asn Gli Arg Met Met Arg Asp Arg Leu Ile Asn His Ala Ile			
	260	265	270
Arg Gln Ala Cis Glu Asp Lis Leu Gli Ala Asp Gln Gln Pro Ala Fen			
	275	280	285
Val Leu Tir Leu Glu Ile Asp Pro His Gln Val Asp Val Asn Val His			
	290	295	300
Pro Ala Lis His Glu Val Arg Fen His Gln Ser Arg Leu Val His Asp			
305	310	315	320
Fen Ile Tir Gln Gli Val Leu Ser Val Leu Gln Gln Gln Tre Glu Tre			
	325	330	335
Ala Leu Pro Leu Glu Glu Ile Ala Pro Ala Pro Arg His Val Gln Glu			
	340	345	350
Asn Arg Ile Ala Ala Gli Arg Asn His			
	355	360	

(2) INFORMAÇÃO PARA A SEQ ID NO: 2:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 538 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIAL: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) SEQUÊNCIA OESCRPTION: SEQ ID NO: 2:

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

Gli Ser Lis Leu Met Val Gli Arg Fen Pro Leu Ala Val Ile His Ile
 275 280 285
 His Ile Asp Pro Tir Leu Ala Asp Val Asn Val His Pro Tre Lis Gln
 290 295 300
 Gln Val Arg Ile Ser Lis Gln Lis Gln Leu Met Tre Leu Val Ser Gln
 305 310 315
 Ala Ile Ala Asn Ser Leu Lis Gln Tre Leu Ile Pro Asp Ala Leu
 325 330 335
 Gln Asn Leu Ala Lis Ser Tre Val Arg Asn Arg Gln Lis Val Gln Gln
 340 345 350
 Tre Ile Leu Pro Leu Ser Fen Pro Gln Leu Fen Fen Gli Gln Met
 355 360 365
 His Gli Tre Tir Leu Fen Ala Gln Gli Arg Asp Gli Leu Tir Ile Ile
 370 375 380
 Asp Gln His Ala Ala Gln Gln Arg Val Lis Tir Gln Tir Arg Gln
 385 390 395
 Ser Ile Gli Asn Val Asp Gln Ser Gln Gln Gln Leu Val Pro Tir
 405 410 415
 Ile Fen Gln Fen Pro Ala Asp Asp Ala Leu Arg Leu Lis Gln Arg Met
 420 425 430
 Pro Leu Leu Gln Gln Val Gli Val Fen Leu Ala Gln Tir Gli Gln Asn
 435 440 445
 Gln Fen Ile Leu Arg Gln His Pro Ile Trp Met Ala Gln Gln Ile
 450 455 460
 Gln Ser Gli Ile Tir Gln Met Cys Asp Met Leu Leu Tre Lis Gln
 465 470 475
 Val Ser Ile Lis Lis Tir Arg Ala Gln Leu Ala Ile Met Met Ser Cys
 485 490 495
 Lis Arg Ser Ile Lis Ala Asn His Arg Ile Asp Asp Ser Ala Arg
 500 505 510
 Gln Leu Leu Tir Gln Leu Ser Gln Cys Asp Asn pro Tir Asn Cys Pro
 515 520 525
 His Gli Arg Pro Val Leu Val His Fen Tre
 530 535

(2) INFORMAÇÃO PARA SEQ ID NO:3:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 607 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIAL: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 3:

```
Met Fen His His Ile Glu Asn Leu Leu Ile Glu Tre Glu Lis Arg Cis
1           5           10           15
Lis Gln Lis Glu Gln Arg Tir Ile Pro Val Lis Tir Leu Fen Ser Met
          20          25          30
Tre Gln Ile His Gln Ile Asn Asp Ile Asp Val His Arg Lis Tre Ser
        35          40          45
Gli Gln Val Ile Tre Asp Leu Tre Tre Ala Val Lis Glu Leu Val Asp
        50          55          60
Asn Ser Ile Asp Ala Asn Ala Asn Gln Ile Glu Ile Ile Fen Lis Asp
65          70          75          80
Tir Gli Leu Glu Ser Ile Glu Cis Ser Asp Asn Gli ASp Gli Ile Asp
          85          90          95
Pro Ser Asn Tir Glu Fen Leu Ala Leu Lis His Tir Tre Ser Lis Ile
        100         105         110
Ala Lis Fen Gln Asp Val Ala Lis Val Gln Tre Leu Gli Fen Arg Gli
        115         120         125
Glu Ala Leu Ser Ser Leu Cis Gli Ile Ala Lis Leu Ser Val Ile Tre
        130         135         140
Tre Tre Ser Pro Pro Lis Ala Asp Lis Glu Leu Tir Asp Met Val Gli
145         150         155         160
His Ile Tre Ser Lis Tre Tre Tre Ser Arg Asn Lis Gli Tre Tre Val
          165          170          175
Leu Val Ser Gln Leu Fen His Asn Leu Pro Val Arg Gln Lis Glu Fen
        180         185         190
Ser Lis Tre Fen Lis Arg Gln Fen Tre Lis Cis Leu Tre Val Ile Gln
        195         200         205
Gli Tir Ala Ile Ile Asn Ala Ala Ile Lis Fen Ser Val Trp Asn Ile
```

210	215	220
Tre Pro Lis Gli Lis Lis Asn Leu Ile Leu Ser Tre Met Arg Asn Ser		
225	230	235
Ser Met Arg Lis Asn Ile Ser Ser Val Fen Gli Ala Gli Gli Met Arg		240
	245	250
		255
Gli Glu Leu Glu Val Aap Leu Val Leu Asp Leu Asn Pro Fen Lis Asn		
	260	265
		270
Arg Met Leu Gli Lis Tir Tre Asp Asp Pro Asp Fen Leu Asp Leu Asp		
	275	280
		285
Tir Lis Ile Arg Val Lis Gli Tir Ile Ser Gln Asn Ser Fen Gli Cis		
290	295	300
Gli Arg Asn Ser Lis Asp Arg Gln Fen Ile Tir Val Asn Lis Arg Pro		
305	310	315
		320
Val Glu Tir Ser Tre Leu Leu Lis Cis Cis Asn Glu Val Tir Lis Tre		
	325	330
		335
Fen Asn Asn Val Gln Fen pro Ala Val Fen Leu Asn Leu Glu Leu Pro		
	340	345
		350
Met Ser Leu Ile Asp Val Asn Val Tre pro Asp Lis Arg Val Ile Leu		
	355	360
		365
Leu His Asn Glu Arg Ala Val Ile Asp Ile Fen Lis Tre Tre Leu Ser		
370	375	380
Asp Tir Tir Asn Arg Gln Glu Leu Ala Leu Pro Lis Arg Met Cis Ser		
385	390	395
		400
Gln Ser Glu Gln Gln Ala Gln Lis Arg Leu Leu Tre Glu Val Fen Asp		
	405	410
		415
Asp Asp Fen Lis Lis Met Glu Val Val Gli Gln Fen Asn Leu Gli Fen		
	420	425
		430
Ile Ile Val Tre Arg Lis Val Asp Asn Lis Ser Asp Leu Fen Ile Val		
435	440	445
Asp Gln His Ala Ser Asp Glu Lis Tir Asn Fen Glu Tre Leu Gln Ala		
450	455	460
Val Tre Val Fen Lis Ser Gln Lis Leu Ile Ile Pro Gln Pro Val Glu		
465	470	475
		480
Leu Ser Val Ile Asp Glu Leu Val Val Leu Asp Asn Leu Pro Val Fen		
	485	490
		495
Glu Lis Asn Gli Fen Lis Leu Lis Ile Asp Glu Glu Glu Glu Fen Gli		
	500	505
		510

Ser Arg Val Lis Leu Leu Ser Leu Pro Tre Ser Lis Gln Tre Leu Fen
515 520 525
Asp Leu Gli Asp Fen Asn Glu Leu Ile His Leu Ile Lis Glu Asp Gli
530 535 540
Gli Leu Arg Arg Asp Asn Ile Arg Cis Ser Lis Ile Arg Ser Met Fen
545 550 555 560
Ala Met Arg Ala Cis Arg Ser Ser Ile Met Ile Gli Lis Pro Leu Asn
565 570 575
Lis Lis Tre Met Tre Arg Val Val His Asn Leu Ser Glu Leu Asp Lis
580 585 590
Pro Trp Asn Cis Pro His Gli Arg Pro Tre Met Arg His Leu Met
595 600 605

(2) INFORMAÇÃO PARA A SEQ ID NO: 4:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 2484 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 4:

CTTGGCTCTT CTGGCGCCAA AATGTCGTTC GTGGCAGGGG TTATTCGGCG GCTGGACGAG 60
ACAGTGGTGA ACCGCATCGC GGCGGGGGAA GTTATCCAGC GGCCAGCTAA TGCTATCAAA 120
GAGATGATTG AGAACTGTTT AGATGCAAAA TCCACAAGTA TTCAAGTGAT TGTTAAAGAG 180
GGAGGCCTGA AGTTGATTCA GATCCAAGAC AATGGCACCG GGATCAGGAA AGAAGATCTG 240
GATATTGTAT GTGAAAGGTT CACTACTAGT AAAGTGCAGT CCTTTGAGGA TTTAGCCAGT 300
ATTTCTACCT ATGGCTTTTCG AGGTGAGGCT TTGGCCAGCA TAAGCCATGT GGCTCATGTT 360
ACTATTACAA CGAAAACAGC TGATGGAAAG TGTGCATACA GAGCAAGTTA CTCAGATGGA 420
AAAGTAAAG CCCCTCCTAA ACCATGTGCT GGCAATCAAG GGACCCAGAT CACGGTGGAG 480
GACCTTTTTT ACAACATAGC CACGAGGAGA AAAGCTTTAA AAAATCCAAG TGAAGAATAT 540
GGGATAATTT TGGAAGTTGT TGGCAGGTAT TCAGTACACA ATGCAGGCAT TAGTTTCTCA 600
GTTAAAAAAC AAGGAGAGAC AGTAGCTGAT GTTAGGACAC TACCCAATGC CTCAACCGTG 660
GACAATATTC GCTCCATCTT TGGAAATGCT GTTAGTCGAG AACTGATAGA AATTGGATGT 720
GAGGATAAAA CCCTAGCCTT CAAAATGAAT GGTTACATAT CCAATGCAAA CTACTCAGTG 780

AAGAAGTGCA TCTTCTTACT CTTCATCAAC CATCGTCTGG TAGAATCAAC TTCCTTGAGA 840
AAAGCCATAG AAACAGTGTA TGCAGCCTAT TTGCCCCAAA ACACACACCC ATTCCTGTAC 900
CTCAGTTTAG AAATCAGTCC CCAGAATGTG GATGTTAATG TGCACCCAC AAAGCATGAA 960
GTTCACTTCC TGCACGAGGA GAGCATCCTG GAGCGGGTGC AGCAGCACAT CGAGAGCAAG 1020
CTCCTGGGCT CCAATTCCTC CAGGATGTAC TTCACCCAGA CTTTGCTACC AGGACTTGCT 1080
GGCCCCCTCTG GGGAGATGGT TAAATCCACA ACAAGTCTGA CCTCGTCTTC TACTTCTGGA 1140
AGTAGTGATA AGGTCTATGC CCACCAGATG GTTCGTACAG ATTCCTGGGA ACAGAAGCTT 1200
GATGCATTTT TGCAGCCTCT GAGCAAACCC CTGTCCAGTC AGCCCCAGGC CATTGTCACA 1260
GAGGATAAGA CAGATATTTT TAGTGGCAGG GCTAGGCAGC AAGATGAGGA GATGCTTGAA 1320
CTCCCAGCCC CTGCTGAAGT GGCTGCCAAA AATCAGAGCT TGGAGGGGGA TACAACAAAG 1380
GGGACTTCAG AAATGTCAGA GAAGAGAGGA CCTACTTCCA GCAACCCAG AAAGAGACAT 1440
CGGGAAGATT CTGATGTGGA AATGGTGGA GATGATTCCT GAAAGGAAAT GACTGCAGCT 1500
TGTACCCCCC GGAGAAGGAT CATTAACTC ACTAGTGTTC TGAGTCTCCA GGAAGAAATT 1560
AATGAGCAGG GACATGAGGT TCTCCGGGAG ATGTTGCATA ACCACTCCTT CGTGGGCTGT 1620
GTGAATCCTC AGTGGGCCTT GGCACAGCAT CAAACCAAGT TATACCTTCT CAACACCACC 1680
AAGCTTAGTG AAGAACTGTT CTACCAGATA CTCATTTATG ATTTTGCCAA TTTTGGTGTT 1740
CTCAGGTTAT CGGAGCCAGC ACCGCTCTTT GACCTTGCCA TGCTTGCCCT AGATAGTCCA 1800
GAGAGTGGCT GGACAGAGGA AGATGGTCCC AAAGAAGGAC TTGCTGAATA CATTGTTGAG 1860
TTTCTGAAGA AGAAGGCTGA GATGCTTGCA GACTATTTCT CTTTGAAAT TGATGAGGAA 1920
GGGAFCTGA TTGGATTACC CCTTCTGATT GACAACATG TGCCCCCTTT GGAGGGACTG 1980
CCTATCTTCA TTCTTCGACT AGCCACTGAG GTGAATTGGG ACGAAGAAAA GGAATGTTTT 2040
GAAAGCCTCA GTAAAGAATG CGCTATGTTT TATTCCATCC GGAAGCAGTA CATATCTGAG 2100
GAGTCGACCC TCTCAGGCCA GCAGAGTGAA GTGCCTGGCT CCATTCCAAA CTCCTGGAAG 2160
TGGACTGTGG AACACATTGT CTATAAAGCC TTGCGCTCAC ACATTCTGCC TCCTAAACAT 2220
TTCACAGAAG ATGGAATAT CCTGCAGCTT GCTAACCTGC CTGATCTATA CAAAGTCTTT 2280
GAGAGGTGTT AAATATGGTT ATTTATGCAC TGTGGGATGT GTTCTTCTTT CTCTGTATTC 2340
CGATACAAAG TGTGTATCA AAGTGTGATA TACAAAGTGT ACCAACATAA GTGTTGGTAG 2400
CACTTAACAC TTATACTTGC CTTCTGATAG TATTCCTTTA TACACAGTGG ATTGATTATA 2460
AATAAATAGA TGTGTCTTAA CATA 2484

(2) INFORMAÇÃO PARA A SEQ ID NO: 5:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 756 aminoácidos
- (b) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(11) TIPO DE MOLECULA: proteina

(x1) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 5:

Met Ser Fen Val Ala Gl1 Val Ile Arg Arg Leu Asp Glu Tre Val Val
1 5 10 15
Asn Arg Ile Ala Ala Gl1 Glu Val Ile Gln Arg Pro Ala Asn Ala Ile
20 25 30
Lis Glu Met Ile Glu Asn C1s Leu Asp Ala Lis Ser Tre Ser Ile Gln
35 40 45
Val Ile Val Lis Glu Gl1 Gl1 Leu Lis Leu Ile Gln Ile Gln Asp Asn
50 55 60
Gl1 Tre Gl1 Ile Arg Lis Glu Asp Leu Asp Ile Val C1s Glu Arg Fen
65 70 75 80
Tre Tre Ser Lis Leu Gln Ser Fen Glu Asp Leu Ala Ser Ile Ser Tre
85 90 95
Tir Gl1 Fen Arg Gl1 Glu Ala Leu Ala Ser Ile Ser M1s Val Ala His
100 105 110
Val Tre Ile Tre Lis Tre Ala Asp Gl1 Lis C1s Ala Tir Arg Ala
115 120 125
Ser Tir Ser Asp Gl1 Lis Leu Lis Ala Pro Lis Pro C1s Ala Gl1
130 135 140
Asn Gln Gl1 Tre Gln Ile Tre Val Glu Asp Leu Fen Tir Asn Ile Ala
145 150 155 160
Tre Arg Arg Lis Ala Leu Lis Asn Pro Ser Glu Glu Tir Gl1 Lis Ile
165 170 175
Leu Glu Val Val Gl1 Arg Tir Ser Val His Asn Ala Gl1 Ile Ser Fen
180 185 190
Ser Val Lis Lis Gln Gl1 Glu Tre Val Ala Asp Val Arg Tre Leu Pro
195 200 205
Asn Ala Ser Tre Val Asp Asn Ile Arg Ser Ile Fen Gl1 Asn Ala Val
210 215 220
Ser Arg Glu Leu Ile Glu Ile Gl1 C1s Glu Asp Lis Tre Leu Ala Fen
225 230 235 240
Lis Met Asn Gl1 Tir Ile Ser Asn Ala Asn Tir Ser Val Lis Lis C1s
245 250 255
Ile Fen Leu Leu Fen Ile Asn 8le Arg Leu Val Glu Ser Tre Ser Leu
260 265 270

Arg Lis Ala Ile Glu Tre Val Tir Ala Ala Tir Leu Pro Lis Asn Tre
 275 280 285
 His Pro Fen Leu Tir Leu Ser Leu Glu Ile Ser Pro Gln Asn Val Asp
 290 295 300
 Val Asn Val His Pro Tre Lis His Glu Val His Fen Leu His Glu Glu
 305 310 315 320
 Ser Ile Leu Glu Arg Val Gln Gln His Ile Glu Ser Lis Leu Leu Gli
 325 330 335
 Ser Asn Ser Ser Arg Met Tir Fen Tre Gln Tre Leu Leu Pro Gli Leu
 340 345 350
 Ala Gli Pro Ser Gli Glu Met Val Lis Ser Tre Tre Ser Leu Tre Ser
 355 360 365
 Ser Ser Tre Ser Gli Ser Ser Asp Lis Val Tir Ala His Gln Met Val
 370 375 380
 Arg Tre Asp Ser Arg Glu Gln Lis Leu Asp Ala Fen Leu Gln Pro Leu
 385 390 395 400
 Ser Lis Pro Leu Ser Ser Gln Pro Gln Ala Ile Val Tre Glu Asp Lis
 405 410 415
 Tre Asp Ile Ser Ser Gli Arg Ala Arg Gln Gln Asp Glu Glu Met Leu
 420 425 430
 Glu Leu Pro Ala pro Ala Glu Val Ala Ala Lis Asn Gln Ser Leu Glu
 435 440 445
 Gli Asp Tre Tre Lis Gli Tre Ser Glu Met Ser Glu Lis Arg Gli Pro
 450 455 450
 Tre Ser Ser Asn Pro Arg Lis Arg His Arg Glu AsP Ser Asp Val Glu
 465 470 475 480
 Met Val Glu Asp Asp Ser Arg Lis Glu Met Tre Ala Ala Cis Tre Pro
 485 490 495
 Arg Arg Arg Ile Ile Asn Leu Tre Ser Val Leu Ser Leu Gln Glu Glu
 500 505 510
 Ile Asn Glu Gln Gli His Glu Val Leu Arg Glu Met Leu His Asn His
 515 520 525
 Ser Fen Val Gli Cis Val Asn Pro Gln Trp Ala Leu Ala Gln His Gln
 530 535 540
 Tre Lis Leu Tir Leu Leu Asn Tre Tre Lis Leu Ser Glu Glu Leu Fen
 545 550 555 560
 Tir Gln Ile Leu Ile Tir Asp Fen Ala Asn Fen Gli Val Leu Arg Leu

	565		570		575
Ser Glu Pro Ala	Pro Leu Fen Asp	Leu Ala Met	Leu Ala	Leu Asp	Ser
	580		585		590
Pro Glu Ser Gli	Trp Tre Glu Glu	Asp Gli Pro	Lis Glu Gli	Leu Ala	
	595		600		605
Glu Tir Ile Val	Glu Fen Leu Lis	Lis Lis Ala	Glu Met	Leu Ala	Asp
	610		615		620
Tir Fen Ser Leu	Glu Ile Asp Glu	Glu Gli Asn	Leu Ile	Gli Leu	Pro
625		630		635	640
Leu Leu Ile Asp	Asn Tir Val pro	Pro Leu Glu	Gli Leu	Pro Ile	Fen
	645		650		655
Ile Leu Arg Leu	Ala Tre Glu Val	Asn Trp Asp	Glu Glu	Lis Glu	Cis
	660		665		670
Fen Glu Ser Leu	Ser Lis Glu Cis	Ala Met Fen	Tir Ser	Ile Arg	Lis
	675		680		685
Gln Tir Ile Ser	Glu Glu Ser Tre	Leu Ser Gli	Gln Gln	Ser Glu	Val
	690		695		700
Pro Gli Ser Ile	Pro Asn Ser Trp	Lis Trp Tre	Val Glu	His Ile	Val
705		710		715	720
Tir Lis Ala Leu	Arg Ser His Ile	Leu Pro Pro	Lis His	Fen Tre	Glu
	725		730		735
Asp Gli Asn Ile	Leu Gln Leu Ala	Asn Leu pro	Asp Leu	Tir Lis	Val
	740		745		750
Fen Glu Arg Cis					
	755				

(2) INFORMATION PARA A SEQ ID NO: 6:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 397 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 6:

```
TGGCTGGATG CTAAGCTACA GCTGAAGGAA GAACGTGAGC ACGAGGCACT GAGGTGATTG 60
CCTGAAGGCA CTTCCGTGTA GCATCTAGAC GTTTCCTTGG CTCTTCTGGC GCCAAAATGT 120
CGTTCGTGGC AGGGGTTATT CGGCGGCTGG ACGAGACAGT GGTGAACCGC ATCGCGGCGG 180
GGGAAGTTAT CCAGCGGCCA GCTAATGCTA TCAAAGAGAT GATTGAGAAC TGGTACGGAG 240
GGAGTCGAGC CGGGCTCACT TAAGGGCTAC GACTTAACGG GCCGCGTCAC TCAATGGCGC 300
GGACACGCCT CTTTCCCCGG GCAGAGGCAT GTACAGCGCA TGCCCAACAAC GGCGGAGGCC 360
GCOGGGTTCC CTACGTGCCA TAAGCCTTCT CCTTTTC 397
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 7:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 393 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genômico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 7:

```
AAACACGTTA ATGAGGCACT ATTGTTTGTA TTTGGAGTTT GTTATCATTG CTTGGCTCAT 60
ATTAAAATAT GTACATTAGA GTAGTTGCAG ACTGATAAAT TATTTTCTGT TTGATTTGCC 120
AGTTTAGATG CAAAATCCAC AAGTATTCAA GTGATTGTTA AAGAGGGAGG CCTGAAGTTG 180
ATTGAGATCC AAGACAATGG CACCGGATC AGGGTAAGTA AAACCTCAAA GTAGCAGGAT 240
GTTTGTGCGC TTCATGGAAG AGTCAGGACC TTTCTCTGTT CTGGAAACTA GGCTTTTGCA 300
GATGGGATTT TTTCAGTAA AAATTCAACA CCAACAATAA ATATTTATTG AGTACCTATT 360
ATTGCGGGG CACTGTTTCT GGGATGTGTC ACT 393
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 8:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 352 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 8:

```
TTTCCTGGAT TAATCAAGAA ATGGAATCA AAGAGATTTG GAAAATGAGT AACATGATTA 60
TTTACTCATC TTTTGGTAT CTAACAGAAA GAAGATCTGG ATATTGTATG TGAAAGGTTT 120
ACTACTAGTA AACTGCAGTC CTTTGAGGAT TTAGCCAGTA TTTCTACCTA TGGCTTTTGA 180
GGTGAGGTAA GCTAAAGATT CAAGAAATGT GTAAAATATC CTCCTGTGAT GACATTGTCT 240
GTCATTTGTT AGTATGTATT TCTCAACATA GATAAATAAG GTTTGGTACC TTTTACTTGT 300
TAAATGTATG CAAATCTGAG CAACTTAAT GAACTTTAAC TTTCAAAGAC TG 352
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 9:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 287 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 9:

```
TGGAAGCAGC GCAGATAAC CTTTCCCTTT GGTGAGGTGA CAGTGGGTGA CCCAGCAGTG 60
AGTTTTTCTT TCAGTCTATT TTCTTTTCTT CCTTAGGCTT TGGCCAGCAT AAGCCATGTG 120
GCTCATGTTA CTATTACAAC GAAAACAGCT GATGGAAAGT GTGCATACAG GTATAGTGCT 180
GACTTCTTTT ACTCATATAT ATTCATTCTG AAATGTATTT TGGGCCTAGG TCTCAGAGTA 240
ATCCTGTCTC AACACCAGTG TTATCTTTGG CAGAGATCTT GAGTACG 287
```

(2) INFORMAÇÃO PARA A SEQ ID NO:10:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 336 pares de bases

(B) TIPO: ácido nucleico

- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

- (ii) TIPO DE MOLÉCULA: ADN (genómico)
- (xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 10:

```

TTGATATGAT TTTCTCTTTT CCCCTTGGGA TTAGTATCTA TCTCTCTACT GGATATTAAT   60
TTGTTATATT TTCTCATTAG AGCAAGTTAC TCAGATGGAA AACTGAAAGC CCCTCCTAAA 120
CCATGTGCTG GCAATCAAGG GACCCAGATC ACGGTAAGAA TGGTACATGG GAGAGTAAAT 180
TGTTGAAGCT TTGTTTGTAT AAATATTGGA ATAAAAAATA AAATGCTTC TAAGTTTTC 240
GGGTAATAAT AAAATGAATT TGCACTAGTT AATGGAGGTC CCAAGATATC CTCTAAGCAA 300
GATAAATGAC TATTGGCTTT TTGGCATGGC AGCCTG                               336

```

(2) INFORMAÇÃO PARA A SEQ ID NO: 11:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 275 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

- (ii) TIPO DE MOLÉCULA: ADN (genómico)
- (xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 11:

```

GCTTTTGCCA GGACCATCTT GGGTTTTATT TTCAAGTACT TCTATGAATT TACAAGAAAA 60
ATCAATCTTC TGTTCAAGTG GAGGACCTTT TTTACAACAT AGCCACGAGG AGAAAAGCTT 120
TAAAAAATCC AAGTGAAGAA TATGGGAAAA TTTTGGAAGT TGTTGGCAGG TACAGTCCAA 180
AATCTGGGAG TGGGTCTCTG AGATTGTCA TCAAAGTAAT GTGTTCTAGT GCTCATACAT 240
TGAACAGTTG CTGAGCTAGA TGGTGAAAAG TAAAA                               275

```

(2) INFORMAÇÃO PARA A SEQ ID NO: 12:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 389 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

- (ii) TIPO DE MOLÉCULA: ADN (genómico)
- (xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ 10 NO: 12:

```
CAGCAACCTA TAAAAGTAGA GAGGAGTCTG TGTTTTGACG CAGCACCTTT AGCATTTTTTA 60
TTTGATGAA GTTCTGCTG GTTATTTTT CTGTGGGTAA AATATTAATA GGCTGTATGG 120
AGATATTTTT CTTTATATGT ACCTTTGTTT AGATTACTCA ACTCCACTAA TTTATTTAAC 180
TAAAAGGGGG CTCTGACATC TAGTGTGTGT TTTTGGCAAC TCTTTTCTTA CTCTTTTGTT 240
TTTCTTTTCC AGGTATTCAG TACACAATGC AGGCATTAGT TTCTCAGTTA AAAAAGTAAG 300
TTCTTGTTT ATGGGGGATG GTTTTGT TTTT ATGAAAAGAA AAAAGGGGAT TTTTAATAGT 360
TTGCTGGTGG AGATAAGGTT ATGATGTTT 389
```

- (2) INFORMAÇÃO PARA A SEQ ID NO: 13:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 381 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

- (ii) TIPO DE MOLÉCULA: ADN (genómico)
- (xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 13:

```
ATGTTTCAGT CTCAGCCATG AGACAATAAA TCCTTGTGTC TTCTGCTGTT TGTTTATCAG 60
CAAGGAGAGA CAGTAGCTGA TGTTAGGACA CTACCCAATG CCTCAACCGT GGACAATATT 120
CGCTCCATCT TTGGAAATGC TGTTAGTCGG TATGTCGATA ACCTATATAA AAAAATCTTT 180
TACATTTATT ATCTTGTTTT ATCATTCAT CACATTATTT GGGAACCTTT CAAGATATTA 240
TGTGTGTTAA GAGTTTGCTT TAGTCAAATA CACAGGCTTG TTTTATGCTT CAGATTTGTT 300
AATGGAGTTC TTATTTACAG TAATCAACAC TTTCTAGCTG TATGTAATCT CCTAGATTCT 360
GTGGCGTGAA TCATGTGTT T 381
```

- (2) INFORMAÇÃO PARA A SEQ ID NO: 14:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 526 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 14:

```
ACTGAGTAGC GTAGGTGGGT GAGTGGGTGG GTGGGTGGGT GGGTGGATGG ATGGATGGGA 60
GGATGGGTGG GTGAATGGGT GAACAGACAA ATGGATGGAT GAATGGACAG GCACAGGACG 120
ACCTCAAATG GACCAAGTCT TCGGCGCCCT CATTTACAA AGTTAGTTTA TGGGAAGGAA 180
CCTTGTGTTT TAAATTCGT ATTCTTTTGT AATCTTTGAG TTTTGAGTAT TTTCAAAAGC 240
TTCAGAATCT CTTTCTAAT AGAGAACTGA TAGAAATTGG ATGTGAGGAT AAAACCCCTAG 300
CCTTCAAAAT GAATGGTTAC ATATCCAATG CAAACTACTC AGTGAAGAAG TGCATCTTCT 360
TACTCTTCAT CAACCGTAAG TTAAGGAGAA CCACATGGGA AATCCACTCA CAGGAAACAC 420
CCACAGGGAA TTTTATGGGA CCATGGAAAA ATTCTGAGT CCATAGGTTT GATTAAACAT 480
GGAGAAACCT CATGGCAAAG TTTGGTTTTA TTGGGAAGCA TGTATA 526
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 15:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 434 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 15:

```
ATAGTGGGCT GCAAAGTGGC CACAGGTAAA GGTGCACCTT TCTTCCTGGG GATGTGATGT 60
GCATATCACT ACAGAAATGT CTTTCCTGAG GTGATGTCAT GACTTTGTGT GAATGTACAC 120
CTGTGACCTC ACCCCTCAGG ACAGTTTGA ACTGGTTGCT TTCTTTTAT TGTTTAGATC 180
```

```

GTCTGGTAGA ATCAACTTCC TTGAGAAAAG CCATAGAAAC AGTGTATGCA GCCTATTTGC 240
CCAAAAACAC ACACCCATTC CTGTACCTCA GGTAAATGTAG CACCAAATC CTCAACCAAG 300
ACTCACAAGG AACAGATGTT CTATCAGGCT CTCCTCTTTG AAAGAGATGA GCATGCTAAT 360
AGTACAATCA GAGTGAATCC CATAACCAC TGGCAAAAGG ATGTTCTGTC CCTTCTTACA 420
GGTACAAGGC ACAG 434

```

(2) INFORMAÇÃO PARA A SEQ ID NO: 16:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 458 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 16:

```

CTTACGCAAA GCTACACAGC TCTTAAGTAG CAGTGCCAAT ATTGAACA ACTCAGACTC 60
GAGCCTGAGG TTTTGACCAC TGTGTCATCT GGCCTCAAAT CTTCTGGCCA CCACATACAC 120
CATATGTGGG CTTTTTCTCC CCCTCCCACT ATCTAAGGTA ATTGT'TCTCT CTTATTTTCC 180
TGACAGTTTA GAAATCAGTC CCCAGAATGT GGATGTTAAT GTGCACCCCA CAAAGCATGA 240
AGTTCACTTC CTGCACGAGG AGAGCATCCT GGAGCGGGTG CAGCAGCACA TCGAGAGCAA 300
GCTCCTGGGC TCCAATTCCT CCAGGATGTA CTTACCCAG GTCAGGGCGC TTCTCATCCA 360
GCTACTTCTC TGGGGCCTTT GAAATGTGCC CGGCCAGACG TGAGAGCCCA GATTTTGTCT 420
GTTATTTAGG AACTTTTTTT GAAGTATTAC CTGGATAG 458

```

(2) INFORMAÇÃO PARA A SEQ ID NO: 17:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 618 pares de bases
- (b) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (d) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 17:

```
GATAATTATA CCTCATACTA GCTTCTTTCT TAGTACTGCT CCATTGTTGGG ACCTGTATAT 60
CTATACTTCT TATTCTGAGT CTCTCCACTA TATATATATA TATATATATA TTTTTTTTTT 120
TTTTTTTTTT TAATACAGAC TTTGCTACCA GGACTTGCTG GCCCCTCTGG GGAGATGGTT 180
AAATCCACAA CAAGTCTGAC CTCGTCTTCT ACTTCTGGAA GTAGTGATAA GGTCTATGCC 240
CACCAGATGG TTCGTACAGA TTCCCGGGAA CAGAAGCTTG ATGCATTCTT GCAGCCTCTG 300
AGCAAACCCC TGTCCAGTCA GCCCCAGGCC ATTGTCACAG AGCATAAGAC AGATATTTCT 360
AGTGGCAGGG CTAGGCAGCA AGATGAGGAG ATGCTTGAAC TCCAGCCCC TGCTGAAGTG 420
GcrGCCAAAA ATCAGAGCTT GGAGGGGGAT ACAACAAAGG GGACTTCAGA AATGTCAGAG 480
AAGAGAGGAC CTACTTCCAG CAACCCAGG TATGGCCTTT TGGGAAAAGT ACAGCCTACC 540
TCCTTTATTC TGTAATAAAA CTGCCTTCTA ACTTTGGCTT TTCATGAATC ACTTGCATCT 600
TCTCTCTGCC GACTTCCC 618
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 18:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 478 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 18:

```
CTGTGCTCCA CCACAGGTCA TCCAGCTCTC TAGACCAGCC CAGAGAAGTT GCTTGCTCCC 60
AAATCCAACC CACAAAATTT GGCTAAGTTT AAAACAAGA ATAATAATGA TCTGCACTTC 120
CTTTTCTTCA TTCCAGAAAG AGACATCGGG AAGATTCTGA TGTGAAATG GTGGAAGATG 180
ATTCCCGAAA GGAAATGACT GCAGCTTGTA CCCCCCGAG AAGGATCATT AACCTCACTA 240
GTGTTTTGAG TCTCCAGGAA GAAATTAATG AGCAGGGACA TGAGGGTACG TAAACGCTGT 300
GGCCTGCCTG GGATGCATAG GGCCTCAACT GCCAAGGTTT TGGAATGGA GAAAGCAGTC 360
ATGTTGTCAG AGTGGCACTA CAGTTTTGAT GGGCAAGCTC CTCTTCCTTT ACTAACCAC 420
AATAGCATCA CCTTAAAGAC AATTTTTGAT TGGGAQAAAA GGGAGAAAAT AATCTCTG 478
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 19:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 377 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 19:

```
CAGTTTTAC CAGGAGGCTC AAATCAGGCC TTGCTTACT TGGTGTCTCT AGTTCTGGTC 60
CCTGGTGCTT TGGTCAATGA AGTGGGGTTG GTAGGATTCT ATTACTTACC TGTTTTTTGG 120
TTTTATTTT TGTTCGTCAG TTCTCCGGGA GATGTGTCAT AACCACTCCT TCGTGGGCTG 180
TGTGAATCCT CAGTGGGCCT TGGCACAGCA TCAAACCAAG TTATACCTTC TCAACACCAC 240
CAAGCTTAGG TAAATCAGCT GAGTGTGTGA ACAAGCAGAC CTACTACAAC AATCCTCCAC 300
GGAGCACAGG CACAAAAGCT AAGGAGAGCA CCATGAAGGT AGTTGGGAAG GGCACAGGCT 360
TTGGAGTCAG CACATGT 377
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 20:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 325 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 20:

```
CCCCTCGTTG AAGCGTTGGA ATCCCACTCT TTGGAAGATT GTCTTAGACT CTTAACCAGA 60
TTCCACAGCC ACGCAGAACT ATGTCTGTCT CATCCATGTG TCAGGGATTA CGTCTCCCAT 120
TTGTCCCAAC TGGTTGTATC TCAAGCATGA ATTCAGCTTT TCCTTAAAGT CACTTCATTT 180
TTATTTTCAG TGAAGAACTG TTCTACCAGA TACTCATTTA TGATTTTGCC AATTTTGCTG 240
TTCTCAGGTT ATCGGTAAGT TTAGATCCTT TTCACTTCTG ACATTTC AAC TGACCGCCCC 300
GCAAACAGTA GCTCTCCACT AAATA 325
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 21:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 341 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 21:

```
CATTIATGGT TTCTCACCTG CCATTCTGAT AGTGGATTCT TGGGAATTCA GGCTTCATTT 60
GGATGCTCCG TTAAAGCTTG CTCCTTCATG TTCTTGCTTC TTCCTAGGAG CCAGCACCGC 120
TCTTTGACCT TGCCATGCTT GCCTTAGATA GTCCAGAGAG TGGCTGGACA GAGGAAGATG 180
GTCCCAAAGA AGGACTTGCT GAATACATTG TTGAGTTTCT GAAGAAGAAG GCTGAGATGC 240
TTGCAGACTA TTTCTCTTTG GAAATTGATG AGGTGTGACA GCCATTCTTA TACTTCTGTT 300
GTATTCTCCA AATAAAATTT CCAGCCGGGT GCATTGGCTCA 341
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 22:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 260 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN A (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 22:

```
CAGATAGGAG GCACAAGGCC TGGGAAAGGC ACTGGAGAAA TGGGATTTGT TTAAACTATG 60
ACAGCATTAT TTCTTGTTCC CTTGTCCCTT TTCCTGCAAG CAGGAAGGGA ACCTGATTGG 120
ATTACCCCTT CTGATTGACA ACTATGTGCC CCCTTTGGAG GGACTGCCTA TCTTCATTCT 180
TCGACTAGCC ACTGAGGTCA GTGATCAAGC AGATACTAAG CATTTCGGTA CATGCATGTG 240
TGCTGGAGGG AAAGGGCAAA 260
```

	565	570	575
Ser Glu Pro Ala Pro Leu Fen Asp Leu Ala Met Leu Ala Leu Asp Ser			
	580	585	590
Pro Glu Ser Gli Trp Tre Glu Glu Asp Gli Pro Lis Glu Gli Leu Ala			
	595	600	605
Glu Tir Ile Val Glu Fen Leu Lis Lis Lis Ala Glu Met Leu Ala Asp			
	610	615	620
Tir Fen Ser Leu Glu Ile Asp Glu Glu Gli Asn Leu Ile Gli Leu Pro			
625	630	635	640
Leu Leu Ile Asp Asn Tir Val pro Pro Leu Glu Gli Leu Pro Ile Fen			
	645	650	655
Ile Leu Arg Leu Ala Tre Glu Val Asn Trp Asp Glu Glu Lis Glu Cis			
	660	665	670
Fen Glu Ser Leu Ser Lis Glu Cis Ala Met Fen Tir Ser Ile Arg Lis			
	675	680	685
Gln Tir Ile Ser Glu Glu Ser Tre Leu Ser Gli Gln Gln Ser Glu Val			
	690	695	700
Pro Gli Ser Ile Pro Asn Ser Trp Lis Trp Tre Val Glu His Ile Val			
705	710	715	720
Tir Lis Ala Leu Arg Ser His Ile Leu Pro Pro Lis His Fen Tre Glu			
	725	730	735
Asp Gli Asn Ile Leu Gln Leu Ala Asn Leu pro Asp Leu Tir Lis Val			
	740	745	750
Fen Glu Arg Cis			
	755		

(2) INFORMATION PARA A SEQ ID NO: 6:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 397 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 6:

```
TGGCTGGATG CTAAGCTACA GCTGAAGGAA GAACGTGAGC ACGAGGCACT GAGGTGATTG 60
CCTGAAGGCA CTTCGGTTGA GCATCTAGAC GTTTCCTTGG CTCTTCTGGC GCCAAAATGT 120
CGTTCGTGGC AGGGGTTATT CGGCGGCTGG ACGAGACAGT GGTGAACCGC ATCGCGGCGG 180
GGGAAGTTAT CCAGCGGCCA GCTAATGCTA TCAAAGAGAT GATTGAGAAC TGGTACGGAG 240
GCAGTCGAGC CGGGCTCACT TAAGGGCTAC GACTTAACGG GCCGCGTCAC TCAATGGCGC 300
GGACACGCCT CTTTCCCCGG GCAGAGGCAT GTACAGCGCA TGCCCAACAC GCGGAGGCC 360
GCOGGGTTCC CTACGTGCCA TAAGCCTTCT CCTTTTC 397
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 7:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 393 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genômico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 7:

```
AAACACGTTA ATGAGGCACT ATTGTTTGTA TTTGGAGTTT GTTATCATTG CTGGGCTCAT 60
ATTAAAATAT GTACATTAGA GTAGTTGCAG ACTGATAAAT TATTTTCTGT TTGATTGCCC 120
AGTTTAGATG CAAAATCCAC AAGTATTCAA GTGATTGTTA AAGAGGGAGG CCTGAAGTTG 180
ATTCAGATCC AAGACAATGG CACCGGGATC AGGGTAAGTA AAACCTCAAA GTAGCAGGAT 240
GTTTGTGCGC TTCATGGAAG AGTCAGGACC TTTCTCTGTT CTGGAAACTA GGCTTTTGCA 300
GATGGGATTT TTTCACTGAA AAATTCAACA CCAACAATAA ATATTTATTG AGTACCTATT 360
ATTTGCGGGG CACTGTTTCA GGGATGTGTC AGT 393
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 8:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 352 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 8:

```
TTTCCTGGAT TAATCAAGAA ATGGAATCA AAGAGATTG GAAAATGAGT AACATGATTA 60
TTTACTCATC TTTTGGTAT CTAACAGAAA GAAGATCTGG ATATTGTATG TGAAAGGTTT 120
ACTACTAGTA AACTGCAGTC CTTTGAGGAT TTAGCCAGTA TTTCTACCTA TGGCTTTTCGA 180
GGTGAGGTAA GCTAAAGATT CAAGAAATGT GTAAAATATC CTCCTGTGAT GACATTGTCT 240
GTCATTTGTT AGTATGTATT TCTCAACATA GATAAATAAG GTTTGGTACC TTTTACTTGT 300
TAAATGTATG CAAATCTGAG CAACTTAAT GAACTTTAAC TTCAAAGAC TG 352
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 9:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 287 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 9:

```
TGGAAGCAGC GCAGATAAC CTTTCCCTTT GGTGAGGTGA CAGTGGGTGA CCCAGCAGTG 60
AGTTTTTCTT TCAGTCTATT TTCTTTTCTT CCTTAGGCTT TGGCCAGCAT AAGCCATGTG 120
GCTCATGTTA CTATTACAAC GAAAACAGCT GATGGAAAGT GTGCATACAG GTATAGTGCT 180
GACTTCTTTT ACTCATATAT ATTCATTCTG AAATGTATTT TGGGCCTAGG TCTCAGAGTA 240
ATCCTGTCTC AACACCAGTG TTATCTTTGG CAGAGATCTT GAGTACG 287
```

(2) INFORMAÇÃO PARA A SEQ ID NO:10:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 336 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 10:

```
TTGATATGAT TTTCTCTTTT CCCCTTGGGA TTAGTATCTA TCTCTCTACT GGATATTAAT 60
TTGTTATATT TTCTCATTAG AGCAAGTTAC TCAGATGGAA AACTGAAAGC CCCTCCTAAA 120
CCATGTGCTG GCAATCAAGG GACCCAGATC ACGGTAAGAA TGGTACATGG GAGAGTAAAT 180
TGTTGAAGCT TTGTTTGTAT AAATATTGGA ATAAAAAATA AAATTGCTTC TAAGTTTCA 240
GGGTAATAAT AAAATGAATT TGCACTAGTT AATGGAGGTC CCAAGATATC CTCTAAGCAA 300
GATAAATGAC TATTGGCTTT TTGGCATGGC AGCCTG 336
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 11:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 275 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 11:

```
GCTTTTGCCA GGACCATCTT GGGTTTATT TTCAAGTACT TCTATGAATT TACAAGAAAA 60
ATCAATCTTC TGTTCAAGTG GAGGACCTTT TTTACAACAT AGCCACGAGG AGAAAAGCTT 120
TAAAAAATCC AAGTGAAGAA TATGGGAAAA TTTTGGAAAGT TGTTGGCAGG TACAGTCCAA 180
AATCTGGGAG TGGGTCTCTG AGATTTGTCA TCAAAGTAAT GTGTTCTAGT GCTCATACAT 240
TGAACAGTTG CTGAGCTAGA TGGTGAAAAG TAAAA 275
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 12:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 389 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ 10 NO: 12:

```
CAGCAACCTA TAAAAGTAGA GAGGAGTCTG TGTTTTGACG CAGCACCTTT AGCATTTTAA 60
TTTGGATGAA GTTCTGCTG GTTTATTTTT CTGTGGGTAA AATATTAATA GGCTGTATGG 120
AGATATTTTT CTTTATATGT ACCTTTGTTT AGATTACTCA ACTCCACTAA TTTATTTAAC 180
TAAAAGGGGG CTCTGACATC TAGTGTGTGT TTTTGGCAAC TCTTTTCTTA CTCTTTTGTT 240
TTTCTTTTCC AGGTATTCAG TACACAATGC AGGCATTAGT TTCTCAGTTA AAAAAGTAAG 300
TTCTTGTTTT ATGGGGGATG GTTTTGT TTTT ATGAAAAGAA AAAAGGGGAT TTTTAATAGT 360
TTGCTGGTGG AGATAAGGTT ATGATGTTT 389
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 13:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 381 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 13:

```
ATGTTTCAGT CTCAGCCATG AGACAATAAA TCCTTGTCGTC TTCTGCTGTT TGTTTATCAG 60
CAAGGAGAGA CAGTAGCTGA TGTTAGGACA CTACCCAATG CCTCAACCGT GGACAATATT 120
CGCTCCATCT TTGGAATGC TGTTAGTCGG TATGTCGATA ACCTATATAA AAAAATCTTT 180
TACATTTATT ATCTTGTTTT ATCATTCCAT CACATTATTT GGGAACCTTT CAAGATATTA 240
TGTGTGTTAA GAGTTTGCTT TAGTCAAATA CACAGGCTTG TTTTATGCTT CAGATTTGTT 300
AATGGAGTTC TTATTTACAG TAATCAACAC TTTCTAGCTG TATGTAATCT CCTAGATTCT 360
GTGGCGTGAA TCATGTGTTT T 381
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 14:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 526 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 14:

```
ACTGAGTAGC GTAGGTGGGT GAGTGGGTGG GTGGGTGGGT GGGTGGATGG ATGGATGGGA 60
GGATGGGTGG GTGAATGGGT GAACAGACAA ATGGATeGAT GAATGGACAG GCACAGGACG 120
ACCTCAAATG GACCAAGTCT TCGGCGCCCT CATTTACAA AGTTAGTTA TGGGAAGGAA 180
CCTTGTGTTT TTAAATTCTG ATTCTTTTGT AATCTTTGAG TTTTGAGTAT TTTCAAAAGC 240
TTCAGAATCT CTTTTCTAAT AGAGAACTGA TAGAAATTGG ATGTGAGGAT AAAACCCTAG 300
CCTTCAAAAT GAATGGTTAC ATATCCAATG CAAACTACTC AGTGAAGAAG TGCATCTTCT 360
TACTCTTCAT CAACCGTAAG TTAAAAAGAA CCACATGGGA AATCCACTCA CAGGAAACAC 420
CCACAGGGAA TTTTATGGGA CCATGGAAAA ATTTCTGAGT CCATAGGTTT GATTAAACAT 480
GGAGAAACCT CATGGCAAAG TTTGGTTTTA TTGGGAAGCA TGTATA 526
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 15:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 434 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 15:

```
ATAGTGGGCT GCAAAGTGGC CACAGGTAAA GGTGCACCTT TCTTCCTGGG GATGTGATGT 60
GCATATCACT ACAGAAATGT CTTTCCTGAG GTGATGTCAT GACTTTGTGT GAATGTACAC 120
CTGTGACCTC ACCCCTCAGG ACAGTTTGA ACTGGTTGCT TTCTTTTAT TGTTTAGATC 180
```

```

GTCTGGTAGA ATCAACTTCC TTGAGAAAAG CCATAGAAAC AGTGTATGCA GCCTATTTGC 240
CCAAAAACAC ACACCCATTC CTGTACCTCA GGTAATGTAG CACCAACTC CTCAACCAAG 300
ACTCACAAGG AACAGATGTT CTATCAGGCT CTCCTCTTTG AAAGAGATGA GCATGCTAAT 360
AGTACAATCA GAGTGAATCC CATAACCAC TGGCAAAAGG ATGTTCTGTC CCTTCTTACA 420
GGTACAAGGC ACAG 434

```

(2) INFORMAÇÃO PARA A SEQ ID NO: 16:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 458 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 16:

```

CTTACGCAA GCTACACAGC TCTTAAGTAG CAGTGCCAAT ATTTGAACA ACTCAGACTC 60
GAGCCTGAGG TTTTGACCAC TGTGTCATCT GGCCTCAAAT CTTCTGGCCA CCACATACAC 120
CATATGTGGG CTTTTTCTCC CCCTCCCACT ATCTAAGGTA ATTGT'TCTCT CTTATTTTCC 180
TGACAGTTTA GAAATCAGTC CCCAGAATGT GGATGTTAAT GTGCACCCCA CAAAGCATGA 240
AGTTCAC TTCGACGAGG AGAGCATCCT GGAGCGGGTG CAGCAGCACA TCGAGAGCAA 300
GCTCCTGGGC TCCAATTCCT CCAGGATGTA CTTACCCAG GTCAGGGCGC TTCTCATCCA 360
GCTACTTCTC TGGGGCCTTT GAAATGTGCC CGGCCAGACG TGAGAGCCCA GATTTTTGCT 420
GTTATTTAGG AACTTTTTTT GAAGTATTAC CTGGATAG 458

```

(2) INFORMAÇÃO PARA A SEQ ID NO: 17:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 618 pares de bases
- (b) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (d) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 17:

```
GATAATTATA CCTCATACTA GCTTCTTTCT TAGTACTGCT CCATTGTTGGG ACCTGTATAT 60
CTATACTTCT TATTCTGAGT CTCTCCACTA TATATATATA TATATATATA TTTTTTTTTT 120
TTTTTTTTTT TAATACAGAC TTTGCTACCA GGACTTGCTG GCCCCCTCTGG GGAGATGGTT 180
AAATCCACAA CAAGTCTGAC CTCGTCTTCT ACTTCTGGAA GTAGTGATAA GGTCTATGCC 240
CACCAGATGG TTCGTACAGA TTCCCGGGAA CAGAAGCTTG ATGCATTTCT GCAGCCTCTG 300
AGCAAACCCC TGTCCAGTCA GCCCCAGGCC ATTGTCACAG AGCATAAGAC AGATATTTCT 360
AGTGGCAGGG CTAGGCAGCA AGATGAGGAG ATGCTTGAAC TCCCAGCCCC TGCTGAAGTG 420
GcrGCCAAAA ATCAGAGCTT GGAGGGGGAT ACAACAAAGG GGACTTCAGA AATGTCAGAG 480
AAGAGAGGAC CTAATTCCAG CAACCCAGG TATGGCCTTT TGGGAAAAGT ACAGCCTACC 540
TCCTTTATTC TGTAATAAAA CTGCCTTCTA ACTTTGGCTT TTCATGAATC ACTTGCATCT 600
TCTCTCTGCC GACTTCCC 618
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 18:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 478 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 18:

```
CTGTGCTCCA CCACAGGTCA TCCAGCTCTC TAGACCAGCC CAGAGAAGTT GCTTGCTCCC 60
AAATCCAACC CACAAAATTT GGCTAAGTTT AAAACAAGA ATAATAATGA TCTGCACTTC 120
CTTTTCTTCA TTCCAGAAAG AGACATCGGG AAGATTCTGA TGTGGAAATG GTGGAAGATG 180
ATTCCCGAAA GGAAATGACT GCAGCTTGTA CCCCCGGAG AAGGATCATT AACCTACTA 240
GTGTTTTGAG TCTCCAGGAA GAAATTAATG AGCAGGGACA TGAGGGTACG TAAACGCTGT 300
GGCCTGCCTG GGATGCATAG GGCCTCAACT GCCAAGGTTT TGGAAATGGA GAAAGCAGTC 360
ATGTTGTCAG AGTGGCACTA CAGTTTGTAT GGGCAAGCTC CTCTTCCTTT ACTAACCCAC 420
AATAGCATCA CCTTAAAGAC AATTTTGTAT TGGGAQAAA GGGAGAAAAT AATCTCTG 478
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 19:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 377 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 19:

```
CAGTTTTCAC CAGGAGGCTC AAATCAGGCC TTGCTTACT TGGTGTCTCT AGTTCTGGTC 60
CCTGGTGCTT TGGTCAATGA AGTGGGGTTG GTAGGATTCT ATTACTTACC TGTTTTTTGG 120
TTTTATTTT TGTTCGCGGA GATGTTGCAT AACCACTCCT TCGTGGGCTG 180
TGTGAATCCT CAGTGGGCCT TGGCACAGCA TCAAACCAAG TTATACCTTC TCAACACCAC 240
CAAGCTTAGG TAAATCAGCT GAGTGTGTGA ACAAGCAGAC CTACTACAAC AATCCTCCAC 300
GGAGCACAGG CACAAAAGCT AAGGAGAGCA CCATGAAGGT AGTTGGGAAG GGCACAGGCT 360
TTGGAGTCAG CACATGT 377
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 20:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 325 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genômico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 20:

```
CCCCTCGTTG AAGCGTTGGA ATCCCACTCT TTGGAAGATT GTCTTAGACT CTTAACCAGA 60
TTCCACAGCC ACGCAGAACT ATGTCTGTCT CATCCATGTG TCAGGGATTA CGTCTCCCAT 120
TTGTCCCAAC TGGTTGTATC TCAAGCATGA ATTCAGCTTT TCCTTAAAGT CACTTCATTT 180
TTATTTTCAG TGAAGAACTG TTCTACCAGA TACTCATTTA TGATTTTGCC AATTTTGCTG 240
TTCTCAGGTT ATCGGTAAGT TTAGATCCTT TTTCACTTCTG ACATTTCAAC TGACCGCCCC 300
GCAAACAGTA GCTCTCCACT AAATA 325
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 21:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 341 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 21:

```
CATTTATGGT TTCTCACCTG CCATTCTGAT AGTGGATTCT TGGGAATTCA GGCTTCATTT 60
GGATGCTCCG TTAAAGCTTG CTCCTTCATG TTCTTGCTTC TTCCTAGGAG CCAGCACCGC 120
TCTTTGACCT TGCCATGCTT GCCTTAGATA GTCCAGAGAG TGGCTGGACA GAGGAAGATG 180
GTCCCAAAGA AGGACTTGCT GAATACATTG TTGAGTTTCT GAAGAAGAAG GCTGAGATGC 240
TTGCAGACTA TTTCTCTTTG GAAATTGATG AGGTGTGACA GCCATTCTTA TACTTCTGTT 300
GTATTCTCCA AATAAAATTT CCAGCCGGGT GCATTGGCTCA 341
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 22:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 260 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN A (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 22:

```
CAGATAGGAG GCACAAGGCC TGGGAAAGGC ACTGGAGAAA TGGGATTTGT TTAAACTATG 60
ACAGCATTAT TTCTTGTTCC CTTGTCCTTT TTCCTGCAAG CAGGAAGGGA ACCTGATTGG 120
ATTACCCCTT CTGATTGACA ACTATGTGCC CCCTTTGGAG GGACTGCCTA TCTTCATTCT 180
TCGACTAGCC ACTGAGGTCA GTGATCAAGC AGATACTAAG CATTTGGGTA CATGCATGTG 240
TGCTGGAGGG AAAGGGCAAA 260
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 23:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 340 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 23:

```
CTATATCTTC CCAGCAATAT TCACAGTCCG TTTACAGTTT TAACGCCTAA AGTATCACAT 60
TTCGTTTTT AGCTTTAAGT AGTCTGTGAT CTCCGTTTAG AATGAGAATG TTTAAATTCG 120
TACCTATTTT GAGGTATTGA ATTTCTTTGG ACCAGGTGAA TTGGGACGAA GAAAAGGAAT 180
GTTTTGAAAG CCTCAGTAAA GAATGCGCTA TGTTCTATTC CATCCGGAAG CAGTACATAT 240
CTGAGGAGTC GACCCTCTCA GGCCAGCAGG TACAGTGGTC ATGCACACTG GCACCCAGG 300
ACTAGGACAG GACCTCATAC ATCTTAGGAG ATGAAACTTG 340
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 24:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 563 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 24:

```
AATCCTCTTG TGTTCAGGCC TGTGGATCCC TGAGAGGCTA GCCCACAAGA TCCACTTCAA 60
AAGCCCTAGA TAACACCAAG TCTTTCCAGA CCCAGTGCAC ATCCCATCAG CCAGGACACC 120
AGTGTATGTT GGGATGCAAA CAGGGAGGCT TATGACATCT AATGTGTTXT CCAGAGTGAA 180
GTGCCCTGGCT CCATTCCAAA CTCCTGGAAG TGGACTGTGG AACACATTGT CTATAAGGCC 240
TTGCGCTCAC ACATTCTGCC TCCTAAACAT TTCACAGAAG ATGGAAATAT CCTGCAGCTT 300
```

GCTAACCTGC CTGATCTATA CAAAGTGTTT GAGAGGTGTT AAATATGGTT ATTTATGCAC 360
 TGTGGGATGT GTTCTTCTTT CTCTGTATTC CGATACAAAG TGTGTATCA AAGTGTGATA 420
 TACAAAGTGT ACCAACATAA GTGTTGGTAG CACTTAAGAC TTATACTTGC CTTCTGATAG 480
 TATTCCTTTA TACACAGTGG ATTGATTATA AATAAATAGA TGTGTCTTAA CATAATTCT 540
 TATTTAATTT TATTATGTAT ATA 563

(2) INFORMAÇÃO PARA A SEQ ID NO: 25:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 137 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 25:

CTTGGCTCTT CTGGCGCCAA AATGTCGTTT GTGGCAGGGG TTATTCGGCG GCTGGACGAG 60
 ACAGTGGTGA ACCGCATCGC GGCGGGGGAA GTTATCCAGC GGCCAGCTAA TGCTATCAAA 120
 GAGATGATTG AGAACTG 137

(2) INFORMAÇÃO PARA A SEQ ID NO: 26:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 91 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 26:

TTTAGATGCA AAATCCACAA GTATTCAAGT GATTGTTAAA GAGGGAGGCC TGAAGTTGAT 60
 TCAGATCCAA GACAATGGCA CCGGGATCAG G 91

(2) INFORMAÇÃO PARA A SEQ ID NO: 27:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 99 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (O) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 27:

AAAGAAGATC TGGATATTGT ATGTGAAAGG TTCACTACTA GTAAACTGCA GTCCTTTGAG 60
GATTTAGCCA GTATTCTAC CTATGGCTTT CGAGGTGAG 74

(2) INFORMAÇÃO PARA A SEQ ID NO: 28:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 74 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (DO) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 28:

GCTTTGGCCA GCATAAGCCA TGTGGCTCAT GTTACTATTA CAACGAAAAC AGCTGATGGA 60
AAGTGTGCAT ACAG 74

(2) INFORMAÇÃO PARA A SEQ ID NO: 29:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 73 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 29:

AGCAAGTTAC TCAGATGGAA AACTGAAAGC CCCTCCTAAA CCATGTGCTG GCAATCAAGG 60
GACCCAGATC ACG 73

(2) INFORMAÇÃO PARA A SEQ ID NO: 30:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 92 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (O) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 30:

GTGGAGGACC TTTTTTACAA CATAGCCACG AGGAGAAAAG CTTTAAAAAA TCCAAGTGAA 60
GAATATGGGA AAATTTTGGA AGTTGTTGGC AG 92

(2) INFORMAÇÃO PARA A SEQ ID NO: 31:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 43 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 31:

GTATTCAGTA CACAA TGCAG GCATT AGTTT CTCAGTTAAA AAA 43

(2) INFORMAÇÃO PARA A SEQ ID NO: 32:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 89 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 32:

CAAGGAGAGA CAGTAGCTGA TGTTAGGACA CTACCCAATG CCTCAACCGT GGACAATATT 60
CGCTCCATCT TTGGAAATGC TGTTAGTCG 89

(2) INFORMAÇÃO PARA A SEQ ID NO: 33:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 113 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 33:

AGAACTGATA GAAATTGGAT GTGAGGATAA AACCTAGCC TTCAAAATGA ATGGTTACAT 60
ATCCAATGCA AACTACTCAG TGAAGAAGTG CATCTTCTTA CTCTTCATCA ACC 113

(2) INFORMAÇÃO PARA A SEQ ID NO: 34:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 94 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:34:

```
ATCGTCTGGT AGAATCAACT TCCTTGAGAA AAGCCATAGA AACAGTGTAT GCAGCCTATT 60
TGCCCAAAAA CACACACCCA TTCCTGTACC TCAG                                     94
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 35:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 154 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 35:

```
TTTAGAAATC AGTCCCCAGA ATGTGGATGT TAATGTGCAC CCCACAAAGC ATGAAGTTCA 60
CTTCCTGCAC GAGGAGAGCA TCCTGGAGCG GGTGCAGCAG CACATCGAGA GCAAGCTCCT 120
GGGCTCCAAT TCCTCCAGGA TGTACTTCAC CCAG                                     154
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 36:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 371 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:36:

```

ACTTTGCTAC CAGGACTTGC TGGCCCCTCT GGGGAGATGG TTAAATCCAC AACAACTCTG 60
ACCTCGTCTT CTACTTCTGG AAGTAGTGAT AAGGTCTATG CCCACCAGAT GGTTCGTACA 120
GATTCCCGGG AACAGAAGCT TGATGCATTT CTGCAGCCTC TGAGCAAACC CCTGTCCAGT 180
CAGCCCCAGG CCATTGTCAC AGAGGATAAG ACAGATATTT CTAGTGGCAG GGCTAGGCAG 240
CAAGATGAGG AGATGCTTGA ACTCCAGCC CCTGCTGAAG TGGCTGCCAA AAATCAGAGC 300
TTGGAGGGGG ATACAACAAA GGGGACTTCA GAAATGTCAG AGAAGAGAGG ACCTACTTCC 360
AGCAACCCCA G 371

```

(2) INFORMAÇÃO PARA A SEQ ID NO: 37:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 149 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 37:

```

AAAGAGACAT CCGGAAGATT CTGATGTGGA AATGGTGGAA GATGATTCCC GAAAGGAAAT 60
GACTGCAGCT TGTACCCCCC GGAGAAGGAT CATTAACTC ACTAGTGTTC TGAGTCTCCA 120
GGAAGAAATT AATGAGCAGG GACATGAGG 149

```

(2) INFORMAÇÃO PARA A SEQ ID NO: 38:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 109 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 38:

```
TTCTCCGGGA GATGTTGCAT AACCACTCCT TCGTGGGCTG TGTGAATCCT CAGTGGGCCT 60
TGCCACAGCA TCAAACCAAG TTATACCTTC TCAACACCAC CAAGCTTAG          109
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 39:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 64 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 39:

```
TGAAGAACTG TTCTACCAGA TACTCATTTA GATTTTGCC AATTTTGGTG TTCTCAGGTT 60
ATCG                                          64
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 40:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 165 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 40:

```
GAGCCAGCAC CGCTCTTTGA CCTTGCCATG CTGCGCTTAG ATAGTCCAGA GAGTGGCTGG 60
ACAGAGGAAG ATGGTCCCAA AGAAGGACTT GCTGAATACA TTGTTGAGTT TCTGAAGAAG 120
AAGGCTGAGA TGCTTGCAGA CTATTTCTCT TTGAAATTO ATGAG 165
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 41:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 93 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 41:

```
GAAGGGAACC TGATTGGATT ACCCCTTCTG ATTGACAACT ATGTGCCCCC TTTGGAGGGA 60
CTGCCTATCT TCATTCTTCG ACTAGCCACT GAG 93
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 42:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 114 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 42:

```
GTGAATTGGG ACGAAGAAAA GGAATGTTTT GAAAGCCTCA GTAAAGAATG CGCTATGTTT 60
TATTCCATCC GGAAGCAGTA CATATCTGAG GAGTCGACCC TCTCAGGCCA GCAG 114
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 43:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 360 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 43:

```
AGTGAAGTGC CTGGCTCCAT TCCAAACTCC TGGAAGTGGG CTGTGGAACA CATTGTCTAT 60
AAAGCCTTGC GCTCACACAT TCTGCCTCCT AAACATTTCG CAGAAGATGG AAATATCCTG 120
CAGCTTGCTA ACCTGCCTGA TCTATACAAA GTCTTTGAGA GGTGTTAAAT ATGTTTATTT 180
ATGCACTGTG GGATGTGTTT TTCTTTCTCT GTATTCCGAT ACAAAGTGTT GTATCAAAGT 240
GTGATATACA AAGTGTACA ACATAAGTGT TGGTAGCACT TAAGACTTAT ACTTCCCTTC 300
TGATAGTATT CCTTTATACA CAGTGGATTG ATTATAAATA AATAGATGTG TCTTAACATA 360
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 44:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (D) OUTRA INFORMAÇÃO: / nota = "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 44:

AGGCACTGAG GTGATTGGC 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 45:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= " iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 45:

TCGTAGCCCTTAAGTGAGC 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 46:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: /nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 46:

AATATGTACA TTAGAGTAGT TG 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 47:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 47:

CAGAGAAAGG TCCTGACTC 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 48:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 22 pares de bases
- (B) TIPO: ácido nucleico

- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= " iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 48:

AGAGATTGG AAAATGAGTA AC 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 49:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 49:

ACAATGTCAT CACAGGAGG 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 50:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 50:

AACCTTCCC TTTGGTGAGG 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 51:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 51:
GATTACTCTG AGACCTAGGC 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 52:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 52:
GATTTTCTCT TTTCCCCTTG GG 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 53:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 23 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1
(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 53:
CAAACAAAGC TTCAACAATT TAC 23

(2) INFORMAÇÃO PARA A SEQ ID NO: 54:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 26 pares de bases
(B) TIPO: ácido nucleico
(C) ESTRUTURA HELICOIDAL: simples
(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature
(B) LOCALIZAÇÃO: 1
(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 54:
GGGTTTTATT TTCAAGTACT TCTATG 26

(2) INFORMAÇÃO PARA A SEQ ID NO: 55:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 26 pares de bases
(B) TIPO: ácido nucleico
(C) ESTRUTURA HELICOIDAL: simples
(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 55:

GCTCAGCAAC TGTTC AATGT ATGAGC 26

(2) INFORMAÇÃO PARA A SEQ ID NO: 56:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 56:

CTAGTGTGTG TTTTGGC 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 57:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 57:
CATAACCTTA TCTCCACC 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 58:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 23 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 58:
CTCAGCCATG AGACAATAAA TCC 23

(2) INFORMAÇÃO PARA A SEQ ID NO: 59:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 21 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 59:

GGTTCCCAA TAATGTGATG G 21

(2) INFORMAÇÃO PARA A SEQ ID NO: 60:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 60:

(2) INFORMAÇÃO PARA A SEQ ID NO: 61:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 23 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 61:

CTGTGGGTGT TTCCTGTGAG TGG 23

(2) INFORMAÇÃO PARA A SEQ ID NO: 62:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 24 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 62:

CATGACTTTG TGTGAATGTA CACC 24

(2) INFORMAÇÃO PARA A SEQ ID NO: 63:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 24 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 63:

GAGGAGAGCC TGATAGAACA TCTG 24

(2) INFORMAÇÃO PARA A SEQ ID NO: 64:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 20 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 64:

GGGCTTTTTC TCCCCCTCCC 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 65:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 18 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 65:

AAAATCTGGG CTCTCACG 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 66:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 19 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 66:
AATTATACCT CATACTAGC 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 67:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 23 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 67:
GTTTTATTAC AGAATAAAGG AGG 23

(2) INFORMAÇÃO PARA A SEQ ID NO: 68:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 68:

AAGCCAAAGT TAGAAGGCA 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 69:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 69:

TGCAACCCAC AAAATTTGGC 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 70:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 70:

CTTTCTCCA TTTCCAAAACC 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 71:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 71:
TGGTGTCTCT AGTTCTGG 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 72:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 72:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (d) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 72:
CATTGTTGTA GTAGCTCTGC 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 73:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 18 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 73:
CCCATTTGTC CCAACTGG 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 74:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 19 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 74:

CGGTCAGTTG AAATGTCAG 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 75:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 22 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 75:

CATTTGGATG CTCCGTTAAA GC 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 76:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 23 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 76:
CACCCGGCTG GAAATTTTAT TTG 23

(2) INFORMAÇÃO PARA A SEQ ID NO: 77:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 77:
GGAANGGCAC TGGAGAAATG GG 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 78:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 25 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 78:

CCCTCCAGCA CACATGCATG TACCG 25

(2) INFORMAÇÃO PARA A SEQ ID NO: 79:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 79:
TAAGTAGTCT GTGATCTCCG 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 80:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 80:
ATGTATGAGG TCCTGTCC 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 81:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 81:

GACACCAGTG TATGTTGG 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 82:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 20 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 82:

GAGAAAGAAG AACACATCCC 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 83:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 38 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 83:

TGTAAAACGA CGGCCAGTCA CTGAGGTGAT TGGCTGAA 38

(2) INFORMAÇÃO PARA A SEQ ID NO: 84:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 19 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 84:

TAGCCCTTAA GTGAGCCCG 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 85:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 38 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 85:
 TGTAAAACGA CGGCCAGTTA CATTAGAGTA GTTGCAGA 38

(2) INFORMAÇÃO PARA A SEQ ID NO: 86:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 86:
 AGGTCCTGAC TCTTCATG 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 87:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 40 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 87:

TGTAAAACGA CGGCCAGTTT GGAAAATGAG TAACATGATT 40

(2) INFORMAÇÃO PARA A SEQ ID NO: 88:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 88:

(2) INFORMAÇÃO PARA A SEQ ID NO: 89:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 38 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 89:

TGTAACACGA CGGCCAGTCT TTCCCTTTGG TGAGGTGA

38

(2) INFORMAÇÃO PARA A SEQ ID NO: 90:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 90:
TACTCTGAGA CCTAGGCCCA 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 91:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 40 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 91:
TGTAACACGA CGGCCAGTTC TCTTTTCCCC TTGGGATTAG 40

(2) INFORMAÇÃO PARA A SEQ ID NO: 92:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 23 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 92:

ACAAAGCTTC AACAATTTAC TCT 23

(2) INFORMAÇÃO PARA A SEQ ID NO: 92:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 46 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 93:

TGTAAACGA CGGCCAGTGT TTTATTTTCA AGTACTTCTA TGAATT 46

(2) INFORMAÇÃO PARA A SEQ ID NO: 94:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 26 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 94:
CAGCAACTGT TCAATGTATG AGCACT 26

(2) INFORMAÇÃO PARA A SEQ ID NO: 95:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 36 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 95:
TGTAACACGA CGGCCAGTGT GTGTGTTTTT GGCAAC 36

(2) INFORMAÇÃO PARA A SEQ ID NO: 96:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 96:

AACCTTATCT CCACCAGC 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 97:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 41 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 97:

(2) INFORMAÇÃO PARA A SEQ ID NO: 98:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 98:

TCCCAAATAA TGTGATGGAA TG 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 99:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 37 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 99:

TGTAAAACGA CGGCCAGTAA GCTTCAGAAT CTCTTTT 37

(2) INFORMAÇÃO PARA A SEQ ID NO: 100:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 23 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 100:

TGGGTGTTTC CTGTGAGTGG ATT 23

(2) INFORMAÇÃO PARA A SEQ ID NO: 101:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 42 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 101:

TGTAAAACGA CCGCCAGTAC TTTGTGTGAA TGTACACCTG TG 42

(2) INFORMAÇÃO PARA A SEQ ID NO: 102:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 24 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 102:

GAGAGCCTGA TAGAACATCT GTTG 24

(2) INFORMAÇÃO PARA A SEQ ID NO: 103:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 39 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 103:

TGTAAACGA CGGCCAGTCT TTTTCTCCCC CTCCCCACTA 39

(2) INFORMAÇÃO PARA A SEQ ID NO: 104:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 17 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 104:

TCTGGGCTCT CACGTCT 17

(2) INFORMAÇÃO PARA A SEQ ID NO: 105:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 105:

CTTATTCTGA GTCTCTCC 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 106:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 35 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 106:

(2) INFORMAÇÃO PARA A SEQ ID NO: 107:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 21 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 107:
GATGSTTCGT ACAGATTCCC G 21

(2) INFORMAÇÃO PARA A SEQ ID NO: 108:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 41 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 108:
TGTAACACGA CGGCCAGTTT ATTACAGAAT AAAGGAGGTA G 41

(2) INFORMAÇÃO PARA A SEQ ID NO: 109:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 39 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 109:
TGTAACACGA CGGCCAGTAA CCCACAAAAT TTGGCTAAG 39

(2) INFORMAÇÃO PARA A SEQ ID NO: 110:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 110:

TCTCCATTTT CAAAACCTTG 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 111:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 18 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 111:

TGTCTCTAGT TCTGGTGC 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 112:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 38 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 112:

TGTAAAACGA CGGCCAGTTG TTGTAGTAGC TCTGCTTG 38

(2) INFORMAÇÃO PARA A SEQ ID NO: 113:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 113:

ATTGTCCCA ACTGGTTGTA 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 114:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 39 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 114:

TGTAAAACGA CGGCCAGTTC AGTTGAAATG TCAGAAGTG 39

(2) INFORMAÇÃO PARA A SEQ ID NO: 115:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 115:

TGTAAAACGA CGGCCAGT 18

(2) INFORMAÇÃO PARA A SEQ ID NO: 116:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 23 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 116:

CCGGCTGGAA ATTTTATTTG GAG 23

(2) INFORMAÇÃO PARA A SEQ ID NO: 117:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 41 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 117:

TGTAAAACGA CGGCCAGTAG GCACTGGAGA AATGGGATTG 41

(2) INFORMAÇÃO PARA A SEQ ID NO: 118:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 26 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 118:

TCCAGCACAC ATGCATGTAC CGAAAT 26

(2) INFORMAÇÃO PARA A SEQ ID NO: 119:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 20 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 119:

GTAGTCTGTG ATCTCCGTTT 20

(2) INFORMAÇÃO PARA A SEQ ID NO: 120:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 36 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: misc_feature

(B) LOCALIZAÇÃO: 1

(O) OUTRA INFORMAÇÃO: / nota= "iniciadores
dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 120:

TGTAAAACGA CGGCCAGTTA TGAGGTCCTG TCCTAG 36

(2) INFORMAÇÃO PARA A SEQ ID NO: 121:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 19 pares de bases

- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 121:
 ACCAGTGTAT GTTGGGATG 19

(2) INFORMAÇÃO PARA A SEQ ID NO: 122:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 39 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: misc_feature
- (B) LOCALIZAÇÃO: 1
- (O) OUTRA INFORMAÇÃO: / nota= "iniciadores dirigidos ao ADN do intrão genómico "

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 122:
 TGTAAAACGA CGGCCAGTGA AAGAAGAACA CATCCCACA 39

(2) INFORMAÇÃO PARA A SEQ ID NO: 123:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 770 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID Nº 123:

Met	Ser	Leu	Arg	Ile	Lis	Ala	Leu	Asp	Ala	Ser	Val	Val	Asn	Lis	Ile
1				5				10						15	
Ala	Ala	Gli	Glu	Ile	Ile	Ile	Ser	Pro	Val	Asn	Ala	Leu	Lis	Glu	Met
			20					25					30		
Met	Glu	Asn	Ser	Ile	Asp	Ala	Asn	Ala	Tre	Met	Ile	Asp	Ile	Leu	Val
		35					40					45			
Lis	Glu	Gli	Gli	Ile	Lis	Val	Leu	Gln	Ile	Tre	Asp	Asn	Gli	Ser	Gli
	50					55				60					
Ile	Asn	Lis	Ala	Asp	Leu	Pro	Ile	Leu	Cis	Glu	Arg	Fen	Tre	Tre	Ser
65				70						75				80	
Lis	Leu	Gln	Lis	Fen	Glu	Asp	Leu	Ser	Gln	Ile	Gln	Tre	Tir	Gli	Fen
			85						90					95	
Arg	Gli	Glu	Ala	Leu	Ala	Ser	Ile	Ser	His	Val	Ala	Arg	Val	Tre	Val
			100						105					110	
Tre	Tre	Lis	Val	Lis	Glu	Asp	Arg	Cis	Ala	Trp	Arg	Val	Ser	Tir	Ala
		115					120					125			
Glu	Gli	Lis	Met	Leu	Glu	Ser	Pro	Lis	Pro	Val	Ala	Gli	Lis	Asp	Gli
	130					135						140			
Tre	Tre	Ile	Leu	Val	Glu	Asp	Leu	Fen	Fen	Asn	Ile	Pro	Ser	Arg	Leu
145			150							155				160	
Arg	Ala	Leu	Arg	Ser	His	Asn	Asp	Glu	Tir	Ser	Lis	Ile	Leu	Asp	Val
			165						170					175	
Val	Gli	Arg	Tir	Ala	Ile	His	Ser	Lis	Asp	Ile	Gli	Fen	Ser	Cis	Lis
		180							185					190	
Lis	Fen	Gli	Asp	Ser	Asn	Tir	Ser	Leu	Ser	Val	Lis	Pro	Ser	Tir	Tre
		195						200						205	

Val Gln Asp Arg Ile Arg Tre Val Fen Asn Lis Ser Val Ala Ser Asn	210	215	220
Leu Ile Tre Fen His Ile Ser Lis Val Gln Asp Leu Asn Leu Gln Ser	225	230	235
Val Asp Glt Lis Val Cts Asn Leu Asn Fen Ile Ser Lis Lis Ser Ile	245	250	255
Ser Leu Ile Fen Fen Ile Asn Asn Arg Leu Val Tre Cts Asp Leu Leu	260	265	270
Arg Arg Ala Leu Asp Ser Val Tir Ser Asn Tir Leu Pro Lis Glt Fen	275	280	285
Arg Pro Fen Ile Tir Leu Glt Ile Val Ile Asp Pro Ala Ala Val Asp	290	295	300
Val Asn Val His Pro Tre Lis Arg Gln Val Arg Fen Leu Ser Gln Asp	305	310	315
Gln Ile Ile Gln Lis Ile Ala Asn Gln Leu His Ala Gln Leu Ser Ala	325	330	335
Ile Asp Thr Ser Arg Tre Fen Lis Ala Ser Ser Ile Ser Tre Asn Lis	340	345	350
Pro Gln Ser Leu Ile Pro Fen Asn Asp Tre Ile Gln Ser Asp Arg Asn	355	360	365
Arg Lis Ser Leu Arg Gln Ala Gln Val Val Gln Asn Ser Tir Tre Tre	370	375	380
Ala Asn Ser Gln Leu Arg Lis Ala Lia Arg Gln Gln Asn Lis Leu Val	385	390	395
Arg Ile Asp Ala Ser Gln Ala Lis Ile Tre Ser Fen Leu Ser Ser Ser	405	410	415
Gln Gln Fen Asn Fen Gln Glt Ser Ser Tre Lis Arg Gln Leu Ser Gln	420	425	430
Pro Lis Val Tre Asn Val Ser His Ser Gln Gln Ala Gln Lis Leu Tre	435	440	445
Leu Asn Gln Ser Gln Pro Arg Asp Ala Asn Tre Ile Asn Asp Asn	450	455	460
Asp Leu Lis Asp Gln Pro Lis Lis Lis Gln Lis Gln Leu Glt Asp Tir	465	470	475
Lis Val Pro Ser Ile Ala Asp Asp Gln Lis Asn Ala Leu Pro Ile Ser	485	490	495
Lis Asp Glt Tir Ile Arg Val Pro Lis Gln Arg Val Asn Val Asn Leu			

500	505	510
Tre Ser Ile Lis Lis Leu Arg Glu Lis Val Asp Asp Ser Ile His Arg		
515	520	525
Glu Leu Tre Asp Ile Fen Ala Asn Leu Asn Tir Val Gli Val Val Asp		
530	535	540
Glu Glu Arg Arg Leu Ala Ala Ile Gln His Asp Leu Lis Leu Fen Leu		
545	550	555
Ile Asp Tir Gli Ser Val Cis Tir Glu Leu Fen Tir Gln Ile Gli Leu		
565	570	575
Tre Asp Fen Ala Asn Fen Gli Lis Ile Asn Leu Gln Ser Tre Asn Val		
580	585	590
Ser Asp Asp Ile Val Leu Tir Asn Leu Leu Ser Glu Fen Asp Glu Leu		
595	600	605
Asn Asp Asp Ala Ser Lis Glu Lis Ile Ile Ser Lis Ile Trp Asp Met		
610	615	620
Ser Ser Het Leu Asn Glu Tir Tir Ser Ile Glu Leu Val Asn Asp Gli		
625	630	635
Leu Asp Asn Asp Leu Lis Ser Val Lis Leu Lis Ser Leu Pro Leu Leu		
645	650	655
Leu Lis Gli Tir Ile Pro Ser Leu Val Lis Leu Pro Fen Fen Ile Tir		
660	665	670
Arg Leu Gli Lis Glu Val Asp Trp Glu Asp Glu Gln Glu Cis Leu Asp		
675	680	685
Gli Ile Leu Arg Glu Ile Ala Leu Leu Tir Ile Pro Asp Met Val Pro		
690	695	700
Lis Val Asp Tre Leu Asp Ala Ser Leu Ser Glu Asp Glu Lis Ala Gln		
705	710	715
Fen Ile Asn Arg Lis Glu Mis Ile Ser Ser Leu Leu Glu His Val Leu		
725	730	735
Fen Pro Cis Ile Lis Arg Arg Fen Leu Ala Pro Arg His Ile Leu Lis		
740	745	750
Asp Val Val Glu Ile Ala Asn Leu Pro Asp Leu Tir Lis Val Fen Glu		
755	760	765
Arg Cis		
770		

(2) INFORMAÇÃO PARA A SEQ ID NO: 124:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 64 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: Proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 124:

```
Val Asn Arg Ile Ala Ala Gli Glu Val Ile Gln Arg Pro Ala Asn Ala
1           5           10           15
Ile Lis Glu Met Ile Glu Asn Cys Leu Asp Ala Lis Fen Tre Ser Ile
          20           25           30
Gln Val Ile Val Lis Glu Gli Gli Leu Lis Leu Ile Gln Ile Gln Asp
          35           40           45
Asn Gli Tre Gli Ile Arg Lis Glu Asp Leu Asp Ile Val Cys Glu Arg
          50           55           60
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 125:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 64 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: Proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 125

```
Val Asn Arg Ile Ala Ala Gli Glu Val Ile Gln Arg Pro Ala Asn Ala
1           5           10           15
Ile Lis Glu Met Ile Glu Asn Cys Leu Asp Ala Lis Ser Tre Ser Ile
          20           25           30
```

Gln Val Ile Val Lis Glu Gli Gli Leu Lis Leu Ile Gln Ile Gln Asp
35 40 45
Asn Gli Tre Gli Ile Arg Lis Glu Asp Leu Asp Ile Val Cis Glu Arg
50 55 60

(2) INFORMAÇÃO PARA A SEQ ID NO: 126:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 52 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 126

Pro Ala Asn Ala Ile Lis Glu Met Ile Glu Asn cis Leu Asp Ala Lis
1 5 10 15
Ser Tre Asn Ile Gln Val Val Val Lis Glu Gli Gli Leu Lis Leu Ile
20 25 30
Gln Ile Gln Asp Asn Gli Tre Gli Ile Arg Lis Glu Asp Leu Asp Ile
35 40 45
Val Cis Glu Arg
50

(2) INFORMAÇÃO PARA A SEQ ID NO: 127:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 64 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: Proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 127

```
Val Asn Lis Ile Ala Ala Gli Glu Ile Ile Ile Ser Pro Val Asn Ala
1           5           10           15
Leu Lis Glu Met Met Glu Asn Ser Ile Asp Ala Asn Ala Tre Met Ile
           20           25           30
Asp Ile Leu Val Lis Glu Gli Gli Ile Lis Val Leu Gln Ile Tre Asp
           35           40           45
Asn Gli Ser Gli Ile Asn Lis Ala Asp Leu Pro Ile Leu Cis Glu Arg
           50           55           60
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 128:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 64 aminoácidos

(B) TIPO: aminoácido

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: Proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 128

```
Val His Arg Ile Tre Ser Gli Gln Val Ile Tre Asp Leu Tre Tre Ala
1           5           10           15
Val Lis Glu Leu Val Asp Asn Ser Ile Asp Ala Asn Ala Asn Gln Ile
           20           25           30
Glu Ile Ile Fen Lis Asp Tir Gli Leu Glu Ser Ile Glu Cis Ser Asp
           35           40           45
Asn Gli Asp Gli Ile Asp Pro Ser Asn Tir Glu Fen Leu Ala Leu Lis
           50           55           60
```

(2) INFORMAÇÃO PARA A SEQ ID NO: 129:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 64 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: Proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 129

Ala	Asn	Gln	Ile	Ala	Ala	Gli	Glu	Val	Val	Glu	Arg	Pro	Ala	Ser	Val
1				5					10					15	
Val	Lis	Glu	Leu	Val	Glu	Asn	Ser	Leu	Asp	Ala	Gli	Ala	Tre	Arg	Ile
				20					25					30	
Asp	Ile	Asp	Ile	Glu	Arg	Gli	Gli	Ala	Lis	Leu	Ile	Arg	Ile	Arg	Asp
				35				40						45	
Asn	Gli	Cis	Gli	Ile	Lis	Lis	Asp	Glu	Leu	Ala	Leu	Ala	Leu	Ala	Arg
				50			55						60		

(2) INFORMAÇÃO PARA A SEQ ID NO: 130:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 64 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 130

Ala	Asn	Gln	Ile	Ala	Ala	Gli	Glu	Val	Val	Glu	Arg	Pro	Ala	Ser	Val
1				5					10					15	
Val	Lis	Glu	Leu	Val	Glu	Asn	Ser	Leu	Asp	Ala	Gli	Ala	Tre	Arg	Val
				20					25					30	
Asp	Ile	Asp	Ile	Glu	Arg	Gli	Gli	Ala	Lis	Leu	Ile	Arg	Ile	Arg	Asp
				35				40						45	

Asn Gli Cis Gli Ile Lis Lis Glu Glu Leu Ala Leu Ala Leu Ala Arg
 50 55 60

(2) INFORMAÇÃO PARA A SEQ ID NO: 131:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 64 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: Proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 131

Ala Asn Gln Ile Ala Ala Gli Glu Val Ile Glu Arg Pro Ala Ser Val
 1 5 10 15
 Cis Lis Glu Leu Val Glu Asn Ala Ile Asp Ala Gli Ser Ser Gln Ile
 20 25 30
 Ile Ile Glu Ile Glu Glu Ala Gli Leu Lis Lis Val Gln Ile Tre Asp
 35 40 45
 Asn Gli His Gli Ile Ala His Asp Glu Val Glu Leu Ala Leu Arg Arg
 50 55 60

((2) INFORMAÇÃO PARA A SEQ ID NO: 132:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 2687 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 132

CCATGGAGCG AGCTGAGAGC TCGAGTACAG AACCTGCTAA GGCCATCAAA CCTATTGATC 60
 GGAAGTCAGT CCATCAGATT TGCTCTGGGC AGGTGGTACT GAGTCTAAGC ACTGCGGTAA 120
 AGCAGTTAGT AGAAAACAGT CTGGATCCTG CTGCCACTAA TATTGATCTA AAGCTTAAGG 180
 ACTATGGAGT GGATCTTATT GAAGTTTCAG ACAATGGATG TGGGGTAGM GAAGAAAAGT 240
 TCCAAGGCTT AACTCTCAAA CATCACACAT CTAAGATTCA AGAGTTTGCC CACCTAACTC 300
 AGGTTGAAAC TTTTGGCTTT CGGGGGGAAG CTCTGAGCTC ACTTTGTGCA CTGAGCGATG 360
 TCACCATTTC TACCTGCCAC GCATCGGCGA AGGTTGGAAC TCGACTCATG TTTGATCACA 420
 ATCGGAAAAT TATCCAGAAA ACCCCCTACC CCCGCCCCAG AGGGACCACA CTCAGCGTGC 480
 AGCAGTTATT TTCCACACTA CCTGTGCGCC ATAAGGAATT TCAAAGGAAT ATTAAGAAGG 540
 AGTATGCCAA AATGGTCCAG GTCTTACATG CATACTGTAT CATTTCAGCA GGCATCCGTG 600
 TAAGTTGCAC CAATCAGCTT GGACAAGGAA AACGACAGCC TGTGGTATGC ACAGGTGGAA 660
 GCCCAGCAT AAAGGAAAAT ATCGGCTCTG TGTGTGGGCA GAAGCAGTTG CAAAGCCTCA 720
 TTCTTTTGT TCAGCTGCCC CCTAGTGACT CCGTGTGTGA AGAGACGGT TTGAGCTGTT 780
 CGGATGCTCT CATAATCTT TTTTACATCT CAGGTTTCAT TTCACAATGC ACGCATGGAG 840
 TTAAGGAG TTCAACAGAC AGACAGTTTT TCTTTATCAA CCGGCGGCCT TGTGACCCAG 900
 CAMGGTCTG AGATCGTG AAGAGGTCT ACCACATGTA AATCGACAC CAGTATCCAT 960
 TTGTTGTTCT TAACATTCTT GTTGATTGAG AATGCGTTGA TATCAATGTT ACTCCAGATA 1020
 AAAAGCAAAT TTGCTACAA GAGGAAAAGC TTTTGTGGC AGTTTTAAAG ACCTCTTTGA 1080
 TAGGAATGTT TGATAGTGAT GTCAACAAGC TAAATGTCAG TCAGCAGCCA CTGCTGGATG 1140
 TTGAAGGTAA CTTAATAAAA ATGCATGCAG CGGATTTGGA AAAGCCCATG GTAGAAAAGC 1200
 AGGATCAATC CCCTTCATTA AGGACTGGAG AAGAAAAAAA AGACGTGTCC ATTTCCAGAC 1260
 TGCAGAGAGG CTTTTCTCTT CGTCACACAA CAGAGAACAA GCCTCACAGC CCAAAGACTC 1320
 CAGAACCAAG AAGGAGCCCT CTAGGACAGA AAAGGGGTAT GCTGTCTTCT AGCACTTCAG 1380
 GTGCCATCTC TGACAAAGGC GTCCTGAGAT CTCAGAAAGA GGCAGTGAGT TCCAGTCACG 1440
 GACCCAGTGA CCCTACGGAC AGAGCGGAGG TGGAGAAGGA CTCGGGGCAC GGCAGCACTT 1500
 CCGTGGATTC TGAGGGGTTT AGCATCCAG ACACGGGCAG TCACTGCAGC AGCGAGTATG 1560
 CGGCCAGCTC CCCAGGGGAC AGGGGCTCGC AGGAACATGT GGACTCTCAG GAGAAAGCGC 1620
 CTGAAACTGA CGACTCTTTT TCAGATGTGG ACTGCCATTC AAACCAGGAA GATACCGGAT 1680
 GTAAATTTCT AGTTTTGCCT CAGCCAACTA ATCTCGCAAC CCCAAACACA AAGCGTTTTA 1740
 AAAAAAGA AATTCTTTCC AGTTCTGACA TTTGTCAAAA GTTAGTAAAT ACTCAGGACA 1800
 TGTGAGCCTC TCAGGTGAT TGAGCTGTGA AAATTAATAA GAAAGTTGTG CCCCTGGACT 1860
 TTTCTATGAG TTCTTTAGCT AAACGAATAA AGCAGTTACA TCATGAAGCA CAGCAAAGTG 1920
 AAGGGGAACA GAATTACAGG AAGTTTAGGG CAAAGATTTG TCCTGGAGAA AATCAAGCAG 1980
 CCGAAGATGA ACTAAGAAAA GAGATAAGTA AAACGATGTT TGCAGAAATG GAAATCATTG 2040
 GTCAGTTTAA CCTGGGATTT ATAATAACCA AACTGAATGA GGATATCTTC ATAGTGGACC 2100
 AGCATGCCAC GGACGAGAAG TATAACTTCG AGATGCTGCA GCAGCACACC GTGCTCCAGG 2160
 GGCAGAGGCT CATAGCACCT CAGACTCTCA ACTTAACTGC TGTTAATGAA GCTGTTCTGA 2220
 TAGAAAATCT GGAAATATTT AGAAAGAATG CCTTTGATTT TGTTATCGAT GAAAATGCTC 2280
 CAGTCACTGA AAGGGCTAAA CTGATTTCTT TGCCAACTAG TMAAACTGG ACCTTCGGAC 2340
 CCCAGGACGT CGATGAACTG ATCTTCATGC TGAGCGACAG CCCTGGGGTC ATGTGCCGCC 2400
 CTTCCCGAGT CAAGCAGATG TTTGCCTCCA GAGCCTGCCG GAAGTCGGTG ATGATTGGGA 2460

CTGCTCTCAA CACAAGCGAA TGAAGAACT GATCACCCAC ATGGGGGAGA TGGGCCACCC 2520
 CTGGAAGTGT CCCCATGGAA GGCCACCATG AGACACATCG CCAACCTGGG TGTCAATTTCT 2580
 CAGAACTGAC CGTAGTCACT GTATGGAATA ATTGGTTTAA TCGCAGATTI TTATGTTTTG 2640
 AAACACAGAG TCTTCACTAA CCTTTTTTGT TTTAAATGA AACCTGC 2687

(2) INFORMAÇÃO PARA A SEQ ID NO: 133:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 862 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: Proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 133

Met	Glu	Arg	Ala	Glu	Ser	Ser	Ser	Tre	Glu	Pro	Ala	Lis	Ala	Ile	Lis
1				5					10					15	
Pro	Ile	Asp	Arg	Lis	Ser	Val	His	Gln	Ile	Cis	Ser	Gli	Gln	Val	Val
				20				25						30	
Leu	Ser	Leu	ser	Tre	Ala	Val	Lis	Glu	Leu	Val	Glu	Asn	Ser	Leu	Asp
				35				40						45	
Ala	Gli	Ala	Tre	Asn	Ile	Asp	Leu	Lis	Leu	Lis	Asp	Tir	Gli	Val	Asp
				50				55						60	
Leu	Ile	Glu	Val	Ser	Asp	Asn	Gli	Cis	Gli	Val	Glu	Glu	Glu	Asn	Fen
65					70					75					80
Glu	Gli	Leu	Tre	Leu	Lis	His	His	Tre	Ser	Lis	Ile	Gln	Glu	Fen	Ala
				85						90				95	
Asp	Leu	Tre	Gln	Val	Glu	Tre	Fen	Gli	Fen	Arg	Gli	Glu	Ala	Leu	Ser
				100						105				110	
Ser	Leu	Cis	Ala	Leu	Ser	Asp	Val	Tre	Ile	Ser	Tre	Cis	His	Ala	Ser
				115						120				125	
Ala	Lis	Val	Gli	Tre	Arg	Leu	Met	Fen	Asp	His	Asn	Gli	Lis	Ile	Ile
				130						135				140	

Gln Lis Tre Pro Tir Pro Arg Pro Arg Gli Tre Tre Val Ser Val Gln
 145 150 155 160
 Gln Leu Fen Ser Tre Leu Pro Val Arg His Lis Glu Fen Gln Arg Asn
 165 170 175
 Ile Lis Lis Glu Tir Ala Lis Met Val Gln Val Leu His Ala Tir Cis
 180 185 190
 Ile Ile Ser Ala Gli Ile Arg Val Ser Cis Tre Asn Gln Leu Gli Gln
 195 200 205
 Gli Lis Arg Gln Pro Val Val Cis Ile Gli Gli Ser Pro Ser Ile Lis
 210 215 220
 Glu Asn Ile Gli Ser Val Fen Gli Gln Lis Gln Leu Gln Ser Leu Ile
 225 230 235 240
 Pro Fen Val Gln Leu Pro Pro Ser Asp Ser Val Cis Glu Glu Tir Gli
 245 250 255
 Leu Ser Cis Ser Asp Ala Leu His Asn Leu Fen Tir Ile Ser Gli Fen
 260 265 270
 Ile Ser Gln Cis Tre His Gli Val Gli Arg Ser Ser Tre Asp Arg Gln
 275 280 285
 Fen Fen Fen Ile Asn Arg Arg Pro Cis Asp Pro Ala Lis Val Cis Arg
 290 295 300
 Leu Val Asn Glu Val Tir His Met Tir Asn Arg His Gln Tir Pro Fen
 305 310 315 320
 Val Val Leu Asn Ile Ser Val Asp Ser Glu Cis Val Asp Ile Asn Val
 325 330 335
 Tre Pro Asp Lis Arg Gln Ile Leu Leu Gln Glu Glu Lis Leu Leu Leu
 340 345 350
 Ala Val Leu Lis Tre Ser Leu Ile Gli Met Fen Asp Ser Asp Val Asn
 355 360 365
 Lis Leu Asn Val Ser Gln Gln Pro Leu Leu Asp Val Glu Gli Asn Leu
 370 375 380
 Ile Lis Met His Ala Ala Asp Leu Glu Lis Pro Met Val Glu His Gln
 385 390 395 400
 Asp Gln Ser Pro Ser Leu Arg Ile Gli Glu Glu Lis Lis Asp Val Ser
 405 410 415
 Ile Ser Arg Leu Arg Glu Ala Fen Ser Leu Arg His Tre Tre Glu Asn
 420 425 430
 Lis Pro His Ser Pro Lis Tre Pro Glu Pro Arg Arg Ser Pro Leu Gli

435	440	445
Gln Lis Arg Gli Met Leu Ser Ser Ser Tre Ser Gli Ala Ile Ser Asp		
450	455	460
Lis Gli Val Leu Arg Ser Gln Lis Glu Ala Val Ser Ser Ser His Gli		
465	470	475
Pro Ser Asp Pro Tre Asp Arg Ala Glu Val Glu Lis Asp Ser Gli His		
	485	490
Gli Ser Tre Ser Val Asp Ser Glu Gli Fen Ser Ile Pro Asp Tre Gli		
	500	510
Ser His Cis Ser Ser Glu Tir Ala Ala Ser Ser Pro Gli Asp Arg Gli		
	515	520
Ser Gln Glu His Val Asp Ser Gln Glu Lis Ala Pro Glu Tre Asp Asp		
530	535	540
Ser Fen Ser Asp Val Asp Cis His Ser Asn Gln Glu Asp Tre Gli Cis		
545	550	555
Lis Fen Arg Val Leu Pro Gln Pro Ile Asn Leu Ala Tre Pro Asn Tre		
	565	570
Lis Arg Fen Lis Lis Glu Glu Ile Leu Ser Ser Ser Asp Ile Cis Gln		
	580	585
Lis Leu Val Asn Tre Gln Asp Met Ser Ala Ser Gln Val Asp Val Ala		
595	600	605
Val Lis Ile Aan Lis Lis Val Val Pro Leu Asp Fen Ser Met Ser Ser		
610	615	620
Leu Ala Lis Arg Ile Lis Gln Leu His His Glu Ala Gln Gln Ser Glu		
625	630	635
Gli Glu Gln Asn Tir Arg Lis Fen Arg Ala Lis Ile Cis Pro Gli Glu		
	645	650
Asn Gln Ala Ala Glu Asp Glu Leu Arg Lis Glu Ile Ser Lis Tre Met		
	660	665
Fen Ala Glu Met Glu Ile Ile Gli Gln Fen Asn Leu Gli Fen Ile Ile		
675	680	685
Tre Lis Leu Asn Glu Asp Ile Fen Ile Val Asp Gln His Ala Tre Asp		
690	695	700
Glu Lis Tir Asn Fen Glu Met Leu Gln Gln His Tre Val Leu Gln Gli		
705	710	715
Gln Arg Leu Ile Ala Pro Gln Tre Leu Asn Leu Tre Ala Val Asn Glu		
	725	730
		735

Ala Val Leu Ile Glu Asn Leu Glu Ile Fen Arg Lis Asn Gli Fen Asp
740 745 750
Fen Val Ile Asp Glu Asn Ala Pro Val Tre Glu Arg Ala Lis Leu Ile
755 760 765
Ser Leu Pro Tre Ser Lis Asn Trp Tre Fen Gli Pro Gln Asp Val Asp
770 775 780
Glu Leu Ile Fen Met Leu Ser Asp Ser Pro Gli Val Met Cis Arg Pro
785 790 795 800
Ser Arg Val Lis Gln Met Fen Ala Ser Arg Ala Cis Arg Lis Ser Val
805 810 815
Met Ile Gli Tre Ala Leu Asn Tre Ser Glu Met Lis Lis Leu Ile Tre
820 825 830
His Met Gli Glu Met Gli His Pro Trp Asn Cis Pro His Gli Arg Pro
835 840 845
Tre Met Arg His Ile Ala Asn Leu Gli Val Ile Ser Gln Asn
850 855 860

(2) INFORMAÇÃO PARA A SEQ ID NO: 134:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 903 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: Proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 134:

Met Fen His His Ile Glu Asn Leu Leu Ile Glu Tre Glu Lis Arg Cis
1 5 10 15
Lis Gln Lis Glu Gln Arg Tir Ile Pro Val Lis Tir Leu Fen Ser Met
20 25 30
Tre Gln Ile His Gln Ile Asn Asp Ile Asp Val His Arg Ile Tre Ser
35 40 45
Gli Gln Val Ile Tre Asp Leu Tre Tre Ala Val Lis Glu Leu Val Asp

50	55	60
Asn Ser Ile Asp Ala Asn Ala Asn Gln Ile Glu Ile Ile Fen Ile Asp		
65	70	75
Tir Gli Leu Glu Ser Ile Glu Cis Ser Asp Asn Gli Asp Gli Ile Asp		80
	85	90
Pro Ser Asn Tir Glu Fen Leu Ala Leu Lis His Tir Tre Ser Lis Ile		95
	100	105
Ala Lis Fen Gln Asp Val Ala Lis Val Gln Tre Leu Gli Fen Arg Gli		110
	115	120
Glu Ala Leu Ser Ser Leu Cis Gli Ile Ala Lis Leu Ser Val Ile Tre		125
	130	135
Tre Tre Ser Pro Pro Lis Ala Asp Lis Leu Glu Tir Asp Met Val Gli		140
145	150	155
His Ile Tre Ser Lis Tre Tre Ser Arg Asn Lis Gli Tre Tre Val Leu		160
	165	170
Val Ser Gln Leu Fen His Asn Leu Pro Val Arg Gln Lis Glu Fen Ser		175
	180	185
Lis Tre Fen Lis Arg Gln Fen Tre Lis Cis Leu Tre Val Ile Gln Gli		190
	195	200
Tir Ala Ile Ile Asn Ala Ala Ile Lis Fen Ser Val Trp Asn Ile Tre		205
	210	215
Pro Lis Gli Lis Lis Asn Leu Ile Leu Ser Tre Met Arg Asn Ser Ser		220
225	230	235
Met Arg LiB Asn Ile Ser Ser Val Fen Gli Ala Gli Gli Met Fen Gli		240
	245	250
Leu Glu Glu Val Asp Leu Val Leu Asp Leu Asn Pro Fen Lis Asn Arg		255
	260	265
Met Leu Gli Lis Tir Tre Asp Asp Pro Asp Fen Leu Asp Leu Asp Tir		270
	275	280
Lis Ile Arg Val Lis Gli Tir Ile Ser Gln Asn Ser Fen Gli Cis Gli		285
	290	295
Arg Asn Ser Lis Asp Arg Gln Fen Ile Tir Val ASn Ile Arg Pro Val		300
305	310	315
Glu Tir Ser Tre Leu Leu Lis Cis Cis Asn Glu Val Tir Lis Tre Fen		320
	325	330
Asn Asn Val Gln Fen Pro Ala Val Fen Leu Asn Leu Glu Leu Pro Met		335
	340	345
		350

Ser Leu Ile Asp Val ASn Val Tre Pro Asp Lis Arg Val Ile Leu Leu
355 360 365
His Asn Glu Arg Ala Val Ile Asp Ile Fen Lis Tre Tre Leu Ser Asp
370 375 380
Tir Tir Asn Arg Gln Glu Leu Ala Leu Pro Lis Arg Met Cis Ser Gln
385 390 395 400
Ser Glu Gln Gln Ala Gln Lis Arg Leu Lis Tre Glu Val Fen Asp Asp
405 410 415
Ser Glu Ser Asn Gln Ser Asn His Ala His Fen Asn Ser Tre Tre Gli
435 440 445
Val Ile Asp Lis Ser Asn Gli Tre Glu Leu Tre Ser Val Met Asp Gli
450 455 460
Asn Tir Tre Asn Val Tre Asp Val Ile Gli Ser Glu Cis Glu Val Ser
465 470 475 480
Val Asp Ser Ser Val Val Leu Asp Glu Gli Asn Ser Ser Tre Pro Tre
485 490 495
Lis Lis Leu Pro Ser Ile Lis Tre Asp Ser Gln Asn Leu Ser Asp Leu
500 505 510
Asn Leu Asn Asn Fen Ser Asn Pro Glu Fen Gln Asn Ile Tre Ser Pro
515 520 525
Asp Lis Ala Arg Ser Leu Glu Lis Val Val Glu Glu Pro Val Tir Fen
530 535 540
Asp Ile Asp Gli Glu Lis Fen Gln Glu Lis Ala Val Leu Ser Gln Ala
545 550 555 560
Asp Gli Leu Val Fen Val Asp Asn Glu Cis His Glu His Tre Asn Asp
565 570 575
Cis Cis His Gln Glu Arg Arg Gli Ser Tre Asp Ile Glu Gln Asp Asp
580 585 590
Glu Ala Asp Ser Ile Tir Ala Glu Ile Glu Pro Val Glu Ile Asn Val
595 600 605
Arg Tre Pro Leu Lis Asn Ser Arg Lis Ser Ile Ser Lis Asp Asn Tir
610 615 620
Arg Ser Leu Ser Asp Gli Leu Tre His Arg Lis Fen Glu Asp Glu Ile
625 630 635 640
Leu Glu Tir Asn Leu Ser Tre Lis Asn Fen Lis Glu Ile Ser Lis Asn
645 650 655
Gli Lis Gln Met Ser Ser Ile Ile Ser Lis Arg Lis Ser Glu Ala Gln

660	665	670
Glu Asn Ile Ile Lis Asn Lis Asp Glu Leu Glu Asp Fen Glu Gln Gli		
675	680	685
Glu Lis Tir Leu Tre Leu Tre Val Ser Lis Asn Asp Fen Lis Lis Met		
690	695	700
Glu Val Val Gli Gln Fen Asn Leu Gli Fen Ile Ile Val Tre Arg Lis		
705	710	715
Val Asp Asn Lis Ser Lis Leu Fen Ile Val Asp Gln His Ala Ser Asp		
725	730	735
Glu Lis Tir Asn Fen Glu Tre Leu Gln Ala Val Tre Val Fen Lis Ser		
740	745	750
Gln Lis Leu Ile Ile Pro Gln Pro Val Glu Leu Ser Val Ile Asp Glu		
755	760	765
Leu Val Val Leu Asp Asn Leu Pro Val Fen Glu Lis Asn Gli Fen Lis		
770	775	780
Leu Lis Ile Asp Glu Glu Glu Glu Fen Gli Ser Arg Val Lis Leu Leu		
785	790	795
Ser Leu Pro Tre Ser Lis Gln Tre Leu Fen Asp Leu Gli Asp Fen Asn		
805	810	815
Glu Leu Ile His Leu Ile Lis Glu Asp Gli Gli Leu Arg Arg Asp Asn		
820	825	830
Ile Arg Cis Ser Lis Ile Arg Ser Met Fen Ala Met Arg Ala Cis Arg		
835	840	845
Ser Ser Ile Met Ile Gli Lis Pro Leu Asn Lis Lis Tre Met Tre Arg		
850	855	860
Val Val His Asn Leu Ser Glu Leu Asp Ile Pro Trp Asn Cis Pro His		
865	870	875
Gli Arg Pro Tre Met Arg His Leu Met Glu Ile Arg Asp Trp Ser Ser		
885	890	895
Fen Ser Lis Asp Tir Glu Ile		
900		

(2) INFORMAÇÃO PARA A SEQ ID NO: 135:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 2577 pares de bases

- (B) TIPO: ácido nucleico
(C) ESTRUTURA HELICOIDAL: simples
(D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 135:

```

TTCCGGCCAA TGCTATCAAA GAGATGATAG AAAACTGTTT AGATGCAAA TCTACAAATA    60
TTCAAGTGGT TGTTAAGGAA GGTGGCCTGA AGCTAATTCA GATCCAAGAC AATGGCACTG   120
GAATCAGGAA GGAAGATCTG GATATTGTGT GTGAGAGGTT CACTACGAGT AACTGCAGA   180
CTTTTGAGGA TTTAGCCAGT ATTTCTACCT ATGGCTTTTCG TGGTGAGCAT TTGGCAAGCA   240
TAAGTCATGT GGCCCATGTC ACTATTACAA CCAAAACAGC TGATGGGAAA TGTGCGTACA   300
GAGCAAGTTA CTCAGATGGA AAGCTGCAAG CCCCTCCTAA ACCCTGTGCA GGCAACCAGG   360
GCACCCTGAT CACGGTGGAA GACCTTTTTT ACAACATAAT CACAAGGAGG AAAGCTTTAA   420
AAAATCCAAG TGAAGAGTAC GGAAAAATTT TGGAAGTTGT TGGCAGGTAT TCAATACACA   480
ATTCAAGCAT TAGTATCTCA GTTAAAAAAC AAGGTGAGAC AGTATCTGAT GTCAGAACAC   540
TGCCCAATGC CACAACCGTG GACAACATTC GCTCCATCTT TGGAAATGCG GTTAGTCGAG   600
AACTGATAGA AGTTGGGTGT GAGGATAAAA CCCTAGCTTT CAAAATGAAT GGCTATATAT   660
CGAATGCAAA GTATTCAGTG AAGAAGTGCA TTTTCCTACT CTTTCATCAAC CACCGTCTGG   720
TAGAATCAGC TGCCTTGAGA AAAGCCATTG AAAGTGTATA TGCAGCATAC TTGCCAAAAA   780
CACACACCCA TTCCTGTACC TCAGTTTGAA ATCAGCCCTC AGAACGTGAC GTCAATGTAC   840
ACCCACCAA GACAGAAGTT CATTTTCTGC ACGAGGAGAG CATTCCTGCAG CGTGTGCAGC   900
AGCACATTGA GAGCAAGCTG CTGGGCTCCA ATTCCTCCAG GATGTATTTT ACCCAGACCT   960
TGCTTCCAGG ACTTGCTGGG CCTCTGGGGA GGCAGCTAGA CCCACGACAG GGGTGGCTTC  1020
CTCATCCACT AGTGGAAGTG GCGACAAGGT CTACGCTTAC CAGATGTCGC GTACGGACTC  1080
CCGGGATCAG AAGCTTGACG CCTTTCTGCA GCCTGTAACC AGCCTTGTGC CCAGCCAGCC  1140
CCAGGACCCT CGCCCTGTCC GAGGGGCCAG GACAGAGGGC TCTCCTGAAA GGGCCACGCG  1200
GGAGGATGAG GAGATGCTTG CTCTCCAGC CCCCGCTGAA GCAGCTGCTG AGAGTGAGAA  1260
CTTGGAGAGG GAATCACTAA TGGAGACTTC AGACGCAGCC CAGAAAGCGG CACCCACTTC  1320
CAGTCCAGGA AGCTCCAGAA AGAGTCATCG GGAGGACTCT GATGTGGAAA TGGTGGAAAA  1380
TGCTTCCGGG AAGGAAATGA CAGCTGCTTG CTACCCAGG AGGAGGATCA TTAACCTCAC  1440
CAGCGTCTTG AGTCTCCAGG AAGAGATTAG TGAGCGGTGC CATGAGACTC TCCGGGAGAT  1500
ACTCCGTAAC CATTCCTTTG TGGGCTGTGT GAATCCTCAG TGGGCCTTGG CACAGCACCA  1560
GACCAAGCTA TACCTCCTCA AACTACCAA GCTCAGTGAA GAGCTGTTCT ACCAGATACT  1620
CATTTATGAT TTTGCCAACT TTGGTGTCTT GAGGTTATCG GAACCAGCGC CACTCTTCGA  1680
CCTGGCCATG CTGGCTTAGA CAGTCTGAA AGTGCTGGA CAGAGGACGA CGGCCCCAAG  1740
AAGGGCTTGC AGAGTACATT GTCGAGTTT TGAAGAGAAG CGAGATGCTT CCAGACTATT  1800
CTCTGTGAGA TCGATGAGAA GGGAACCTGA TTGATTACTC TTCTGATGAC AGCTATGTGC  1860
CACCTTTGGA GGGACTGCCT ATCTTCATTC TTCGACTGGC CACTGAGCTG AATTGGGTGA  1920
AGAAAAGGAG TGTTTTGAAA GTCTCAGTAA AGAATGTGCT ATGTTTTACT CCATTCGGAA  1980
GCAGTATATA CTGGAGGAGT CGACCTCTC AGGCCAGCAG AGTGACATGC CTGGCTCCAC  2040

```

GTCAAAGCCC TGGAAGTGA CTGTGGAGCA CATTATCTAT AAAGCCTTCC GCTCACACCT 2100
 CCTACCTCCG AAGCATTTC AAGAAGATGG CAATGTCCTG CAGCTTGCCA ACCTGCCAGA 2160
 TCTATACAAA GTCTTTGAGC GGTGTTAAAT ACAATCATAG CCACCGTAGA GACTGCATGA 2220
 CCATCCAAGG CGAAGTGTAT GGTACTAATC TGGAAGCCAC AGAATAGGAC ACTTGGTTTC 2280
 AGCTCCAGGG TTTTCAGTCC TCACTATTCT TGTTCTGTAT CCCAGTATTG GTGCTGCAAC 2340
 TTAATGTACT TCACCTGTGG ATTGGCTGCA AATAAACTCA CGTGTATTGG AAAAAOOAA 2400
 TTCCTGCAGC CCGGGGGATC CACTAGTTCT AGAGCGGCCG CCACCGGTGG AGCTCCAGCT 2460
 TTTGTTCCCT TTAGTGAGGG TTAATTTCA GCTTGGCGTA ATCATGGTCA TAGCTGTTTC 2520
 CTGTGTGAAA TTGTTATCCG CTCACAATC CACACAACAT ACGAGCCGGA AGCATAA 2577

(2) INFORMAÇÃO PARA A SEQ ID NO: 136:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 728 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: Proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 136:

Pro	Ala	Asn	Ala	Ile	Lis	Glu	Met	Ile	Glu	Asn	Cis	Leu	Asp	Ala	Lis
1				5				10				15			
Ser	Tre	Asn	Ile	Gln	Val	Val	Val	Lis	Glu	Gli	Gli	Leu	Lis	Leu	Ile
				20				25				30			
Gln	Ile	Gln	Asp	Asn	Gli	Tre	Gli	Ile	Arg	Lis	Glu	Asp	Leu	Asp	Ile
				35				40				45			
Val	Cis	Glu	Arg	Fen	Tre	Tre	Ser	Lis	Leu	Gln	Tre	Fen	Glu	Asp	Leu
				50				55				60			
Ala	Ser	Ile	Ser	Tre	Tir	Gli	Fen	Arg	Gli	Glu	His	Leu	Ala	Ser	Ile
65				70				75				80			
Ser	His	Val	Ala	His	Val	Tre	Ile	Tre	Tre	Ile	Tre	Ala	Asp	Gli	Lis
				85				90				95			
Cis	Ala	Tir	Arg	Ala	Ser	Tir	Ser	Asp	Gli	Lis	Leu	Gln	Ala	Pro	Pro
				100				105				110			
Lis	Pro	Cis	Ala	Gli	Asn	Gln	Gli	Tre	Leu	Ile	Tre	Val	Glu	Asp	Leu

115	120	125
Fen Tir Asn Ile Ile Tre Arg Arg Lis Ala Leu Lis Asn Pro Ser Glu		
130	135	140
Glu Tir Gli Lis Ile Leu Glu Val Val Gli Arg Tir Ser Ile His Asn		
145	150	155
Ser Gli Ile Ser Ile Ser Val Ile Lis Gln Gli Glu Tre Val Ser Asp		
165	170	175
Val Arg Tre Leu Pro Aen Ala Tre Tre Val Asp Asn Ile Arg Ser Ile		
180	185	190
Fen Gli Asn Ala Val Ser Arg Glu Leu Ile Glu Val Gli Cis Glu Asp		
195	200	205
Lis Tre Leu Ala Fen Lis Met Asn Gli Tir Ile Ser Asn Ala Lis Tir		
210	215	220
Ser Val Lis Lis Cis Ile Fen Leu Leu Fen Ile Asn His Arg Leu Val		
225	230	235
Glu Ser Ala Ala Leu Arg Lis Ala Ile Glu Tre Val Tir Ala Ala Tir		
245	250	255
Leu Pro Lis Tre His Tre His Ser Cis Tre Ser Val Glx Asn Gln Pro		
260	265	270
Ser Glu Arg Asp Val Asn Val His Pro Tre Lis Tre Glu Val His Fen		
275	280	285
Leu His Glu Glu Ser Ile Leu Gln Arg Val Gln Gln His Ile Glu Ser		
290	295	300
Lis Leu Leu Gli Ser Asn Ser Ser Arg Met Val Fen His Pro Asp Leu		
305	310	315
Ala Ser Arg Tre Cis Trp Ala Ser Gli Glu Ala Ala Arg Pro Tre Tre		
325	330	335
Gli Val Ala Ser Ser Ser Tre Ser Gli Ser Gli Asp Lis Val Tir Ala		
340	345	350
Tir Gln Met Ser Arg Tre Asp Ser Arg Asp Gln Lis Leu Asp Ala Fen		
355	360	365
Leu Gln Pro Val Ser Ser Leu Val Pro Ser Gln Pro Gln Asp Pro Arg		
370	375	380
Pro Val Arg Gli Ala Arg Tre Glu Gli Ser Pro Glu Arg Ala Tre Arg		
385	390	395
Glu Asp Glu Glu Met Leu Ala Leu Pro Ala Pro Ala Glu Ala Ala Ala		
405	410	415

Glu Ser Glu Asn Leu ala Arg Glu Ser Leu Met Glu Tre Ser Asp Ala
 420 425 430
 Ala Gln Lis Ala Ala Pro Tre Ser Ser Pro Gli Ser Ser Arg Lis Ser
 435 440 445
 His Arg Glu Asp Ser Asp Val Glu Met Val Glu Asn Ala Ser Gli Lis
 450 455 460
 Glu Met Tre Ala Ala Cis Tir Pro Arg Arg Arg Ile Ile Asn Leu Tre
 465 470 475 480
 Ser Val Leu Ser Leu Gln Glu Glu Ile Ser Glu Arg Cis His Glu Tre
 485 490 495
 Leu Arg Glu Ile Leu Arg Asn His Ser Fen Val Gli Cis Val Asn Pro
 500 505 510
 Gln Trp Ala Leu Ala Gln His Gln Tre Lis Leu Tir Leu Leu Asn Tre
 515 520 525
 Tre Lis Leu Ser Glu Glu Leu Fen Tir Gln Ile Leu Ile Tir Asp Fen
 530 535 540
 Ala Asn Fen Gli Val Leu Arg Leu Ser Glu Pro Ala Pro Leu Fen Asp
 545 550 555 560
 Leu Ala Met Leu Ala Cis Tre Val Leu Lis Val Ala Gli Gln Arg Tre
 565 570 575
 Tre Ala Arg Arg Arg Ala Cis Arg Val His Cis Arg Val Ser Glu Glu
 580 585 590
 Lis Arg Asp Ala Cis Arg Leu Fen Ser Val Arg Ser Met Arg Arg Glu
 595 600 605
 Pro Asp Glx Leu Leu Fen Glx Glx Gln Leu Cis Ala Met Fen Gli Gli
 610 615 620
 Tre Ala Tir Leu His Ser Ser Tre Gli His Glx Gli Glu Leu Gli Glu
 625 630 635 640
 Glu Lis Glu Cis Fen Glu Ser Leu Ser Lis Glu Cis Ala Met Fen Tir
 645 650 655
 Ser Ile Arg Lis Gln Tir Ile Leu Glu Glu Ser Tre Leu Ser Gli Gln
 660 665 670
 Gln Ser Asp Met Pro Gli Ser Tre Ser Lis Pro Trp Lis Trp Tre Val
 675 680 685
 Glu His Ile Ile Tir Lis Ala Fen Arg Ser His Leu Leu Pro Pro Lis
 690 695 700
 His Fen Tre Glu Asp Gli Asn Val Leu Gln Leu Ala Asn Leu Pro Asp

705	710	715	720
Leu	Tir	Lis	Val
Fen	Glu	Arg	Cis
725			

(2) INFORMAÇÃO PARA A SEQ ID NO: 137:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 3065 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 137:

```

CGGTGAAGGT CCTGAAGAAT TTCCAGATTC CTGAGTATCA TTGGAGGAGA CAGATAACCT 60
GTCGTCAGGT AACGATGGTG TATATGCAAC AGAAATGGGT GTTCCTGGAG ACGCGTCTTT 120
TCCCGAGAGC GGCACCGCAA CTCTCCCGCG GTGACTGTGA CTGGAGGACT CCTGCATCCA 180
TGCAGCAAAC CGAAGGCGTG AGTACAGAAT GTGCTAAGGC CATCAAGCCT ATTGATGGGA 240
AGTCAGTCCA TCAAATTTGT TCTGGGCAGG TGATACTCAG TTTAAGCACC GCTGTGAAGG 300
AGTTGATAGA AAATAGTGTA GATGCTGGTG CTACTIONAT TGATCTAAGG CTAAAGACT 360
ATGGGGTGGA CCTCATTGAA GTTTCAGACA ATGGATGTGG GGTAGAAGAA GAAAACCTTG 420
AAGGTCTAGC TCTGAAACAT CACACATCTA AGATTCAAGA GTTTGCCGAC CTCACGCAGG 480
TTGAAACTTT CGGCTTTTCG GGGGAAGCTC TGAGCTCTCT GTGTGCACTA AGTGATGTCA 540
CTATATCTAC CTGCCACGGG TCTGCAAGCG TTGGGACTCG ACTGGTGTTC GACCATAATG 600
GGAAATCAC CCAGAAACT CCCTACCCCC GACCTAAAGG AACCACAGTC AGTCTGCAGC 660
ACTTATTTTA TACTACTACC GTGCGTTACA AAGAGTTTCA GAGGAACATT AAAAAGGAGT 720
ATTCCAAAAT GGTGCAGGTC TTACAGGCCT ACTGTATCAT CTCAGCAGGC CTCCGTGTAA 780
GCTGCACTAA TCACCTCGGA CAGGGGAACC GGCACGCTGT GCTGTCCACA AGCGGCACGT 840
CTGGCATGAA GGAAATATC GGGTCTGTGT TTGGCCAGAA GCAGTTCCAA AGCCTCATTC 900
CTTTTGTTCA GCTGCCCCCT AGTGACGCTG TGTGTGAAGA GTACOOCTG AGCACTTCAG 960
GACGCCACAA AACCTTTTCT ACGTTTTTCG GCTTCATTTT ACAGTGCACG CACGGCGCCG 1020
GGAGGAGTGC AACAGACAGG CAGTTTTTCT TCATCAATCA GAGGCCCTGT GACCCAGCAA 1080
AGGTCTCTAA GCTTGTCATG GAGGTTTATC ACATGTATAA CCGGCATCAG TACCCATTTG 1140
TCGTCTTAA CGTTTCCGTT GACTCAGAAAT GTGTGGATAT TAATGTAACCT CCAGATAAAA 1200
GGCAAATTCT ACTACAAGAA GAGAAGCTAT TGCTGGCCGT TTTAAAGACC TCCTTGATAG 1260
GAATGTTTGA CAGTGATGCA AACAAGCTTA ATGTCAACCA GCAGCCACTG CTAGATGTTG 1320
AAGGTAACCT AGTAAAGTCG CATACTGCAG AACTAGAAAA GCCTGTGCCA GGAAAGCAAG 1380

```

ATAACCTCTCC TTCACTGAAG AGCACAGCAG ACGAGAAAAG GGTAGCATCC ATCTCCAGGC 1440
 TGAGAGAGGC CTTTTCTCTT CATCCTACTA AAGAGATCAA GTCTAGOOGT CCAGAGACTG 1500
 CTGAACTGAC ACGGAGTTTT CCAAGTGAGA AAAGGGGCGT GTTATCCTCT TATCCTTCAG 1560
 ACGTCATCTC TTACAGAGGC CTCCGTGGCT CGCAGGACAA ATTGGTGAGT CCCACGGACA 1620
 GCCCTGGTGA CTGTATGGAC GAGAGAAAA TAGAAAAAGA CTCAGGGCTC AGCAGCACCT 1680
 CAGCTGGCTC TGAGGAAGAG TTCAGCACCC CAGAAGTGGC CAGTAGCTTT AGCAGTGA CT 1740
 ATAACGTGAG CTCCCTAGAA GACAGACCTT CTCAGGAAAC CATAAACTGT GGTGACCTGC 1800
 TGCCGTCCTC CAGGTACAGG ACAGTCCTTG AAGCCAGAAG ACCATGGATA TCAATGCAAA 1860
 GCTCTACCTC TAGCTCGTCT GTCACCCACA AATGCCAAGC GCTTCAAGAC AGAGGAAGAC 1920
 CCTCAAATGT CAACATATCT CAAAGATTGC CTGGTCCTCA GAGCACCTCA GCAGCTGAGG 1980
 TCGATGTAGC CATAAAAATG AATAAGAGAT CGTGCTCCTC GAGTTCTCTA GCTAAGCGAA 2040
 TGAAGCAGTT ACAGCACCTA AAGGCGCAGA ACAAACATGA ACTGAGTTAC AGAAAATTTA 2100
 GGGCCAAGAT TTGCCCTGGA GAAAACCAAG CAGCAGAAGA TGAAGTCAGA AAAGAGATTA 2160
 GTAAATCGAT GTTTCAGAG ATGGAGATCT TGGGTCAGTT TAACCTGGGA TTTATAGTAA 2220
 CCAAACCTGAA AGAGGACCTC TTCCTGGTGG ACCAGCATGC TCGGATGAG AAGTACAACT 2280
 TTGAGATGCT GCAGCAGCAC ACGGTGCTCC AGGCGCAGAG GCTCATCACG TGGGTGCACA 2340
 CAGGCTTCAG AGTTCCCAGA CCCAGACTC TGAACCTAAC TGCTGTCAAT GAAGCTGTAC 2400
 TGATAGAAAA TCTGGAAATA TTCAGAAAGA ATGGCTTTGA CTTTGTCAAT GATGAGGATG 2460
 CTCCAGTCAC TGAAAGGGCT AAATTGATTT CTTACCAAC TAGTAAAAAC TGGACCTTTG 2520
 GACCCCAAGA TATAGATGAA CTGATCTTTA TGTTAAGTGA CAGCCCTGGG GTCATGTGCC 2580
 GGCCCTCACG AGTCAGACAG ATGTTTGCTT CCAGAGCCTG TCGGAAGTCA GTGATGATTG 2640
 GAACGGCGCT CAATGCGAGC GAGATGAAGA AGCTCATCAC CCACATOOGT GAGATGGACC 2700
 ACCCCTGGAA CTGCCCCCAC GGCAGGCCAA CCATGAGGCA CGTTGCCAAT CTGGATGTCA 2760
 TCTCTCAGAA CTGACACACC CCTTGTAGCA TAGAGTTTAT TACAGATTGT TCGGTTCCGA 2820
 AAGAGAAGGT TTTAAGTAAT CTGATTATCG TTGTACAAA ATTAGCATGC TGCTTTAATG 2880
 TACTGGATCC ATTTAAAAGC AGTGTTAAGG CAGGCATGAT GGAGTGTTCC TCTAGCTCAG 2940
 CTACTTGGGT GATCCGGTGG GAGCTCATGT GAGCCAGGA CTTTGAGACC ACTCCGAGCC 3000
 ACATTCATGA GACTCAATTC AAGGACAAAA AAAAAAGAT ATTTTGAAG CCTTTTAAAA 3060
 AAAAA 3065

(2) INFORMAÇÃO PARA A SEQ ID NO: 138:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 864 aminoácidos
- (B) TIPO: aminoácido
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 138:

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

178

				565					570					575			
Asp	Arg	Gli	Arg	Pro	Ser	Asn	Val	Asn	Ile	Ser	Gln	Arg	Leu	Pro	Gli		
				580					585					590			
Pro	Gln	Ser	Tre	Ser	Ala	Ala	Glu	Val	Asp	Val	Ala	Ile	Lis	Met	Asn		
				595					600					605			
Lis	Arg	Ser	Cis	Ser	Ser	Ser	Ser	Leu	Ala	Lis	Arg	Met	Lis	Gln	Leu		
				610					615					620			
Gln	His	Leu	Lis	Ala	Gln	Asn	Lis	His	Glu	Leu	Ser	Tir	Arg	Lis	Fen		
				625					630					635			640
Arg	Ala	Lis	Ile	Cis	Pro	Gli	Glu	Asn	Gln	Ala	Aia	Glu	Asp	Glu	Leu		
				645					650					655			
Arg	Lis	Glu	Ile	Ser	Lis	Ser	Met	Fen	Ala	Glu	Met	Glu	Ile	Leu	Gli		
				660					665					670			
Gln	Fen	Asn	Leu	Gli	Fen	Ile	Val	Tre	Lis	Leu	Lis	Glu	Asp	Leu	Fen		
				675					680					685			
Leu	Val	Asp	Gln	His	Ala	Ala	Asp	Glu	Lis	Tir	Asn	Fen	Glu	Met	Leu		
				690					695					700			
Gln	Gln	His	Tre	Val	Leu	Gln	Ala	Gln	Arg	Leu	Ile	Tre	Trp	Val	His		
				705					710					715			720
Tre	Gli	Fen	Arg	Val	Pro	Arg	Pro	Gln	Tre	Leu	Asn	Leu	Tre	Ala	Val		
				725					730					735			
Asn	Glu	Ala	Val	Leu	Ile	Glu	Asn	Leu	Glu	Ile	Fen	Arg	Lis	Asn	Gli		
				740					745					750			
Fen	Asp	Fen	Val	Ile	Asp	Glu	Asp	Ala	Pro	Val	Tre	Glu	Arg	Ala	Lis		
				775					760					765			
Leu	Ile	Ser	Leu	Pro	Tre	Ser	Lis	Asn	Trp	Tre	Fen	Gli	Pro	Gln	Asp		
				770					775					780			
Ile	Asp	Glu	Leu	Ile	Fen	Met	Leu	Ser	Asp	Ser	Pro	Gli	Val	Met	Cis		
				785					790					795			800
Arg	Pro	Ser	Arg	Val	Arg	Gln	Met	Fen	Ala	Ser	Arg	Ala	Cis	Arg	Lis		
				805					810					815			
Ser	Val	Met	Ile	Gli	Tre	Ala	Leu	Asn	Ala	Ser	Glu	Met	Lis	Lis	Leu		
				820					825					830			
Ile	Tre	His	Met	Gli	Glu	Met	Asp	His	Pro	Trp	Asn	Cis	Pro	His	Gli		
				835					840					845			
Arg	Pro	Tre	Met	Arg	His	Val	Ala	Asn	Leu	Asp	Val	Ile	Ser	Gln	Asn		
				850					855					860			

(2) INFORMAÇÃO PARA A SEQ ID NO: 139:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 29 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 139:

CTTGATTCTA GAGCITCNCC NCKRAANCC 29

(2) INFORMAÇÃO PARA A SEQ ID NO: 140:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 29 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 140:

AGGTCGGAGC TCAARGARIT NGTINGANAA 29

(2) INFORMAÇÃO PARA A SEQ ID NO: 141:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 15 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 141:

ACTTGTGGA TTTTGC 15

(2) INFORMAÇÃO PARA A SEQ ID NO: 142:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 15 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 142:

ACTTGTGAAT TTTGC 15

(2) INFORMAÇÃO PARA A SEQ ID NO: 143:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 143:

TTCGGTGACA GATTTGTAAA TG 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 144:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 16 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(x1) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 144:
TTTACGGAGC CCTGGC 16

(2) INFORMAÇÃO PARA A SEQ ID NO: 145:

(1) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(x1) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 145:
TCACCATAA AATAGTTCC CG 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 146:

(1) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) ESTRUTURA HELICOIDAL: simples
- (D) TOPOLOGIA: linear

(x1) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 146:
TCCTGGATCA TATTTCTGA GC 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 147:

(1) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 22 pares de bases
- (B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 147:

TTTCAGGTAT GTCCTGTTAC CC 22

(2) INFORMAÇÃO PARA A SEQ ID NO: 148:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 22 pares de bases

(B) TIPO: ácido nucleico

(C) ESTRUTURA HELICOIDAL: simples

(D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 148:

TGAGGCAGCT TTAAGAAAC TC 22

Lisboa, 13 de Setembro de 2006

REIVINDICAÇÕES

1. Processo de diagnóstico da propensão para cancro do cólon de um indivíduo, caracterizado pelo facto de compreender:

a detecção da presença de uma mutação numa sequência de ácido nucleico de hMLH1, que se pode obter por:

(i) preparação de um primeiro iniciador degenerado correspondendo à sequência de aminoácidos GFRGEA e um segundo iniciador degenerado comportando a sequência da SEQ ID n° 140

(ii) realização de uma RCP (reação em cadeia de polimerase) utilizando os referidos iniciadores no ADNc obtido a partir de culturas de fibroblastos primários humanos para gerar um produto de RCP e;

(iii) isolamento da sequência completa de hMLH1 de uma biblioteca de ADNc humano utilizando o referido produto como sonda,

em que a referida mutação se detecta por comparação de uma sequência de hMLH1 isolada de um tecido desse indivíduo com uma sequência de hMLH1 de tipo selvagem, sendo a presença da mutação indicativa da susceptibilidade do indivíduo para cancro do cólon e em que a mutação é caracterizada por uma mutação da transição de C para T que produz uma substituição não conservadora de aminoácido na posição 44 da proteína hMLH1 (SEQ ID N° 5).

2. Processo de acordo com a reivindicação 1, caracterizado pelo facto de a sequência de ácidos nucleicos de hMLH1 ter a sequência da SEQ ID N° 4.

3. Processo de diagnóstico de propensão para o cancro do cólon de um indivíduo, caracterizado pelo facto de compreender:

a detecção da presença de uma mutação numa sequência de ácidos nucleicos de hPMS1 que comporta a sequência da SEQ ID N° 132,

em que a referida mutação é detectada comparando uma sequência de ácidos nucleicos de hPMS1 isolada de um tecido do indivíduo com uma sequência de hPMS1 de tipo selvagem, respectivamente, sendo a presença de uma mutação indicativa de que o indivíduo tem propensão para o cancro do cólon.

4. Processo de diagnóstico da propensão para o cancro do cólon de um indivíduo, caracterizado pelo facto de compreender:

a detecção da ligação de um anticorpo que se liga especificamente a um polipéptido de hMLH1 comportando a sequência da SEQ ID N° 5 ou um polipéptido de hPMS1 comportando a sequência da SEQ ID N° 133, com uma amostra obtida do indivíduo,

sendo a quantidade de ligação do referido anticorpo indicativa de que o indivíduo tem propensão para o cancro do cólon.

5. Processo de acordo com a reivindicação 1 ou com a reivindicação 2, caracterizado pelo facto de a etapa de detecção compreender:

a amplificação de um segmento da sequência de hMLH1 isolada;

a comparação do segmento amplificado com um segmento análogo da sequência de hMLH1 de tipo selvagem; e

a detecção de uma diferença entre o segmento amplificado e o segmento análogo, sendo a diferença indicativa da mutação da transição de C para T na sequência de hMLH1.

6. Processo de acordo com a reivindicação 3 caracterizado pelo facto de a etapa de detecção compreender:

a amplificação de um segmento da sequência de hPMS1 isolado;

a comparação do segmento amplificado com um segmento análogo da sequência de hPMS1 de tipo selvagem; e

a detecção de uma diferença entre o segmento amplificado e o segmento análogo, sendo a diferença indicativa de uma mutação na sequência de hPMS1.

7. Processo de acordo com a reivindicação 6, caracterizado pelo facto de a diferença na sequência de nucleótidos ser seleccionada no grupo que consiste em eliminações de pelo menos um dos nucleótidos, inserções de pelo menos um dos nucleótidos, substituições de pelo menos um dos nucleótidos e re-arranjos dos nucleótidos.

8. Processo de acordo com a reivindicação 5 ou com a reivindicação 6, caracterizado pelo facto de a etapa de amplificação compreender:

a transcrição reversa de todo ou de uma parte do produto do gene de ARN para o ADN; e
a amplificação de um segmento do ADN produzido por transcrição reversa.

9. Processo de acordo com a reivindicação 8, caracterizado pelo facto de a etapa de amplificação compreender:

a selecção de um par de iniciadores de oligonucleótido capazes de hibridarem com as estruturas helicoidais opostas do segmento de ADN e na orientação oposta; e
a realização de uma reacção em cadeia de polimerase utilizando os iniciadores de oligonucleótido de tal modo que o ácido nucleico que intervém entre os iniciadores é amplificado para se tornar num segmento amplificado.

10. Processo de acordo com uma qualquer das reivindicações precedentes caracterizado pelo facto de a etapa de detecção de uma mutação compreender a determinação se a diferença entre o segmento amplificado e o segmento análogo causa um fenótipo afectado.

11. Polinucleótido isolado caracterizado pelo facto de codificar a proteína de reparação do desemparelhamento de ADN e compreender a sequência da SEQ ID N° 132.

12. Polipéptido de reparação do desemparelhamento de ADN isolado, caracterizado pelo facto de ser codificado pelo polinucleótido isolado de acordo com a reivindicação 11.
13. Polipéptido de reparação do desemparelhamento de ADN isolado, caracterizado pelo facto de comportar a sequência da SEQ ID N° 133.
14. Anticorpo purificado caracterizado pelo facto de se ligar especificamente a um polipéptido de acordo com a reivindicação 12 ou com a reivindicação 13.
15. Anticorpo, de acordo com a reivindicação 14, caracterizado pelo facto de o anticorpo ser um anticorpo monoclonal.
16. Sonda caracterizada pelo facto de compreender:

uma sequência de nucleótidos com menos de 500 pb capazes de se ligarem especificamente a bases complementares da SEQ ID N° 4, por meio do emparelhamento de Watson/Crick, com uma T_m maior do que 55 °C; e

uma parte marcada ligada à sequência.

em que a parte marcada tem uma propriedade seleccionada no grupo que consiste em fluorescência, radioactividade e quimio-luminescência.

17. Sonda caracterizada pelo facto de compreender:

uma sequência de nucleótidos capazes de se ligarem especificamente a bases complementares da

SEQ ID N° 132, por meio do emparelhamento de Watson/Crick, com uma Tm maior do que 55 °C; e

uma parte marcada ligada à sequência, em que a parte marcada tem uma propriedade seleccionada no grupo que consiste em fluorescência, radioactividade e quimio-luminescência.

Lisboa, 13 de Setembro de 2006

processo para avaliar outros genes neste processo de
reparação de erros de emparelhamento.

1/24

Guia para o isolamento e caracterização dos genes de *PMS1* e *MLH1* de mamíferos

- Etapa 1 Preparação de conjuntos de oligonucleótidos degenerados para RCP.
- Etapa 2 Transcrição reversa e RECP em poli A + ARNm selecionado, isolados de células humanas
- Etapa 3 Clonagem e sequenciação de fragmentos gerados na RCP; identificação de dois fragmentos de genes que representam *PMS1* e *MLH1* humanos.
- Etapa 4 Isolamento de clones de ADNc de *PMS1* e *MLH1* humanos e de rato utilizando os fragmentos de RCP como sondas
- Etapa 5 Isolamento de clones genômicos de *PMS1* e *MLH1* humanos e de rato
- Etapa 6 Mapeamento posicional dos cromossomas de genes de *PMS1* e *MLH1* humanos e de rato por hibridação *in situ* por fluorescência
- Etapa 7 Utilização de sequências genômicas e de ADNc para identificar mutações nos genes *PMS1* e *MLH1* de famílias com CCNPH
- Etapa 8 Preparação de vectores para atingir genes para romper os genes *PMS1* e *MLH1* de ratos em células IE (indiferenciadas embriónicas); estudo de ratos deficientes na reparação de desemparelhamentos

Figura 1

2/24

```

MutL      1      10      20
          KPIQVLPDQLANQIAAGEVVERPASVVK - SEQ. ID NO: 1
HexB      1      10      20
          KSHIIEIPEMLANQIAAGEVIERPASVCK - SEQ. ID NO: 2
Pms1      1      10      20      30      40      50      60
          KFMHIEHLLIETEKRCCKOEORYIPVKYLFSTOIHQINDIDVHRITSQCVITDLTAVK - SEQ. ID NO: 3

MutL      40      50      60      70      80
          ELVENSIDAGATRVVDIDIERGGAKLIRIRDNCGGKKEELALALARMHATSKIASLDGLEA
          .....
HexB      40      50      60      70      80
          ELVENAIDAGSSQIIIEIEEAGLKCVQITDNCHGIAHDEVELALRRHATSKIKHQADLFR
          .....
Pms1      40      50      60      70      80      90      100      110      120
          ELVDNSIDAMNQIEIIFKDYCLEIECSDNCGDIDPSNYEFALQHYTSKIAKFQDVAK
          .....

MutL      100     110     120     130     140
          IISLCFRGEALASISSVSRLTILTSRTAEQAEAWQAYAEGRDMOVTVKPAHPVGTILEVL
          .....
HexB      100     110     120     130     140
          IRTLGRGEALPSIASVSVLTLLTAVDGCASHCTKLVARCGEVE.EVIPATSPVGTKVCVE
          .....
Pms1      100     110     120     130     140     150     160     170
          VQTLGRGEALSSLOGIAKLSVITTTSPPKADKELYDMVGHIT.SKTTTSRNKGTIVLVS
          .....

MutL      160     170     180     190     200
          DLFYNTPARRK.FMRTETETFMHIOEIIIRRIALAFDVTILSHNGKLVQYTRAVAKDQ
          .....
HexB      160     170     180     190     200
          DLFYNTPARLK.YKCSQAELSHIIDIYNRGLAHPEISFSLISDGKENTR...TAGTCQ
          .....
Pms1      160     170     180     190     200     210     220     230
          QLFHNLPRQKEFSKIFKQFTKCLTVIQGYAIIAAIKFSVWNITPKGKQNLILSTMRN
          .....

MutL      220     230     240     250     260
          KERRLGAIQCTPFLEQALAEWQHGDITLRGWVADPNHTTALTETIQCYVNGRMHRDL
          .....
HexB      220     230     240     250     260
          LRQAIAAGIYGLVSAKKHIEIENSIDLFEISGFVSLPELTRANRNYISL.FINGRYIKHFL
          .....
Pms1      220     230     240     250     260     270     280     290
          SSKRMJN.ISSVFCAGCMRGELEVVLVDLNPFFKRMILCKYTDOPDFLDLDYKTRVKCYIS
          .....

MutL      280     290     300
          INHAIHQACEDKGA.....DQOPAFVLYLEIDPHQGV
          .....
HexB      280     290     300
          INRAILDGFGSKLV.....GRFFLAVIHIXIDPYLADV
          .....
Pms1      280     290     300     310     320     330     340     350
          QNSFGCGRNSKQRFIYVNRKPVEYSTLLKCCNEVYKTFNNVQFFAVFLNLELPSLIDV
          .....

MutL      310     320     330     340     350     360
          NVHPAGHEVRFHOSRLVHDFIYQGVLSVLOQQTETALPLEELAPAPRHVQENRIAGRMH
          .....
HexB      310     320     330     340     350     360
          NVHPTKQEVRISEKELMTLVSEALANSLEQTLIPDALENLAKSTVRNREKVEQITLPL
          .....
Pms1      310     320     330     340     350     360     370     380     390     400     410
          HVTPOKRVILLHNERAVID.IFKITLSQYNNRQELALPKRMCSQSEQQAQKRLKTEVFDD
          .....

HexB      450     460     470     480     490     500
          SFPELEFFQGHNGTYLFA...QGROGLYIIDQHAAQERVYEEYRESIGNVDOSQOQLL
          .....
Pms1      450     460     470     480     490     500     510     520     530     540     550     560
          DFKKHEVVGQFNLGFIIVTRKVDNKSDFIVDQHASDEKYNFETLQAVTVF...KSQKLI
          .....

HexB      510     520     530     540     550     560
          VPIYIEFPADDALRLKERMPLLEEVGVFLAEYGENQFILRENPIWMAEEEIESGIYEMCD
          .....
Pms1      510     520     530     540     550     560     570     580     590     600     610
          IPQPVLSVIDELVLDNLVPVFEKNGFKIKIDEEEFGRSVKLISLPTSKQTLFDLGDPM
          .....

HexB      570     580     590     600     610
          MLLITKEVSIKKYRAELA.....IMSKRSIRKAMRIDHMSARQLLYQLSQCDNPFY
          .....
Pms1      570     580     590     600     610     620     630     640     650     660     670
          ELINLIKEDGGLRRDNIRCSKIRSHFAMRACRSSIMICKPLNKKTMTRVVHNLSELDKPM
          .....

HexB      620
          NCPHGRPVLVHFT
          .....
Pms1      620
          NCPHGRPTMRHLM
          .....
          680      690

```

Figura 2

3/24

CTTGGCTCTTCTGGCGCAAAATGCTGCTGGCAGGGGTTATTCGGCGGCTGGACGAG 60 — SEQ. ID NO: 4
 K S F V A G V I R R L D E
 ACAGTGGTGAACCCGATCGCGCGGGGGAAGTTATCCAGCGGCCAGCTAATGCTATCAAA 120 — SEQ. ID NO: 5
 T V V N R I A A G E V I O R E A N A I K
 GAGATGATTGAGAAGTGTGTAGATGCAAAATCCACAAGTATTCAAGTGATTGTTAAAGAG
 E H I E N C L D A N S T S I Q V I V K E 180
 GGAGGCTGAAGTTGATTGAGATCCAGCAATGGCACCGGGATCAGGAAGAAGATCTG
 G G L K L I O I O D N G T G I R K E D L
 GATATTGTATGTGAAGGTTCACTACTAGTAACTGCAGTCTTTGAGGATTAGCCAGT 300
 D I V C E R F T T S K L O S F E D L A S
 ATTTCTACCTATGGCTTCGAGGTGAGCTTTGGCCAGCATAAGCCATGTGGCTCATGTT 360
 I S T Y G F R G E A L A S I G H V A H V
 ACTATTACAAAGAAACAGCTGATGGAAAGTGTCATACAGAGCAAGTTACTCAGATGGA 420
 T I T T K T A D G K C A Y R A S Y S D G
 AAACGTAAAGCCCTCTTAACCATGTGCTGGCAATCAAGGGACCCAGATCACGGTGGAG 480
 F L K A P P K P C A G H G I G T V E
 GACCTTTTACACATAGCCAGGAGGAAAGCTTTAAAAAATCCAGTGGAAGATAT 540
 D L F Y N I A T R R R R A L K N P S E E Y
 GGGAAATTTTGAAGTTGTTGGCAGGTATTCACTACACATGAGGCAATTAAGTTCTCA 600
 C K I L E V V G R Y S V E N A G I S F S
 GTTAAAJAACAGGAGAGACAGTACGTGATGTTAGGACACTACCCAAATGCTTCAACCGT 660
 V K K Q G E T V A D V R K T L P N A S T V
 GACATATTGCTGCTATCTTTGGAATGCTGTTACTGAGAACGTAGAGAAATGAGATGT 720
 D H I R S I F G N A V S R E L I E I G C
 GAGGATAAAACCTTAGGCTTCAAAATGAATGGTTACATATCCAAATCACTACTCAGTG 780
 E D F T L A F K M H G Y I S H A N Y S V
 AAGAAAGTCACTCTTCTACTCTCACTACAGCATGTCTGTGAAGATCACTTCTTGGA 840
 K F C I F L P F I N E R L V E S T S L R
 AAAGCCATAGAAACAGGTATGAGGCTATTGGCCAAACACACACCCATCTCTGAC 900
 K A I E T V Y A A Y L P K N T H P F L Y
 CTCAGTTAGAAATCAGTCCCAAGATGGATGTTAATGTGACCCCAACAGCATGA 960
 L S L E I S P Q N V D V N V H P T K H E
 GTTCACTCTCTGACGAGGAGGAGCATCTGCGAGGGGTGACGACACATCGAGAGCAAG 1020
 V H F L S E E S I L E R Y O Q H I E S K
 CTCCTGGGCTCCAAATCTCCGAGGATGATCTTACCCAGACTTTGCTACGAGGACTGCT 1080
 L L G S N S S R H Y P T Q T L L P O L A
 GGGCTCTGCGGAGATGTGTAATCCACAAAGTGTGACCTCTCTTCTACTCTGGA 1140
 C P S G E N R S T S L T S S S T S G
 AGTAGCATAGAGTCTATCCGACGAGATGTCTGTACAGATTCCCGGGAAGAGAGCTT 1200
 S E D V Y A E O M V R T D S E Q L
 GATGCAATTTCTGAGGCTCTGAGCAACCCCTGTCCAGTCAAGCCCAAGGCTATGTCA 1260
 D A F L O F L S K E L S S O P Q A I V T
 GAGGATAGACAGATATTTCTAGTGGCAGGGCTAGCCAGCAAGATGAGGAGATGCTTGA 1320
 E D X T D I S S G R A R O Q D E E H L E
 CTCGCCGCCCTGCTGAGTGGCTGCCAAAJATCAGAGCTTGGAGGGGATACAAACAG 1380
 L P A P A E V A A K N Q S L E G D T T K
 GGCAGCTCAGAAATGTCAGAGAGAGAGGAGCACTATCTCCAGCAACCCCAAGAGACAT 1440
 C T S E M S E K R G P T S S N P R K R H
 CGGAAAGATTCTGATGTGAAATGGTGGAAAGATGATTCGGGAAAGGAAATGACTGCACT 1500
 R E D S D V E M V E D D S E R E H T A
 TGTACCCCGCGAGAGGATCATTAACCTCACTAGTGTGTTGAGTCTCCAGAGAGAAAT 1560
 C T P R R R I I N L T S V L S L Q E E I
 AATGACAGGGACATGAGGTTCTCCGAGGATGTTGCAATACCACTCTCTCTGGGCTGT 1620
 N E O G H E V L R E M L R N H S F V G C
 GTGAATCTCACTGAGGCTTGGCACAGCATCAAAACAAAGTTATACCTTCTCAACACCC 1680
 V N P Q W A L A Q H O T K L Y L L N T T
 AAGCTTAGTGAAGACTGTGTTACTACAGATACCTATTATGATTTTGGCAATTTTGGTGT 1740
 K L S E E L F Y Q I L I Y D F A N F G V
 CTCAGGTTATCGGAGCCAGCACCGCTCTTTGACCTTGCCATGCTTGGCTTAGATAGTCCA 1800
 L R L S E P A P L F D L A H L A L D S P
 GAGATGGCTGGACAGAGGAGATGGTCCAAAGAAAGGACTTGGTAAATACATTGTTGAG 1860
 E S G H T E E D O P K E G L A E Y I V E
 TTTCTGAAGAGAGAGGCTGAGATGCTTGCAGACTATTTCTCTTTGGAAATGATGAGGAA 1920
 F L K K K A E M L A D Y F S L E I D E E
 GGGAACTGATTGGATTACCCCTTCTGATTGACAACTATGTGCCCCCTTTGGAGGGACTG 1980
 G N L I G L P L L I D N Y V P P L E G L
 CCTATCTTCACTTCTGACTAGCCACTGAGGTGAATTGGGACGAAGAAAGGAATGTTTT 2040
 P I F I L R L A T E V N W D E E K E C F
 GAAAGCTCAGTAAAGAAATGCGCTATGTTCTATTCCATCCGGAAGCAGTACATATCTGAG 2100
 E S L S K E C A M F Y S I R K O Y I S E
 GAGTGCAGCCCTCAGGCCAGCAGAGTGAAGTGCCTGGCTCCATTCCAAACTCTCGAAG 2160
 E S T L S G O Q S E V P G S I P N S W R
 TGGACTGTGAACACATTTGCTATAAAGCCTTGGGCTCAGACATTTGCTCTCTTAACAT 2220
 N T V E H I V V K A L R S H I L P P K H
 TTCACAGAGATGGAATATCTGCGAGCTTGTAAAGCTGCTGATCTATACAAAGCTTTT 2280
 F T E D G N I L O L A N L P D L Y K V P
 GAGAGGCTGAATATGCTTATTTATGACCTGGGAGTGTGTTCTTTCTCTCTGTATTC 2340
 E R C
 CGATCAAAAGTGTGTATCAAAAGTGTGATATACAAAGCTGTACCAACATAAGTGTGGTAG 2400
 CACTTAAGACTTATACTTGGCTTCTGATAGTATTCCTTATACACAGTGGATGTGATATA 2460
 AATAATAGATGTGCTTAACATA 2484

Nucleótido de ADNc
 de MLH1 humana e
 sequência da
 proteína

Figura 3

4/24

#1: 18442 a 19109 (-21 a 116)

TGGCTGGATGCTAAGCTACAGCTGAAGGAAGAAGCTGAGCACGagggcactgaggt
gattggcTGAAGGCACTTCCGTTGAGCATCTAGACGTTTccttgctcttctggc
ggcaaaatgctcgttcgtggcaggggttattcggcggctggacgagacagtggga
accgcatcgcggcgggggaagtattccagcggccagctaattgctatcaaagagat
gattgagaactgTACGGAGGGAGT'CGAGCCGGgctcacttaagggtacgaCTT
AACGGGCCGCGTCACTCAATGGCGCGGACACGCTCTTCCCGGGCAGAGGCAT
GTACAGCGCATGCCACAACGGCGGAGGCCGCCGGGTTCCTACGTGCCATAAGC
CTTCTCCTTTTC

SEQ. ID
NO: 6

SEQ. ID NO: 25

#2: 19689 a 19688 (117 a 207)

AAACACGTTAATGAGGCACTATTGTTTGTATTGGAGTTTGTATCATTGCTTGG
CTCATATTAAaataatgtacattagagtagttgCAGACTGATAAATTATTTCTGT
TTGATTGGCAGTtttagatgcaaaatccacaagtattcaagtgtattgtaaaagag
ggaggcctgaagttgattcagatccaagacaatggcaccgggatcaggGTAAGTA
AAACCTCAAAGTAGCAGGATGTTTGTGCGCTTCATGGAAgagtcaggacctttct
ctgTTCTGGAAACTAGGCTTTTGCAGATGGGATTTTTCCTGAAAAATTCAACA
CCAACAATAAATATTTATTGAGTACCTATTATTGCGGGGCACTGTTACGGGGAT
GTGTCAGT

SEQ. ID
NO: 7

SEQ. ID NO: 26

#3: 19687 a 19786 (208 a 306)

TTTCCTGGATTAATCAAGAAATGGAATTCAAagagatttggaataatgagtaacAT
CATTATTTACTCATCTTTTGGTATCTAACAGaaagaagatctggataattgtatg
tgaaagggttcactactagtaaaactgcagtcctttgaggatttagccagttattct
acctatggctttcagaggtgagGTAAGCTAAAGATTCAAGAAATGTGTAATAATc
ctcctgtgatgacattgtCTGTCAATTGTTAGTATGTATTCTCAACATAGATAA
ATAAGGTTTGGTACCTTTTACTTGTAAATGTATGCAAACTGAGCAAACCTTAAT
GAACPTTAACTTTCAAAGACTG

SEQ. ID
NO: 8

SEQ. ID NO: 27

#4 18492 a 18421 (307 a 380)

TGGAAGCAGCAGNCAGATaaccctttcccttttggtgaggtGACAGTGGGTGACCCA
GCAGTGAGTTTTTCTTTCAGTCTATTTCTTTCTTCTTACGgctttggccagca
taagccatgtggctcatgttactattacaacgaaaacagctgatggaaagtgtgc
atacagGTATAGTGCTGACTTCTTTTACTCATATATATTCAATTCTGAAATGTATT
TTGGgcttaggtctcagagtaateCTGTCTCAACACCAGTGTATCTTTNNNGGC
AGAGATCTTGAGTACG

SEQ. ID
NO: 9

SEQ. ID NO: 28

Figura 4A - 1

5/24

#5: 18313 a 18179 (381 a 453)

TTGATATgattttctctttttcccttgggATTAGTATCTATCTCTCTACTGGATA
TTAATTTGTTATATTTTCTCATTAGagcaagttactcagatggaaaactgaaagc
ccctcctaaccatgtgctggcaatcaagggaccagatcacgGTAAGAATGGTA
CATGGGAGAgtaaattgttgaagctttgtttgTATAAATATTGGAATAAAAAATA
AAATTGCTTCTAAGTTTTCAGGGTAATAATAAAATGAATTTGCACTAGTTAATGG
AGGTCCCAAGATATCCTCTAAGCAAGATAAATGACTATTGGCTTTTNNNTGGCATG
GCAGCCTG

SEQ. ID
NO: 10

SEQ. ID NO: 29

#6: 18318 a 18317 (454 a 545)

GCTTTTGCCAGGACCATCTTgggtttttattttcaagtacttctatgAATTTACAA
GAAAAATCAATCTTCTGTTTCAGgtggaggaccttttttacaacatagccacgagg
agaaaagctttaaaaaatccaagtgaaagaatatgggaaaattttggaagttgttg
gcagGTACAGTCCAAAATCTGGGAGTGGGTCTCTGAGATTGTTCATCAAAGTAAT
GTGTTCTAGTgctcatatcattgaacagttgctgagcTAGATGGTGAAAAGTAAAA

SEQ. ID
NO: 11

SEQ. ID NO: 30

#7: 19009 a 19135 (546 a 588)

CAGCAACCTATAAAAGTAGAGAGGAGTCTGTGTTTTGACGCAGCACCTTTAGCAT
TTTTATTGGATGAAGTTTCTGCTGGTTTATTTTTCTGTGGGTAAAAATATTAATA
GGCTGTATGGAGATATTTTTCTTTATATGTACCTTTGTTTAGATTACTCAACTCC
ACTAATTTATTTAACTAAAAGGGGGCTCTGACATctagtgtgtgtttttggcAAC
TCTTTTCTTACTCTTTTGTCTTTTCCAGgtatttcagtcacacaatgcaggca
ttagtttctcagttaaaaaaGTAAGTTCTTGGTTTATGGGGGATGGTTTTGTGTTT
ATGAAAAGAAAAAAGGGGATTTTAAATAGTTTGCTggtggagataaggttatgAT
GTTT

SEQ. ID
NO: 12

SEQ. ID NO: 31

#8: 18197 a 18924 (589 a 677)

ATGTTTCAGTctcagccatgagacaataaaatccTTGTGTCTTCTGCTGTTTGTTT
ATCAGCaaggagagacagtagctgatgttaggacactacccaatgcotcaaccgt
ggacaatatcgtccatcttttggaaatgctgttagtcgGTATGTCGATAACCTA
TATAAAAAAATCTTTTACATTTATTTATCTTGGTTTATCATTccatcacattattt
gggaaacctTTCAAGATATTATGTGTGTTAAGAGTTTGCTTTAGTCAAATACACAG
GCTTGTTTTATGCTTCAGATTTGTTAATGGAGTTCTTATTTACGTAATCAACAC
TTTCTAGGTGTATGTAATCTCCTAGATTCTGTGGCGTGAATCATGTGTTCT

SEQ. ID
NO: 13

SEQ. ID NO: 32

Figura 4A - 2

6/24

#9: 18765 a 18198 (678 a 790)

ACTGAGTAGGGTAGGTGGGTGAGTGGGTGGGTGGGTGGGTGGGTGGATGGATGGA
TGGGAGGATGGGTGGGTGAATGGGTGAACAGACAAATGGATGGATGAATGGACAG
GCACAGGAGGACCTCAAATGGACCAAGTCTTCGGGGCCCTCATTTCACAAAGTTA
GTTTATGGGAAGGAACCTTGTTGTTTTTAAATTCTGATTCTTTTGTAAATGTTTGAG
TTTTGAGTATTTTcaaaagcttcagaatctcTTTTCTAATAGagaactgatagaa
attggatgtgaggataaaacccctagccttcaaaatgaatggttacatatccaatg
caaaactactcaagtgaagaagtgcatcttcttactcttcatcaaccGTAAGTTAAA
AAGAACCACATGGGAAATccactccaggaacacacacagGGAATTTTATGGGA
CCATGGAAAAATTTCTGAGTCCATAGGTTTGATTAAACATGGAGAAACCTCATGG
CAAAGTTTGGTTTTATTGGGAAGCATGTATA

SEQ. ID
NO: 14

SEQ. ID NO: 33

#10: 18305 a 18306 (791 a 884)

ATAGTGGGCTGGAAAGTGGCCACAGCTAAAGGTGCACCTTTCTTCCTGGGGATGT
GATGTGCATATCACTACAGAAATGTCTTTCTGAGGTGATGTCatgacttttgtgt
gaatgtacacctGTGACCTCACCCTCAGGACAGTTTGAAGTGGTTGCTTTCTT
TTTATTGTTTAGatcgtctggtagaatcaacttccttgagaaaagccatagaaac
agtgtatgcagcctatttggcccaaaacacacacacccattcctgtacctcagGTAA
TGTAGCACCAAACTCCTCAACCAAGACTCACAAGGAacagatgttctatcagcct
ctctctTTGAAAGAGATGAGCATGCTAATAGTACAATCAGAGTGAATCCCATAC
ACCACTGGCAAAAGGATGTTCTGTCCCTTTTACAGGTACAAGGCACAG

SEQ. ID
NO: 15

SEQ. ID NO: 34

#11: 18182 a 19041 (885 a 1038)

CTTACGCAAAGCTACACAGCTCTTAAGTAGCAGTGCCAATATTGAAACACACTCA
GACTCGAGCCTGAGGTTTTGACCACTGTGTCATCTGGCCTCAAATCTTCTGGCCA
CCACATACACCATATGTgggctttttctccccctcccACTATCTAAGGTAAATGT
TCTCTCTTATTTTCTGACAGtttagaaatcagtcgccagaaatgtggatgttaat
gtgcacccacaaagcatgaagtttacttcctgcacgaggagagcatcctggagc
gggtgcagcagcatcgagagcaagctcctgggctccaattcctccaggatgta
cttcaccagGTGAGGGCGCTTCTCATCCAGCTACTTCTCTGGGGCCTTTGAAAT
GTGCCCCGCCAGcgtgagagcccagattttTGCTGTTATTTAGGAACTTTTTTT
GAAGTATTACCTGGATAG

SEQ. ID
NO: 16

SEQ. ID NO: 35

Figura 4A - 3

7/24

#12: 18579 a 18178 (1039 a 1409)

GATaattatacctcactagcTTCTTTCTTAGTACTGCTCCATTTGGGGACCTG
TATATCTTACTTCTTATTCTGAGTCTCTCCACTATATATATATATATATA
TTTTTTTTTTTTTTTTTTTTTAAATACAGaottttgctaccaggacttgctggcccc
tctggggagatggttaaattccacaacaagtctgacctgctcttctacttctggaa
gtagtataaggtctatgcccacagatggttcgtacagattcccgggaacagaa
gcttgatgcatttctgcagcctctgagcaaacctctgcccagtcagcccaggcc
attgtcacagaggataagacagatatcttctagtggcagggttaggcagcaagatg
aggagatgcttgaactcccagcccctgctgaagtggctgccaaaaatcagagctt
ggagggggatacaacaaagggaacttcagaaatgtcagagaagagaggacctact
tccagcaaccccagGTATGGCCTTTTGGGAAAAGTACAGCCTAcctcctttattc
tgtaataaaaacTGCCTTCTAACTTTGGGCTTTTCATGAATCACTTGCATCTTCTCT
CTGCCGACTTCCC

SEQ. ID
NO: 17

SEQ. ID NO: 36

Crê-se que o sítio aceitador tenha 21 T's

#13: 18420* a 18443 (1410 a 1558)

CTGTGCTCCAGCACAGGTCATCCAGCTCTGTAGACCAGCGCAGAGAAGTTGCTTG
CTCCCAAAtgcaaccccaaaaatttggcTAAGTTTAAAAACAAGAATAATAATGA
TCTGCACTTCCTTTTCTTCATTGCAGaaagagacatcggaagattctgatgtgg
aaatggtggaagatgattcccgaaggaatgactgcagcttgtaacccccggag
aaggatcattaacctcaactagtgttttgagctctccaggaagaaattaatgagcag
ggacatgaggGTACGTAAACGCTGTGGCCTGCCCTGGGATGCATAGGGCCTCAACT
GCCAaggttttggaaatggagaaaagCAGTCATGTTGTGAGAGTGGCACTACAGTT
TTGATGGGCAAGCTCCTCTTCTTTACTAACCACAAATAGCATCAGCTTAAAGAC
AATTTTGTATTGGGAGAAAAGGGAGAAAATAATCTCTG

SEQ. ID
NO: 18

SEQ. ID NO: 37

#14: 19028 a 18897 (1559 a 1667)

CAGTTTTCACCAGGAGGCTCAAATCAGGCNCTTTGCTTACTtggtgtctctagt
tctggTGCCTGGTCTTTGGTCAATGAAGTGGGGTTGGTAGGATTCTATTACTTA
CCTGTTTTTTGGTTTTATTTTTTGTTCAGTtctccgggagatggtgcataac
cactccttcgtgggtgtgtgaatcctcagtgggccttggcacagcatcaaacca
agttataccttctcaacaccaccaagcttagGTAAATCAGCTGAGTGTGTGAACA
AgcagagctactacaacaatgGTCCAGGGAGCACAGGCACAAAAGCTAAGGAGAG
CAGCATGAAGGTAGTTGGGAAGGGCACAGGCTTTGGAGTCAGCACATGT

SEQ. ID
NO: 19

SEQ. ID NO: 38

Figura 4A - 4

8/54

#15: 19025 a 18575 (1668 a 1731)

CCCCCTGGTTGAAGCGTTGGAATCCCCTCTTTGGANNNNNNAGATTGTGTAGTA
CTGTTAACCAGATTCCACAGCCAGGCAGAACTATGTCGTCTCATCCATGTGTCA
GGGATTACGTCTCCCATTTGTCCCACTGGTTGTATCTCAAGCATGAATTCAGCT
TTTCCTTAAAGTCACCTTCATTTTATTTTTCAGTgaagaactgttctaccagatac
tcatttatgattttgcaattttggtgttctcaggtatcgGTAAGTTTAGATCC
TTTTCACTTctgacatttcaactgcccGCCCGCAAACAGTAGCTCTCCACTAAA
TA

SEQ. ID
NO: 20

SEQ. ID NO: 39

-19 do sítio aceitador da secção é A nalgumas pessoas.
Outros são heterozigóticos para A e G. GTCACCTTC ou CTCGCTTC
(Polimorfismo)

#16: 18184 a 18314 (1732 a 1896)

CATTTATGGTTTCTCACCTGCCATTCTGATAGTGGATTCTTGGGAATTCAGGCTT
catttgatgctccgttaaagcTTGCTCCTTCATGTTCTTGCTTCTTCTAGGag
ccagcacgcgtotttgaccttgccatgcttgcccttagatagtcagagagtggtct
ggacagaggaagatggtccaaaggaaggacttgctgaatacattgttgagtttct
gaagaagaaggctgagatgcttgacagactatttctctttggaaattgatgagGTG
TGACAGCCATTCTTATACTTCTGTGTATCTCcaaaataaaatttccagccgggt
gCATTTGGCTCA

SEQ. ID
NO: 21

SEQ. ID NO: 40

#17: 18429 a 18315 (1897 a 1989)

CAGATAGGAGGCACAAGGCCTGggaaggcactggagaaatgggATTGTTTAAA
CTATGACAGCATTATTTCTGTTCCCTTGTCCTTTTCTGCAAGCAGgaagggg
acctgattggattaccctcttgattgacaactatgtgccccctttggagggact
gcctatcttcattcttcgactagccactgagGTCAGTGATCAAGCAGATACTAAG
CATTTcgggtacatgcatgtgtgctggagggAAAGGGCAAA

SEQ. ID
NO: 22

SEQ. ID NO: 41

#18: 18444 a 18581 (1990 a 2103)

CTATATCTTCCAGCAATATTCACAGTCCGTTTACAGTTTTAACGCCTAAAGTAT
CACATTTTCGTTTTTTAGCTTTaagtagtctgtgatctccgTTTGAATGAGAAATG
TTTAAATTCGTACCTATTTTGAGGTATTGAATTTCTTGGACCAGgtgaattggg
acgaagaaaaggaatgttttgaagcctcagtaagaatgcgctatgttctattc
catccgggaagcagtacatatctgaggagtcgaccctctcagggccagcagGTACAG
TGGTGATGCACACTGGCACCCAGGACTAggacagggacctcatatCTTAGGAG
ATGAAACTTG

SEQ. ID
NO: 23

SEQ. ID NO: 42

Figura 4A - 5

9/24

#19: 18638 a 18637 (2104 a 2271). 2463 é o fim de
ADNC

AATCCTCTTGTTTCAGGCGCTGGATCCCTGAGAGGCTAGCCCACAAGATCCAC
TTCAAAGCCCTAGATAACACCAAGTCTTTCCAGACCCAGTGACATCCCATCAG
CCAGGacaccagtgtatgttggGATGCAAACAGGGAGGCTTATGACATCTAATGT
GTTTTCCAGagtgaagtgcctgggtccattccaaactcctggaagtggactgtgg
aacacattgtctataaagccttgcgtcacacattctgcctcctaaacatttcac
agaagatggaaatatcctgcagcttgcctaaccctgcctgatctatacaaaagtcttt
gagaggtgtfaaatatggttattttatgcactgtgggatgtgtttctttctctg
tattccgatacaaaagtgtgtatcaaaagtgtgatatacaaaagtgtaccaacataa
gtgttggtagcacttaagacttatacttgccttctgatagtattcctttatacac
agtggattgattataaataaataagatgtgtcttaacataATTTCTTATTTAATTT
TATTATGTATATA

SEQ. ID
NO: 24

SEQ. ID NO: 43

Figura 4A - 6

INICIADORES DE AMPLIFICAÇÃO DO EXÃO hMLH1

Iniciador de amplificação
do primeiro estágio

	SEQ. ID NO.	Iniciador de amplificação do segundo estágio	SEQ. ID NO.
Exão 1			
N-18442- 5'aggcacigagg'gattggc	44	N-19295- 5'gtaaaaaacgacggccagctgaggtgattggctgaa	83
C-19109- 5'tcgtagcccttaagtgagc	45	C-19446- *5'tagoccttaagtgagcccg	84
Exão 2			
N-19689- 5'aatagtacattagagtagttg	46	N-18685- 5'gtaaaaacgacggccagttacattagagtagtgcaga	85
C-19688- 5'cagaganaaggctcctgactc	47	C-19067- *5'aggctcctgactcttccatg	86
Exão 3			
N-19687- 5'agagatttggaaaaatgagtaac	48	N-18687- 5'gtaaaaacgacggccagtttggaaaaatgagtaacattgatt	87
C-19786- 5'acaatgtcatcacagaggagg	49	C-19068- *5'gtcatcacaggaggatat	88
Exão 4			
N-18492- 5'aacctttccctttggtgagg	50	N-19294- 5'gtaaaaacgacggccagctcttccctttggtagagtgaa	89
C-18421- 5'gattactctgagaccctagggc	51	C-19077- *5'tactctgagaccctagggccca	90
Exão 5			
N-18313- 5'gattttctcttttcccttgggg	52	N-19301- 5'gtaaaaacgacggccagttctcttttcccttgggatttag	91
C-18179- 5'caacaaagcttcaacaatttac	53	C-19046- *5'acaaagcttcaacaatttacct	92

Figura 4B - página 1

10/24

Iniciador de amplificação do primeiro estágio	SEQ. ID <u>NO.</u>	Iniciador de amplificação do segundo estágio	SEQ. ID <u>NO.</u>
Exão 6			
N-18318- 5'gggtttttttcaagtacttctatg	54	N-19711- 5'tgtaaaacgacggccagtggttttttcaagtacttctatgaatt	93
C-18317- 5'gtcagcaactgttcaatgtatgagc	55	C-19079- *5'cagcaactgttcaatgtatgagcact	94
Exão 7			
N-19009- 5'ctagtgtgttttttggc	56	N-19293- 5'tgtaaaacgacggccagtggtgtgtttttggcaac	95
C-19135- 5'cataaccttattccacc	57	C-19435- *5'aaccttattctccaccagc	96
Exão 8			
N-18197- 5'ctcagccatgagacaataaatcc	58	N-19329- 5'tgtaaaacgacggccagtgccatgagacaataaatccttg	97
C-18924- 5'ggttcaccaataatgtagagg	59	C-19450- *5'tcccaataatgtagagggaatg	98
Exão 9			
N-18765- 5'caaaagcttcagaatctc	60	N-19608- 5'-tgtaaaacgacggccagtgagcttcagaatctctttt	99
C-18198- 5'ctgtgggtgtttctgtgagtg	61	C-19449- *5'-tgggtgtttctgtgtgagtggaatt	100
Exão 10			
N-18305- 5'catgactttgtgtgaagtacacc	62	N-19297- 5'tgtaaaacgacggccagtgactttgtgtgaagtacacctgtg	101
C-18306- 5'gaggagagccgatagaacatctg	63	C-19081- *5'gagagccgatagaacatctgttg	102

Figura 4B - página 2

22/24

Iniciador de amplificação do primeiro estágio	SEQ. ID NO:	Iniciador de amplificação do segundo estágio	SEQ. ID NO:
Exão 11			
N-18182- 5'gggctttttcccccctcc	64	N-19486- 5'tgtaaaacgacggccagctttttcccccctccacta	103
C-19041- 5'aaaatcgggctctcacg	65	C-19455- 5'tcgggctctcacgct	104
Exão 12 (Ver nota no fim)			
N-18579- 5'aattataccatactagc	66	N-20546- 5'cttattctgagctctcc	105
C-18178- 5'gtttattacagaataaaggagg	67	C-20002- 5'tgtaaaacgacggccagtggttcagaggctgc	106
		N-19829- 5'gatggctcgtacagattcccg	107
		C-19385- 5'tgtaaaacgacggccagtttattacagaataaaggaggtag	108
Exão 13			
N-18420- 5'tgcaaccacaaaatttggc	69	N-19300- 5'tgtaaaacgacggccagtaaccacaaaatttggctaag	109
C-18443- 5'ctttccatttccaaaacc	70	C-19078- 5'tctccatttccaaaacctg	110
Exão 14			
N-19028- 5'tggctctctagttctgg	71	N-19456- 5'tgtctctagttctggctc	111
C-18897- 5'cattgtgttagtagctctgc	72	C-19472- 5'tgtaaaacgacggccagttgtgttagtagctctgcttg	112
Exão 15			
N-19025- 5'ccatttgcaccaactgg	73	N-19697- 5'atttgcaccaactggttga	113
C-18575- 5'cggtcagttgaaatgtcag	74	C-19466- 5'tgtaaaacgacggccagttcagttgaaatgtcagaagtg	114

Figura 4B - página 3

12/24

Iniciador de amplificação do primeiro estágio	SEQ. ID <u>NO:</u>	Iniciador de amplificação do segundo estágio	SEQ. ID <u>NO:</u>
Exão 16			
N-18184- 5'catttggagctccgttaaagc	75	N-19269- 5'tgtaaaacgacggccagt	115
C-18314- 5'cacccggctggaaattttatttg	76	C-19047- *5'ccggctggaaattttatttgag	116
Exão 17			
N-18429- 5'ggaaagcactgggagaaatggg	77	N-19298- 5'tgtaaaacgacggccagtaggcactggagaaatgggatttg	117
C-18315- 5'ccctccagcacacatgcatgtaccg	78	C-19080- *5'tccagcacacatgcatgtaccgaaat	118
Exão 18			
N-18444- 5'taagtagctgtagtccg	79	N-19436- *5'gtagctgtagtccggtt	119
C-18581- 5'atgtatgaggtcctgtcc	80	C-19471- 5'tgtaaaacgacggccagtaggaggtccgtcctag	120
Exão 19			
N-18638- 5'gacaccagtgtagtgg	81	N-19447- *5'accagtgtagtgggaag	121
C-18637- 5'gagaaagaagaacacatccc	82	C-19330- 5'tgtaaaacgacggccagtagaagaagaacacatcccaca	122

Todas as sequências lêem-se de 5' para 3'. Os números de identificação do iniciador estão listados antes de cada sequência de iniciador. N indica o iniciador no lado 5' do exão. C indica o iniciador no lado 3' do exão.
* Indica que o nucleótido de 5' é biotinilado

Figura 4B - página 4

HUMAN 100
YEAST 97
HUMAN 200
YEAST 197
HUMAN 296
YEAST 296
HUMAN 380
YEAST 396
HUMAN 479
YEAST 481
HUMAN 571
YEAST 581
HUMAN 663
YEAST 676
HUMAN 756
YEAST 769

SEQ. ID NOS:

5 e 123

Figura 5

15/24

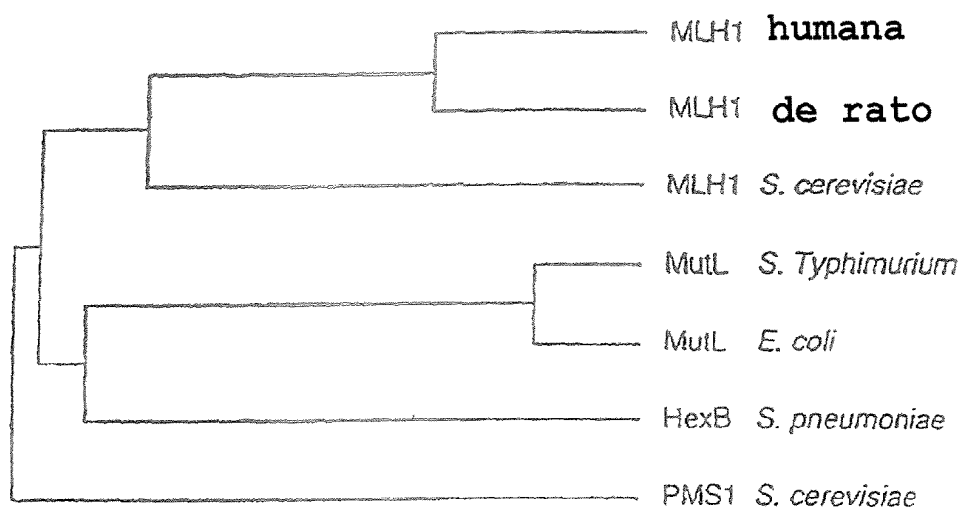


Figura 6

16/24

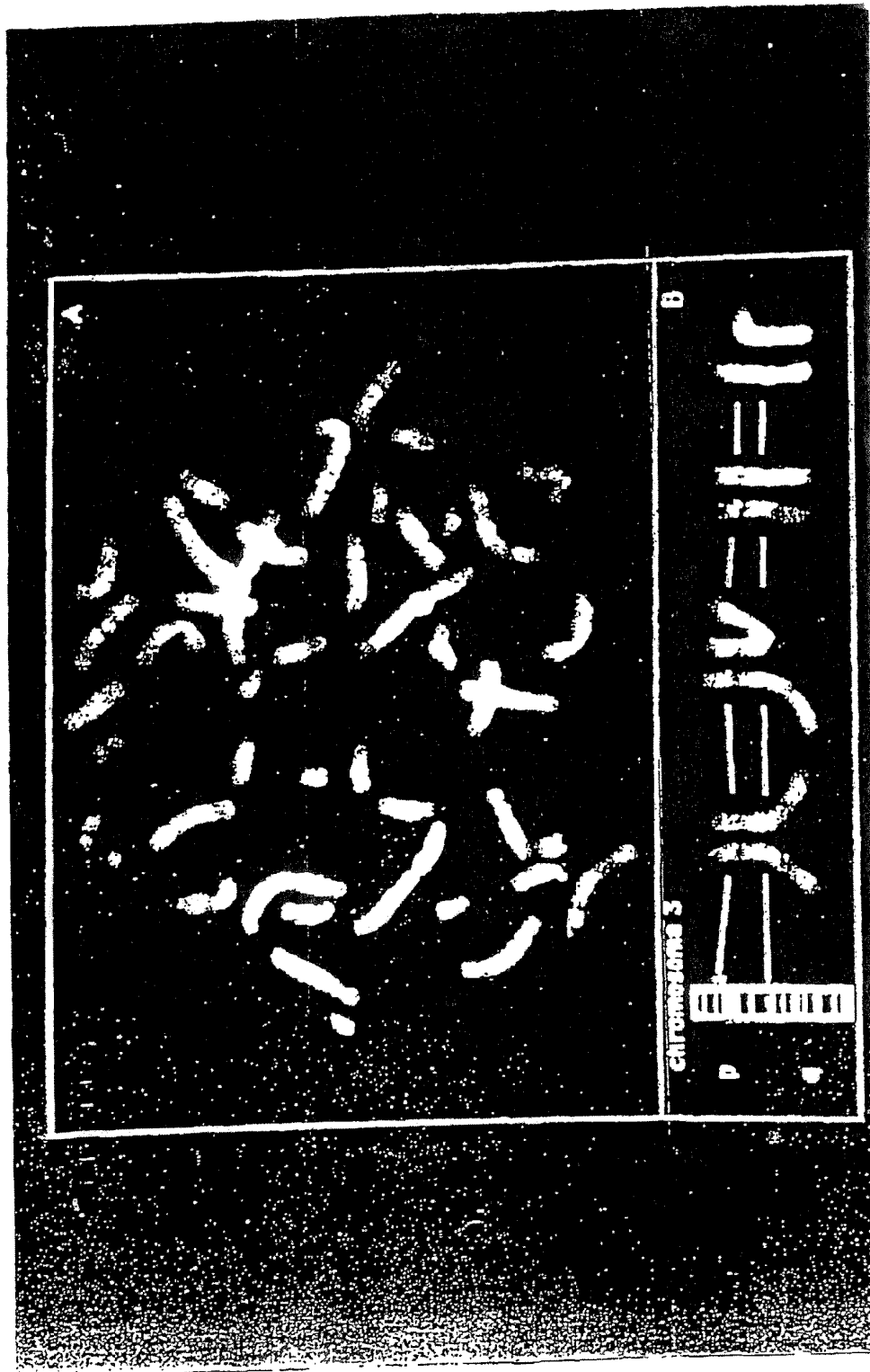


Figura 7

17/24

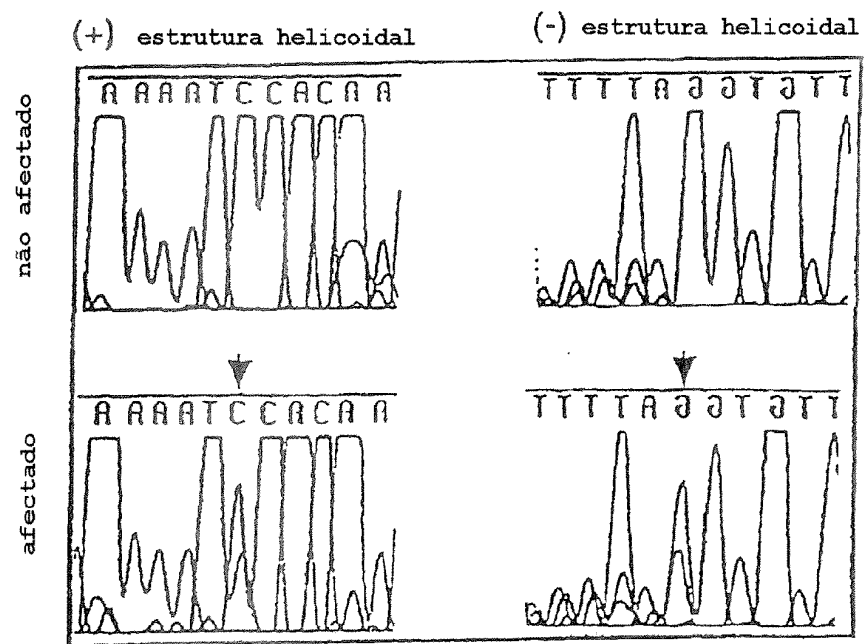


Figura 8

Sequência de nucleótido de EMS1 humana. O possível início (ATG) e os códons de
paragem (TGA) estão sublinhados

	10	20	30	40	50	60	70	80	
1	CCATCGAGCG	AGCTGAGAGC	TCGAGTACAG	AACCTGCTAA	GGCCATCAAA	CCTATTGATC	GGAACTCAGT	CCATCAGATT	80
81	TGCTCTGGGC	AGGTGGTACT	GAGTCTAAGC	ACTGCGGTAA	AGGAGTTAGT	AGAAAACAGT	CTGGATGCTG	GTGCCACTAA	160
161	TATTGATCTA	AAGCTTAAGG	ACTATGGAGT	GGATCTTATT	GAAGTTTCAG	ACAATGGATG	TGGGGTAGAA	GAAGAAAATC	240
241	TCGAAGGCTT	AACCTGAAA	CATCACACAT	CTAAGATYCA	AGAGTTTGCC	GACCTAACCT	AGGTTGAAAC	TTTTGGCTTT	320
321	CGGGGGGANG	CTCTGAGCTC	ACTTTGTGCA	CTGAGCGATG	TCACCATTTT	TACCTGCCAC	GCATCGGCGA	AGGTTGGAAC	400
401	TCGACTGATG	TTTGATCACA	ATGGGAAAAT	TATCCAGAAA	ACCCCTTACC	CCCCCCCCAG	AGGGACCACA	GTGAGCGTGC	480
481	AGCAGTTATT	TTCCACACTA	CCTGTCCGCC	ATAAGGAATT	TCAAAGGAAT	ATTAAGAAGG	AGTATGCCAA	AATGGTCCAG	560
561	GTCTTACATG	CATCTGTAT	CATTTCAGCA	GGCATCCGTG	TAACTTGCAC	CAATCAGCTT	GGACAAGGAA	AACGACAGCC	640
641	TGTGCTATGC	ACAGGTGGAA	GGCCGACGAT	AAAGGAAAAT	ATCGGCTCTG	TGTTTGGGCA	GAAGCAGTTG	CAAGGCTTCA	720
721	TTCTTTTGT	TCAGCTGCCC	CCTAGTGACT	CCGTGTGTGA	AGAGTACGGT	TTGAGCTGTT	CGGATGCTCT	GCATAATCTT	800
801	TTTTACATCT	CAGGTTTCAT	TTCACAATGC	ACGCATGGAG	TTGGAAGGAG	TTCAACAGAC	AGACAGTTTT	TCCTTATCAA	880
881	CCCCCGGCTT	TGTGACCCAG	CAAAGGTCTG	CAGACTCGTG	AATGAGGTCT	ACCACATGTA	TAATCGACAC	CAGTATCCAT	960
961	TTCTTTCTCT	TAACATTTCT	GTGATTCAG	AATGCGTTGA	TATCAATGTT	ACTCCAGATA	AAAGGCCAAT	TTTCTACAAA	1040
1041	CAGGAAAAGC	TTTTGTTGGC	AGTTTAAAG	ACCTCTTTGA	TAGGAATGTT	TGATAGTGAT	GTCAACAAGC	TAAATGTGAG	1120
1121	TCAGCAGCCA	CTGCTGGATG	TTGAAGGTAA	CTTAATAAAA	ATGCATGCAG	CGGNTTGGGA	AAAGCCCATG	GTAGAAAAGC	1200
1201	AGGATCAATC	CCCTTCATTA	AGGACTGGAG	AAGAAAAAAA	AGACGTGTCC	ATTTCCAGAC	TGCGAGAGGC	CTTTTCTCTT	1280
1281	CGTCACACAA	CAGAGAACAA	GCCTCACAGC	CCAAAGACTC	CAGAACCAAG	AAGGAGCCCT	CTAGGACAGA	AAAGGGGTAT	1360
1361	CCTGTCTTCT	AGCACTTCAG	GTGCCATCTC	TGACAAAAGC	GTCTTGAGAT	CTCAGAAAGA	GGCAGTGAGT	TCCAGTCACG	1440
1441	GACCCAGTGA	CCCTACGGAC	AGAGCGGAGG	TGGAGAAGGA	CTCGGGGCAC	GGCAGCACTT	CCCTGGATTG	TGAGGGGTTC	1520
1521	AGCATCCACG	ACACGGGGCAG	TCACCTGCAGC	AGCGAGTATG	CGCCAGCTC	CCCGGGGGAC	AGGGCTCCGC	AGGAACATGT	1600
1601	CGACTCTCCAG	GAGAAAGCGC	CTGAAACTGA	CGACTCTTTT	TCAGATGTGG	ACTGCCATTC	AAACCAGGAA	GATACCGGAT	1680
1681	GTAAATTTTCG	AGTTTTCGCT	CAGCCAACTA	ATCTCGCAAC	CCCAACACAA	AAGCGTTTTA	AAAAAGAGA	AATTCTTTCC	1760
1761	AGTTCTGACA	TTTGTCAAAA	GTTAGTAAAT	ACTCAGGACA	TGTCAGCCTC	TCAGGTGAT	GTAGCTGTGA	AAATTAATAA	1840
1841	GAAAGTTCTG	CCCCTGGAAT	TTTCTATGAG	TTCTTTAGCT	AAACGAATAA	AGCAGTTACA	TCATGAAGCA	CAGCAAAAGT	1920
1921	AAGCGGACAA	GAATTACAGG	AATTTTAGGG	CAAGATTTG	TCCTGGAGAA	AATCAAGCAG	CCGAGATGA	ACTAAGAAAA	2000
2001	GAGATAACTA	AAACGATGTT	TGCAGAAATG	GAAATCATTG	GTCACTTTAA	CCTGGGATTT	ATAATAACCA	AACCTGAATGA	2080
2081	GGATATCTTC	ATAGTGGACC	AGCATGCCAC	GGACGAGAA	TATAACTTCG	AGATGCTGCA	GCACACACCC	GTCTCCAGG	2160
2161	GGCAGAGGCT	CATAGCACCT	CAGACTCTCA	ACTTAACTGC	TGTTAATGAA	GCTGTTCTGA	TAGAAAAATC	CGAAATATTT	2240
2241	AGAAAGAATG	GCTTTGATTT	TGTTATCGAT	GAAAAATGCTC	CAGTCACTGA	AAGGGCTAAA	CTGATTTTCT	TGCCAACTAG	2320
2321	TAAAACTGG	ACCTTCGGAC	CCCAGGACGT	CGATGAACCTG	ATCTTCATGC	TGAGCGACAG	CCCTGGGGTC	ATGTGCCGCC	2400
2401	CTTCCGAGT	CAAGCAGATG	TTTGCTCCCA	GAGCCTGCCG	GAAGTCGGTG	ATGATTGGGA	CTGCTCTCAA	CACAAGCGAA	2480
2481	TGAAGAACT	GATCACCAC	ATGGGGGAGA	TGGGCCACCC	CTGGAACTGT	CCCATGGAAA	GGCCACCATG	AGACACATCG	2560
2561	CCAACTGGG	TGTCAATTTCT	CAGAACTGAC	CGTAGTCACT	GTATGGAATA	ATTGTTTTTA	TGCCAGATTT	TTATGTTTTG	2640
2641	AAAGACAGAG	TCTTCACTAA	CCTTTTTTGT	TTTAAATGA	AACCTGC				2687

SEQ. ID NO: 132

Figura 10

Reporte do alinhamento de PMS1, utilizando o processo de Clustal com o quadro ponderado do resíduo de PAM250

Quarta-feira, 26 de Janeiro de 199 4:51 PM

```

PMS1_HUMAN M-ERASSSTEPAM-----ATKPTDRKSVHCTCSGOVLSISTAVKELWENSIDAGATNIDLKLDYQVDLIEVSDNCCVVEENF 80
PMS1_YEAST MFHHIENLLIETETRCCKQKEQRYIPVKYLFSTQTHQNDIDVHRTSGQITITITAVKELWONSIDANANQETIIFKDYGLESIHCSDNCCGIDPSY 100

PMS1_HUMAN EGUTLKHHTSKIQERALTQVEITFGFGEALSSICALSQVTISICHASAVGTRIMFDHNGHITQKTPYFPPGTTVSQCLESTLPVHKEHQRNHKE 180
PMS1_YEAST EFLALKVYTSKIAKQDVAVMQLGFRGEALSSICIAKLSVITTSPPHAD-KLEYDMVGHITSKITTSENGGTVLSGLEHNLVVBQKESKTFNRQ 199

PMS1_HUMAN YAMVQVLAHVOITISGIRVSCINQLGQKRPVVCTGGSPAIKENISVFGQKQLSI--IPFI-QIPI-SDSVCEEGLSCSDALHNIFF---ISFI 273
PMS1_YEAST FTCLINIQQYALINAIKFSVNIITPKGKVLILSTMRNNSMRKNISVFGAGCMRGLIEVDILDLNFKNRMICKITDD--PDFLDLQKIRVQYI 297

PMS1_HUMAN SQCTHVGHSSTDRORFFINRRICDPKVCRLVNEVMMYRHOVHFVNLISVDECVINVTDPKRQILLQESKLLAVLKTSITGMFSDVANKINVS 373
PMS1_YEAST SONSFCGRNSNDROFIYVNRVVEYSTLLKCCNEVKTFTNVQFEAFLELPSLIIMVVTDPKRVILLHNRRAVIDIFKITLSDYNNRQ--ELALP 395

PMS1_HUMAN QQPLLDVGNLII-MHAADLEKPMVEKQDQSPSLTIGEEKDVISIRLEAFSLRHTTENPHSPKTPPERRSPLOKRGMLSSTSGAISDKVIRSQK 472
PMS1_YEAST KRCMSQSEQQAGVRLKTEVDDRSTHESDENVHARSESNOG---HAFNSTGTVIDSNGTELTSMVDGNYTNVTDVIGSECEVSDSSVLDEN 492

PMS1_HUMAN EAVSSSHQPDPTDRAEVEKDSGHGSTSVDSSEGFSPITGSHCSSEYAASSPCGRSQEHVDSQEKAPETDSSSVVCHSNQETGCKFFVLPQHINLA 572
PMS1_YEAST SSTPTKKLPITKISQNLSDNLNANFNPEFONITSPIKARSLEKVVPEPVYITIDGKFOEKAVLSQADGLMVINECHEHTNCCHQEHRGSTDIQD 592

PMS1_HUMAN TPNTKRFKKEIILSSS-----DICQLVNTQDMSAQVDVAVKINKKVVHDFSVSS-----LAMRIQLHHEAQQSEGHQNYRKFRAMICPENQAAD 662
PMS1_YEAST DEADSIYAETIPVEINVRTPLKNSPSTISKDNYRSLSDGLTHRKFEDEI-EYNLSTKNFKEISNCKQMSSIISKRKSEAQENIIRKDELEDFEQGK 691

PMS1_HUMAN EIRKEISHTMFAEMETICQFNLCFIIT--ALNE--EIPVDQHATDEKYNFEMLOHIVLQGRLLAPQTLNITAVNEALLENIEIRKNGDFVIDEN 758
PMS1_YEAST VITLVSHNCKKMEVVGQFNLCFIIVTRVNDKSLFIVDOHNSDEKYNFETLQAVTFKSKLITIPVVELSVIDEIIVLDNUPVVEKNGHKLKIDBE 791

PMS1_HUMAN APVTEPAKILSLPTSNWITGPOVDLEIFMLSDSHVM---CHPSRVKQMFASRACKSVMTGALTSEYKKLITHMGEMGHFWNCPHGRPTMRHIAN 855
PMS1_YEAST EEFCSFVKILSLPTSHQTLBLCQFNELIHLIKEDGLRDNHICSKIRSFAMRACHSSMTICKPLNKKTVTRVVHNLSELDKFWNCPHGRPTMRHIME 891

PMS1_HUMAN L---GVTFQN 862
PMS1_YEAST IRDWSSEKDYEI 904

```

Figura 11

20/5/94

Sequência parcial de nucleótidos do ADNc de MH1 de rato. O putativo codão de paragem (TAA) está sublinhado

	1	10	20	30	40	50	60	70	80
1	TTCCGGCCAA	TGCTATCAAA	GAGATGATAG	AAAGCTGTT	AAAGCTGTT	AGATGCAAAA	TCTACAAATA	TTCAGTGGT	TGTTAAGGAA
81	GGTGGCCCTGA	AGCTAAATCA	GATCCAAGAC	ATGGCAGCTG	ATGGCAGCTG	GAATCAGGAA	GGAGATCTCG	GATATGCTG	GTGAGAGGTT
161	CACATACAGT	AAACTGCAGA	CTTTTGAGAC	TTTAGCCACTG	TTTAGCCACTG	ATTCTTACCT	ATGGCTTTTC	TGGTGAGCAT	TTGCCAAGCA
241	TAACTCATGT	GGCCCATGTC	ACTATTACNA	CCAAACACGC	CCAAACACGC	TGATGGGAAA	TGTGCGTACA	GAGCAAGTTA	CTCAGATGGA
321	AAGCTGCAAG	CCCTCTCTAA	ACCTGTGCA	CCAAACACGC	CCAAACACGC	GCACCCCTGAT	CAGGTTGGA	GACCTTTTTT	ACAACATAAT
401	CACAAGGAGG	AAAGCTTTAA	ACCTGTGCA	TGAAGAGTAC	TGAAGAGTAC	GGAAAAATTT	TGGAAGTTGT	TGCCCAATGC	CACACACCTG
481	ATTCAGCAT	TAGTATCTCA	GTTAAAAAAC	AAAGTGAGAC	AAAGTGAGAC	AGTATCTGAT	TGCAGAACAC	TGCCCAATGC	CACACACCTG
561	GACAACATTC	GCTCCATCTT	TGGAATGCG	AAAGTGAGAC	AAAGTGAGAC	AAAGTGAGAC	AGTATGCTG	TGCCCAATGC	CACACACCTG
641	CAAAATGAAT	GGCTATATAT	CGAATGCAAA	GTTATTCAGT	GTTATTCAGT	AAAGTGAGAC	AGTATGCTG	TGCCCAATGC	CACACACCTG
721	TAGAATCAGC	TCCCTTGAGA	AAAGCCATTC	AAAGTGAGAC	AAAGTGAGAC	AAAGTGAGAC	AGTATGCTG	TGCCCAATGC	CACACACCTG
801	TCAGTTTGAA	ATCAGCCCTC	AGAACGTGAC	AAAGTGAGAC	AAAGTGAGAC	AAAGTGAGAC	AGTATGCTG	TGCCCAATGC	CACACACCTG
881	CATTCCTCAG	CGTGTGCAGC	AGCACATGGA	GAGCAAGCTG	GAGCAAGCTG	ATCCCTCCAG	ATTCCTCCAG	GATCTATTTT	ACCCAGGAGG
961	TGCTTCCAGG	ACTTGTCTGG	CCTCTGGGA	GAGCAAGCTG	GAGCAAGCTG	ATCCCTCCAG	ATTCCTCCAG	GATCTATTTT	ACCCAGGAGG
1041	CGGACAAAGT	CTACGCTTAC	CAGATGTCG	GTACGGACTC	GTACGGACTC	CGGGATCAG	AGCTTGAGC	CCTTCTGCA	GCCTGTAACC
1121	AGCCTTGTGC	CCAGCCAGCC	CCAGGACCTT	CGCCCTGTCC	CGCCCTGTCC	GAGGGGCGAG	GACAGAGGGC	CCTTCTGCA	GCCTGTAACC
1201	CGAGGATGAG	GAGATGCTTG	CTCTCCAGC	CGCCCTGTCC	CGCCCTGTCC	GAGGGGCGAG	GACAGAGGGC	CCTTCTGCA	GCCTGTAACC
1281	TGGAGACTTC	AGACGAGCC	CAGAAAGCGG	CACCCACTTC	CACCCACTTC	CAGTCCAGGA	AGCTCCAGAA	AGAGTCTATC	GGGCACTTAC
1361	GATGTGGAAA	TGGTGGAAAA	TGCTTCCGGG	AAGGAAATGA	AAGGAAATGA	CAGCTGCTTG	CTACCCCAAG	AGAGTCTATC	GGGCACTTAC
1441	CAGCGTCTTG	AGTCTCCAGG	AAGAGATTAG	TGAGCGGTGC	TGAGCGGTGC	CATGAGACTC	TCCGGGAGAT	AGTCTCTAAC	CATTCTTTTG
1521	TGGGCTGTGT	GAATCCTCAG	TGGGCTTGG	CACAGCACCA	CACAGCACCA	GACCAAGCTA	TACCTCCCTCA	ACACTACCAA	GCTCAGTGAA
1601	GAGCTGTTCT	ACCAGATACT	CATTTATGAT	TTTGCCAACT	TTTGCCAACT	TTGGTGTCTT	GAGGTTATCG	GAACCAAGCG	CACCTCTTCA
1681	GTCGAGTTTC	TGAAGAGAGG	CAGTCTCGAA	AGTGGCTGGA	AGTGGCTGGA	CAGAGGACGA	CGGCCCAAG	AAGGGCTTGC	AGAGTACATT
1761	TTCTGATGAC	AGCTATGTGC	CACCTTTGGA	GGGACTGCCT	GGGACTGCCT	CTCTGTGAGA	TGGATGAGAA	GGGAACCTGA	TTGATTAATC
1841	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA
1921	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA
2001	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA
2081	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA
2161	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA
2241	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA
2321	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA
2401	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA
2481	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA
2561	AGAAAGGAGG	TGTTTGTGAA	GTCTCAGTAA	AGAATGTGCT	AGAATGTGCT	ATCTTCTATC	TTCGACTGCG	CAGTCTATTA	AATTGGGTGA

SEQ. ID NO: 135

Figura 12

(Nota: as secções de leitura são colocadas de modo a incluir as que têm uma forte

```

de ratol .....PANAIKEMIENCLDAKSTNIQVWKEGGLKLIQDNGTGIRKEDLDIVCERFTTSKLOTFEDLASISTYGR 73 136
|||||
humana 1 MSFVAGVIRRLDETWNRIAGEVIGRANAIKEMIENCLDAKSTSIQVIVKEGGLKLIQDNGTGIRKEDLDIVCERFTTSKLOTFEDLASISTYGR 100

```

[illegible]
$$\frac{22}{24}$$

Figura 13

Sequência de nucleótidos de PMS1 de rato.
Os códons do suposto início (ATG) e da
suposta paragem (TGA) estão sublinhados

SEQ. ID NO: 137

Figura 14

	1	10	20	30	40	50	60	70	80
1	CGTGAAGGT	CCTGANGAT	TTCCAGATTC	CTGAGTATCA	TTGGAGGACA	CAGATAAACC	GTGTCAGGT	AACGATGGTG	80
81	TATATGCAAC	AGAAATGGGT	GTTCCTGGAG	ACGGCTCTTT	TCGGAGAGGC	GGCACCCCAA	CTCTCCCGCG	GTGACTGTGA	160
161	TTTGACGAGT	CCTGCATCCG	TGGAGCAAAC	CGAAGGCGTG	AGTACAGAA	GTGCTAAGGC	CATCAAGCCT	ATTGATGGGA	240
241	AGTCAGTCCA	TCAAATTTGT	TCTGGGCAGG	TGATACTCAG	TTTAAGCACC	GCTGTGAAGG	AGTTGATAGA	AAATAGTGTG	320
321	GATCCTCGTG	CTACTACTAT	TGATCTAAGG	CTTAAAGACT	ATGGGGTGA	CCTCATTTGAA	GTTCAGACA	ATGCATGTGG	400
401	GCTAGAAGAA	GAAACTTTG	AAGTCTAGC	TCTGAACAT	CACNCATCTA	AGATTCAAGA	GTTTGCCGAC	CTCAGCGAGG	480
481	TTGAAACTTT	CGGCTTTGG	GGGGAAGCTC	TGAGCTCTCT	GTGTGCACTA	AGTGAATGTA	CTATATCTAC	GTGCCACGGG	560
561	TCTGCAGGCG	TTGGGACTCG	ACTGGTGTCT	GACCAATAG	GGAAATCAC	CCAGAAACT	CCCTACCCCG	GACCTAAAGG	640
641	AACCAAGCTC	AGTGTGAGC	ACTTATTTTA	TACACTACCC	GTGCGTTACA	AAGACTTTCA	GAGCAACATT	AAAAAGCACT	720
721	ATTCCAAAT	GCTGCAGGTC	TTACAGGCGT	ACTGTATCAT	CTCAGCAGGC	GTCCGTGTAA	GCTGCACATA	TGAGCTCGGA	800
801	CAGGGGAGC	GGCACCGCT	GCTGTGCACA	AGCGGCACGT	CTGGCATGAA	GGAAATATC	GGGTCTGTGT	TTGCCAGAA	880
881	GCAGTTGCAA	AGCCTCATTC	CTTTGTGTCA	GCTGGCCCTT	AGTGACGCTG	TGTTGTGAAGA	GTACGGGCTG	AGCACTTCAG	960
961	GACGCCACAA	AACCTTTTCT	AGCTTTTCGG	GCTTCATTTT	ACAGTGCACG	CAGCGCGCGG	GGAGGAGTGC	AACAGCAGG	1040
1041	CAGTTTTTCT	TGATCAATCA	GAGGCCCTGT	GACCCAGCAA	AGTCTCTTAA	GCTTGTCAAT	GAGGTTTTATC	ACATGTATTA	1120
1121	CGGCGATCAG	TACCATTTTG	TGCTCTTAA	CGTTTCCGTT	GACTCAGAA	GTGTGATAT	TAATGTAACT	CCAGATAAAA	1200
1201	CGCAATTTCT	ACTACAGAA	GAGAAGCTAT	TGCTGGCCGT	TTTAAAGACC	TGCTTGTATG	GAATGTTTGA	CAGTGATGCA	1280
1281	AGCAAGCTTA	ATGTCAACCA	GCAGCCACTG	CTAGATGTTG	AAAGTAACCT	AGTAAATCG	CATACCTGCG	AACTAGAAAA	1360
1361	GGCTGTGACA	CGAAGCAAG	ATAACTCTCC	TTCACTGAAG	AGCACAGCAG	ACGAGAAAAG	GGTAGCATCC	ATCTCCAGCG	1440
1441	TGAGTGAGCG	CTTTTCTCTT	CATCTACTA	AAGAGTCAA	GTCTAGGGGT	CCAGAGACTG	CTGAATGAC	ACGGAGTTTT	1520
1521	CTAAGTGAAG	AAGGGGGCGT	GTATCTCTCT	TATCCTTCAG	AGTCTATCTC	TTACAGAGGC	CTCCGTGGCT	CCGACGACAA	1600
1601	ATTGCTGAGT	CCACCGACCA	GGCTGGTGA	CTGTATGAC	AGAGAGAAAA	TAGAAAAAGA	CTCAGGGCTC	AGCAGCACTT	1680
1681	CAGCTGGCTC	TGAGGAGAG	TTACGACCC	CAGAAATGGC	CAGTAGCTTT	AGCAGTGACT	ATAAGCTGAG	CTCCCTAGAA	1760
1761	GAAGAGACCT	CTCAGGAAC	CATAAATCT	GCTGACCTGC	TGCGGTCTCT	CAGGTACAGG	ACAGTCTTGT	AAGCCAGAA	1840
1841	ACCATGATA	TCATGCAAA	GCTTACCTC	TAGCTGCTCT	GTCAACCA	AATGCCAAGC	GCTTCAAGAC	AGAGGAAGAC	1920
1921	CCTCAATGT	CACATATCT	CAAGATTGC	CTGGTCTCTA	GAGCACTCA	GCAGCTGAGG	TGATGTGAGC	CATAAAATG	2000
2001	ANTAGAGAT	CGTCTCTCT	GAGTCTCTTA	GCTAAGCGAA	TGAAGCAGTT	ACAGCACTTA	AAGGCCGAGA	ACAAACATGA	2080
2081	ACTGAGTTAC	AGAAATTTA	GGGCCAAGAT	TTGCCCTGGA	GAAACCAAG	CAGCGAAGA	TGAATCTAGA	AAAGAGATTA	2160
2161	GTAATCGAT	GTTCGAGAG	ATGAGATCT	TGGGTCACTT	TAACCTGGGA	TTTATAGTAA	CCAACTGAA	AGAGGACCTC	2240
2241	TTGCTGGTGG	ACCAGCATGC	TGCGGATGAG	AGGTACAACT	TTGAGATGCT	GCAGCAGCAC	ACGGTGTCTC	AGGCGCAGAG	2320
2321	GCTCATCAG	TGGGTGCACA	CAGGCTTCAG	AGTTCCGAGA	CCCCAGACTC	TGAATCTAAC	TGCTGTCAAT	GAAGCTGTAC	2400
2401	TGATAGAAAA	TCTGGAATA	TTAGAAAAA	ATGGCTTTGA	CTTTGTCAAT	GATGAGGATC	CTCCAGTCAC	TGAAGGGCT	2480
2481	AAMTGTATT	CCTTACCAAC	TAGTAAAAAC	TGGAGCTTTG	GACCCCAAGA	TATAGATGAA	CTGATCTTTA	TGTTAAGTGA	2560
2561	CAGCCCTGGG	GTGATGTGCC	GGCCCTCAGC	AGTCAGACAG	ATGTTTGGCT	CCAGAGCCTG	TGGAAGTCA	GTGATGATTG	2640
2641	GAGGGGCGCT	CAATGCGAGC	GAGATGAAGA	AGCTCATCAC	CCACATGGGT	GAGATGGACC	ACCCCTGGAA	CTGCCCCAC	2720
2721	GGGAGGCCAA	CCATGAGGCA	CGTGTCCAAT	CTGATGTCA	TCTCTCAGAA	CTGACACACC	CCTTGTAGCA	TAGAGTTTAT	2800
2801	TACAGATGCT	TGGTTTCCCA	AAGAGAGGTT	TTAAGTAAAT	CTGATTATCG	TTGTACAAAA	ATTAGCATGC	TGCTTTAATG	2880
2881	TACTGATGCC	ATTAAAGAG	AGTGTAAAGG	CAGGCTGAT	GGAGTGTGCC	TCTAGCTCAG	CTACTTGGGT	GATCCGGTGG	2960
2961	GAGCTCATGT	GAGCCGAGCA	CTTTGAGACC	ACTCCGAGCC	ACATTCATGA	GACTCAATTC	AAGGACAAAA	AAAAAAGAT	3040
3041	ATTTTTGAAG	CCTTTTAAAA	AAAAA						3065

23/24

24/24

Comparação das sequências de aminoácidos previstas para as proteínas rPMS1 e hPMS1. As linhas verticais indicam identidades de aminoácidos

de rato 1 NEOTEGVSTECARAIKPIDGKSVHQICSGOVLSTAVKELIENSVDAGATTIDRLKDYGVDLIEVSDNGCGVEENFEGLAKHHTSKIQEFADLTQ 100 138
humana 1 HIRAESSTTEPAKAIKPIDRKSVMHQICSGQVLSLSTAVKELVENSIDAGATNIDKLKDYGVDLIEVSDNGCGVEENFEGLTLKHHTSKIQEFADLTQ 100
101 VETFGFRGEALSSLCALSDVTISTCHGSASVGRVLRVFDHNGKITQKTPYPRPKGTTVSQHLFYTLTPVRYKEFORNIKEYSKMVQLQAYCIIISAGVRV 200
101 VETFGFRGEALSSLCALSDVTISTCHASAKVGTRLMFHDHNGKIIQKTPYPRPTGTTVSQQLFSTLTPVRHKEFORNIKEYAKMVQLHAYCIIISAGIRV 200
201 SCTNQLGQGRHRAVVCVTSCTGSMKENIGSVFGQKQLQSLIPFVQLPSPDAVCEYGLSTSGRHKTFTSFSGFISQCTHAGAGRSATDRQFFFINQRPCDPA 300
201 SCTNQLGQGRKQPVVCTGGSPSIKENIGSVFGQKQLQSLIPFVQLPSPDSVCEYGLSCSDALHNLFYISQFISQCTHGVGRSSTDQFFFINRRPCDPA 300
301 KVKSLVNEVYIMYNRRHQYPFVVLNVSDSECDVINVTEDKQRIILQSEKLLAVLKTSLIGHFSDANKLVNVNQPLLDVEGNLVKSHTAELKVPVPGKQ 400
301 KVKRLVNEVYIMYNRRHQYPFVVLNISVDSECDVINVTEDKQRIILQSEKLLAVLKTSLIGHFSDVNVKLVNSQQPLLDVEGNLVKHHADLEKPMVEKQ 400
401 DMSPLSKSTADEKNVASISRLREAFSLHPTKEIKSRGPETAELTRSPSEKRGVLSSYPSDVISYRGLRGSQDKLVSETDSPGDCMDREKIEKDSGLST 500
401 EGPFSRL-TGEEKDVVISRLREAFSLRHTTENKPHSPKTPPEPRRPLGQKRGHLSSTSGAISDKGVLRSQKEAVSSSHGSPSDFTDRAEVEKDSGHGST 499
501 SAGSEEEFSTPEVASSFSDDYVNVSSLEDPSQETINC-----GDLLPSSRYATVLEARREWISMQSSSTSSSVTHKCOALQDRGRPSNVNISORLPGP 593
500 SVDS-EGFSIPDTGSHCSSEYAASSPGDRGSQEHVDSOEKAPETDPSFSDVDCHSNQEDTGCKFRVLPQPTNLA-TPNTRFKKEEILSSSDICQKLVT 597
594 QSTSAAEVDVAIKNKR-----SCSSSSLAKRMKQLQHLKAQNKHLSYRKFRAKICPGENQAAEDELARKEISKSMFAEHEILGQFNLGFIIVTKLKEDLFL 689
598 QDMSASQVDVAVKINKKVPLDFSMSSLAKRIKQLHHEAQQSEGEQYRKFRAKICPGENQAAEDELARKEISKTHFAEHEIIGQFNLGFIITKLNEIFI 697
690 VQQAHADEKYNFEMLQHTVLAQORLITWVHTGFRVPRPQTNLNTAVNEAVLIENLEIFRKNQDFVIDEAPVTERAKLISLPTSKNWTFGPDIDELI 789
698 VQHAATDEKYNFEMLQHTVLAQORLITWVHTGFRVPRPQTNLNTAVNEAVLIENLEIFRKNQDFVIDEAPVTERAKLISLPTSKNWTFGPDIDELI 787
790 FNLSDSPGVNCRPSRVQMFASRACRKSVMIGTALNASMKKLITHGEMDHPWNCPHGRPTMRHVANLDVISON 864
788 FNLSDSPGVNCRPSRVQMFASRACRKSVMIGTALNTSEMKKLITHNGEMGHFWNCPHGRPTMRHIANLGVISON 862

Figura 15